Understanding the genetic mechanisms of *Clostridium difficile* toxin regulation and clinical relapse.

Michelle Mary Lister, BSc.

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy.

February 2018

Abstract

Clostridium difficile is the leading cause of health care associated diarrhoea and remains a burden for the NHS. Disease symptoms can range from mild diarrhoea through to fulminant pseudomembranous colitis, resulting in mortality for some patients. Recurrence is a major problem and estimates are that 20% of all patients with disease will either relapse (with the same strain) or have a re-infection (with a different strain).

Arguably, the main virulence factors are toxins A (TcdA) and toxin B (TcdB) which cause disease symptoms. The genes encoding TcdA and TcdB are located within the pathogenicity locus (PaLoc) along with three accessory genes; *tcdR*, *tcdE* and *tcdC*. The regulatory network has been studied but we aimed to add to this knowledge by using two under investigated strains R20291 a so-called hypervirulent strain and VPI 10463 a strain known to produce higher levels of toxin.

Two different methods of investigation were employed during this study to improve our understanding of both the regulation of TcdA / TcdB but also the genetic mechanisms behind clinical relapse. These methods were; using forward and reverse genetic analysis to assess phenotypic differences and using bioinformatics to identify genes and / or single nucleotide variants (SNP) that may play a role.

Using a combination these methods we have identified potential regulators of toxin production in both strains. We have also identified unique genes and SNPs that might provide a fitness benefit to strains of *C. difficile* that were isolated from patients who had suffered relapse episodes.

Acknowledgements

I dedicate this thesis to my mum, Mary Elisabeth Patterson. Without her sacrifice I would not be where I am today. Her understanding and tolerance knows no bounds and her unconditional love and support are always present.

I would firstly like to thank my ever-understanding husband Darren, whose constant supply of tea has helped me through my writing. Even though he "still has no idea what I do" his support has been unwavering and has been incredibly patient. His ability to make me laugh has made the tough times feel better and I cannot express how grateful I am.

A special thank you goes to Sarah Kuehne, Michelle Kelly, Emma Stevenson, Louise Dynes and James Turton. Without their ears to chew there would have been several laboratory based meltdowns. I am grateful for their support and wisdom and I hope I am a better scientist for knowing them.

To my fellow colleagues within phase two, it was thoroughly enjoyable working with everyone. You will all be missed.

Lastly, to all my friends and family, thank you for your support!

Declaration

Unless otherwise acknowledged, the work presented in this thesis is my own. No part has been submitted for another degree in the University of Nottingham or any other institute of learning.

Michelle Lister

February 2018.

Publications relating to these studies

Lister, M., Stevenson, E., Heeg, D., Minton, N.P. & Kuehne, S.A. (2014). Comparison of culture based methods for the isolation of *Clostridium difficile* from stool samples in a research setting. *Anaerobe* **28**, 226–229

Publications in writing relating to these studies.

Whole genome analysis of *Clostridium difficile* VPI 10463 revels potential genes involved in increased toxin production, a large genome rearrangement and two type I restriction modification systems. Lister, M., Stevenson, E., Minton, N.P. & Kuehne, S.A.

Prevalence of co-infecting strains of *Clostridium difficile* within the Nottingham University Hospitals NHS Trust and understanding the genetics behind recurrent *Clostridium difficile* infection. Lister, M., Stevenson, E., Giles, M., Slone, T., Diggle, M., Spiller, R., Minton, N.P. & Kuehne, S.A.

Abbreviations

AFLP	Amplified fragment length polymorphism
AI-2	Autoinducer-2
AIP	Autoinducing peptides
BHIS	Supplemented Brain Heart Infusion medium
BLAST	Basic Local Alignment Search Tool
BoNT	Botulinum neurotoxin
bp	Basepair(s)
BW	Brendan Wren
CA	Community acquired
CCD	Charge coupled device
CCEY	Cycloserine cefoxitin egg yolk medium
CCFA	Cycloserine cefoxitin fructose agar
CCFB	Cycloserine cefoxitin fructose broth
CCMB-TAL	Cycloserine cefoxitin mannitol broth with taurocholate and lysozyme.
CCR	Carbon catabolite repression
CDI	*Clostridium difficile* infection
c-di-GMP	Cyclic di-guanosyl-5'monophosphate
CDMM	*Clostridium difficile* minimal media
CDRN	*Clostridium difficile* ribotype network
CDS	Coding sequences
CDT	*Clostridium difficile* transferase
CFU	Colony forming units
COG	Functional Clusters of Orthologous Groups
CRISPR	Clustered regularly interspaced short palindromic repeats
CROPS	C-terminal combined repetitive oligopeptides
CRP	C-reactive protein
CTN	Cytotoxin Neutralisation
CWP	Cell wall proteins
Da	Dalton

ddNTPs	didexoyNTPs
DIG	Digoxigenin
DMEM	Dulbecco's modified Eagles medium
DNA	Deoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
FC	5-fluorocytosine
FU	5-fluorouracil
g	Grams
GDH	glutamate dehydrogenase
GDT	glucosyltransferase
GSPG4	chondroitin sulphate proteoglycan 4
HA	Hospital acquired
HGT	horizontal gene transfer
IMG/MS	Integrated Microbial Genomes & Microbiome Samples
$InsP_6$	Inositol hexakisphosphate
ITRs	Inverted terminal repeats
IV	Intravenous
JGI	Joint Genome Institute
Kbp	Kilo base pairs
kg	Kilo gram
L	Litres
LB	Luria-Bertani
LCT	Large clostridial cytotoxins
mg	Milli grams
MIC	Minimum inhibitory concentration
MICE	Minimum inhibitory concentrations evaluators
mL	Milli litres
MLST	Multilocus sequence typing
MLVA	Multilocus variable number tandem repeat analysis
Mm	Milli meter
MNV	Multiple nucleotide variants

