Understanding *Clostridium difficile* Infection Outcomes, through Host Clinical Variables, and Bacterial Whole Genome and Phenotypic Analysis

A thesis submitted by Emma Butt to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Sciences
In February 2013

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University

Abstract

Clostridium difficile is a clinically problematic pathogen and continues to persist within the healthcare system. Presentation of disease symptoms ranges from mild to severe diarrhoea, through to fulminant pseudomembranous colitis. Approximately 20% of patients will suffer from recurrent episodes and of all patients who die from *C. difficile* related causes, approximately 41% of death certificates mention *C. difficile* as an underlying cause of death, and this poses a significant burden on healthcare facilities.

Three methods of investigation were employed to develop a more comprehensive understanding of both the host and isolate association with the outcomes of *C. difficile* infection; mortality and recurrence. These methods were; analysing patient clinical data to try and identify host markers of infection outcomes, evaluating *C. difficile* type association with infection outcomes, and genetically and phenotypically characterising the clinically relevant *C. difficile* isolates associated to these outcomes.

During this study statistical analysis of clinical data revealed that there were four variables; white cell count, serum albumin, C-reactive protein and respiratory rate, which were prognostic of mortality in patients with *C. difficile* infection. Threshold levels of these variables were used to create a clinical prediction rule to classify patients with *C. difficile* infection who were more 'at risk' from mortality, with statistical significance in both a derivation and validation cohort. However, analysis was unable to determine variables prognostic of recurrent infection.

Due to small sample sizes of some groups of isolates, no groups of *C. difficile* isolates were significantly associated with increased recurrent infection or mortality during this study. Some groups of isolates were associated with primary only infection and/or low mortality. There was a non-significant trend in particular *C. difficile* isolate groups being associated with infection outcomes; a panel of representative isolates was therefore chosen to be characterised in more detail.

Phenotypic and genetic analysis of a panel of sixteen *C. difficile* isolates revealed isolate specific differences in toxin production, conservation of transposable elements and SNP abundances, which may have played a role in infection outcomes. Isolate motility and antibiotic resistance profiles were not statistically significantly different between isolates within a particular group of *C. difficile* types.

One hypothesis from the collective results obtained during this study suggests that the phenotypic and genotypic changes in isolates may have facilitated differences in their interaction with the host. In turn, the host specific inflammatory response to the infecting *C. difficile* isolate may have played a role in host outcomes.

Research conducted during this study has begun to assess which host specific responses may be important in determining the outcome of *C. difficile* infection, and which *C. difficile* isolate characteristics may in part also contribute. However, the assessment of both host and isolate association to infection outcomes would benefit from further investigation in a larger cohort, in order to prove or refute conclusively any hypothesises generated in this study.

Acknowledgements

I would first and foremost like to dedicate this PhD thesis to my granddad, Alec (John) Ireland; he never got to see me finish the research on which this PhD thesis is based. As a great chemist, his enthusiasm for science was bestowed upon me, and without which, I am sure I would not have pursued a career in science with such passion as I now do.

You are always in my thoughts granddad.....

To all those who made this PhD possible I wish to sincerely thank you. I am sincerely grateful to my principle supervisor Dr Steve Michell, your continued support has been invaluable to me. It has allowed me to develop as scientist and as a result of this, have found a research area which I now wish to pursue to a higher level.

Thanks to my second supervisor Dr Ed Keedwell, who has provided much needed statistical support and whose patience was greatly appreciated.

To everyone on the 4th floor of Geoffrey Pope, I am extremely grateful for your help and guidance in the lab and for making a great working environment at the University of Exeter. I have thoroughly enjoyed working with you all.

Finally, I would like to thank my parents.... You let me move back home to do this PhD and for that I am eternally grateful. I have loved every minute of living back home with you! My fiancée Jamie..... You are the light when the moments are dark, you have been patient, kind and have accepted all that has come with this PhD. I love you so much and am very lucky to have you in my life.

Table of Contents

Content **Page Number**

Page Number

Page Number

Page Number

Page Number

Page Number

Page Number

Page Number

List of Figures and Tables **Page Number**

Figures

Chapter One

Chapter Two

Chapter Three

Chapter Four

Page Number

Chapter Five

Chapter Six

Page Number

Chapter Seven

Chapter Eight

Page Number

Page Number

Page Number

Tables **Page Number**

Chapter One

Chapter Two

Chapter Three

Chapter Four

Page Number

Page Number

Chapter Nine

Concluding Discussion

Chapter Ten **Page Number**

Appendices

List of Manuscripts Submitted/Prepared as a Result of Research Conducted during this PhD

Submitted Manuscripts

Derivation and Validation of a Simple, Accurate and Robust Prediction Rule for Risk of Mortality in Patients with *Clostridium difficile* Infection. Emma Butt[1], Jane AH Foster[2], Edward Keedwell[3], Julia EA Bell[4], Richard W Titball[1], Aneel Bhangu[5], Stephen L Michell[1] and Ray Sheridan[6]. *BMC Infectious diseases. Under review*

***Clostridium difficile*: Evaluation of disease severity and the potential role of Faecal Calprotectin as a biomarker. Have the UK guidelines got it right?** Dr Jane AH Foster[1], Emma Butt[2], Dr Edward Keedwell[3], Dr Stephen L Michell[2], Dr Tariq Ahmed[1], Dr Ray Sheridan[1]. *In preparation.*

Characterisation of Clinically Relevant *Clostridium difficile* Isolates from different Infection Outcomes. [1]Emma Butt, [2]Ed Keedwell, and [1]Stephen .L. Michell. *Journal of Medical Microbiology. In preparation.*

Abbreviations

ADP	Adenine tri phosphate
H_1	Alternative hypothesis
AUC	Area under the curve
bp	Base pairs
bpm	Beats per minute
BHI	Brain heart infusion
mcL	Cells/micro litre
χ^2	Chi-Square value
CDI	*Clostridium difficile* infection
CDS	Coding Sequences
Cfu/s	Colony forming units
df	Degrees of freedom
DNA	Deoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
g	Grams
x g	Gravitational force
Galß1-4GlcNAc	N-Acetyllactosamine
GDH	Glutamate Dehydrogenase
GDP	Guanosine 5'-diphosphate
GTP	Guanosine 5'-Triphosphate
HPA	Health Protection Agency
HCAI	Healthcare acquired infection
hr/hrs	hour/hours
IL	Interleukin
Kbp	Kilo base pairs
kg	Kilo gram
L	Litres
MAP	Mitogen Activated Protein
µl	Micro litres
µM	Micro molar
mg	Milli grams
ml	Milli litres
mm	Milli metre
mM	Milli molar
MIC	Minimum inhibitory concentration
Mins	Minutes
MLST	Multi locus sequence typing
mAb/s	Monoclonal Antibody/ies
Nf-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
ng	Nano grams
nm	Nano metre
NGS	Next generation sequencing

NS	Non-synonymous
H_0	Null hypothesis
PaLoc	Pathogenicity locus
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
ROC	Receiver operating curves
REA	Restriction enzyme analysis
Rpm	Revolutions per minute
SEM	Standard error of the mean
S	Synonymous
Tcdx	*Clostridium difficile* Toxin (x)= A, B, C, R, or E
Tn	Transposon
TE	Tris- EDTA
TAE	Tris-acetate EDTA
TBS	Tris-buffered saline
TBST	Tris-buffered saline
TY	Tryptone yeast
TNF	Tumour necrosis factor
Yrs	Years

Chapter One. General Introduction

Introduction

This introductory Chapter will provide an overview of *Clostridium difficile* and its emergence as a clinically problematic pathogen. It will highlight the challenges imposed by this pathogenic bacterium at both at a molecular and clinical level, and describe how the complex interaction of *Clostridium difficile* and the host is developing research methodologies. Finally, it will outline the aim and structure of this thesis.

1.1 *Clostridium difficile*

Clostridium difficile (*C. difficile*) is a Gram positive, anaerobic, spore forming, rod shaped bacterium, approximately 0.3-2 x 1.5-2 μm in size. It falls into the following taxonomic classification; Kingdom: Bacteria, Division: Firmicutes, Class: Clostridia, Genus: *Clostridium*, Species: *difficile*[183].

C. difficile belongs to the low G+C (≤ 50%) Gram positive group of bacteria that generally have the following characteristics; shape: rods, cocci or irregular, and ability to form endospores. *C. difficile* endospore formation is key to its transmission from contaminated surfaces in hospitals to patients. Sporulation usually commences in response to environmental stresses such as nutrient limitation[183]. Spores result from asymmetric cell division; have a thick coat composed of several protein layers and a cortex made of peptidoglycan. Upon germination, the spores lose these external layers and resume vegetative growth.

1.2 History of *C. difficile* and Clinical Disease Presentation

In the 1960s, anaerobes were widely recognised as the predominant organisms in the large bowel, and by the early 1970s, pharmaceutical companies had deduced that a conventional treatment for bowel bacteria i.e. kanamycin[155], was not effective, and began using clindamycin as an alternative treatment. Observations of diarrhoea were reported for patients

undergoing clindamycin treatment and, in a number of cases this progressed to severe inflammation of the colonic mucosa, also known as pseudomembranous colitis (PMC)[150].

PMC had previously been described by Finney[70], and in the 1970s an interest in the disease was sparked by the correlation of its incidence and increased use of clindamycin[150].

In 1935, *C. difficile* was characterised by Hall and O'Toole[89], as being part of the normal microflora of neonates. However, it was not until the 1970s, that an association between *C. difficile*, disease, and more specifically PMC was recognised. The association of *C. difficile* with PMC was established between 1975-1978, and was proposed via three convergent lines of study by; Tedesco *et al.*[211], Green[86] and Hafiz[88]. Hafiz's PhD thesis on *C. difficile*[88] was the most comprehensive report on *C. difficile* at this time[25]. Work to establish a connection between *C. difficile* and disease, proceeded rapidly between 1975 and 1980[25], and it was Larson *et al.*[135] who identified that the cause of PMC might be due to a bacterial toxin. Studies by Rifkin *et al.*[189] confirmed the observations of Larson *et al.*[135], but also found that the toxin which was present in the stools of patients with antibiotic-associated colitis, was neutralised by antitoxin from *Clostridium sordellii* (*C. sordellii*), indicating that this toxin might be closely related to other toxins from the Clostridia genus. They found that the effect of toxin on animal models and on tissue culture cells, was identical to those observed by Hambre *et al.*[90] during World War II and Green's work in guinea pig and hamster models of penicillin induced colitis[86]. The work by Hambre *et al.*[90] also highlighted the importance of treatment with oral vancomycin, as it was provided as a therapy to patients with antibiotic-associated colitis, and correlated with loss of toxin from the stool samples[189].

By 1979 Bartlett *et al.*[24] had evidence to suggest that *C. difficile* was the causative agent of PMC. A prospective study on 189 patients was performed; using a cytotoxicity assay, to deduce if the toxin found in stools could be neutralised by *C. sordellii* antitoxin. This study identified that 26 out of 27 patients with PMC and 9 out of 63 patients with antibiotic-

associated diarrhoea (but no documented PMC) were toxin positive, but healthy controls, patients with ulcerative colitis (UC) and neonatal necrotizing enterocolitis were toxin negative. Toxin neutralisation was observed with *C. sordellii* antitoxin, but *C. difficile* was the only species of Clostridia to be isolated from PMC patients. *C. difficile* was isolated from 6 of the 8 specimens in a culture positive assay, but not from healthy patients. These results, along with previous studies by Bartlett *et al.* in 1977[27] led to the conclusion that *C. difficile* was the major cause of PMC.

Researchers now knew *C. difficile* to be the main cause of infectious diarrhoea that develops after hospitalisation and treatment with antibiotics[193] and, as such, investigation began to deduce the molecular mechanisms of pathogenicity for this bacteria.

1.3 Pathogenicity of *C. difficile*

Antibiotic use in hospitals is major risk factor for development of CDI. Disruption of the normal gut microbiota by antibiotic therapy allows *C. difficile* to colonise the colon and cause disease[193]. CDI is the result of *C. difficile* spore germination, colonisation of the gut[55], and production of two highly potent toxins; toxin A (TcdA) and toxin B (TcdB). The notoriously effective transmission of *C. difficile* is associated with excretion of spores in the faeces of infected patients[136].

TcdA and TcdB are transcribed from a pathogenicity locus (PaLoc) and are now accepted as being the major cause of disease symptoms[193]. Work in the late 1980s and early 1990s had identified the DNA sequences of *tcdA* and *tcdB* [23, 220], and it was in 1995 that Hammond and Johnson characterised the PaLoc from *C. difficile* VPI 10463, which consists of a 19.6 Kbp region and encodes five genes; *tcdR, tcdB, tcdE, tcdA* and *tcdC* (Figure 1.1)[91]. TcdA and TcdB are the primary factors causing symptoms of CDI however, the accessory proteins, TcdR, TcdE and TcdC are all required for *C. difficile* pathogenicity, as the PaLoc is a highly

conserved element in all toxigenic strains of *C. difficile* and is replaced by a 115 bp sequence in non-toxigenic strains[91].

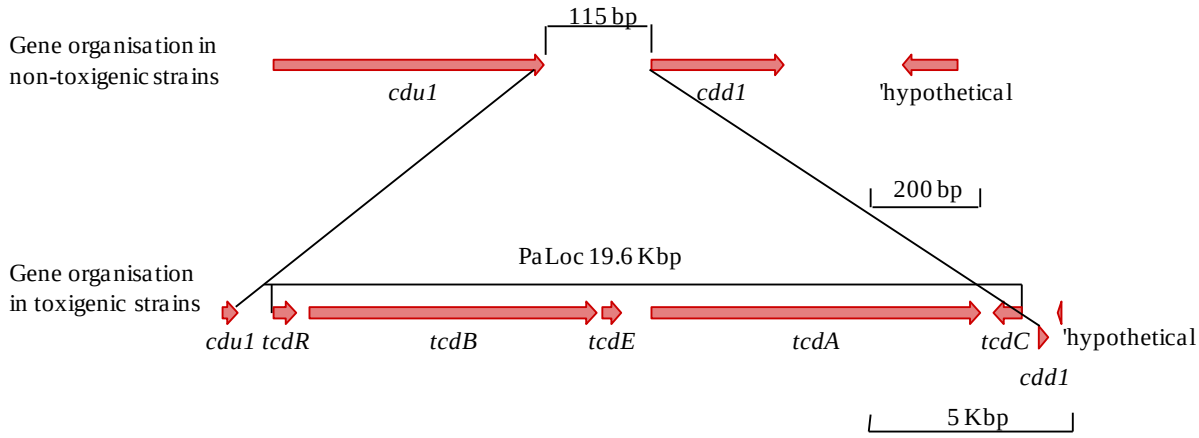


Figure 1.1- Genetic map of the 19.6 Kbp PaLoc of *C. difficile*, showing the deletion of the PaLoc in non-toxigenic strains.

Some strains of *C. difficile* produce a binary toxin (CDT). The toxigenic effect of CDT has been observed in hamster and rabbit models of infection. *C. difficile* $TcdA^{-}TcdB^{-}CDT^{+}$ strains are able to colonise the gut of hamsters following clindamycin treatment, but do not induce diarrhoea or death[79]. However, upon inoculation into rabbit ileal loops, there is increased non-haemorrhagic fluid accumulation. Both these observations suggest that CDT may have enterotoxic properties and may serve as a concomitant in pathogenesis in $TcdA^{+}TcdB^{+}$ strains. The evidence for clinical significance of CDT in disease outcomes will be discussed further in Chapter Seven. CDT is a dimeric ADP-ribosyltransferase encoded by two genes *cdtA* and *ctdB*[207] which show high sequence identity to *Clostridium perfringens* Iota toxin[180].

1.4 TcdA and TcdB: Sequence Structure and Cell Binding Properties

TcdA and TcdB belong to a class of toxins named the large clostridial cytotoxins (LCTs), due to their high molecular weight, cytopathic action on cells in culture, and their glucosyltransferase activity[221]. Sequence analysis and protein predictions of LCTs from

Clostridial species has enabled a tripartite structure to be deduced[107], consisting of an N-terminal catalytic domain, a hydrophobic translocation domain and a C-terminal ligand binding domain (Figure 1.2).

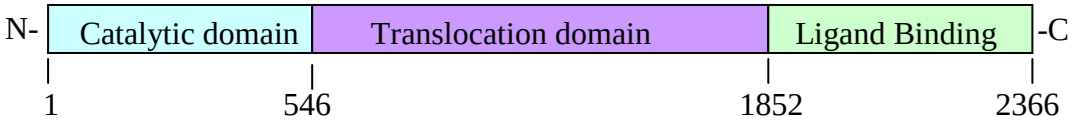


Figure 1.2- Representative diagram of the tripartite amino acid structure of LCTs.

TcdA and *tcdB* show 49% identity at the DNA sequence level, 63% homology at the amino acid level, and 14% of the DNA sequence shows conserved substitutions[117]. Both toxins have conserved C and N-terminal domains and while experimental evidence by some[184, 222] show that these domains are functionally the same, TcdB is more heterogeneous in structure[184]. Sequence alignments of LCTs show that there is a common DXD motif conserved within the N-terminal domain of all Clostridial toxins, and appears to be essential to glucosyltransferase activity[41].

The C-terminal domain of LCTs contain combined repetitive oligopeptides (CROPs)[221]. The CROPs are involved in ligand binding to cell surface receptors: TcdA shows specificity for carbohydrate antigens X, I and Y, which have a common Galß1-4GlcNAc core[215]. As yet there is no data to deduce the receptor type for TcdB, only that is has an affinity for the basolateral side of host cells[193] and it is highly probable that the CROPs are involved in this targeting[222]. TcdA and TcdB enter host cells via receptor mediated endocytosis[172] and from there, the N-terminal glucosyltransferase domains are translocated into the cytosol[177] via an autocatalytic process[188] .

TcdA and TcdB are monoglucosyltransferases which catalyse the incorporation of glucose into threonine 37 of Rho A, via the cosubstrate UDP-glucose[117,118,116]. The modification of Rho A results in multiple downstream effects, including altering epithelial cell structure, which in turn evokes an inflammatory response in the host, and facilitates clinical

presentation of disease, such as diarrhoea[123]. Further discussion of the role of toxin expression and gene regulation in disease outcomes will be discussed in Part Three, Chapter Seven.

1.5 Genetic Analysis of *C. difficile* to Understand Pathogenicity

Mechanisms to investigate genes that are linked to the virulence of bacteria are now common practice in molecular microbiology laboratories. They involve targeted or non-specific gene knock outs, or insertional gene knock downs, in a manner that has an origin in the 19th century form of Koch's postulate[69].

Pre 2007, gene knock out systems for *C. difficile* proved unstable[169]. In 2007, Heap *et al*.[95] developed ClosTron: a universal gene knock-out system for *Clostridium* which uses a bacterial group II intron containing an antibiotic resistance gene, which is in turn interrupted by a self-splicing group I intron. This technology has since led to the construction of several *C. difficile* mutants; in genes encoding sporulation initiation factors[217], TcdA and TcdB[130], and has helped to confirm the importance of both toxins in disease progression. Clearly, mutant construction will facilitate a better understanding of these key virulence determinants. Nevertheless, there is still a greater depth of genetic and evolutionary understanding that could perhaps only come from sequencing whole genomes of *C. difficile* isolates.

1.6 Diagnosis of *C. difficile* Infection

Several tests exist for the diagnosis of CDI, and there are numerous studies that have evaluated their robustness[227]. There is also a huge effort to deduce significant risk factors associated with the onset of CDI, mortality, severity, and recurrent infection, in order that more effective preventative methods/treatments can be defined. If the data obtained from a truly robust and accurate diagnostic test, could be combined with predictors of CDI,

mortality, severity and recurrence, then perhaps this would aid in the control of CDI in the clinical setting, reduce mortality rates and therefore reduce the burden caused by *C. difficile*.

Around the globe there is much variation of the type of tests utilised to diagnose CDI, due to the cost and ease by which tests can be performed. Until recently the 'Gold Standard' tests for *C. difficile* detection were either the Cytotoxin Neutralisation (CTN) assay[26]; which detects the toxins produced by *C. difficile* in the supernatants of patient faeces, using both antitoxin-protected and nonprotected tissue culture monolayers or the cytotoxigenic culture test, which utilises the application of culture supernatents (as opposed to faecal supernatents) to antitoxin-protected and nonprotected tissue culture monolayers[65]. Both tests have a very high positive predictive value (PPV), frequently ≥ 90%, but are very labour intensive and time consuming[65], and as of 2008, only 1% of labs were using this test, due to the their high complexity and labour intensive processes[196]. A table adapted from a review by Bartlett[26] summarising other diagnostic methods is shown below (Table 1.1). All these diagnostic tests vary in sensitivity and specificity, and currently there is no uniformity between laboratories regarding which tests should be preferentially used.

Table 1.1- List of diagnostic testing methods for CDI [26].

Test	Detection	Time required	Comment
Cytotoxin	Toxin B	24–48 hrs	Relatively sensitive and specific, but delayed results and technically complex
Toxin-culture	Toxigenic *C. difficile*	3–5 days	Sensitive, but delayed results and technically complex
EIA-toxin A or A/B	Toxin A or A/B	Hours	Rapid, inexpensive, specific but relatively insensitive
EIA-GDH	*C. difficile*	Hours	Rapid, inexpensive, relatively sensitive, but not specific, especially in nosocomial cases
EIA-GDH + EIA toxin A/B	*C. difficile* & toxins	Hours	Rapid, inexpensive, need alternative toxin test for pos/neg result
RT-PCR toxin B gene	Toxigenic *C. difficile*	Hours	Rapid, very sensitive, need clinical correlations for positive tests

1.7 *C. difficile* Strain Typing

Once a diagnosis of CDI has been confirmed there may be a requirement to type the infecting strain, especially if an increase in the number of infections and/or increased severity, mortality and recurrence outcomes have been observed[229]. Typing of *C. difficile* strains has increased since 2007 due to an epidemic caused by a 'hypervirulent' BI/NAPI/027 strain, that has been associated with increased mortality, severity and recurrent infection[73].

Typing techniques vary in ease, cost and user time, and the most widely utilised UK *C. difficile* typing assay is PCR ribotyping. This typing technique is based on PCR amplification of the multiple alleles; ranging from 250-600 bp, that exist between the 16S-23S ribosome intergenic spacer regions[55]. More than four hundred different types have been identified in this manner[149], which was refined for use in routine clinical typing laboratories by O'Neill *et al.*[170]. Many other methods have been developed to type *C. difficile* and these are summarised in Table 1.2. Whilst PCR ribotyping is the most common method, it does not distinguish between sub-types of *C. difficile*, and this is also true of MLST techniques[139]. Dawson *et al.* (2009) make a poignant observation, remarking that our inability to distinguish *C. difficile* subtypes within a particular ribotype represents a current gap in our efforts to understand individual strains, and bacterial factors, in relation to disease severity and onset of CDI. They also remark that this inability to distinguish *C. difficile* subtypes indicates the need for deeper genetic analysis if subtle changes between strains that do and do not cause disease are to be fully understood[55]. The importance of ribotype surveillance in order to better understand *C. difficile* epidemiology, and ribotype association to infection outcomes will be discussed in Chapter Five.

Table 1.2– Summary of *C. difficile* typing techniques[55].

Technique	DNA target	Advantages	Disadvantages
MLVA (Multi-locus variable number of tandem repeats analysis)	Randomly distributed DNA repeat units in the genome	Able to sub-type strains unlike PCR ribotyping	Sometimes too discriminatory
AFLP (amplified fragment restriction polymorphism)	*Pst*I and *Mse*I restriction sites	Based on PCR amplification, therefore DNA degradation proves less of an issue (Klaassen *et al*.,[127]).	Recognises only primary types, uses fluorescent label, requires appropriate analysis equipment, does not distinguish sub-types
slpAST (Surface layer protein A gene Sequence typing)	Surface layer protein A gene	PCR amplification is straightforward and transferable between laboratories. Distinguishes sub-types	Utilises the sequence of only one locus
PCR ribotyping	16S–23S spacer region	Gold standard, recognises primary types, easily reproducible and portable between laboratories	Limited discriminatory potential, does not distinguish sub-types
REA (Restriction enzyme analysis)	*Hind*III restriction sites	Distinguishes sub-type strains unlike PCR ribotyping	Inter-laboratory data comparison is difficult, labour intensive
MLST (Multi locus sequence typing)	Housekeeping loci (usually 7)	Easily reproducible between laboratories	Limited discriminatory potential, time consuming, does not distinguish sub-types
PFGE (Pulse field gel electrophoresis)	*Sma*I restriction sites	Widely used molecular fingerprinting technique for typing of clinical isolates, distinguishes sub-types	Time consuming, inter-laboratory comparison may be difficult. Degradation of DNA can result in samples being un-typeable

1.8 Risk Factors for *C. difficile* Infection

CDI is associated with asymptomatic carriage through to clinical presentation of disease, with symptoms ranging from mild diarrhoea to moderate, severe and fulminant PMC. Presentation of disease is usually signified by watery diarrhoea, fever, abdominal pain and/or tenderness[161]. Antibiotics such as the fluoroquinolones, cephalosporins and clindamycin[42], which are used to treat other infections, are known to facilitate the onset of CDI[26]. The most

commonly used and/or UK accepted and licensed treatments for CDI are, oral or intravenous vancomycin, metronidazole, or fidaxomicin[133] but alternative therapies are also available. Much research has focused on identifying the significant risk factors for development of CDI in order to provide effective treatment regimens. Significant risk factors for the onset of CDI are, hospitalisation and treatment with antibiotics[114]. Other important risk factors include, old age (≥ 65 yrs), and the use of[140], and dose dependant effect of proton pump inhibitors (PPIs)[100]. Although there has been on-going debate about the risk incurred by PPI use[167,73]. Chronic and serious illness, underlying health conditions, as well as agents that alter gut microbiota; such as chemotherapy agents[162], haemodialysis, non-surgical admission and duration of stay in an intensive care unit[75] have also been identified as risk factors for CDI.

Perhaps a more significant challenge for researchers is identifying risk factors for outcomes of infection, such as, severity of disease, recurrent infection and mortality. There is a current research focus on defining risk factors for the prognosis of such outcomes for CDI, and also combining these prognostic variables into scoring systems which can identify patients who may be more at risk of certain infection outcomes. Prediction rules which combine information such as, patient characteristics, test results and other disease characteristics are intended to aid a clinician in making medical decisions which can facilitate the way a patient is treated if increased risk for an outcome is identified[213]. A further discussion of risk factors associated with *C. difficile* related mortality and recurrent infection will be presented in Chapter Two.

1.9 Summary

This general introduction has briefly outlined how C. *difficile* has emerged as a problematic HCAI, and introduced the steps that researchers are taking to facilitate control of infection within a clinical setting.

Improving control of CDI may require a combination of rapid and more robust diagnostic tests, simple, accurate prediction rules of clinical outcomes, and phenotypic and genetic analysis of the *C. difficile* strains that are causing infection, to better understand their possible role in disease outcomes.

Combining complex datasets to elucidate what contribution host and pathogen factors may have on the outcome of an infection is a formidable task. The use of statistical software will facilitate the identification of host risk factors for CDI outcomes, and powerful bioinformatic programmes will facilitate our understanding of *C. difficile* at the whole genome sequence level. Classical microbiological techniques will continue to play an important role in defining phenotypes of such a divergent collection of *C. difficile* ribotypes.

1.10 Aim

With the details of *C. difficile* pathogenicity and burden on the healthcare setting outlined in this thesis, it is envisaged that the aim of this PhD will be to develop a more comprehensive knowledge of both host and strain association to CDI outcomes such as mortality and recurrent infection, through data analysis of clinical host variables and genetic and phenotypic characterisation of clinically relevant isolates.

This thesis is organised into three parts in order to convey how different methods of research have been conducted to try and gain a more comprehensive understanding of the host and bacterial factors that contribute to the outcomes of *C. difficile* infection.

- Part One of this thesis will introduce host clinical risk factors that are associated with outcomes of CDI, highlight a possible novel host marker of infection, and introduce how the development of clinical prediction rules may facilitate early identification of outcomes of infection such as, mortality and recurrence.

- Part Two will highlight how acquisition of certain ribotypes may be associated with the outcome of an infection. It will also introduce how bacterial and host factors could be implicated in the change in ribotype prevalence and association to infection outcomes.

- Part three will focus on characterising some of the genetic and phenotypic differences between different ribotypes; in bacterial virulence mechanisms, such as toxin production, sporulation, antibiotic resistance and motility. This will then be followed by a discussion on how these differences may facilitate ribotype association to different infection outcomes.

Part One
Identifying Host Markers of *C. difficile* Infection Outcomes and Development of Clinical Prediction Rules

Chapter Two. A Review of Clinical Risk Factors and Prediction Rules for Mortality and Recurrent *C. difficile* Infection

2.1 Introduction

Chapter One highlighted the risk factors which are prognostic of the onset of CDI. However, identifying patients with adverse outcomes of infection, such as recurrent CDI and a high mortality risk is important, especially as better treatments for CDI are emerging, which are more costly than older, less effective therapies[45]. Health care associated infections (HCAIs) caused by *C. difficile* represent a major clinical burden due to the increased expenditure on patient management and implementation of infection control measures[223]. The expenditure is magnified as many patients may suffer from recurrent infection[63], which is defined as a return of clinical symptoms of CDI after a period of wellness, along with a positive stool test for *C. difficile* at any evaluation period[77]. The increased patient expenditure is not however to supersede the burden caused by morbidity and mortality as a result of CDI[193]. There have been few studies to generate simple classification rules for patients with CDI who may be more at risk of adverse outcomes such as mortality and/or recurrence[30, 32, 45]; considering that a recent publication by J.S Brazier stated that "*C. difficile* has transcended the confines of the specialist anaerobic scientific literature and has taken its place alongside other modern day hospital superbugs that are well and truly in the public domain"[35].

This chapter will highlight the risk factors that have been identified for outcomes of infection such as mortality and recurrence, introduce how novel markers of CDI may be used as prognostic variables, and also highlight how some studies have been combining risk factors of outcomes to develop clinical prediction rules to identify patients in these 'at risk' groups.

2.2 *C. difficile* and Mortality

Antibiotic resistant strains of *Staphylococcus aureus*, *Enterococcus* spp. and *C. difficile* have been widely publicised as a particular problem in the healthcare system[232].

In the US, patients developing CDI have been estimated to cost the healthcare industry an additional $3669-7234 per hospitalisation[162] and in the UK, as of 1996, this cost is estimated at £4000/case[228].

Internationally, CDI related mortality has risen steadily over the last 10 years and has been attributed to over 300000 deaths per annum in the US in 2006, up from >150000 cases per annum in 2000. It is estimated that at present, there are approximately 15000-20000 CDI related deaths per annum in the US[187]. Across Europe incidences of CDI have also increased, with Germany seeing a rise in cases from between 1.7-3.8 cases per 100000 per annum in 2002 to 14.8 cases per 100000 per annum in 2006, and in Spain the number of CDI cases in discharged patients >65 yrs has risen threefold between 1997 and 2005[193].

Mortality rates among CDI patients vary greatly in the literature, with rates between 4.7 % and 57% being reported[163]. According to the Office for National Statistics, deaths from CDI in England and Wales accounted for 1.1% of all deaths between 2006 and 2010. In England and Wales cases of death certificates mentioning *C. difficile* rose periodically from 1999 to 2006, and peaked in 2007 due to an epidemic strain of *C. difficile* which is now predominant in the clinical setting. Since 2007, cases of cases of death certificates mentioning *C. difficile* have fallen (Figure 2.1).

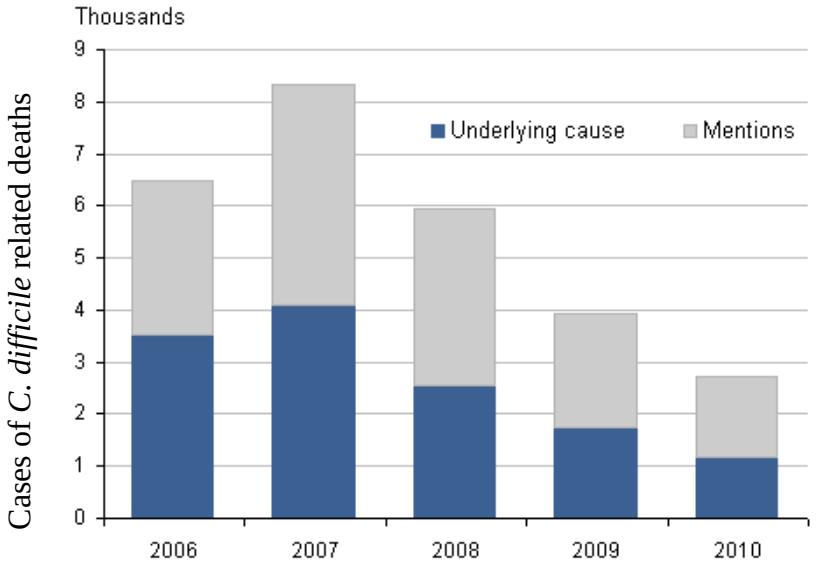


Figure 2.1- Number of death certificates mentioning *C. difficile* by underlying cause and other mentions, from 2006-2010 in England and Wales. Source: Office for National Statistics licensed under the Open Government Licence v.1.0.[205].

In England, from the period 2007-2010, within cases of CDI related deaths, the percentage of death certificates mentioning *C. difficile* as an underlying cause, ranged from 49%-41%. (Table 2.1).

Table 2.1- Number of deaths in England where *C. difficile* was mentioned on the death certificate as the underlying cause of death, 2007-2010.

	2007	**2008**	**2009**	**2010**
Death Certificates mentioning *C. difficile*	7,916	5,465	3,550	2,335
Death Certificates where *C. difficile* was the underlying cause of death	3,875	2,298	1,510	946
Percentage of mentions selected as underlying cause	49	42	43	41

Table adapted from data from the Office for National Statistics licensed under the Open Government Licence v.1.0.[205]

Although deaths due to CDI seem to be a declining, this should not overshadow the number of cases of CDI reported as underlying cause of death, as compared to deaths from other well-known HCAIs such as, methicillin resistant *Staphylococcus aureus* (MRSA), which ranged from 29-16% over the same period[204]. A review by Jones *et al*. report that CDI is 12-

14 times more common than MRSA bacteraemia, therefore the mortality burden imposed by acquiring *C. difficile* is clearly high[114]. This emphasises the need to facilitate the elimination and/or control of its transmission within the healthcare setting.

2.3 *C. difficile* and Recurrent Infection

Recurrent infection is known to occur in approximately 20% of patients following withdrawal of antibiotics[112], however, this may rise to 65% if a patient has a history of prior CDI[102]. Recurrent infection may occur due to relapse (with the same isolate); possibly due to persistent spores remaining in the gut and re-colonising after antibiotic withdrawal, or re-infection from an exogenous source. The increased healthcare associated cost of recurrent infection is due to the re-admittance of patients to hospital, and the re-administration of antibiotics and diagnostic tests[63]. Re-infection of patients from non-clinical sources is of growing concern, and ribotypes such as 078, that are capable of infecting both humans and animals[85] are particularly problematic as they potentially increase the pool of community acquired infection sources. These community infection sources may not only be pivotal to re-infection of 'at risk' patients, but could facilitate the evolution of an epidemic ribotype that is non-nosocomially associated, and is able to infect and colonise hosts who are not usually susceptible to infection[125].

Although incidences of relapse versus re-infection have not been well documented, approximately 33%-75% cases of recurrent CDI have been attributed to re-infection with another isolate[111], which could possibly be explained by the duration of time before subsequent infections occur. In 1989, a study by Johnson *et al*. found that the mean time between the cessation of antibiotic therapy and relapses was 14.5 days, whereas for re-infections it was 42.5 days, indicating that the increased time between infections may be indicative of re-infection from an exogenous source[113]. Barbut *et al*. also imply that re-infections occur later than relapses, and thus emphasises the need to eliminate spore

contaminated environments to reduce the chances of re-infection[21]. Identifying host risk factors associated with recurrent infection could facilitate the way these patients are treated and the way that infection is managed, especially in cases where CDI is severe and may require surgical intervention.

2.4 Clinical Prediction of *C. difficile* Infection Outcomes

Although identifying the potential risk factors for CDI is useful (Described in Chapter One), early identification of those at risk from more adverse infection outcomes is also important, in order for clinicians to effectively assess patient management options, including both medical and surgical intervention[22]. It is therefore surprising that there have been few studies identifying risk factors and generating prediction rules for infection outcomes such as mortality and recurrence. In 2012, a systematic review of prediction rules for infection outcomes in patients with CDI, found only a small number of studies (N=13) had derived rules for outcomes such as, severity of disease (N=5), mortality, (N=5), recurrence, (N=2) and response to therapy (N=1), however, only two of these have been prospectively validated[45].

2.5 Clinical Risk Factors and Prediction of Mortality

In 2012, Bloomfield *et al*. published a systematic review on clinical variables which were associated with increased mortality in patients with CDI[32]. After exclusion criteria, this study identified 26 publications which looked at risk factors for mortality in patients with CDI. The review identified that the majority of studies found older age, higher white blood cell count (WCC), higher creatinine levels, low albumin levels and, to a lesser extent, corticosteroid use, as being more frequently associated with mortality. To facilitate early identification of patients at risk from mortality in clinical settings, a study by Bhangu *et al*. identified significant clinical variables associated with mortality, and then developed a scoring system

to identify patients with CDI, who may be more at risk from death during the course of infection[30]. In a cohort of 158 patients, this study identified; age ≥ 80, clinically severe disease, WCC ≥ 20 ($x10^9$) or CRP ≥ 150 (mg/L), urea ≥ 15 (mmol/L) and serum albumin ≤ 20 (g/L), as variables that could be used in combination to score the relative risk of death in patients with CDI. By using a score from 0-5 (evaluated within the first 72 hrs of a toxin positive stool sample) the risk of death increased as the score increased, where a score of 0-1; indicated a 22% risk of mortality, a score of 2-3; indicated a 55% risk of mortality and a score of 4-5; indicated an 89% risk of mortality. The study conducted by Bhangu *et al*.[30] relied on five variables which could be easily measured by treating physicians. However, one variable; severity of disease, is further defined by three more variables including sepsis, peritonitis and ≥ 10 episodes of diarrhoea in 24 hrs. The diagnosis of sepsis is further defined by the presence of diarrhoea with at least two other parameters that could include tachycardia (≥ 90 bpm), pyrexia (temperature ≥ 38^{o}C), tachypnea (≥ 20 breaths per minute) or new onset hypotension. The addition of all these clinical parameters requires a more complicated and prolonged analysis that might be undertaken in the time constraints of a busy ward round, and highlights the need for a simpler prediction rule to guide clinical patient management.

2.6 Clinical Risk Factors and Prediction of Recurrent *C. difficile* Infection

Since the clinical burden imposed by recurrent CDI is a significant one, it is surprising that there have been few studies that have evaluated the associated risk factors. This is exemplified by a meta-analysis of studies into recurrent CDI that identified only twelve studies, spanning 24 years which assessed risk factors for recurrent CDI[77].

In 1999, McFarland *et al*. identified two significant risk factors for recurrent CDI using statistical analysis on prospective and retrospective data[160]. Logistic regression analysis revealed two independent variables; older age (64.8 ± 1.6) and a lower quality of health index (42.9 ± 17.8) as risk factors for recurrent CDI (P=0.01). In 2008, a meta-analysis study

confirmed the data showing that age was a significant risk factor (P=0.0012) and also highlighted other significant variables such as, continued use of non *C. difficile* associated antibiotics after diagnosis (P<0.001) and concomitant receipt of antacids (P=0.019)[77]. The latter variable is however controversial, with data showing antacids having both negative and positive correlation with CDI[77]. The authors of the meta-analysis also emphasise its limitations, due to the inclusion of studies with small sample sizes and limited publication of data identifying other possible variables which could contribute to increased incidence of recurrent disease. Such variables include, initial severity of disease, presence of a hypervirulent strain, immune dysfunction and gastrointestinal surgery.

A novel study by Hu *et al.*[102] has taken the identification of factors associated with recurrent CDI one step further, and used the risk factors identified in an initial prospective study (derivation cohort) to develop a prediction rule for recurrent infection that could be tested in an independent cohort (validation cohort). The identifiers obtained from the derivation cohort (via multivariate logistical analysis) were, age (>65 yrs), severe or fulminant disease and additional antibiotic use after discontinuation of CDI therapy, and were used to develop a clinical prediction rule. Their study is unique in that it combined immunological data from antitoxin A IgG measurements, with the risk factors identified from clinical observations, to develop a combined prediction rule. The combined rule was tested in both cohorts to explore any added value that IgG data may provide. Sensitivity and specificity of the clinical and combined prediction rules were compared using receiver operating characteristic curves (ROC) and the area under the curve (AUC) was used to determine if the rules were statistically significant at predicting who would develop recurrent CDI. The analysis showed that the performance of the clinical prediction rule was nearly identical in the derivation and validation cohort (AUC= 0.83 vs. 0.80) but for the combined rule, its performance was lower in the validation cohort, compared to the derivation cohort (AUC= 0.89 vs. 0.62). In the Hu *et*

al.[102] study, the poorer performance of the combined rule in the validation cohort as compared to the derivation cohort was attributed to a number of factors including, small sample sizes in both cohorts, different timing of antibody measurements, and possible changing epidemiology of the *C. difficile* strains that could lead to a change in IgG response to toxin A. However, while detection of antitoxin A IgG was not beneficial to the prediction rule in the study; it does emphasise the possibility of using other biological markers as novel host markers of CDI, and this is something which will be discussed below.

2.7 Faecal Calprotectin as a Biomarker for *C. difficile* Infection Outcomes

The previous sections have highlighted risk factors for CDI outcomes such as, mortality and recurrence. However, there has been little research into additional markers that could be useful predictors of these outcomes. A clinically useful biomarker which has been implemented in assessing levels of gastrointestinal inflammation is calprotectin[3]. For most patients, CDI results in an inflammatory response in the host, and to this end the use of calprotectin was investigated in this thesis, as a putative marker for CDI outcomes.

Calprotectin is a 36 KDa heterotrimeric calcium binding protein which is widely distributed in myelomonocytic and epithelial cells[110]. It is a major protein in neutrophillic granulocytes and macrophages[192], and high levels of calprotectin are found in association with diseases such as cystic fibrosis, rheumatoid arthritis, ulcerative colitis and bacterial infections, in a variety of organic fluids including serum/plasma, urine, faeces, oral fluid, cerebral spinal fluid, synovial fluid and empyema fluid[110]. A methodological study by Roseth *et al.*[192] discovered calprotectin in the faeces of patients with irritable bowel disease (IBD) whilst looking for a marker of faecal leukocytes and indeed, these results have been reported in other studies[52, 80]. In acute phase inflammatory reactions, calprotectin can be detected at elevated levels, correlating with elevated levels of neutrophils and granulocytes, or other inflammatory markers, such as CRP[110]. Calprotectin has also been correlated with the

excretion of ^{111}-indium labelled leucocytes which is said to be the 'gold standard' for measuring intestinal inflammation[52] and it has also been implicated in the assessment of disease severity in patients with ulcerative colitis[191] and Crohns disease[80]. Further clarification of its usefulness in assessing intestinal inflammation is highlighted by Costa *et al*., who revealed that calprotectin is a stronger indicator of relapse of IBD than ulcerative colitis, and that calprotectin levels may reflect the inflammatory state of the intestinal mucosa[52]. The use of calprotectin as a biomarker for intestinal inflammation has been pursued with scientific interest, as it provides a non-invasive means of reflecting intestinal inflammation and by passes the difficulty of assessing faecal leukocytes[51]. Fagerhol *et al*.[68] also propose that in general medical practice there is a need for a simple test to assist in the selection of patients with prolonged diarrhoea and/or other non-specific gastrointestinal symptoms.

2.8 Study Aim

The aim of Part One of this study is to try and identify both known and novel host predictors of disease outcomes such as mortality, and recurrent CDI, in order that they might be combined and used in simple scoring rules to identify patients in these 'at risk' groups. Simple rules that can be applied to a case within a clinical setting are of great interest to healthcare system. A scoring rule may allow the accurate prediction of outcomes such as recurrence and mortality, thus allowing attending clinicians to effectively monitor patients' treatment regimen.

Chapter Three. Clinical Outcomes of *C. difficile* Infection- Mortality and Recurrence; Risk factors and Clinical Prediction Rules

3.1 Introduction and Aim

Chapter Two highlighted that identifying risk factors for CDI outcomes such as mortality and recurrence, and combining them into simple scoring rules is of great interest to the healthcare system. These rules could allow the prediction of outcomes such as recurrence and mortality which can then be used to effectively tailor patients' treatment regimen[22].

The aim of the work in this chapter is to use retrospectively collected clinical data to identify host markers of infection outcomes such as, mortality and recurrence, and to use these variables to obtain a prediction rule to classify patients with CDI who may be at more risk of these outcomes.

3.2 Study Cohort

Detailed descriptions of methods used in this chapter can be found in the Materials and Methods chapter (Chapter Ten, Section 10.4). The definition of all-cause mortality in this study is; patients who have died from CDI and non-CDI related causes as an inpatient, or have been discharged and have then died ≤ 30 days. Recurrent CDI is defined in this study as; the onset of clinical symptoms within 30 days of withdrawal of treatment of antibiotics, once stools have returned to type IV according to the Bristol Stool chart. The threshold for recurrent CDI was determined for this study as, that which would capture true recurrent CDI, and not a prolonged infection or two completely separate CDIs. Univariate, multivariate and decision tree analysis was used to analyse over 186 variables; retrospectively collated from clinical data of 213 patients, including those with multiple infections, which gives 245 cases in total. Cohort inclusion criteria was defined as those who had a frozen sample of *C. difficile* toxin A/B positive stool confirmed using the TechLab® C.diff Quik Chek Complete™

Enzyme linked Immuno-assay (Alere Ltd, UK), and were recent inpatients on a *Clostridium difficile* cohort ward at the Royal Devon and Exeter (RD&E) hospital.

3.3 Patient Demographics

54% (115) of the cohort were female. 71% (174) of cases were admitted for acute medical care, 10% (25) of cases were admitted due to community acquired CDI and the remaining cases were admitted for; elective surgery (7), emergency surgery (15), trauma (12), planned procedures (1) or renal care (11). 78% (191) of cases were admitted from the home environment, 13.5% (33) of cases were admitted from another hospital and 8.6% (21) of cases were admitted from long term care facilities. All but three cases (who were later transferred to an intensive care ward) were admitted and treated in the *Clostridium difficile* cohort ward. 29 patients had >1 CDI, which is a recurrence rate of 14%. 26 patients presented with two episodes of CDI, and three patients presented with three episodes. Out of the 29 patients listed as having recurrent CDI, only 16 cases were truly recurrent according to the definition in Section 3.2. Of the 16 patients who had recurrent CDI, 14 had relapsing infection (same ribotype) and two had a secondary episode with a different ribotype. From the final 16 patients who had true recurrent infection, six patients received additional antibiotics between primary and seconday infections. Five of the six patients with concurrent antibiotics listed between infections; had been discharged after the first infection, re-admitted within 30 days, then had a subsequent infection. Four patients were discharged, and then re-admitted, and had a subsequent infection with no concurrent antibiotics noted in the case data. One patient had recurrent CDI in the same hospital stay and had concurrent antibiotics listed in their case notes. Six people had recurrent CDI in the same hospital stay, and were not given concurrent antibiotics before the onset of a seconday infection.

Of the 245 cases, 225 had data available for antibiotics given prior to admission. 34 different antibiotics were prescribed to this patient cohort. Patients were prescribed between 1-8

antibiotics prior to the onset of CDI (median= 2), and four patients had no antibiotics prior to the onset of CDI. The 34 different antibiotics could be grouped into 20 classes, and the main class of antibiotic prescribed to patients were the penicillins. No antibiotic was clearly prescribed more than any other.

Of the 245 cases, 232 had data available for post CDI treatments. A total of 11 CDI treatment options were identified, with most patients receiving between 1-5 treatments (median= 1), and 10 patients receiving no treatment, either due to palliative care, or other reasons. Oral metronidazole (45.1%) and oral vancomycin (35.6%) accounted for most of the CDI treatments, but other treatments included, intravenous metronidazole and vancomycin, fluconazole, fusidic acid and rifampicin.

3.4 Mortality Outcome Statistics

In this study cohort, 76% (162) of patients survived to discharge, with the remaining 24% (51) either surviving to discharge then dying within 30 days (40), or dying within 30 days of admission (11) (all-cause mortality). The overall mortality rate for this cohort was 24% and *C. difficile* related mortality was 12.6%. Of the 51 patients who died at any stage after admission, 53% (27) had *C. difficile* listed in part one or two of their death certificates, according to the Department of Health guidelines[1] (CDI related mortality). Of these 27 deaths, 9 (33%) were due to CDI (listed in part one of the death certificate, according to the Department of Health report[1]) and 18 (66%) were not directly due to CDI, but it was mentioned on the death certificate (listed in part two of the death certificate, according to the Department of Health report[1]). The remaining 24 deaths were not attributed in any way to CDI (non-CDI related mortality). Of the 245 cases, 197 had data available for co-morbidities. A total of 15 co-morbidities were found amongst patients in this cohort. Most patients had between 1-5 co-morbidities (median= 1) and 43 patients had no co-morbidities. Ischaemic

heart disease and/or congestive cardiac failure (23.7%) were the most prevalent co-morbidities, followed by active cancer (14%) and diabetes (12.4%). There were only four cases of inflammatory bowel disease within this cohort.

3.5 Severity of Infection Outcome Statistics

Severity of infection was noted for all cases (using the Department of Health guidelines[1]) and there were 31.4% (77) cases of mild infection, 13.9% (34) cases of moderate infection, 47.3% (116) cases of severe infection and 7.3% (18) cases of life threatening infection.

3.6 Site of Infection Acquisition Statistics

67% (164) of this cohort had hospital acquired CDI in, 16% (40) had community acquired CDI, and 17% (41) of cases had community acquired CDI, but had been an in-patient in the last 30 days. 57.7% (123) of patients admitted were on gastric acid suppressing drugs, and four patients were on cancer chemotherapy agents. Baseline clinical measurements for the cohort are presented Table 3.1.

Table 3.1– Baseline clinical characteristics of the RD&E cohort.

	N	Min.	Max.	Mean	Std. Error. Mean	Std. Deviation	Skewness	Skewness Std. Error.
Age on Admission (yrs)	245	21	100	77.75 (Median = 80)	.853	13.346	-1.468	.156
Length of Stay in hospital (days)	245	3	158	38.61	1.872	29.302	1.388	.156
Pulse (BPM)	234	56	149	88.50	1.065	16.292	.469	.159
Systolic Blood pressure (mm/Hg)	235	60	190	123.09	1.497	22.943	.070	.159
Diastolic blood pressure (mm/Hg)	234	30	114	68.24	.872	13.339	.291	.159
Temperature (^{o}C)	232	35	39	36.80	.041	.620	.785	.160
Respiratory rate (resps/min)	227	11	35	17.62	.267	4.028	1.615	.162
Stool Calprotectin Levels (µg/g)	168	42	12989	1912.16	162.798	2110.104	2.665	.187
Percent Rise In Creatinine from Baseline	242	44.22	427.08	32.687	3.3962	52.832	3.418	.156
CRP Levels (mg/L)	228	3	354	107.25	5.375	81.166	.933	.161
White Blood Cell Count (x 10^3 mcL)	243	0	173	14.81	.948	14.781	6.455	.156
Neutrophil Count (cells/ul)	241	0	82	11.29	.568	8.819	3.159	.157
Platelet Levels (mcL)	243	14	1124	306.69	9.472	147.655	1.338	.156
Serum Albumin Levels (g/L)	204	15	46	30.20	.394	5.622	.202	.170
Lactate Levels (IU/L)	10	1	3	1.70	.260	.823	.687	.687

N=number of cases

Data from Table 3.1 and histographic displays of scalar variable measurements (data not shown) indicated that some vararibles were normally or near normally distributed (systolic blood pressure and serum albumin) but others were positively skewed (pulse, diastolic blood pressure, temperature, respiratory breaths per minute, C- reactive protein (CRP) levels, white cell count (WCC), neutrophil count, percent rise in creatinine levels, platelet levels, stool calprotectin levels and length of stay in hospital). One variable; age, was negatively skewed.

In order to deduce whether parametric or non-parametric tests should be used to test for significance of variables to the outcomes, Shapiro-Wilk tests for normality were performed to calculate how much deviance each variable showed from the normal distribution. Serum albumin and systolic blood pressure were the only two variables which did not deviate significantly from the normal distribution (P=0.216 and 0.718 respectively). Variable groups were then tested for normality in relation to the outcome measures all-cause mortality and recurrent CDI.

3.7 Univariate Analysis

All the demographic and baseline clinical characteristic data outlined above was used in univariate and multivariate analysis to try and identifiy variables significantly associated with mortality and recurrent CDI. Analysis was conducted on one set of patient data (N=213). For analysis of differences in variable groups for the mortality outcome, data analysis was performed on all cases with primary only CDI, and on one set of data from those with recurrent CDI based on the last data entry, to account for those which may have had a worse outcome on subsequent infections. To detect significance of variables to the outcome recurrence, data analysis was performed on all cases with primary only CDI and on one set of data from those with recurrent CDI, based on the index case data entry. A potential limitation of the data analysis procedure conducted on this cohort may have arisen due to the positive and negative predictive value (PPV and NPV) power of the TechLab® C.diff Quik Chek Complete™ Enzyme linked Immuno-assay test (Alere Ltd, UK). The PPV and NPV of this test can vary with differences in CDI prevalence rates[178]. This may have meant that some true positive patient data was excluded from the initial data collection and some false positive patient data was incorporated.

3.7.1 Variables Independently Associated with All-Cause Mortality

All-cause mortality was subdivided into three groups to establish if there were significant differences between means and medians of variables that were prognostic of either CDI-related mortality (N=27), or non-CDI related mortality (N=24), versus those who survived (comparator group; N=162).

All variables were tested for significance in relation to the all-cause mortality outcome by either, Independent-median tests, One-Way ANOVA or Chi-squared tests. Variables which remained statistically significantly associated with the outcomes (P=0.05) after Bonferroni correction (where P=0.05/N; N=number of variables being tested) were used in the multinomial regression model.

3.7.1. i. Parametric tests

One-way ANOVA tests for differences between means of groups with respect to the outcome measure all-cause mortality (0=survived; 1= CDI related mortality, 2= non-CDI related mortality) were used for normally distributed interval data. Serum albumin was the only normally distributed variable to remain statistically significantly associated with all-cause mortality.

The difference in mean serum albumin levels across the groups was determined to be significantly different (P<0.001; F=12.45; *df*=178). Results of Bonferroni post-hoc analysis revealed that the significant differences in mean serum albumin levels were between, the survival group (31.28 g/L), and the groups; death was related to CDI (25.60 g/L; P<0.001) and the group whose death was not related to CDI (27.37 g/L; P=0.011).

3.7.1. ii. Non-Parametric Tests

Independent Samples K-median tests were used for variables which were non-normally distributed with respect to the all-cause mortality outcome. Respiratory rate, CRP and WCC

were the only variables that remained statistically significantly associated with all-cause mortality.

The difference in median respiratory rate across the mortality groups was determined to be significantly different ($\chi^2(2)=11.8$; P=0.003) from the grand median (16 resps/min). Pair wise comparison of all groups revealed that there was a statistical difference in median respiratory rate between the survival group (16 resps/min) and the group whose death was related to CDI (18.5 resps/min) (P=0.005) but not the group whose death was not related to CDI (18 resps/min).

The difference in median CRP levels across the groups was determined to be significantly different ($\chi^2(2)=14.2$; P=0.001) from the grand median (89.50 mg/L). Pair wise comparison of all groups revealed that there was a statistical difference in median CRP levels between the survival group (79 mg/L) and the group whose death was related to CDI (180 mg/L) (P<0.003) but not the group whose death was not related to CDI (116 mg/L).

The difference in median WCC levels across the groups was determined to be significantly different ($\chi^2(2)=11.4$; P=0.003) from the grand median (12 $x10^3$ mcL). Pair wise comparison of all groups revealed that there was a statistical difference in median WCC levels between the survival group (11 $x10^3$ mcL) and the group whose death was related to CDI (17 $x10^3$ mcL) (P=0.002) but not the group whose death was not related to CDI (10.5 $x10^3$ mcL).

3.7.2. Variables Independently Associated with Recurrence

Because the actual number of true recurrent cases (relapse and re-infection) in this study was so small after the case exclusion criteria that was defined in Section 3.2 (N=16), all index cases of proposed recurrent CDI (N=29) were included in the analysis procedure.

Chi-Squared tests, Independent median tests and One-way ANOVA tests were used to deduce if any variables were associated with recurrent CDI. Tests revealed that no variables were independently associated with recurrent CDI.

3.8 Multivariate Analysis of Clinical Variables Independently Associated with All-cause Mortality

Variables independently associated with the CDI related mortality (N=19, dummy code =1) and non-CDI related mortality (N= 18 dummy code =2)) outcomes, (versus the survival group (N=124, dummy code= 0); were subject to multinomial logistic analysis. All significant variables were entered using a block entry method. Results revealed that the variables serum albumin (P<0.001), respiratory rate (P=0.002), CRP (P=0.034) and WCC (P=0.049) were all significant predictors of mortality (Table 3.2). Further breakdown of the model (Table 3.3) revealed that respiratory rate and serum albumin were significant predictors of both non-CDI related mortality (P=0.007 and P=0.004 respectively) and CDI related mortality (P=0.003 and P=0.001 respectively), whilst CRP (P=0.025) and WCC (P=0.020) remained statistically significant specifically for CDI related mortality. The model showed an overall classification of 80.7% as seen in Table 3.4.

Table 3.2- Statistically significant predictor variables left in the multinomial logistic regression equation after block entry.

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	*df*	Sig.
Intercept	166.666	0.438	2	P=0.803
Respiratory Rate	178.521	12.293	2	P=0.002
Serum Albumin	185.052	18.824	2	P<0.001
White Blood Cell Count	172.279	6.050	2	P=0.049
C-Reactive Protein	173.006	6.778	2	P=0.034

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

Table 3.3- Parameter estimates for the all-cause outcome measure; CDI related mortality and non-CDI related mortality.

All-cause Mortality		β	Std. Error	Wald	*df*	Sig.	Exp(β)	95% Confidence Interval for Exp(β)	
								Lower Bound	Upper Bound
CDI related mortality	**Intercept**	-1.217	2.119	.330	1	P=0.566			
	Respiratory Rate	.200	.068	8.791	1	P=0.003	1.222	1.070	1.395
	Albumin	-.221	.067	10.850	1	P=0.001	0.801	0.703	0.914
	White Blood Cell Count	.045	.020	5.016	1	P=0.025	1.046	1.006	1.088
	C-Reactive Protein	.009	.004	5.388	1	P=0.020	1.009	1.001	1.017
Non-CDI related mortality	**Intercept**	-.967	1.941	.248	1	P=0.618			
	Respiratory Rate	.171	.063	7.322	1	P=0.007	1.186	1.048	1.342
	Albumin	-.170	.059	8.174	1	P=0.004	0.844	0.751	0.948
	White Blood Cell Count	.005	.030	.024	1	P=0.877	1.005	0.947	1.066
	C-Reactive Protein	.007	.004	3.196	1	P=0.074	1.007	0.999	1.015

-2Log Likelihood =166.3 (χ^2= 58.6, *df*=8, P=0.000), Nagelkerke R^2=40.5, Pearson Chi-square Goodness of Fit Test; P=0.996.

Table 3.4- Classification table for multinomial regression model for all-cause mortality outcome.

Observed (N)	Predicted			
	Survived	**CDI related mortality**	**Non-CDI related mortality**	**Percent Correct**
Survived	121	3	0	97.6%
CDI related mortality	9	9	1	47.4%
Non-CDI related mortality	12	6	0	.0%
Overall Percentage	88.2%	11.2%	.6%	80.7%

Variables which were found to be significant predictors of all-cause mortality were entered into a decision tree model in order to obtain threshold levels which could accurately classify cases into the categories, *survived* (N=194 dummy code=1) and *died* (N=51, dummy code =0) of the mortality outcome.