MRSA	Methicillin resistant *Staphylococcus aureus*
NAAT	Nucleic acid amplification test
NADPH	2,4-dienoyl-CoA reductase
ng	Nano grams
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
nm	Nano metre
NM	Nigel Minton
NS	Non-synonymous
NTC	No template control
NUH	Nottingham University Hospitals
OB-fold	Oligonucleotide-binding fold
OD	Optical density
ORF	Open reading frame
PaLoc	Pathogenicity locus
PBS	Phosphate buffer solution
PCR	Polymerase chain reaction
PFGE	Pulse-field gel electrophoresis
PHASTER	Phage Search Tool Enhanced Release
PHE	Public Health England
PMC	Pseudomembranous colitis
PNPG	p-nitrophenyl-β-D-glucopyranoside
PPI	Proton pump inhibitors
PROVEAN	Protein Variation Effect Analyzer
PTS	Phosphotransferase system
QSs	Quadruplex structures
RAM	Retrotranspositin-activated selectable marker
RAST	Rapid Annotation Subsystem Technology
REA	Restriction endonuclease analysis
REBASE	Restriction Enzyme Database
RFLP	Restriction fragment length polymorphisms

RNA	Ribonucleic acid
RNP	Ribonuclear protein
rpm	Revolutions per minute
RT	Ribotype
*slpA*ST	Surface layer protein A sequence typing
SMRT	Single Molecule Real Time Sequencing
SNP	Single nucleotide variants
Tcd*x*	*Clostridium difficile* toxin (A, B, C, E or R)
Tn	Transposon
TY	Tryptose Yeast Extract Medium
v	Volume
VNTR	Variable number tandem repeat
VRE	Vancomycin Resistant Enterococci
w	Weight
WCAA	Wilkins - Chalgren Anaerobe Agar
x g	Gravity
μL	Micro litres

Table of Contents

Table of Figures

Table of Tables

Chapter One Introduction

1.1 *Clostridium difficile*

1.1.1 Clostridium difficile and its emergence as a human pathogen

The genus *Clostridium* is comprised of approximately 100 described species, the majority of which are benign[1]. They are ubiquitous in the environment and can be found in soil, marine sediments, sewage and the intestinal tracts of both humans and animals[2]. Several species have been shown to have useful applications in a variety of fields. *Clostridium acetobutylicum* and *Clostridium beijerinckii* are both examples of clostridia that are used for the industrial production of biofuels due to their ability to ferment organic compounds[3]. Spores of *Clostridium sporogenes* and *Clostridium novyi* have shown to have great potential as tumour delivery vehicles for chemotherapeutic agents[4]. Conversely several *Clostridium* species have the ability to cause disease in humans under favourable conditions of which *Clostridium perfringens* and *Clostridium difficile* are arguably the most notorious. *C. perfringens* is known to cause gas gangrene, gastrointestinal infections (after consumption of contaminated foods) and necrotic enteritis in infants[5]. However, in recent years the spotlight has been reserved for *C. difficile*. It was coined a "superbug" by the media, along with methicillin resistant *Staphylococcus aureus* (MRSA)[6], due to the increased incidence of *Clostridium difficile* infection (CDI) in hospitals and associated morbidity. *C. difficile* has been implicated as the cause of extracolonic diseases, including reactive arthritis, bacteraemia and wound infections[7], but its most prevalent site of infection is the gastrointestinal tract.

C. difficile is an anaerobic, Gram-positive bacillus, of 0.3-2 x 1.5-2 μm in size, and which is able to form endospores. It was first described in the literature as part of the neonatal gastrointestinal microflora in 1935 and named *Bacillus difficilis*[8]. Over time, as knowledge improved for the class Bacilli, it was noted that certain genera within the class were

aerobic/facultatively anaerobic and others were obligate anaerobic. This prompted the division of the class into Bacilli and Clostridia.

C. difficile was first recognised as a human pathogen in 1978 when a number of pseudomembranous colitis cases were found to be associated with the use of clindamycin[9]. Screening of symptomatic patient faeces showed that *C. difficile* toxins were present and were likely the cause of symptoms. However, the true effect of clindamycin on CDI was not fully understood until 1999. Through the analysis of four distinct outbreaks, using susceptibility testing and typing of *C. difficile* strains, in combination with clinical data, Johnson *et al.*[10] were able to demonstrate that the use of clindamycin increased the risk of CDI. This was due to a clindamycin resistant *C. difficile* strain which was epidemic at the time. During the 2000's there was a steady rise in the frequency of outbreaks of CDI throughout Europe, North America and some parts of Asia[11]. *C. difficile* is now recognised as a major cause of hospital acquired diarrhoea and has become a considerable burden for healthcare services due to treatment costs, patient isolation, and ward closures. In Europe, the estimated incremental cost per patient associated with CDI is between £4577 and £8843[12].

1.1.2 Clostridium difficile risk factors

The risk of developing CDI is most significantly increased by the use of antibiotics. This risk increases with prolonged duration of antibiotic treatment and by the number of different antibiotics received[13]. The antibiotics associated with the highest risk are the broad-spectrum second-generation (and higher) cephalosporins, clindamycin and fluoroquinolones [10,14–16], but over the years almost all antibiotics have been linked to CDI[16]. Patients are at highest risk during therapy and for the first month after therapy is ceased, however, this risk starts to decline between month one and three[17].

Other important risk factors include advanced age; patients over the age of 65 have an increased risk of acquiring CDI. It is also believed that for each year of age, after 18, the risk increases by 2% each subsequent year for health-care acquired CDI[11]. Prior hospitalisation and/or residents of long term care facilities also have an increased risk[18]. Other risk factors associated with CDI are the use of antacids, particularly proton pump inhibitors (PPI)[11,14,19]. Although this is still controversial and the mechanism by which PPIs increase the risk is not well described. There are a number of other conditions/treatments which act as risk factors. These include, irritable bowel disease, end stage renal disease, chronic liver disease, nasogastric feeding, chemotherapy and immunosuppression[14,16,20].