3.9 Model Derivation

Due to the small number of patients in both the CDI-related mortality group and the non-CDI related mortality groups, for the purpose of decision tree classification analysis these groups were merged into the single category outcome; *died*, totalling 51 deaths in the cohort of 213,

and resulting in a 24% mortality rate. All the variables showing significant association with both the CDI-related and non-CDI related mortality were entered in a decision tree model to look for the measurements at which the tree partitioned the data into each outcome. The prediction rule was derived by inputting all 245 cases of data (all data for patients with primary only and recurrent CDI) into the analysis. Data was then analysed using a (50:50) training and test data split, in order that a sufficient amount of data was captured for analysis in each data portion. A classification and regression tree (CRT) model was used to predict values of the dependent (target) variable based on values of the independent (predictor) variables (SPSS Inc). The CRT model was chosen as a growing method. It uses a recursive partitioning method and builds classification and regression trees for predicting categorical variables (classification). All results for derivation of threshold values are reported on test data, and detailed explanations of the model method can be seen in the Materials and Methods chapter (Chapter 10, Section 10.4.iii).

Variables and their respective threshold values, which were found during multivariate and CRT analysis, were assigned a score to create a prediction rule which was then applied to the derivation and a validation cohort.

3.10 Decision Tree Classification Results for All Cause Mortality

The variables, respiratory rate (resps/min), CRP (mg/L), WCC ($x10^3$ mcL) and serum albumin (g/L) were entered into a CRT model to look for measurements at which the tree split the data into the outcome groups; *died* and *survived*. The model used all four variables to classify the test data (N=120), and the overall accuracy of the model at classifying the data was 75%.

The model correctly predicted 47% of cases into the death outcome, and 80% of the survival outcome cases correctly, using these four variables (Table 3.5). Inaccuracy in classifying the

test data may be due to the fact that there were only a small percentage of cases for the death outcome in the cohort.

Table 3.5 - Classification of test data by the CRT model.

Sample	Observed	Predicted		
		Died	Survived	Percent Correct
Test	Died	9	10	47.4%
	Survived	20	81	80.2%
	Overall Percentage	24.2%	75.8%	75.0%

Growing Method: CRT, Dependent Variable: Mortality Outcome

The model generated normalised importance of variables (Figure 3.1).

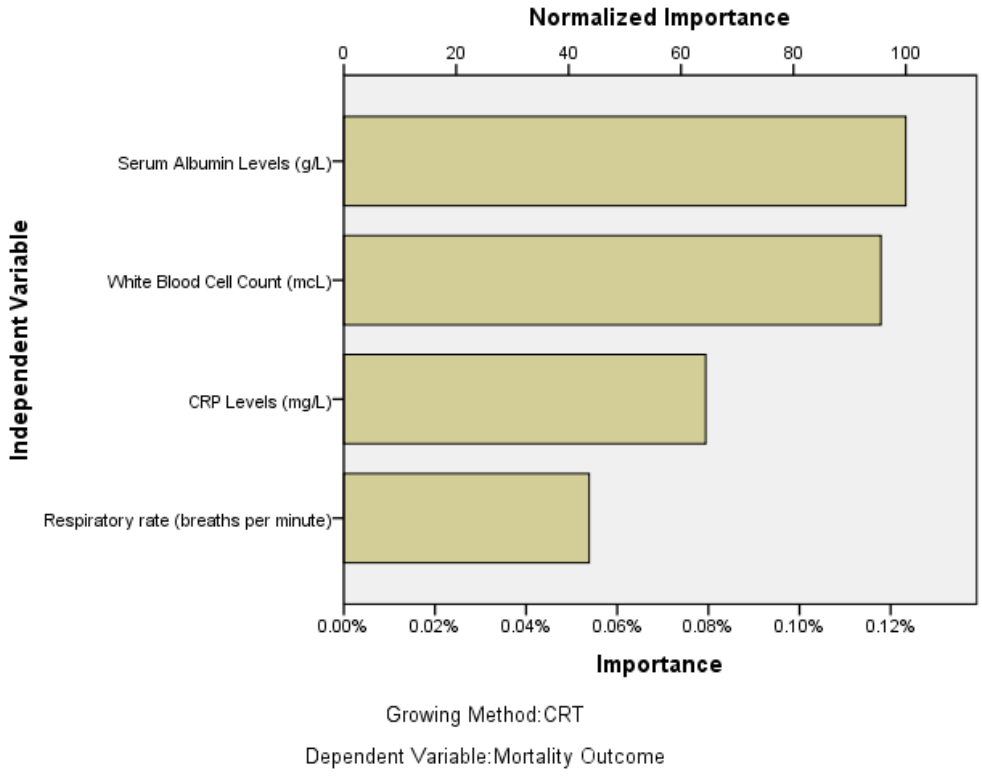


Figure 3.1- Normalised importance of the variables in the mortality outcome model.

Threshold levels for each of the predictor variables were generated from the model (Figure 3.2). Classification rules (data not shown) were also generated along with the model diagram.

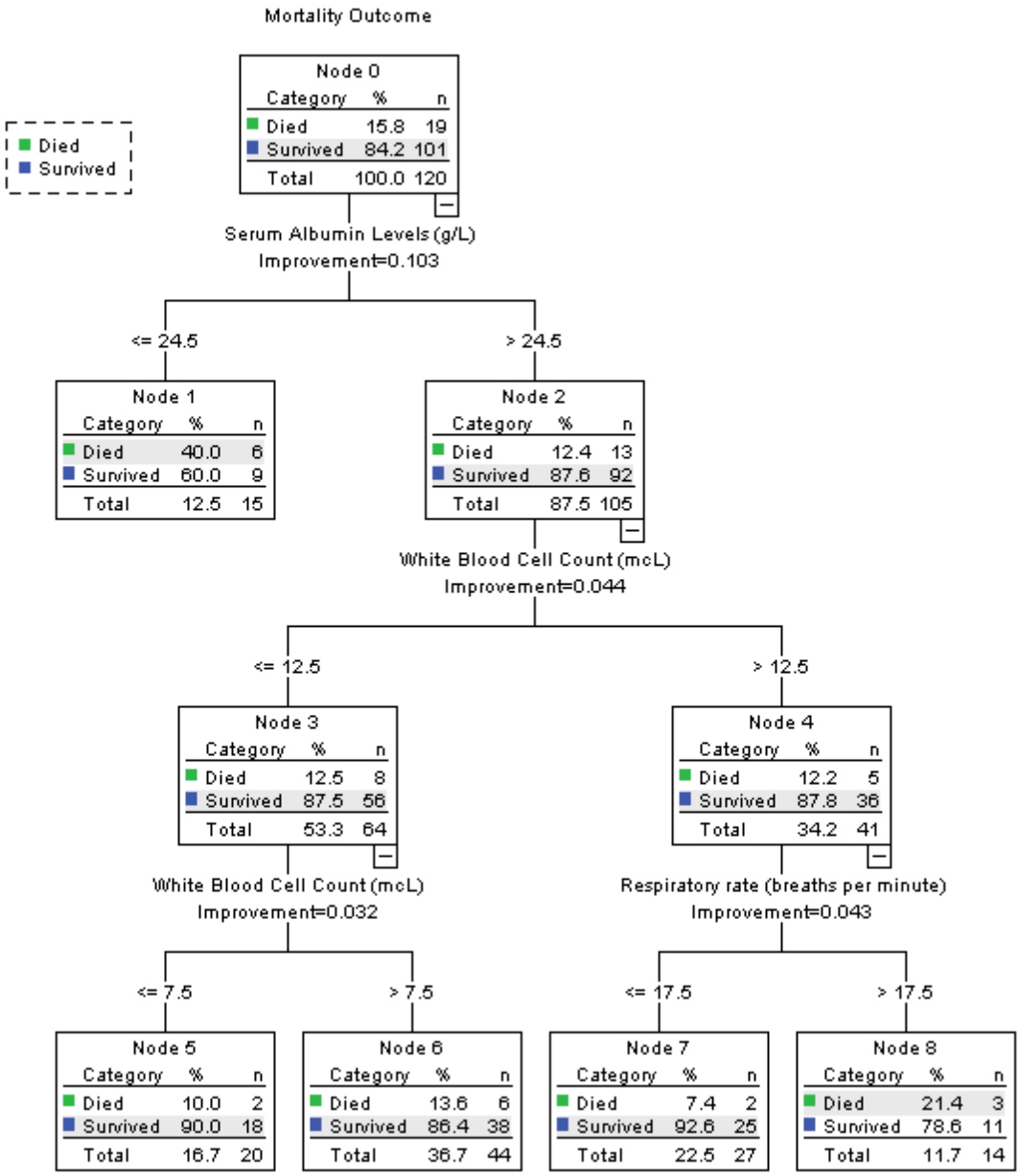


Figure 3.2 - CRT diagram for test data for the outcome all-cause mortality.

Threshold values generated on the test data, for each variable were–

- Serum Albumin (g/L) ≤ 24.5
- C- reactive protein (mg/L) >228
- Respiratory rate (resps/min) >17
- White cell count ($x10^3$ mcL) >12

Normalised importance values indicated that serum albumin and WCC were the most important predictors, followed by CRP and respiratory rate (Figure 3.1). The decision tree classification rules (not shown here) identified that low serum albumin (≤ 24.5 g/L) or high CRP levels (>228 mg/L) were both associated with the increased probability of death as an outcome. However, if serum albumin levels were >24.5 g/L, and CRP was < 228 mg/L then, increased WCC (>12 $x10^3$ mcL) and increased respiratory rate (>17 resps/min) were associated with the increased probability of death as an outcome.

3.11 Prediction Rule for Mortality in Patients with *C. difficile* Infection

Based on the threshold levels indicated by the CRT model, a prediction rule was developed, which assigned one point for a serum albumin level ≤ 24.5 g/L, one point for a CRP level > 228 mg/L and one point for a combination of WCC >12 $x10^3$ mcL and respiratory rate > 17 resps/min as this proved more discriminatory than both values alone (Figure 3.3).

Variables-

- One point for a Serum Albumin level ≤ 24.5 (g/L),
- One point for a CRP level >228 (mg/L)
- One point for a combination of WCC >12 ($x10^3$ mcL) and respiratory rate >17 (resps/min).

Figure 3.3- Summary of the prediction rule developed during this research.

The summation of any variables (Figure 3.3), resulted in a score from 0-3 which could classify the cases into the groups *survived* or *died*, in the all-cause mortality outcome.

The prediction rule was then applied to cases in the cohort (N=244) and the scores for each case were cross tabulated with the groups in the all-cause mortality outcome. The results are presented in Table 3.6a and 3.6b, and show that the increased risk of mortality follows an increasing score (0, 1, 2, 3) in patients with CDI.

Table 3.6a- Cross tabulation of score for risk of mortality against the actual morality outcome.

		Score for Mortality				Total
		0	**1**	**2**	**3**	
Died	**Count**	16	21	10	3	50
	(%) within Score for Mortality	9.5	36.8	66.7	100.0	20.5
Survived	**Count**	153	36	5	0	194
	(%) within Score for Mortality	90.5	63.2	33.3	.0	79.5
Total	**Count**	169	57	15	3	244*
	(%) within Score for Mortality	100.0	100.0	100.0	100.0	100.0

*One missing case due to lack of data for all variables.

Table 3.6b – Summary of mortality risk with increasing prediction rule score.

Score	Mortality Risk (%)	Count (number of cases)
0	9.5	16/169
1	36.8	21/57
2	66.7	10/15
3	100	3/3

The overall accuracy of the prediction rule tested in the derivation cohort, was evaluated by ROC characteristics using the area under the curve (AUC). The prediction rule yielded an AUC of 0.754 (P<0.001; 95% CI: 0.670-0.837), thus deducing that the model showed a better ability to predict risk of mortality in patients with CDI (Figure 3.4, blue line) than chance alone, which is reflected by the green line in Figures 3.4, 3.5 and 3.6 respectively.

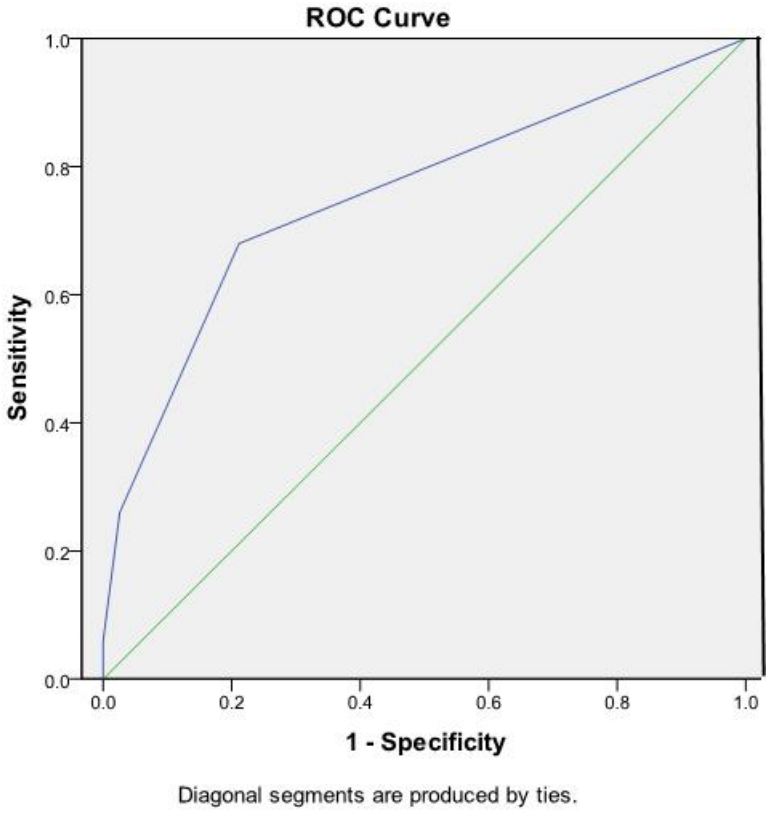


Figure 3.4- ROC curve for the prediction rule, to classify those at risk of mortality, as tested in the derivation cohort.

3.12 Prediction Rule Validation

The prediction rule shown in Figure 3.3 was then validated on independent data from 158 patients with CDI, obtained from a UK teaching hospital[30]. Respiratory rate data was not available for this cohort, and so the prediction rule derived during this study was run again, omitting respiratory rate as a predictor. Thus, one full point was allocated to WCC >12 ($x10^3$ mcL). The simplified prediction rule was tested on both data sets, and compared to the results from a rule previously derived by Bhangu *et al*.[30], (Table 3.7).

Table 3.7 - Prediction rule scoring system (excluding respiratory rate) tested on derivation cohort and validation cohort versus the prediction rule derived by Bhangu *et al*.[30]

	Prediction rule Risk Score on Derivation Cohort (N=244)		**Prediction rule Risk Score on Validation Cohort (N=154).**		**Bhangu *et al* [30] prediction rule (N= 151)**		
Score	**Mortality Risk**	**Count (number of cases)**	**Mortality Risk**	**Count (number of cases)**	**Score**	**Count (number of cases)**	**Mortality risk**
0	10.4 %	13/125	20.9 %	9/43	**0-1**	19/86	22% (Low)
1	23.3 %	20/86	37.1 %	23/62	**2-3**	31/56	55% (medium)
2	42.9%	12/28	54.3%	25/46	**4-5**	8/9	89% (high)
3	100 %	5/5	66.7 %	2/3			

ROC curves were used to evaluate the original prediction rule and the prediction rule (excluding respiratory rate) on both the derivation and validation cohort (Figures 3.5 and 3.6). The original prediction rule yielded an AUC on the derivation cohort of, 0.754 (P<0.001; 95% CI: 0.670-0.837). The simplified prediction rule (excluding respiratory rate), tested on the derivation cohort, resulted in an AUC of 0.704 (P<0.001; 95% CI: 0.619-0.790) (Figure 3.5, blue line). The AUC of the simplified prediction rule was reduced by a further 5% from 0.704 (as found in the derivation cohort), to 0.653 (P=0.001 95% CI: 0.565-0.741), when tested on the validation cohort (Figure 3.6, blue line), but remained statistically significant, thus demonstrating the robustness of the prediction rule when tested on new data.

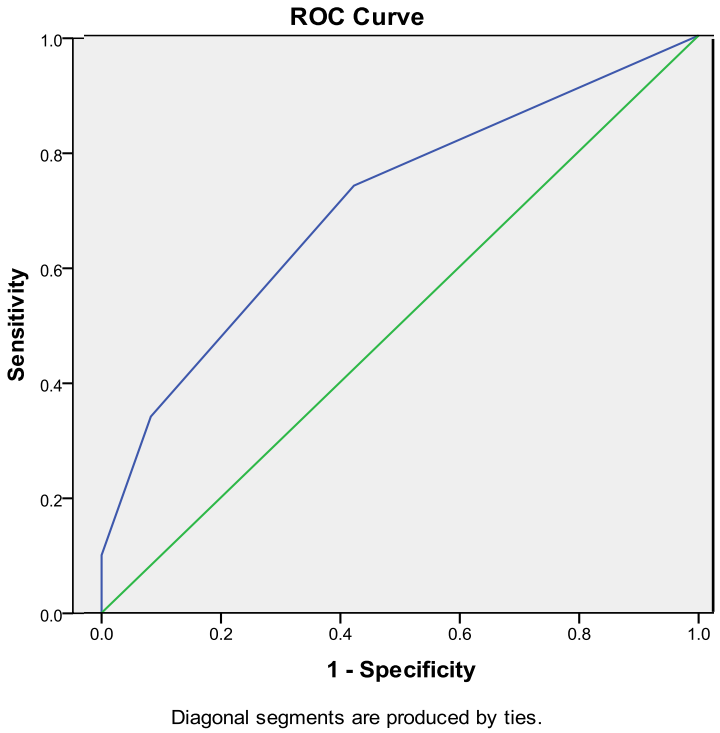


Figure 3.5- ROC curve for the prediction rule (excluding respiratory rate), to classify those at risk of mortality, as tested in the derivation cohort.

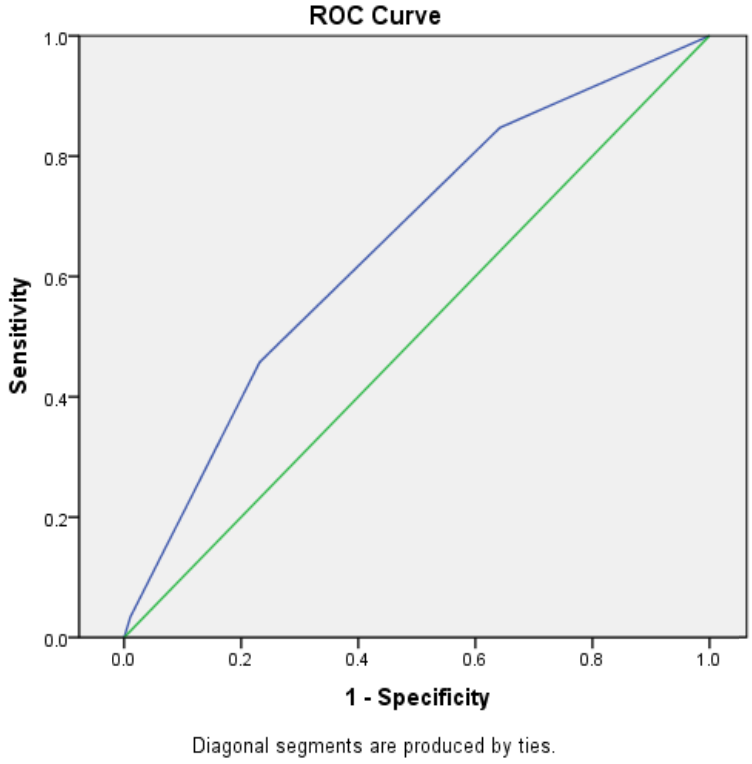


Figure 3.6- ROC curve for the prediction rule (excluding respiratory rate), to classify those at risk of mortality, as tested in the validation cohort.

3.13 Discussion

3.13.1 Discussion of Mortality Data Analysis

The aim of the work in this chapter was to try and identify host variables that could be used to derive prediction rules for CDI outcomes such as, mortality and recurrence. Discussed below are the key findings of this work and its applicability to a clinical setting.

3.13.2 Prediction Rule Generation

A review by Grobman and Stamilo in 2006 has highlighted the advantages and disadvantages of using one particular kind of statistical method to generate a clinical prediction rule[87] (Table 3.8). The present study has combined three methods to derive a rule for predicting risk of mortality in patients with CDI. This method perhaps has the advantage of; first eliminating variables which are not independently associated with mortality via univariate analysis (and could thus over complicate a model if they were all included in CRT analysis), next; determining the significance of the variables in predicting mortality via multivariate analysis, and finally allowing unbiased selection of threshold levels for variables, via CRT analysis, which could then classify patients into the mortality outcome.

Table 3.8- Techniques for the development of clinical prediction rules with corresponding advantages and disadvantages.

Technique	Advantages/strengths	Disadvantages/limitations
Univariate analysis	Simple statistic methods Easy clinical application	Reduced accuracy
Multivariable analysis (e.g., logistic regression)	Improved accuracy Relatively ease of clinical application	More involved statistical methods May miss complex variable relationships
Neural network	Improved accuracy Incorporation of complex variable relationships	Difficult clinical application and dissemination Less intuitive Unknown effect of any single variable
Predictive nomogram	Improved accuracy Ease of clinical application	Advanced statistics
CART/CRT analysis	Improved accuracy Ease of clinical application Intuitive partitioning	Advanced statistics

Other studies, such as that by Bhangu *et al*., used ROC curves to dichotomise continuous variables into their most significant/clinically relevant levels, and assigned them a score based on the odds ratios (ORs) from multivariate analysis[30]. Allowing a CRT model to use a portion of data in training to generate a rule that can then be validated on a test data set allows for an unbiased, higher accuracy classifier model. In this study the model was able to classify 75% of the test data into the correct outcome and a recent publication by Adams and Leveson suggests that rules generated by CRT methods are generally easily understood and translatable into everyday clinical practice[4].

The application of the prediction rule to the cohort data revealed that there was a clear increase in mortality with increased score, and that the rule was significantly better than chance (AUC= 0.754; P<0.001) at classifying patients who may be more at risk from death than others. The evaluation of this prediction rule (excluding the respiratory rate variable) in

the validation cohort by ROC curves, confirmed that the prediction rule was robust when tested on a new data set; even though a key variable was missing, as the prediction rule remained statistically significant even though the AUC values were reduced. The use of only four variables in this prediction rule as opposed to other studies, which use more variables[30], may be more desirable in a clinical setting, as it may save valuable time for the clinician who is providing treatment.

3.13.3 Predicting Mortality in Patients with *C. difficile* Infection

Mortality rates among CDI patients vary greatly in the literature, with rates between 4.7 % and 57% being reported[163]. These differences are representative of the way that deaths in cohorts of CDI patients are reported; ranging from deaths directly attributed to CDI, all-cause mortality, 30-day mortality and deaths post discharge. Differences may also be due to the underlying heterogeneity of the *C. difficile* infected cohort, who are generally aged, frail, and present with multiple co-morbidities. The overall mortality rate for the cohort in this study, was 24%, which is lower than the results published by McGowan *et al.*[163]; who report all-cause 30-day mortality rates between 30.7%-53.5%, within seven community trust hospitals between 2002-2008, and also lower than the study by Bhangu *et al.*[30], which reported an overall mortality rate of 38%. Although the all-cause mortality rate was lower compared to findings by others, CDI related mortality was 12.6 %. This rate is higher than a pooled mortality rate of 8.03%, reported in meta-analysis of other studies since 2000, although this was not UK-specific[120]. In this cohort, the increased rate of deaths attributed to CDI may be due to them being more accurately recorded in the specialised ward, coupled with potential under-reporting elsewhere. The specialist cohort ward setting may also be responsible for a lower all-cause mortality rate, and in turn may have promoted data selection bias, which could have influenced the data analysis procedure. However, the prediction rule derived in this study has been made applicable to patients from more general hospital environments, as

it has been validated on an independent cohort[30], and is shown to be consistent in the classification of patients with increased mortality risk in this cohort[30] (Table 3.7). Multinomial logistic regression analysis revealed that the four variables entered in the model were significant predictors of all-cause mortality (Table 3.3). It was also possible to distinguish two variables which were significantly associated with CDI related mortality rather than non-CDI related mortality. These two variables were; elevated WCC (P=0.025; OR 1.046, CI 1.006-1.008) in CDI related mortality vs. P=0.877 in non-CDI related mortality) and elevated CRP (P=0.020 OR 1.0009, CI 1.001-1.017) in CDI related mortality vs. P=0.074 in non-CDI related mortality). The small ORs for all these variables suggest that chances of mortality only increase slightly with increasing/decreasing levels (Table 3.4). However, small changes in ORs are also reflected in the findings of Bhangu *et al.*[30] and variable ORs are reported in the Bloomfield *et al.*[32] study. These small ORs could be due to the relatively small sample sizes from which the risk factors were derived, and/or because of exclusion of data by the analysis software. Despite the small ORs in this data set, the variables were seen to have a significant association to the increased risk of mortality, and thus were included for derivation of the prediction rule.

It is pertinent to note that whilst scores of zero in this study, and that of Bhangu *et al.*[30] are still associated with a higher mortality than that reflected in other clinically established prediction rules, such as, the CURB-65 rule[145]; where, scores of zero represent 0-0.5% mortality, but the actual mortality in the cohort was around 9.5%. The mortality rates for *C. difficile* cohorts used during this study were, 24% (derivation cohort) and 38% (validation cohort). In the Lim *et al*. study[145] the following mortality percentages are quoted for CURB-65 score (in the validation cohort) 0=0%, 1=0%, 2=8.3%, 3=21.4%, 4=26.3, 5=33.3%. Analysis of this result showed that a score of two or less results in a lower risk of mortality than the mean (9.34%) and those above two, an increased risk. In comparison, the prediction

rule score developed during the present study (in the validation cohort) is as follows; 0=20.9%, 1=37.1%, 2=54.3%, 3=66.7%. The same conclusions hold true for this approach, albeit for higher overall mortality rates, with points zero and one resulting in lower than average (38%) risk of mortality and points two and three demonstrating increased risk. Thus, whilst the CURB-65 score is undoubtedly more discriminatory at the lower end of mortality than this proposed approach, the characteristics of this study data which has a much higher mean risk of mortality, implies that the proposed rule is better at the higher end of mortality risk, which is perhaps more helpful to someone who is first attending a patient presenting with CDI. Finally, it should be noted that although the zero score has an attendant mortality rate that is significantly higher than 0%, it is also significantly lower than the actual mean mortality in both derivation and validation cohorts (9.5% vs. 24% in the derivation cohort and 20.8% vs. 38% in the validation cohort).

3.13.4 Clinical Relevance of Variables

The four variables; serum albumin (g/L) (P<0.001), CRP (mg/L) (P=0.034), WCC ($x10^3$ mcL) (P=0.049) and respiratory rate (resps/min) (P=0.002), were identified by univariate and multivariate analysis, as being significant predictors of mortality in patients with CDI (Table 3.3). Clinical measurements of all these variables could be readily taken at the time of CDI diagnosis, and are likely to be taken from any patient who is in hospital, making this scoring system very applicable in a clinical setting. Both this study and the study by Bhangu *et al*.[30] found serum albumin, WCC and CRP levels to be important risk factors of mortality in patients with CDI. The conclusions from a systematic review identifying significant risk factors associated with mortality from CDI[32] corroborate well with findings in this study, in that higher CRP, WCC and lower serum albumin were found to be associated with increased mortality, while high temperature, diarrhoea severity, presence of renal failure, diabetes, cancer and nasogastric use did not appear to be associated. A recent publication has indicated

that serum albumin, WCC and CRP are important prognostic variables for short term mortality in patients with CDI[225]. Other studies have found serum albumin to be a significant risk factor for CDI in nursing home patients[11] and Anthony *et al*. found that a fall in serum albumin level was consistent with the onset of CDI[14]. They also state that liver function testing for serum albumin is cheap (£1.54 at current prices, of which serum albumin is approximately £0.25) and quick to process (average turnaround is ~1.5 hrs from receipt in the laboratory). They conclude that as serum albumin is part of liver function testing, and that most patients have liver function tests along with renal tests on admission, it should be used as a prognostic factor of CDI. It is also worthy to note that both higher urea and lower serum albumin levels were significantly associated with mortality in the Bhangu *et al*. study[30]. Other studies have suggested elevated urea as a marker of risk of mortality, however, urea levels were not evaluated in this study, as emphasis was placed on the percentage increase in creatinine from a baseline reading as specified by the Department of Health report[5]. Elevated creatinine levels were not found to be a significant predictor of mortality in this study, which might be attributed to particular emphasis being placed on maintaining hydration in the patients on the cohort ward, while in other clinical settings patients at CDI diagnosis are not often managed by a specialist, at least initially. The role of urea will be re-evaluated in a prospective study to determine whether it adds statistical strength to the prediction rule.

Both CRP and WCC are clinical markers of infection and/or inflammation. CRP is an acute-phase reactant, and measurements of its levels are frequently used to aid in the diagnosis of bacterial infections[49]. CRP is synthesised by the liver, mainly in response to IL-6, which is produced not only during infection, but also in many types of inflammation[198] and may indicate why this protein is overly elevated in CDI patients. WCC levels are also used as a marker of inflammation and bacterial infection[56]. Thus, the results found during this study are supported by the literature which indicates that both CRP and WCC levels are elevated in

patients with CDI, and are therefore may be more significantly associated with mortality due to CDI than mortality from other causes. There may be no certainty that an increase/decrease in clinical variables such as respiratory rate, WCC, CRP and serum albumin are specifically due to CDI; as these patients are usually older, and have multiple co-morbidities. However, it is generally observed that these markers have usually returned to baseline levels, before a later rise, which occurs around the time of *C. difficile* diagnosis, which could be up to 1-2 weeks after cessation of antibiotic treatment for a previous condition. Thus, an acute rise/decline in these markers, around the time of infection diagnosis may be generally attributed to *C. difficile* infection. Furthermore, the combination of variables significantly associated specifically with CDI-related mortality (WCC and CRP) with variables shown to be associated with all-cause mortality outcomes; elevated respiratory rate (which is also known to be associated with increased mortality in patients with community acquired pneumonia[145]) and low serum albumin, provide a strong basis for being able to identify CDI patients with increased risk of mortality.

The outcome of the CRT model for test data found that levels of serum albumin (g/L) ≤ 24.5, CRP (mg/L) >228 , respiratory rate (resps/min) >17 and WCC ($x10^3$ mcL) >12 were the most effective at classifying cases into the groups of the all-cause mortality outcome. Levels of serum albumin <20-30 g/L have been reported as indicators of CDI, as well as severity of infection[11, 30, 231] and the levels found in this study corroborate well with these published levels. CRP levels in healthy individual are normally less <10 mg/L; however, in disease states, this level increases in the first 6 to 8 hrs of infection, and can reach peak levels approaching 350–400 mg/L, after approximately 48 hrs, in response to acute active inflammation or bacterial infection[49]. This study identified a CRP threshold of >228 mg/L, which fits in well with these published measurements of active infection and/or inflammation. Rules from the model identified serum albumin ≤ 24.5 (g/L) or CRP (mg/L)

>228 as important predictors of mortality, thus giving strong evidence for equal use of these variables in a scoring system for risk of mortality in patients with CDI. The model rules also found serum albumin (> 24.5 g/L) or CRP ≤ 228 (mg/L) along with a combined increase in WCC (>12 $x10^3$ mcL) and increased respiratory rate (>17 resps/min) were all associated with the increased probability of death as an outcome, thus the variables respiratory rate (resps/min)>17 and WCC ($x10^3$ mcL)>12 were assigned one point together. The clear increase in mortality with increased score (Table 3.7) shows that this rule classifies patients into those who may be more at risk than others. Also, as only four variables were used in this model as opposed to that of Bhangu *et al.*[30] who used six, is more applicable in a clinical setting as it may save valuable time for the clinician who is providing treatment.

Within the clinical setting, being able to simply and accurately identify CDI patients who are at increased risk from death will facilitate clinical teams with therapeutic choice[22] and aid communication with individual patients and relatives.

3.13.5 Discussion of Recurrent *C. difficile* Infection Data Analysis

No significant variables were found associated with recurrent CDI after Bonferroni correction. This may have resulted from the small sample population that actually had true recurrent disease (N=16), and not just on going infection, or a second infection after 30 days of stool resolution from the first infection. Of the 16 patients who had true recurrent CDI, 14 patients relapsed, and five had an infection with another ribotype. Detailed exploration of the data associated with each case of recurrent CDI revealed that 6/16 patients with true recurrent CDI had additional antibiotics listed in case notes between the index case of CDI and the subsequent infections. Five of the six patients with additional antibiotics listed were discharged 30 days prior to re-admission for the recurrent CDI episode. It may be possible that these patients were prescribed antibiotics in the community, and this was responsible for the secondary infection. Accounting for the six cases where the possible re-administration of

antibiotics caused recurrent infection, there were only 10 cases where re-administration of antibiotics was not the probable cause of recurrent CDI. Thus, this might play an important part in why it was not possible to find any useful variables associated with recurrent CDI. The lack of variables associated with recurrent CDI has also been demonstrated in a study by Choi *et al*. which found that the only independent predictor of recurrent CDI in a cohort of 84 was co-infection with vancomycin resistant *enterococci* (VRE), despite looking at patients on gastric acid suppressors, and also other variables similar to this study such as, WCC, serum albumin levels and CRP levels[48]. Other studies have implied that the presence of certain *C. difficile* strains may also be associated with recurrent CDI[102, 206]. Ribotype association to infection outcomes will be discussed further in Part Two of this thesis.

In order to find significant host variables for predicting recurrent CDI, studies would benefit from much larger cohorts, as exemplified by the study of Eyre *et al*.[67]. Their study was based on data from 1678 patients, with 363 (22%) experiencing recurrent CDI. They found that increasing age, previous total hours admitted to hospital and CRP levels at first CDI, were all associated with increased recurrent CDI, which was assessed as a 4-month recurrence risk[67].

With the advent of next generation sequencing it is also not surprising that some researchers have been looking into host genetic markers for recurrent CDI, along with clinically measurable variables. In 2010, Garey *et al*. reported that that a common single nucleotide polymorphism (SNP) in the IL-8 gene promoter was an independent predictor of recurrent CDI[76]. They concluded that in a prospective study cohort of 96 patients with CDI, 10 (39%) out of 26 patients with an A/A genotype in the IL-8 promoter gene experienced recurrent CDI, compared with 13 (19%) out of 70 patients with an A/T or T/T genotype (OR, 2.7; 95% CI, 1.01–7.4) (P=0.043). They also proposed that these results coupled with therapies to prevent recurrent CDI, such as monoclonal antibody therapy or antibiotic therapy with

rifaximin, fidaxomicin, tapered or pulsed vancomycin, or probiotics could serve the basis of future research into this clinically problematic outcome.

3.14 Summary

This large retrospective cohort study has identified four easily measurable clinical variables: serum albumin level ≤ 24.5 g/L, CRP level >228 mg/L and combination of WCC >12 $x10^3$ mcL and respiratory rate >17 resps/min, that are capable of predicting the risk of mortality in patients with CDI. The prediction rule is simple, accurate and robust and the variables are not defined by other parameters, which could further complicate its use. The prediction rule has been validated through an internal split sample procedure, and on an independent cohort, thus making it applicable to patients from more general clinical environments.

No variables were found to be significantly associated with recurrent CDI. Defining a robust clinical prediction rule for recurrent infection is perhaps limited due to the small sample sizes. In conjunction with increasing sample sizes to identify strong predictors of recurrent disease, there may be other/novel host factors, and/or bacterial factors that might influence the outcome of an infection. Identifying potential novel biomarkers for outcomes of CDI is something that has been explored in this study, and will be discussed in the next chapter.

Chapter Four. Investigating the use of Calprotectin as a Novel Host Marker of *C. difficile* Infection Outcomes

4.1 Introduction

Chapter Three concluded that there were no routine clinical markers that could be used as predictors of recurrent infection. However, prediction of severity, recurrence and mortality would provide useful information to guide management of CDI affected patients[168]. Therefore, the discovery of novel biomarkers for these infection outcomes might be clinically useful. As highlighted in Chapter Two, a clinically useful biomarker which has been implemented in assessing levels of gastrointestinal inflammation is calprotectin. CDI is associated with an inflammatory response in the colon, and thus the potential role of calprotectin as marker for CDI outcomes was investigated.

Aim

Faecal calprotectin (FC) is a strong indicator of the inflammatory status of the intestinal mucosa, and assessing faecal calprotectin provides a non-invasive test which is more robust than determining faecal leukocyte counts[51]. Thus, all available stools were tested for levels of FC using the calprotectin ELISA (Materials and Methods, Chapter 10, section 10.8) to determine whether calprotectin is a) elevated in patients with CDI compared to published calprotectin levels, and b) if higher levels are detected in patients with recurrent disease, compared to those who only had one infection. Additionally, differences in median calprotectin levels were statistically tested for groups with other outcome measures such as severity of infection and all-cause mortality (which includes patients who died from CDI and non-CDI related causes). Calprotectin levels found in this study could only be compared to published ranges, as faecal samples were retrospectively collected, and there was no control group included in the initial sample collection.

Method

The outcome measures defined for this cohort were severity, recurrence and all-cause mortality. Clinical information was retrospectively collected from the patient notes to enable severity scores to be calculated based on the UK Department of Health (DoH) scoring system[5]. The laboratory parameters that were recorded are shown in Table 4.1 and were those recorded nearest to the time of collection of the CDI positive stool.

Table 4.1- Laboratory parameters recorded.

C-reactive protein
White blood count
Neutrophils
Creatinine (baseline)
Creatinine (at CDI diagnosis)
Platelets
Albumin

Faecal samples were stored at -20°C. Patients without a stored stool sample and those with a low volume sample of < 50 mg were excluded.

4.2 Results

4.2.1 Patient Demographics

A total of 245 cases were collected for statistical analysis, which equated to 213 patients, due to duplicate entries for patients with recurrent CDI. Following the exclusion criteria listed above, 168 cases were available for statistical analysis. Patient demographics and descriptive statistics for this study cohort are presented in Table 4.2 and 4.3 respectively. The overall median level of calprotectin in faeces of this cohort was 1224.46 µg/g.

Table 4.2- Patient demographics for cases available for statistical analysis.

Patient Demographics		Frequency and %	N
Gender	Male	70 (47%)	
	Female	78 (53%)	148
Mortality Outcome	Died	37 (22%)	
	Survived	131 (78%)	168
Score For Severity	Mild	56 (33.3%)	
	Moderate	22 (13.1%)	
	Severe	78 (46.4%)	
	Life Threatening	12 (7.1%)	168
Number of Episodes	One episode	121 (81%)	
	Multiple Episodes	29 (19%)	150

Table 4.3– Descriptive statistics for variables.

	N	Minimum	Maximum	Mean	Median	Std. Deviation
Age on Admission (yrs)	168	21	100	77.55		13.393
Length of Stay in hospital	168	3	138	38.10		28.506
Stool Calprotectin Levels	168	42	12989	1912.16	1244.46	2110.104

4.2.2 Investigating Calprotectin levels in Patients with Primary only and Recurrent *C. difficile* Infection

121 (72%) cases presented with one infection and 47 (28%) cases presented with recurrent/ multiple infections. Of the 47 cases of multiple infections there were a total of 29 patients, 26 of those having 2 episodes of CDI and 3 having 3 episodes. The median calprotectin level for cases in which the outcome was only one episode of CDI was 1176.23 µg/g, and for cases in which the outcome was multiple episodes of CDI was 1282.26 µg/g. The distribution of calprotectin levels in patients with primary only and recurrent CDI is presented in Figure 4.1. Independent samples median tests revealed that median FC levels were not significantly different across categories of primary infection and recurrence (P=0.864; χ^2=0.030; *df*=1; N=168).

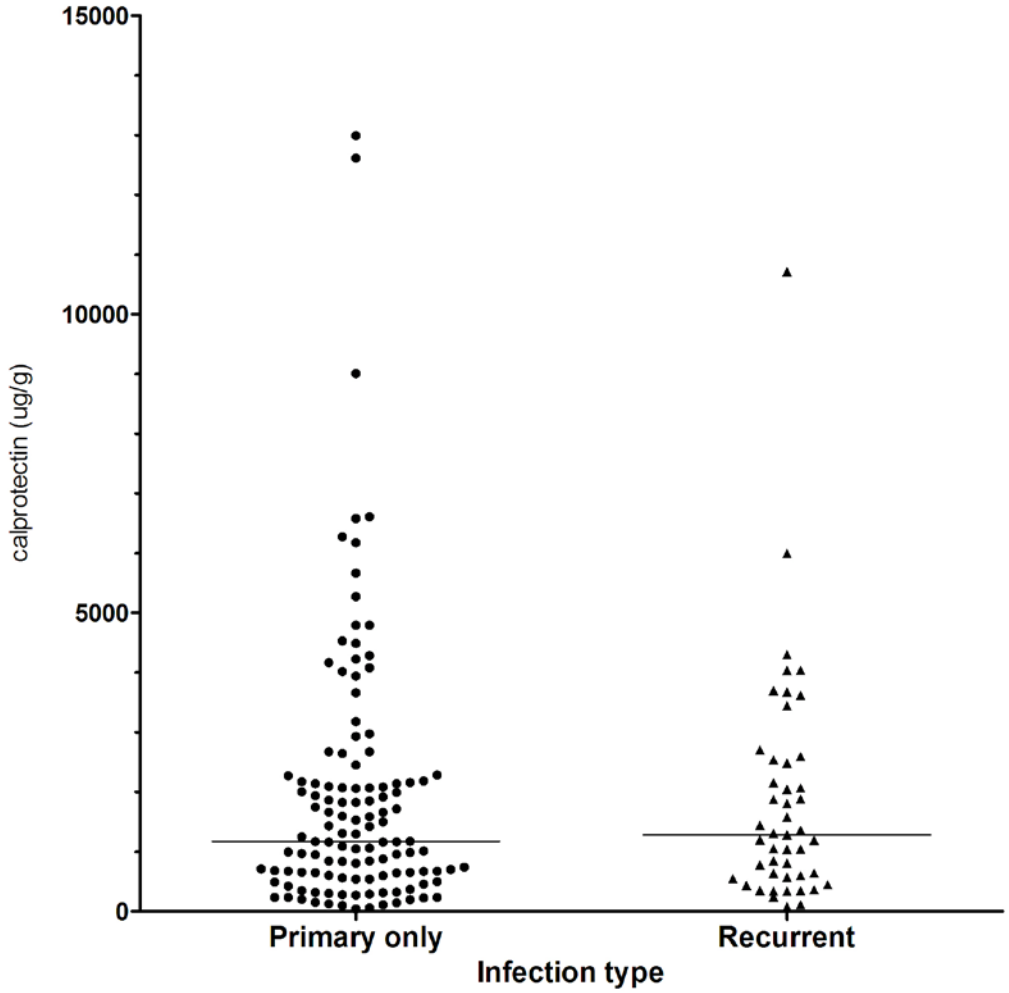


Figure 4.1- Levels of calprotectin within different infection outcome groups. Bars represent median calprotectin levels.

4.2.3 Investigating Calprotectin levels in Patients with Varying Severity of *C. difficile* Infection

Of the 168 cases included in this study 56 (33.3%) presented with mild CDI, 22 (13.1%) presented with moderate CDI, 78 (46.4 %) presented with severe CDI and 12 (7.1%) presented with life threatening CDI, according to the DoH scoring system (Table 4.1). Median calprotectin levels for each group in the severity outcome are presented in Table 4.4 below.

Table 4.4- Number of cases and median stool calprotectin levels for groups of patients in the severity outcome.

Score for Severity of Disease	N	Median Calprotectin levels (µg/g)
Mild	56	1172.67
Moderate	22	1839.55
Severe	78	1161.49
Life Threatening	12	1315.64
Total	168	1224.46

Independent samples median tests revealed that median FC levels were not significantly different across categories of severity, as classified by the DoH scorecard for severity of CDI (P=0.308; χ^2=3.6; *df*=3; N=168).The distribution of calprotectin levels in patients with different severity grading is presented in Figure 4.2.

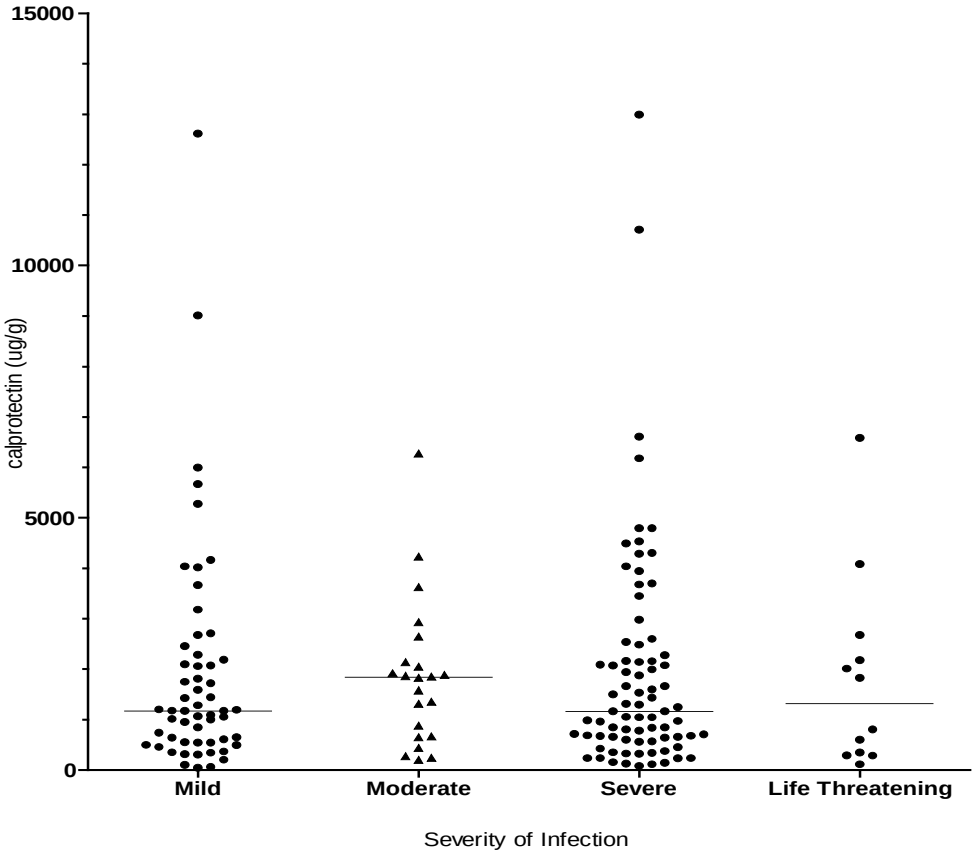


Figure 4.2 - Levels of calprotectin within different severity groups. Bars represent median calprotectin levels.

4.2.4 Investigating Calprotectin Levels in Patients who Died versus those who Survived

Due to the small samples sizes for cases of CDI related and non-CDI cases within this cohort, these categories were merged into one group to determine calprotectin levels. From the 168 cases available in this study, there were 131 (78%) cases survival and 37 (22%) cases of death. The median calprotectin level for cases in which death was the outcome was 1166.73 µg/g and for those who survived was 1247.69 µg/g. Independent samples median tests revealed that median FC levels were not significantly different across categories of all-cause mortality (P=0.852; χ^2=0.035; *df*=1; N=168). The distribution of calprotectin levels across the all-cause mortality outcome is presented in Figure 4.3.

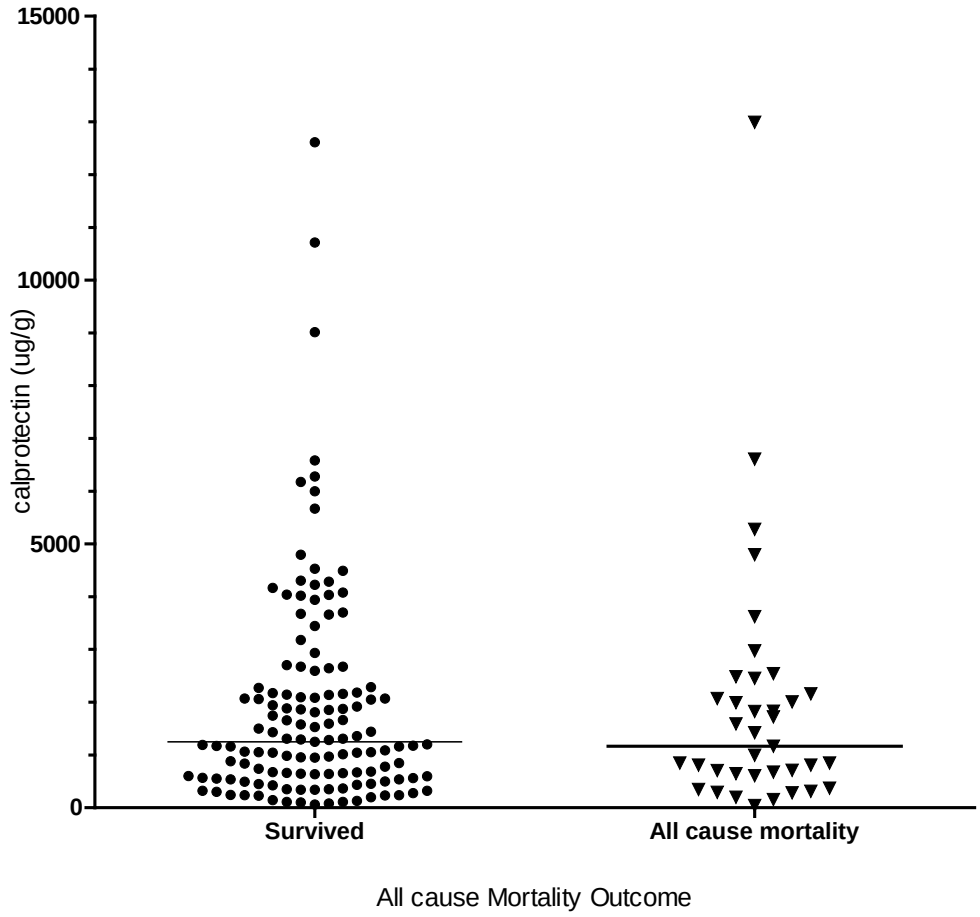


Figure 4.3- Levels of calprotectin within different mortality outcome groups. Bars represent median calprotectin levels.

4.2.5 Results for Analysis of Stool Calprotectin Levels with Respect to Department of Health Scorecard Measures.

Although calprotectin levels were not significantly different across the groups for each outcome, it is known to be a marker of inflammation, especially in IBD patients[80]. Bivariate correlation was used to analyse FC levels against the variables; CRP, WCC, neutrophils, creatinine (as a percent rise in baseline at CDI diagnosis), platelets and serum albumin, as these measurements are taken at the time nearest to a *C. difficile* toxin positive stool sample; with the aim of identifying if FC was correlated to any variables which are markers of clinical infection. Pearson correlation indicated that WCC was positively correlated (R^2= 0.322) with FC levels (P=0.05; N=166), but FC was not significantly correlated with any other variable after Bonferroni correction.

4.2.6 Testing Calprotectin Levels Against Stool Storage Duration

A potential limitation of this study is that FC extraction was performed on stools that were stored at -20^{o}C for a prolonged period. There is evidence to show that FC is stable for one year with storage at -20^{o}C[214], but no data is available for longer storage periods.

Approximate storage time in years was taken from the time that *C. difficile* toxin A/B positive stool samples were stored, to when the assay was performed. Approximate duration of storage of faecal samples analysed in this study is presented in Table 4.5, where storage temperature was -20^{o}C. Median calprotectin levels for each duration period are presented in Table 4.6.

Table 4.5- Approximate duration of storage of stool samples.

Duration of Stool storage (yrs)	Frequency	Percent	Valid Percent	Cumulative Percent
One Year	29	11.8	17.3	17.3
Two Years	63	25.7	37.5	54.8
Three Years	76	31.0	45.2	100.0
Total	168	68.6	100.0	

Table 4.6- Median calprotectin levels for storage periods.

Duration of Stool storage (yrs)	N	Median Stool Calprotectin Levels (µg/g)	Grouped Median
One Year	29	1294.84	1294.84
Two Years	63	1176.23	1176.23
Three Years	76	1241.74	1241.74
Total	168	1224.46	1224.46

Independent median tests were used to assess whether there were any significant differences in median calprotectin levels for each duration period, from the grand median (1224.46 µg/g). The tests revealed that there were no significant differences in calprotectin levels across the groups of different storage duration (P=0.975; *df*=2; N=168).

4.3 Discussion

As this was a retrospective study, some stool samples would have been collected three years prior to analysis with the calprotectin ELISA assay. Thus, a potential limitation of this study is reflected in reports that propose calprotectin is not stable in stool samples that have been frozen for long periods of time (email communication with Dr Ray Sheridan, Royal Devon and Exeter Foundation Trust). A study by Ton *et al.*[214] tested the stability of calprotectin in 50–100 mg of stool samples that were stored for six months at -20^{o}C, and concluded that there were no significant changes in calprotectin levels before and after storage (P=0.16). Independent median tests of levels of calprotectin in stool samples that were stored at -20^{o}C for different time periods during the present study, revealed that there were no significant differences (P=0.975) across the groups, therefore duration of storage is not likely to have affected calprotectin levels in this study, and corroborates findings by Ton *et al.*[214]. However, this study was not able to determine the effect of freeze/thaw cycles on calprotectin levels which may also have had an effect on calprotectin levels in faeces at the time of analysis[129].

The lack of statistical difference in calprotectin levels between groups in each outcome (Sections 4.4, 4.5 and 4.6); suggests that it would not be useful as a prognostic marker for

outcomes such as severity, recurrence and mortality. Ranges of calprotectin levels in stool samples measured by the Calpro Calprotectin ELISA Test (ALP, CALPRO AS, Norway) have been reported by Johne *et al.*[110] and Røseth *et al.*[192]. These studies suggest that calprotectin levels considered as 'normal' fall between 5–50 µg/g and that elevated levels are considered as those which are >50 µg/g. Median values in patients with symptomatic colorectal cancers are reported to be around 350 µg/g and levels in active, symptomatic inflammatory bowel disease range between 200–40,000 µg/g. In this study, it is interesting that while there were no significant differences in median calprotectin levels between groups in each outcome, the overall median calprotectin level in the cohort was 1224.46 µg/g, and corresponds to a level indicative of active, symptomatic inflammatory bowel disease, when compared to the findings of Johne *et al.*[110] and Røseth *et al.*[192] highlighted above.

C. difficile is known to induce an inflammatory response in the host[123, 181]. Thus, the high median calprotectin level found within this cohort may indicate that patients may have active infection with *C. difficile*. However, lack of a significant difference in calprotectin levels across the outcome groups, may mean there are other factors, such as the host immune system, or acquisition of a particular *C. difficile* strain, which contribute significantly to the outcome of infection.

The idea that calprotectin may be of additional use to identify those with CDI, is supported by data in Section 4.7, which shows that calprotectin levels are correlated with WCC. Increased WCC has also been associated with onset of CDI[186] and outcomes of infection (Chapters 1-3). Gisbert *et al.* also note that calprotectin is useful in discriminating between organic and functional intestinal diseases[81], thus, its correlation with the level of active inflammation/disease, may be useful for identifying those with CDI, and not other functional digestive diseases.

The correlation of increased calprotectin levels and onset of CDI is however, only speculation. As this study was retrospective, one of its limitations was that there were no control groups to use for data comparison.

4.3.1 Conclusion

As this was a retrospective study, caution must be used when making a presumption about using calprotectin as an indicator of inflammatory response to *C. difficile*; as it is found elevated in many clinical disease manifestations such as neoplasia, IBD, general bacterial infections and also in those who are on non-steroidal anti-inflammatory drugs[68]. Only results tallied with detailed patient history could give a truly representative indication of the predictive ability of calprotectin to distinguish those with CDI from those without. To further clarify the potential use of calprotectin as a marker of CDI, a prospective cohort study that looked at calprotectin levels in healthy individuals (healthy controls), those with CDI in the absence of any co-morbidity, such as, IBD, IBS or cancer, and those with a combination of CDI and another inflammatory conditions, would need to be conducted.

4.4 Part One Summary

- Four variables were significant predictors of mortality in patients with CDI; serum albumin (g/L) ($P<0.001$), CRP (mg/L) ($P=0.034$), WCC ($x10^3$ mcL) ($P=0.049$ and respiratory rate (resps/min) ($P=0.002$), and could be used to generate a prediction rule for classifying those more at risk of mortality.

- Statistical analysis of cohort data was unable to deduce significant variables prognostic of recurrent CDI.

- Faecal calprotectin was elevated in patients with CDI compared to published literature, but median levels were not significantly different across any infection outcome.

As there were so few variables associated with mortality and recurrent CDI, there may be other host and/or bacterial factors that are playing a role in the outcome of an infection. Some studies have suggested that the ribotype causing CDI may play a role in the outcome of infection[58, 102, 175, 206]. The emergence of the 'hypervirulent' BI/NAPI/027 strain, which has been associated with increased mortality[175], severity[73] and recurrence, has led researchers to focus on the association of the acquisition of this particular ribotype to infections outcomes. However, the prevalence of ribotypes across the globe is changing, and thus some researchers have begun to assess the contribution of other ribotypes to the outcome of infection[58, 83, 206]. The changing epidemiology of ribotypes and their association to outcomes of CDI such as, mortality and recurrence will be discussed in Part Two of this thesis.

Part Two
Assessing *C. difficile* Ribotype Epidemiology and Association with Outcomes of Infection

Chapter Five. A Review of *C. difficile* Epidemiology and Emerging *C. difficile* Ribotypes

5.1 Introduction

Part One of this thesis highlighted the importance of identifying host markers of infection for outcomes such as, mortality and recurrent CDI. Over 186 host variables were analysed during this study, so it is somewhat surprising that only four variables were found to be prognostic markers of mortality, and no markers were found to be prognostic of recurrent infection. The identification of only a small number of host variables associated with these infection outcomes may suggest that there are bacterial factors which may be associated with the outcome of CDI. Typing of *C. difficile* strains that have caused an infection has increased since 2007, when an epidemic 'hypervirulent' BI/NAPI/027 ribotype very rapidly disseminated worldwide and was associated with more severe infection outcomes[73]. This has led researchers to focus on assessing ribotype epidemiology in order to identify strains that could cause more severe infection outcomes.

The diversity of ribotypes within the clinical setting is changing, and there are ribotypes emerging which are associated with severe clinical outcomes, similar to the 'hypervirulent' BI/NAPI/027 ribotype[42, 58, 114]. This chapter will highlight the changing ribotype epidemiology at a global and regional level and introduce the emergent ribotypes that have been linked to more severe clinical outcomes of infection.

5.2 Global Epidemiology of *C. difficile*

There has been a recent focus on the need to assess the epidemiology of *C. difficile* ribotypes on both national and international scales, due to the emergence of epidemic strains. The particular importance of tracking ribotype epidemiology is highlighted by the creation of the European *Clostridium difficile* Infection Study (ECDIS) by the European Centre of Disease

and Prevention and Control (ECDC), whose aim is to enable enhanced CDI detection and ribotype surveillance across Europe[28], and Figure 5.1 represents the European wide distribution of ribotypes during 2008.

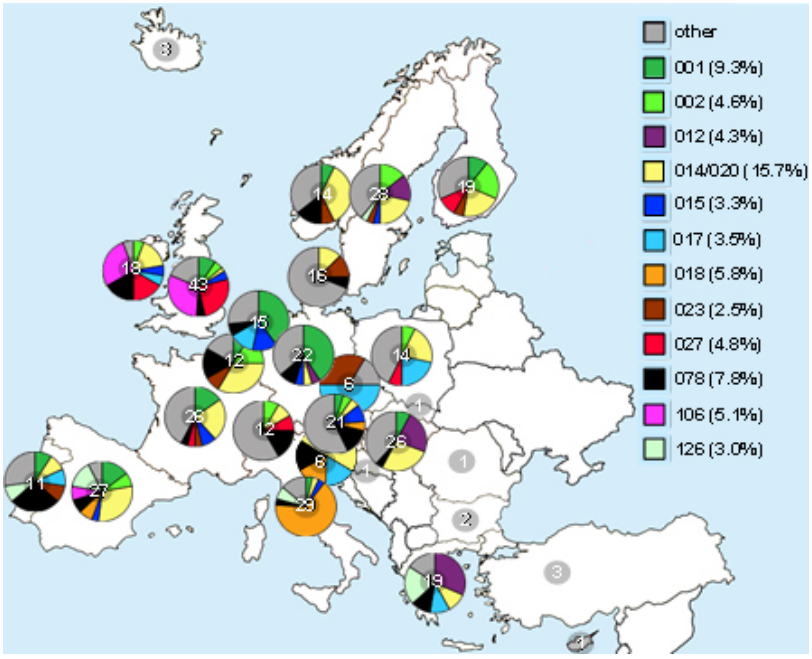


Figure 5.1- Geographic distribution of *Clostridium difficile* PCR ribotypes in European countries with more than five typable isolates, as of November 2008. Pie charts show proportion of most frequent PCR ribotypes per country. The number in the centre of the pie chart is the number of typed isolates in this country[28].

A study by Cheknis *et al.*[46] recorded the distribution of strains from a North American, European and Australian trial for treatment of CDI between 2005 and 2007. 775 isolates were screened, and 50% of theses strains consisted of three REA groups; groups BI (Ribotype 027), J (001) and Y (Ribotypes 014/020/220/221), with 24 % of isolates being accounted for by REA group BI (Ribotype 027). In Europe, the most common type was J (Ribotype 001). Surprisingly in their study, ribotypes not found in the seven REA types investigated (BI, J, Y, K, G, BF and CF) made up 67% of the Australian ribotypes. The difference in global distribution of ribotypes emphasises the need to not only track ribotype epidemiology, but to also identify what selective pressures are driving this ribotype diversity.

In 2007, the Health Protection Agency (HPA) formulated the *Clostridium difficile* Ribotyping Network (CDRN) to track the changes in ribotype frequencies across the England and Northern Ireland (NI), in order that it may help try to prevent future *C. difficile* epidemics, and provide timely data of ribotypes to infection prevention teams. This could then enable targeted interventions and resources to high-incidence CDI settings[229]. This network is comprised of eight regional laboratories which routinely type and assess antimicrobial susceptibility of *C. difficile*. Figure 5.2 represents the prevalence of *C. difficile* ribotypes in England between 2007-2010 as reported by the HPA[6]. They report that the top ten ribotypes circulating are, 027, 106, 001, 002, 104/020, 015 078, 005, 023 and 026. However, the frequency of each ribotype can vary from one year to the next, and also from region to region, as exemplified by the regional pattern of the South West (Figure 5.3). The establishment of both national and international surveillance groups emphasises the importance of global and regional ribotype surveillance to facilitate the detection and control of *C. difficile* infection.

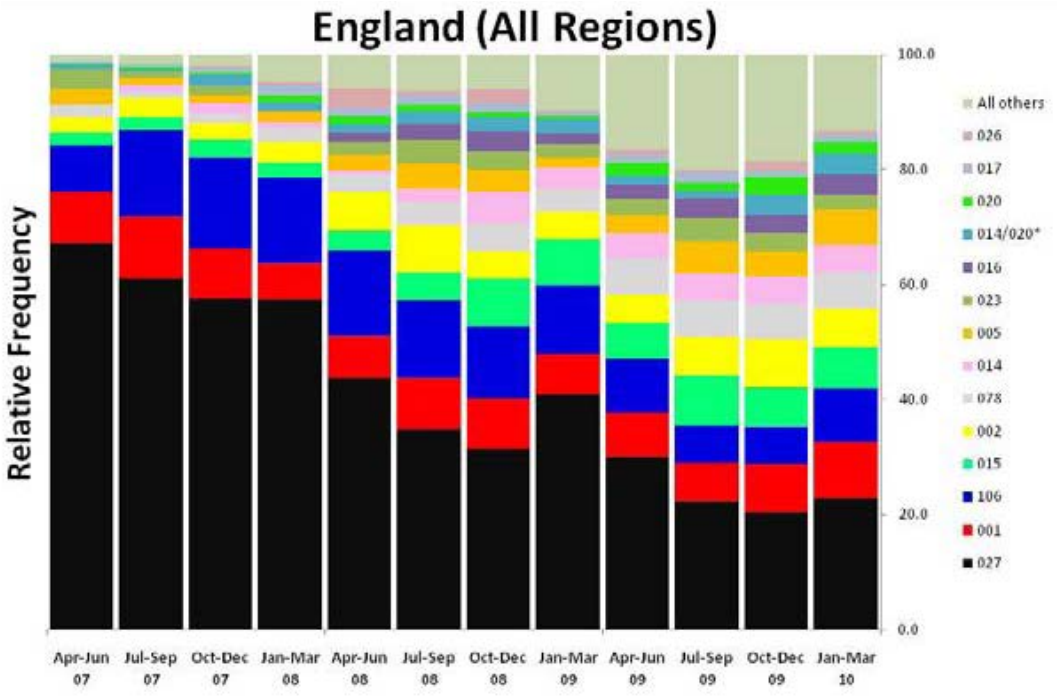


Figure 5.2 - Distribution of *C. difficile* ribotypes in England from 2007-2010[6].

5.3 Distribution of Ribotypes in the South West

Section 5.2 highlighted the prevalence of particular ribotypes within England, with the caveat that these could vary from region to region. Figure 5.3, from the CDRN 2009/2010 report[6] shows the ribotype distribution for the South West region from April 2007 to March 2010 and also highlights the increased diversity of ribotypes from one year to the next. At the time of this report the ten most common ribotypes distributed in the South West were 027, 001, 106, 015 002, 078, 014, 005, 023 and 020.

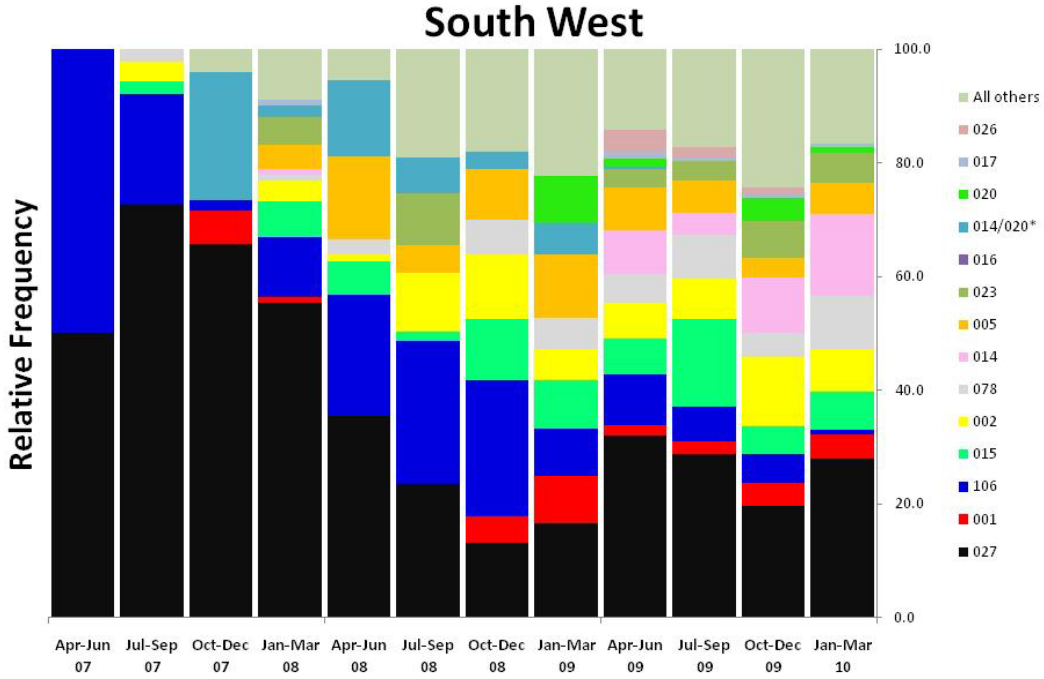


Figure 5.3- Distribution of *C. difficile* ribotypes in the South West of England, from 2007-2010[6].

Whilst the 'hypervirulent' BI/NAPI/027 ribotype still remains most prevalent, there has been a shift in ribotype distribution across England (Table 5.1) which shows an overall decrease in its prevalence, and an increase in other ribotypes such as 002, 015, 078, 005, 023 and 026, over a 3 yr period. It is clear that there are factors driving this ribotype diversity and the next section will highlight the impact that some of these emergent ribotypes are having on the outcome of CDI.

Table 5.1- Changing prevalence of most common (top 10) *C. difficile* ribotypes detected by CDRN in 2007/08, 2008/09 and 2009/10[6].

Ribotype	2007/08 (n,%)	2008/09 (n,%)	2009/10 (n,%)	Prevalence change [07/08 and 09/10](%)
027	1152 (55.3)	1468 (36.1)	1102 (22.1)	- 33.2
106	270 (13.0)	517 (12.7)	364 (7.3)	- 5.7
001	181 (8.7)	297 (7.3)	371 (7.4)	- 1.3
002	57 (2.7)	231 (5.7)	302 (6.0)	+ 3.3
014/020	57 (2.8)	218 (5.4)	128 (2.6)	- 0.2
015	50 (2.4)	215 (5.3)	330 (6.6)	+ 4.2
078	37 (1.8)	144 (3.5)	285 (5.7)	+ 3.9
005	29 (1.4)	118 (2.9)	213 (4.3)	+ 2.9
023	21 (1.0)	109 (2.7)	149 (3.0)	+ 2.0
026	5 (0.2)	87 (2.1)	41 (0.8)	+ 0.6

5.4 Clinical Relevance of Emerging Types

The previous sections have indicated that the prevalence of certain *C. difficile* ribotypes is changing. The increasing drive to type *C. difficile* isolates is due to the fact that some strains possess virulence mechanisms which give them the potential to cause severe clinical disease. Changes in prevalence of ribotypes have had serious implications on the general population, as was exemplified by sudden the increase in prevalence of ribotype 027, which was facilitated by its resistance to the fluoroquinolone antibiotics[93]. The dissemination of ribotype 027 resulted in a worldwide epidemic between 2005 and 2008, and acquisition of this ribotype was associated with increased mortality, severity and recurrent infection[73]. Recently the increase in prevalence of ribotype 078 has also gathered more research interest, as this ribotype is capable of infecting both human and animal hosts[106], thus increasing the pool of infection sources. Goorhuis *et al*. reported that isolation of ribotype 078 from patients in the Netherlands has increased from 3% to 13% between 2005-2008, and that while 078 and 027 ribotypes present with similar severity, 078 ribotypes were isolated from younger patients, and cases were more associated with community acquisition[83]. The community acquired CDI association could be linked to the fact that this ribotype is commonly isolated from animal

reservoirs and that transmission from one reservoir to another is very likely[114]. Although *C. difficile* is predominantly a HCAI, there is increasing concern over community acquired infection, and more specifically those who are asymptomatic carriers, who then come into the healthcare environment and facilitate the transmission of infection. Developing a better knowledge of non-clinical infection sources, such as those in elderly care homes and asymptomatic carriers, will perhaps facilitate better infection control and also provide a more comprehensive understanding of the role that community acquired CDI has on the emergence of different ribotypes[114].

MLST analysis of *C. difficile* isolates by Stabler *et al*.[203] has found five distinct clades of *C. difficile*, two of which contain ribotypes already seen as emerging strains (027 and 078) and two of which contain ribotypes very diverse from other clades. One of these clades contains ribotype 017, which is a natural $TcdA^{-}TcdB^{+}$ variant, and the other contains ribotype 023; prevalence of both these ribotypes is increasing across Europe, and both have been associated with more severe CDI[42, 58]. A recent study by Stewart *et al*. has also proposed that ribotypes which produce CDT, such as 027, 023 and 078, are also associated with recurrent infection[206].

The recent recognition of *C. difficile* as a prominent HCAI has led to research which has focused on either evaluating the 'hypervirulent' BI/NAPI/027 ribotype association to particular infection outcomes, or assessing geographic ribotype distributions. As such, there has been a shift away from focusing on other ribotypes that may have the potential to cause more severe outcomes of infection.

The prevalence of these emerging ribotypes is likely to be attributable to a variety of genetic factors that increase transmissibility, their ability to survive, and be resistant to the most commonly used treatments for other hospital infections. Therefore, comparative studies of

these ribotypes may contribute to our understanding of how these virulence mechanisms are associated with increased virulence and poorer clinical outcomes[22].

5.5 Aim

Statistical analysis of clinical data in Part One of this thesis was used to deduce if there were any host markers of CDI infection outcomes such as, recurrent CDI and mortality. Four variables were found to be host prognostic markers of mortality but none were found to be prognostic recurrent infection. Clearly, there are other factors contributing to the outcome of infection, and one of these factors could be the *C. difficile* ribotype causing infection. This chapter has provided evidence that the epidemiology of *C. difficile* ribotypes is changing, and that there are ribotypes emerging which have the potential to cause more adverse clinical outcomes. The aim of work conducted in Chapter Six of this thesis is to assess the distribution of ribotypes within the study cohort defined in Chapter Four, and to evaluate whether there is any ribotype association to CDI outcomes such as, mortality and recurrent infection.

Chapter Six. Cohort Ribotype Distribution and Association with the Infections Outcomes; Recurrence and Mortality

6.1 Introduction and Aim

Statistical analysis of cohort data in Chapters Three and Four demonstrated that there were four host variables which were significant predictors of mortality, but none were found to be significant predictors of recurrent CDI. Chapter Five highlighted the importance of surveying the distribution of specific *C. difficile* ribotypes across the globe, to track the changing ribotype epidemiology and emergence of possible epidemic *C. difficile* ribotypes, that may be associated with increased severity, mortality or recurrent CDI[42]. However, there have been relatively few studies which have investigated both host and ribotype association to CDI outcomes[84, 199, 208, 224]. Given that in this study, there were so few host markers of infection for mortality and no markers of recurrent CDI, the aim of work in this chapter is to assess which ribotypes are predominant in the study cohort, and assess whether acquisition of a particular ribotype is also associated with the outcome of a patient with CDI.

6.2 Results

6.2.1 Ribotype Distribution for the RD&E Cohort

In order to determine if there was any ribotype association to infection outcomes, such as mortality and/or recurrent CDI, *C. difficile* was first isolated from 226 stool samples; collected from patients with CDI, and ribotyped. PCR reactions were analysed by agarose gel electrophoresis as outlined in the Materials and Methods chapter (Chapter Ten, Section 10.7). Gel images of *C. difficile* banding patterns were analysed using the Bionumerics software (Applied Maths) and a representative gel is presented in Figure 6.1.

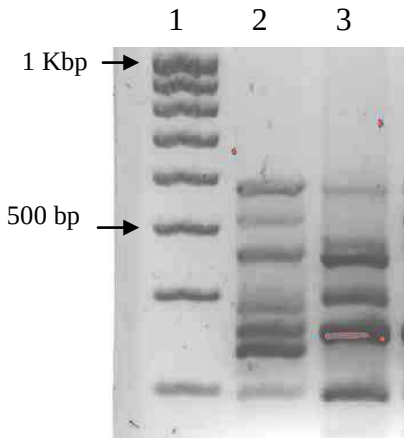


Figure 6.1- Representative gel image of the ribotype patterns that were produced for all isolates in the study cohort. 1) SupperLadder Low (100 bp), 2) Ribotype pattern of *C. difficile* 630Δ*erm* (012), 3) Ribotype pattern of *C. difficile* R20291 (027).

Total cohort ribotype distribution is presented in Figure 6.2. A total of 28 ribotypes were found among cases of this cohort, 14 of which were found in cases of recurrent CDI. Ribotype 027 (N=57) and ribotype 012 (N=51) were the two most predominant types within this cohort for the study period (2007-2009). Of the top ten ribotypes that are predominant in England and Northern Ireland (NI) (Chapter Five, Table 5.1) eight (027, 001, 002, 014/020, 015, 078, 005 & 023) are found in the top ten ribotypes of the RD&E cohort. In the 2009/2010 CDRN report[6] ribotypes 106 and 026 are in the top ten ribotypes predominant in England and NI, however, in the study cohort, ribotype 106 was only found once, and ribotype 026 was not found at all. Ribotype 012 was also predominant in the RD&E cohort, but it is not in the top ten predominant ribotypes at a national level, nor is it a South West specific isolate

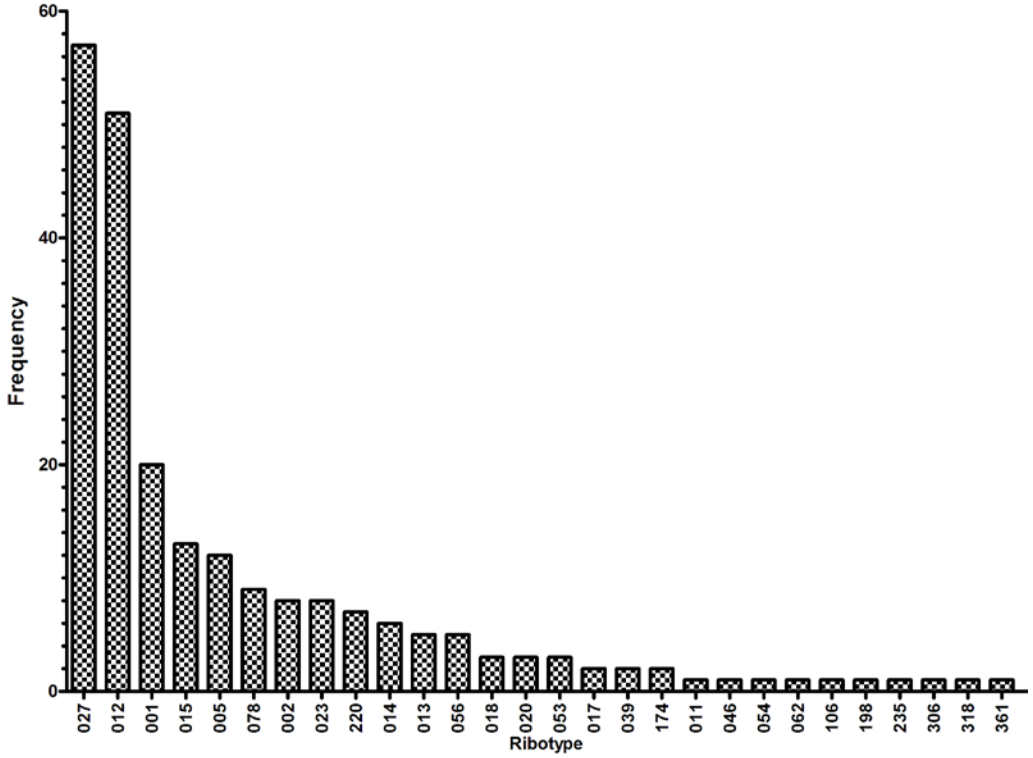


Figure 6.2- Histogram of ribotype distribution within CDI cases from the Royal Devon and Exeter Hospital from 2007-2009.

Chapter Five (Table 5.1) also highlighted the change in prevalence of ribotypes in England and NI from the period April 2007-March 2010. The study period for collection of *C. difficile* isolates from the RD&E was January 2007 until December 2009 and the relative frequency distribution of ribotypes, by year was investigated (Figure 6.3).

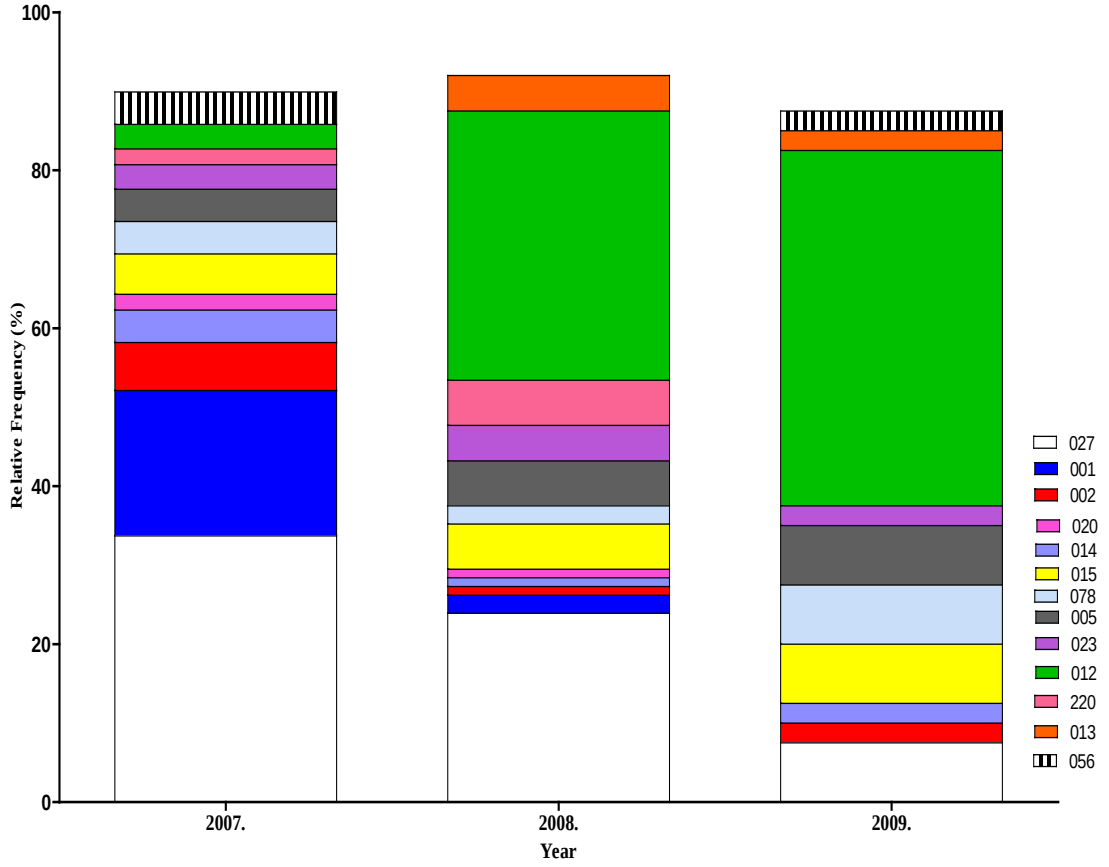


Figure 6.3- Relative frequency distribution of *C. difficile* ribotypes in the Royal Devon and Exeter Hospital from 2007-2009. Fifteen ribotypes made the remaining relative frequency data and were present ≤ 4% over the study period.

Figure 6.3 shows the prevalence of ribotypes, by year, over the study period. There are clear changes in prevalence of ribotypes 027 and 012, and more subtle changes for other ribotypes. In order to compare the trends in prevalence of ribotypes in the RD&E cohort to trends seen nationally, the percent change in prevalence of ribotypes over the study period was calculated, and compared to the percent change seen at a national level, as determined by the 2009/2010 CRDN report[6] (Table 6.1). Generally, the prevalence change patterns of ribotypes

seen in the study cohort show similarity to the national changes over the period April 2007 to March 2010 (Table 6.1). But there are some differences which will be outlined below.

Table 6.1- Yearly change in prevalence of ribotypes in the RD&E cohort as compared to the national change in prevalence reported by the 2009/2010 CDRN report[6]

Ribotype	Relative Frequencies (%, N)			Prevalence change in study cohort from 2007-2009 (%)	Prevalence change as defined in the CDRN 2009/2010 report 2007-2009	Percent (%) difference in prevalence between the study cohort and the CDRN report
	2007	2008	2009			
027	33.70 (33)	23.90 (21)	7.50 (3)	-26.2	-19.2	7 (-)
001	18.40 (18)	2.30 (2)	0 (0)	-18.4	-1.4	17 (-)
012	3.1 (3)	34.1 (30)	45 (18)	+ 15	n/a	n/a
015	5.10 (5)	5.70 (5)	7.50(0)	+2.4	+2.9	0.5 (+)
002	6.10(6)	1.10 (1)	2.50 (1)	-3.6	+3	6.6 (-)
078	4.10(4)	2.30 (2)	7.50(3)	+3.4	+1.7	1.7 (+)
***014/020**	6.10(6)	2.20 (2)	2.50(1)	-3.6	+2.6	6.2 (-)
005	4.10 (5)	5.70 (5)	7.50 (3)	+3.4	+1.5	1.9 (+)
023	3.10 (3)	4.50 (4)	2.50 (1)	-0.6	+1.7	2.3 (-)
220	28.6 (2)	71.4 (5)	0 (0)	-28.6	n/a	n/a
013	0 (0)	80 (4)	20 (1)	+20	n/a	n/a
056	80 (4)	0 (0)	20 (1)	-20	n/a	n/a

* Inclusion of ribotype 020 as it is often grouped with ribotype 014. +/- = increase or decrease in ribotype prevalence compared to the 2009/2010 CDRN report[6].

According to results in Table 6.1; over the study period in the RD&E cohort, there has been a decrease in prevalence of 26% for ribotype 027; this is a further decrease of 7% from the national decrease of 19%. Prevalence of ribotype 001 has decreased by 18.4% in the RD&E cohort which is a further decrease of 17% from the 1.4% national decrease. Data also indicates an increase in prevalence of 2.4% for ribotype 015, which is similar to the reported increase by the HPA for the same period. An increase of 3.4% for ribotype 005 in this cohort is over double the national prevalence rate as reported by the HPA for the same period. Contrasting data from Table 6.2 shows a decrease of 3.6% of ribotype 002 for the study cohort as compared to a national increase of 3% as reported by the HPA, and decrease of

0.6% for ribotype 023 in the study cohort against a national increase of 1.7%. Ribotype 013 showed a change in prevalence in this cohort, but is not seen as an emerging strain by the HPA. In the RD&E cohort some ribotypes are present at low frequencies; therefore care must be taken when interpreting changing epidemiology.

The prevalence of certain *C. difficile* ribotypes is changing at a national and regional level[229]. Given that there was a rapid dissemination of the 'hypervirulent' BI/NAPI/027 ribotype across the globe and it was associated with increased mortality, severity and recurrent infection[148, 218], there have been some studies that have tried to deduce if there is any association between outcomes of infection and the acquisition of a particular *C. difficile* ribotype[7, 9, 199, 206, 208]. Therefore, the next step of the analysis was to determine if any of the ribotypes identified during this study were significantly associated with mortality or recurrent CDI outcomes. However, this analysis has limitations of power to deduce statistical significance of ribotype associations to recurrent CDI and mortality, due to the moderate to small sample sizes of ribotypes identified during this study. All associations to outcomes deduced from this analysis were therefore be interpreted with caution and led to generation of hypotheses only.

6.2.2 Distribution of Cases of Recurrent Infection within Ribotypes

Recurrent infection imposes a significant burden on the healthcare healthcare system[63, 77]. Results from Chapter Three indicated that for the study cohort examined in this thesis, there were no variables associated with recurrent CDI. Therefore, research in this chapter aimed to investigate whether CDI caused by a particular *C. difficile* ribotype was associated with recurrent CDI, as there is some literary evidence to suggest that isolates containing binary toxin (CDT) may be predictors of this outcome[206].

Figure 6.4 shows the percentage of relapsing infection cases (the ribotype that caused the primary infection was the same as the subsequent infections) and primary only infection cases for each ribotype, where there were >3 cases of any ribotype. The distribution of cases of relapse and primary only CDI for different ribotypes was then subjected to binomial analysis to see if any ribotypes were associated with increased recurrent CDI.

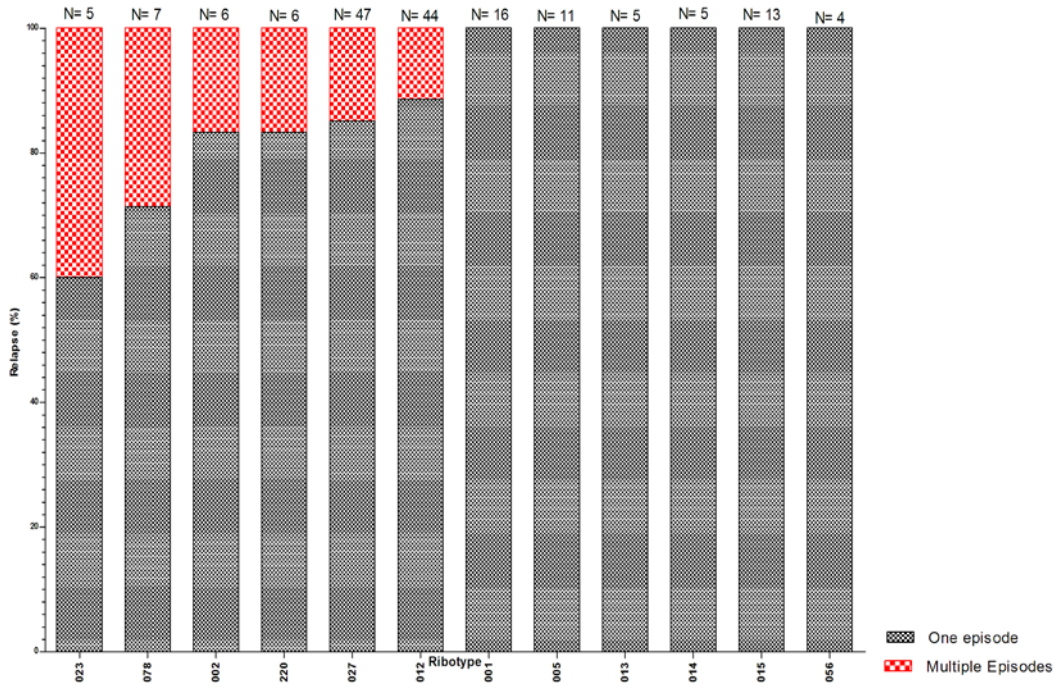


Figure 6.4– Distribution of percent recurrent and primary CDI cases for specific ribotypes. N= number of patients. Some ribotypes are missing as they were present at ≤ 3 times in the population.

6.2.2. i. Binomial Analysis Results

Binomial analysis was conducted on patient data to test for ribotype association to recurrent infection. A test proportion statistic of 0.2 was assigned for the probability of being in the multiple episodes of CDI group, given that the approximate number of recurrent episodes reported in the literature is 20%[111, 146].

Two analyses were performed-

1) H_0- A test statistic of 0.8 was assigned for the probability of being in the primary only CDI group and H_1- is where the probability of being in the primary only CDI group is <0.8. Indicating an increased association to recurrent CDI.

2) H_0- A test statistic of 0.2 was assigned for the probability of being in the multiple CDI episodes group, and H_1- is where the probability of being in the multiple CDI episodes group is <0.2. Indicating an increased association to primary only CDI.

Analysis of cases of recurrent CDI within different ribotypes revealed that there were no ribotypes significantly associated with recurrent CDI. The percentage of cases of recurrent CDI for ribotypes 023 and 078 is > 20% (Figure 6.4), therefore, it is possible that ribotypes 023 and 078 might be more associated with the multiple episode outcome if a larger sample population was available.

The outcome of the analysis two showed that ribotypes, 001, 005, 013, 014, 015, and 056 were significantly associated with cases of primary infection (Table 6.2, P≤0.002). However, due the relatively small sample sizes the actual significance of association with primary only CDI must be interpreted with caution.

Table 6.2– Ribotypes significantly associated with one episode of CDI.

C. difficile **Ribotype**	**Category**	**N**	**Observed Proportion.**	**Test Proportion.**	**Exact Sig. (1-tailed)**
001	**Primary CDI**	16	1.0	0.2	P<0.001
	Total	16	1.0		
005	**Primary CDI**	11	1.0	0.2	P<0.001
	Total	11	1.0		
013	**Primary CDI**	5	1.0	0.2	P<0.001
	Total	5	1.0		
014	**Primary CDI**	5	1.0	0.2	P<0.001
	Total	5	1.0		
015	**Primary CDI**	13	1.0	0.2	P<0.001
	Total	13	1.0		
056	**Primary CDI**	4	1.0	0.2	P=0.002
	Total	4	1.0		

a. Alternative hypothesis states that the proportion of cases in the first group < 0.2. b. Based on Z Approximation. N= Number of patients.

6.2.3 Distribution of Cases of Mortality within Ribotypes

In 2010, the Office for National Statistics reported that for those who died from CDI related causes, the percentage of death certificates mentioning *C. difficile* was 41%. There have been may studies identifying host variables associated with mortality[32] and numerous studies have found the 'hypervirulent' BI/NAPI/027 ribotype, to be associated with increased mortality[218, 229]. Results from Chapter Four deduced that there were four host variables associated with mortality in patients with CDI; however host factors may only contribute in part to the outcome of an infection. Based on research that has linked CDI caused by certain *C. difficile* ribotype with mortality, distribution of mortality for different *C. difficile* ribotypes was assessed. Figure 6.5 presents the percentage of cases of survival, CDI related mortality and non-CDI related mortality for each ribotype, where there were >3 cases of any ribotype.

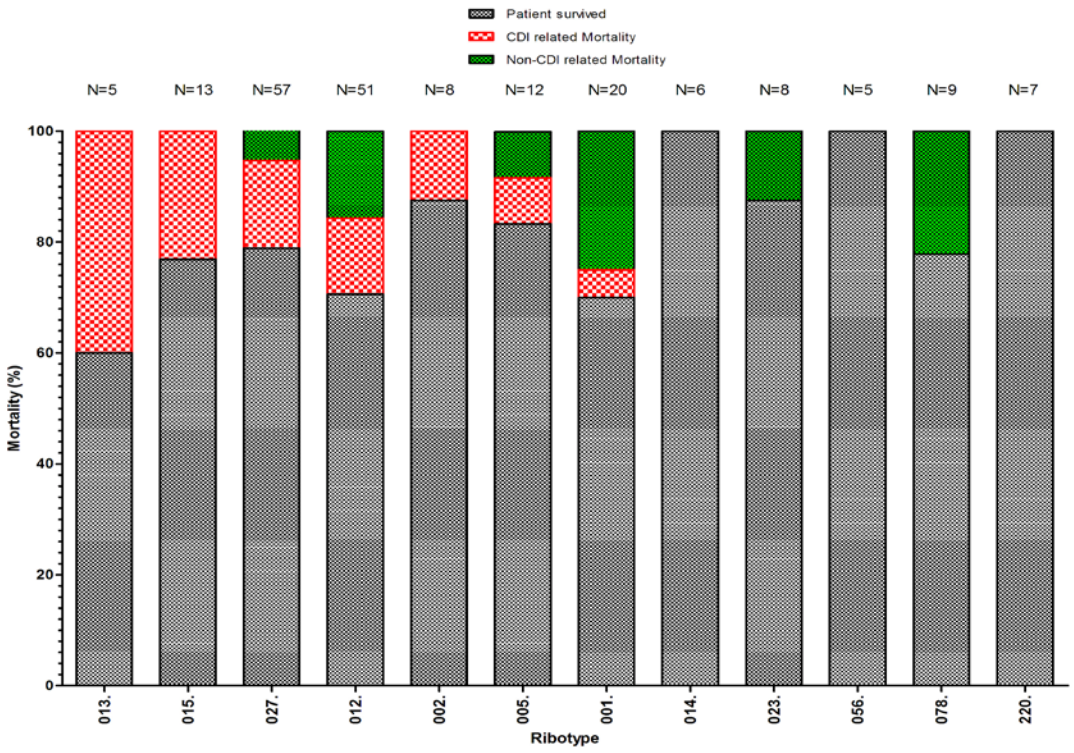


Figure 6.5- Distribution of percent mortality for specific ribotypes. N= number of cases. Some ribotypes are missing as they were present at ≤ 3 times in the population. Grey hatched boxes represent survival, red hatched boxes represent CDI related morality and green hatched boxed represent non-CDI related mortality.

6.2.3. i. Binomial Analysis Results

Binomial analysis was conducted on all cases excluding those who had non-CDI related mortality (N=189). Analysis was performed using the following test statistics based on 2010 data from the Office for National Statistics which indicates that of patients who died from CDI related causes; 41% of death certificates mentioned *C. difficile* as the underlying cause of death.

Two analyses were performed-

1) H_0- A test statistic of 0.59 was assigned for the probability of being in the survival group where H_1- is where the probability of survival is $\neq 0.59$

2) H_0- A test statistic of 0.41was assigned for the probability of being in the CDI related mortality group where H_1- is where the probability of CDI related mortality is <0.41. Indicating an increased association to lower mortality.

Results from analysis one revealed that there were no ribotypes significantly associated with mortality above 41%. However, two ribotypes (013 and 015) did show mortality rates above or equal to the cohort average of 24% (Figure 6.5). Results from analysis two (Table 6.3) identified nine ribotypes that were significantly associated with lower mortality with $P \leq 0.03$.

Table 6.3– Ribotypes significantly associated with lower mortality.

C. difficile Ribotype	Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)	Asymp. Sig. (1-tailed)
001	CDI- related Mortality	1	.07	0.41	P=0.004[a]	
	Patient survived	14	.93			
	Total	15	1.00			
005	CDI- related Mortality	1	.09	0.41	P=0.026[a]	
	Patient survived	10	.91			
	Total	11	1.00			
012	CDI- related Mortality	7	.16	0.41		P<0.001[a,b]
	Patient survived	36	.84			
	Total	43	1.00			
014	Patient survived	6	1.00	0.41	P=0.005	
	Total	6	1.00			
023	Patient survived	7	1.00	0.41	P=0.002	
	Total	7	1.00			
027	CDI- related Mortality	9	.17	0.41		P<0.001[a,b]
	Patient survived	45	.83			
	Total	54	1.00			
056	Patient survived	5	1.00	0.41	P=0.012	
	Total	5	1.00			
078	Patient survived	7	1.00	0.41	P=0.002	
	Total	7	1.00			
220	Patient survived	7	1.00	0.41	P=0.002	
	Total	7	1.00			

a. Alternative hypothesis states that the proportion of cases in the first group <0.41. b. Based on Z Approximation.

Given that there were two ribotypes (013 and 015) which were associated with cases of mortality rates above or equal to the cohort average, binomial analysis one was repeated again with a test statistic of 0.76 to deduce if any ribotypes were significantly associated with mortality above the cohort average of 24%. Results of this analysis deduced that these ribotypes were still not significantly associated with increased mortality. When binomial analysis two was repeated with a test statistic of 0.24, five ribotypes (014, 023, 056, 078 & 220) were significantly associated with lower mortality ($P \leq 0.001$). However, due to the relatively small sample sizes for some ribotypes, the actual significance of association to lower mortality must be interpreted with caution.

6.3 Discussion

Assessing the epidemiology of *C. difficile* ribotypes provides a useful indication of possible regional differences and time trends to identify potential emergent or declining *C. difficile* ribotypes[229]. The overall prevalence of ribotypes in this cohort did fit the general national trend of ribotype prevalence (Chapter Five, Table 5.1), but there were some differences which will be discussed in the next section.

6.3.1 Ribotype Distribution

A recent publication by Wilcox *et al.*[229] demonstrated that the South East and West regions had significant year-on-year increases of ribotypes, 002, 015 and 078 ($P<0.05$) but no significant change in prevalence of ribotype 027. The results from ribotype distribution analysis of this study cohort supports the data by Wilcox *et al.*[229], in that ribotypes 015 and 078 were seen to increase during the study period. Over the period in which the study samples were collected there were differences in the trend of prevalence for ribotypes 001, 002, 014, 023 and 027 compared to the national trend. On a national scale across the study period, ribotypes 002, 014 and 023 showed an increase of 3, 2.6 and 1.7% respectively, but in this cohort showed a decrease of 3.6, 3.6 and 0.6% respectively. There was also a bigger decrease of ribotype 001 in this study cohort (18.4%), relative to the national decrease (1.4%).

In conjunction with findings from the CDRN report[6], ribotype 027 decreased in prevalence in this cohort. However, this differs to findings by Wilcox *et al.*, who report that there was no significant change in prevalence of this ribotype in the South West[229].

There were two ribotypes (012 & 013) prevalent in this cohort which are not recognised as prevalent ribotypes in the England by the 2009/2010 CDRN report[6]. Wilcox *et al.*[229] report that the incidences of ‘other’ ribotypes; not considered in the top ten ribotypes in England,

have increased on a national scale. According to the 2010/2011 CDRN report, ribotype 012 is the 15th most common ribotype in England[7] and in Europe it is considered a predominant ribotype[6]. The proportion of ribotype 012, relative to all the others, over a four year period is less than 0.02[7]. Therefore, whilst in this study these ribotypes were common, their prevalence at a national and regional level is more likely to be captured in the 2010/2011 CDRN report data, as an increase in 'other' ribotypes. Why the 012 ribotype is found more frequently in this cohort of patients is something which could be investigated further if more time were available.

Differences in prevalence of the above mentioned ribotypes, from national trends could be due to the fact that in the South West, prevalence of these ribotypes fluctuates over time (Chapter Five, Figure 5.3). There will also always be differences in ribotype distribution, when comparing a single clinical facility to national and regional trends. Some differences may also come from potential bias that is brought about by the increased sample submission to the CDRN, which may then lead to an amplification of the relative contribution of other *C. difficile* ribotypes, including emergent strains, among recorded cases[229].

Whilst the overall distribution of ribotypes in the study cohort shows similarity to the national and regional trends, differences highlight the potential effect that individual geographic locations and clinical settings may have on ribotype diversity. There is an increased need for vigilance of the *C. difficile* ribotypes predominating within clinical settings, especially when *C. difficile* has the potential to cause such a debilitating infection. Coupled with ribotype surveillance, is the need to understand factors which could be facilitating the emergence of certain ribotypes, such as bacterial antibiotic resistance and sporulation, the immune status of the host and the ability of the normal gut microflora to re-colonise after antibiotic therapy.

Antibiotic resistance is one factor which may contribute to the persistence of *C. difficile* in the clinical setting. It has been reported by the HPA that epidemic lineages show decreased susceptibility to metronidazole[6] and that there is a need for heightened susceptibility testing due to the possible acquisition of antibiotic resistance from other enteric bacteria that may harbour resistance genes. Sporulation of *C. difficile* is another factor which could contribute to the prevalence of certain ribotypes within the clinical setting. Spore production is the key mode of transmission and persistence of *C. difficile* within the hospital environment. The rate at which spores are formed could help understand why certain ribotypes are predominating within the clinical setting. Ribotype specific differences in bacterial virulence mechanisms that could be facilitating the emergence of certain ribotypes, and may play a role in the outcome of infection will be discussed further in Part Three.

6.3.2 Ribotype Association with Infection Outcomes

Due to the small ribotype sample sizes found within this study cohort the statistical analysis procedure in this chapter has limited power to deduce the true significance of ribotype association with infection outcomes. A recent publication by Walker *et al* (2013) concluded that approximately "700-800 cases are needed to detect an absolute mortality increase above 8% (across clades), with 80% power"[225]. Therefore the following discussion is speculative and hypothesis driven, based on the findings during this study.

In this study ribotype 027 was found most frequently, however binomial statistical analysis revealed that it was not significantly associated with increased mortality, or with recurrent disease, but was statistically significantly associated with lower mortality ($P<0.001$). This is contradictory to evidence shown in the 2009/2010 CDRN report which suggests that ribotype 027 is associated with increased mortality[6]. There is varying opinion on whether this ribotype is still 'hypervirulent' due to confounding reports on its association to increased mortality,

severity and recurrent CDI[131, 199]. Ribotype 027 is found by some, to produce more toxins and spores[219] than other types, which could indicate why this ribotype has remained endemic in the hospital population. In this study no ribotypes were associated with increased recurrent infection. Although ribotypes 023 and 078 show a > 20% recurrence rate (Figure 6.4), the number of patients presenting with CDI with these ribotypes was low (N=5 & 7 respectively). Two out of five patients presenting with ribotype 023 had > 1 CDI, and three out of seven patients presenting with ribotype 078 had > 1 CDI episode. Both these ribotypes encode genes for binary toxin (CDT), thus, it might be argued that if the cohort sample size were to increase, then ribotype numbers would also increase and they might then show an increased association to recurrent CDI. This would correlate with work by Stewart *et al*. who found that CDT was a predictor of recurrent CDI[206]. Ribotypes 023 and 078 also showed significant association to lower mortality (023; P=0.002 and 078; P=0.002), which corroborates with the 2009/2010 CDRN report[6], as it shows that percentage residual mortality rates are lower for these ribotypes (Figure 6.6) compared to others.

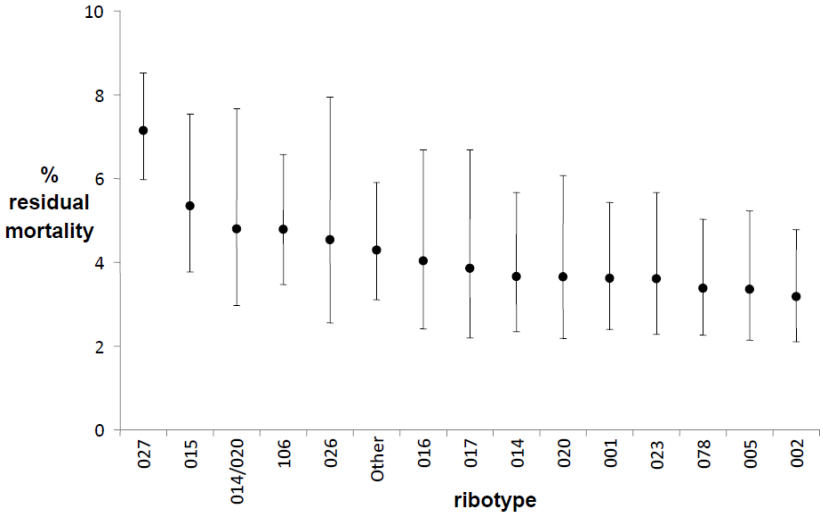


Figure 6.6- Percent residual mortality rates for clinical ribotypes[6].

Ribotypes 015 and 013, were also both associated with increased primary CDI (P<0.001) and had mortality rates that were equal to, or higher than the cohort average of 24%, but not than

the percentage of CDI related deaths where *C. difficile* was mentioned as an underlying cause (41%). In the 2009/2010 CDRN report[6], ribotype 015 has the second highest residual mortality (Figure 6.6), and whilst there was no overall association of this ribotype to increased mortality in the CDRN report, evidence from results in this thesis may suggest that in the RD&E, it might have contributed to the mortality infection outcome.

Such differences in outcomes associated with acquisition of a certain ribotype could be attributed to a variety of bacterial virulence characteristics such as, toxin production, sporulation, motility/adhesion and antibiotic resistance and also the interaction with the host. The interaction between *C. difficile* and the host immune system is complex and not fully understood, but it is an important driving factor for the outcome of infection[123]. The production of TcdA and TcdB and CDT from *C. difficile* elicits an inflammatory response via the NF-κB and MAP kinase pathways, which then mediates activation of monocytes and macrophages, as well as pro-inflammatory cytokines such as Il-1β, TNF-α and IL-8[123]. The host antibody response to this inflammatory process comes in the form of IgG and IgM serum antibodies to TcdA and TcdB, and this inflammatory/immune response is thought to contribute to CDI outcomes such as severity, recurrence and mortality[123]. A study by Kyne *et al*., found that IgM antibodies to TcdA and TcdB on day three after the onset of CDI were higher in cases of primary only CDI than recurrent CDI, and on day 12 of infection IgG antibodies to TcdA were also higher in primary only cases[132]. This led to the conclusion that a decreased IgG, IgM immune response to toxin could be indicative of recurrent CDI. A more recent study by Katchar *et al*., also concluded that the IgG2 and IgG3 subclasses were deficient in those suffering from recurrent CDI[121].

It is interesting that in both studies by Kyne *et al.*[132] and Hu *et al.*[102], severe CDI was observed as a marker for recurrent CDI. Many studies have indicated that inflammatory responses such as elevated WCC and neutrophil count are predictive of more severe

outcomes for CDI[45]. During the present study, host inflammatory markers such as, WCC and CRP, were predictive of mortality in patients with CDI (Chapter Three). Therefore, perhaps there is a fine balance between the level of inflammation evoked by *C. difficile* and the strength of the host immune response, in facilitating the outcome of infection.

The complex host-pathogen interaction and the importance of the host immune response to infection is further emphasised by work produced by Jiang *et al.*[108, 109], and Garey *et al.*[76] who show that the A/A genotype in the -251 promoter region of the Il-8 gene is predictive of recurrent CDI. These studies report that subjects with the A/A genotype produce more faecal levels of Il-8 and they show lower rates of host immunoglobulin response to TcdA. One could thus hypothesise, that a possible explanation for the outcome of association findings during the present study come from both bacterial toxin production, and host immunostatus. For example, ribotypes that are showing an association (although not significant) to mortality such as, ribotypes 015 and 013 could be producing more toxin, eliciting a prolific inflammatory response in the host, who then either have a high immune response but are unable to control the infection or do not respond and are unable to fight infection. However, if the patient survives CDI with these ribotypes there is little chance of it causing another infection due to the protective nature of the antibody response (if the patient has an immune response) or possible compromised pathology of the colon; making another infection improbable. Thus, these ribotypes may also show an association to primary only CDI. Ribotypes which are more frequently found in cases of recurrent CDI, such as 023 and 078 (although not significantly) may produce less toxin, or employ other mechanisms to avoid eliciting an increased inflammatory response, which leads to a lowered host immune response, which is not protective and may then predispose the host for another infection. However, this is only a hypothesis and there is clearly a need to investigate such interactions in more detail, as well as investigating other mechanisms that account for infection outcomes.

6.4 Conclusion

It is clear from literary evidence that in Chapter Five, that in England there are some ribotypes which are prevalent (027, 001, 002, 005) or emerging more progressively (014, 015, 023, 078) than others within the clinical setting. Limited statistical findings in this chapter suggest that there are some ribotypes which are significantly associated with primary only CDI and higher mortality, and some which show some association to recurrent infection and low mortality (although not significantly). There are also some ribotypes which show different trends from the rest. The reason for the prevalence of certain ribotypes within the clinical setting, and their association to clinical outcomes is likely to be driven by a combination of, clinical infection management procedures, host inflammatory/immune response to infection, and bacterial virulence mechanisms, which include antibiotic resistance, sporulation, toxin production and gut colonisation efficiency.

6.5 Summary

Part One of this thesis has highlighted the important host variables which contribute to the outcome CDI. Part Two of this thesis has highlighted that there could also be an association between the ribotype causing infection, with the outcome of an infection. Given that there are certain ribotypes which appear to be frequently associated with recurrent infection and mortality, there may be important differences in the virulence mechanisms of these ribotypes, which may explain these associations.

In order to deduce if differences in any bacterial virulence mechanisms could play a role in infection outcomes, sixteen isolates (Table 6.4) from the ribotypes 078, 023, 027, 014, 015 and 013, were chosen for whole genome sequence analysis and phenotypic characterisation. These isolates were selected on the basis of literary evidence in Chapter Five which suggests they are from ribotypes which are clinically prevalent and/or are seen as emerging, and

evidence in this chapter which links them to the CDI outcomes such as, mortality and recurrent CDI. The number of isolates that were able to be characterised during this study was limited by time constraints and thus the results presented in Part Three are not representative of specific ribotype groups and their association with certain CDI outcomes. However, the panel consists of isolates from both mortality and recurrent outcomes, from different years and different severities of infections to try to start to understand what virulence characteristics may contribute to the outcome of infection to which the individual isolate was associated. Part Three of this thesis will introduce *C. difficile* virulence mechanisms and present the results of whole genome sequence analysis and phenotypic characterisation studies for the clinical isolates seen in Table 6.4. A more detailed patient history corresponding to each isolate in this panel can be seen in Appendix One.

Table 6.4- Clinical isolates selected for whole genome and phenotypic analysis.

Clinical Isolates	Ribotype	Severity	Cause of Death*	Number of Episodes	Patient
FA08012693	013	Mild	N/A	One episode	1
FA08006290	013	Severe	CDI was part of or mentioned in death	One episode	2
FA07007469	014	Life Threatening	N/A	One episode	3
FA07004464	014	Mild	N/A	One episode	4
FA09007583	015	Severe	CDI was part of or mentioned in death	One episode	5
FA07011498	015	Severe	N/A	One episode	6
FA07003485	023	Moderate	N/A	Multiple Episodes	7
FA07004080	023	Mild	N/A	Multiple Episodes	7
FA08006661	023	Severe	CDI was not the cause of death	One episode	8
FA08005864	023	Severe	N/A	One episode	9
FA07007522	078	Severe	N/A	Multiple Episodes	10
FA07008490	078	Severe	N/A	Multiple Episodes	10
FA09004991	078	Life Threatening	N/A	One episode	11
FA08006656	078	Severe	CDI was not the cause of death	One episode	12
FA07001994	027	Mild	N/A	Multiple Episodes	13
FA07003754	027	Severe	Patient survived	Multiple Episodes	13

* N/A- Patient survived.

Part Three
Genetic and Phenotypic Characterisation of *C. difficile* Isolates Associated with Outcomes of Infection

Chapter Seven. Molecular Basis of *C. difficile* Virulence

7.1 Introduction

Part Three of this thesis will report on the genetic and phenotypic characterisation of the clinically predominant ribotypes selected in Part Two. This approach has been taken based on the idea that clinical ribotypes may prevail due to phenotypic differences in virulence mechanisms. Comparative studies of these ribotypes may contribute to our understanding of the differences in mechanisms that could be associated with differences in virulence[22], and how they may contribute to the outcome of an infection. This introductory chapter will give a brief overview of the applicability of next generation sequencing (NGS) to aid the study of globally important pathogens, and highlight the current major virulence mechanisms of *C. difficile*.

7.2 Whole Genome Analysis to Identify Bacterial Virulence Determinants

The capabilities of NGS and advanced bioinformatic programming has facilitated the rapid elucidation of microbial genomes, and has made it possible to deduce significant findings in microbial pathogenicity, virulence and evolution[154]. It is the nature of many disease causing pathogens to evolve via mutation and sequence exchange. These mutations allow pathogens to better adapt to the host, evade the immune response and become more resistant to antibiotic therapy[142]. Thus, the spectrum of mutations acquired by pathogenic bacteria are likely to reflect the genetic mechanisms of pathogenesis *in vivo*[142]. Over the last five years many research groups have used NGS platforms to facilitate bacterial whole genome comparisons in order to compile information on virulence mechanisms, in addition to evolutionary data from phylogenetic analysis[29, 98, 101]. Studies using NGS platforms significantly reduce the labour intensive process of mutational mapping and perhaps facilitate more efficient targeting of genes in pathogenic bacteria that may have the potential for therapeutic use.

Whilst looking at whole genome sequences allows novel insights into bacterial evolution and virulence, it is also important to link the changes seen at the genomic level, with *in vitro* phenotypic traits, in order to gain a more comprehensive insight into possible host-pathogen relationships.

7.3 Whole Genome Sequencing of *C. difficile*; Linking Phenotypic and Genotypic Data

In 2006 the genome of *C. difficile* 630; a hypervirulent strain, isolated from a PMC patient in Zurich, was published[197]. The genome consists of a 4.3 Mbp chromosome and an 8 Kbp plasmid (pCD630). The chromosome is predicted to have ~3,776 coding DNA sequences (CDS) whilst the plasmid is predicted to contain 11 CDSs, none of which have any assigned function. Sequence alignments of Clostridial genomes revealed that 50% of CDSs are unique to *C. difficile* and may encode accessory functions and mobile elements (Figure 7.1). These mobile elements make up ~11% of the genome and it is proposed that this may reflect the bacteria's niche in the gut, therefore providing insight into possible virulence determinants[197].

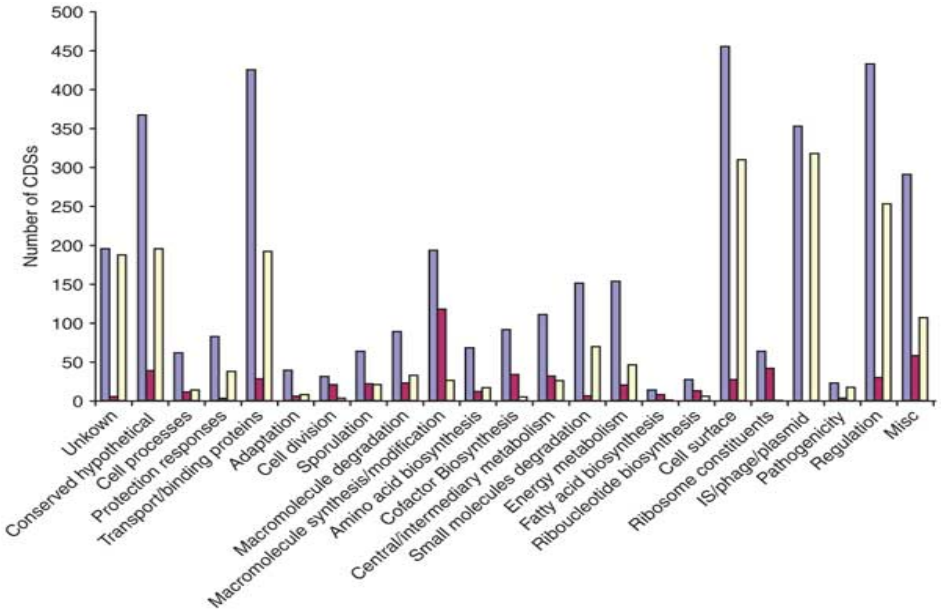


Figure 7.1- Representation of the distribution, by functional categories, of the shared and unique *C. difficile* CDSs relative to the sequenced Clostridial genomes. Blue: *C. difficile* CDSs, purple: *C. difficile* CDSs shared with the sequenced clostridia, yellow: CDSs unique to *C. difficile* relative to the sequenced clostridia[197].

The publication of the *C. difficile* 630 genome sequence, and the progression of NGS platforms have led to the publication of many more *C. difficile* genomes, thus facilitating genome comparison studies. These studies have provided useful insights into the genetic changes between strains which have led to altered motility, toxicity and antibiotic resistance phenotypes[202]. The idea that understanding the genetic background of a bacterium, allows a deeper understanding of *in vitro* phenotypic characteristics, is something that underpins the approach employed during this study. Thus, the aim of this part of the PhD research is to couple both bacterial phenotypic and genotypic data, in order to better understand the complex interaction of *C. difficile* with the host.

7.4 Role of *C. difficile* Virulence Factors in Disease

7.4.1 *C. difficile* Virulence Characteristics

Prevalence and transmission of *C. difficile*, and clinical disease presentation, are likely to be driven by factors including, toxin production, sporulation and antibiotic resistance. Most

strains of *C. difficile* produce two toxins (TcdA and TcdB) which cause disease symptoms, and all strains are capable of forming highly resilient spores[173], which are excreted in faeces and contaminate the environment. Within the clinical environment, CDI is exacerbated by the use of broad spectrum antibiotics such as, penicillins, cephalosporins, and fluoroquinolones[42], which disrupt the normal gut flora, enabling *C. difficile* spores to germinate, vegetative cell growth to resume and thus allows subsequent toxin production. Recent publications have also implicated bacterial flagellar as an important factor in *C. difficile* pathogenicity, contributing to adherence of bacteria to epithelial cells in the gut[59] and the organisms motility[218]. Flagellar gene regulation has also recently been associated with toxin expression[15]. In order to facilitate a greater understanding of the complex repertoire of *C. difficile* virulence factors, whole genome sequencing and mutational analysis of the genome is pivotal.

7.4.2 Toxin Mediated Virulence

The mechanism of *C. difficile* pathogenesis via TcdA was initially proposed by Hecht *et al.*[96] who concluded that the action of TcdA is to increase cell permeability to ions and hexose sized molecules via tight junction modification. Others have also deduced that TcdB is able to alter cellular phenotype via redistribution of filamentous actin[97]. It is the disruption of cytoskeletal and tight junction structure that leads to retraction of microfilaments from the plasma membrane, altered cell shape and local accumulation of condensed actin[105]. The alteration of epithelial cell phenotype, induced by the toxins then leads to the recruitment of neutrophils to gut epithelia. Work by Pothoulakis *et al*. has shown that TcdA is chemoactic to human neutrophils[182], and later work by Kelly *et al*. has shown that a TcdA receptor is present on rabbit neutrophils[122]. Work by Kelly *et al*. also demonstrates the importance of neutrophil recruitment for the enterotoxicity of TcdA, by showing that enterotoxicity is inhibited by an antibody directed against the leukocyte adhesion molecule CD18[122]. Pothoulakis suggests that TcdA mediated enterotoxicity is initiated through cytokines

released from intestinal epithelia on response to TcdA[181]. The effect of toxins on gut epithelia evokes a host inflammatory response, which facilitates intestinal damage, and results in the clinical presentation of disease, such as diarrhoea and PMC[123]. However, the resulting immune response mounted by the host may also have a part to play in the outcome of clinical disease, whether it be severity of infection, mortality or recurrent CDI[123].