A further discussion of risk factors and their association with recurrent infection will be presented in Chapter Five.

1.1.3 Clostridium difficile disease

It is the ability of *C. difficile* to produce endospores that enables it to be a major nosocomial pathogen. The spore form can stay in the environment for long periods of time, resist many common cleaning practices, be transmitted between patients and can cause recurring infection[21]. After ingestion of spores it is not known at what stage they germinate and return to vegetative cells, although the bile salts present in the intestine (e.g., taurocholate) play a role in this process[22]. In the mouse model, germination occurs in the small intestine and caecum, while in the hamster model 80% of germination occurs in the small intestine[23,24]. It can be assumed that it is the same in the human host. The environment has to be favourable for colonisation to occur, such as after antibiotic treatment when the normal gut microbiota has been disrupted and the colonisation resistance it affords ablated.

The disease state is caused by the production of toxins A and B. These toxins enter the intestinal epithelial cells and affect the organisation of the cell cytoskeleton resulting in cell death[25]. It is this process that leads to inflammation and the clinical symptoms of CDI. This will be discussed in more detail in section 1.2.

The clinical manifestation of disease can vary depending on the severity of infection. In mild to moderate infection stools are watery and the patient may exhibit signs of colitis such as, lower abdominal cramps, fever, leucocytosis and hypoalbuminemia[20]. Severe CDI is associated with a white blood cell count of >15 x 10^9/L, a temperature of >38.5°C, an acute rising serum creatinine and evidence of sever colitis[20,26]. In <5% of cases fulminant disease occurs, in which symptoms include severe abdominal pain, perfuse diarrhoea, hypotension, ileus, pseudomembranous colitis and/or toxic megacolon[20,26,27]. Recurrent infection occurs in 19 – 35% of all cases, this may be re-emergence of the original strain or infection with a new strain, usually within 30 days of completing treatment for CDI[28–30]. After one episode of recurrent CDI the risk of further relapses increases[29]. Recurrent infection will be discussed in more detail in Chapter Five.

1.1.4 Treatments

The recommendations for treatment of CDI within the UK were last updated in 2013 by Public Health England (PHE)[31]. It is recommended that all implicated antimicrobial treatments are discontinued at the earliest possible timepoint. Antibiotic treatment is dependent on the severity of disease. In the case of mild and moderate disease oral Metronidazole is recommended as it has been shown to be as effective as Vancomycin, is cheaper and will not increase the risk of selection of Vancomycin Resistant Enterococci (VRE)[32,33]. Oral Vancomycin is reserved for severe cases as a number of studies have shown it to be superior to Metronidazole[31]. If patients are at high risk of recurrent infection Fidaxomicin should be considered. It has been shown to be more effective at preventing

recurrence over Vancomycin[34–36]. Patients who are not responding to Vancomycin should either receive Fidaxomicin or high dosage oral Vancomycin plus intravenous (IV) Metronidazole[31]. In life threatening cases, such as toxic megacolon, septic shock or perforation, colectomy may be required in an attempt to prevent mortality[26,27]. Rifaximin, when given as a "chaser" therapy after traditional CDI antibiotics, has been shown to reduce recurrence rates[37,38].

There are concerns about resistance, a few cases have been described where *C. difficile*, cultured after Rifaximin treatment, have shown high minimum inhibitory concentrations (MIC) to the drug[39,40].

1.1.5 Clostridium difficile surveillance in the UK

Surveillance first started in England and Wales in 1990 as a voluntary scheme with Northern Ireland joining in 2001. Over a 15-year period there was a steady increase in the number of cases of CDI (Fig 1.1)[41]. As a result, in 2004 a mandatory scheme was put in place in England for reporting CDI in patients over 65 years of age. Since April 2007 it has been mandatory for all NHS trusts in England to report all cases of CDI in all patients over 2 years of age in an attempt to accurately assess the frequency of CDI[42,43].

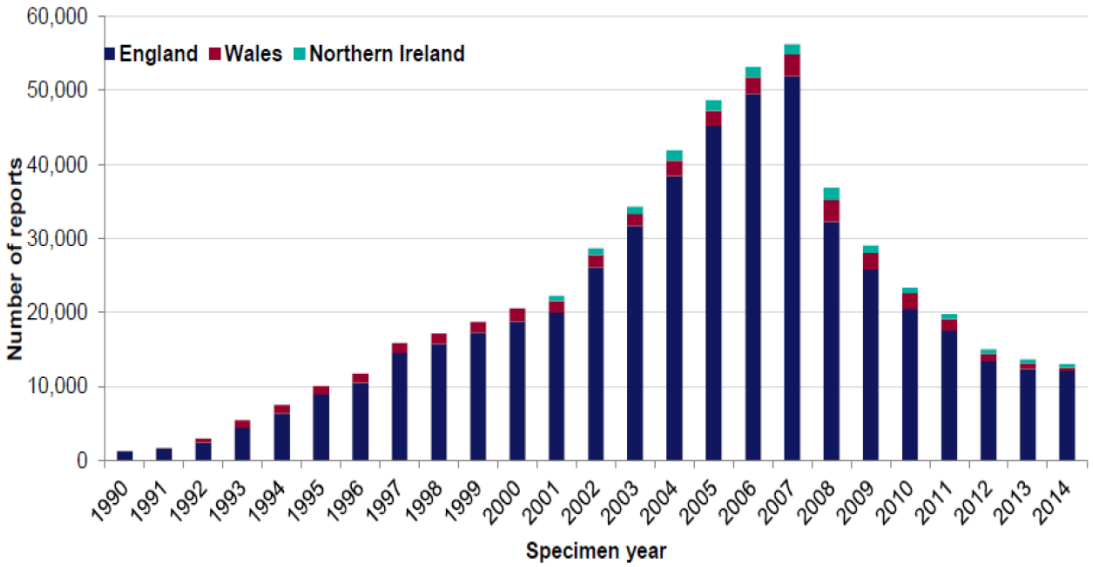


Figure 1-1: Number of positive *C. difficile* faecal samples reported under the voluntary reporting scheme for England, Wales and Northern Ireland* 1990-2014. *Northern Ireland joined in 2001[41].