7.5 Role of Toxins in Disease

There has been much debate into which toxins are essential for pathogenicity of *C. difficile*. Lyras *et al*. demonstrated that TcdB is essential for pathogenicity in the hamster disease model, and they state that this can be supported by the increasing incidence of $TcdA^{-}TcdB^{+}$ strains being isolated from disease cases, yet cases of naturally occurring $TcdA^{+}TcdB^{-}$ strains have never been reported[151]. However, a study by Kuehene *et al*. demonstrated that isogenic mutants of *C. difficile*, producing either TcdA or TcdB could cause fulminant disease in the hamster model of infection, thus emphasising a role for both toxins in pathogenesis[130]. *C. difficile* also produces binary toxin (CDT). Due to the lack of robust genetic mutational tools available to genetically manipulate *C. difficile*, and the fact that most CDT^{+} strains are also $TcdA^{+}TcdB^{+}$, only natural *C. difficile* variant types have been used to gather limited information on the role of CDT in pathogenesis[79]. CDT has similar functions to TcdA and TcdB, in that its role is to modify the actin cytoskeleton of a cell, thus increasing cell permeability[180].

CDT is found in the epidemic 'hypervirulent' BI/NAPI/027 strain and emerging ribotypes such as, 078 and 023[207], and has been proposed by some to be an additional virulence factor[176]. A study by Stubbs *et al*. analysed a number of ribotypes for the presence of *tcdA*, *tcdB*, *cdtA* and *cdtB*. Their study found that there was a correlation between ribotypes with significant changes in *tcdA* and *tcdB*; such as, ribotypes 027 and 078, as compared to *C. difficile* VP10463, and presence of CDT[207].

7.6 Transcriptional Regulation and Translation of the Pathogenicity Locus; Role in Disease Outcomes

Whilst expression of TcdA and/or TcdB is essential for *C. difficile* disease, the transcriptional regulation of toxin genes is also important, as changes in regulation have been seen in ribotypes that are capable of causing more severe disease[158]. The *C. difficile* PaLoc is constitutively transcribed during all stages of growth, as two transcriptional units; *tcdA, B, R* and *E* in one direction and *tcdC* in the opposite direction[104] (Figure 7.2).

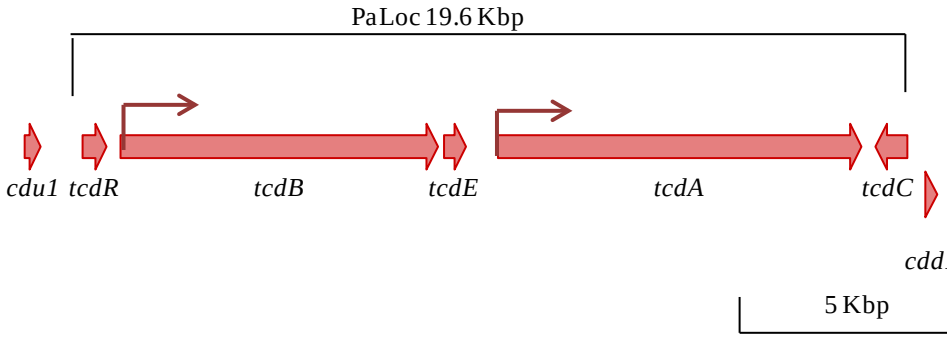


Figure 7.2- Genetic map of the 19.6 Kbp PaLoc of *C. difficile*. Direction of the block arrows indicates the direction of transcriptional regulation of the toxin genes during early log stages of growth. The small arrows above TcdA and TcdB represent independent transcription from their own promoters, during late log and stationary growth phases.

During late log stages of growth there is increased independent transcription of *tcdA* and *tcdB*. This is thought to be due to environmental stress responses, regulating monocistronic transcription from their own promoters[104]. Observations of *tcdC* transcript levels, show it is highly up regulated during the early exponential phase of growth, where transcription of *tcdA, B, R* and *E* is low. This is reversed during late exponential growth, possibly indicating the action of TcdC as a negative regulator of TcdA, B, R and E at this stage[156]. Further work on *tcdA* and *tcdB* regulation has elucidated the role of TcdR and TcdC accessory proteins, as positive and negative regulators respectively, of expression during different growth phases[33].

During early exponential growth phase, TcdC negatively regulates *tcdA* and *tcdB* by destabilising the TcdR-haloenzyme, resulting in diminished binding to its respective promoters[156]. Negative regulation of TcdA and TcdB by TcdC is interrupted by TcdR during stationary phase growth. BLAST analysis of TcdR amino acid sequences by Mani and Dupuy have shown that it is an alternative sigma (σ) factor with the ability to interact with a variety of core enzyme RNA polymerases (RNAP)[153]. Interaction of TcdR with RNAP results in the formation of the TcdR-haloenzyme complex, which is then able to bind to the promoter sequences for *tcdA* and *tcdB* and increase transcription during stationary phase. Regulation of toxin and accessory protein expression is governed by a multitude of host, bacterial and environmental stress/starvation factors[64]. TcdR expression is thought to be sensitive to glucose levels in media[153] and genetic changes in TcdC are thought to increase/prolong toxin production in some *C. difficile* strains[165].

The regulatory role of TcdC in toxin synthesis and association with infection outcomes has recently come under scrutiny. The emergence of the hypervirulent' BI/NAPI/027 ribotype, which has been associated with increased mortality and disease severity, was attributed to many factors, including robust toxin production[165]. Upon analysis of the PaLoc genes, two changes in the *tcdC* repressor gene were found, an 18 bp deletion, and a deletion at 117 bp resulting in a frame shift mutation, truncating TcdC. These changes were thought to be sufficient to inhibit the negative regulatory action of TcdC; increasing toxin production[53], thus contributing to the poor outcome of patients infected with this ribotype. A study by Dingle *et al*. has recently shown that as many as twenty-six *tcdC* gene variants exist, but *tcdC* truncation mutations are conserved in ribotypes 027, 078 and 023, from three distinct clades [58]. Dingle *et al*. state that truncation of *tcdC* occurs by two different mechanisms; a single nucleotide deletion (in some clade 2 isolates (containing 027 ribotypes)) and a single nucleotide substitution (common to all members of clade 3 (containing 023 ribotypes) and 5

(containing 078 ribotypes)) and propose that the evolution of this truncation at least twice, may indicate evolutionary convergence, due to a common selective advantage[58]. In their study these three clades were associated with more clinically severe disease, relative to clades 1 and 4 which contained other ribotypes with predominantly 'wild-type' *tcdC*. However, their study did not look at TcdA and TcdB production in any of the *tcdC* variant isolates, so it could not be deduced what effects the variant alleles were having on toxin production. Toxin production of isolates in different media and with 'wild type' and *tcdC* truncation genotypes was investigated during this PhD research and will be discussed further in Chapters Eight and Nine.

Some researchers have shown that ribotypes with variations in *tcdC* may not be associated with poorer clinical outcomes. A study by Goldenberg *et al*. did not find any association of ribotypes with variations in *tcdC*, to clinical outcomes such as severity of disease, ICU admission, mortality, recurrence and length of stay in hospital[82]. They did however; find significant differences in serum CRP concentrations and total peripheral white cell counts in patients infected with ribotypes with variations in *tcdC*, versus those who were not. This suggests that infection with these strains induces a stronger host inflammatory response, possibly due to hyper-production of toxin. They also found a significant difference in outcomes between patients who were infected with ribotypes containing CDT, as opposed to those who were not. The study by Goldenberg *et al*.[82] suggests that even if the strains with variations in *tcdC* were to produce more toxin, the effect alone is not enough to determine the outcome of an infection and highlights the importance of other factors in defining the outcome of infection. The contradictory findings highlighted above emphasise the importance of identifying both host markers and bacterial characteristics when trying to deduce predictors of CDI outcomes.

7.7 Non- toxin Mediated Virulence Factors; Antibiotic Resistance

Antibiotic resistant bacteria pose a significant burden on the healthcare system. Events highlighting this burden include the emergence of methicillin-resistant *Staphylococcus aureus* (MRSA)[143] within two years of the introduction of methicillin to clinical treatment regimen[66]. The acquisition of antibiotic resistance was also the cause of a worldwide *C. difficile* epidemic[93]. The 'hypervirulent' BI/NAPI/027 ribotype which caused the worldwide epidemic is resistant to newer fluoroquinolones such as moxifloxacin, levofloxacin and gatifloxacin, and is thought by some to be associated with increased disease severity and mortality[148]. In order to deduce why this ribotype had become resistant to newer fluoroquinolones, whole genome sequence data from, a pre-epidemic 027 ribotype (CD196), the 'hypervirulent' BI/NAPI/027 strain (R20291), and *C. difficile* 630, were compared, to investigate possible mutations in GyrA and/or GyrB, as resistance to the new fluoroquinolones is encoded by mutations in the quinolone resistance-determining region (QRDR) of these proteins[103]. *In vitro* phenotypic data was then used to determine the antibiotic resistance profiles of all three strains[202]. Sequence data revealed that certain mutations in *gyrA* were conserved in all three strains; however, a single amino acid substitution (Thr82Ile) in GyrA, was only detected in the epidemic R20291 strain. Experimental work verified that R20291 was resistant (MIC $\geq$32 mg/l) to three fluoroquinolones tested (gatifloxacin, moxifloxacin and levofloxacin), and that CD196 and *C. difficile* 630 were fluoroquinolone sensitive (MIC $\leq$ 1.5, 2 and 3 mg/l respectively)[202]. Through sequencing of a global collection of 027 ribotypes, it is now known that two distinct epidemic *C. difficile* lineages emerged with fluoroquinolone resistance. Both lineages independently acquired the Thr82Ile mutation and spread globally in distinct patterns[93].

Treatment of CDI has changed little since the introduction of metronidazole and vancomycin to therapeutic regimen. Fidaxomicin is the first CDI treatment to be approved by the US Food

and Drug Administration in more than 20 years[133]. Resistance to metronidazole is conferred by nitroimidazole genes, first identified in *Bacteroides fragilis* and *Helicobacter pylori*. Homologs of these genes have been found in *C. tetani* and *C. bifermentans* but as yet have not been identified in *C. difficile*[103]. Currently there is some concern over emergence of *C. difficile* strains that show reduced susceptibility to metronidazole[18]. This has prompted a shift, by some, to use vancomycin preferentially to treat CDI[6]. Vancomycin resistance genes have been acquired by other enteric bacteria such as, *Enterococcus* spp.[152], and vancomycin resistant *enterococci* (VRE) is becoming a clinically problematic HCAI[209]. Genes showing similar sequence identity to vancomycin resistance genes found in VRE have been identified in *C. difficile*[197] but whether these genes are expressed and/or are functional remains a question, as there have been no documented cases of vancomycin resistant *C. difficile* isolates. The use of vancomycin is associated with increased incidences of VRE[10], and this might mean that it may become more difficult to find the delicate balance between using antibiotics effectively to treat CDI, whilst not promoting other clinically problematic pathogens.

The emergence of *C. difficile* strains that are resistant to the fluoroquinolones clearly shows that the genome of *C. difficile* is under antibiotic selection pressure, and that acquisition of resistance to commonly used antibiotics is a major feature in the evolution and prevalence of *C. difficile* in the clinical setting[93]. It may, therefore be important, to combine genetic and phenotypic data to gain a more comprehensive understanding of the mechanisms by which resistance occurs. Despite the effectiveness of current antibiotics to treat initial *C. difficile* infections, it is clear that other preventative/treatment methods are needed, such that, if ever resistance to treatment antibiotics were to occur, there would still be multiple effective treatment options for CDI. More thorough whole genome sequence analysis of *C. difficile*

isolates may hold the key to elucidating possible target genes and pathways which could then be exploited to facilitate the development of alternative treatment therapies.

7.8 Non- toxin Mediated Virulence Factors; Sporulation

It has been mentioned previously that transmission of *C. difficile* in the environment is facilitated by its ability to form highly heat labile, alcohol tolerant spores which are able to persist in the environment for prolong periods of time[217]. In 2008, Sorg and Sonenshein proposed a model for spore germination which utilised primary bile salts, cholate derivatives and glycine as co-geminants[200]. In patients on an antibiotic regimen the normal microflora of the gut is altered therefore the number of bacteria capable of reducing primary bile salts is reduced[195]. This bacterial reduction may lead to increased primary bile salts and decreased secondary metabolites in the caecum, providing a good environment for *C. difficile* growth. Similarly the increase in primary bile acids in the large bowel may facilitate the germination of spores and *C. difficile* growth and survival in the colon[200]. Research into sporulation characteristics of different *C. difficile* isolates increased, when it was shown that the 'hypervirulent' BI/NAPI/027 epidemic strain produced more spores than other non-epidemic types[165] thus, possibly making it more efficient at spreading in healthcare facilities. Many groups have since investigated the sporulation characteristics of epidemic and non-epidemic strains. Some groups have seen no differences between them[38], and some groups have even shown that there are sporulation differences between isolates of the same ribotype[39], leading to the belief that sporulation characteristics are diverse and non-ribotype specific. Numerous studies have linked sporulation of *C. difficile* with toxin production[119, 217, 219, 8, 199]. However, there are also conflicting reports on the correlation of toxin production and sporulation, with some studies showing that isolates with robust toxin production also show increased sporulation[219], and others who show that robust toxin producers are poor spore formers[9, 199].

A more recent study by Rosenbusch *et al*. even suggests that sporulation is not linked to toxin production at all[190].

Even though there is no uniform agreement of the sporulation characteristics of *C. difficile* isolates, what is clear, is that *C. difficile* sporulation is key to effective transmission within the clinical setting. Understanding sporulation characteristics of individual isolates will perhaps be facilitated by using whole genome sequence data to characterise all the genes involved in sporulation, and also, by studying *in vitro* sporulation phenotypes to assess the rate at which clinically problematic isolates are forming spores. Using this strategy to investigate sporulation of clinically problematic isolates is something which was employed during research for this thesis and will be highlighted in Chapters Eight and Nine.

7.9 Non- toxin Mediated Virulence Factors; Bacterial Motility

Bacterial motility is inherently important to most bacterial species for the colonisation of abiotic or biotic surfaces, to optimise growth and survival and to evade desiccation in a hostile host environment[92]. Swimming and swarming motility of bacteria requires the assembly of complex machinery called the flagellar[115] which is also known to play a role in bacterial adherence host cells[12]. Bacterial motility has been linked to the virulence of many bacterial species, and a recent publication by Aubry *et al*. has found that early stage toxin production in *C. difficile* is modulated by the flagellar operon, providing evidence that links flagellar driven motility with toxin production[15]. Genes that govern the motility of *C. difficile* 630 are organised into three distinct flagellar operons[202] (Table 7.1).

Table 7.1- Positions and genes of the flagellar operons in the *C. difficile* 630 genome[202].

Flagellar regulon	Position (bp)	Gene region	Genes encoded
F1 (late stage)	293002-304049	CD0226-CD0240	*fliC, fliD,* and a putative glycosyltransferase
F2	304766-308251	CD0241-CD0244	Flagellar biosynthetic glycan genes
F3 (early genes)	309272-333020	*flgB*- CD0272	*fliF, fliG, fliM flhR, fliR fliA*- σ^{28}homolog

Whole genome comparison studies of *C. difficile* isolates have been used to identify differences in the F1 and F2 regions which might affect motility[202] and virulence[216]. Chapters Eight and Nine will discuss further, the role that flagellar genes may play in *C. difficile* pathogenicity.

7.10 Summary

There is a wealth of genetic information still to be de-convoluted in order to understand *C. difficile* virulence factors, strain evolution, and host-pathogen interaction. It is hoped that the genomic data provided by NGS of multiple strains, from multiple sources will allow detailed discovery of *C. difficile* population structure, facilitate identification of virulence determinants and thus alert clinicians to potential problematic *C. difficile* strains, so that timely intervention methods and clinical care practices can be implemented[42]. However, Stabler *et al*.[202] stipulate that while NGS may facilitate the identification of virulence factors, their possible role in disease would need to be tested both *in vitro* and *in vivo* in order for the data to be conclusive.

7.11 Aim

Results in Part Two of this thesis concluded that there were *C. difficile* ribotypes which were associated with certain infection outcomes, and that are of broader clinical interest in terms of their prevalence in the clinical setting. The aim of the research presented in Chapters Eight and Nine is to gain a deeper insight into these clinically problematic isolates, through whole genome sequencing and phenotypic analysis of virulence mechanisms detailed in this chapter. Whole genome analysis will provide a comprehensive dataset which will be analysed to see if there are genetic changes in virulence related mechanisms. Phenotypic characterisation of these isolates will allow direct *in vitro* observation of the effect of these genetic changes. By aligning whole genome sequences, looking at SNP differences and phenotypic data of

isolates from different ribotypes; that have also come from different infection outcomes, it may be possible to gain a better understanding of differences between *C. difficile* isolates that have been found in cases of recurrent CDI and mortality from CDI.

Chapter Eight. Whole Genome Sequence Analysis of Clinically Relevant Isolates

8.1 Introduction

Statistical analysis of clinical data in Part One of this thesis identified very few clinical markers for certain CDI outcomes. Therefore, statistical analysis was used to investigate whether acquisition of a particular ribotype had any association to the outcome of infection. In Part Two (Chapter Six), statistical analysis revealed that there were some ribotypes significantly associated ($P \leq 0.02$) with primary infection (001, 005, 013, 014, 015 & 056), and others that were found more frequently in cases of recurrent CDI (027, 023 & 078). Some ribotypes were found significantly associated ($P \leq 0.026$) with lower mortality (001, 005, 012, 014, 023, 027 056, 078 & 220) and others were found more frequently in cases of higher mortality (013, 015, 012). It is possible that genetic and phenotypic differences between different ribotypes and between isolates of the same ribotype, may play a role in the outcome of CDI. The previous chapter also highlighted the importance of coupling whole genome sequence data with *in vitro* phenotypic observations to better understand the virulence characteristics of *C. difficile*.

8.1.1 Isolate Selection

Based on the outcome of statistical analysis from Chapter Six and findings in published literature from Chapter Five, (that highlight the emergence of certain ribotypes within the clinical setting), sixteen clinical isolates (Table 8.1) and six reference isolates (Table 8.2) comprising of six ribotypes were chosen for whole genome sequencing.

Ribotype 027 Isolates

Ribotype 027 is the most prevalent ribotype within clinical settings in the UK[7] and is reported to be associated with increased mortality, severity and recurrent infection[73, 229]. Two isolates (FA07001994 & FA07003754) were chosen from a patient with recurrent CDI in 2007. Isolate FA07001994 was associated with mild CDI and isolate FA07003754 was associated with severe CDI (Table 8.1).

Ribotype 078 Isolates

Ribotype 078 is a clinically emerging ribotype, that has been frequently isolated from human, animal and environmental sources[106], is more frequently associated with community acquired CDI, and has also been associated with increased mortality[83]. Four isolates (Table 8.1) were chosen from this ribotype, based on clinical relevance and their association to clinical outcomes such as low mortality (Chapter Six, Table 6.3) and recurrent CDI.

Ribotype 023 Isolates

Ribotype 023 is also seen as an clinically emerging ribotype[7] and has been seen by some, to be associated with more adverse clinical outcomes[42, 58]. This ribotype was frequently seen in (but not significantly associated with) cases of recurrent CDI. In this study cohort, 023 isolates were associated with lower mortality (Chapter Six, Table 6.3), which is consistent with findings from the 2009/2010 CDRN report, which show that this ribotype has a lower percentage residual mortality compare to the other ribotypes[6]. Four isolates were chosen from this ribotype (Table 8.1).

Ribotype 015 Isolates

In the 2009/2010 CDRN report ribotype 015 was found to have the second highest residual mortality after ribotype 027[6], and has had a year-on-year increase in prevalence in the South East and West regions[229]. The same increasing prevalence was seen in this study cohort and

the mean percentage mortality associated with this ribotype was equal to the cohort average of 24% (Chapter Six, Section 6.5). Two isolates from this ribotype were selected for SNP and whole genome sequence analysis. One isolate (FA07011498) was taken from a patient in 2007 who survived the CDI, and one isolate (FA09007583) was taken from a patient in 2009 who died as a result of the CDI. Both these isolates came from patients who had only one episode of CDI as there were no cases of recurrent CDI associated with this ribotype (Table 8.1).

Ribotype 014 Isolates

Ribotype 014 is again seen as clinically emerging, and also has a lower residual mortality than other ribotypes[6]. In Chapter Six this ribotype was found to be significantly associated with lower mortality and primary only CDI. Two isolates from this ribotype (FA07007469 & FA07004464) were selected for whole genome sequence alignment and SNP analysis. Both patients survived the episodes of CDI. However, one patient infected with isolate FA07004464, had a case of mild CDI and the other patient, infected with isolate FA07007469, had a life threatening case of CDI. Both these isolates came from patients who had only one episode of CDI as there were no cases of recurrent CDI associated with this ribotype (Table 8.1).

Ribotype 013 Isolates

Two isolates from this ribotype were selected on the basis that they were significantly associated with increased primary only infection and had a case percentage mortality that was higher than the cohort average of 24% (Chapter Six, Figure 6.5). One isolate (FA08006290) was taken from a patient who died as a result of CDI and the other isolate (FA08012693) was taken from a patient who survived. Both these isolates came from patients who had only one

episode of CDI as there were no cases of recurrent CDI associated with this ribotype (Table 8.1).

8.1.2 Aim

These isolates were selected for whole genome sequence analysis to assess-

1. Whether there were any collective genetic similarities between isolates from different infection outcomes.
2. What types of genes were collectively undergoing mutation.
3. Whether there were any genetic differences in virulence mechanisms (as highlighted in Chapter Seven).
4. Whether there were large scale genetic rearrangements in genomes.
5. Whether there were any SNP changes in isolates during the course of recurrent infection.

Table 8.1- Clinical isolates selected for whole genome sequence analysis.

Clinical Isolates	Ribotype	Severity	Cause of Death*	Number of Episodes	Patient
FA08012693	013	Mild	N/A	One episode	1
FA08006290	013	Severe	CDI was part of or mentioned in death	One episode	2
FA07007469	014	Life Threatening	N/A	One episode	3
FA07004464	014	Mild	N/A	One episode	4
FA09007583	015	Severe	CDI was part of or mentioned in death	One episode	5
FA07011498	015	Severe	N/A	One episode	6
FA07003485	023	Moderate	N/A	Multiple Episodes	7
FA07004080	023	Mild	N/A	Multiple Episodes	7
FA08006661	023	Severe	CDI was not the cause of death	One episode	8
FA08005864	023	Severe	N/A	One episode	9
FA07007522	078	Severe	N/A	Multiple Episodes	10
FA07008490	078	Severe	N/A	Multiple Episodes	10
FA09004991	078	Life Threatening	N/A	One episode	11
FA08006656	078	Severe	CDI was not the cause of death	One episode	12
FA07001994	027	Mild	N/A	Multiple Episodes	13
FA07003754	027	Severe	Patient survived	Multiple Episodes	13

*N/A- Patient survived

Table 8.2- Isolates selected for use as reference genomes during whole genome sequence analysis.

Reference Isolates	Ribotype	Mutation Phenotype / isolation information	NCBI RefSeq/GenBank ID
630Δ*erm*	012	Erythromycin sensitive derivative of 630	(630 strain NC_009089.1)
R20291	027	*Clostridium difficile* R20291 was isolated in Stoke Mandeville Hospital, UK in 2006.	NC_013316.1
M120	078	*Clostridium difficile* M120 is a human strain isolated in the United Kingdom in 2007	FN_665653
CD305	023	Isolated from a 74 year old male with severe symptoms in 2008	N/A
TL174	015	Isolated in 2009	N/A
TL176	014	Isolated in 2009	N/A

Isolates from this panel were then phenotypically characterised to determine whether genetic changes presented in this chapter were having an effect that could be observed *in vitro*, (Phenotypic results are presented in the following chapter).

8.1.3 Method

The isolates listed in Table 8.1 were subjected to whole genome sequencing to determine SNP and large scale genetic differences between them, and reference isolates of the same and different ribotype. Genome sequences from isolates belonging to the 027, 078 and 013 ribotypes were aligned against published sequences (R20291-027: RefSeq: NC_013316.1, M120-078: GenBank: FN665653 & 630-012 RefSeq: NC_009089.1 respectively). Unpublished reference sequences for 014, 023 and 015 ribotypes were kindly provided by T. Lawley at the Wellcome Trust Sanger Centre, to which sequences of clinical isolates of these ribotypes were aligned. SNP detection was performed using SAMtools[141] (mpileup, bcftools and vcftools commands). All SNP sites called were based on homozygous changes (depicted by 1/1 in the Variant Call Format (VCF) file) with a depth of coverage of ≥ 6 reads. Whole genome consensus sequences for clinical isolates were generated using SAMtools, annotated using RAST and aligned to *C. difficile* 630, R20291 and their respective reference sequences, using Mauve[54], to detect any gross changes within the genome sequences. Sequence reads that were not mapped to reference sequences using Bowtie[134], were assembled *de novo* using Velvet[233] and the resulting contigs were annotated by RAST[16].

Comprehensive detail of the sequence analysis pipeline can be found in the Materials and Methods chapter (Chapter Ten, Section 10.14). Details of the Bowtie alignment pipeline, mapping statistics, non-synonymous SNP data, and RAST annotated sequences derived from assembly of unmapped sequence reads can be found in Appendix Two (CD).

8.2 SNP Detection for Ribotype 027 Isolates

Ribotype 027 is the most prevalent ribotype within clinical settings in the UK[7] and is reported to be associated with increased mortality, severity and recurrent infection[73, 229]. Two isolates (FA07001994 & FA07003754) from a patient with recurrent CDI were chosen from this

ribotype, to assess if any genes were undergoing genetic mutation as compared to the reference isolate, and to deduce whether there was any rapid evolution of *C. difficile* isolates between primary and secondary infections, which might give a survival advantage within the host.

Sequences were aligned to the RefSeq NC_013316.1 (R20291, 027) sequence. A total of 64 SNPs were called. BAM/BCF files were used to verify each SNP manually in Artemis. 56 SNPs were either intergenic or arose from ambiguity due to the presence of multiple copies of a transposase B-like gene. Compared to the published reference genome (R20291), six non-synonymous (NS) SNPs (Table 8.3) and two synonymous (S) SNPs were found in genes CDR20291_t6; encoding an Asp tRNA and CDR20291_3497; encoding a hypothetical protein. SIFT (J. Craig Venter Institute™) protein prediction software was used to check the tolerance of the amino acid change caused by all of the NS SNPs within the genes. The software predicted that the amino acid change (V31K) resulting from the T>G SNP in CDR20291_1593; a putative arsenic pump membrane protein precursor, would affect protein function. All amino acid substitutions resulting from other NS SNPs were predicted to be tolerated within the proteins.

Table 8.3- Non-synonymous changes found within two 027 *C. difficile* isolates from one patient with recurrent CDI.

Position(bp) in the Reference	Reference	Change	Region/ genes	Codon and Amino acid change
1187241	C	A	CDR20291_0967 putative solute binding lipoprotein (1186156-1187241)	gac>gaA N>K
1568676	C	A	CDR20291_1323 putative ruberythrin (1568265-1568807)	caa > Aaa; Q>K
1876233	A	C	CDR20291_1593 putative arsenical pump membrane protein (1876139-1877401)	gtt > gGt; V>K
2160266	T	G	CDR20291_1848 putative peptidase (2159345-2160517)	tct > Gct; S>A
2195190	T	A	CDR20291_1879 two-component sensor histidine kinase (2194462-2195370)	aat > aaA D>E
3065836	A	G	CDR20291_r22 23s rRNA (3063277-3066229)	

8.3 Whole Genome Sequence Alignment of Ribotype 027 Isolates

Whole genome consensus sequences of 027 isolates that were annotated by RAST were aligned using Mauve, to *C. difficile* 630 and R20291 sequences. Whole genome alignment revealed that there was no coverage of a 16 Kbp region in clinical isolates relative to *C. difficile* R20291 (Figure 8.1). This region is comprised of twenty one genes (CDR20291_1744–CDR20291_1764) which includes; a two-component response regulator (CDR20291_1748); putative lantibiotic transporters and a toxin/antitoxin module; which may be important at regulating mobile element stability.

Figure 8.1- A small region of whole genome alignment of two *C. difficile* genomes from a patient with recurrent CDI. Green box= magnified region of genetic variability; no sequence coverage 16 Kbp region in clinical isolates.

In order to deduce if genes in the 16 Kbp region were really absent in the clinical isolates, or just poorly mapped, unmapped reads from outputs of Bowtie were assembled *de novo* using Velvet (1.0.18) and contigs were annotated using RAST. RAST analysis identified five putative protein sequences from the contigs of the isolate from the primary infection

(FA07001994) and three from the isolate from the secondary infection (FA07003754). The predicted function of these putative proteins were a glycerol-3-phosphate regulon repressor of the DeoR family, three hypothetical proteins and a phage intergrase from isolate FA0701994, and a phage intergrase and two hypothetical proteins from isolate FA0703754. The genes encoding these proteins were not part of the 16 Kbp region.

8.4 SNP Detection for Ribotype 078 Isolates

Ribotype 078 is a clinically emerging ribotype, that has been frequently isolated from human, animal and environmental sources[106], is more frequently associated with community acquired CDI, and has also been associated with increased mortality[83]. Four isolates (Table 8.1) were chosen from this ribotype, based on clinical relevance and their association to clinical outcomes such as mortality and recurrent CDI (detailed in Chapter Six). Sequences were aligned to the GenBank FN_665653 (M120-078). Results of SNP and whole genome alignment will be presented in this section. Table 8.4 contains SNP site differences for each isolate in the panel relative to the reference genome.

Table 8.4- Summary of SNP site differences in four 078 isolates.

078 ribotype SNP observations	S	NS
Total SNP sites called	101	
Ambiguous SNPs	56	
Differing SNP sites	8	16
SNPs shared	4	16
SNPs in FA07007522 1 (1st recurrent isolate)	7	26
SNPs in FA07008490 (2nd recurrent isolate)	7	26
SNPs in FA08006656 (1o only isolate- non CDI related death)	4	15
SNPs in FA09004991 (1o only isolate- Survived CDI)	8	24

There were only a small number of SNPs in clinical isolates compared to the reference sequence. Across all four genomes there were four synonymous SNP sites and 16 NS SNPs; ten within transposase genes and six (Table 8.5) within putative genes encoding a

phosphopentomutase, a probable secreted protein homolog of yjcM/yhbB in *B. subtilis*, an argininosuccinate lyase, a membrane component of multidrug resistance system, an inosose isomerise and a two-component response regulator.

Table 8.5- Non-synonymous changes found within all 078 *C. difficile* isolates.

Position (bp) in the Reference	Reference	Change	Putative Gene Site	Codon and Amino acid change
1330473	T	G	Phosphopentomutase	ata -> aGa; I >R
1473978	C	G	Probable secreted protein homolog of yjcM/yhbB *B. subtilis*	aca -> aGa; T > R
2751817	A	G	Argininosuccinate lyase	tcg -> Ccg; S >P
2760534	T	G	Membrane component of multidrug resistance system	ttt -> Gtt; F > V
3003590	T	C	Inosose isomerase	act -> Gct; T > A
3621740	G	T	Two-component response regulator	gcg -> gAg; A > E

8.4.1 SNP Detection in Ribotype 078 Sequential Isolates

Sequential isolates analysed in this study came from one patient with two CDIs within a one month period, where infection one occurred in July 2007 and infection two occurred in August 2007. This patient survived both CDI episodes. Looking at the genomes of isolates from recurrent infection could help deduce if they are undergoing rapid mutational change between primary and secondary infections, which might give a survival advantage within the host.

Downstream SNP analysis revealed that all NS and S SNP sites identified (33 in total) in the primary isolate (FA07007522) were present in the secondary isolate (FA07008490). Seven NS SNPs were unique to these sequential isolates (Table 8.6) in genes encoding a putative glycolate dehydrogenase, three hypothetical proteins, an acetylornithine aminotransferase, a trehalose-6-phosphate hydrolase, and a dihydropyrimidine.

Table 8.6 Non-synonymous changes found within two 078 *C. difficile* isolates from one patient with recurrent CDI.

Position (bp) in the Reference	Reference	Change	Putative Gene Site	Codon and Amino acid change
1281180	T	C	Glycolate dehydrogenase	gtg -> gCg; V > A
1772387	C	T	Hypothetical protein	cct -> Tct; P > S
2147524	G	T	Hypothetical protein	cca -> Aca; P> T
2209318	C	A	Acetylornithine aminotransferase	gga -> Tga G > Stop
2284263	C	A	Hypothetical protein	gca -> gTa A > V
3231744	T	C	Trehalose-6-phosphate hydrolase	tat -> tGt; Y > C
3526448	G	A	Dihydropyrimidine dehydrogenase	aca -> aTa; T > I

8.4.2 SNP Detection in Ribotype 078 Isolates from Primary Only Infections

Two isolates were chosen from patients with primary only CDI episodes. One isolate came from a patient who had an infection in 2008 (FA08006656) with an outcome of death, although *C. difficile* was not listed as contributing to death in the death certificate. The other isolate (FA09004991) came from a patient who had an infection in 2009 and survived.

Downstream SNP analysis revealed that in FA08006656, there was only one synonymous SNP in a putative ribonuclease reductase class II gene. All other SNPs present in this isolate were found among all other 078 isolates. This isolate had less SNPs than the other three isolates (Table 8.4).

SNP analysis also revealed that there were eight SNPs unique to FA09004991. Five of these SNPs were NS (Table 8.7) and were found in putative genes encoding a cupin 2 conserved barrel domain protein, an alpha/beta superfamily hydrolase (possible chloroperoxidase), a two-component response regulator, a glucosamine-6-phosphate amino transferase and Fe-S-cluster containing hydrogenase component 2.

Table 8.7- Non-synonymous changes found within FA09004991

Position (bp) in the Reference	Reference	Change	Putative Gene Site	Codon and Amino acid change
2485002	C	T	Cupin 2C conserved barrel domain protein	gga -> Aga; G > R
3206513	G	A	Alpha/beta superfamily hydrolase (possible chloroperoxidase)	cag -> Tag; E > STOP
3621965	C	T	Two-component response regulator	aga -> aAa; R > K
3645874	A	G	Glucosamine-fructose-6-phosphate aminotransferase [isomerising]	ata -> aCa; I > T
3697708	G	A	Fe-S-cluster-containing hydrogenase components 2	gca -> gTa; A > V

8.5 Whole Genome Sequence Alignment of Ribotype 078 isolates

Whole genome consensus sequences of 078 isolates that were annotated by RAST were aligned using Mauve, to *C.difficile* 630, R20291 and M120 sequences. The alignment revealed that there was a 106 Kbp insertion in the M120 reference genome (420000bp-530000bp), relative to the *C. difficile* R20291 sequence (Figure 8.2). *C. difficile* M120 and 630 shared 37 Kbp of this region, and in *C. difficile* 630 it is annotated as conjugative transposon 2 (CTn*2*). It is therefore more likely that CTn*2* has been deleted in *C. difficile* R20291 rather than being inserted in *C. difficile* M120 and 630. All four clinical isolates had no sequence coverage of a 60 Kbp region relative to *C. difficile* M120. This includes the 37 Kbp region annotated as CTn*2* in *C. difficile* 630. Isolate FA08006656 contained 46 Kbp of the 106 Kbp region (420000bp-470000bp) relative to *C. difficile* M120 (Figure 8.2). The 46 Kbp region found in *C. difficile* M120 and FA08006656 contained genes that showed similarity to the region, CDR20291_1425-CDR20291_1464 in *C. difficile* R20291, which contains genes encoding for phage associated proteins and a virulence related protein.

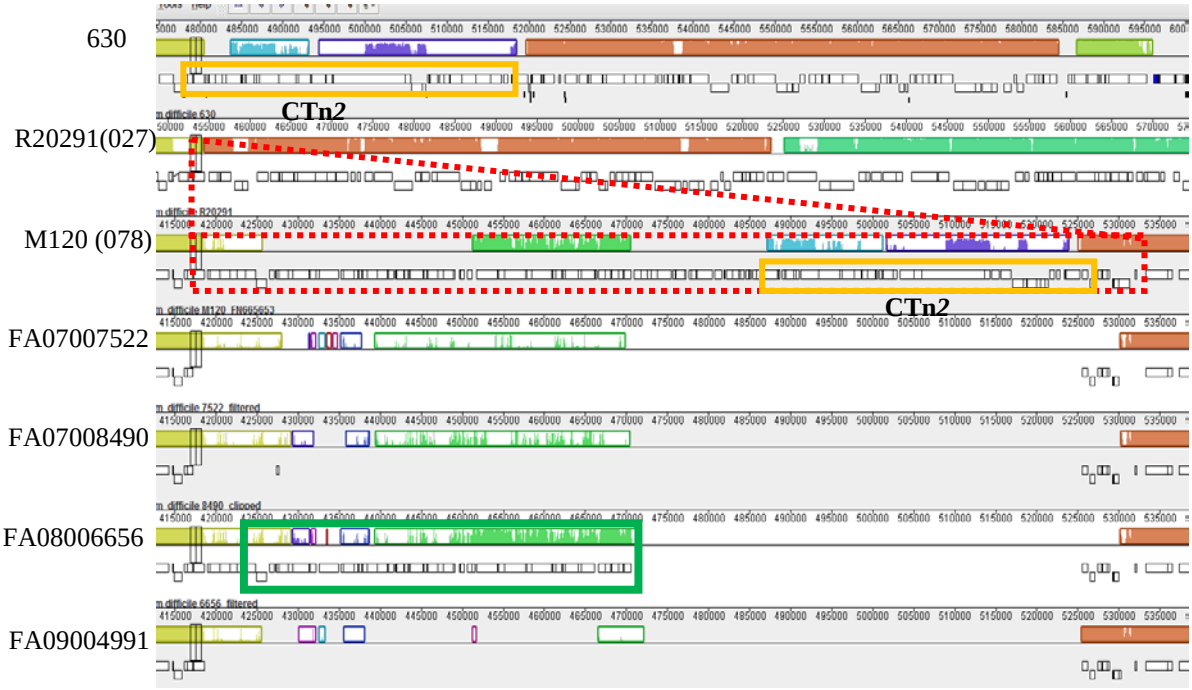


Figure 8.2- Part of whole genome alignment of 078 *C. difficile* genomes. Red hatched box represents the 106 Kbp region in M120 that is not mapped in three clinical genomes (However some genes from region are found in the unaligned contig files). The green box represents the 46 Kbp region contained only in FA08006656 and the reference M120 genome. The orange boxes represent a 37 Kbp region (annotated as CTn*2* in *C. difficile* 630) found in both *C. difficile* 630 and M120.

Unmapped sequence reads from the Bowtie output were assembled *de novo* using Velvet (1.0.18) and the resulting contigs were annotated using RAST in order to deduce whether the genes annotated in the 106 Kbp region of M120 were either poorly mapped, or not present in the clinical isolates. Putative genes annotated in the unmapped contig files of all isolates, included numerous hypothetical proteins, transcriptional regulation genes, phage genes, genes involved in metabolism, *tetM*, a putative tetracycline resistance gene, a lantibiotic transport ATP-binding protein; *srtF*, a spectinomycin 9-O-adenylyltransferase, an aminoglycoside nucleotidyltransferase gene, and two putative virulence genes; encoding a VirB2-like protein and a programmed cell death toxin YdcE (Details of putative genes derived from unmapped sequences can be seen in Appendix Two). These genes could serve as additional virulence determinants.

The putative genes found in the unmapped contig files of isolates FA07007522, FA07008490 and FA09004991 were of similar type to the majority of genes annotated in the 60 Kbp region of M120, indicating that these three isolates may have retained the CTn*2* region. However, these genes were not found in the unmapped contigs file from FA08006656, suggesting that this isolate has not retained the genetic region containing CTn*2*.

FA07007522, FA07008490 and FA09004991 did not contain genes specific to the 46 Kbp region conserved in FA08006656 and *C. difficile* M120.

Other genetic changes in the sequences of the 078 isolates relative to *C. difficile* 630 and R20291 included the loss of the large (F3) flagellar operon (~25 Kbp). The F3 operon has been replaced by a transposase gene (Figure 8.3). A smaller flagellar (F1) operon encoding approximately 10 genes which precedes the region containing the large operon is conserved in the 078 isolates, relative to the *C. difficile* 630 and R20291 reference isolates.

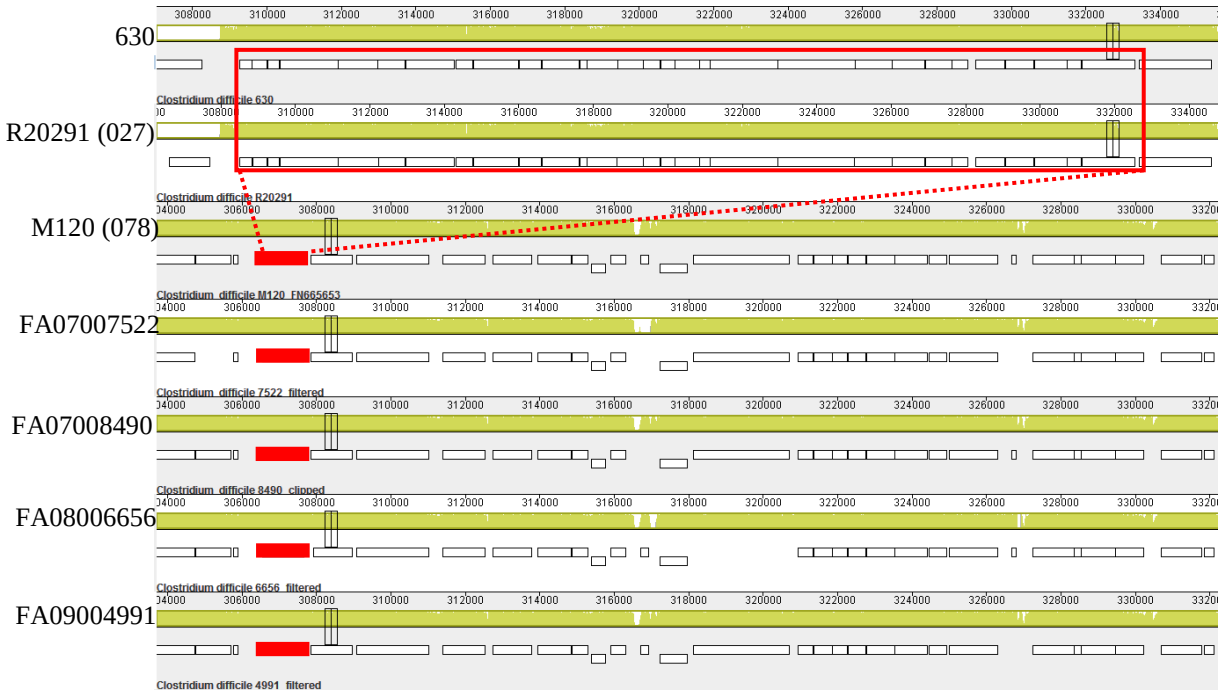


Figure 8.3- A small region of the whole genome alignment of 078 *C. difficile* genomes depicting the loss of the F3 flagellar operon (highlighted in red) from the 078 isolates as compared to the reference isolates. Solid red box indicates the transposase gene which has replaced the F3 operon in 078 genomes.

8.6 SNP Detection for Ribotype 023 Isolates

Ribotype 023 is also seen as a clinically emerging ribotype[7] and has been observed by some, to be associated with more adverse clinical outcomes[42, 58]. This ribotype was frequently seen in (but not significantly associated with) cases of recurrent CDI, in this study cohort. The 023 isolates were also associated with lower mortality (Chapter Six, Table 6.3), which is consistent with findings from the 2009/2010 CDRN report, which show that this ribotype has a lower percentage residual mortality compared to the other ribotypes[6]. Four isolates were chosen from this ribotype (Table 8.1) to look at SNP and large scale genome changes. Results are outlined in the following paragraphs.

A total of 1456 SNPs sites were called across the four genomes relative to the CD305 reference sequence. SNP site differences for each isolate can be seen in Table 8.8.

Table 8.8- Summary of SNP site differences in four 023 isolates.

023 ribotypes		
SNP observations	**S**	**NS**
Total SNP sites called	1456	
Ambiguous SNPs	388	
Differing SNP sites	183	91
SNPs shared	593	201
SNPs in FA07003485 (1st recurrent isolate)	750	273
SNPs in FA07004080 (2nd recurrent isolate)	750	273
SNPs in FA08006661 (1o only isolate- non CDI related death)	755	275
SNPs in FA08005864 (1o only isolate- Survived CDI)	611	212

NS SNP sites found in all four genomes (Appendix Two) include those in numerous flagella genes, *VanW*; a vancomycin B-type resistance gene, a gene encoding a choline binding protein-A (cell surface protein) and *MviM*; a virulence factor.

8.6.1 SNP Detection in Ribotype 023 Sequential Isolates

Sequential ribotype 023 isolates analysed in this study came from one patient with two CDIs within one month in April 2007. This patient survived both CDI episodes.

Downstream SNP analysis revealed that all NS and S SNP sites identified (1023 in total) in the primary isolate (FA07003485) were present in the secondary isolate (FA07004080). There were 7 NS SNPs unique to the sequential isolates, in genes encoding three hypothetical proteins, a glycolate dehydrogenase, phage tail fibers, a biphenyl-22C3-diol-12C2-dioxygenase III-related protein and a cobyric acid synthase.

8.6.2 SNP Detection in Ribotype 023 Isolates from Primary Only Infections

Two isolates were chosen from patients with primary only CDI episodes. Isolate FA08006661 was taken from a patient who died as a result of CDI, although *C. difficile* was not listed on the death certificate. Isolate FA08005864 came from a patient who survived the CDI.

Isolate FA08005864 contained 823 SNPs (212 NS & 611 S) in comparison to the reference sequence (CD305). Five NS SNPs were unique to this isolate and were located in genes encoding a putative imidazole glycerol phosphate synthase amidotransferase subunit, a transcriptional regulator2C of the MecI family, a putative peptidase, phage tail fibre protein, and a hypothetical protein (Table 8.9).

Table 8.9- Non-synonymous changes found within FA08005864.

Position (bp) in the Reference	Reference	Change	Putative Gene Site	Codon and Amino acid change
1726832	G	A	Imidazole glycerol phosphate synthase amidotransferase subunit	gta -> Ata; V > I
2254967	G	A	Transcriptional regulator2C MecI family	cca -> cTa P > L
2769431	T	C	Putative peptidase	ata -> atG I > M
3008689	G	T	Phage tail fibre protein	aca -> aAa T > K
3185734	C	T	Hypothetical protein	gac -> Aac D > N

1030 SNP sites (275 NS & 755 S) were found in isolate FA08006661 in comparison to the reference sequence. A total of 15 NS SNPs were unique to this isolate (Table 8.10).

Table 8.10- Non-synonymous changes found within FA08006661

Position (bp) in the Reference	Reference	Change	Putative Gene Site	Codon and Amino acid change
439233	C	C	Hypothetical protein	acc -> aAc T > N
807428	G	T	Hypothetical protein	aga -> aTa R > I
1513581	A	G	Putative phage tail fiber protein	act -> Gct T > A
2432698	G	T	Hypothetical protein	tgc -> tTc C > F
2791249	G	A	Membrane component of multidrug resistance system	gtt -> Att; V > I
2861532	T	C	hypothetical protein	cca -> Tca P > S
2893887	G	A	Peptidoglycan N-acetylglucosamine deacetylase	aca -> aAa T > K
3172711	C	T	dTDP-4-dehydrorhamnose 32C5-epimerase	gtt -> Att V > I
3215053	C	A	Gluconate permease	gta -> Tta V > L
3713580	C	T	Two-component response regulator	tgc -> tAc; C > Y
3806997	G	T	23S rRNA (Uracil-5-) -methyltransferase RumA	cac -> Aac; H > N
3845704	T	C	Uroporphyrinogen-III methyltransferase/ Uroporphyrinogen-III synthase	aaa -> Gaa; K > E
4252518	A	G	Phage terminase2C small subunit	agt -> Ggt; S > G
4252791	G	A	Phage terminase2C small subunit	gaa -> Aaa; E > K

8.7 Whole Genome Sequence Alignment of Ribotype 023 Isolates

Whole genome consensus sequences from 023 isolates that were annotated by RAST were aligned using Mauve to *C. difficile* 630, R20291 and CD305 sequences. A 140 Kbp insertion was seen upstream of a gene encoding a 50S ribosomal protein (CDR20291_3541; CD3680) relative to the *C. difficile* 630 and R20291 sequences (Figure 8.4). In the sequence of CD305 (023 reference sequence) this region consists of many genes annotated as phage related, and hypothetical related genes, however this region is poorly mapped in clinical sequences, as indicated in the alignment view (Figure 8.4).

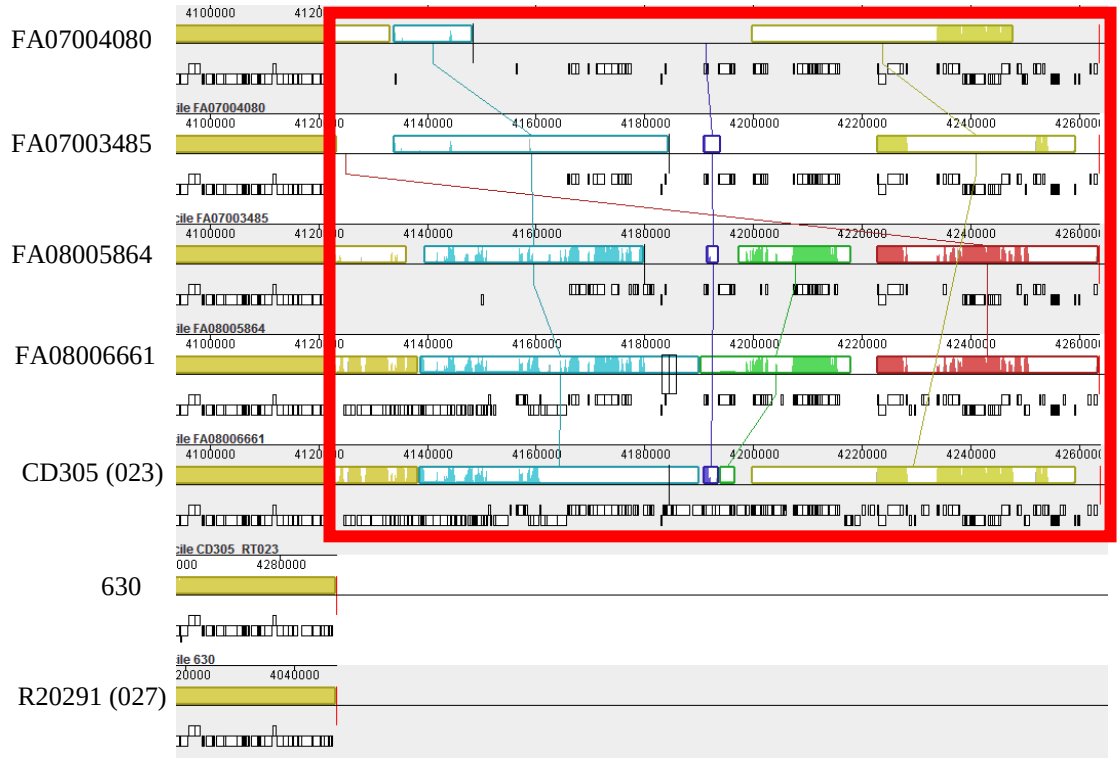


Figure 8.4- Mauve alignment view of all four 023 isolate genomes in comparison to the 630 and R20291 reference genomes. The region highlighted in the red represents the probable insertion of a 140 Kbp region comprised of phage and hypothetical protein related genes.

In order to determine whether genes were not actually represented, or just poorly mapped in the clinical isolates, sequence reads which were not mapped in the original Bowtie alignment were assembled *de novo* using Velvet (1.0.18) and then annotated using RAST. Putative genes that were annotated, encoded numerous hypothetical proteins and a phage related protein. Other genes included *vanW*; encoding a vancomycin B-type resistance protein, multiple flagellar related proteins, *mviM*; a virulence factor, and death on curing protein, doc toxin (Details of putative genes derived from unmapped sequences can be seen in Appendix Two). These genes may be part of the 140 Kbp insertion observed in the reference as they were of similar type.

Sequence alignment also revealed the insertion of a 6 Kbp region between *tcdA* and *tcdB* (Figure 8.5), composed of nine genes and include, a putative transposon intergrase (Tn916

ORF3-like); a Tn916 putative transcriptional regulator, two hypothetical proteins, *merR*, an oxidoreductase of the aldo/keto reductase family, a flavodoxin and gene encoding a putative transcriptional regulator.

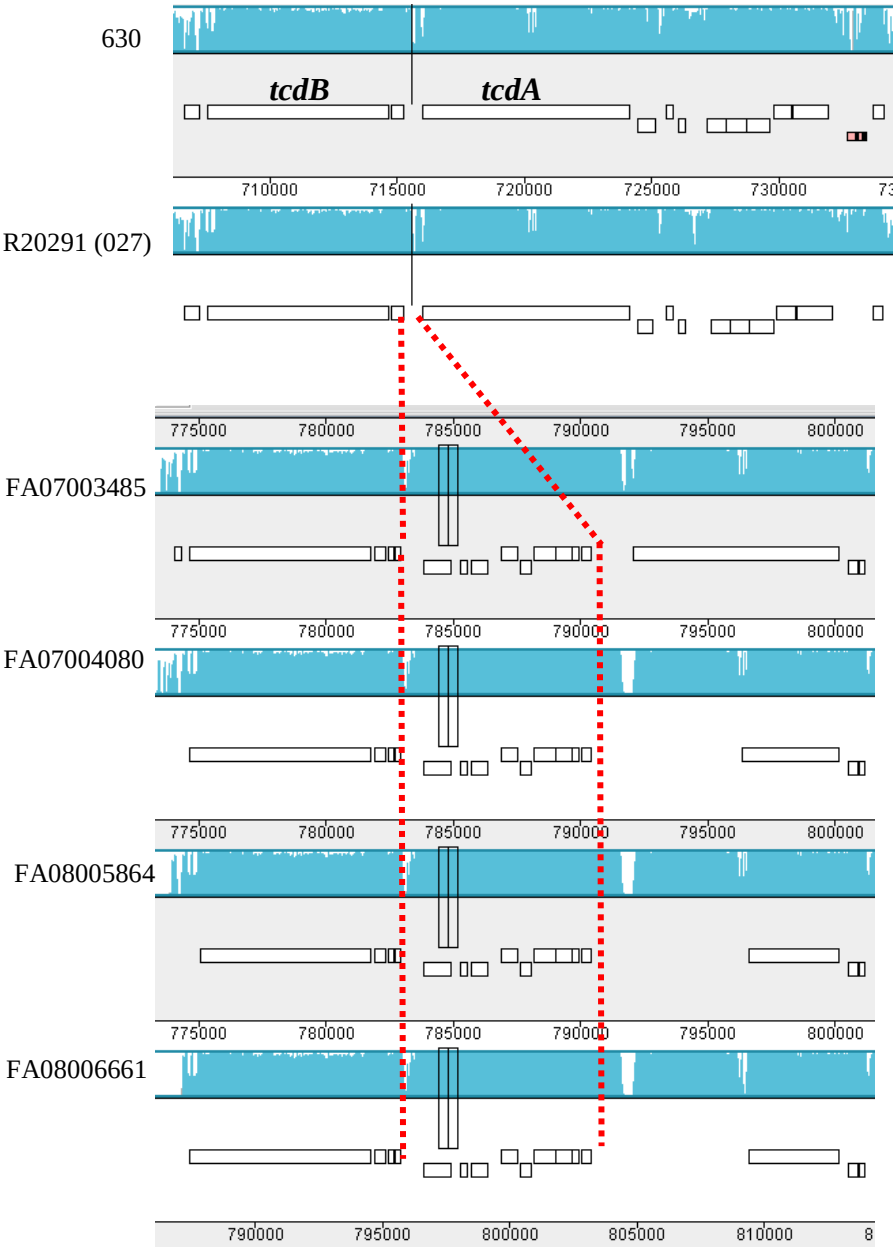


Figure 8.5- Magnified region of the mauve alignment view of all four 023 isolate genomes in comparison to the 630 and R20291 reference genomes. Red hatched lines represent the probable insertion of a 6 Kbp region between *tcdA* and *tcdB* of 023 isolates

8.8 SNP Detection for Ribotype 015 Isolates

In the 2009/2010 CDRN report ribotype 015 was found to have the second highest residual mortality after ribotype 027[6], and has had a year-on-year increase in prevalence in the South East and West regions of England[229]. The same increasing prevalence was seen in this study cohort and the mean percentage mortality associated with this ribotype equal to the cohort average of 24% (Chapter Six, Section 6.5). Two isolates of this ribotype were selected for SNP and whole genome sequence analysis. One isolate (FA07011498) was taken from a patient in 2007 who survived the CDI and one isolate (FA09007583) was taken from a patient in 2009 who died as a result of the CDI. Both these isolates came from patients who had only one episode of CDI as there were no cases of recurrent CDI associated with this ribotype.

A total of 3056 SNP sites were called for the two 015 isolates as compared to the reference genome TL174 (Table 8.11).

Table 8.11- Summary of SNP site differences in the 015 isolates.

015 Ribotypes		
SNP observations	**S**	**NS**
Total SNP sites called	3056	
Ambiguous SNPs	627	
Differing SNP sites	1610	796
SNPs shared	20	10
SNPs in FA07011498 (1^o only isolate-Survived CDI)	21	16
SNPs in FA09007583 (1^o only isolate-CDI related death)	1618	798

Interestingly, there were many more SNPs (S and NS) found in the isolate which came from the patient who died (FA09007583; 1634 SNPs) compared to the isolate from the patient who survived (FA07011498; 819 SNPs). NS SNPs for these isolates can be seen in Appendix Two. In both isolates, 10 NS SNPs were found in genes encoding phage portal proteins, hypothetical proteins, and phage tail fibres (Table 8.12).

Table 8.12- Non-synonymous changes found within both 015 *C. difficile* isolates.

Position (bp) in the Reference	Reference	Change	Putative Gene Site	Codon and Amino acid change
1258843	T	A	Phage portal protein	tca -> Aca; S > T
1262752	A	G	Phage capsid and scaffold	aca -> Gca; T > A
1263588	A	G	Phage capsid and scaffold	atg -> Gtg; M > V
1263612	A	G	Phage capsid and scaffold	atg -> Gtg; M > V
1264025	A	C	Hypothetical protein	gat -> gCt; D > A
1264132	G	A	Hypothetical protein	gta -> Ata; V > I
1264193	T	C	Hypothetical protein	atc -> aCc; I > T
1264319	A	G	Hypothetical protein	ata -> Gta; I > V
4317454	A	G	Phage tail fibers	aat -> aGt; N > S
4330009	T	C	Hypothetical protein	tca -> Cca; S > P

8.9 Whole Genome Sequence Alignment of Ribotype 015 Isolates

Whole genome consensus sequences of 015 isolates that were annotated by RAST were aligned using Mauve, against *C. difficile* 630, R20291 and TL174 sequences. Whole genome analysis revealed the insertion of many small phage and transposable regions when compared to the *C. difficile* 630 and R20291 reference sequences. A large scale 150 Kbp insertion was seen upstream of a gene encoding a 50S ribosomal protein (CDR20291_3541; CD3680) in comparison to *C. difficile* 630 and R20291 reference sequences (Figure 8.6). This region was poorly mapped in clinical isolates. In order to determine if genes were not actually represented, or just poorly mapped in the clinical isolates, sequence reads which were not mapped in the original Bowtie assembly were assembled *de novo* using Velvet (1.0.18) and then annotated using RAST. Putative genes that were annotated, encoded numerous hypothetical proteins and phage related proteins which could be part of this inserted region as

they were of similar type (Details of putative genes derived from unmapped sequences can be seen in Appendix Two).

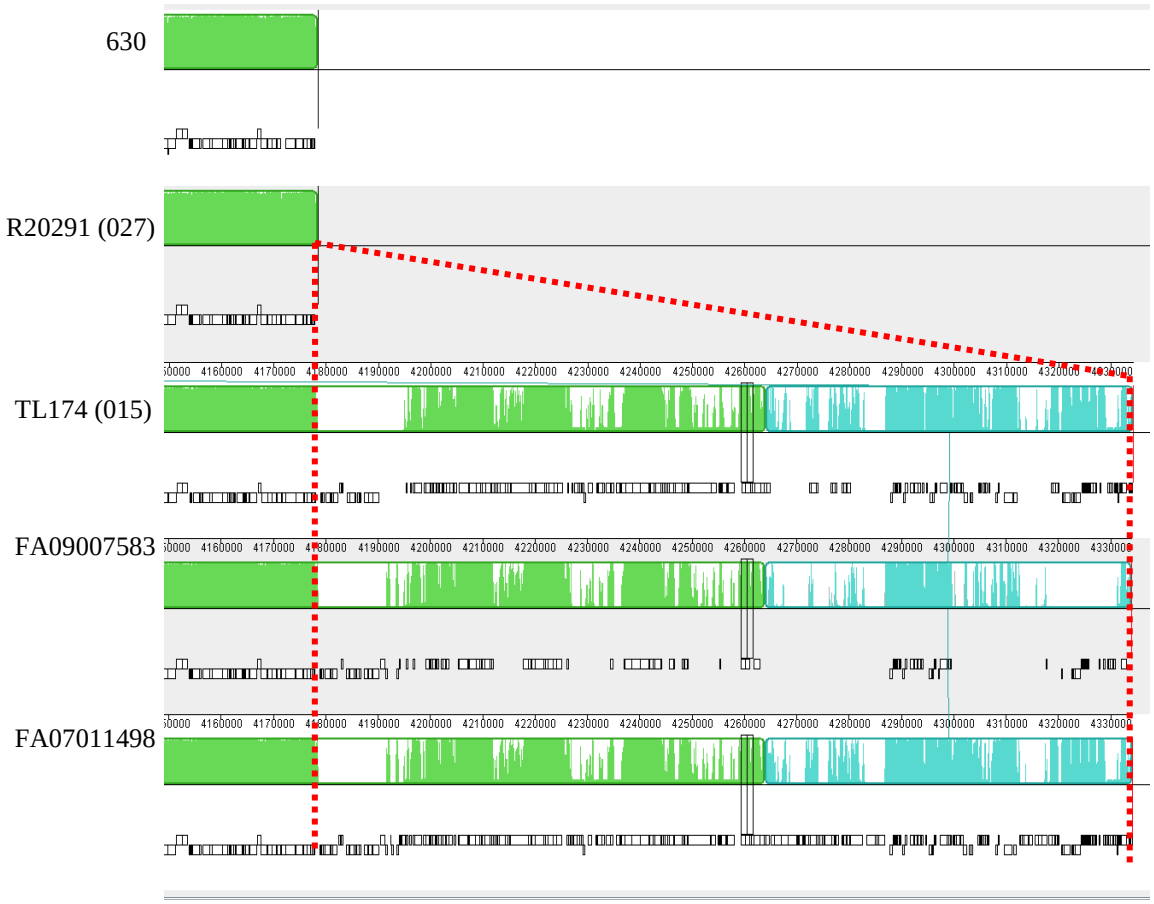


Figure 8.6- Mauve alignment view of ribotype 015 isolate genomes in comparison to the 630 and R20291 reference genomes. Red hatched lines represent the probable insertion of a 150 Kbp region comprised of phage and hypothetical protein related genes.

8.10 SNP Detection for Ribotype 014 Isolates

Ribotype 014 is also seen as clinically emerging, and also has a lower residual mortality than other ribotypes[6]. In Chapter Six this ribotype was found to be significantly associated with lower mortality and primary only CDI. Two isolates from this ribotype (FA07007469 & FA07004464) were selected for whole genome sequence alignment and SNP analysis. Both patients survived the episodes of CDI. However, one patient infected with isolate FA07004464, had a case of mild CDI and the other patient, infected with isolate

FA07007469, had a life threatening case of CDI. A total of 10634 SNP sites were called for the two 014 isolates (Table 8.13) as compared to the reference, TL176.

Table 8.13- Summary of SNP site differences in 014 isolates.

014 ribotypes		
SNP observations	**S**	**NS**
Total SNP sites called	10634	
Ambiguous SNPs	2067	
Differing SNP sites	183	91
SNPs shared	all	2668
SNPs in FA07004464 1 (1^o only isolate-Survived mild disease)	5896	2667
SNPs in FA07007469 (1^o only isolate-Survived life threatening disease)	5896	2668

All S SNPS were conserved in both genomes as compared to the reference. All but one NS SNP was conserved in both isolates. The single different NS SNP was found in the isolate with life threatening infection (FA07007469), at position 877186 bp. The SNP resulted in a G>T nucleotide substitution in a phosphotransfer system (PTS) 2C beta-glucoside-specific IIB component gene and changed the amino acid glycine to valine (All SNP sites can be seen in Appendix Two).

8.11 Whole Genome Alignment of Ribotype 014 Isolates

Whole genome consensus sequences of 014 isolates that were annotated by RAST were aligned using Mauve, to *C. difficile* 630, R20291 and TL176 sequences. Whole genome sequence alignment revealed that in the reference sequence (TL176) there was an insertion of a 190 Kbp region upstream of an ATPase gene (CDR20291_3440; CD3602) relative to *C. difficile* R20291 and 630 references (Figure 8.7). This region was poorly mapped in clinical isolates. In the reference sequence (TL176) genes in this region encode many phage and hypothetical proteins. In order to determine if genes were not actually represented or just poorly mapped in the clinical isolates, sequence reads which were not mapped in the original Bowtie assembly were assembled *de novo* using Velvet (1.0.18) and then annotated using

RAST. Putative genes that were annotated, encoded numerous hypothetical proteins and phage related proteins which could be part of this inserted region as they were of similar type (Details of putative genes derived from unmapped sequences can be seen in Appendix Two).

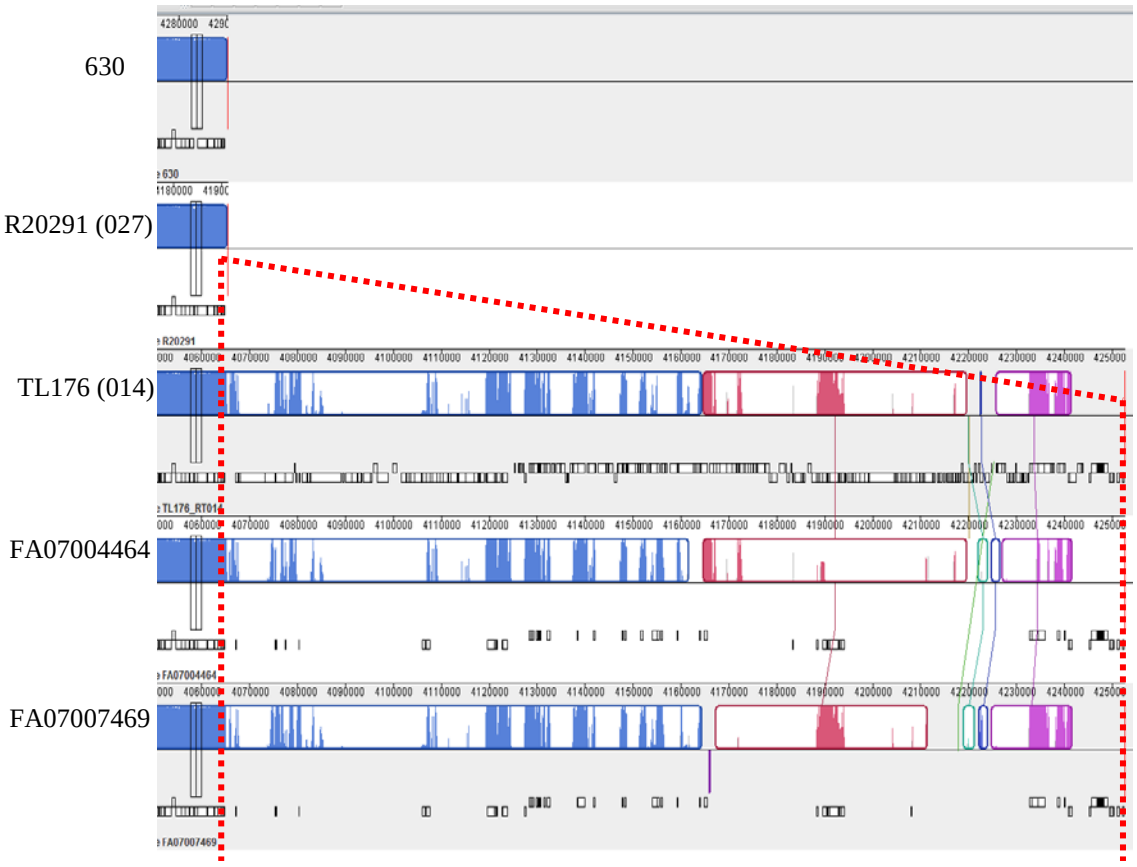


Figure 8.7- Mauve alignment view of ribotype 014 isolate genomes in comparison to the 630 and R20291 reference genomes. Red hatched lines represent the probable insertion of a 190 Kbp region comprised of phage and hypothetical protein related genes.

8.12 SNP Detection for Ribotype 013 Isolates

Two isolates from this ribotype were selected on the basis that they were significantly associated with increased primary only infection and had a case percentage mortality that was higher than the cohort average of 24% (Chapter Six, Figure 6.5). One isolate (FA08006290) was taken from a patient who died as a result of CDI and the other isolate (FA08012693) was taken from a patient who survived.

A total of 11554 SNP sites were called for the two 013 isolates as compared to the *C. difficile* 630 reference sequence (Table 8.14).

Table 8.14- Summary of SNP site differences in 013 isolates.

013 ribotypes		
SNP observations	**S**	**NS**
Total SNP sites called	11554	
Ambiguous SNPs	8916	
Differing SNP sites	573	275
SNPs shared	5687	2381
SNPs in FA08012693 1(1^{o} only isolate- survived)	5884	2468
SNPs in FA08006290 (1^{o} only isolate- CDI related death)	6060	2566

Interestingly, there were more NS and S SNPs (S; 6060, NS; 2566 SNPs) in the isolate (FA07006290) found in patient who died as a result of having CDI, than in the patient who survived (S; 5884, NS; 2468 SNPs respectively) compared to the reference sequence (Details of all SNP sites can be found in Appendix Two).

Eleven of the total number of NS SNPs identified in the 013 isolates were found in putative sporulation related genes. Five of these NS SNPs were unique to the 013 isolates (Table 8.15). Results of sporulation/germination characterisation studies of these isolates (Chapter Nine, Section 9.6, Figure 9.6a) deduced that both isolates did not form heat resistant cfus after 24 hrs; therefore late onset of sporulation could be a possible consequence of these SNPs.

Table 8.15- Non-synonymous SNPs in putative sporulation related genes.

Position (bp) in the Reference	Reference	Change	Putative gene	Codon and Amino acid change
945641	C	G	CD0774 stage V sporulation protein AD	gac -> gaG; D>E
1262598	C	T	CD1068 putative polysaccharide biosynthesis 2F sporulation protein	gca -> Aca; A>T
4081725	C	A	CD3490 stage II sporulation protein E	gca -> Tca; A>S
4281295	C	A	CD3671 stage 0 sporulation protein J	gga -> gTa; G>V
1729574	A	G	CD1492 putative two-component sensor histidine kinase	att -> Gtt; I>V

8.13 Whole Genome Alignment of Ribotype 013 Isolates

Whole genome consensus sequences of 013 isolates that were annotated by RAST were aligned using Mauve, to *C. difficile* 630. Many small scale genetic rearrangements were observed when the genomes were aligned, and this is probably due to the diversity of the genome from *C. difficile* 630. The small gaps in the genomes compared to the *C. difficile* 630 reference were found in genetically mobile regions containing conjugative transposons (Ctns) (CTn*1*, CTn*2*, CTn*3* and CTn*4*). There was also poor coverage of a region between CD0904-CD0979 and CD2898-CD3156 as compared to the reference sequences, these two regions are annotated as phage 1 and phage 2 in *C. difficile* 630. Some of the genes in these regions were contained in the RAST annotated gene file derived from the assembly of unmapped sequence reads (Appendix Two).

8.14 Sequence Analysis of Toxin Genes

Genetic changes in the genes of the PaLoc (Chapter Seven, Section 7.7) have been correlated with phenotypic changes in toxin production[53]. Ribotypes with variant *tcdA* and *tcdB* genes have also been correlated with the presence of genes encoding CDT[207]. Genomes of all isolates were analysed for the presence of *tcdA*, *tcdB* and CDT genes using Mauve and were marked for their presence or absence (+/-). Multiple alignment of *tcdC* DNA sequences from all isolates were made using the Global-Ref alignment software in Clone Manager Professional Suite version 8 (Scientific & Educational Software, USA) to see if there were any genetic changes, relative to the *C. difficile* 630 and R20291 genomes, that may possibly affect *in vitro* toxin production. *TcdC* variants were assessed using the same nomenclature as indicated by Dingle *et al.*[58]. Results of genetic changes are presented in Table 8.16.

All clinical isolates encoded genes for TcdA and TcdB, and only isolates from ribotypes 023, 078 and 027 encoded genes for CDT. Sequence analysis of *tcdC* revealed that isolates from

ribotypes 023 and 078 showed a single nucleotide substitution (C>T) at sequence position 183 bp which results in a premature stop codon at amino acid 62 (TAAStop 62), and isolates from ribotypes 027 show a single nucleotide deletion at position 117 bp resulting in a premature stop codon at amino acid 66 (Δ1Stop66) (Table 8.16). These premature stop codons result in truncated TcdC[58] and this could have implications on toxin production as highlighted in Chapter Seven. Toxin production of all isolates was investigated, and results are presented in Chapter Nine (Figures 9.2, 9.3 and 9.4).

Table 8.16- Presence/absence of toxin genes and *tcdC* genotypes that were found in clinical isolates.

Isolate ID	Ribotype	*tcdA*	*tcdB*	CDT genes	TcdC variants
FA07003485	023	+	+	+	TAAStop 62
FA07004080	023	+	+	+	TAAStop 62
FA08005864	023	+	+	+	TAAStop 62
FA08006661	023	+	+	+	TAAStop 62
FA07007522	078	+	+	+	TAAStop 62
FA07008490	078	+	+	+	TAAStop 62
FA08006656	078	+	+	+	TAAStop 62
FA09004991	078	+	+	+	TAAStop 62
FA07001994	027	+	+	+	Δ1Stop 66
FA07003754	027	+	+	+	Δ1Stop 66
FA07011498	015	+	+	-	WT
FA09007583	015	+	+	-	WT
FA07004464	014	+	+	-	WT
FA07007469	014	+	+	-	WT
FA08012693	013	+	+	-	WT
FA08006290	013	+	+	-	WT

8.15 Sequence Analysis of Antibiotic Resistance Genes

Within the clinical environment *C. difficile* infection is exacerbated by inappropriate use of antibiotics such as clindamycin and the fluoroquinolones, such as ciprofloxacin, levofloxacin and moxifloxacin[42]. Ciprofloxacin was prescribed to some patients in the study cohort and therefore may have facilitated CDI. Current therapies for treatment of CDI at the time of this

study were metronidazole and vancomycin. There is concern that strains/genotypes of *C. difficile* are emerging that show reduced susceptibility to metronidazole[7] and that overuse of vancomycin is promoting increased incidences of VRE[10]. Resistance to fluoroquinolones results from mutations in the quinolone resistance-determining region (QRDR) of DNA gyrase subunits GyrA or GyrB[103]. Genomes of all clinical isolates (Table 8.1) were analysed for the presence of known amino acid changes in GyrA and GyrB by multiple alignment of amino acid sequences, using the Global-Ref alignment software in Clone Manager Professional Suite version 8 (Scientific & Educational Software, USA). Whole genome consensus sequences annotated by RAST were aligned using Mauve, and were used to identify other putative genes which could be involved in antibiotic resistance. Amino acid changes in gyrase genes and presence of other genes that could possibly affect antibiotic sensitivity of isolates are presented in Table 8.17.

The Thr82Ile amino acid substitution in GyrA relative to the sequence of *C. difficile* 630, which has been proposed to confer fluoroquinolone resistance[60, 202], was found in all ribotype 027 and 014 isolates. Two other previously defined amino acid substitutions; Leu406Ile and Asp468Asn[202], were also found in the 027 isolates, but not in any other isolate. Two novel amino acid mutations (Asp205Glu and Lys413Asn) were found in GyrA of ribotype 023 and ribotype 078 isolates respectively.

A previously described amino acid substitution (Asp426Val) in GyrB[61] which was present in a fluoroquinolone resistant $TcdA^{-}TcdB^{+}$ strain, was not found in any isolate in this study, but two other amino acid substitutions; Ser366Val[61] and Ser416Ala[201], not thought to be involved in fluoroquinolone resistance[61, 201], were present in all 078 isolates. Three novel amino acid substitutions in GyrB (015-Val130Ile, 078-Gln160His & Ser416Ala), were found in all 015 and 078 isolates. Amino acid changes in GyrA and GyrB could have implications in the

resistance of isolates to newer fluoroquinolones such as levofloxacin, and moxifloxacin, and will be discussed further in Section 8.16.

Table 8.17- Amino acid substitutions in GyrA and GyrB and presence of putative genes that could be involved in antibiotic resistance.

Isolate ID	Ribotype	Amino Acid Substitutions		Presence of Putative Vancomycin Resistance Genes	Presence of Putative Metronidazole Resistance Genes
		GyrA	GyrB		
FA07003485	023	Asp205Glu	n/a	*VanW* (x2) & *VanZ**	5-nitroimidazole reductase
FA07004080	023	Asp205Glu	n/a	*VanW* (x2) & *VanZ**	5-nitroimidazole reductase
FA08005864	023	Asp205Glu	n/a	*VanW* (x2) & *VanZ**	5-nitroimidazole reductase
FA08006661	023	Asp205Glu	n/a	*VanW* (x2) & *VanZ**	5-nitroimidazole reductase
FA07007522	078	Lys413Asn	Gln160His Ser366Val* Ser416Ala*	*VanW* (x2) & *VanZ**	5-nitroimidazole reductase
FA07008490	078	Lys413 Asn	Gln160His Ser366Val* Ser416Ala*	*VanW* (x2) & *VanZ**	5-nitroimidazole reductase
FA08006656	078	Lys413 Asn	Gln160His Ser366Val* Ser416Ala*	*VanW* (x2) & *VanZ**	5-nitroimidazole reductase
FA09004991	078	Lys413 Asn	Gln160His Ser366Val* Ser416Ala*	*VanW* (x2) & *VanZ**	5-nitroimidazole reductase
FA07001994	027	Thr82Ile* Leu406Ile* Asp468Asn*	n/a	*VanG**, *VanTG**, *VanR**, *VanS**, *VanW* (x2), & *VanZ**	5-nitroimidazole reductase (CD_R20291_1308)
FA07003754	027	Thr82Ile* Leu406Ile* Asp468Asn*	n/a	*VanG**, *VanTG**, *VanR**, *VanS**, *VanW* (x2),& *VanZ**	5-nitroimidazole reductase
FA07011498	015	n/a	Val130Ile	*VanG**, *VanTG**, *VanR**, *VanS**, & *VanW* (x2)	5-nitroimidazole reductase
FA09007583	015	n/a	Val130Ile	*VanG**, *VanTG**, *VanR**, *VanS**, *VanW* (x2) & *VanZ**	5-nitroimidazole reductase
FA07004464	014	Thr82Ile*	n/a	*VanG**, *VanTG**, *VanR**, *VanS**, & *VanW* (x2)	5-nitroimidazole reductase
FA07007469	014	Thr82Ile*	n/a	*VanG**, *VanTG**, *VanR**, *VanS**, & *VanW* (x2)	5-nitroimidazole reductase
FA08012693	013	n/a	n/a	*VanG**, *VanTG**, *VanR**, *VanS**, *VanW* (x2), & *VanZ**	5-nitroimidazole reductase
FA08006290	013	n/a	n/a	*VanG**, *VanTG**, *VanR**, *VanS**, *VanW* (x2), & *VanZ**	5-nitroimidazole reductase

* Known amino acid substitutions and putative resistance genes[61, 197, 201, 202].

RAST annotated whole genome sequence alignments of all isolates revealed the presence of a putative *vanG*-like operon, which is also present in the genome of *C. difficile* 630[197]. This operon consists of *vanR*/*vanS*; encoding a regulator-sensor kinase, *vanG*; encoding a putative vancomcyin/teicoplanin resistance protein and *vanTG*; encoding a serine/alanine racemase. This operon was conserved in all isolates from ribotypes 013, 014, 015 and 027. Two copies of *vanW*; a gene found in vancomycin resistant *Enterococcus faecalis*, but assigned unknown function[164]; and one copy *vanZ*; encoding a putative teicoplanin resistance gene, were found in all isolates from ribotypes 013, 027, 023 078 and one isolate (FA09007583) from ribotype 015 (Table 8.17). All isolates also contained a gene encoding a putative 5-nitroimidazole reductase, which showed over 95% DNA sequence identity to genes CDR20291_1308 and CD1459 from R20291 and *C. difficile* 630 respectively (Table 8.17).

8.16 Discussion

The aim of work in this chapter was to compare whole genome sequences of clinical isolates to one another, and their respective reference counterparts, to try and identify any SNP differences in specific virulence genes, other genes, and/or large scale genomic rearrangements that could lead to different *in vitro* phenotypes. Discussed below are the key findings of this whole genome sequence analysis.

8.16.1 General SNP Inferences

One of the aims of genetic analysis of isolates in this study was to determine whether the genomes of *C. difficile* isolates from patients with recurrent CDI were undergoing accelerated mutational changes in genes that could facilitate their survival in the gut. Current literature suggests that some bacteria might be able to acquire antibiotic resistance during the course of an infection, which results in a prolonged infection. This is exemplified in a study by Hornsey *et al*., who show that an increasing problematic HCAI pathogen; *Acinetobacter*

baumanii, may have acquired resistance to tigecycline during the course of therapy[99]. This group found 18 SNP differences from the isolate that caused the first infection compared to the isolate at the end of the infection. One of these SNPs was in *AdeS* and this was believed to be responsible for up-regulation of the AdeABC efflux system, which could have been responsible for tigecycline resistance. However, isolate heterogeneity at the time of bacterial isolation could also have influenced the outcome. Evidence from research during the present study suggests that the genome of *C. difficile* is not under short term selection pressure. The analysis of six genomes from three ribotypes, from patients with recurrent infection, revealed that there were no additional SNPs in the genome of the secondary isolate compared to the genome of the primary isolate. This data corroborates well with findings by Kennemann *et al*., who found only one point mutation in the genome of a *Helicobacter pylori* isolate sequenced at the beginning of an infection, and then three months later[124]. They deduced that other genetic changes seen during the course of infection were the result of population heterogeneity and the time of initial infection, and perhaps highlights the potential problem of this phenomenon when analysing genomes of *C. difficile* isolates that have come from different stages of recurrent infection.

Throughout the genomes of different ribotypes, NS SNPs were found within specific groups of genes in the population, which could indicate the how the niche environment in which *C. difficile* resides is playing a role in the selective pressure on the genome. These SNPs were seen in genes such as histidine kinases, response regulators, which are sensory mechanisms, resistance genes and genes involved in metabolism. This is exemplified by the NS SNP that was conserved in CDR20291_1593, encoding a putative arsenic pump membrane protein of the 027 isolates. The amino acid change resulting from this SNP was predicted to affect

protein function. Thus, a mutation in this type of gene may indicate its role in the tolerance of isolates to heavy metals in a host, or clinical environment.

There were also interesting NS SNPs found in putative sporulation genes that were unique to the 013 isolates (Table 8.15). Results of sporulation characterisation experiments (Chapter Nine, Figure 9.6a) revealed that ribotype 013 isolates showed delayed sporulation/germination. This could possibly be explained by the sporulation genes in which the SNPs are located. Underwood *et al*.[217] proposed that the histidine kinase gene, CD1492, is likely to be involved in activation of *Spo0A*, relaying signals from environmental stimuli to *Spo0A* via the two-component signal transduction pathway. Other sporulation related genes such as CD3671, encoding a stage 0 sporulation protein J and CD3490, encoding a stage II sporulation protein E, have been implicated in cellular partitioning[137] and direction of early forespore-specific gene expression respectively in *Bacillus subtilis*[19, 44]. Therefore, it is possible that the cumulative effect of NS mutations in these sporulation related genes could be delaying the onset of sporulation/germination in these isolates. However, this is only speculation, and detailed analysis of the effect that the amino acid changes would have on the proteins would need to be conducted to test this hypothesis. Within the genome of the 013 isolates, eleven putative sporulation genes contained NS SNPs that were not unique to these isolates. However, these SNPs were not found in any of the other isolates that showed delayed onset of sporulation/germination.

A NGS study of *C. difficile* isolates performed by He *et al*.[94] revealed that between divergent lineages of isolates there is strong purifying selection; however, this strong purifying selection is not enforced for recently diverged lineages suggesting that NS SNPs are not purged efficiently or that the effects of purifying selection are delayed, and may be due to the narrow host range for this species. SNP differences observed between isolates could be the result of delayed purifying selection, but may highlight what types of genes are undergoing

change. He *et al.*[94] also investigated genes under positive selection and found a total of twelve genes, including predicted membrane and exported proteins, and response regulators, which they believe are under host immune derived selection. Changes in genes of this type occur commonly in isolates under investigation in the present study, and perhaps signify the importance of the host-pathogen interaction that drives the diversity of the *C. difficile* genome. He *et al.*[94] also make an important observation about their analysis, in that they only identified genes under positive selection in the core CDSs, but there could be more genes under selection in the accessory or 'pan-genome' which is something reflected in data from this study. During the current study, the presence of SNPs conserved between isolates from different time periods may suggest that these SNPs have been fixed in the population and therefore could be advantageous to their survival.

Within ribotypes 015 and 013 there seemed to be many more NS SNPs in the isolates from patients whose death was a consequence of CDI, than the isolates from patients who survived (Table 8.18).

Table 8.18 – Number of NS SNPs in isolates from different mortality outcomes.

Ribotype \ Outcome	Number of NS SNPs	
	Died	**Survived**
015 (N=2)	798	16
013 (N=2)	2566	2468

However, in other ribotypes, where isolates had come from patients who had died from non-CDI related causes (FA08006656 & FA08006661), there was either a decrease in the number of SNPs (FA08006656) or only a small increase in the number of SNPs (FA08006661) compared to isolates from patients who did not die.

The increased abundance of SNPs in isolates that had come from patients who died as a consequence of CDI, compared to isolates from patients who either did not die, or died from non-CDI related causes could, in part, be due to factors such as, an increased selective force

that is driving change in certain genes of isolates associated with CDI related death. However, more isolates would need to be sequenced in order for data to be conclusive. The increase in SNPs could equally be due to the timescale of isolation of clinical samples compared to their respective references. This could mean that there had been insufficient time for SNPs to be purged in isolates from later time periods with respect to their reference sequences.

SNP differences in sequences of two isolates (FA07004664 & FA07007469) from different severity outcomes were compared to establish if SNP markers of severe infection identified by Forgetta *et al.*[71] were present. It was surprising to see that both genomes only differed by one SNP. The isolate that caused the life threatening CDI (FA07007469) had a SNP in a gene encoding a PTS 2C beta-glucoside-specific IIB component, which is not likely to account for the difference in infection outcome in patients from which these isolates were taken. Forgetta *et al.*[71] studied the genomes of fourteen *C. difficile* isolates to try and identify markers of severe infection, and identified candidate SNPs in genes such as CD1269 (part of a two-component gene system) and CD1265 (part of a ABC transporter gene system), which were found to associate more closely with disease severity than currently used diagnostic markers. These SNPs were not found in genomes from ribotype 014 isolates relative to *C. difficile* 630, nor were they found in any ribotype 015 isolates associated with severe disease. However, one SNP was found in ribotype 013 isolates that caused both mild and severe disease, and numerous SNPs were found in all isolates from 023, 027 and 078, ribotypes, regardless of the infection outcome to which they were associated. This indicates that perhaps these SNP markers are not so useful for determining the likelihood of severe infection, and rather that they are indicative of diverse *C. difficile* lineages.

While it is relatively difficult to attribute any of the SNP changes observed during the present study, to a particular outcome, this kind of data may be useful to track the type of genes in

which SNPs are accumulating, and thus identify possibly regulatory pathways that are exploited by *C. difficile*.

Given this difficulty, and the necessary validation required to assess if any SNPs could be attributed to a certain infection outcome, alignment of whole genome sequences was performed to see if there were any inter/intra, gross genetic changes in genome sequences which might contribute to different *in vitro* phenotypes.

8.16.2 Inferences from Whole Genome Sequence Alignment

The genome of *C. difficile* is highly mobile, and sequence annotation of *C. difficile* 630 has revealed that it contains mobile genetic elements including bacteriophages, insertion (IS) elements, IStrons (a chimera of an IS element and a group I intron) and putative CTns: CTn*1*, CTn*2*, CTn*4*, CTn*5*, CTn*6* and CTn*7*[37, 157, 197]. These mobile elements are believed by some to give *C. difficile* the ability to acquire and disseminate a large variety of accessory genes via horizontal gene transfer, and thus sample the metagenome of bacteria residing in the gut[36]. Whole genome sequence alignment of isolates relative to one another and their reference counter parts has revealed that the insertion of large phage elements is conserved within isolates of the same ribotype, but diverse between different ribotypes.

Large genetic regions encoding phage related proteins are seen to be inserted at different sites between ribotypes. In the *C. difficile* 630 genome there are two annotated prophages (P1; CD0904-CD0979 and P2; CD2889-CD2952) totaling 102 Kbp[157]. Regions of similar size are found integrated at the end of the genomes in ribotypes, 023, 015, and 014 (Figures 8.5, 8.8 & 8.9) compared to 078, 027 and 013 ribotypes, which have smaller regions of phage related genes integrated at different sites.

More subtle differences in insertion sites were observed for the integration of transposable elements. Whole genome alignment of clinical 027 sequences against R20291 (Figure 8.1)

shows the loss of a 16 Kbp region, recently called Tn6104[37]. Tn6104 encodes a two-component response regulator, putative lantibiotic transporters and a toxin/antitoxin system. Tn6104 is part of a larger phage island, more recently termed Tn6103[37]. Tn6103 comprises of a further twenty four genes (CDR20291_1765_CDR20291_1788); split into two transposable elements (Tn6105 and Tn6106). Tn6103 is hypothesised to be attributed to the virulence of the epidemic 'hypervirulent' BI/NAPI/027 strain[202], as it is missing from *C. difficile* 630. FA07001994 and FA07003754 have coverage of both Tn6105 and Tn6106 sequences; therefore it would be interesting to see whether other 027 ribotypes also showed the loss of Tn6104, and whether its loss was correlated with decreased virulence of this ribotype.

Within the genomes of the 078 isolates there was differential conservation of a 106 Kbp region (Figure 8.2). In comparison to the *C. difficile* M120 (078) reference sequence, isolates FA07007522, FA07008490 and FA09004991 have retained a 60 Kbp portion of this region, encoding a *C. difficile* 630 CTn*2*-like region, which is missing in isolate FA08006656. However, the remaining 46 Kbp region (420000bp-470000bp) identified in *C. difficile* M120 is conserved in isolate FA08006656. Genes in this region were not mapped, nor found in the RAST annotated unaligned contig file of the isolates containing the 60 Kbp region. The 46 Kbp region conserved in isolate FA08006656 encodes genes for methylases, a σ^{-70} factor, phage proteins, a virulence related protein and a holin/toxin secretion/phage lysis protein. The 078 ribotypes have been liked to more severe infection outcomes[53], and the 46 Kbp region conserved in FA08006656 does contain toxin and virulence related genes, which may increase its virulence. This isolate also came from a patient with CDI who died, and although the death was not attributed to CDI directly, conservation of the 46 Kbp region may have facilitated a more the more aggressive infection outcome. Thus, the insertion of these transposable elements may contribute to the observation of increased hypervirulence of the 078 ribotype.

Further genetic difference of the 078 ribotypes compared to *C. difficile* R20291 and 630 was the deletion of the large F3 flagellar operon, which may explain the non-motile phenotype of these isolates (Chapter Nine, Figure 9.1). This operon consists of the early stage genes; including *fliF, fliG, fliM, flhR, fliR* and *fliA*; a σ^{-28} homolog, which governs the late genes found in the F1 operon (*fliC, fliD* and a glycosylation gene). The F1 operon and a small chemotaxis operon are conserved among this ribotype, however the global regulator of these genes; *fliA*, a σ^{-28} homolog, is missing.

Whether the F1 operon is redundant or expressed is unclear. However, a possible reason for its conservation could be due to the link that has been observed between flagellar gene regulation and toxin gene expression[15]. Work by Aubry *et al.,* suggests that genes involved in the regulation of the flagellar operon, also regulate toxin production[15]. If flagellar and toxin gene regulation were coupled, then the regulation of the small conserved F1 and chemotaxis operon in the 078 ribotypes may be coupled to toxin production, thus necessitating their evolutionary conservation. Another possible explanation for the conservation of the F1 operon is the role it may play in adherence to colonic epithelia. Contradictory results pertaining to the role of FliC and FliD in *C. difficile* adherence has been demonstrated by Tasteyre *et al*.[210] . They concluded that *in vitro*, these genes may have a role in adherence, as recombinant FliC and FliD proteins were able to bind to murine mucus, but the observation that they were implicated in adherence was not repeatable *in vivo* when axenic mice were challenged with flagellated and non-flagellated strains. However, the group do suggest that choice of mouse for challenge may have had an impact on the results, as axenic mice do not have an epithelial barrier and thus could have facilitated adherence of the non-flagellated mutants.

Studies by Lillehoj *et al*. concluded that FliC (flagellin) is necessary for adhesion of *Pseudomonas aeruginosa* to Muc1 mucin on lung epithelia[144]. They also conclude that

adherence is independent of motility by showing that adherence to CHO-Muc1 cells; by wild type bacteria rendered non-motile by treatment with Gramicidin D, is the same as adherence for fully motile wild type strains. This observation may in part explain the retention of the F1 operon containing both *fliC* and *fliD*, in the 078 ribotypes. If adherence is independent of motility, then there may be a small possibility that the genes of the F3 operon are not needed for motility and only the small F1 gene operon is necessary for adhesion of this ribotype to colonic epithelia. It then follows that an alternative regulatory pathway for the F1 genes must also be in place, as the *fliA* σ^{28} homolog which governs regulation of the late genes is missing. The link between motility and adhesion has been studied in *C. difficile*. Adherence of a non-motile *C. difficile fliC* mutant to differentiated Caco-2 tissue culture cells was increased compared to wild type *C. difficile*, in studies conducted by Dingle *et al*.[59]. They believe that repression of motility might be a pathogenic strategy employed by *C. difficile* to establish infection within the host. It would be interesting to establish if the small F1 operon was being expressed *in vitro* by the 078 isolates, to further clarify their regulatory status. It would also be interesting to repeat the adherence studies, as conducted by Dingle *et al*.[59], with natural non-motile and motile *C. difficile* isolates. This would be to test whether adherence of *C. difficile* to Caco-2 tissue culture cells was a flagellar dependent process, and not specific to deletion of *fliC* or *fliD,* as the polar effect observed by the deletion of *fliC* or *fliD* was the loss of flagellar, which rendered the mutants non-motile.

In isolates from the 023 ribotype, there was a 6 Kbp insertion between *tcdA* and *tcdB* (Figure 8.5). This insertion is composed of nine genes that include a putative transposon intergrase (Tn916 ORF3-like), a Tn916 putative transcriptional regulator, two hypothetical proteins, *merR*, an oxidoreductase of the aldo/keto reductase family, a flavodoxin and a gene encoding a putative transcriptional regulator. Personal communication with Dr Kate Dingle (Department of Microbiology, Experimental Medicine Division, Nuffield Department of

Medicine, University of Oxford) has revealed that these genes are all involved in the transfer of electrons during metabolic processes and they may have a functional role in adaptation to oxidative stress. The insertion of these genes between *tcdA* and *tcdB* may affect the production of TcdA and TcdB. Toxin production data (Chapter Nine, Figure 9.4c) from this ribotype revealed that TcdA is being produced later, and at lower levels than all the other ribotypes, and this insertion could possibly explain this phenomenon. However, gene deletion experiments would be necessary to test this hypothesis.

8.16.3. Inferences from Sequence Analysis of Toxin and Antibiotic Resistance Genes

Whole genome sequence alignments and alignments of genes from the PaLoc revealed that *tcdB* and *tcdA* were conserved in all genomes, and three ribotypes (023, 027 and 078) also contained genes for CDT (Table 8.16). This is consistent with findings by Stubbs *et al*. who also show that these ribotypes have CDT, while ribotypes 013, 014 and 015 do not[207]. Stubbs *et al*. hypothesise that only ribotypes showing gross variations in *tcdA* and *tcdB* have the presence of CDT[207]. Goldenberg *et al*., was not able to link *C. difficile* isolates with the *tcdC* truncation genotype to any mortality outcome; however, they did find that there was an association between strains with CDT and an increased host inflammatory response[82]. Genetic evidence from the current study suggests that isolates with CDT also have *tcdC* truncation genotypes, and there is literary evidence to show that *tcdC* truncations lead to prolonged toxin production[22, 53]. This data would suggest that the ribotypes with the *tcdC* truncation genotypes (027, 023 and 078), would be producing toxin at earlier stages of growth, and would also have CDT. This may also mean that these isolates might be correlated with increased levels of inflammation[82]. Genetic analysis of *tcdC* from the 027, 023 and 078 ribotypes revealed they do possess CDT. However, results from phenotypic investigation of toxin production in this study (Chapter Nine, Figures 9.2 and 9.3) suggests

that not all isolates with *tcdC* truncation genotypes are associated with early/increased toxin production. This is something that will be discussed further in Chapter Nine.

There were putative NS SNPs in, and conservation of, genes encoding putative vancomycin and teicoplanin resistance mechanisms among all isolates (Table 8.19). SNPs were found in, *vanW*, a gene of unknown function, that has been identified in vancomycin resistant *Enterococcus faecalis*[164], a σ^{-54} dependent transcriptional regulator downstream of *vanW*, *vanS*, a sensor histidine kinase, *vanG*; a putative vancomcyin/teicoplanin resistance protein, and *vanZ*, a putative teicoplanin resistance gene.

Table 8.19 - Conserved putative vancomycin/teicoplanin resistance genes, and presence/absence of putative SNPs within them.

Clinical Isolate	Ribotype	NS SNPs in Putative vancomycin/teicolplanin resistance genes					Conservation of Putative vancomycin/teicolplanin resistance genes		
		vanW	σ^{54} Transcription factor	*vanS*	*vanG*	*vanZ*	*vanG*-like Operon	*vanW*	*vanZ*
FA07001994	027	-	-	-	-	-	+	x2	+
FA07003754	027	-	-	-	-	-	+	x2	+
FA07007522	078	-	-	-	-	-	-	x2	+
FA07008490	078	-	-	-	-	-	-	x2	+
FA08006656	078	-	-	-	-	-	-	x2	+
FA09004991	078	-	-	-	-	-	-	x2	+
FA07003485	023	+	+	-	-	-	-	x3	+
FA07004080	023	+	+	-	-	-	-	x3	+
FA08005864	023	+	+	-	-	-	-	x2	+
FA08006661	023	+	+	-	-	-	-	x3	+
FA07011498	015	-	-	-	-	-	+	x2	-
FA09007583	015	-	-	-	-	+	+	x2	+
FA07004464	014	+	-	+	-	-	+	x2	-
FA07007469	014	+	-	+	-	-	+	x2	-
FA08012693	013	-	-	-	+	-	+	x2	+
FA08006290	013	-	-	-	+	-	+	x2	+

+ presence of SNPs/conservation of genes, - absence of SNPs or absence of genes

A *vanG*-like operon was also conserved in some isolates (Table 8.18) and has also been identified in *C. difficile* 630[197]. The operon in clinical isolates consists of *vanR*/*vanS*, which encodes a regulator-sensor kinase, *vanG*, a putative vancomcyin/teicoplanin resistance protein and *vanTG*, a serine/alanine racemase. The organisation of the genes in the *vanG*-like

operon in the clinical isolates shows similarity to the *vanG* operon found in *Enterococcus faecalis*[57]. In *Enterococcus faecalis* this operon confers moderate *vanG*-type resistance to vancomycin, but not teicoplanin, and contains two additional genes (*vanX* and *vanX/Y*) relative to *C. difficile* 630 and the clinical isolates. These two key enzymes are needed for the prevention of interaction of glycopeptides with their normal targets[57] and the deletion of *vanX* and *vanY* has been shown to decrease vancomycin resistance in *Enterococcus faecium*[138]. As these genes are missing from the clinical isolates, it could explain why isolates harbouring the operon still exhibit vancomycin sensitivity, if the operon was expressed.

There is speculation over whether the other putative vancomycin/teicoplanin resistance genes are expressed in *C. difficile* 630[197]. During the present study, there was no significant difference in vancomycin susceptibility for any isolates (Chapter Nine, Table 9.3); therefore in these isolates these genes may also not be expressed. However, it is not to say that conservation of these genes, and the SNPs within them, will play no role in resistance of *C. difficile* isolates to vancomycin, or other glycopeptides in the future. Or that the use of vancomycin may be facilitating selective pressure on the genomes of all isolates.

All clinical isolates in this study cohort were resistant to ciprofloxacin, (Chapter Nine, Table 9.3). Four clinical isolates, from ribotypes 027 (FA07001994 & FA07003754), and 014, (FA07007466 & FA07004464), harboured the Thr82Ile amino acid substitution in GyrA, relative to *C. difficile* 630. This mutation has been proposed to confer resistance to the newer fluoroquinolones; gatifloxacin, moxifloxacin and levofloxacin[202]. While no other isolate harboured the Thr82Ile mutation, all isolates (except ribotype 013 isolates) harboured novel amino acid substitutions in either GyrA and/or GyrB (Table 8.17). The presence of novel amino acid substitutions in isolates from ribotypes 023, 078 and 015, might suggest the possibility that they have increased resistance to the newer fluoroquinolones, gatifloxacin, moxifloxacin and levofloxacin, whilst isolates from ribotype 013 do not. Antibiotic

susceptibility testing of all isolates, using these three fluoroquinolones would need to be investigated to deduce the contribution of known and novel amino acid substitutions to resistance profiles.

No isolate in this study cohort was resistant to metronidazole (Chapter Nine, Table 9.3). The presence of *nim* genes, which confer metronidazole resistance in some *Bacteroides fragilis* isolates[147], were not found in any isolate; although a putative 5-nitroimidazole-reducatse was conserved in all isolates (Table 8.17). There is increasing concern over the emergence of *C. difficile* isolates with reduced susceptibility to metronidazole[18] and even though conventional mechanisms of resistance, i.e. through *nim* genes, were not present in isolates in this study, others have deduced that some bacteria of the intestinal tract, may employ other mechanisms, such as, metabolic inactivation of metronidazole, to resist the bactericidal activities of these compounds[185]. These alternative mechanisms may, one day, also be employed by *C. difficile* to tolerate metronidazole.

8.16.4 Conclusion

Whole genome sequence analysis on the panel of clinical isolates has revealed many interesting results. The different abundance of SNPs within some isolates of the same ribotype over the study period may be suggestive of their recent diversification from a common ancestor. The genes in which SNPs occur could indicate what role the host/clinical environment is having on genome diversity. The abundance of SNPs makes it hard to try and decipher whether or not they could contribute to the outcome of an infection, but they may be helpful to understand what type of genes are mutating, and help focus researchers on their quest to understand the evolution of this bacterial species and identify new therapeutic targets.

The genome *C. difficile* is highly mobile[197] and this is clearly reflected by the insertion of mobile elements and phage regions observed between different ribotypes. The stable integration and conservation of mobile elements at certain sites within a ribotype, reflected in this data, suggests that they may be advantageous for the bacteria in their niche environment. This is an idea which Brouwer *et al*. also share, as they believe that many of these putative mobile elements have been conserved due to the accessory modules that they possess[37]. They suggest that the majority of accessory genes carried by these putative mobile elements encode ABC transporters and efflux pumps; which may have a role in antibiotic resistance or function in resistance to antimicrobial peptides, bile salt hydrolases and alternative σ factors. It may thus prove important to study the pan-genome of *C. difficile* isolates in order to determine the absolute role of these accessory genes and how they may impact on host-pathogen interaction. Looking at the locality of mobile element insertion may also facilitate our understanding of the impact they are having on gene regulation. If mobile elements were to insert between ORFs they may result in transcriptional effects in the locality of the insertion site which could fundamentally alter the phenotype of the bacteria[37].

However, as summarised by Knetsch *et al*.[128] the complex insertion sites of phage regions and transposable elements serves as a reminder that finding stably integrated strain specific markers would be very challenging, given the highly mobile nature of genomes of *C. difficile*. Perhaps a shift towards understanding the role of these insertions at the transcriptional level could give more insight into the complex host-pathogen interplay that governs CDI outcomes such as mortality and recurrence.

8.17 Summary

Emphasised in Chapter Seven was the need to couple inferences from whole genome analysis with *in vitro* phenotypic data, in order to better understand the complex host-pathogen relationship.

The genome of *C. difficile* is evolving and adapting to host and clinical environments. Results from this chapter provides evidence that suggests that there are SNP differences in virulence genes and large scale genomic changes that may give rise to phenotypic differences that can be observed *in vitro*. Consequently, these differences may contribute to the emergence of certain ribotypes within the clinical environment and may also explain their association to certain outcomes of infection.

Phenotypic characterisation of the isolates outlined in Table 8.1 and Table 8.2 was conducted to deduce if any link could be made between the genetic changes observed in virulence genes and isolate *in vitro* phenotypes, which could then be related back to association with infection outcomes such as recurrent CDI or mortality. This data will be presented in Chapter Nine.

Chapter Nine. Phenotypic Characterisation of Outcome Isolates

9.1 Introduction and Aim

Chapter Seven highlighted some of the virulence mechanisms that may contribute to the prevalence of *C. difficile* in a clinical setting, and that may also play a role in determining the outcome of infection. The previous chapter presented genomic analysis of isolates selected in this study and reported gross and subtle genetic differences which may translate to different *in vitro* phenotypes, and alterations in toxin production, antibiotic resistance, sporulation and motility. This chapter will present data resulting from phenotypic characterisation of these clinical isolates (Table 9.1) along with reference isolates (Table 9.2) to assess if the genetic differences found in Chapter Eight are translated to observable differences *in vitro*. Isolates were subjected to antibiotic MIC assays via the E-test method, spore germination studies, motility assays and semi-quantitative toxin immune-dot blot assays. Finally, I will discuss how these phenotypic findings, combined with the genetic findings may be linked to the infection outcomes.

Table 9.1- Clinical isolates selected for phenotypic analysis.

Clinical Isolates	Ribotype	Severity	Cause of Death*	Number of Episodes	Patient
FA08012693	013	Mild	N/A	One episode	1
FA08006290	013	Severe	CDI was part of or mentioned in death	One episode	2
FA07007469	014	Life Threatening	N/A	One episode	3
FA07004464	014	Mild	N/A	One episode	4
FA09007583	015	Severe	CDI was part of or mentioned in death	One episode	5
FA07011498	015	Severe	N/A	One episode	6
FA07003485	023	Moderate	N/A	Multiple Episodes	7
FA07004080	023	Mild	N/A	Multiple Episodes	7
FA08006661	023	Severe	CDI was not the cause of death	One episode	8
FA08005864	023	Severe	N/A	One episode	9
FA07007522	078	Severe	N/A	Multiple Episodes	10
FA07008490	078	Severe	N/A	Multiple Episodes	10
FA09004991	078	Life Threatening	N/A	One episode	11
FA08006656	078	Severe	CDI was not the cause of death	One episode	12
FA07001994	027	Mild	N/A	Multiple Episodes	13
FA07003754	027	Severe	Patient survived	Multiple Episodes	13

Table 9.2- Reference isolates selected as controls.

Reference Isolates	Ribotype	Mutation Phenotype / isolation information	NCBI RefSeq/GenBank ID
R20291	027	*Clostridium difficile* R20291 was isolated in Stoke Mandeville Hospital, UK in 2006.	NC_013316.1
630 Δ *SpoOA*::*erm*	012	Insertional mutation in the *Spo0A* gene using an erythromycin cassette. This isolate is unable to form spores and is erythromycin resistant	N/A
M120	078	*Clostridium difficile* M120 is a human strain isolated in the United Kingdom in 2007	FN_665653
CD305	023	Isolated from a 74 year old male with severe symptoms in 2008	N/A
TL174	015	Isolated in 2009	N/A
TL176	014	Isolated in 2009	N/A
CF5	017	Isolated in 1995. This is a natural *TcdA*$^-$*TcdB*$^+$ variant	N/A
07-584	Unknown	This is a toxin negative isolate	N/A

9.2 Results

9.2.1 *C. difficile* growth

Growth of all isolates in BHIS broth was measured by changes in optical density (OD_{590}) and cfu counts over 48 hrs to eliminate the possibility that differences in growth were the cause of different phenotypes investigated in this study (Appendix Three, Figures A1.1-A1.3). More specifically, Two-way ANOVA of mean cfu counts between all isolates was determined not to be significantly different ($P<0.05$) at the time points used for harvest of culture supernatant and subsequent toxin detection as described in section 9.4 below. Therefore, it was concluded that levels of TcdA and TcdB detection in BHI were the result of true isolate specific differences, and not differences in cfu counts. Due to time constraints in this PhD it was not possible to determine growth of all isolates in TY broth over 48 hrs. Two-way ANOVA of mean OD_{590} for all isolates, over the 120 hr time period revealed that there was a small overall significant difference in clinical isolate growth in TY broth ($P=0.044$). Bonferroni post-hoc comparison tests revealed this was only at certain time points between certain isolates.

9.2.2 Characterisation of Isolate Swimming Motility

Flagellar driven motility has been attributed to the virulence of many bacterial species including *Campylobacter jejuni*, *Vibrio cholerae*, and *Helicobacter pylori*[15]. Chapter Eight provided evidence that the F3 flagellar operon of the ribotype 078 isolates, relative to *C. difficile* 630 and R20291 was absent. Motility of all isolates was investigated to assess any general differences in isolate motility, and to determine if the absence of the F3 flagellar operon in all of the 078 isolates affected motility. Motility assays were carried out according to a method modified from Valiente *et al.*[218]. *C. difficile* isolates were cultured in 0.1% BHI agar for 48 hrs in anaerobic conditions. Fully sequenced isolates M120 (non-motile) and R20291 (highly motile) were used as controls, along with reference isolates for 023, 014 and

015 ribotypes. Motility was measured as the distance from inoculation site to the maximum stalactite position (Appendix Three, Figures A3.4–A3.6). Graphical representation of mean motility distances are presented in Figure 9.1, and are the result of a minimum of three independent experiments.

One-Way ANOVA analysis of mean motility revealed that there was an overall significant difference between some isolates ($P<0.001$; $df=20$; $R=0.91$; $F= 24.71$). Tukey's post-hoc multiple comparison tests revealed that the difference in mean motility was most significant between all 078 isolates (Figure 9.1) and all other isolates ($P<0.001$). There were no significant differences between the mean motility of the reference isolates (M120, CD305, TL174 & TL176) with their respective clinical isolates of the same ribotype (Figure 9.1). The exception to this was the detection of a significant difference in mean motility between R20291 and FA07001994 (both 027 ribotypes) ($P<0.001$). FA07001994 is the primary isolate in a pair of sequential isolates from a patient with recurrent CDI and it was more motile than R20291. There were no significant differences between the mean motility of clinical isolates of the same ribotype.

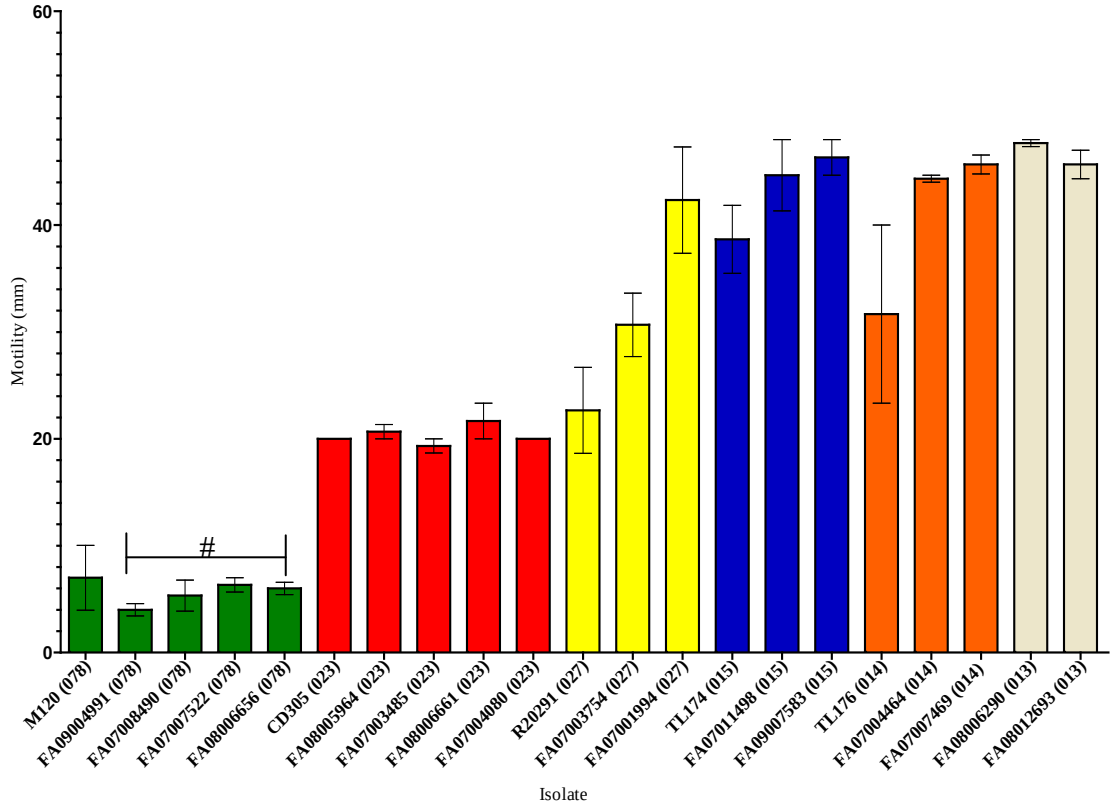


Figure 9.1- Graphical representation of mean *C. difficile* isolate motility. Motility was scored as <10 mm (non-motile), 10-20 mm (motile) and >20 mm (highly motile). Green bars= 078 ribotypes, Red bars= 023 ribotypes, Yellow bars= 027 ribotypes, Blue bars= 015 ribotypes, Orange bars= 014 Ribotypes, Beige bars= 013 ribotypes.

All clinical and reference 013, 014 and 015 ribotypes and clinical 027 isolates were more motile than the highly motile control isolate (R20291) and all 023 isolates were as motile as R20291 (Figure 9.1).

All 078 isolates were non-motile, possibly due to the loss of the F3 flagellar operon (Chapter Eight, Figure 8.3). The mean motility of the control 027 isolate (R20291) differed significantly from the 015 reference isolate (TL174) $P<0.05$, and the 078 clinical isolates ($P<0.05$). The mean motility of the reference 015 isolate (TL174) differed significantly from the 078 clinical isolates and the 023 clinical isolates ($P<0.05$). The mean motility of the reference 014 isolate (TL176) differed significantly from the clinical 078 isolates. The difference in motility of some isolates as compared to R20291 may mean that they have

different host colonisation capabilities, or given that flagellar regulation couples motility and possible toxin expression, some of the more motile isolates could be producing more toxin at earlier stages of growth. Thus, toxin production of *C. difficile* isolates was investigated.

9.2.3 Characterisation of *in vitro* TcdA and TcdB Production

Differences in toxin production by *C. difficile* can be partially explained by environmental growth conditions such as glucose concentration[153], mutations in TcdC[58] and possibly changes in genes of the flagellar operon[15]. Results from Chapter Eight highlighted genetic changes in *tcdC*, and other regions of the PaLoc, that could lead to *in vitro* differences in toxin production for some ribotypes. The panel of sixteen clinical isolates (Table 9.1) and seven reference isolates (Table 9.2), were assessed for the *in vitro* production of TcdA and TcdB in BHI and TY broth, over a period of 120 hrs by immuno dot-blots as outlined in the Materials and Methods chapter (Chapter Ten, Section 10.12), to assess if there were differences in toxin production between ribotypes. Culture supernatants from all isolates, at all time points were blotted together. A toxin negative control isolate, 07-584, (kindly provided by Beata Walter from the University of Ljubljana, Ljubljana, Slovenia) and a natural $TcdA^{-}TcdB^{+}$ isolate (CF5, Ribotype 017) were used to determine the specificity of monoclonal antibodies (mAbs) for TcdA and TcdB. mAbs specific for TcdA did not cross react with supernatants from the negative control or the CF5 isolate, demonstrating the specificity of the TcdA mAb for TcdA. Detection of TcdA and TcdB was performed on a minimum of two independent samples from each media. The enhanced chemiluminescent (ECL) detection system (GE Healthcare) was used to detect toxin in BHI broth and infra red emission spectra, on the Odyssey Clx system (LI-COR, Germany) was used to detect toxin in TY broth. Figures 9.2a, 9.2b and 9.3 depict representative blots.

9.2.3. i Detection of TcdA

For most ribotypes, there was uniform detection of TcdA at later time points (48 and 120 hrs) in both media tested (Figures 9.2 and 9.3), indicating that glucose levels of 0.2 % in BHI may not be having a significant repressing effect on toxin synthesis at this stage of growth. The exception to this uniform level of detection of TcdA was seen in the 023 ribotypes (Figure 9.3c) where TcdA production appeared to be less at 8, 24 and 48 hrs.

The most evident differences in production of TcdA between different isolates was observed at late log stages of growth (8 hrs) in both media tested (Figures 9.2 and 9.3). Generally, detection at 8 hrs was less pronounced in BHI than TY broth (Table 9.3), and this may be indicative of a small repressing effect of glucose, and/or differences in the sensitivity of detection method (chemiluminescence versus infrared) used. Equally, these differences could be due to differential transcriptional regulation of TcdA by each *C. difficile* isolate. The level of detection of TcdA for each isolate at 8 hrs is summarised in Table 9.3. Five isolates showed strong detection of TcdA at 8hrs, and the remaining eleven isolates showed low or no detection of TcdA at this time point.

There is also some evidence to suggest that the non-motile ribotype 078 isolates start producing TcdA later than the other (more motile) isolates (Table 9.3 and Figure 9.3b).

Table 9.3- Level of detection of TcdA in all clinical isolates after 8 hrs of growth in TY and BHI broth.

Clinical Isolates	Ribotype	Detection of TcdA at 8hrs in TY Broth	Detection of TcdA at 8hrs in BHI Broth
FA08012693	013	++	+
FA08006290	013	+	n/a
FA07007469	014	+	n/a
FA07004464	014	+	n/a
FA09007583	015	+	n/a
FA07011498	015	++	+
FA07003485	023	+	n/a
FA07004080	023	n/a	n/a
FA08006661	023	++	n/a
FA08005864	023	n/a	n/a
FA07007522	078	n/a	n/a
FA07008490	078	n/a	n/a
FA09004991	078	n/a	n/a
FA08006656	078	+	n/a
FA07001994	027	++	+
FA07003754	027	++	++

n/a= No Detection, += Low/less pronounced detection, ++Strong detection

Results from these experiments indicate that there is also intra-ribotype variation in levels of TcdA produced at early growth stages (8hrs). This is exemplified between isolates from the 015, 013 ribotypes (9.2a & 9.2c) and 023 ribotypes (in TY broth only) (Figure 9.3c). The 015 (FA09007583, Figure 9.2a) and 013 (FA08006290, Figure 9.2c) isolates which were associated with death had less detection of TcdA at early time points (8 hrs) than isolates which came from patients who survived. This phenomenon is inversed for ribotype 023 and 078 isolates, where the isolates found in patients who died (023; FA08006661, Figure 9.3c & 078; FA08006656, Figure 9.3b) (but not from CDI related causes) showed early detection of TcdA, and isolates found in patients who survived showed detection of TcdA later.