In April 2007 the *Clostridium difficile* Ribotyping Network (CDRN) was also established to enhance the mandatory surveillance of CDI by providing epidemiological information that assists in the recognition and control of epidemic strains. Since its inception the network has grown from six to nine laboratories, with one being in Northern Ireland. Samples are provided depending on local clinical need but are usually due to one of the following scenarios[42]:

- Increased frequency of cases or high baseline rates of CDI
- Increased severity/complications of cases of CDI
- Increased mortality associated with CDI
- Increased recurrence rate of CDI

Since the introduction of the CDRN the rates of CDI have fallen significantly (Fig 1.1), possibly due to better epidemiology allowing improved infection control and case management.

1.1.6 Typing methods for Clostridium difficile

Typing is used as an epidemiological tool to cluster isolates with regards to either genotypes or phenotypes. There are multiple different typing methods available for *C. difficile*, historically phenotypic characteristics were used. These included antibiotic resistance profiles, slide agglutination methods, Western immunoblotting and soluble protein patterns[44]. These methods were acceptable for local use and showed that *C. difficile* was transmissible between patients[6]. However, as the technology available evolved so did the typing methods. This improved reproducibility between laboratories and enabled the epidemiological monitoring of *C. difficile* within countries and globally.

Killgore *et al.*[44] compared seven different typing methods; pulse-field gel electrophoresis (PFGE), amplified fragment length polymorphism (AFLP), multilocus sequence typing (MLST), restriction endonuclease analysis (REA), surface layer protein A sequence typing (*slpA*ST), multilocus variable number tandem repeat analysis (MLVA) and polymerase chain reaction (PCR)-ribotyping. The study found that MLVA and REA were highly discriminatory between strains, where both methods were able to separate the so called hypervirulent BI/NAPI/027 strains by source location. These methods are likely to be too discriminatory to be used to track routine epidemiology of CDI. MVLA is, however, used as an enhanced service by the CDRN in cases where high rates of CDI are recorded within a Trust, in those instances where there has been a failure to meet CDI targets and/or in outbreak cases after agreement with the local health protection team[45]. It has also been suggested that MLVA should be used for the phylogeny of *C. difficile*[46]. Methods such as PFGE and REA are labour intensive and it can be difficult to compare data between laboratories, a failing partially addressed through the use of computer software[44]. *slpA*ST and MLST are sequence based methods that both produce highly transferable data, but which are relatively expensive in comparison to other methods[47]. The method of typing

currently used in the UK by the CDRN is PCR-ribotyping. This method employs PCR to amplify the intergenic spacer region between the 16s rRNA gene and the 23s rRNA gene [48]. The genetic differences, copy number and size, in these regions are enough to produce a distinctive "fingerprint" for each strain[6]. Kilgore *et al.*[44] found that even when, at two separate laboratories, different primer sets were used there was only one disagreement within the 41 tested over which ribotype (RT) to assign. There are hundreds of different ribotypes and each has been designated a unique three-digit code starting at RT001[48]. Another typing method, toxinotyping, was not covered by Kilgore *et al.*[44], uses selected fragments of the *tcdA* and *tcdB* genes to look for differences in restriction fragments to a reference strain (VPI 10463) when digested with specific restriction enzymes [49]. These differences are known as restriction fragment length polymorphisms (RFLP). There are over 30 different toxinotypes and these have been shown to correlate with serotypes[49–51]. The advantages and disadvantages for each method are described in Table 1-1.

Table 1-1: Commonly used *Clostridium difficile* typing methods and their associated advantages and disadvantages. Information gathered from[44,52,53].

Typing method	Advantages	Disadvantages
Pulse-field gel electrophoresis (PFGE)	Offers good discriminatory power. Inexpensive.	Labour intensive. Consensus between restriction endonucleases needed. Data not easily transferred between laboratories.
Amplified fragment length polymorphism (AFLP)	Highly discriminatory. Can choose the number of loci within one reaction. No prior knowledge of DNA sequence is required.	Specialised equipment required if process is automated. Complex protocol.
Multilocus sequence typing (MLST)	Data for comparison available via the internet. Reproducibility and repeatability high.	Low discriminatory power. Expensive. Complex protocol. Requires a skilled user.
Restriction endonuclease analysis (REA)	Offers good discriminatory power. Reproducible.	Labour intensive. Interpretation can be difficult. Data not easily transferred between laboratories.
Surface layer protein A sequence typing (slpAST)	Relevant to vaccine development.	Not widely used.
Multilocus variable number tandem repeat analysis (MLVA)	Highly discriminatory. Reactions can be multiplexed using fluorescent probes and capillary PCR.	Unsuitable for long-term epidemiological surveillance as loci may evolve too quickly. Labour intensive. Expensive.
Ribotyping	Most commonly used typing method of *C. difficile* in the UK. Offers good discriminatory power. Results are reproducible.	Labour intensive. Specialised equipment required (if performing capillary electrophoresis). Data not easily transferred between laboratories. Reference database not easily accessible.
Toxinotyping	Highly reproducible.	Low resolution.

1.2 *Clostridium difficile* pathogenesis

1.2.1 Structure and function of toxin A and B

Toxin A (TcdA) and toxin B (TcdB) are well described as the main virulence factors of *C. difficile*. It has been shown that non-toxigenic strains of *C. difficile* are able to colonise the colon but are not able to cause CDI, because of this these strains are being considered as a potential treatment for recurrent CDI infections[54]. TcdA and TcdB are large (308 and 270 kDa, respectively), homologous (47% DNA identity and 63% amino acid identity) proteins[55]. They consist of four functional domains; the N-terminal glucosyltransferase (GDT) is the enzymatic component plus the C-terminal combined repetitive oligopeptides (CROPS), delivery/pore-forming and the autoprotease domains (Fig 1.2)[56]. They belong to the large clostridial cytotoxins (LCTs) due to their high molecular weight, glycosylating activity and cytopathic effect on cells[57].

Binding and internalisation of TcdA and TcdB occurs in a number of steps termed the ABCD model (A, biological activity; B, binding; C, cutting; D, delivery). Firstly, the CROPS bind to receptors on the cell surface after which endocytosis is initiated. TcdA shows specificity to carbohydrates with a Galß1-4GlcNAc core although the exact ligand is still unknown[58,59]. TcdB has recently had a receptor identified, chondroitin sulphate proteoglycan 4 (GSPG4)[60]. After binding, toxins are internalised into endosomes and acidification induces a conformational change in the delivery domain resulting in insertion into the membrane[61]. The autoprotease domain is then activated by binding to inositol hexakisphosphate ($InsP_6$) resulting in cleavage of the GTD into the host cell cytosol[62]. The toxins can then modify and inactivate the Rho family GTPases Rho, Rac and Cdc42 resulting in cell rounding, disaggregation of the actin cytoskeleton, loss of intestinal epithelium barrier function and cell death[55] .

Figure 1-2: Schematic representations of TcdA and TcdB adapted from[56]. The N-terminal glucosyltransferase domain (GTD) (red), oligopeptides (CROPs) (green), delivery/pore-forming (yellow), delivery domain (orange) and autoprotease (blue).

Historically, using pure toxin extracts, it was established that TcdA could elicit a disease response in hamsters when administered alone, whereas TcdB could not. Although, TcdB could cause disease symptoms if a sub-lethal dose of TcdA was administered at the same time[63]. This caused confusion over how strains with only functional TcdB were causing clinical infections as this did not fit the common conception that either both toxins or TcdA alone were required for disease[64]. In the hamster model it has been shown that TcdB alone can cause disease [65,66]. Lyras *et al.*[65] used homologous recombination to create single crossover mutants with either TcdA or TcdB deficient phenotypes. They showed that in the hamster model A^-B^+ mutants caused CDI and animals were more likely to die than when infected with A^+B^- mutants. Kuehne *et al.*[67] were able to create stable, isogenic mutants of *C. difficile* using ClosTron technology[68,69], meaning that not only A^-B^+ and A^+B^- mutants were constructed but for the first time a A^-B^- mutant was generated. Using these mutants their findings corresponded with those of Lyras *et al.* in that TcdB alone could cause a disease state in hamsters. However, they also showed that TcdA alone could also cause disease in hamsters. The difference in outcomes between the two studies has been hypothesised to be due either to different endpoints used in the animal models or the presence of secondary mutations in the progenitor strains used for mutant construction[66]. In both cases, the creation of the mutants required the prior independent isolation of erythromycin sensitive derivatives of *C. difficile* 630 through repeated subculture. The derivative used by Kuehne *et al* was 630Δ*erm*, while that made by Lyras *et al.* was

designated 630E. It is suggested that mutations may have arisen during the extensive subculturing undertaken during their isolation and these may have effected their virulence phenotype[66]. Thus, for instance, strain 630E has become non-motile[66].

In 2014 Kuehne *et al.*[70] created similar A^-B^+, A^+B^- and A^-B^- mutants in the RT027 strain R20291. Using the same experimental model as used previously[67] they found that the A^+B^- and A^-B^+ mutants were both able to cause disease, although the former was less virulent. This was followed by Carter *et al.*[71] creating similar mutants in a different RT027 strain M7404. In this study using three different animal models they also found that A^+B^- mutants were less virulent than A^-B^+ mutants. Histological findings showed that A^+B^- mutants caused more superficial and localised damage when compared to the A^-B^+ counterparts. It appears from these data that both TcdA and TcdB have a role in pathogenesis. This is validated by the recent discovery of a TcdA positive/TcdB negative strain of *C. difficile* which was isolated from a clinical case in Cambrai Hospital, France[72].

1.2.2 The Pathogenicity Locus

The pathogenicity locus (PaLoc) is a19.6kb region that encodes the genes for TcdA and TcdB (*tcdA* & *tcdB*) along with three accessory genes; *tcdR*, *tcdE* and *tcdC*[73] (Fig 1.3). The PaLoc was first sequenced in 1995 from strain VPI 10463 and in non-toxigenic strains a 127bp fragment resides in its place[73].

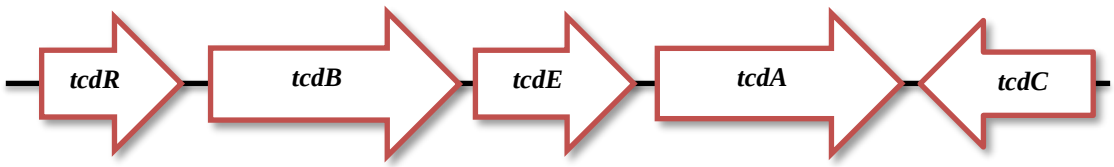


Figure 1-3: Schematic of the PaLoc from type strain VPI 10463. Arrows indicate direction of transcription, figure not to scale.

TcdR, encoded by *tcdR*, is believed to be a positive regulator of toxin production. It is a 22kDa protein that is part of the group 5 of the sigma 70 family and it binds directly to the RNA polymerase core enzyme[74]. It is thought that *tcdR* has two promoters, one which is

independent and allows a basal level of toxin gene transcription and another which is more potent and is subject to growth phase[75]. To date its role in toxin production is undisputed[76].