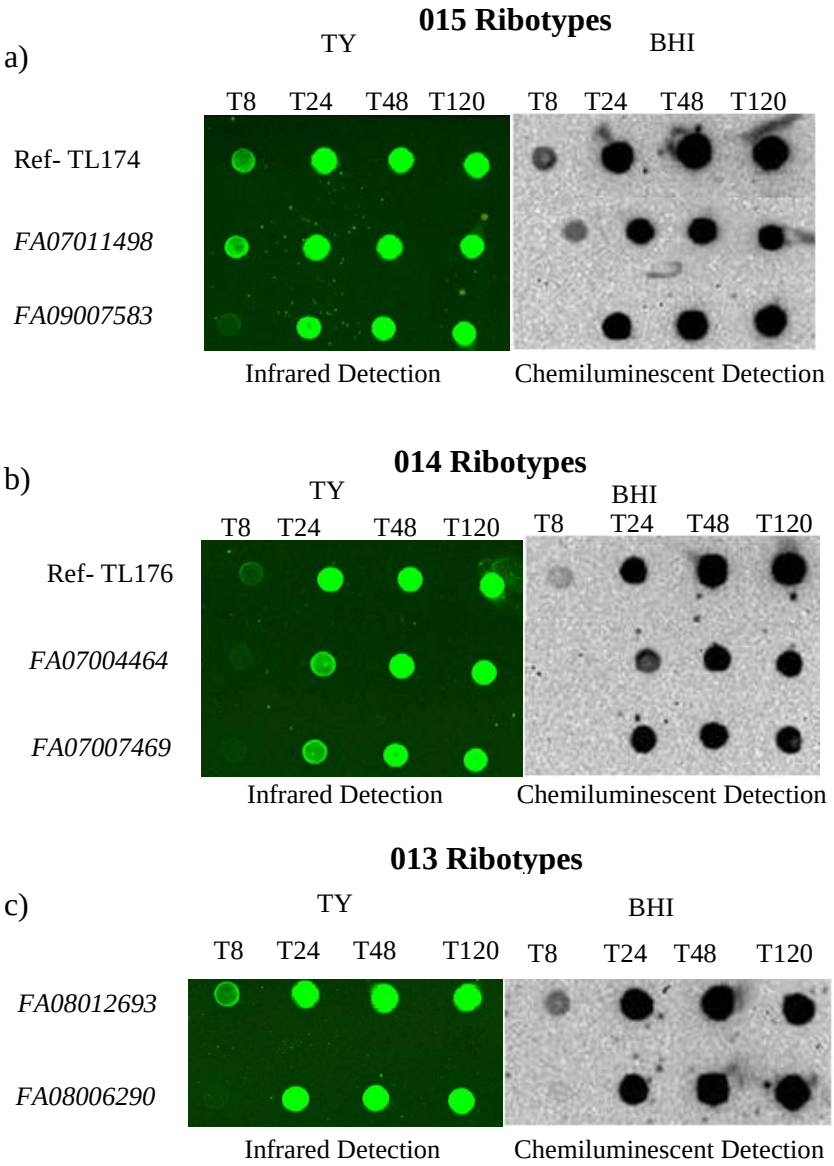


Figure 9.2- Immuno-dot blot for *in vitro* TcdA production *C. difficile* isolates. Culture supernatant from the toxin negative and CF5 strain was blotted with all isolates. No toxin was detected with the mAb specific for TcdA. Ref= Reference isolate. T= Growth time in hrs.

At 8 hrs there was a difference in infrared detection of TcdA between two 023 isolates from one patient with recurrent CDI. The primary isolate from this infection (FA07003485, Figure 9.3c) appeared to be producing TcdA at an earlier time point (8hrs) than the secondary isolate (FA07004080, Figure 9.3c).

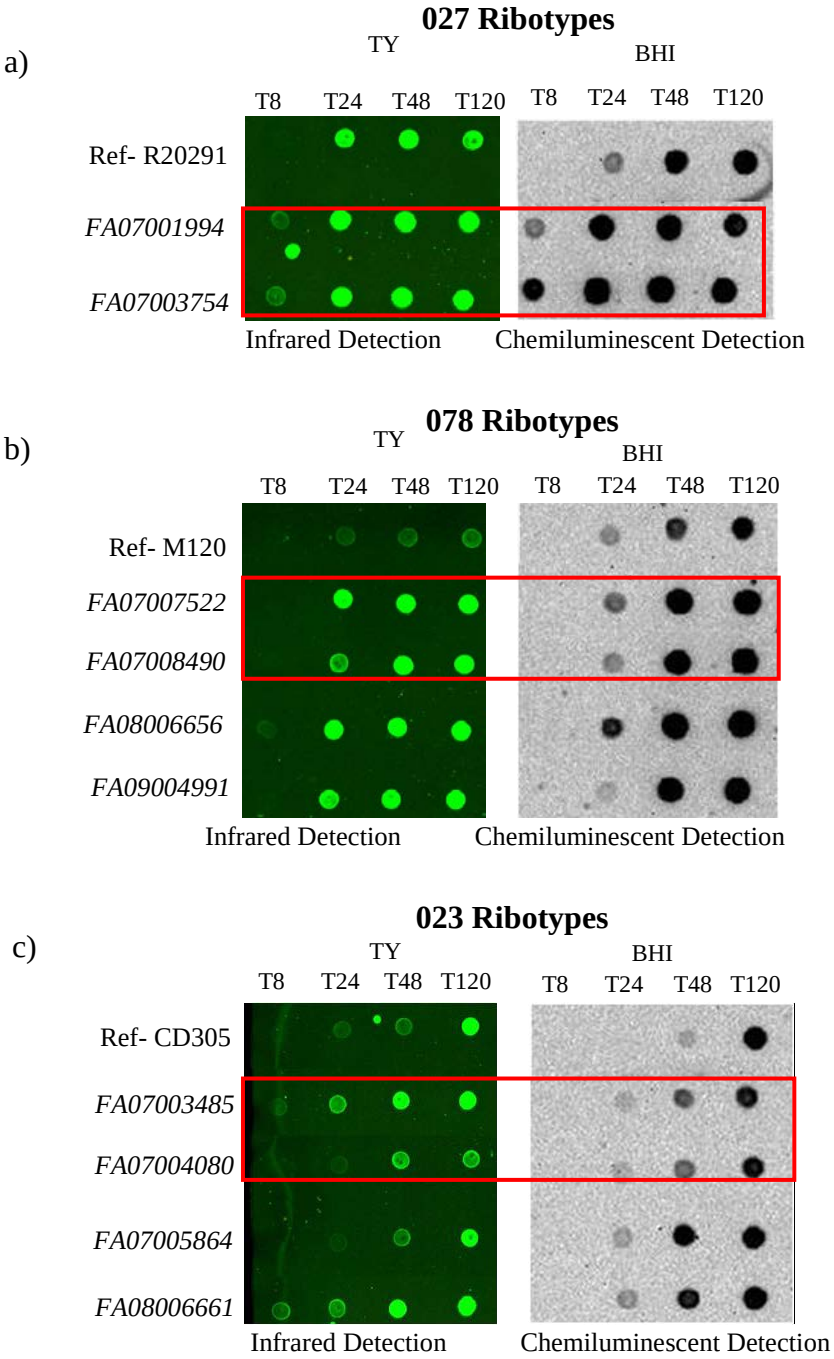


Figure 9.3- Immuno-dot blot for *in vitro* TcdA production for *C. difficile* isolates. Red boxes represent toxin dots from recurrent infection (same patient). Culture supernatant from the toxin negative and CF5 isolates was blotted with all isolates. No toxin was detected with the mAb specific for TcdA. Ref= Reference isolate. T= Growth time in hrs.

9.2.3. ii Detection of TcdB

Regardless of the media used in this experiment, detection of TcdB was only observed in isolates of the 015 and 013 ribotype from 24 hrs onwards in TY broth, and 48 hrs onwards in BHI broth (Figures 9.4a & 9.4b). Detection of TcdB (above background) was only seen in the 015 reference isolate (TL174) and isolates from each ribotype which came from patients who survived the CDI (FA08012693 & FA07011498). In TY broth, infrared intensity was consistent for the three time points, 24, 48 and 120 hrs, whereas in BHI broth, chemiluminescent intensity peaked at 48 hrs then decreased thereafter. A second monoclonal antibody was used to try and detect TcdB from *C. difficile* R20291 and CF5 however; this also failed to detect any TcdB, in either BHI of TY broth.

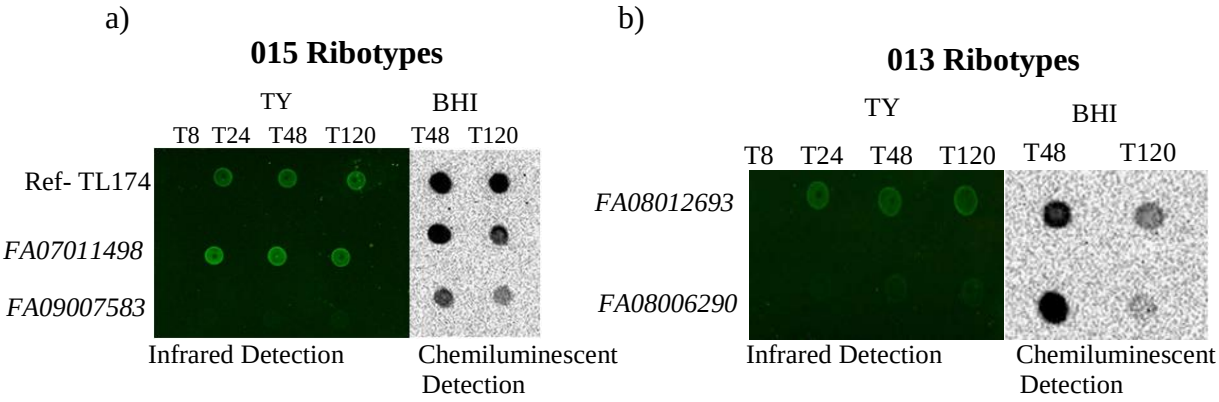


Figure 9.4- Immuno-dot blot for TcdB for *C. difficile* isolates. Culture supernatant from the toxin negative strain was blotted with all isolates and no toxin was detected. Ref= Reference isolate. T= Growth time in hrs

9.2.4 Characterisation of Isolate Antibiotic Resistance Profiles

Within the clinical environment *C. difficile* infection is exacerbated by inappropriate use of antibiotics such as clindamycin and the fluoroquinolones, such as ciprofloxacin, levofloxacin and moxifloxacin[42]. Current therapies for treatment of CDI at the time of this study were metronidazole and vancomycin. There is current concern that isolates are emerging which show reduced susceptibility to metronidazole[7] and there is also concern that overuse of vancomycin is promoting increased incidences of VRE[10]. In Chapter Eight, genetic analysis

of the isolates outlined in Table 9.1 revealed that most isolates had changes in *gyrA* and/or *gyrB* which could possibly give rise to fluoroquinolone resistance. Genetic analysis also revealed the presence of putative vancomycin resistance genes. Thus, isolate resistance to ciprofloxacin, metronidazole and vancomycin was investigated.

Isolates were subject to MIC analysis by the E-test method (Biomerieux) as described in the Materials and Methods chapter (Chapter Ten, Section 10.9), in order to evaluate susceptibilities to clinically relevant antibiotics. Results are summarised in Table 9.4 and Figures 9.5a, 9.5b and 9.5c below, and are based upon three independent experiments. Antibiotic susceptible positive control strains (as recommended by the manufactures specifications) *Staphylococcus aureus* ATCC 29213 and *Bacteroides fragilis* ATCC 25285 were tested with all *C. difficile* isolates.

Mean MICs of all isolates were compared by One-Way ANOVA followed by Bonferroni post-hoc comparison tests. There was an overall significant difference in mean antibiotic MICs between some isolates, for ciprofloxacin ($P<0.001$) and metronidazole ($P<0.001$) but not vancomycin. Post-hoc comparison tests revealed that the significant difference in mean MICs for ciprofloxacin resistance were between isolate FA08006656 (Ribotype 078, Table 9.4) and most other clinical isolates, with the exception of two 078 isolates (FA07007522 & FA07008490) that were from recurrent infection. The MIC of isolate FA08006656 was 8 µg/ml, which was lower than those isolates of the same ribotype and all other ribotypes.

Table 9.4- Mean antibiotic MICs of all clinical cohort isolates.

		Mean Antibiotic MIC (µg/ml)		
Isolate ID	**Ribotype**	**MET MIC**	**VANC MIC**	**CIP MIC**
FA07003485	023	0.047	1	≥ 32
FA07004080	023	0.028	1	≥ 32
FA08005864	023	0.028	0.91	≥ 32
FA08006661	023	0.039	1.16	≥ 32
FA07007522	078	0.042	1.3	24
FA07008490	078	0.052	1.16	10
FA08006656*	078	0.032	1.16	8*
FA09004991	078	0.062	0.91	≥ 32
FA07001994*	027	0.146*	1.33	≥ 32
FA07003754	027	0.066	1.33	≥ 32
FA07011498	015	0.064	0.91	≥ 32
FA09007583	015	0.068	1.33	≥ 32
FA07004464	014	0.074	1.33	≥ 32
FA07007469	014	0.058	0.91	≥ 32
FA08012693	013	0.068	1.67	≥ 32
FA08006290	013	0.049	1.67	≥ 32

MET- Metronidazole, VANC= Vancomycin, CIP= Ciprofloxacin. *= Isolates with significantly different MICs for some antibiotics.

Post-hoc comparison tests also revealed that the significant differences in mean MICs for metronidazole were between isolate FA07001994 (027 ribotype; mean 0.146 µg/ml) and all clinical and reference 078 ribotypes, except one (FA09004991); all clinical and reference 023 ribotypes, and isolate FA08006290 (Ribotype 013). Isolate FA07001994 was from a patient with recurrent CDI and was isolated from the primary infection.

Antibiotic levels at which *C. difficile* is determined to be clinically resistant (breakpoint resistance) were identified, and are listed in Table 9.5. No isolate showed resistance to either metronidazole (≥4 µg/ml), or vancomycin (≥2 µg/ml) as indicated by these breakpoint values. All isolates showed resistance (≥6 µg/ml) to ciprofloxacin.

Table 9.5- Range of MIC values from study isolates with the resistance breakpoints used. (Adapted from[34, 62, 166] and personal communication with Dr Val Hall).

Antibiotic	MIC range (µg/ml)	MIC_{50} (µg/ml)	MIC_{90} (µg/ml)	Breakpoint (µg/ml)	Resistance (%)
Vancomycin	0.91–1.67	1.167	1.33	≥2	0
Metronidazole	0.028–0.146	0.055	0.067	≥4	0
Ciprofloxacin	8–≥32	≥32	≥32	≥6	100%

One-way ANOVA analysis of mean MICs revealed that there were no significant differences between isolates that had come from primary infections and the secondary infection for cases of recurrent CDI. Nevertheless, it can be seen that the mean MIC for ciprofloxacin for 078 isolates FA07007522 (24 µg/ml) from a primary infection was higher than that from the isolate recovered from the secondary infection (FA07008490; 10 µg/ml) (Figure 9.5a, Table 9.4). This trend was also the same for mean MICs to vancomycin (FA07007522; 1.3 µg/ml vs. FA07008490; 1.17 µg/ml) (Figure 9.5c, Table 9.4) but reversed for metronidazole (FA07007522; 0.043 µg/ml vs. FA07008490; 0.053 µg/ml) (Figure 9.5b, Table 9.4).

For isolates of the 023 (FA07003485 & FA07004080) and 027 (FA07001994 & FA07003754) ribotype, there were no differences in mean MICs for ciprofloxacin between the isolates that came from primary infections compared to isolates that came from secondary infections (Figure 9.5a). However, there was a decrease in MIC for metronidazole for isolates found in the secondary infection compared to those found in the primary infection (023; FA07003485 and FA07004080; 0.047 µg/ml vs. 0.029 µg/ml) (027; FA07001994 and FA07003754; 0.146 µg/ml vs. 0.067 µg/ml) (Figure 9.5c).

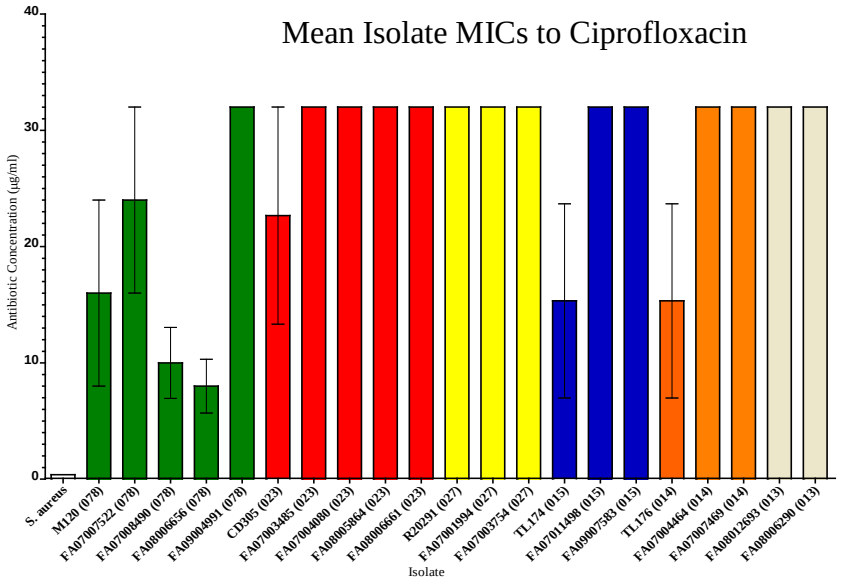


Figure 9.5a- Resistance of *C. difficile* isolates to ciprofloxacin. Values of 32 on the scale represent a value of > 32 µg/ml on the E-test strip. Error bars= SEM. MIC for control strain (*S. aureus*) for ciprofloxacin falls between 0.125–0.5 µg/ml according to manufacturer specification. Green bars= 078 Ribotypes, Red bars= 023 Ribotypes, Yellow bars= 027 ribotypes, Blue bars= 015 ribotypes, Orange bars= 014 Ribotypes, Beige bars= 013 ribotypes.

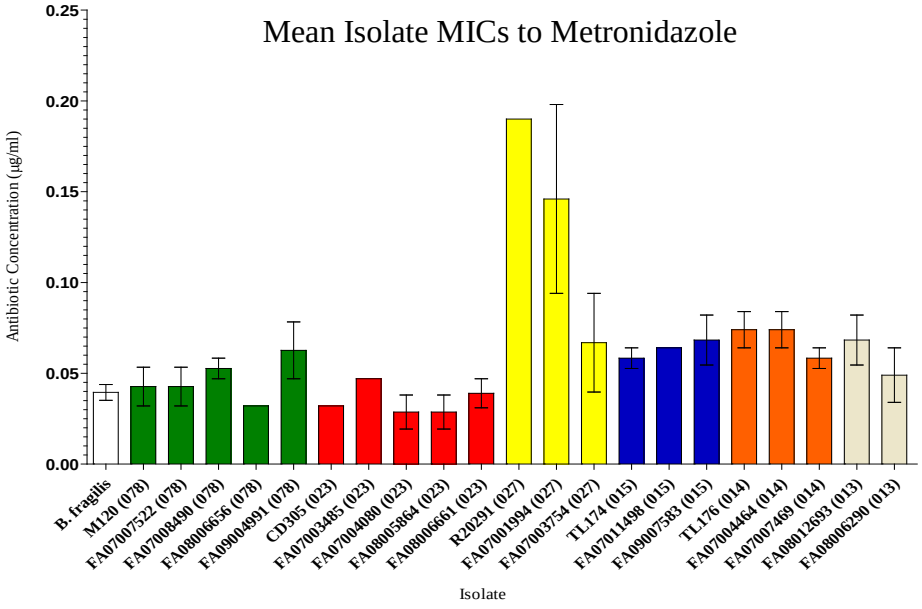


Figure 9.5b- Resistance of *C. difficile* isolates to metronidazole. Error bars= SEM. MIC for control strain (*B. fragilis*) for metronidazole falls between 0.25–1 µg/ml according to manufacturer specification. Green bars= 078 Ribotypes, Red bars= 023 Ribotypes, Yellow bars= 027 ribotypes, Blue bars= 015 ribotypes, Orange bars= 014 Ribotypes, Beige bars= 013 ribotypes.

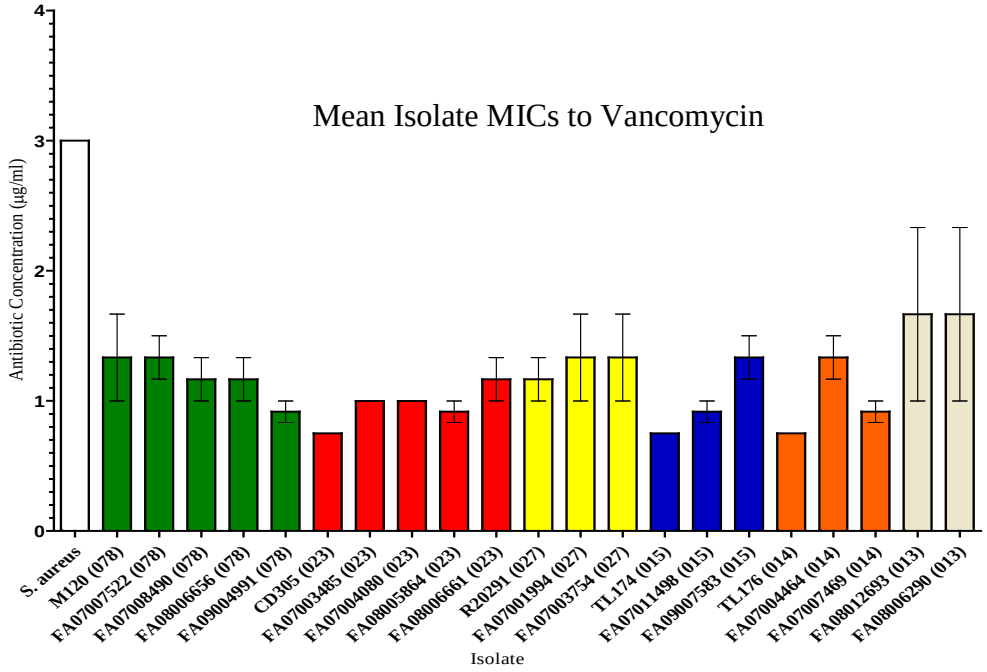


Figure 9.5c- Resistance of *C. difficile* isolates to vancomycin. Error bars= SEM. MIC for control strain (*S. aureus*) for vancomycin falls between 0.5–2 µg/ml, according to manufacturers specification. Green bars= 078 Ribotypes, Red bars= 023 Ribotypes, Yellow bars= 027 ribotypes, Blue bars= 015 ribotypes, Orange bars= 014 Ribotypes, Beige bars= 013 ribotypes.

9.2.5 Characterisation of Isolate Spore Germination Profiles

Sporulation of *C. difficile* is vital to its transmission in the clinical setting. However there are conflicting reports on the rate of sporulation between epidemic and non-epidemic strains[39] which could explain the prevalence of certain ribotypes within a clinical setting. There have also been conflicting reports on the link between toxin production and sporulation and whether they are correlated in any way[119, 190].

Whole genome analysis results from Chapter Eight (Table 8.15) also revealed that the 013 ribotype showed a number of non-synonymous (NS) SNPs in putative sporulation genes which may affect sporulation. Isolates were subjected to spore germination assays in order to assess if there were any differences in sporulation rates of that may possibly contribute to their establishment in the clinical setting. Sporulation was investigated according to a

protocol modified from Burns *et al*.[40]. Briefly, 5 ml BHIS broth was inoculated to an OD_{590} 0.05 and grown to OD_{590} 0.1-0.2 (~1-5x10^7 cfu). Broth from these cultures were standardised to an OD_{590} of 0.1 and 100 µl (~1x10^5 cfu) of the standardised culture was used to inoculate 10 ml of fresh BHIS broth (0 hrs). Standardisation of the starter culture ensured that the resulting spore formation was not the result of different cfu densities at 0 hrs. Cultures were grown for five days in an anaerobic environment. A 'hypervirulent' BI/NAPI/027 ribotype (R20291), and a *C. difficile* 630 mutant isolate, which is unable to form spores (630Δ*spo0A::erm*, Table 10.4) were used as controls. Two-way ANOVA followed by Bonferroni post-hoc comparison tests, revealed there were no significant differences between mean total cfu counts of any isolates at 0 hrs (P=0.6882), therefore any differences in sporulation were isolate specific, and not the result of differing cfu density at the start of the experiment. Experiments were repeated until there were no spores present at 0 hrs (data not shown), thus ensuring that all spore formation started from cells which were vegetative. Results presented in Figures 9.6a, 9.6b and 9.6c represent mean heat resistant cfu counts at 24, 48 and 120 hrs, from a minimum of two independent experiments.

Results from two-way ANOVA, followed by Bonferroni post-hoc comparison tests failed to find any statistically significant differences between mean heat resistant cfu counts of different clinical isolates, at 24 and 48 hrs. However, there are clearly gross differences in the production of heat resistant cfus between some clinical isolates. The onset of sporulation at 24 hrs differs most between all 013 isolates, isolates FA07008490, FA08006656 (078 Ribotypes), and the FA07003754 isolate (027 Ribotype). These isolates did not produce heat resistant cfus at 24 hrs whereas all other isolates did (Figure 9.6a). As mentioned previously, whole genome analysis of the 013 ribotype isolates revealed NS SNPs in sporulation specific genes which could explain the late onset of sporulation.

The primary isolate (FA09001994) from the 027 recurrent pair formed heat resistant cfus at 24 hrs while the second isolate (FA07003754) did not (Figure 9.6a). This may be suggestive of differences in transcriptional regulation of sporulation genes of the secondary isolate, or that the secondary isolate is a genetically unique isolate and that the patient from which the isolates were taken had a heterogeneous population of 027 isolates. This phenomenon was also observed for isolates from a 078 ribotype pair, where the primary isolate (FA07007522) formed heat resistant cfus, but the secondary (FA07008490) did not. At 24 hrs the 078 isolates were forming less heat resistant cfus than the cohort average (Figure 9.6a, dashed line).

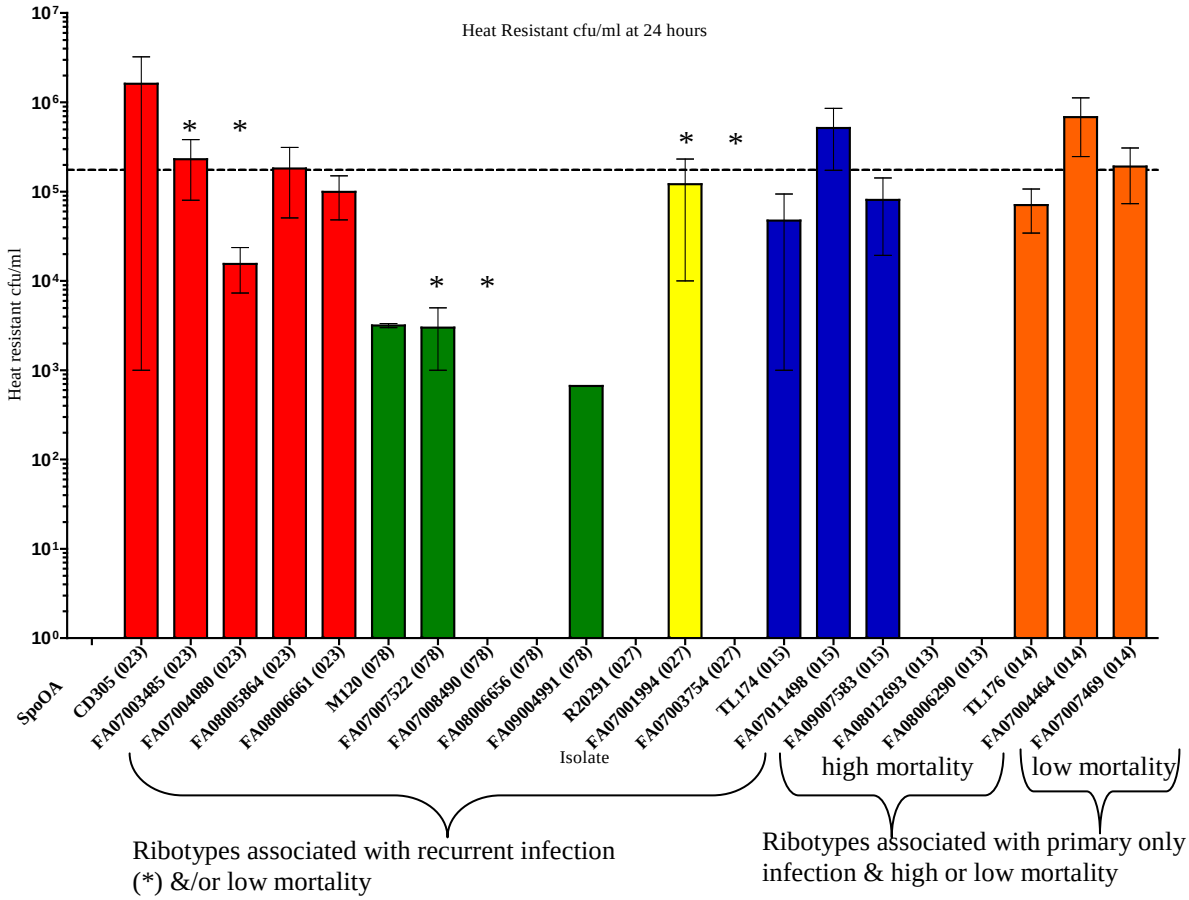


Figure 9.6a- Heat resistant cfu at 24 hrs for clinical and reference isolates under investigation. Error bars= SEM of a minimum of two independent experiments. Dashed line represents mean heat resistant cfu for this time point. Red bars= 023 ribotypes, green bars= 078 ribotypes, yellow bars= 027 ribotypes, blue bars= 015 ribotypes and orange bars= 014 ribotypes.

By 48 hrs heat resistant cfu counts became more uniform between isolates (Figure 9.6b) and all but one 078 isolate (FA08006656), was seen to form heat resistant cfus. The heat resistant cfu count for the 078 isolates at 48 hrs superseded, or were level with isolates from other ribotypes (015 and 014), indicating a rapid sporulation onset, with heat resistant cfu titres going from ~10^3 at 24 hrs to ~10^6 heat resistant cfus at 48 hrs for the 078 ribotype. This is also true for the 013 isolates, where at 24 hrs there were no heat resistant cfus, but by 48 hrs ~ 10^5 heat resistant cfus were observed.

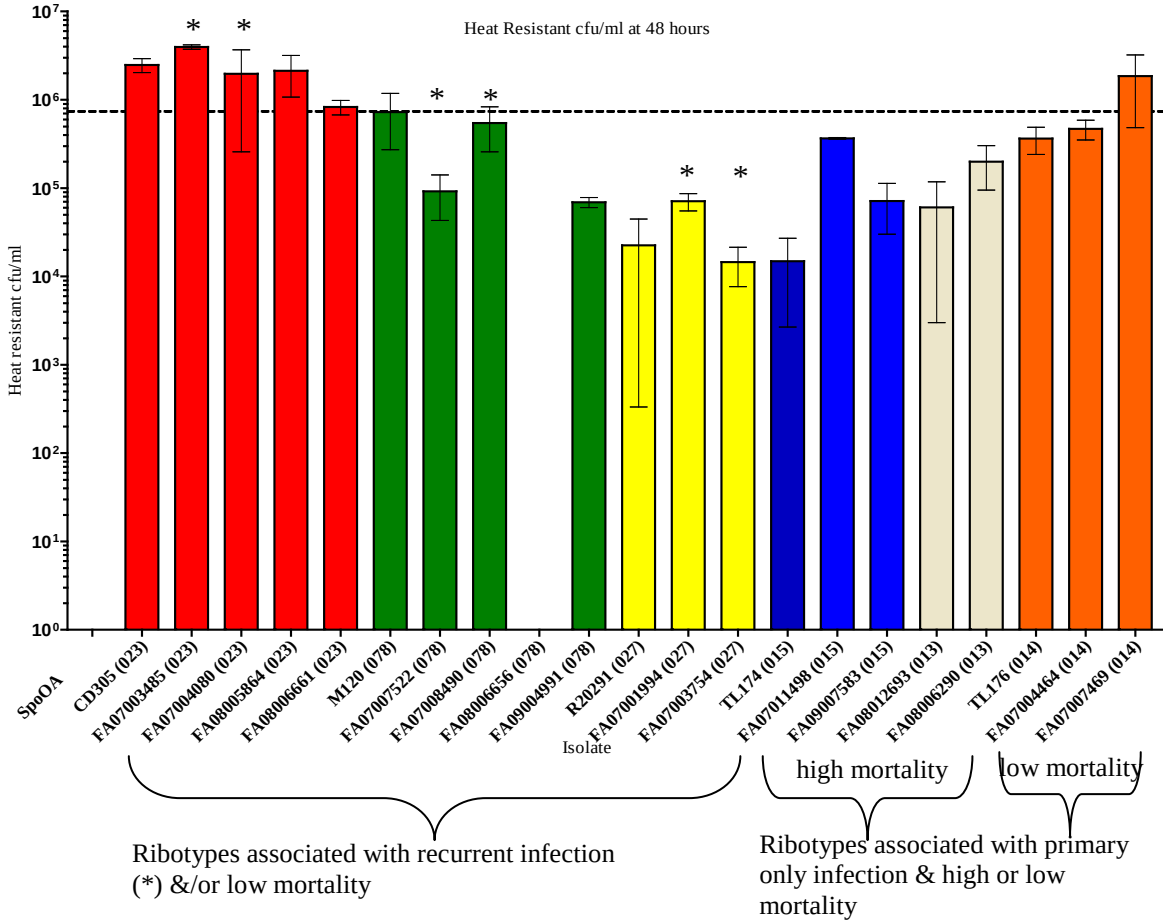


Figure 9.6b- Heat resistant cfu at 48 hrs for clinical and reference isolates under investigation. Error bars= SEM of a minimum of two independent experiments. Dashed line represents mean heat resistant cfu for this time point. Red bars= 023 ribotypes, green bars= 078 ribotypes, yellow bars= 027 ribotypes, blue bars= 015 ribotypes, orange bars= 014 ribotypes and beige bars= 013 ribotypes.

By 120 hrs nearly all clinical isolates were producing heat resistant cfus close to the cohort mean (Figure 9.6c, dashed line). Isolate FA08012693 produced the most heat resistant cfus at 120 hrs (Mean= $1.52x10^7$). Two-way ANOVA analysis of mean heat resistant cfu counts followed by Bonferroni post-hoc comparison tests revealed that isolate FA08012693 (013 Ribotype) had a significantly higher mean heat resistant cfu count ($P<0.01$) than all other clinical isolates (Figure 9.6c). The mean heat resistant cfu count for isolate FA08006290 (013 Ribotype) also differed significantly from isolates FA08006656 ($P<0.05$) and FA09004991 ($P<0.01$) (078 ribotypes).

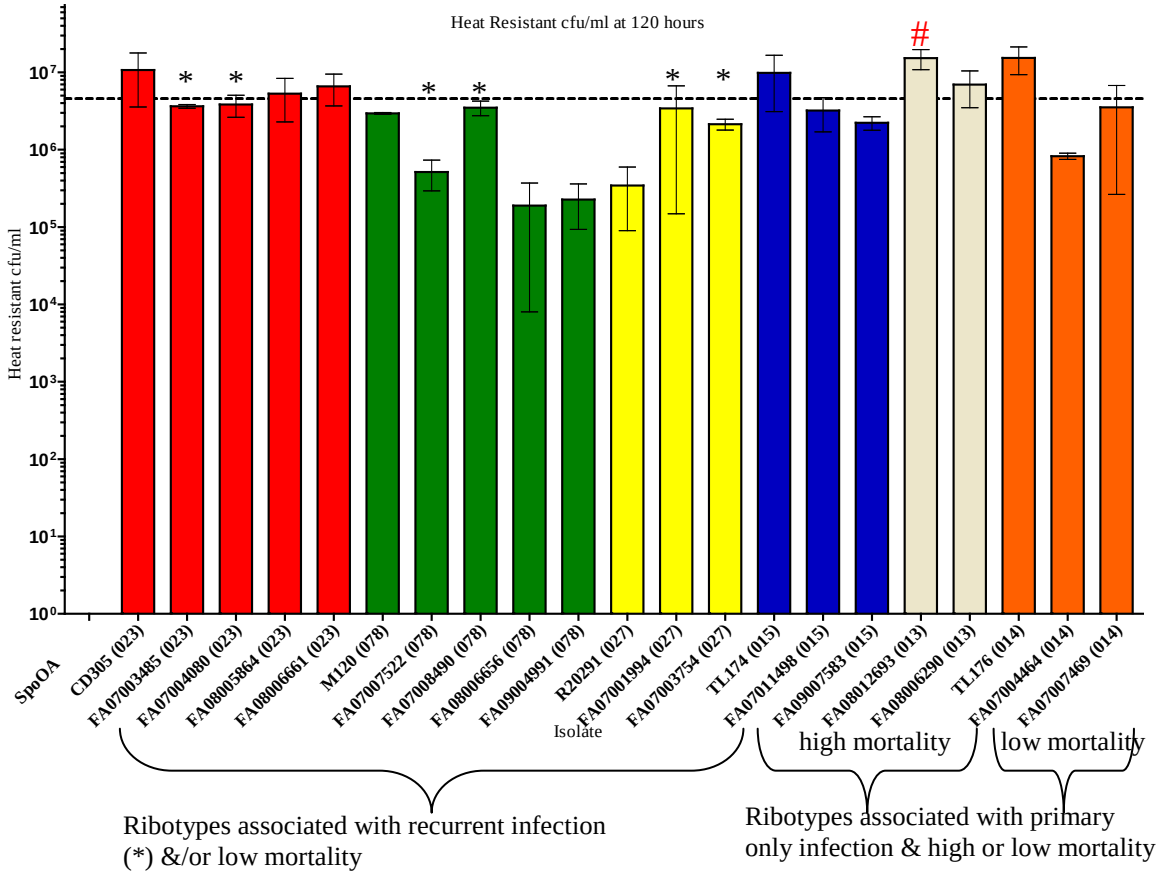


Figure 9.6c- Heat resistant cfu at 120 hrs for clinical and reference isolates under investigation. Error bars= SEM of a minimum of two independent experiments. Red bars= 023 ribotypes, green bars= 078 ribotypes, yellow bars= 027 ribotypes, blue bars= 015 ribotypes, orange bars= 014 ribotypes and beige bars= 013 ribotypes. #= Isolate which was significantly different from most other clinical isolates.

At all time points there did not seem to be any significant differences in mean heat resistant cfus between clinical isolates from recurrent infection (023, 078 and 027 ribotypes) and isolates which came from patients with only one infection (015, 014 & 013 ribotypes) (24 hrs; P=0.291, 48 hrs P=0.218, 120 hrs; P=0.318).

9.3 Discussion

Results from this study aimed to deduce whether there were any intra or inter ribotype differences in virulence mechanisms that correlated with genetic changes seen in Chapter Eight.

9.3.1 Bacterial Motility

Data from this study indicates that bacterial motility is conserved within ribotypes as post-hoc comparison tests revealed that there were no significant differences between the mean motility of the control isolates (M120, CD305, TL174 & TL176) with their respective clinical ribotypes from this cohort. The exception to this was the detection of a significant difference in mean motility between R20291 and FA07001994 (P <0.001). All isolates from the 013, 014 and 015 ribotypes were more motile than the highly motile R20291 isolate, and these were the isolates that were less frequently associated with recurrent CDI. There was one ribotype (023) which was as motile as the highly motile R20291. Consistent with published data, the 078 ribotypes were non-motile. The non-motility of the 078 ribotypes may be attributed to the deletion of a large flagellar operon (F3 region) within the genome (Chapter Eight, Figure 8.3). Flagellar genes are key to bacterial motility as indicated by Aubry *et al*. [15], who constructed *C. difficile* 630 mutants with disruptions in key genes of the F1 (*fliC &* CD0204*)* and F3 (*fliF, fliG, flhb/fliR, fliA & fliM)* regions, that resulted in loss of motility. According to Aubry *et al*. mutations in the genes that rendered *C. difficile* 630 non-motile were also linked to changes in toxin expression. Mutation of *fliC* correlated with increased

gene expression of *tcdR, tcdB tcdA* and *tcdE*, after 4, 7 and 24 hrs of growth, which corresponded to increased toxin level detection in culture supernatants. Conversely, mutation in any one of the F3 genes including the σ^{28}- homolog *fliA* appeared to decrease expression of *tcdR, tcdB tcdA* and *tcdE*, at 4, 7 and 24 hrs of growth, which also correlated with decreased toxin detection in culture supernatants[15]. The present study indicates clear differences in motility of ribotypes, and this could be linked to differences in toxin production as Aubry *et al*.[15] suggests.

9.3.2 *In vitro* detection of TcdA and TcdB

Qualitative observations for detection of TcdA suggest that levels are uniform by 120 hrs for all isolates, and this is possibly due to bacterial cell lysis, resulting in the release of intracellular toxin from the cell, or saturation of the detection system. For all isolates except, those of ribotype 023, TcdA levels are also uniform by 48 hrs. The different toxin phenotype for the 023 isolates (Figure 9.3c) could be due to changes in regulation of toxin expression by TcdC, or the insertion of a 6 Kbp region in-between *tcdA* and *tcdB* (Chapter Eight, Figure 8.5). The lower intensity of detection of TcdA, and lack of detection for TcdB for 023 isolates could, in part, go some way to explaining why this ribotype was found more frequently (although not significantly) associated with recurrent CDI and low mortality. There may be less of an inflammatory response to toxin of the 023 isolates, therefore less damage to the gut epithelia and this might make a more conducive environment for recurrent CDI. Detection of TcdA for the 027 ribotype suggests that they are more robust/earlier toxin producers, as TcdA was detected at 8 hrs of growth (Figure 9.3a). These *in vitro* findings are consistent with published data by Merrigan *et al*.[165], highlighted by others[22, 74], and also supported by studies of this ribotype in the human gut model[47, 72]. Data from this PhD research did not support the hypothesis that ribotypes with a truncated *tcdC* genotype were more robust/earlier toxin producers than those ribotypes which harboured normal *tcdC*.

Figures 9.2 and 9.3 show that seven isolates of four different ribotypes produced TcdA at early stages of growth (8hrs in TY broth), and these isolates harboured both normal *tcdC* and truncated *tcdC* genes (Table 9.6) as discussed in Chapter Eight (Section 8.16.3). The isolates in Table 9.6 came from different infection outcomes and therefore there is some evidence to suggest that, although some isolates are capable of early TcdA synthesis, regardless of whether they have truncated *tcdC* genotypes; there are other factors, which could include the host response to infection, are important at defining the outcome of infection.

Table 9.6- TcdC Genotypes and patient outcomes of isolates that produced TcdA at 8 hours of growth in TY broth.

Isolate	Ribotype	*tcdC* Genotype	Mortality Outcome
FA07001994	027	Δ1Stop 66	Survived
FA07003754	027	Δ1Stop 66	Survived
FA08012693	013	WT	Survived
FA07011498	015	WT	Survived
FA08006656	078	TAAstop 62	Died
FA07003485	023	TAAstop 62	Survived
FA08006661	023	TAAstop 62	Died

It is clear from this study that there are intra-ribotype differences in the time at which TcdA and TcdB are able to be detected. This *in vitro* intra-ribotype diversity of toxin production has been observed in a number of studies[31,9, 199], and some of these suggest that lack of correlation between clinical outcome and the ribotype causing infection could be due to the highly variable *in vitro* toxin levels that are produced within and between different ribotypes[31,9]. This intra ribotype diversity of TcdA production is exemplified most between isolates of ribotypes 013 and 015 (Figures 9.2a & 9.2c). Within these ribotypes, the isolates associated with the patient who died produced TcdA later than the isolate from the patient who survived. There is also some variation between isolates of ribotypes 023 and 078 (Figures 9.3b & 9.3c), where the isolates associated with the patient who died produced TcdA earlier than those isolated from the patients who did not die. Whether or not the changes in

toxin production alone could have an impact on the outcome of CDI remains unclear. But what is known is that the immune/inflammatory response of the host to toxin is an important factor in the outcome of disease[123]. Therefore, it may be likely that for patients in this study; the unique virulence characteristics of the infecting *C. difficile* isolate (such as toxin production) and the individual patients' immune/inflammatory response would have both contributed to the outcome of infection. Evidence from this study data implies that some of the more motile isolates produce toxin at earlier stages of growth compared to the 078, non-motile isolates, which corresponds with data from Aubry *et al.*[15] who showed that loss of motility as a result of deletion of the F1 flagellar genes (as seen in ribotype 078 isolates) resulted in decreased toxin levels at early stages of growth.

Methods employed in this study were unable to detect TcdB in isolates of the 027 ribotype and other ribotypes (014, 078 & 023). The 027 ribotype has been shown to produce both TcdA and TcdB, and it unclear why this assay failed to detect TcdB, especially when this method was adapted from work by Sirard *et al.*[199] and used the exact same mAb as described in their methods. DNA sequence analysis of all isolates concluded that *tcdB* was present in all isolates, therefore, lack of *tcdB* could not account for lack of detection of TcdB. There are a number of possible reasons that may account for the lack of TcdB detection. One possibility was the specificity of the mAb to TcdB. To address this, two mAbs were used to try and detect TcdB in 027 isolates. Both mAbs were able to detect TcdB in other isolates (Figure 9.3) but not the 027 isolates. Thus, it can be concluded that the mAbs used in this study were not specific to epitopes of TcdB from all *C. difficile* ribotypes. A study by Pruitt *et al.*[184] deduced that TcdB is more heterogeneous in its structure than TcdA, and work by Dingle *et al.*[58] showed the carboxy terminus receptor binding domain (RBD) of TcdB is heterogeneous between different ribotypes. These observations, coupled with literature that suggests the RBDs of both TcdA and TcdB are most effective at generating toxin neutralising antibodies[17,

[126], could explain why the mAbs used in this study are not detecting TcdB from all ribotypes. Differing TcdB epitopes may be sufficiently divergent as to not be recognised by the mAbs used in this study. Amino acid sequence analysis of TcdB revealed distinct patches of different amino acid sequence conserved within ribotypes, and this could explain the lack of affinity of the mAbs for TcdB in some ribotypes. Another possible reason is that levels of production of TcdB by some isolates were too low to be detected by both mAbs. A polyclonal antibody was also used to detect TcdB from different ribotypes; however it was non-specific and detected the presence of cross-reactive proteins in the toxin negative strain.

To further clarify the reasons for lack of TcdB detection by this assay it may be necessary to validate expression of *tcdB* via quantitative PCR (qPCR) in order to determine if gene transcription is occurring in all isolates. If *tcdB* mRNA levels were measurable then the next step would be to try and purify toxin from culture supernatants and re-probe with antibody specific for TcdB. In parallel, a sandwich ELISA could be utilised, to try and quantify the levels of TcdA and TcdB in the culture supernatants, thus providing more quantitative data. However, due to time constraints of this PhD project it was not possible to carry out these experiments.

9.3.3 Antibiotics MIC profiles

Other important factors that could influence the outcome of CDI is the emergence of isolates that are resistant to CDI treatment therapies such as metronidazole and vancomycin, and the exacerbation of disease due to inappropriate use of antibiotic classes such as the fluoroquinolones and cephalosporins[78]. There is increasing concern over the use of vancomycin in the health care system due to its effect on increasing cases of infection with vancomycin resistant *enterococci* (VRE)[48, 174]. There have also been studies showing decreased susceptibilities or ribotype 001 isolates to metronidazole[18] and that in *C. difficile* populations there may be hetero-resistance to metronidazole[174]. Fluoroquinolones were

prescribed to patients in the study cohort, with ciprofloxacin being the most frequently prescribed antibiotic in this class. All isolates chosen for MIC analysis were resistant to ciprofloxacin (≥ 6 μg/ml), which is consistent with findings by others[13, 201]. Most isolates were highly resistant to ciprofloxacin (≥ 32 μg/ml) except isolates from ribotype 078, whose MIC levels varied between. All 078 isolates shared the same mutations in GyrA and GyrB (Chapter Eight, Table 8.4); therefore the varying resistance could be due to either varying transcriptional control of *gyrA* and *gyrB* genes or mutations in other genes i.e. topoisomersases that lead to altered resistance to ciprofloxacin.

Genetic analysis of *gyrA* and *gyrB* genes revealed that the 027 and 014 ribotypes shared the Thr82Ile amino acid substitution in GyrA which confers resistance to newer fluoroquinolones such as gatifloxacin, moxifloxacin and levofloxacin[202]. However, no other isolates shared this amino acid substitution, and instead, had different amino acid substitutions at different locations (Chapter Eight, Table 8.17). These amino acid substitutions could be conferring resistance to the newer fluoroquinolones and further antibiotic susceptibility testing of all isolates, to newer fluoroquinolones would need to be investigated to deduce the contribution of known and novel amino acid substitutions to resistance profiles. The increased resistance of isolates to ciprofloxacin may have facilitated CDI infection within this cohort and a review by Gerding highlights that the overuse of the fluoroquinolones in a clinical setting are part of the driving factor for CDI outbreaks[78]. Within England, a significant decrease ($P<0.05$) in the use of fluoroquinolones was noted in CDI cases from 2007-2010. Therefore steps are being taken to improve hospital prescribing policy of these antibiotics. All of the isolates tested in this study were susceptible to metronidazole and vancomycin, although there were two 027 isolates that had a higher level of metronidazole resistance than all others (R20291; 0.19 and FA0701994; 0.146 μg/ml respectively). Whole genome analysis of all isolates did reveal the presence of a putative orphan gene; 5-nitroimadazole reductase (Chapter Eight, Table 8.17),

which could be linked to the metabolism of 5-nitroimidazole drugs[43], but given that these isolates were all susceptible to metronidazole they are not likely to be involved in metronidazole resistance. Slight differences in MICs could be attributable to slight differences in environmental conditions between biological experimental repeats or the choice of E-test MIC method for determining antibiotic susceptibilities to metronidazole. Whilst the agar dilution MIC determination method for anaerobes is recommended by the National Committee for Clinical Laboratory Standards (NCCLS)[179] the E-test has been used as a simple alternative for MIC testing in *C. difficile*, with results the correlate well with the reference method[179]. The exception to this correlation is found when trying to deduce metronidazole susceptibilities in anaerobes. Many studies have found that resistance to metronidazole in anaerobes and specifically *C. difficile* is lower via the E-test method than the agar dilution method[50, 171, 179]. This lowered detection of resistance may have hidden the true isolate resistance breakpoints in this study and thus, the resistance of isolates to metronidazole would need to be re-evaluated using the agar dilution method, if more time were available.

Chapter Eight (Section 8.16.3) discussed the findings that 013, 014, 015 and 027 contained putative genes organised into a *vanG*-like operon which, in *Enterococcus faecalis* encodes moderate vancomycin resistance[57]. Although no isolate from the cohort in this study was resistant to vancomycin, there may be selective pressure on genomes of these isolates which could lead to the emergence of resistant isolates.

There were no statistical differences in mean antibiotic MICs for primary isolates and secondary isolates which came from patients with recurrent infection. This, coupled with genetic evidence (Chapter Eight) which found that there were no SNP differences in antibiotic resistance genes, or any genes involved in resistance mechanisms, suggests that selection for antibiotic resistance is not occurring during the course of infection.

Attributing cause and effect of individual antibiotics for CDI cases is difficult[229] as many patients are often on multiple antibiotics for other co-morbidities and often (as is the case in this study) sample cohorts are of insufficient size and there are many with inadequate controls[212]. During this study there were no significant differences in mean antibiotic MICs for ribotypes found significantly associated with primary only infection (013, 014 & 015) versus those that came from patients with recurrent infection (027, 078 & 023), and no significant differences between mean antibiotic MICs for ribotypes that were found to be significantly associated with low mortality (014, 078, 023 & 027) versus those who were not (015, 013). There were also no significant differences in mean MICs between isolates that had come from a primary infection versus isolates that had come from secondary infections. Therefore, evidence may suggest that in patients in this study cohort, resistance of *C. difficile* to certain antibiotics may have had a general role in contributing to the onset of CDI, and its ability to reside within clinical facilities, rather than having an impact on infection outcomes such as recurrence and/or mortality.

9.3.4 Sporulation Characteristics

The formation of heat stable spores which are resilient to a multitude of chemical and environmental detergents facilitates the transmission of *C. difficile* within a clinical environment. The efficiency of sporulation may explain why certain ribotypes are emerging within the clinical setting. In this study, after 120 hrs of growth, spore titres for the majority of clinical isolates are uniform, and this is something which is supported by Burns *et al*.[39]. At all time points there did not seem to be any significant differences between the mean production of heat resistant cfus between groups of clinical isolates from recurrent infection (023, 078 and 027 ribotypes) and isolates which came from patients with only one infection (015, 014 & 013 ribotypes) (24 hrs; P= 0.291, 48 hrs P=0.218, 120 hrs; P=0.318). This data also implies that sporulation/germination is ribotype specific, or even isolate specific. The

isolate specific differences in heat resistant cfu counts are clear at early (24 hr) time points (Figure 9.6a), with some isolates either not producing spores, or spores not germinating at this stage, and some which do. These finding are perhaps not surprising given that every clinical *C. difficile* isolate investigated in this study has come from and individual patient whose dynamic intrinsic gut environment is going to have a significant part to play in bacterial sporulation/germination.

Both Akerlund *et al*.[9] and Sirard *et al*.[199] looked at clinical association of strain type, toxin production and sporulation. Akerlund *et al*.[9] observed a general trend that linked low spore count, but high vegetative growth and high toxin titre. During the current study some isolates that produced high spore titres at earlier stages of growth (015 and 014 ribotypes), were still capable of robust TcdA production at early stages of growth. This differs to what is currently hypothesised about spore production and toxin production as exemplified by Akerlund *et al*.[9] and Sirard *et al*.[199]. Data from this research does (in part) support the findings from Rosenbusch *et al*.[190] who found that sporulation was not coupled to toxin production, as in their study there was no statistical difference between toxin titres of *C. difficile* 630 and *C. difficileΔspo0A* isolates. The high sporulation titres and high detection of toxin for some isolates under investigation during this PhD study, could perhaps explain in part, the recent emergence of these ribotypes in the clinical setting. (Chapter Five, section 5.3).

The sporulation experiment was not able to deduce if both 013 isolates, isolate FA07003754 and FA08006656 were either not producing spores at 24 and/or 48 hrs, or producing spores that were unable to germinate. If more time were available it would be necessary to perform the experiment again, enumerating spores by phase contrast microscopy before heat treatment at the specified time points. Performing this count would provide the ability to distinguish whether it was lack of spore formation (no spores visible under phase constant) or the inability of spores to germinate (spores visible under phase contrast) which was giving rise to

the data. Synergistic analysis of the genome of 013 ribotypes highlighted NS SNPs sporulation specific genes (Chapter Eight, Table 8.15) that might account for observations that this ribotype did not produce heat resistant cfus after 24 hrs. Discussion of the SNP changes was detailed in Chapter Eight (Section 8.16.1) and if these changes were found to affect protein function, it would imply that they might have an effect on sporulation of this ribotype.

9.4 Conclusion

Isolates investigated during this PhD study do show differences in motility, toxin production and formation of heat resistant cfus at 24 and 48 hrs. Some of these differences might be attributed to changes in the genome, thus, SNP validation coupled with mutational analysis would need to be conducted to verify these findings. Phenotypic differences of isolates are less apparent when looking at antibiotic resistance profiles to current CDI therapies and ciprofloxacin.

While collectively isolates from the same ribotype show genomic similarity (Discussed in Chapter Eight), phenotypically, some isolates from the same ribotype behave slightly differently. Several studies have looked at phenotypic traits in isolation, or in conjunction with very few other data sets and have been unable to link their observations with any clinical outcome of disease[159]. Thus, the collective conclusion from phenotypic experiments, coupled with what is known about the genetic basis of these ribotypes (Chapter Eight), and the response in the host, may provide more of an insight into the host/pathogen interaction which determines an infection outcome.

The accumulation of data from all methods of investigation applied during this study will be summarised further in a final discussion. This will attempt to convey the general characteristics of each individual *C. difficile* isolate studied during this research, and how

these characteristics might have influenced the infection outcome to which they were associated.

9.5 Concluding Summary and Discussion

9.5.1 Summary

The aim of this PhD was to develop a more comprehensive understanding of both host and isolate association to CDI outcomes such as, mortality and recurrent infection, through data analysis of clinical host variables, and genetic and phenotypic characterisation of clinically relevant isolates. The research detailed in this thesis focused on the mortality and recurrent CDI outcomes, as they pose a significant burden on the healthcare setting.

Three methods of investigation were employed to deduce both host and isolate association to infection outcomes. These included; analysing patient clinical data to try and identify host markers of infection outcomes, evaluating ribotype association to infection outcomes, and genetically and phenotypically characterising the clinically relevant *C. difficile* ribotypes associated with these outcomes, to look for differences in bacterial virulence mechanisms such as those that regulate, toxin production, antibiotic resistance, sporulation and motility.

9.5.1.i Variables Prognostic of *C. difficile* Infection Outcomes

Statistical analysis of patient data from this study was able to identify four variables prognostic of mortality as a result of CDI, but was unable to deduce variables that were associated with recurrent disease. The ability to deduce variables prognostic of recurrent CDI within this study was likely due to the small number of true recurrent infections and the inability to distinguish what impact concomitant antibiotics had on the recurrence of CDI.

In this cohort, the variables derived from clinical data, that were prognostic of mortality included, increased WCC, respiratory rate and CRP, as well as decreased serum albumin. A four point clinical prediction rule was derived from threshold levels of these variables, determined as levels which best classified patients with CDI who were more ‘at risk’ from death. The use of a simple prediction rule in a clinical setting could facilitate the way in

which clinicians are able to effectively manage a patient with CDI, perhaps prompting a different treatment regime if a high risk of mortality is identified. The prediction rule could be used by non-specialists to consider mortality risk when assessing a patient presenting with CDI within or around 48 hrs of diagnosis. It may also be useful in communication within teams and between teams, for example in discussion with microbiology doctors, as well as giving the patient and relatives information that is evidence-based. The clinical prediction rule derived in this study would benefit from prospective evaluation.

Three of the variables found to be prognostic of mortality in patients with CDI are markers of a host inflammatory response, possibly as a consequence of *C. difficile* infection. Statistical analysis of data collated during this study was not able to deduce any ribotypes that were associated with increased mortality and recurrent infection, but some ribotypes, were found to be significantly associated with primary only infection and decreased mortality. However, a limitation of this study was that the power to detect true ribotype specific associations with mortality and recurrent infection was low, because there were only a small numbers of isolates from some ribotypes. Therefore the significant findings deduced in this study generated hypotheses which could be investigated further. And results also emphasises the need for bigger clinical cohorts to investigate ribotype association with infection outcomes.

9.5.1.ii *In vitro* Isolate Virulence Characteristics

As there were so few/no host markers of recurrent CDI and mortality and even though no direct associations could be made between the ribotype association with infection outcomes; a small selection of the isolates that caused CDI in patients during this study were characterised to try and establish if isolate specific differences in virulence mechanisms may have contributed to the outcome of infection to which they were associated. Presented in

Table 9.7 and Table 9.8 is a summary of the phenotypic and genetic observations identified for each isolate studied during this period of research.

While it was not possible to draw direct conclusions about the association of an individual isolates association with an infection outcome due to the small sample size used for isolate to isolate comparison, and the exhaustive characterisation that is needed to establish such as link, there are hypothesises which might be generated from the data, and investigated further.

Within this isolate panel, each isolate within a ribotype appears to have the same general *in vitro* characteristics for the antibiotics tested during this study. Each isolate of a ribotype also shares the same motility characteristics. Thus, the lack of difference in these specific virulence characteristics, between each isolate of a ribotype, may not have had an effect on the outcome of the patient from which the isolates was found.

Toxin production at 8 hrs and differences in production of heat resistant cfu at 24 hrs were the most evident phenotypic differences between ribotypes, and between isolates of the same ribotype. Differences in the production of heat resistant cfus are known to be variable, even between isolates of the same ribotype[38]. This makes it very difficult to assess the contribution of isolate specific differences in generation of heat resistant cfu observed during this study, to the outcome of infection, and emphasises the need for bigger cohort studies to observe true effects that differences in sporulation/germination may have.