TcdC, encoded by *tcdC*, is believed to be a negative regulator of toxin production. This belief is based both on the observed inverse correlation between transcription of *tcdC* and the toxin genes and on biochemical data[77]. TcdC is a 26kDa anti-sigma factor[74] which binds to TcdR-RNA polymerase core enzyme to prevent it binding to the promoter regions of the PaLoc[77]. The *tcdC* gene is encoded on the reverse strand of the other four genes on the PaLoc and is located downstream in relation to them (Fig 1.3). There are however doubts about the role of TcdC due to two different observations. Firstly, there is variability of toxin production in hyper-virulent strains even though they frequently carry mutations in the *tcdC* gene[76,78]. Secondly, there have been studies that show increased TcdC production that coincides with increased toxin production[76,78]. One recent study claims that mutations in the *tcdC* gene are an "important factor in the development of hypervirulence in epidemic *C. difficile* isolates"[79] postulating that the mutations in the *tcdC* gene lead to increased toxin production. The study was performed using autonomous plasmids to complement mutant strains carrying dysfunctional *tcdC* genes. The use of data derived from such studies should, however, be treated with caution has the multicopy nature of plasmids can affect outcomes due to the high gene dosage of the complementing gene. In another recent study, Cartman *et.al.*[80] used allelic exchange to both create and repair mutant strains carrying dysfunctional *tcdC* genes. This study demonstrated that there was no significant difference in the level of toxin produced between the isogenic strains. As all the mutations were performed on the chromosome it reflects a more natural environment. With conflicting evidence from different studies, it is difficult to speculate what role TcdC has within the bacterium. Indeed, evidence presented by van Leeuwen *et al.*[81] suggests that TcdC may not even bind to the PaLoc at all. Their data showed that

TcdC forms an oligonucleotide-binding fold (OB-fold) that binds to quadruplex structures (QSs). However, none of the recognition sites are found within the PaLoc. Although they do not rule out that the binding determinant may associate with a different structure. The regulation of toxin gene expression will be discussed in more detail in Chapter Three.

The method in which toxins are released from the bacterial cell is usually due to a C-terminal or N-terminal signal peptide, a Tat-signal peptide or some other alternative definable signal peptide. *C. difficile* toxins A and B do not possess any of these signal peptides and therefore are exported to the extra cellular space by some other means[82]. It has been hypothesised that the *tcdE* gene and its product TcdE play a role in the excretion of TcdA and TcdB[83]. TcdE is a 19kDa protein that is highly hydrophobic, it is predicted to contain three transmembrane domains and shares structural and sequence similarities to class 1 holin proteins[82,84]. Holins are usually produced by bacteriophages and are small membrane proteins designed to release intracellular phage's after development. They form lethal holes within the membrane by oligomerisation which results in cell lysis[82–84]. There is conflicting evidence as to whether TcdE acts as a secretion system for TcdA and TcdB. One study showed that functional inactivation of the *tcdE* gene did not affect any of their measured clostridial characteristics, namely growth, sporulation or release of TcdA and TcdB and that their release correlated with bacteriolysis which was independent of TcdE production[84]. Another study showed that TcdE was essential for TcdA and TcdB release and did so by secreting the proteins rather than lysing the bacterial cell[82]. Both studies used ClosTron mutagenesis to create insertional mutations in *tcdE* at position 234 so it is unclear why there were two different outcomes but this might be due to using two different erythromycin sensitive strains of CD630 (as described earlier) and/or different experimental methods. A more recent study not only appears to support the hypothesis that TcdE is required for transporting toxins out of the cell but also gives an explanation

as to why Olling *et al*.[84] found that the cells lysed more easily. By creating *tcdE* mutants in both R20291 and CD646, Govind *et al*.[85] were able to show that R20291 mutants were deficient in releasing toxin into culture medium. However, in CD646 low levels of toxin were produced and the cells lysed more rapidly. Consequently, toxins from CD646 are released into the culture medium independent of TcdE. These data complements with both studies in that low toxin producing strains (i.e., CD646 and 630Δ*erm*) are lysed and higher toxin producing strains (i.e., R20291 and 630E) are unable to efficiently transport toxins from the cell[82,84,85]. These findings demonstrate that our understanding of species should not be based on individual strains.

1.2.3 Other Clostridium difficile virulence factors

In addition to TcdA and TcdB some strains of *C. difficile* also produce a third toxin known as binary toxin or *Clostridium difficile* transferase (CDT), an actin-specific ADP-ribosylating toxin[86]. The exact role of CDT in virulence remains unclear, however, it is interesting that the toxin is found in strains associated with severe CDI such as RT027 and RT078[87]. Studies have shown that CDT may have a role in colonisation and adhesion[88], but there is conflicting evidence as to whether it can[70,89] or cannot[90] cause disease in hamsters when administered alone.

Adherence to intestinal epithelial cells is essential for colonisation and the establishment of infection. Putative virulence factors that have shown to have a role in adherence or colonisation are cell wall proteins (CWP), the heat shock protein GroEL, the surface layer (S- layer), fibronectin binding proteins, fimbriae and the flagella[91].

Endospores are the infectious particle and are a required step in the life cycle of *C. difficile*. The spores are excreted in faeces and contaminate the environment. They are able to resist radiation, heat and many alcohol and chemical based disinfectants[92] and can persist in the

environment for extended periods of time. It has also been demonstrated in the murine model by Lawley *et al.*[93] that colonised immunocompetent subjects can become high shedders of spores after receiving antibiotic treatment, increasing the risk to immunocompromised subjects.

1.3 Genetic tools for the manipulation of the *Clostridium difficile* genome

Molecular biology is widely used to inactivate genes that are proposed to have an effect of pathogenicity or virulence and look for a measurable difference in phenotype. In 1988 Falkow[94] suggested that the study of singular genes required a form of molecular Koch's postulate, in that there has to be a well-defined effect from the loss or gain of the gene in question and that restoration of this gene reverts the discovered phenotype.

1.3.1 ClosTron insertional inactivation

Initial mutants made in *C. difficile* were found to be unstable[95]. In 2007, Heap *et al.*[68] developed ClosTron: A universal gene knock-out system for the genus *Clostridium*. This system employs a bacterial group II intron which carries an antibiotic resistance gene which itself is interrupted by a group I intron. When the ClosTron plasmid has been conjugated into the *C. difficile* host and the group II intron is transcribed, then the LtrA protein (which is located elsewhere on the ClosTron plasmid) binds to the transcript after which a ribonuclear protein (RNP) is formed. The group I intron is then spliced out resulting in an intact antibiotic resistance gene, the RNP recognises and binds to the target site, nicks the DNA and inserts the RNA in the chromosome. The LtrA protein then synthesises the complementary DNA strand and host polymerases degrade the RNA and synthesise new DNA. Ligases repair the two gaps resulting in a functional antibiotic resistance gene inserted in the gene of interest.

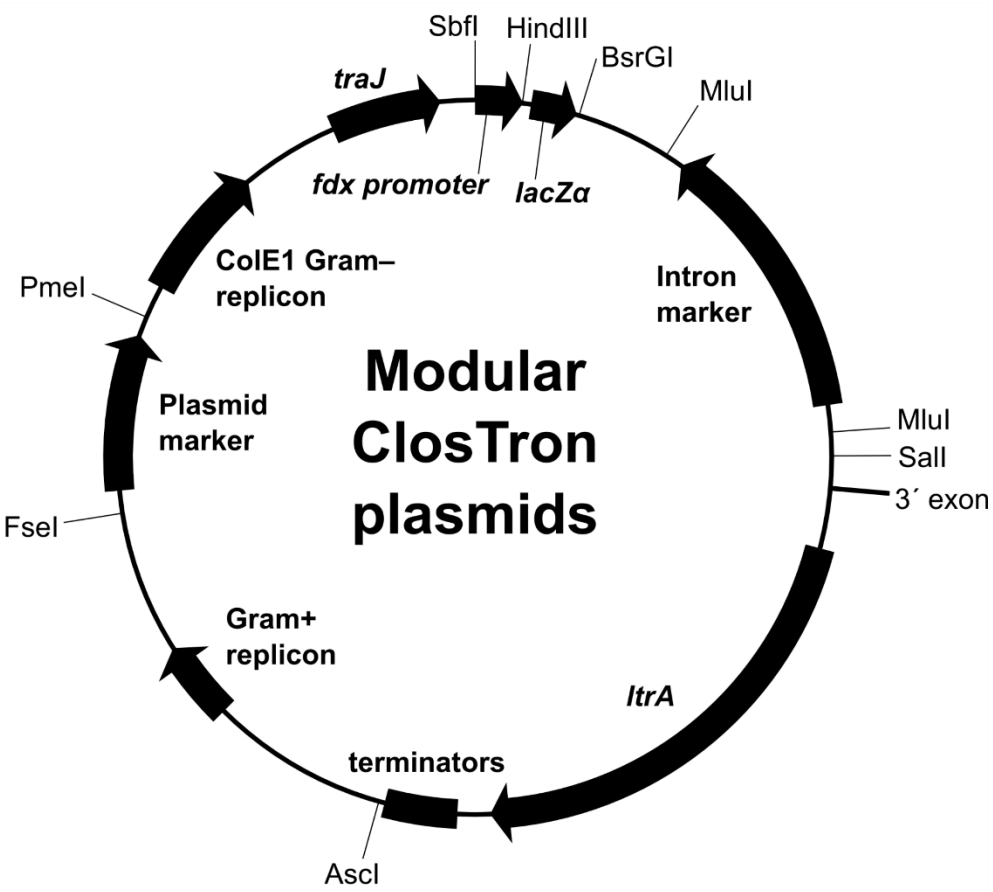


Figure 1-4: Schematic of ClosTron plasmid pMTL007C-E2[69]. The plasmid contains a Gram-positive replicon (pCB102), a thiamphenicol resistance gene (*catP*) for transconjugant selection, a Gram-negative replicon (ColE1) and the conjugal transfer function (*traJ*). The intron marker contains the group II intron and the erythromycin resistance gene (*ermB*) which is inactivated by the group I intron. The ltrA gene facilitates the excision of the group I intron and the insertion of the group II intron into the chromosome.

This technology was improved in 2010 which resulted in the ability to create multiple mutations in a single strain by using marker recycling[69]. Fig 1.3 shows a schematic of the pMTL007C-E2 plasmid. What enables the system to be used in multiple locations throughout the genome is the Perutka Method[96] of group II intron design, this alters the intron RNA to complement different DNA targets. An online algorithm (www.clostron.com) has automated the process of intron design where after imputing a target sequence a list of potential target sites is produced along with a score. A higher score increases the likelihood a successful insertion will occur.

1.3.1 Mariner Transposon random mutagenesis

In 2010, Cartman and Minton[97] established a method of creating random mutants in *C. difficile* by utilising a *mariner*-transposon. Other systems had been previously described[98,99] but the transposons employed (Tn*916* and Tn*5397*) showed a strong target site preference and multiple insertions in single clones respectively. These are both undesirable features for the study of random mutant libraries. The *mariner*-transposable element *Himar1* uses a cut and paste mechanism for DNA transfer and is the only element required for transposition[100]. The transposon is flanked by inverted terminal repeats (ITRs) and inserts in to TA target sites which is suitable for a low-GC content organism like *C. difficile*. Fig 1.5 shows a schematic of the pMTL-SC1 plasmid. This method has been successfully used in our laboratory to identify genes involved in auxotrophic, sporulation and germination deficiency[97]. It has also been used to identify genes involved in toxin A and B deficiency and this will be discussed in Chapter Three of this thesis.