Whilst early onset of toxin production by some *C. difficile* isolates has been proposed to be a contributing factor to poorer outcomes of patients in some studies[22], within this panel of isolates, there did not appear to be a link between early onset of toxin production and the outcome of the patient in which the isolate was found. Thus, one hypothesis for such differences in infection outcomes may be better explained by how the host responds to the infecting *C. difficile* isolate, rather than the actual onset of toxin production by *C. difficile*, or differences in other bacterial virulence factors. This hypothesis is partly support by the data

obtained from cross-tabulation of the clinical prediction rule scores of each patient (derived from the variables found to be prognostic of mortality during this study) with the isolates that caused CDI (Table 9.7). The clinical prediction rule score is derived from some variables which are indicative of an inflammatory response in the host (serum albumin, CRP, WCC). Data from Table 9.7 data indicates that for the majority of isolates which were associated with cases where a patient had died, the clinical prediction rule score for that patient was higher in comparison to a patient who did not die, when infected with the same ribotype. This trend was observed in all ribotypes where one patient had died and at least one patient did not. The data for patients in this study could therefore again suggest that the host specific response to the isolate causing infection plays a larger role in the outcome of infection than perhaps the specific bacterial characteristics alone play. However, this is purely speculative and based on a small data set. This hypothesis would need to be investigated on a bigger scale in the remaining cohort of patients identified during this study in order to gain sufficient ribotype sample sizes for isolate to isolate comparison and association to outcomes.

Table 9.7- Overview of phenotypic and genetic characteristics of cohort clinical isolates.

Isolate ID	Ribotype	Mortality Outcome	Recurrent CDI outcome	Presence or Absence of toxin genes			TcdC variants	Motility	TcdA production relative to R20291 at 8 hrs	TcdA production relative to R20291 at 24 hrs	TcdA production relative to R20291 at 48 hrs	TcdB detection	Sporulation at 24 hours	Sporulation at 48 hours	Prediction Rule Score
				TcdA	TcdB	CDT									
FA07003485	023	S	+	+	+	+	TAAstop 62	M	+	–	–	–	+	+	0
FA07004080	023	S	+	+	+	+	TAAstop 62	M	N/A	–	–	–	+	+	0
FA08005864	023	S	–	+	+	+	TAAstop 62	M	N/A	–	–	–	+	+	1
FA08006661	023	D^	–	+	+	+	TAAstop 62	M	+	–	+	–	+	+	3
FA07007522	078	S	+	+	+	+	TAAstop 62	NM	N/A	–	+	–	+	+	0
FA07008490	078	S	+	+	+	+	TAAstop 62	NM	N/A	–	+	–	–	+	0
FA08006656	078	D^	–	+	+	+	TAAstop 62	NM	+	+	+	–	–	–	1
FA09004991	078	S	–	+	+	+	TAAstop 62	NM	N/A	–	–	–	+	+	2
FA07001994	027	S	+	+	+	+	Δ1Stop 66	M	+	+	+	–	–	+	0
FA07003754	027	S	+	+	+	+	Δ1Stop 66	M	+	+	+	–	+	+	1
FA07011498	015	S	–	+	+	–	WT	HM	+	+	+	+	+	+	1
FA09007583	015	D	–	+	+	–	WT	HM	+	+	+	+	+	+	2
FA07004464	014	S	–	+	+	–	WT	HM	+	+	+	–	+	+	0
FA07007469	014	S	–	+	+	–	WT	HM	+	–	+	–	+	+	0
FA08012693	013	S	–	+	+	–	WT	HM	+	+	+	+	–	+	0
FA08006290	013	D	–	+	+	–	WT	HM	+	+	+	+	–	+	1

Mortality Outcome; S= survived, D- Died (^)= not as the result of CDI, Motility; M= motile, HM= highly motile, NM= non-motile. N/A- not producing toxin. +/- = presence or absence of toxin genes. TcdA production and /or sporulation.

Table 9.8- Overview of Antibiotic MICs of cohort clinical isolates.

		Antibiotic MIC µg/ml		
Isolate ID	**Ribotype**	**Mean MET MIC**	**Mean VANC MIC**	**Mean CIP MIC**
FA07003485	023	0.047	1	≥ 32
FA07004080	023	0.028	1	≥ 32
FA08005864	023	0.028	0.91	≥ 32
FA08006661	023	0.039	1.16	≥ 32
FA07007522	078	0.042	1.3	24
FA07008490	078	0.052	1.16	10
FA08006656	078	0.032	1.16	8
FA09004991	078	0.062	0.91	≥ 32
FA07001994	027	0.146	1.33	≥ 32
FA07003754	027	0.066	1.33	≥ 32
FA07011498	015	0.064	0.91	≥ 32
FA09007583	015	0.068	1.33	≥ 32
FA07004464	014	0.074	1.33	≥ 32
FA07007469	014	0.058	0.91	≥ 32
FA08012693	013	0.068	1.67	≥ 32
FA08006290	013	0.049	1.67	≥ 32

MET= Metronidazole, VANC= Vancomycin, CIP= Ciprofloxacin

9.5.1.iii Whole Genome Sequence Analysis

Even though isolate to isolate sequence comparison was based on a small sample selection it was clear that there was an abundance of SNP differences between isolates of the same ribotype. However, isolate specific SNPs were not seen in the specific antibiotic resistance genes investigated during this research; such as *gyrA* and/or *gyrB,* which may have given a selective advantage against the fluoroquinolones, putative vancomycin or metronidazole resistance genes. Nor were there any SNP differences between isolates that had come from a primary infection versus isolates that had come from secondary infections. There also did not appear to be any isolate specific SNP differences in spore and motility related genes. This information suggests the possibility that while the genome of each individual *C. difficile* isolate may undergo selection in different genes, there are common SNPs occurring within a particular ribotype, which may give them a selective advantage in their niche environment and would warrant future investigation from a bigger selection of isolates.

The conservation of transposable and phage elements by different ribotypes was observed during this study. However, only one isolate (FA08006656, ribotype 078) showed a large scale genetic change compared to other isolates of the same ribotype. This isolate was associated with a patient with CDI, whom died, but again, the conservation of this region could not be directly linked to the outcome of the patient, and it would thus prove useful to investigate the conservation of this genetic region in a wider panel of 078 isolates to deduce if there could be a link. It may then also be useful to identify the specific role of the genes which are conserved in this region. Generally, no other gross genetic changes were observed between isolates of the same ribotype that may have accounted for different clinical outcomes. The conservation of insertion sites of these transposable elements within a ribotype is something that could be investigated further on a wider panel of isolates as there may be a selective advantage of the position of these elements within the genome and how they affect surrounding genes.

During this research it was not possible to deduce the impact of all the SNP differences and smaller gene changes that were found in this data set. Given the relative difficulty of trying to investigate how each SNP and gene change could contribute to the outcome of infection it may prove useful for the future, to investigate the types of genes in which the SNPs occur. This may provide insight into the pathogenic strategy employed by individual *C. difficile* isolates to cause different infection outcomes.

9.5.2 Conclusion

What is clear from this research is the importance of identifying both host and isolate determinants of infection outcomes. All the variables analysed in this study clearly only play a small role in defining an infection outcome. There are other factors such as, the host immune response and the gut microbiome, which may play a more significant role in infection outcomes such as, recurrence and mortality. No doubt there will always be studies

looking into these clinical outcomes, as they are of significant importance to clinicians[226]. If studies are to find significant variables that are predictive of outcomes, they must be extremely comprehensive and include data from the host, including; the host immune response to infection, gut microbial constitution at the onset of, and during infection, and after the cessation of CDI treatment, clinical blood markers, host DNA (to identify susceptibility markers) as well as demographic parameters analysed in this study. The studies should also include information about the ribotype that caused the infection, which can then be linked to its potential virulence, with regards to toxin production, sporulation, antibiotic resistance and motility. Knowing the genetic background of these isolates could be used in conjunction with bacterial phenotypic characteristics and host response to elucidate bacterial markers of infection outcomes. Study samples must be large enough to capture sufficient data for the intended outcome measure, and definitions of parameters for outcome measures must be clear before sample collection commences. Perhaps the establishment of a comprehensive study protocol that all researchers could use when attempting to conduct a study of this type would provide uniformity of the way in which studies into identifying infection outcomes are conducted, thus allowing comparisons of multiple data sets in order to deduce true predictors of infection outcomes.

9.6 Future Work

Outlined in this section are future experiments that would be conducted to clarify some of the remaining questions resulting from research conducted during this PhD. It also includes a wider plan of experiments that could be conducted, and might be necessary, to facilitate an even greater depth of understanding of *C. difficile* isolates investigated in this study.

9.6.1 Expanding the Panel of Isolates Chosen for phenotypic and Genetic Analysis

Phenotypic and genetic analysis was conducted on a small panel of isolates during this study and therefore data interpretation is limited. It would therefore be interesting to characterise more isolates of the ribotypes investigated during this study, in order to identify true isolate specific differences which may have contributed to the outcome of infection. There were also two ribotypes that were not investigated during this study (027 and 012). Given the change in prevalence of these ribotypes over the study period, it would be interesting to investigate the phenotypic and genetic characteristics of these strains in relation to changes in characteristics of the clinical cohort in order to try and identify possible reasons for their prevalence change.

9.6.2 Further Characterisation of Isolates Studied During this Research

9.6.2. i Toxin Gene Expression, Protein Purification and Cytotoxicity Assays

- qPCR of PaLoc, *spo0A* and F1 flagellar genes; 078 isolates only, to determine if and when they are being expressed during differential growth stages
- *In vitro* cytotoxicity tests of *C. difficile* culture supernatants to determine the toxicity of the isolates characterised during this study.
- Purification of TcdA and TcdB from bacterial cultures.
- Sandwich ELISA to determine toxin concentration in bacterial cultures.

9.6.2. ii Sporulation Studies

- Microscopic determination of spore numbers at multiple time points to establish if lack of heat resistant cfus at 24 and 48 hrs for some isolates was the result of delayed sporulation and/or germination.

9.6.2. iii Antibiotic Resistance to Levofloxacin, Moxifloxacin and Gatifloxacin

- Repeat antibiotic MIC experiments to determine if any of the amino acid changes observed in GyrA and GyrB could contribute to resistance of *C. difficile* isolates to the newer fluoroquinolones.

9.6.2. iv Whole Genome Sequence Analysis

- PCR confirmation of SNPs in sporulation and virulence genes to determine if the SNP sites found during genetic analysis were real.

9.6.2. v Adhesion Assays

- Adhesion assays to determine the ability of flagellated and non-flagellated isolates to adhere to Caco-2 tissue culture cells.

9.6.2. vi Transcriptomics

- *In vitro* and *in vivo* RNA-sequencing of clinical isolates to determine their gene expression profiles.

9.6.2. vii *In vivo* Infection of Mice and Hamsters

- *In vivo* mouse and hamster infection model; Challenging mice and hamsters with emerging isolates to determine the severity of infection they cause, their *in vivo* colonisation abilities and to determine their transmission efficacy by enumerating spore titres in faeces.

Chapter Ten. Materials and Methods

10.1 Culture Media

Table 10.1- Culture Media.

Culture Media	Composition	Supplier
BHI[a]	Brain infusion solids (12.5 g/L), Beef heart infusion solids (5.0 g/L), Proteose peptone (10.0 g/L), Glucose (2.0 g/L), Sodium chloride (5.0 g/L) and Disodium phosphate (2.5 g/L),	Oxoid
BHIS Broth	As above but with the addition of yeast (5 mg/ml) and L-cysteine (0.1% w/v).	**BHI**- Oxoid **L-Cysteine**- Sigma-Aldrich **Sodium taurocholate** - Sigma-Aldrich **Yeast Extract**- Oxoid
BHIS (Sodium taurocholate) Broth	As above but with the addition of sodium taurocholate (0.1% w/v)	As above
BHI Sodium taurocholate Agar	As BHI but with the addition of sodium taurocholate (0.1% w/v) and Agar (6.1 g/400ml)	**Number 2 Bacteriological Agar** – Lab M Limited
Cycloserine-cefoxitin egg-yolk (CCEY) Braziers Agar	Peptone Mix (23.0 g/L), Sodium chloride (5.0 g/L), Soluble Starch (1.0 g/L), Agar (12.0 g/L), Sodium bicarbonate (0.4 g/L), Glucose (1.0 g/L), Sodium pyruvate (1.0 g/L), Cysteine HCl (0.5 g/L), Haemin (0.01 g/L), Vitamin K (0.001 g/L), L-arginine (1.0 g/L), Soluble pyrophosphate (0.25 g/L), Sodium succinate (0.5 g/L), Cholic acid (1.0 g/L), p-Hydroxyphenylacetic acid (1.0 g/L), D-cycloserine (250 mg/L), Cefoxitin (8 mg/L) and Egg Yolk (40 ml/L)	**Braziers Agar** – Oxoid **Cycloserine/Cefoxitin**- Oxoid **Egg Yolk**- Oxoid
Tryptose-Yeast broth (TY)	Bacto™ Tryptose (Enzymatic digest of protein 8 g/L), Yeast extract (5 g/L), NaCl (5 g/L)	**Bacto™ Tryptone**- Becton, Dickinson U.K. Limited **NaCl**- Sigma-Aldrich **Yeast Extract**- As above

a- Supplemented with agar if necessary

10.2 Buffers

Table 10.2- Buffers.

Buffer	Composition	Supplier
Tris buffered saline (TBS)	NaCl (8 g/L), KCl (0.2 g/L) and Tris Base (3 g/L) adjusted to pH 7.6 for 1X solution	All chemicals from Sigma-Aldrich
Tris buffered saline with Tween-20 (TBST)	NaCl (8 g/L), KCl (0.2 g/L), Tris Base (3 g/L) and Tween-20 (1 ml/L) adjusted to pH 7.6 for 1X solution	All chemicals from Sigma-Aldrich
Phosphate buffered saline (PBS)	1 PBS tablet per 200 ml water	Sigma-Aldrich
Tris Acetic acid (TAE)	Tris- Acetate (40 mM), EDTA (1 mM) for 1X solution	Sigma-Aldrich

10.3 Bacterial Isolates

Table 10.3- Antibiotic susceptible positive control strains

Strain Name	Source	Depositor and History
***Staphylococcus aureus* subsp. *aureus* Rosenbach ATCC 29213**	LGC Standards, UK	Micro-Media Systems, Inc. ATCC <<--Micro-Media Systems, Inc.<<-- E.H. Gerlach strain Wichita
***Bacteroides fragilis* (Veillon and Zuber) Castellani and Chalmers ATCC 25285**	LGC Standards, UK	WE Moore ATCC <<--WE Moore<<--A. Sonnenwirth EN-2 <<--- NCTC 9343 <<--- L. Garrod

Reference *C. difficile* strains used in this study are presnted Table 10.4 below.

Table 10.4- List of reference *C. difficile* strains used in this study.

Strain Name	**Ribotype**	**Source**	**Mutation Phenotype / isolation information**	**NCBI RefSeq/GenBank ID**
C. difficile* 630Δ*erm	012	Dr Steve Michell	Erythromycin sensitive derivative of 630 This strain is the epidemic type X variant that has been extensively studied in research and clinical laboratories. It produces both toxin A, and B and was isolated from a case of PMC in Zurich.	(630 strain NC_009089.1)
R20291	027	David Burns-University of Nottingham	*Clostridium difficile* R20291 was isolated in Stoke Mandeville Hospital, UK in 2006 and is closely related to the North American hypervirulent BI strains.	NC_013316.1
C. difficile* 630 Δ *SpoOA*:: *erm	012	David Burns-University of Nottingham	Insertional mutation in the *SpoOA* gene using an eyrthromycin cassette This isolate is unable to form spores and is erythromycin resistant	N/A
***C. difficile* M120**	078	Dr Trevor Lawley-Wellcome Trust Sanger Centre	*Clostridium difficile* M120 is a human strain isolated in the United Kingdom in 2007	FN_665653
***C. difficile* CD305**	023	Dr Trevor Lawley-Wellcome Trust Sanger Centre	Isolated from a 74 year old male with severe symptoms in 2008	N/A
***C. difficile* TL174**	015	Dr Trevor Lawley-Wellcome Trust Sanger Centre	Isolated in 2009	N/A
***C. difficile* TL176**	014	Dr Trevor Lawley-Wellcome Trust Sanger Centre	Isolated in 2009	N/A
***C. difficile* CF5**	017	Dr Trevor Lawley-Wellcome Trust Sanger Centre	Isolated in 1995. This is a natural $tcdA^-$ $tcdB^+$ variant	N/A
***C. difficile* 07-584**	Unknown	Beata Walter from the University of Ljubljana,	This is a toxin negative isolate	N/A

10.4. Statistical Analysis Methods

10.4.1 Study Cohort

Royal Devon and Exeter Hospital (RD&E) Trust is a 49 ward, 797 inpatient-bed hospital. The mean CDI rate was 6.55 trust apportioned cases per 10,000 bed days from the period March 2007- April 2009 as reported by the Health Protection Agency. Data was

retrospectively collected from 213 patients, including those with recurrent CDI, and thus a total of 245 cases were available for analysis. Inclusion criteria for this study were patients with a *C. difficile* toxin A/B positive stool sample confirmed using the TechLab® C.diff Quik Chek Complete™ Enzyme linked Immuno-assay (Alere Ltd, UK), who were transferred to the specialised *C. difficile* cohort ward between 2007 and 2009. A potential limitation of this test is that its positive and negative predictive value (PPVs and NPVs) power can vary with differences in CDI prevalence rates[178]. This may have meant that some true positive patient data was excluded from the initial data collection and some false positive patient data was incorporated.

The *C. difficile* cohort ward is a 19 bed ward with 7 double doored, single occupancy side rooms and cohorted bays with closed doors. Patients were managed by a multi-disciplinary specialist team including a physician, microbiologist, antibiotic pharmacist, infection control nurse and physiotherapists. Having identified the study population, the clinical notes were reviewed for the admission period involving the positive culture result. Information parameters collected include: biographical details, co-morbidities, concurrent medications, pre-admission place of residence, antibiotic administration in the previous two months, faecal calprotectin results, routine blood tests and observations at time of diagnosis. Clinical blood measurements and routine observations i.e. blood pressure, respiratory rate etc. were those first taken within 48 hrs of *C. difficile* toxin A/B positive test. Formal statistical advice for data analysis was sought to try and establish a suitable cohort sample size to assess a) variables significantly associated with mortality and recurrent CDI and b) ribotypes that were significantly associated with mortality and recurrent CDI. A power calculation was difficult to determine as this study was undertaken retrospectively, therefore only a limited number of clinical data and stool samples were available for analysis, and also due to the fact that mortality and recurrent CDI rates may vary over time depending on multiple variables which

may include, the infecting *C. difficile* isolate, patient comorbidity and sites of outbreaks. Nevertheless, attempts were made to deduce variables prognostic of mortality and recurrent CDI and also to try and identify possible ribotype associations with infection outcomes, using statistical analysis as outlined below.

Data was anonymised and inputted in to PASW Statistics 18 (SPSS Inc, IBM, USA) software. Nominal variables dichotomised using binary format unless dates were used as an input value. Descriptive statistics or frequency statistics/histograms were generated for the appropriate variables on a case or patient bases.

Variable groups were then tested for normality for each outcome measure. Tests for normality were performed using the Shapiro-Wilk test where a $P>0.05$ was associated with normally distributed data. Data was excluded on a test basis.

Univariate and multivariate identification of variables significantly associated with mortality and recurrent CDI was conducted on one case of patient data (N=213). For analysis of differences in variable groups for the mortality outcome, data analysis was performed on all cases with primary only CDI, and on one set of data from those with recurrent CDI based on the last data entry, to account for those which may have had a worse outcome on subsequent infections. To detect significance of variables to the outcome recurrence, data analysis was performed on all cases with primary only CDI and on one set of data from those with recurrent CDI, based on the index case data entry. Not all cases were used in the initial analysis in order to leave a portion of data to then be incorporated in the classification and regression tree (CRT) analysis at a later point

All variables were tested for significance using Bonferroni correction (where $p=0.05/N$; N=number of variables being tested) on a test by test basis.

10.4.2 Univariate analysis

All variables were tested for significance in relation to outcome measures by Independent t-tests, Independent samples median tests, One-Way ANOVA and Chi-squared tests. Levenes test for equality of variances were used to deduce significance.

10.4.3 Multivariate Analysis

Before significant variables were entered into the regression analysis they were partially correlated with other variables to assess if they were truly independent predictors. Bivariate correlation was also used to correlate calprotectin with other potential markers of CDI. Independent predictors of outcome measures were analysed via multinomial statistical regression using a block entry method.

10.4.4 Classification Models

A Classification and Regression Tree (CRT) criteria was chosen as a growing method as it uses a recursive partitioning method and builds classification and regression trees for predicting categorical predictor variables (classification). The prediction rule was derived by inputting all 245 cases of data (patients with primary only and recurrent CDI) into the analysis, which was then split into training and test data (50:50) in order that a sufficient amount of data was captured in each portion. The model used an automatic maximum tree depth from the root node of 5, a minimum number of cases in a parent node of 10 which if split into further groups (child nodes), would contain a minimum number of 5 cases, to ensure outcome measurements were based on a sufficient amount of data. A misclassification cost of 3 was given for the prediction of death where the outcome was survival and a cost of 6 was given for the prediction of survival where the outcome was death, to account for the sample size bias. The CRT growing method attempts to maximize within-node homogeneity. The extent to which a node does not represent a homogenous subset of cases is an indication

of impurity. For the categorical dependent variable a Gini impurity measure was selected in which splits are found that maximize the homogeneity of child nodes with respect to the value of the dependent variable. Gini is based on squared probabilities of membership for each category of the dependent variable. It reaches its minimum (zero) when all cases in a node fall into a single category (SPSS Inc).

10.4.5 Binomial Analysis

Binomial analysis was conducted on patient data to test for ribotype association to recurrent infection and CDI related mortality. Two analyses were performed and are detailed in the appropriate chapter.

10.5 Culturing *Clostridium difficile* from Frozen Faecal Samples

Frozen stool samples were inoculated onto cycloserine-cefoxitin egg-yolk (CCEY) braziers agar (Oxoid) and incubated in an anaerobic cabinet (Don Whitley Scientific) in an environment with the composition; CO_2 10%, H_2 10%, and N_2 80% for 48 hrs. Colonies of typical *C. difficile* morphology (yellow, with granulated edges) were sub-cultured onto brain heart infusion (BHI) agar plates until pure culture plates were obtained. Isolates were stored at 4^{o}C on Amies with charcoal transport swabs (Sterilin) or grown overnight in BHI broth, from which an 800 µl aliquot was frozen at -80^{o}C in 200 µl (80%) glycerol.

10.6 *Clostridium difficile* Growth Curves

C. difficile grown on BHI agar plates were sub-cultured into 10 ml BHI broth supplemented with L-Cysteine (0.1% w/v) and yeast (5 mg/ml) (BHIS broth). Cultures were grown anaerobically overnight and the OD_{590} was read. 50 ml BHIS broth was inoculated to an approximate OD_{590} of 0.05 with the overnight cultures which equated to 0 hrs of growth. Samples were taken from cultures at 1 hr intervals at 0 to 10 hrs, then at 24 and 48 hrs, and were serially diluted (1:10). Each dilution (10 µl) was inoculated onto BHI agar, and colony

forming units (cfu) were counted. CFU/ml was estimated by the equation (1/mean cfu)/volume plated.

10.7 Ribotyping

Ribotyping was carried out according to a standard operating procedure (ARUSOP 039), used by the Anaerobe Reference Laboratory, Cardiff. DNA was extracted from *C. difficile* using chelex 100 resin (Sigma-Aldrich). *C. difficile* colonies were isolated from BHI agar plates and heated at 100^{o}C for 12 mins in a 5% chelex suspension. Aqueous crude extract was transferred into a 0.5 ml eppendorf tube and 5 µl was used immediately for PCR. PCR amplification was performed on the crude DNA extract using 0.5 µl (50 µM) primer P3 = 5′-CTG GGG TGA AGT CGT AAC AAG G-3’ and 0.5 µl (50 µM) primer P5’ = 5′-GCG CCC TTT GTA GCT TGA CC-3’, 10µl dNTPs (10 mM each; Promega), 1.5 µl $MgCl_2$ (1.5 mM; NEB), 0.25 µl Taq polymerase (5000 u/ml; NEB) and 32.25 µl water. Cycling conditions of 95˚C for 2 min (initial denaturing step), 92˚C for 1 min, 55˚C for 1 min, 72˚C for 1 min 30 seconds (for 30 cycles), 95^{o}C for 1 min, 55˚C for 45 sec, 72˚C for 5 min and 4˚C for infinity were used to amplify target region using a BioRad thermocycler. PCR products were transferred to 1.5 ml eppendorfs and concentrated to ~ 20 µl by heating at 75^{o}C (with lids open) for 45 mins. PCR products were mixed with gel loading dye, blue (6X, NEB) and run on a chilled (4^{o}C) 3% metaphor agarose gel (Lonza, UK), in chilled (4^{o}C) 1X TAE at 60mA, 200 volts for 3 hrs with 5 µl Superladder Low 100bp Marker (Fisher). Gels were stained with ethidium bromide (1:20 dilution in water; Sigma-Aldrich) for 20 min and destained for 5-10 min in deionised water. Gel images were captured on a BioRad gel doc image analysis system with Quantity One™ software and banding patterns on gels were analysed using GelCompare (Bionumerics, Applied-Maths, NV) analysis software at the Anaerobe Reference Laboratory, Cardiff.

10.8 Calprotectin ELISA

The calprotectin ELISA assay (Buhlmann Laboratories AG) was carried out according to manufacturer's specifications. 50-100 mg stool was weighed into a 2 ml universal container using a 5 µl inoculation loop. Once the desired volume of stool had been weighed the inoculation loop was broken, leaving the lower portion in the universal tube. Extraction buffer was added at 49X the weight of the stool sample. Samples were homogenised in extraction buffer at 20^{o}C for 45 mins at 400 rpm. Extracts were transferred to a clean 2 ml eppendorf tube and centrifuged for 5 mins at 3000 xg. Sample supernatants were transferred to clean 2 ml eppendorfs and used immediately in the extended range ELISA procedure which allows quantification of calprotectin from 30-1800 µg. Stool extracts were diluted 1:150 with incubation buffer and equilibrated for 5 mins at 18^{o}C-28^{o}C for 5 mins. A 96-well ELISA plate was washed, twice for 20 seconds with 300 µl. Wash buffer was emptied and the plate blotted onto a suitable absorbent source. 100 µl incubation buffer was pipette into wells A1 and A2 (blank), 100 µl calibration buffer A was pipetted into wells B1 and B2, 100 µl calibrator B was pipette into wells C1 and C2 etc., until calibrator B was added to wells F1 and F2. 100 µl low and high controls were pipetted into wells G1 and G2 and H1 and H2 respectively. 100 µl of each diluted sample was pipette in duplicate, into subsequent wells. The plate was covered with plate sealer and incubated for 30±5 mins at 18^{o}C-20^{o}C at 500 rpm in a shaking incubator. The wells were emptied and the wash procedure was repeated three times. 100 µl enzyme label was pipette into all wells and the plate was covered with plate sealer and incubated for 30±5 mins at 18^{o}C-20^{o}C at 500 rpm in a shaking incubator. The wells were emptied and the wash procedure was repeated five times. 100 µl pre equilibrated TMB substrate solution was added to all wells. The plate was covered with plate sealer and incubated for 15±2 mins at 18^{o}C-20^{o}C at 500 rpm in a shaking incubator. 100 µl stop solution was then added to all wells. Absorbance was measured at 450 nm in a microtiter plate reader.

Samples which gave an initial concentration higher that the concentration of calibrator E, were diluted and assayed again. Data resulting from the final concentration of calprotectin, taking in to consider any dilutions was graphically visualised using GraphPad Prism 5.01 and statistical analysis with regards to each CDI infection outcome measure was analysed using PASW Statistics 18 (SPSS Inc).

10.9 Antibiotic Minimum Inhibitory Concentration (MIC) Assay

MIC via the E-test method (Biomerieux) was carried out according to the method of Babut *et al*.[20]. Overnight colonies of *C. difficile* were re-suspended in PBS to a McFarland standard of 1 (Biomerieux). Cultures were swabbed in three directions onto pre reduced BHI agar plates and air dried for 15 mins. E-test strips of ciprofloxacin; metronidazole and vancomycin were overlaid onto the plates and then incubated under anaerobic conditions for 24 hrs. MICs were read at the point that the zone of complete inhibition intersected with the MIC scale. Antibiotic susceptible positive control strains (as recommended by the manufactures specifications) *Staphylococcus aureus* ATCC 29213 and *Bacteroides fragilis* ATCC 25285 (Table 10.4) were tested with all *C. difficile* isolates.

10.10 Motility Assay

C. difficile was grown on BHI agar plates, overnight in an anaerobic environment. Fresh 0.1% BHI agar was aliquot into 30 ml universal tubes, air dried for 40 mins and reduced in an anaerobic environment for a minimum of 2 hrs. A 1 µl loop of bacteria was inoculated into the top 2-5 mm of BHI agar and incubated for 48 hrs.

10.11 Sporulation Assay

This assay was adapted from Burns *et al*.[38].

BHI broth supplemented with L-Cysteine (0.1% w/v) and yeast (5 mg/ml) (BHIS broth), plus 0.1% sodium taurocholate was inoculated with colonies grown on BHI agar and incubated in

an anaerobic cabinet for 24 hrs. 5 ml BHIS broth was inoculated to an OD_{590} 0.05 and grown to OD_{590} 0.1-0.2 (~1-5 10^7 cfu). Broth from the starter cultures were standardised to an OD_{590} of 0.1 and 100 µl (~ $1x10^5$ cfu) of the standardised starter culture was used to inoculate 10 ml of fresh BHIS broth (0 hrs). This was to try and ensure that no spores were present in the culture at 0 hrs and that differences in cell densities at 0 hrs did not affect resulting heat resistant cfu counts. Cultures were grown for five days in an anaerobic environment. Total cfus (spores and vegetative cells) were enumerated at 0 hrs and 120 hrs by serially diluting (1:10) the culture in phosphate buffered saline (PBS), plating out dilutions onto BHI plates supplemented with 0.1% Sodium taurocholate, which is known to facilitate spore outgrowth [230], and then grown anaerobically for 24 hrs.

Heat resistant cfu were obtained by heat treating 100 µl of culture for 25 min at 65^{o}C (to kill vegetative cells). The culture was cooled, centrifuged at 9000 rpm and the supernatant removed. The cell pellet was re-suspended in 100 µl cold PBS+ 0.05% tween-80, and the process was repeated as necessary, up to a maximum of two times. Heat resistant cfu/ml were enumerated by serially diluting (1:10) the culture in PBS and plating out dilutions onto BHI plates supplemented with 0.1% (w/v) sodium taurocholate, and grown anaerobically for 24 hrs at 37^{o}C. Data was input into GraphPad Prism 5.01 to determine total cfus (spores and vegetative cells) and heat resistant cfu count at 0 hrs and 120 hrs and heat resistant cfu count at 0, 24, 48 and 120 hrs. Experimental data is based on a minimum of two biological repeats.

10.12 *In vitro* Toxin A and B Immunodot Blots

10.12.1 Bacterial growth

This method was adapted from a protocol described by Sirard *et al.*[199]. *C. difficile* grown on BHI agar was inoculated into 10 ml BHIS broth and grown overnight. 10 ml BHI/TY broth was inoculated to an OD_{590} 0.05 (~10^6-10^7 bacterial cells) with the overnight culture. Cultures were grown anaerobically for 120 hrs. At 8 hrs, 24 hrs, 48 hrs and 120 hrs 500 µl of culture

was aliquot into 1.5 ml eppendorf tubes, and centrifuged at 13000 rpm for 1 minute. 400 µl culture supernatant was aliquot into clean eppendorf tubes and frozen at -20^{o}C until further use.

10.12.2 Immunoblot

Sample supernatants were thawed on ice and 10 µl neat culture supernatant was spotted onto Hybond-ECL nitrocellulose membrane (GE Healthcare). Membranes were allowed to dry for 30 mins and then probed for TcdA and TcdB using primary IgG monoclonal antibodies (mAbs) (Meridian Biosciences) for TcdA (C70517M) and TcdB (C70888M, C65426M) at 1:3000 and 1:1000 dilutions in TBS with 0.1% tween-20 (TBST) and 3% skimmed milk (Oxoid), respectively. A goat anti-mouse IgG-Horse radish peroxidise (HRP) conjugated 2^{o} antibody (1:3000 dilution in TBST and 3% Skimmed milk) was used for chemiluminescent detection according to the ECL Plus Western Blotting Detection Reagents guidelines (GE Healthcare). Alternatively goat anti-mouse IR Dye 800CW 2^{o} antibodies (LI-COR, Germany) (1:10000 dilution in TBST and 3% Skimmed milk) were used, and toxin was detected using infra-red emission spectra on the Odyssey Clx system (LI-COR, Germany). Images were visualised with Image Studio version 2.1 software available with the Odyssey system.

Membranes were blocked overnight at 4^{o}C in TBST with 5% skimmed milk. After the overnight blocking step, membranes were washed twice for 15 mins in TBST and once for 15 mins in TBS. Membranes were then incubated in 1^{o} antibody for 1hr at room temperature on an orbital shaker. Membranes were washed twice for 15 mins in TBST and once for 15mins in TBS then incubated in the 2^{o} antibody for 1 hr at room temperature on an orbital shaker. Membranes were washed five times in TBST for 10 mins. A 1:1 mixture of ECL (Thermo Scientific Pierce ECL) western blotting solution A and B was prepared on saran wrap and the membrane placed protein side down onto the mix. Membranes were incubated for 3 mins. Excess ECL reagent was drained from the membrane and it was placed between clear plastic

covers for development. Toxins were detected using a BioRad ChemiDoc transluminator system, for 10 mins with 120 second exposure intervals. Toxins were then semi-quantified using density plots performed by Quantity One™ software (BioRad). Alternatively membranes were placed directly onto the Odyssey scanning system (LI-COR, Germany) and visualised with the appropriate software as detailed above.

10.13 Genomic DNA extraction

Genomic DNA was extracted from a lawn of *C. difficile* grown anaerobically overnight on BHI agar plates according to the Wizard Genomic DNA Purification Kit (Promega). One lawn of *C. difficile* was re-suspended with a sterile swab in 1 ml PBS. Samples were centrifuged at 13000 rpm for 2 mins, the supernatant was discarded and the pellet re-suspended in 480 µl EDTA (50 mM). 120 µl lysozyme (10 mg/ml) was added to the re-suspended pellet and incubated at 37^{o}C for 1 hr. Samples were centrifuged for 2 mins at 13000 rpm and the supernatant removed. Pellets were then re-suspended in 600 µl nucleic acid lysis solution and incubated at 80^{o}C for 5 mins. Samples were cooled to room temperature and 3 µl RNase was added to the lysate. Samples were inverted 5 times and incubated at 37^{o}C for 1 hr. Samples were cooled to room temperature, 200 µl protein precipitation solution was added, and samples were vortexed at maximum speed for 20 seconds. Samples were incubated on ice for 5 min and then centrifuged at 13000 rpm for 3 mins. Sample supernatant was transferred to a clean 1.5 ml eppendorf and 600 µl room temperature isopropanol was added. Samples were gently inverted until thread like strands were visible. Samples were centrifuged at 13000 rpm for 2 mins, supernatant carefully drained on absorbent towel and 600 µl ethanol (70%) was added. Samples were inverted gently until the pellet was floating and then centrifuged at 13000 rpm for 2 mins. All ethanol was carefully aspirated and excess drained on absorbent towel. Samples were air dried for 10-

15 mins then rehydrated by adding 100 µl rehydration solution. Samples were incubated at 65^{o}C for 1 hr and periodically mixed by gentle agitation.

10.14 Whole Genome Sequencing

Whole genome sequencing was performed by the Wellcome Trust Sanger Institute using, Illumina GA2 Genetic analysers (Illumina Inc) and the HiSeq 2000 at the University of Exeter. Briefly; paired end reads (76 bp and 100 bp) were first filtered to remove adapter sequences using fastq-mcf sequence quality filter, clipping and processor available freely from http://code.google.com/p/ea-utils/. Filtered reads were remapped to an appropriate reference genome sequence (Table 10.3) using Bowtie 0.12.7[134]. SNPs were called using the SAMtools (0.18) utilities[141], and visualised using Variant Call Format (VCF) 4.1[2] and Artemis (13.2.0)[194]. The VCF SNP files were further filtered to reduce the number of low quality calls based on a cut-off value of 60 in the VCF filter column of the text file. All SNP sites called were based on homozygous changes (depicted by 1/1 in the Variant Call Format (VCF) file) with a depth of coverage of ≥ 6 reads. A custom Perl script from Dr David Studholme was used to determine location and codon change (if any) for the SNPs indicated in the VCF file. Reference sequence files for 023, 014 and 015 ribotypes were kindly given by the Wellcome Trust Sanger Centre (Table 10.4), but were not fully annotated. Automated annotation of 023, 014, and 015 reference sequences generated by (Rapid Annotation using Subsystem Technology (RAST)[16], were used for Mauve alignments where published genomes were not available.

Whole genome alignments were displayed in Mauve multiple genome alignment software[54] by generating consensus sequences for the assembled genomes using SAMtools. Unmapped reads were filtered and combined using a Perl script designed by Dr Konrad Paszkiewicz and assembled *de novo* using Velvet 1.0.18[233] using a Kmer length optimisation step performed

by a VelvetOptimiser Perl script (Velvet 1.0.18). Contiguous sequences were annotated using RAST[16] to determine putative gene identities.

Detailed commands and Perl scripts used for genome analysis, genome assembly statistics, output of SNP analysis and velvet assembly of unmapped sequence reads are presented in Appendix Two.

References

1. **Department of Health and Health Protection Agency**: *Clostridium difficile* infection: How to deal with the problem. 2009.
2. **The 1000 Genomes Project Consortium**: A map of human genome variation from population-scale sequencing. 2010. *http://dx.doi.org/10.1038/nature09534*
3. **Aadland E., Fagerhol M. K.**: Faecal calprotectin: a marker of inflammation throughout the intestinal tract. *European Journal of Gastroenterology & Hepatology* 2002, **14**(8):823-825.
4. **Adams S. T., Leveson S. H.**: Clinical prediction rules. *BMJ* 2012, **344**:doi: 10.1136/bmj.d8312
5. **Department of Health and the Health Protection Agency**: *Clostridium difficile* infection: How to deal with the problem. 2008. *http://www.hpa.org.uk/web/HPAwebFile/HPAweb_C/1232006607827*
6. **Department of Health and the Health Protection Agency**: *Clostridium difficile* Ribotyping Network (CDRN) for England and Northern Ireland 2009/10 2011. *Report.http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1296681529717*
7. **Department of Health and the Health Protection Agency**: *Clostridium difficile* Ribotyping Network (CDRN) for England and Northern Ireland 2010/11 Annual report. 2012. *Report.http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317133396963*
8. **Akerlund T., Persson I., Unemo M., Noren T., Svenungsson B., Wullt M., Burman L.**: Increased sporulation rate of epidemic *Clostridium difficile* Type 027/NAP1. *J Clin Microbiol* 2008, **46**:1530 - 1533.
9. **Ãkerlund T., Svenungsson B., Lagergren Ã., Burman L. G.**: Correlation of Disease Severity with Fecal Toxin Levels in Patients with *Clostridium difficile*-Associated Diarrhea and Distribution of PCR Ribotypes and Toxin Yields *In Vitro* of Corresponding Isolates. *Journal of Clinical Microbiology* 2006, **44**(2):353-358.
10. **Al-Nassir W. N., Sethi A. K., Li Y., Pultz M. J., Riggs M. M., Donskey C. J.**: Both Oral Metronidazole and Oral Vancomycin Promote Persistent Overgrowth of Vancomycin-Resistant *Enterococci* during Treatment of *Clostridium difficile*-Associated Disease. *Antimicrobial Agents and Chemotherapy* 2008, **52**(7):2403-2406.
11. **Al-Tureihi F. I. J., Hassoun A., Wolf-Klein G., Isenberg H.**: Albumin, length of stay, and proton pump inhibitors: Key factors in *Clostridium difficile*-associated disease in nursing home patients. *Journal of the American Medical Directors Association* 2005, **6**(2):105-108.
12. **Allen-Vercoe E., Woodward M. J.**: The Role of Flagella, But not Fimbriae, in the Adherence of *Salmonella Enterica* Serotype *Enteritidis* to Chick Gut Explant. *Journal of Medical Microbiology* 1999, **48**(8):771-780.
13. **Alonso R., Peláez T., González-Abad M. J., Alcalá L., Muñoz P., Rodríguez-Créixems M., Bouza E.**: *In vitro* activity of new quinolones against *Clostridium difficile*. *Journal of Antimicrobial Chemotherapy* 2001, **47**(2):195-197.
14. **Anthony D. M., Reynolds T., Paton J., Rafter L.**: Serum albumin in risk assessment for *Clostridium difficile*. *Journal of Hospital Infection* 2009, **71**(4):378-379.
15. **Aubry A., Hussack G., Chen W., KuoLee R., Twine S. M., Fulton K. M., Foote S., Carrillo C. D., Tanha J., Logan S. M.**: Modulation of toxin production by the flagellar regulon in *Clostridium difficile*. *Infection and Immunity* 2012.
16. **Aziz R., Bartels D., Best A., DeJongh M., Disz T., Edwards R., Formsma K., Gerdes S., Glass E., Kubal M., Meyer F., Olsen G., Olson R., Osterman A.,**

Overbeek R., McNeil L., Paarmann D., Paczian T., Parrello B., Pusch G., Reich C., Stevens R., Vassieva O., Vonstein V., Wilke A., Zagnitko O.: The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 2008, **9**(1):75.

17. **Babcock G. J., Broering T. J., Hernandez H. J., Mandell R. B., Donahue K., Boatright N., Stack A. M., Lowy I., Graziano R., Molrine D., Ambrosino D. M., Thomas W. D.**: Human Monoclonal Antibodies Directed against Toxins A and B Prevent *Clostridium difficile*-Induced Mortality in Hamsters. *Infection and Immunity* 2006, **74**(11):6339-6347.
18. **Baines S. D., O'Connor R., Freeman J., Fawley W. N., Harmanus C., Mastrantonio P., Kuijper E. J., Wilcox M. H.**: Emergence of reduced susceptibility to metronidazole in *Clostridium difficile*. *J Antimicrob Chemother* 2008:dkn313.
19. **Barák I., Behari J., Olmedo G., Guzmán P., Brown D. P., Castro E., Walker D., Westpheling J., Youngman P.**: Structure and function of the *Bacillus* SpoIIE protein and its localization to sites of sporulation septum assembly. *Molecular Microbiology* 1996, **19**(5):1047-1060.
20. **Barbut F., Decre D., Burghoffer B., Lesage D., Delisle F., Lalande V., Delmee M., Avesani V., Sano N., Coudert C., Petit J.-C.**: Antimicrobial Susceptibilities and Serogroups of Clinical Strains of *Clostridium difficile* Isolated in France in 1991 and 1997. *Antimicrobial Agents and Chemotherapy* 1999, **43**(11):2607-2611.
21. **Barbut F., Richard A., Hamadi K., Chomette V., Burghoffer B., Petit J.-C.**: Epidemiology of Recurrences or Reinfections of *Clostridium difficile*-Associated Diarrhea. *J Clin Microbiol* 2000, **38**(6):2386-2388.
22. **Barbut F., Rupnik M.**: 027, 078 and others: going beyond the numbers (and away from the hypervirulence). *Clinical Infectious Diseases* 2012.
23. **Barroso L. A., Wang S. Z., Phelps C. J., Johnson J. L., Wilkins T. D.**: Nucleotide sequence of *Clostridium difficile* toxin B gene. *Nucleic acids research* 1990, **18**(13):4004.
24. **Bartlett J. G.**: Antibiotic-Associated Pseudomembranous Colitis. *Reviews of Infectious Diseases* 1979, **1**(3):530-539.
25. **Bartlett J. G.**: *Clostridium difficile*: History of Its Role as an Enteric Pathogen and the Current State of Knowledge about the Organism. *Clinical Infectious Diseases* 1994, **18**:265-272.
26. **Bartlett J. G.**: *Clostridium difficile*: progress and challenges. *Annals of the New York Academy of Sciences* 2010, **1213**(1):62-69.
27. **Bartlettt J. G., Onderdonk, A.B., Cisneros, R.L. and Kasper, D.L.**: Clindamycin-associated colitis due to a toxin-producing species of *Clostridium* in hamsters. *J Infect Dis* 1977, **136**:701-705.
28. **Bauer M. P., Notermans D. W., van Benthem B. H. B., Brazier J. S., Wilcox M. H., Rupnik M., Monnet D. L., van Dissel J. T., Kuijper E. J.**: *Clostridium difficile* infection in Europe: a hospital-based survey. *The Lancet* 2011, **377**(9759):63-73.
29. **Beres S. B., Carroll R. K., Shea P. R., Sitkiewicz I., Martinez-Gutierrez J. C., Low D. E., McGeer A., Willey B. M., Green K., Tyrrell G. J., Goldman T. D., Feldgarden M., Birren B. W., Fofanov Y., Boos J., Wheaton W. D., Honisch C., Musser J. M.**: Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proceedings of the National Academy of Sciences* 2011, **107**(9):4371-4376.
30. **Bhangu S., Bhangu A., Nightingale P., Michael A.**: Mortality and risk stratification in patients with *Clostridium difficile*-associated diarrhoea. *Colorectal Disease* 2010, **12**(3):241-246.

31. **Blake J. E., Mitsikosta F., Metcalfe M. A.**: Immunological detection and cytotoxic properties of toxins from toxin A-positive, toxin B-positive *Clostridium difficile* variants. *Journal of Medical Microbiology* 2004, **53**(3):197-205.
32. **Bloomfield M. G., Sherwin J. C., Gkrania-Klotsas E.**: Risk factors for mortality in *Clostridium difficile* infection in the general hospital population: a systematic review. *Journal of Hospital Infection* 2012, **82**(1):1-12.
33. **Bouvet P. J., Popoff M. R.**: Genetic relatedness of *Clostridium difficile* isolates from various origins determined by triple-locus sequence analysis based on toxin regulatory genes tcdC, tcdR, and cdtR. *J Clin Microbiol* 2008, **46**:3703 - 3713.
34. **Brazier J.**: Antimicrobial susceptibility of polymerase chain reaction ribotypes of *Clostridium difficile* commonly isolated from symptomatic hospital patients in the UK. *J Hosp Infect* 2005, **61**:11 - 14.
35. **Brazier J. S.**: *Clostridium difficile*: The anaerobe that made the grade. *Anaerobe* 2012, **18**(2):197-199.
36. **Brouwer M. S. M., Roberts A. P., Mullany P., Allan E.**: *In silico* analysis of sequenced strains of *Clostridium difficile* reveals a related set of conjugative transposons carrying a variety of accessory genes. *Mobile Genetic Elements* 2012, **2**(1):8-12.
37. **Brouwer M. S. M., Warburton P. J., Roberts A. P., Mullany P., Allan E.**: Genetic Organisation, Mobility and Predicted Functions of Genes on Integrated, Mobile Genetic Elements in Sequenced Strains of *Clostridium difficile*. *PLoS One* 2011, **6**(8):e23014.
38. **Burns D. A., Heap J. T., Minton N. P.**: The diverse sporulation characteristics of *Clostridium difficile* clinical isolates are not associated with type. *Anaerobe* 2010, **In Press, Uncorrected Proof**.
39. **Burns D. A., Heeg D., Cartman S. T., Minton N. P.**: Reconsidering the Sporulation Characteristics of Hypervirulent *Clostridium difficile* BI/NAP1/027. *PLoS One* 2011, **6**(9):e24894.
40. **Burns D. A., Minton N. P.**: Sporulation studies in *Clostridium difficile*. *Journal of Microbiological Methods* 2011(0).
41. **Busch C., Hofmann F., Selzer J. r., Munro S., Jeckel D., Aktories K.**: A Common Motif of Eukaryotic Glycosyltransferases Is Essential for the Enzyme Activity of Large Clostridial Cytotoxins. *Journal of Biological Chemistry* 1998, **273**(31):19566-19572.
42. **Cairns M. D., Stabler R. A., Shetty N., Wren B. W.**: The continually evolving *Clostridium difficile* species. *Future Microbiology* 2012, **7**(8):945-957.
43. **Carlier J. P., Sellier N., Rager M. N., Reysset G.**: Metabolism of a 5-nitroimidazole in susceptible and resistant isogenic strains of *Bacteroides fragilis*. *Antimicrobial Agents and Chemotherapy* 1997, **41**(7):1495-1499.
44. **Carniol K., Ben-Yehuda S., King N., Losick R.**: Genetic Dissection of the Sporulation Protein SpoIIE and Its Role in Asymmetric Division in *Bacillus subtilis*. *Journal of Bacteriology* 2005, **187**(10):3511-3520.
45. **Chakra A., Nour A., Pepin J., Valiquette L.**: Prediction Tools for Unfavourable Outcomes in *Clostridium difficile* Infection: A Systematic Review. *PLoS One* 2012, **7**(1):e30258.
46. **Cheknis A. K., Sambol S. P., Davidson D. M., Nagaro K. J., Mancini M. C., Aida Hidalgo-Arroyo G., Brazier J. S., Johnson S., Gerding D. N.**: Distribution of *Clostridium difficile* strains from a North American, European and Australian trial of treatment for *C. difficile* infections: 2005-2007. *Anaerobe* 2009, **15**(6):230-233.

47. **Chilton C. H., Freeman J., Crowther G. S., Todhunter S. L., Nicholson S., Wilcox M. H.**: Co-amoxiclav induces proliferation and cytotoxin production of *Clostridium difficile* ribotype 027 in a human gut model. *Journal of Antimicrobial Chemotherapy* 2012.
48. **Choi H. K., Kim K. H., Lee S. H., Lee S. J.**: Risk Factors for Recurrence of *Clostridium difficile* Infection: Effect of Vancomycin-resistant Enterococci Colonization. *J Korean Med Sci* 2011, **26**(7):859-864.
49. **Clyne B., Olshaker J. S.**: The C-reactive protein. *The Journal of Emergency Medicine* 1998, **17**(6):1019-1025.
50. **Cormican M. G., Erwin M. E., Jones R. N.**: False resistance to metronidazole by E-test among anaerobic bacteria investigations of contributing test conditions and medium quality. *Diagnostic Microbiology and Infectious Disease* 1996, **24**(2):117-119.
51. **Costa F., Mumolo M. G., Bellini M., Romano M. R., Ceccarelli L., Arpe P., Sterpi C., Marchi S., Maltinti G.**: Role of faecal calprotectin as non-invasive marker of intestinal inflammation. *Digestive and Liver Disease* 2003, **35**(9):642-647.
52. **Costa F., Mumolo M. G., Ceccarelli L., Bellini M., Romano M. R., Sterpi C., Ricchiuti A., Marchi S., Bottai M.**: Calprotectin is a stronger predictive marker of relapse in ulcerative colitis than in Crohns disease. *Gut* 2005, **54**(3):364-368.
53. **Curry S. R., Marsh J. W., Muto C. A., O'Leary M. M., Pasculle A. W., Harrison L. H.**: *tcdC* Genotypes Associated with Severe TcdC Truncation in an Epidemic Clone and Other Strains of *Clostridium difficile*. *Journal of Clinical Microbiology* 2007, **45**(1):215-221.
54. **Darling A. C. E., Mau B., Blattner F. R., Perna N. T.**: Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research* 2004, **14**(7):1394-1403.
55. **Dawson L. F., Valiente E., Wren B. W.**: *Clostridium difficile*-A continually evolving and problematic pathogen. *Infection, Genetics and Evolution* 2009, **9**(6):1410-1417.
56. **Delevaux I., Andre M., Colombier M., Albuisson E., Meylheuc F., Begue R.-J., Piette J.-C., Aumaitre O.**: Can procalcitonin measurement help in differentiating between bacterial infection and other kinds of inflammatory processes? *Annals of the Rheumatic Diseases* 2003, **62**(4):337-340.
57. **Depardieu F., Bonora M. G., Reynolds P. E., Courvalin P.**: The *vanG* glycopeptide resistance operon from *Enterococcus faecalis* revisited. *Molecular Microbiology* 2003, **50**(3):931-948.
58. **Dingle K. E., Griffiths D., Didelot X., Evans J., Vaughan A., Kachrimanidou M., Stoesser N., Jolley K. A., Golubchik T., Harding R. M., Peto T. E., Fawley W., Walker A. S., Wilcox M., Crook D. W.**: Clinical *Clostridium difficile*: Clonality and Pathogenicity Locus Diversity. *PLoS One* 2011, **6**(5):e19993.
59. **Dingle T. C., Mulvey G. L., Armstrong G. D.**: Mutagenic Analysis of the *Clostridium difficile* Flagellar Proteins, FliC and FliD, and Their Contribution to Virulence in Hamsters. *Infection and Immunity* 2011, **79**(10):4061-4067.
60. **Drudy D., Kyne L., O'Mahony R., Fanning S.**: gyrA mutations in fluoroquinolone-resistant *Clostridium difficile* PCR-027. *Emerg Infect Dis* 2007, **13**:504 - 505.
61. **Drudy D., Quinn T., O'Mahony R., Kyne L., Ã"'Gaora P., Fanning S. a.**: High-level resistance to moxifloxacin and gatifloxacin associated with a novel mutation in gyrB in toxin-A-negative, toxin-B-positive *Clostridium difficile*. *Journal of Antimicrobial Chemotherapy* 2006, **58**(6):1264-1267.

62. **Drummond L. J., McCoubrey J., Smith D. G. E., Starr J. M., Poxton I. R.**: Changes in sensitivity patterns to selected antibiotics in *Clostridium difficile* in geriatric in-patients over an 18-month period. *Journal of Medical Microbiology* 2003, **52**(3):259-263.
63. **Dubberke E. R., Olsen M. A.**: Burden of *Clostridium difficile* on the Healthcare System. *Clinical Infectious Diseases* 2012, **55**(suppl 2):S88-S92.
64. **Dupuy B., Govind R., Antunes A., Matamouros S.**: *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. *J Med Microbiol* 2008, **57**:685 - 689.
65. **Eastwood K., Else P., Charlett A., Wilcox M.**: Comparison of Nine Commercially Available Clostridium difficile Toxin Detection Assays, a Real-Time PCR Assay for *C. difficile tcdB*, and a Glutamate Dehydrogenase Detection Assay to Cytotoxin Testing and Cytotoxigenic Culture Methods. *Journal of Clinical Microbiology* 2009, **47**(10):3211-3217.
66. **Elston J. W. T., Barlow G. D.**: Community-associated MRSA in the United Kingdom. *Journal of Infection* 2009, **59**(3):149-155.
67. **Eyre D. W., Walker A. S., Wyllie D., Dingle K. E., Griffiths D., Finney J., O'Connor L., Vaughan A., Crook D. W., Wilcox M. H., Peto T. E. A.**: Predictors of First Recurrence of *Clostridium difficile* Infection: Implications for Initial Management. *Clinical Infectious Diseases* 2012, **55**(suppl 2):S77-S87.
68. **Fagerhol M. K.**: Calprotectin, a faecal marker of organic gastrointestinal abnormality. *The Lancet* 2000, **356**(9244):1783-1784.
69. **Falkow S.**: Molecular Koch's Postulates Applied to Microbial Pathogenicity. *Reviews of Infectious Diseases* 1988, **10**:S274-S276.
70. **Finney J. M. T.**: Gastro-enterostomy for cicatrizing ul- cer of the pylorus. *Bull Johns Hopkins Hosp* 1893, **4**:53-55.
71. **Forgetta V., Oughton M. T., Marquis P., Brukner I., Blanchette R., Haub K., Magrini V., Mardis E. R., Gerding D. N., Loo V. G., Miller M. A., Mulvey M. R., Rupnik M., Dascal A., Dewar K.**: Fourteen-Genome Comparison Identifies DNA Markers for Severe-Disease-Associated Strains of *Clostridium difficile*. *Journal of Clinical Microbiology* 2011, **49**(6):2230-2238.
72. **Freeman J., Baines S. D., Saxton K., Wilcox M. H.**: Effect of metronidazole on growth and toxin production by epidemic *Clostridium difficile* PCR ribotypes 001 and 027 in a human gut model. *Journal of Antimicrobial Chemotherapy* 2007, **60**(1):83-91.
73. **Freeman J., Bauer M. P., Baines S. D., Corver J., Fawley W. N., Goorhuis B., Kuijper E. J., Wilcox M. H.**: The Changing Epidemiology of *Clostridium difficile I*nfections. *Clin Microbiol Rev* 2010, **23**(3):529-549.
74. **Freeman J., Fawley W., Baines S., Wilcox M.**: Measurement of toxin production by *Clostridium difficile*. *The Lancet* 2006, **367**(9515):982-983.
75. **Garey K. W., Dao-Tran T. K., Jiang Z. D., Price M. P., Gentry L. O., DuPont H. L.**: A clinical risk index for *Clostridium difficile* infection in hospitalised patients receiving broad-spectrum antibiotics. *Journal of Hospital Infection* 2008, **70**(2):142-147.
76. **Garey Kevin W., Jiang Z.-D., Ghantoji S., Tam Vincent H., Arora V., DuPont Herbert L.**: A Common Polymorphism in the Interleukin Il-8 Gene Promoter Is Associated with an Increased Risk for Recurrent *Clostridium difficile* Infection. *Clinical Infectious Diseases* 2010, **51**(12):1406-1410.
77. **Garey K. W., Sethi S., Yadav Y., DuPont H. L.**: Meta-analysis to assess risk factors for recurrent *Clostridium difficile* infection. *Journal of Hospital Infection* 2008, **70**(4):298-304.

78. **Gerding D. N.**: Clindamycin, Cephalosporins, Fluoroquinolones, and *Clostridium difficile*-Associated Diarrhea: This Is an Antimicrobial Resistance Problem. *Clinical Infectious Diseases* 2004, **38**(5):646-648.
79. **Geric B., Carman Robert J., Rupnik M., Genheimer C. W., Sambol Susan P., Lyerly David M., Gerding Dale N., Johnson S.**: Binary Toxin Producing, Large Clostridial Toxin-Negative *Clostridium difficile* Strains Are Enterotoxic but Do Not Cause Disease in Hamsters. *The Journal of Infectious Diseases* 2006, **193**(8):1143-1150.
80. **Gisbert J. P., Bermejo F., Perez-Calle J. L., Taxonera C., Vera I., McNicholl A. G., Algaba A., Lopez P., Lopez-Palacios N., Calvo M., Gonzalez-Lama Y., Carneros J. A., Velasco M., Mate J.**: Fecal calprotectin and lactoferrin for the prediction of inflammatory bowel disease relapse. *Inflamm Bowel Dis* 2009, **15**(8):1190-1198.
81. **Gisbert J. P., McNicholl A. G.**: Questions and answers on the role of faecal calprotectin as a biological marker in inflammatory bowel disease. *Digestive and Liver Disease* 2009, **41**(1):56-66.
82. **Goldenberg S. D., French G. L.**: Lack of association of tcdC type and binary toxin status with disease severity and outcome in toxigenic *Clostridium difficile*. *Journal of Infection* 2011, **62**(5):355-362.
83. **Goorhuis A., Bakker D., Corver J., Debast Sylvia B., Harmanus C., Notermans Daan W., Bergwerff Aldert A., Dekker Frido W., Kuijper Ed J.**: Emergence of *Clostridium difficile* Infection Due to a New Hypervirulent Strain, Polymerase Chain Reaction Ribotype 078. *Clinical Infectious Diseases* 2008, **47**(9):1162-1170.
84. **Goorhuis A., Debast S. B., Dutilh J. C., van Kinschot C. M., Harmanus C., Cannegieter S. C., Hagen E. C., Kuijper E. J.**: Type-Specific Risk Factors and Outcome in an Outbreak With 2 Different *Clostridium difficile* Types Simultaneously in 1 Hospital. *Clinical Infectious Diseases* 2011.
85. **Goorhuis A., Debast S. B., van Leengoed L. A. M. G., Harmanus C., Notermans D. W., Bergwerff A. A., Kuijper E. J.**: *Clostridium difficile* PCR Ribotype 078: an Emerging Strain in Humans and in Pigs? *Journal of Clinical Microbiology* 2008, **46**(3):1157-1158.
86. **Green R. H.**: The association of viral activation with penicillin toxicity in guinea pigs and hamsters. . *Yale J Biol Med* 1974, **47**:166-181.
87. **Grobman W. A., Stamilio D. M.**: Methods of clinical prediction. *American Journal of Obstetrics and Gynecology* 2006, **194**(3):888-894.
88. **Hafiz S.**: *Clostridium difficile* and its toxins *[dissertation]*. University of Leeds; 1974.
89. **Hall I. C., O'Toole E.**: Intestinal Flora In New-Born Infants: With A Description Of A New Pathogenic Anaerobe, Bacillus Difficilis. *Am J Dis Child* 1935, **49**(2):390-402.
90. **Hambre DM R. G., McKnee CM, MacPhillamy HB**: The toxicity of penicillin as prepared for clinical use. *Am J Med Sci* 1943, **206**:642-653.
91. **Hammond G. A., and Johnson J.L.**: The toxigenic element of *Clostridium difficile* strain VPI 10463 *Microbial Pathogenesis* 1995, **19**(4):210-213.
92. **Harshey R. M.**: Bacterial Motility on a Surface: Many Ways to a Common Goal. *Annual Review of Microbiology* 2003, **57**(1):249-273.
93. **He M., Miyajima F., Roberts P., Ellison L., Pickard D. J., Martin M. J., Connor T. R., Harris S. R., Fairley D., Bamford K. B., D'Arc S., Brazier J., Brown D., Coia J. E., Douce G., Gerding D., Kim H. J., Koh T. H., Kato H., Senoh M., Louie T., Michell S., Butt E., Peacock S. J., Brown N. M., Riley T., Songer G., Wilcox M., Pirmohamed M., Kuijper E., Hawkey P., Wren B. W., Dougan G.,**

Parkhill J., Lawley T. D.: Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat Genet* 2012, **45**:109-113.

94. **He M., Sebaihia M., Lawley T. D., Stabler R. A., Dawson L. F., Martin M. J., Holt K. E., Seth-Smith H. M. B., Quail M. A., Rance R., Brooks K., Churcher C., Harris D., Bentley S. D., Burrows C., Clark L., Corton C., Murray V., Rose G., Thurston S., van Tonder A., Walker D., Wren B. W., Dougan G., Parkhill J.**: Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proceedings of the National Academy of Sciences* 2010, **107**(16):7527-7532.
95. **Heap J. T., Pennington O. J., Cartman S. T., Carter G. P., Minton N. P.**: The ClosTron: A universal gene knock-out system for the genus Clostridium. *Journal of Microbiological Methods* 2007, **70**(3):452-464.
96. **Hecht G., C. Pothoulakis, J. T. Lamont, and J. L. Madara**: *Clostridium difficile* toxin A perturbs cytoskeletal structure and tight junction permeability of cultured human intestinal epithelial monolayers. *J Clin Investig* 1988, **56**:1053-1061.
97. **Hecht G K. A., Pothoulakis C, LaMont JT, Madara JL.**: *Clostridium difficile* toxin B disrupts the barrier function of T84 monolayers. *Gastroenterology* 1992, **102**(2):416-423.
98. **Holt K. E., Parkhill, J., Mazzoni, C.J., Roumagnac,P., Weill, F., Goodhead, I., Rance, R., Baker, S., Maskell, D.J., Wain, J., Dolecek, C., Achtman, M. and Dougan, G.**: High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. *Nat Rev Genet* 2008, **40**(8):987-993.
99. **Hornsey M., Loman N., Wareham D. W., Ellington M. J., Pallen M. J., Turton J. F., Underwood A., Gaulton T., Thomas C. P., Doumith M., Livermore D. M., Woodford N.**: Whole-genome comparison of two *Acinetobacter baumannii* isolates from a single patient, where resistance developed during tigecycline therapy. *Journal of Antimicrobial Chemotherapy* 2011, **66**(7):1499-1503.
100. **Howell M. D., Novack V., Grgurich P., Soulliard D., Novak L., Pencina M., Talmor D.**: Iatrogenic gastric acid suppression and the risk of nosocomial *clostridium difficile* infection. *Archives of Internal Medicine* 2010, **170**(9):784-790.
101. **Hu B., Xie G., Lo C.-C., Starkenburg S. R., Chain P. S. G.**: Pathogen comparative genomics in the next-generation sequencing era: genome alignments, pangenomics and metagenomics. *Briefings in Functional Genomics* 2011, **10**(6):322-333.
102. **Hu M. Y., Katchar K., Kyne L., Maroo S., Tummala S., Dreisbach V., Xu H., Leffler D. A., Kelly C. P.**: Prospective Derivation and Validation of a Clinical Prediction Rule for Recurrent *Clostridium difficile* Infection. *Gastroenterology* 2009, **136**(4):1206-1214.
103. **Huang H., Weintraub A., Fang H., Nord C. E.**: Antimicrobial resistance in *Clostridium difficile*. *International Journal of Antimicrobial Agents* 2009, **34**(6):516-522.
104. **Hundsberger. T. B. V., Weidmann., M.,Leukel P, Sauerborn M and Von Eichel-Streiber C.** : Transcription Analysis of the Genes tcdA-E of the Pathogenicity Locus of *Clostridium difficile*. *European Journal of Biochemistry* 1996, **244**(3):735-742.
105. **Isaura M., Myrna S., Enrique S., Marcelino C.**: Occluding Junctions in MDCK Cells: Modulation of Transepithelial Permeability by the Cytoskeleton. *Journal of Cellular Biochemistry* 1982, **18**(4):407-421.
106. **Janezic S., Ocepek M., Zidaric V., Rupnik M.**: *Clostridium difficile* genotypes other than ribotype 078 that are prevalent among human, animal and environmental isolates. *BMC Microbiology* 2012, **12**(1):48.
107. **Jank T., Aktories K.**: Structure and mode of action of clostridial glucosylating toxins: the ABCD model. *Trends in Microbiology* 2008, **16**(5):222-229.

108. **Jiang Z.-D., DuPont H. L., Garey K., Price M., Graham G., Okhuysen P., Dao-Tran T., LaRocco M.**: A Common Polymorphism in the Interleukin 8 Gene Promoter Is Associated with *Clostridium difficile* Diarrhea. *Am J Gastroenterol* 2006, **101**(5):1112-1116.
109. **Jiang Z. D., Garey K. W., Price M., Graham G., Okhuysen P., Daoâ€"Tran T., LaRocco M., DuPont H. L.**: Association of Interleukin-8 Polymorphism and Immunoglobulin G Anti Toxin A in Patients With *Clostridium difficile* Associated Diarrhea. *Clinical Gastroenterology and Hepatology* 2007, **5**(8):964-968.
110. **Johne B., Fagerhol M. K., Lyberg T., Prydz H., Brandtzaeg P., Naess-Andresen C. F., Dale I.**: Functional and clinical aspects of the myelomonocyte protein calprotectin. *Mol Pathol* 1997, **50**(3):113-123.
111. **Johnson S.**: Recurrent *Clostridium difficile* infection: A review of risk factors, treatments, and outcomes. *Journal of Infection* 2009, **58**(6):403-410.
112. **Johnson S.**: Recurrent *Clostridium difficile* infection: causality and therapeutic approaches. *International Journal of Antimicrobial Agents* 2009, **33**(Supplement 1):S33-S36.
113. **Johnson S., Adelmann A., Clabots C. R., Peterson L. R., Gerding D. N.**: Recurrences of *Clostridium difficile* Diarrhea Not Caused by the Original Infecting Organism. *The Journal of Infectious Diseases* 1989, **159**(2):340-343.
114. **Jones A. M., Kuijper E. J., Wilcox M. H.**: *Clostridium difficile:* A European perspective. *Journal of Infection* 2012, **66**(2):115-128.
115. **Josenhans C., Suerbaum S.**: The role of motility as a virulence factor in bacteria. *International Journal of Medical Microbiology* 2002, **291**(8):605-614.
116. **Just I., Dreger S. C., Genth H.**: Rho-inhibiting toxins. *Toxicology Letters* 2009, **189**(Supplement 1):S18-S19.
117. **Just I., Hofmann F., Genth H., Gerhard R.**: Bacterial protein toxins inhibiting low-molecular-mass GTP-binding proteins. *International Journal of Medical Microbiology* 2001, **291**(4):243-250.
118. **Just I., Wilm M., Selzer J. r., Rex G., von Eichel-Streiber C., Mann M., Aktories K.**: The Enterotoxin from *Clostridium difficile* (ToxA) Monoglucosylates the Rho Proteins. *Journal of Biological Chemistry* 1995, **270**(23):13932-13936.
119. **Kamiya S., Ogura H., Meng X. Q., Nakamura S.**: Correlation between cytotoxin production and sporulation in *Clostridium difficile*. *Journal of Medical Microbiology* 1992, **37**(3):206-210.
120. **Karas J. A., Enoch D. A., Aliyu S. H.**: A review of mortality due to *Clostridium difficile* infection. *Journal of Infection* 2010, **61**(1):1-8.
121. **Katchar K., Taylor C. P., Tummala S., Chen X., Sheikh J., Kelly C. P.**: Association Between IgG2 and IgG3 Subclass Responses to Toxin A and Recurrent *Clostridium difficile*-Associated Disease. *Clinical Gastroenterology and Hepatology* 2007, **5**(6):707-713.
122. **Kelly C. P., Becker, S., Linevsky,J.K., Joshi, M.A., Conor O'Keane, J., Dickey, Thomas LaMont, J., and Pothoulakis, C.**: Neutrophil Recruitment in *Clostridium dfflicile* Toxin A Enteritis in the Rabbit. *J Clin Invest* 1994, **93**:1257-1265.
123. **Kelly C. P., Kyne L.**: The host immune response to *Clostridium difficile*. *Journal of Medical Microbiology* 2011, **60**(8):1070-1079.
124. **Kennemann L., Didelot X., Aebischer T., Kuhn S., Drescher B., Droege M., Reinhardt R., Correa P., Meyer T. F., Josenhans C., Falush D., Suerbaum S.**: *Helicobacter pylori* genome evolution during human infection. *Proceedings of the National Academy of Sciences* 2011, **108**(12):5033-5038.

125. **Khanna S., Pardi D. S., Aronson S. L., Kammer P. P., Orenstein R., St Sauver J. L., Harmsen W. S., Zinsmeister A. R.**: The Epidemiology of Community-Acquired *Clostridium difficile* Infection: A Population-Based Study. *Am J Gastroenterol* 2012, **107**(1):89-95.
126. **Kink J. A., Williams J. A.**: Antibodies to Recombinant *Clostridium difficile* Toxins A and B Are an Effective Treatment and Prevent Relapse of *C. difficile*-Associated Disease in a Hamster Model of Infection. *Infection and Immunity* 1998, **66**(5):2018-2025.
127. **Klaassen C. H. W., van Haren H. A., Horrevorts A. M.**: Molecular Fingerprinting of *Clostridium difficile* Isolates: Pulsed-Field Gel Electrophoresis versus Amplified Fragment Length Polymorphism. *Journal of Clinical Microbiology* 2002, **40**(1):101-104.
128. **Knetsch C. W., Hensgens M. P. M., Harmanus C. l., van der Bijl M. W., Savelkoul P. H. M., Kuijper E. J., Corver J., van Leeuwen H. C.**: Genetic markers for *Clostridium difficile* lineages linked to hypervirulence. *Microbiology* 2011, **157**(11):3113-3123.
129. **Kok L., Elias S. G., Witteman B. J. M., Goedhard J. G., Muris J. W. M., Moons K. G. M., de Wit N. J.**: Diagnostic Accuracy of Point-of-Care Fecal Calprotectin and Immunochemical Occult Blood Tests for Diagnosis of Organic Bowel Disease in Primary Care: The Cost-Effectiveness of a Decision Rule for Abdominal Complaints in Primary Care (CEDAR) Study. *Clinical Chemistry* 2012, **58**(6):989-998.
130. **Kuehne S. A., Cartman S. T., Heap J. T., Kelly M. L., Cockayne A., Minton N. P.**: The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* 2010, **467**(7316):711-713.
131. **Kuijper E. J., Monnet D. L.**: The spread of hypervirulent *Clostridium difficile* PCR Ribotype 027 in Europe. *International Journal of Infectious Diseases* 2008, **12**(Supplement 1):e22-e22.
132. **Kyne L., Warny M., Qamar A., Kelly C. P.**: Association between antibody response to toxin A and protection against recurrent *Clostridium difficile* diarrhoea. *The Lancet* 2001, **357**(9251):189-193.
133. **Lancaster J. W., Matthews S. J.**: Fidaxomicin: The Newest Addition to the Armamentarium Against *Clostridium difficile* Infections. *Clinical Therapeutics* 2012, **34**(1):1-13.
134. **Langmead B., Trapnell C., Pop M., Salzberg S.**: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 2009, **10**(3):R25.
135. **Larson H. E., Parry J. V., Price A. B., Davies D. R., Dolby J., Tyrrell D. A.**: Undescribed toxin in pseudomembranous colitis. *Br Med J* 1977, **1**(6071):1246-1248.
136. **Lawley T. D., Croucher N. J., Yu L., Clare S., Sebaihia M., Goulding D., Pickard D. J., Parkhill J., Choudhary J., Dougan G.**: Proteomic and Genomic Characterization of Highly Infectious *Clostridium difficile* 630 Spores. *J Bacteriol* 2009, **191**(17):5377-5386.
137. **Lee P. S., Lin D. C.-H., Moriya S., Grossman A. D.**: Effects of the Chromosome Partitioning Protein Spo0J (ParB) on *oriC* Positioning and Replication Initiation in *Bacillus subtilis*. *Journal of Bacteriology* 2003, **185**(4):1326-1337.
138. **Lee W. G., Huh J. Y., Cho S. R., Lim Y. A.**: Reduction in Glycopeptide Resistance in Vancomycin-Resistant *Enterococci* as a Result of *vanA* Cluster Rearrangements. *Antimicrobial Agents and Chemotherapy* 2004, **48**(4):1379-1381.

139. **Lemee L., Dhalluin A., Pestel-Caron M., Lemeland J.-F., Pons J.-L.**: Multilocus Sequence Typing Analysis of Human and Animal *Clostridium difficile* Isolates of Various Toxigenic Types. *J Clin Microbiol* 2004, **42**(6):2609-2617.
140. **Leonard J., Marshall J. K., Moayyedi P.**: Systematic Review of the Risk of Enteric Infection in Patients Taking Acid Suppression. *Am J Gastroenterol* 2007, **102**(9):2047-2056.
141. **Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., Subgroup G. P. D. P.**: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, **25**(16):2078-2079.
142. **Lieberman T. D., Michel J.-B., Aingaran M., Potter-Bynoe G., Roux D., Davis M. R., Skurnik D., Leiby N., LiPuma J. J., Goldberg J. B., McAdam A. J., Priebe G. P., Kishony R.**: Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet* 2012, **43**(12):1275-1280.
143. **Liebowitz L. D.**: MRSA burden and interventions. *International Journal of Antimicrobial Agents* 2009, **34**(Supplement 3):S11-S13.
144. **Lillehoj E. P., Kim B. T., Kim K. C.**: Identification of *Pseudomonas aeruginosa* flagellin as an adhesin for Muc1 mucin. *American Journal of Physiology - Lung Cellular and Molecular Physiology* 2002, **282**(4):L751-L756.
145. **Lim W. S., van der Eerden M. M., Laing R., Boersma W. G., Karalus N., Town G. I., Lewis S. A., Macfarlane J. T.**: Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003, **58**(5):377-382.
146. **Linsky A., Gupta K., Lawler E. V., Fonda J. R., Hermos J. A.**: Proton Pump Inhibitors and Risk for Recurrent *Clostridium difficile* Infection. *Arch Intern Med* 2010, **170**(9):772-778.
147. **Löfmark S., Fang H., Hedberg M., Edlund C.**: Inducible Metronidazole Resistance and *nim* Genes in Clinical *Bacteroides fragilis* Group Isolates. *Antimicrobial Agents and Chemotherapy* 2005, **49**(3):1253-1256.
148. **Loo V. G., Poirier L., Miller M. A., Oughton M., Libman M. D., Michaud S., Bourgault A. M., Nguyen T., Frenette C., Kelly M., Vibien A., Brassard P., Fenn S., Dewar K., Hudson T. J., Horn R., Rene P., Monczak Y., Dascal A.**: A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *N Engl J Med* 2005, **353**:2442 - 2449.
149. **Lowe D. O., Mamdani M. M., Kopp A., Low D. E., Juurlink D. N.**: Proton Pump Inhibitors and Hospitalization for *Clostridium Difficile*-Associated Disease: A Population-Based Study. *Clinical Infectious Diseases* 2006, **43**(10):1272-1276.
150. **Lyerly DM K. H., Wilkins TD.**: *Clostridium difficile*: its disease and toxins. *Clin Microbiol Rev* 1988, **1**(1):1-18.
151. **Lyras D., O'Connor J. R., Howarth P. M., Sambol S. P., Carter G. P., Phumoonna T., Poon R., Adams V., Vedantam G., Johnson S., Gerding D. N., Rood J. I.**: Toxin B is essential for virulence of *Clostridium difficile*. *Nature* 2009, **458**(7242):1176-1179.
152. **Malathum K., Murray B. E.**: Vancomycin-resistant *enterococci*: recent advances in genetics, epidemiology and therapeutic options. *Drug Resistance Updates* 1999, **2**(4):224-243.
153. **Mani N., and Dupuy, B.**: Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**:5844-5849.
154. **Mardis E. R.**: Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* 2008, **9**(1):387-402.

155. **Martin W. J., Gardner M., Washington J. A.**: In Vitro Antimicrobial Susceptibility of Anaerobic Bacteria Isolated from Clinical Specimens. *Antimicrobial Agents and Chemotherapy* 1972, **1**(2):148-158.
156. **Matamouros S. E., P. and Bruno Dupuy**: *Clostridium difficile* toxin expression is inhibited bythe novel regulator TcdC. *Molecular Microbiology* 2007, **64**(5):1274-1288.
157. **Mayer M. J., Narbad A., Gasson M. J.**: Molecular Characterization of a *Clostridium difficile* Bacteriophage and Its Cloned Biologically Active Endolysin. *Journal of Bacteriology* 2008, **190**(20):6734-6740.
158. **McDonald L. C., Killgore G. E., Thompson A., Owens R. C., Jr., Kazakova S. V., Sambol S. P., Johnson S., Gerding D. N.**: An Epidemic, Toxin Gene-Variant Strain of *Clostridium difficile*. *N Engl J Med* 2005, **353**(23):2433-2441.
159. **McFarland L. V., Elmer G. W., Stamm W. E., Mulligan M. E.**: Correlation of immunoblot type, enterotoxin production, and cytotoxin production with clinical manifestations of *Clostridium difficile* infection in a cohort of hospitalized patients. *Infection and Immunity* 1991, **59**(7):2456-2462.
160. **McFarland L. V., Surawicz C. M., Rubin M., Fekety R., Elmer G. W., Greenberg R. N.**: Recurrent *Clostridium difficile* Disease: Epidemiology and Clinical Characteristics. *Infection Control and Hospital Epidemiology* 1999, **20**(1):43-50.
161. **McFee R. B.**: Nosocomial or Hospital-acquired Infections: An Overview. *Disease-a-Month* 2009, **55**(7):422-438.
162. **McFee R. B., Abdelsayed G. G.**: *Clostridium difficile*. *Disease-a-Month* 2009, **55**(7):439-470.
163. **McGowan A. P., Lalayiannis L. C., Sarma J. B., Marshall B., Martin K. E., Welfare M. R.**: Thirty-day mortality of *Clostridium difficile* infection in a UK National Health Service Foundation Trust between 2002 and 2008. *Journal of Hospital Infection* 2010, **77**(1):11-15.
164. **McKessar S. J., Berry A. M., Bell J. M., Turnidge J. D., Paton J. C.**: Genetic Characterization of *vanG*, a Novel Vancomycin Resistance Locus of *Enterococcus faecalis*. *Antimicrobial Agents and Chemotherapy* 2000, **44**(11):3224-3228.
165. **Merrigan M., Venugopal A., Mallozzi M., Roxas B., Viswanathan V. K., Johnson S., Gerding D. N., Vedantam G.**: Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as robust toxin production. *J Bacteriol* 2010, **192**(19):4904-4911.
166. **Mutlu E., Wroe A. J., Sanchez-Hurtado K., Brazier J. S., Poxton I. R.**: Molecular characterization and antimicrobial susceptibility patterns of *Clostridium difficile* strains isolated from hospitals in south-east Scotland. *Journal of Medical Microbiology* 2007, **56**(7):921-929.
167. **Nerandzic M. M., Pultz M. J., Donskey C. J.**: Examination of Potential Mechanisms To Explain the Association between Proton Pump Inhibitors and *Clostridium difficile* Infection. *Antimicrobial Agents and Chemotherapy* 2009, **53**(10):4133-4137.
168. **Noblett S. E., Welfare M., Seymour K.**: The role of surgery in *Clostridium difficile* colitis. *BMJ* 2009, **338**:b1563.
169. **O'Connor J. R., Lyras, D., Farrow, K. A., Adams, V., Powell, D.R., Hinds. J., Cheung, J. K., & J. I. Rood**: Construction and analysis of chromosomal Clostridium difficile mutants. *Molecular Microbiology* 2006, **61**(5):1335-1351.
170. **O'Neill G. L., Ogunsola F. T., Brazier J. S., Duerden B. I.**: Modification of a PCR Ribotyping Method for Application as a Routine Typing Scheme for *Clostridium difficile*. *Anaerobe* 1996, **2**(4):205-209.

171. **Palmer J., Chen S., Gottlieb T., Schiog F., Gilbert G.**: False resistance to metronidazole of anaerobic bacteria using the E test. *Journal of Antimicrobial Chemotherapy* 1994, **34**(4):598-600.
172. **Papatheodorou P., Zamboglou C., Genisyuerek S., Guttenberg G., Aktories K.**: *Clostridial* glucosylating toxins enter cells via clathrin-mediated endocytosis. *PLoS ONE* 2010, **5**(5):e10673.
173. **Paredes-Sabja D., Bond C., Carman R. J., Setlow P., Sarker M. R.**: Germination of spores of *Clostridium difficile* strains, including isolates from a hospital outbreak of *Clostridium difficile*-associated disease (CDAD). *Microbiology* 2008, **154**(8):2241-2250.
174. **Pelaez T., Cercenado E., Alcala L., Marin M., Martin-Lopez A., Martinez-Alarcon J., Catalan P., Sanchez-Somolinos M., Bouza E.**: Metronidazole Resistance in *Clostridium difficile* Is Heterogeneous. *J Clin Microbiol* 2008, **46**(9):3028-3032.
175. **Pepin J., Valiquette L., Cossette B.**: Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *Canadian Medical Association Journal* 2005, **173**(9):1037-1042.
176. **Perelle S., Gibert M., Bourlioux P., Corthier G., Popoff M. R.**: Production of a complete binary toxin (actin-specific ADP- ribosyltransferase) by *Clostridium difficile* CD196. *Infect Immun* 1997, **65**(4):1402-1407.
177. **Pfeifer G., Schirmer J. r., Leemhuis J., Busch C., Meyer D. K., Aktories K., Barth H.**: Cellular Uptake of *Clostridium difficile* Toxin B. *Journal of Biological Chemistry* 2003, **278**(45):44535-44541.
178. **Planche T., Aghaizu A., Holliman R., Riley P., Poloniecki J., Breathnach A., Krishna S.**: Diagnosis of *Clostridium difficile* infection by toxin detection kits: a systematic review. *The Lancet Infectious Diseases* 2008, **8**(12):777-784.
179. **Poilane I., Cruaud P., Torlotin J. C., Collignon A.**: Comparison of the E test to the reference agar dilution method for antibiotic susceptibility testing of *Clostridium difficile*. *Clinical Microbiology and Infection* 2000, **6**(3):155-156.
180. **Popoff M. R., Rubin E. J., Gill D. M., Boquet P.**: Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infect Immun* 1988, **56**(9):2299-2306.
181. **Pothoulakis C.**: Effects of *Clostridium difficile* Toxins on Epithelial Cell Barrier. *Annals of the New York Academy of Sciences* 2000, **915**:347-356.
182. **Pothoulakis C., Sullivan, R., Melnick, D.A., Triadafilopoulos, G., Gadenne, A.S., Meshulam, T. and LaMont, J.T .** *Clostridium difficile* toxin A stimulates intracellular calcium release and chemotactic response in human granulocytes. *The Journal of Clinical Investigation* 1988, **81**(6):1741-1745.
183. **Prescott L. M., Harley J. P., Klein D. A.**: Microbiology, Fifth Edition edn: McGraw-Hill Companies Inc; 2002.
184. **Pruitt R. N., Chambers M. G., Ng K. K.-S., Ohi M. D., Lacy D. B.**: Structural organization of the functional domains of *Clostridium difficile* toxins A and B. *Proceedings of the National Academy of Sciences* 2010, **107**(30):13467-13472.
185. **Rafii F., Wynne R., Heinze T. M., Paine D. D.**: Mechanism of metronidazole-resistance by isolates of nitroreductase-producing *Enterococcus gallinarum* and *Enterococcus casseliflavus* from the human intestinal tract. *FEMS Microbiology Letters* 2003, **225**(2):195-200.
186. **Raveh D., Rabinowitz B., Breuer G. S., Rudensky B., Yinnon A. M.**: Risk factors for *Clostridium difficile* toxin-positive nosocomial diarrhoea. *International Journal of Antimicrobial Agents* 2006, **28**(3):231-237.

187. **Redelings M. D., Sorvillo, F., & Mascola, L.**: Increase in *Clostridium difficile*–related Mortality Rates, United States, 1999–2004. *Emerg Infect Dis [serial on the Internet]* 2007.
188. **Reineke J., Tenzer S., Rupnik M., Koschinski A., Hasselmayer O., Schrattenholz A., Schild H., von Eichel-Streiber C.**: Autocatalytic cleavage of *Clostridium difficile* toxin[thinsp]B. *Nature* 2007, **446**(7134):415-419.
189. **Rifkin G. D., Feketya, F.R., Silva, J., & Sackb, J.R. and R. B**: Antibiotic-Induced Colitis Implication Of A Toxin Neutralised By *Clostridium Sordellii* Antitoxin. *The Lancet* 1977, **310**(8048):1103-1106.
190. **Rosenbusch K. E., Bakker D., Kuijper E. J., Smits W. K.**: *C. difficile* 630Δerm Spo0A Regulates Sporulation, but Does Not Contribute to Toxin Production, by Direct High-Affinity Binding to Target DNA. *PLoS One* 2012, **7**(10):e48608.
191. **Roseth A. G., Aadland E., Jahnsen J., Raknerud N.**: Assessment of Disease Activity in Ulcerative Colitis by Faecal Calprotectin, a Novel Granulocyte Marker Protein. *Digestion* 1997, **58**(2):176-180.
192. **Roseth A. G., Fagerhol M. K., Aadland E., Schjonsby H.**: Assessment of the Neutrophil Dominating Protein Calprotectin in Feces: A Methodologic Study. *Scandinavian Journal of Gastroenterology* 1992, **27**(9):793-798.
193. **Rupnik M., Wilcox M. H., Gerding D. N.**: *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Micro* 2009, **7**(7):526-536.
194. **Rutherford K., Parkhill J., Crook J., Horsnell T., Rice P., Rajandream M., Barrell B.**: Artemis: sequence visualization and annotation. *Bioinformatics* 2000, **16**:944 - 945.
195. **Samuel P., Holtzman C. M., Meilman E., Sekowski I.**: Effect of Neomycin and Other Antibiotics on Serum Cholesterol Levels and on 7{alpha}-Dehydroxylation of Bile Acids by the Fecal Bacterial Flora in Man. *Circ Res* 1973, **33**(4):393-402.
196. **Schmidt M. L., Gilligan P. H.**: *Clostridium difficile* testing algorithms: What is practical and feasible? *Anaerobe* 2009, **15**(6):270-273.
197. **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., Holden M. T. G., Wright A., Churcher C., Quail M. A., Baker S., Bason N., Brooks K., Chillingworth T., Cronin A., Davis P., Dowd L., Fraser A., Feltwell T., Hance Z., Holroyd S., Jagels K., Moule S., Mungall K., Price C., Rabbinowitsch E., Sharp S., Simmonds M., Stevens K., Unwin L., Whithead S., Dupuy B., Dougan G., Barrell B., Parkhill J.**: The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet* 2006, **38**(7):779-786.
198. **Simon L., Gauvin F., Amre D. K., Saint-Louis P., Lacroix J.**: Serum Procalcitonin and C-Reactive Protein Levels as Markers of Bacterial Infection: A Systematic Review and Meta-analysis. *Clinical Infectious Diseases* 2004, **39**(2):206-217.
199. **Sirard S., Valiquette L., Fortier L.-C.**: Lack of Association between Clinical Outcome of *Clostridium difficile* Infections, Strain Type, and Virulence-Associated Phenotypes. *Journal of Clinical Microbiology* 2011, **49**(12):4040-4046.
200. **Sorg J. A., Sonenshein A. L.**: Bile Salts and Glycine as Cogerminants for *Clostridium difficile* Spores. *J Bacteriol* 2008, **190**(7):2505-2512.
201. **Spigaglia P., Barbanti F., Mastrantonio P., Brazier J. S., Barbut F., Delmée M., Kuijper E., R. Poxton I., difficile o. b. o. t. E. S. G. o. C.**: Fluoroquinolone resistance in *Clostridium difficile* isolates from a prospective study of *C. difficile* infections in Europe. *Journal of Medical Microbiology* 2008, **57**(6):784-789.
202. **Stabler R., He M., Dawson L., Martin M., Valiente E., Corton C., Lawley T., Sebaihia M., Quail M., Rose G., Gerding D., Gibert M., Popoff M., Parkhill J.,**

Dougan G., Wren B.: Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biology* 2009, **10**(9):R102.

203. **Stabler R. A., Dawson L. F., Valiente E., Cairns M. D., Martin M. J., Donahue E. H., Riley T. V., Songer J. G., Kuijper E. J., Dingle K. E., Wren B. W.**: Macro and Micro Diversity of *Clostridium difficile* Isolates from Diverse Sources and Geographical Locations. *PLoS One* 2012, **7**(3):e31559.
204. **Office for National Statistics**: Deaths involving MRSA: England and Wales, 2006 to 2010. 2011. *Bulletin.http://www.ons.gov.uk/ons/rel/subnational-health2/deaths-involving-mrsa/2006-to-2010/statistical-bulletin.html#tab-Results*
205. **Office for National Statistics.**: Deaths involving *Clostridium difficile*: England and Wales, 2006 to 2010. 2011. *Bulletin.http://www.ons.gov.uk/ons/rel/subnational-health2/deaths-involving-clostridium-difficile/2006-to-2010/statistical-bulletin.html*
206. **Stewart D., Berg A., Hegarty J.**: Predicting Recurrence of *C. difficile* Colitis Using Bacterial Virulence Factors: Binary Toxin Is the Key. *Journal of Gastrointestinal Surgery* 2012:1-8.
207. **Stubbs S., Rupnik M., Gibert M., Brazier J., Duerden B., Popoff M.**: Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile*. *FEMS Microbiology Letters* 2000, **186**(2):307-312.
208. **Sundram F., Guyot A., Carboo I., Green S., Lilaonitkul M., Scourfield A.**: *Clostridium difficile* ribotypes 027 and 106: clinical outcomes and risk factors. *Journal of Hospital Infection* 2009, **72**(2):111-118.
209. **Tacconelli E., Cataldo M. A.**: Vancomycin-resistant enterococci (VRE): transmission and control. *International Journal of Antimicrobial Agents* 2008, **31**(2):99-106.
210. **Tasteyre A., Barc M.-C., Collignon A., Boureau H., Karjalainen T.**: Role of FliC and FliD Flagellar Proteins of *Clostridium difficile* in Adherence and Gut Colonization. *Infection and Immunity* 2001, **69**(12):7937-7940.
211. **Tedesco F. J., Barton, F.W. & Alpers, J.H.**: Clindamycin-assocated cloitis. A prospective study. *Ann Intern Med* 1974, **81**:429-433.
212. **Thomas C., Stevenson M., Riley T. V.**: Antibiotics and hospital-acquired *Clostridium difficile*-associated diarrhoea: a systematic review. *Journal of Antimicrobial Chemotherapy* 2003, **51**(6):1339-1350.
213. **Toll D. B., Janssen K. J. M., Vergouwe Y., Moons K. G. M.**: Validation, updating and impact of clinical prediction rules: A review. *Journal of Clinical Epidemiology* 2008, **61**(11):1085-1094.
214. **Tøn H., Brandsnes Ø., Dale S., Holtlund J., Skuibina E., Schjønsby H., Johne B.**: Improved assay for fecal calprotectin. *Clinica Chimica Acta* 2000, **292**(1-2):41-54.
215. **Tucker K. D., Wilkins T. D.**: Toxin A of *Clostridium difficile* binds to the human carbohydrate antigens I, X, and Y. *Infect Immun* 1991, **59**(1):73-78.
216. **Twine S., Reid C., Aubry A., McMullin D., Fulton K., Austin J., Logan S.**: Motility and flagellar glycosylation in *Clostridium difficile*. *J Bacteriol* 2009, **191**:7050 - 7062.
217. **Underwood S., Guan S., Vijayasubhash V., Baines S. D., Graham L., Lewis R. J., Wilcox M. H., Stephenson K.**: Characterization of the Sporulation Initiation Pathway of *Clostridium difficile* and Its Role in Toxin Production. *J Bacteriol* 2009, **191**(23):7296-7305.
218. **Valiente E., Dawson L. F., Cairns M. D., Stabler R. A., Wren B. W.**: Emergence of new PCR ribotypes from the hypervirulent *Clostridium difficile* 027 lineage. *Journal of Medical Microbiology* 2012, **61**(1):49-56.

219. **Vohra P., Poxton I. R.**: Comparison of toxin and spore production in clinically relevant strains of *Clostridium difficile*. *Microbiology* 2011, **157**(5):1343-1353.
220. **Von Eichel- Streiber C., Suckau D., Wacher M., Hadding U.**: Cloning and Characterization of Overlapping DNA Fragments of the Toxin A Gene of *Clostridium difficile*. *J Gen Microbiol* 1989, **135**(1):55-64.
221. **von Eichel-Streiber C., Boquet P., Sauerborn M., Thelestam M.**: Large clostridial cytotoxins -a family of glycosyltransferases modifying small GTP-binding proteins. *Trends in Microbiology* 1996, **4**(10):375-382.
222. **von Eichel-Streiber C. S., M and Kuramitsu H.K.**: Evidence for a modular structure of the homologous repetitive C-terminal carbohydrate-binding sites of *Clostridium difficile* toxins and *Streptococcus mutans* glucosyltransferases. *Journal of Bacteriology* 1992, **174**(20):6707-6710.
223. **Vonberg R. P., Reichardt C., Behnke M., Schwab F., Zindler S., Gastmeier P.**: Costs of nosocomial *Clostridium difficile*-associated diarrhoea. *Journal of Hospital Infection* 2008, **70**(1):15-20.
224. **Walk S. T., Micic D., Jain R., Lo E. S., Trivedi I., Liu E. W., Almassalha L. M., Ewing S. A., Ring C., Galecki A. T., Rogers M. A. M., Washer L., Newton D. W., Malani P. N., Young V. B., Aronoff D. M.**: *Clostridium difficile* Ribotype Does Not Predict Severe Infection. *Clinical Infectious Diseases* 2012, **55**(12):1661-1668.
225. **Walker A. S., Eyre D. W., Wyllie D. H., Dingle K. E., Griffiths D., Shine B., Oakley S., O'Connor L., Finney J., Vaughan A., Crook D. W., Wilcox M. H., Peto T. E. A.**: Relationship Between Bacterial Strain Type, Host Biomarkers and Mortality in *Clostridium difficile* Infection. *Clinical Infectious Diseases* 2013.
226. **Wiegand P. N., Nathwani D., Wilcox M. H., Stephens J., Shelbaya A., Haider S.**: Clinical and economic burden of *Clostridium difficile* infection in Europe: a systematic review of healthcare-facility-acquired infection. *Journal of Hospital Infection* 2012(0).
227. **Wilcox M. H.**: Laboratory diagnosis of *Clostridium difficile* infection: in a state of transition or confusion or both? *Journal of Hospital Infection* 2011, **79**(1):1-3.
228. **Wilcox M. H., Cunniffe J. G., Trundle C., Redpath C.**: Financial burden of hospital-acquired *Clostridium difficile* infection. *Journal of Hospital Infection* 1996, **34**(1):23-30.
229. **Wilcox M. H., Shetty N., Fawley W. N., Shemko M., Coen P., Birtles A., Cairns M., Curran M. D., Dodgson K. J., Green S. M., Hardy K. J., Hawkey P. M., Magee J. G., Sails A. D., Wren M. W. D.**: Changing Epidemiology of *Clostridium difficile* Infection Following the Introduction of a National Ribotyping-Based Surveillance Scheme in England. *Clinical Infectious Diseases* 2012, **55**(8):1056-1063.
230. **Wilson K. H., Kennedy M. J., Fekety F. R.**: Use of sodium taurocholate to enhance spore recovery on a medium selective for *Clostridium difficile*. *J Clin Microbiol* 1982, **15**(3):443-446.
231. **Wilson V., Cheek L., Satta G., Walker-Bone K., Cubbon M., Citron D., Gerding D. N., Llewelyn M. J.**: Predictors of Death after *Clostridium difficile* Infection: A Report on 128 Strain-Typed Cases from a Teaching Hospital in the United Kingdom. *Clin Infect Dis* 2010.
232. **Woodford N., Livermore D. M.**: Infections caused by Gram-positive bacteria: a review of the global challenge. *Journal of Infection* 2009, **59**(Supplement 1):S4-S16.
233. **Zerbino D. R., Birney E.**: Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* 2008, **18**(5):821-829.

Appendix One

Clinical Information for Isolate Panel

The clinical information with respect to clinical outcomes of disease and patient demographics; corresponding to each clinical isolate that was characterised in this study is given in the table below.

Over 186 variables for each isolate/ patient were collected and only some are shown in the tables A1.1 and A1.2 below.

Table A1.1 – Infection outcome pattern for each clinical isolate

Isolate ID	Ribotype	Severity	Patient	Cause of Death*	Number or Episodes	Relapse OR Re-infection	*C.difficile* Positive	Resolution to Type 4 Stools For 48 hrs
FA08012693	13	Mild	1	N/A	One episode		18.12.08	
FA08006290	13	Severe	2	CDI was part of or mentioned in death	One episode		30.06.08	05.07.08
FA07007469	14	Life Threatening	3	N/A	One episode		30.07.07	08.08.07
FA07004464	14	Mild	4	N/A	One episode		24.03.07	24.04.07
FA09007583	15	Severe	5	CDI was part of or mentioned in death	One episode		07.08.09	
FA07011498	15	Severe	6	N/A	One episode		09.11.07	16.12.07
FA07003485	23	Moderate	7	N/A	Multiple Episodes	Relapse with same type	10.04.07	19.04.07
FA07004080	23	Mild	7	N/A	Multiple Episodes	Relapse with same type	26.04.07	
FA08006661	23	Severe	8	CDI was not the cause of death	One episode		09.07.08	
FA08005864	23	Severe	9	N/A	One episode		17.06.08	25.06.08
FA07007522	78	Severe	10	N/A	Multiple Episodes	Relapse with same type	31.07.07	10.08.07
FA07008490	78	Severe	10	N/A	Multiple Episodes	Relapse with same type	28.08.07	07.09.07
FA09004991	78	Life Threatening	11	N/A	One episode		26.05.09	23.07.09
FA08006656	78	Severe	12	CDI was not the cause of death	One episode		09.07.08	22.07.08
FA07001994	27	Mild	13	N/A	Multiple Episodes	Relapse with same type	26.02.07	05.03.07
FA07003754	27	Severe	13	Patient survived	Multiple Episodes	Relapse with same type	16.04.07	24.04.07

Table A1.2- Patient demographics for clinical isolate panel

Isolate ID	Sex	Date Of Birth	Age	Admission	Discharge	Duration Of Stay In Hospital	Cause Of Death	Date Of Death
FA08012693	male	05.06.33	75	25.10.08	22.01.09	89		
FA08006290	male	02.12.24	84	25.06.08	05.07.08	10	Mantle Cell Lymphoma	05.07.08
FA07007469	male	23.10.15	92	06.07.07	10.08.07	35		
FA07004464	male	29.04.24	83	04.03.07	24.04.07	51		
FA09007583	female	16.01.22	87	06.08.09	12.08.09	6		12.08.09
FA07011498	male	10.09.21	86	09.11.07	21.12.07	42		
FA07003485	female	09.05.29	78	02.03.07	22.05.07	81		
FA07004080	female	09.05.29	78	02.03.07	07.06.07	97		
FA08006661	female	29.05.30	78	09.07.08	31.07.08	22	Cardiorespiratory failure with pulmonary oedema.	31.07.08
FA08005864	male	28.08.31	77	08.06.08	25.06.08	17		
FA07007522	male	24.11.45	62	29.07.07	10.08.07	12		
FA07008490	male	24.11.45	62	24.08.07	11.09.07	18		
FA09004991	male	05.10.45	64	22.05.09	23.07.09	62		
FA08006656	male	08.12.25	83	07.07.08	24.07.08	17		08.08.08
FA07001994	male	28.09.23	84	15.02.07	21.03.07	34		
FA07003754	male	28.09.23	84	15.04.07	03.05.07	18		

Appendix Three. Growth Curve and Motility Assay Results

C. difficile Growth Curves in BHIS Broth over 48 hours

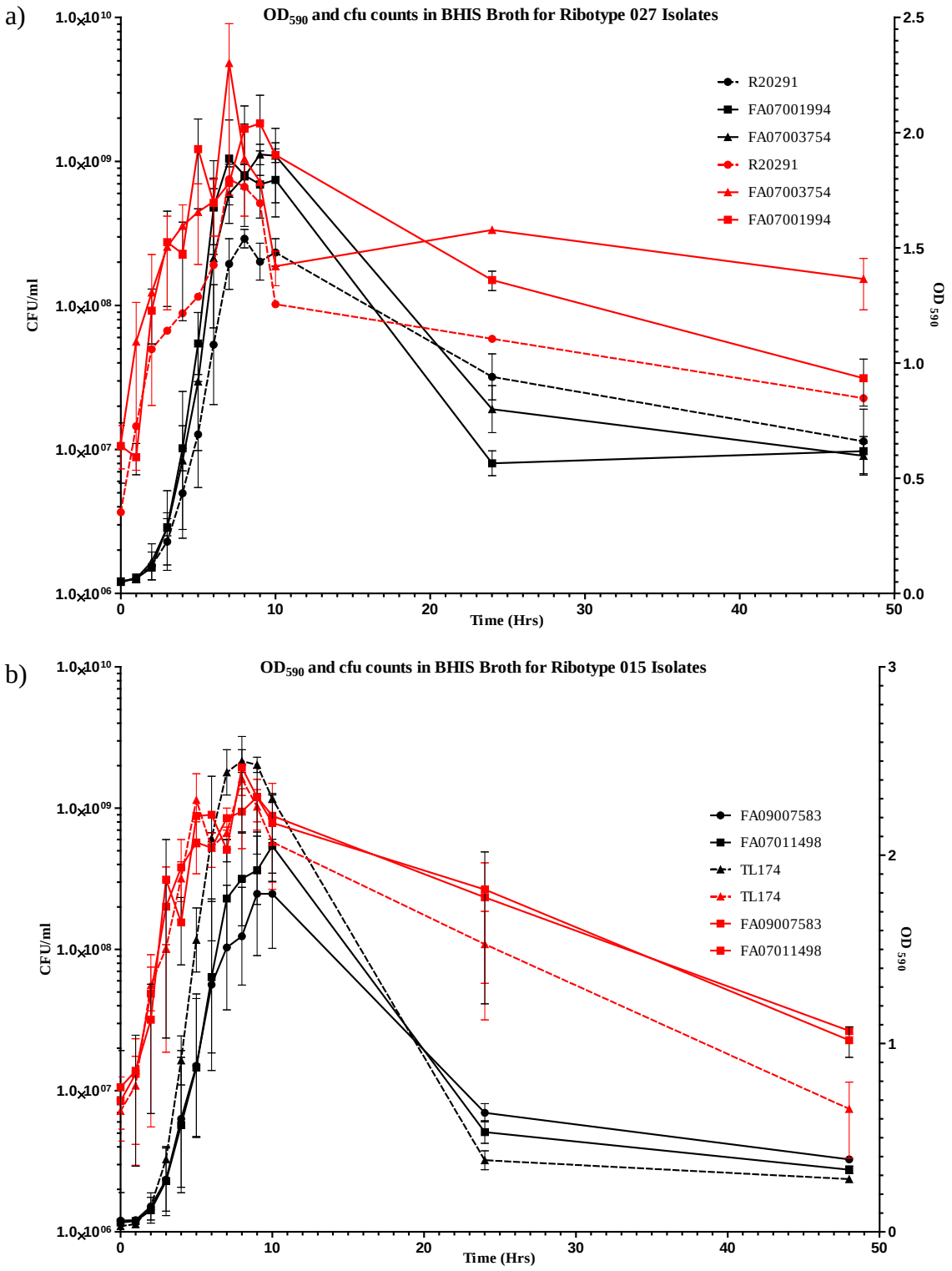
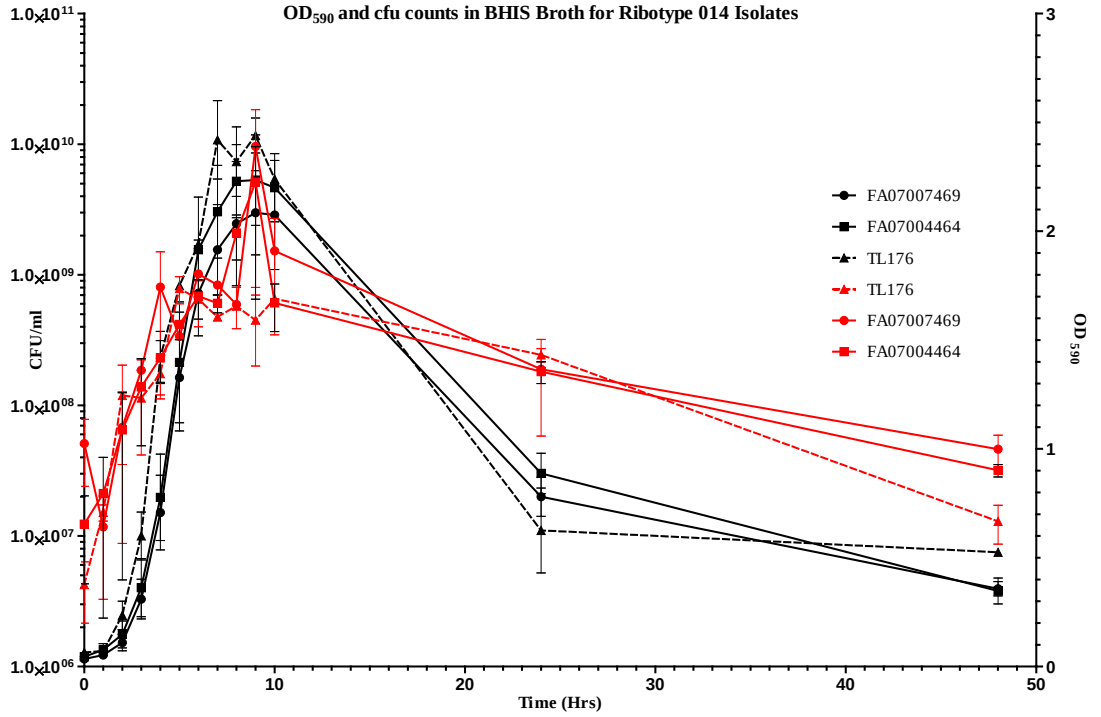


Figure A3.1- Growth curve of *C. difficile* ribotypes a) 027 and b) 015 in BHIS broth over 48 hrs. Black line= OD_{590}, Red line= cfu/ml. Error Bars= SEM; OD_{590} curve-SEM, N=4, cfu/ml curve-SEM, N=2.

a)



b)

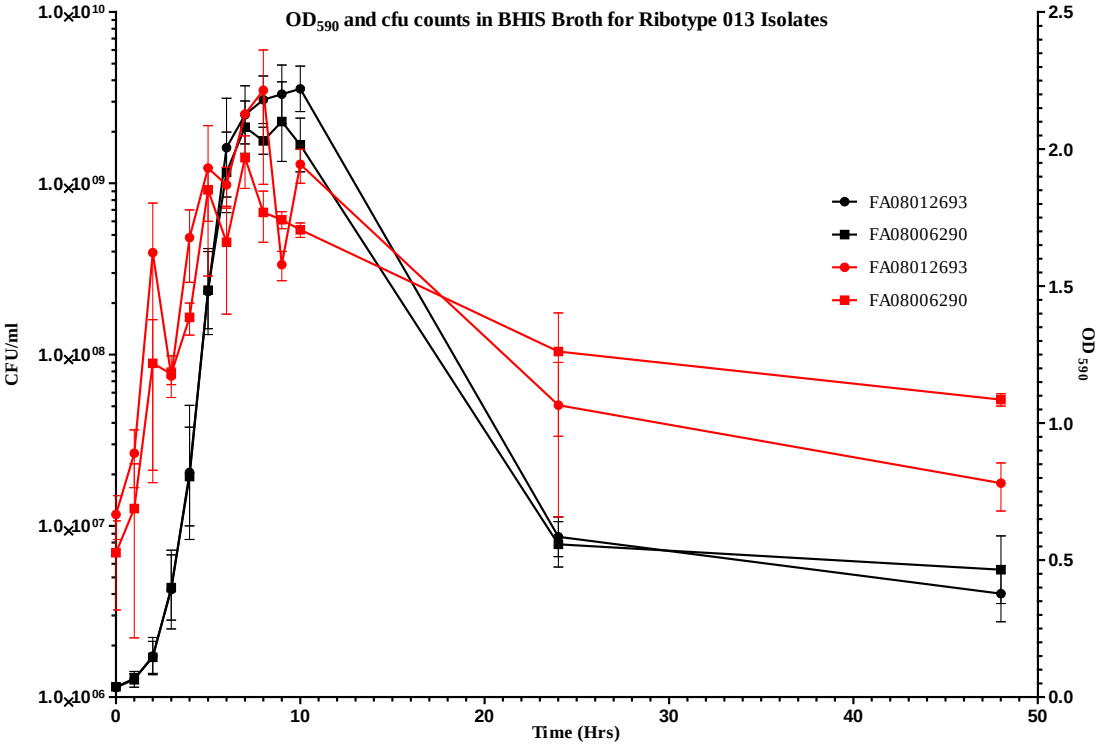
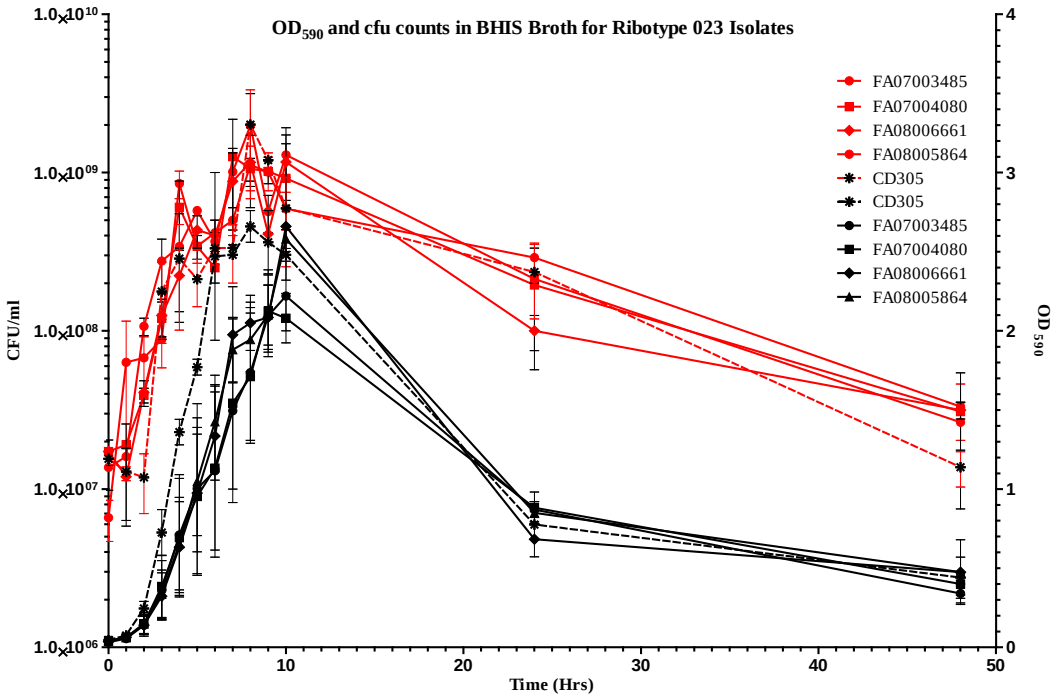


Figure A3.2- Growth curve of *C. difficile* ribotypes a) 014 and b) 013 in BHIS broth over 48 hrs. Black line= OD_{590}, Red line= cfu/ml. Error Bars= SEM; OD_{590} curve-SEM, N=4, cfu/ml curve-SEM, N=2.

a)



b)

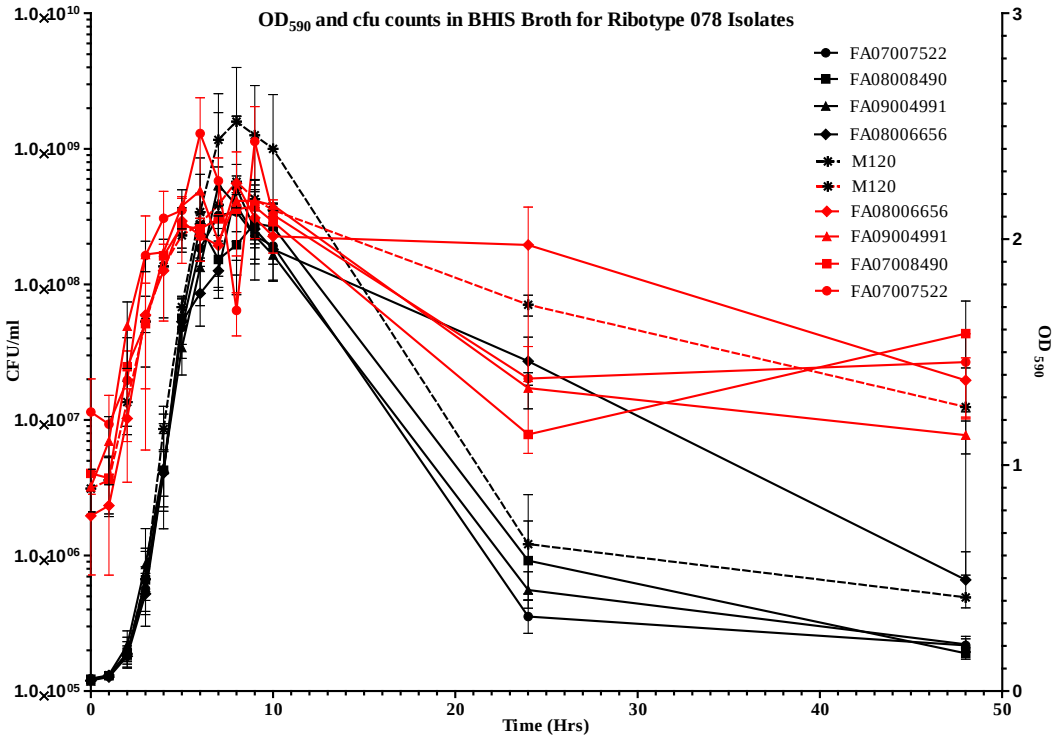


Figure A3.3- Growth curve of *C. difficile* ribotypes a) 023 and b) 078 in BHIS broth over 48 hrs. Black line= OD_{590}, Red line= cfu/ml. Error Bars= SEM; OD_{590} curve-SEM, N=4, cfu/ml curve-SEM, N=2.

Motility of *C. difficile* Isolates in 0.1% BHI Agar

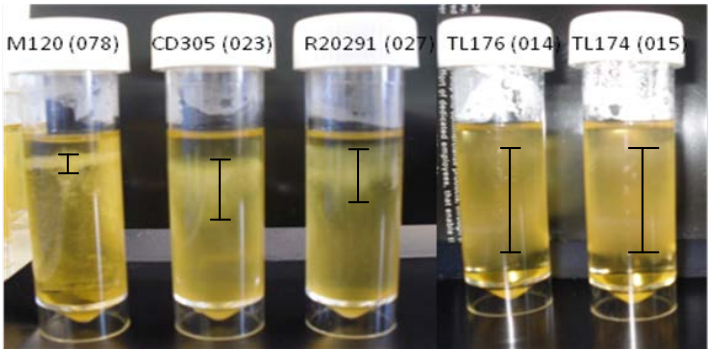


Figure A3.4– Representative motility of reference isolates. Bars indicate points from which motility was measured. A representative isolate was not available for ribotype 013.

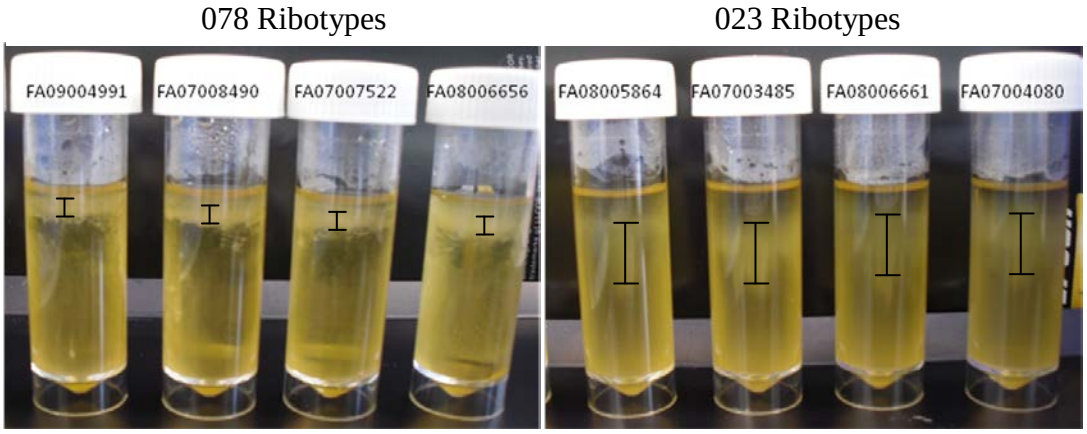


Figure A3.5– Representative motility of 078 and of 023 isolates. Bars indicate points from which motility was measured.

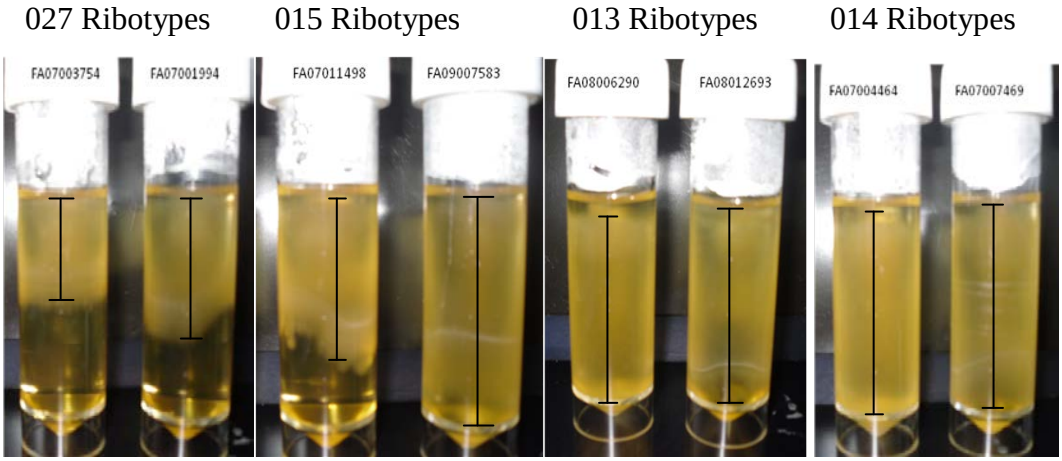


Figure A3.6– Representative motility of 027,015 013 and 014 isolates. Bars indicate points from which motility was measured.