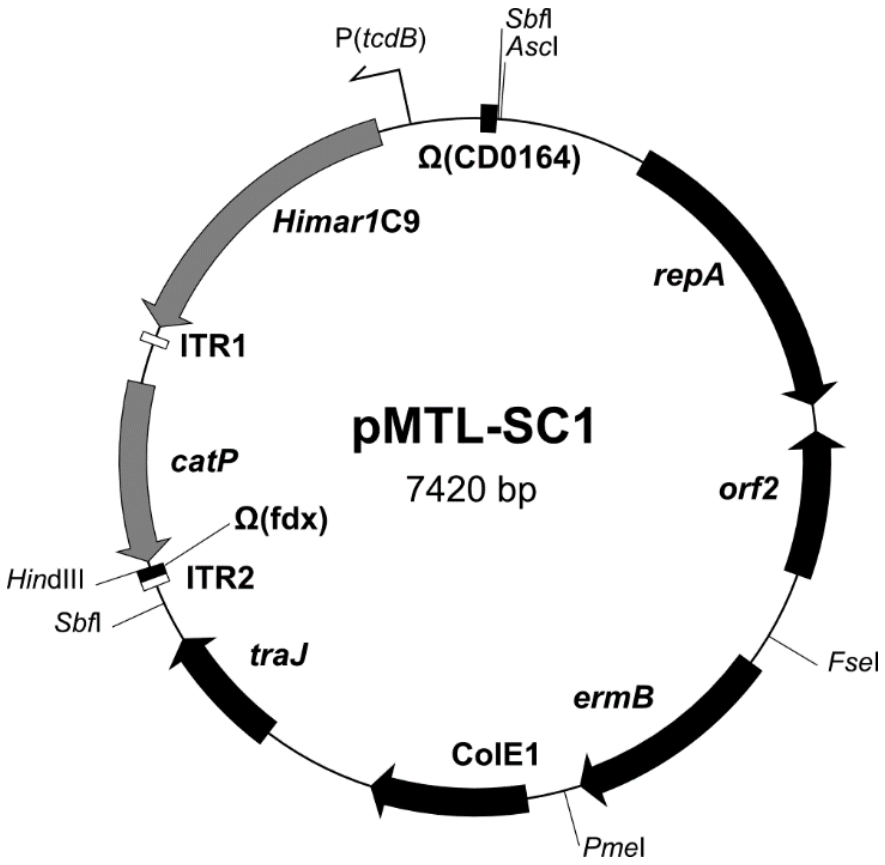


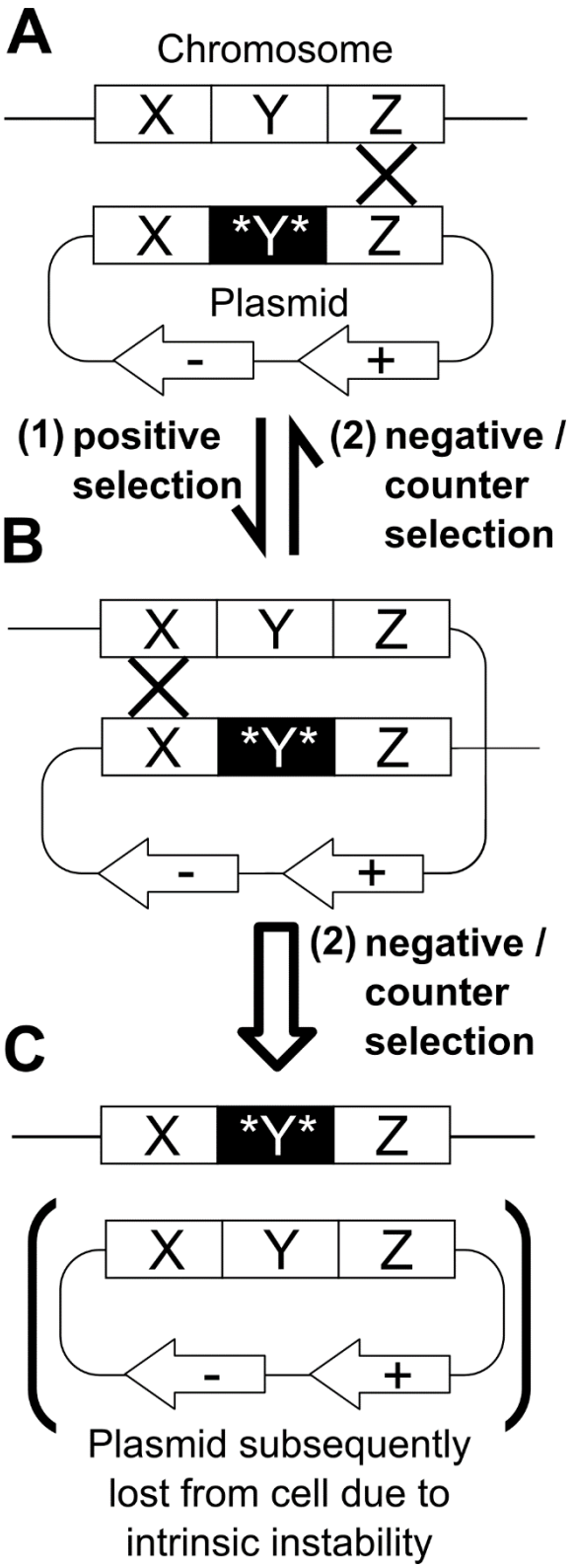
Figure 1-5: Schematic of the *mariner*-transposon plasmid[97]. The plasmid contains a Gram-positive replicon (*repA* & *orf2*) from *C. botulinum*, an erythromycin resistance gene (*ermB*) for transconjugant selection, a Gram-negative replicon (ColE1) and the conjugal transfer function (*traJ*). The mariner-transposon element contains the Himar1C9 upstream of the toxin B promoter PtcdB of *C. difficile*. The transposable element is a thiamphenicol resistance gene (*catP*) and is flanked by two inverted terminal repeats.

1.3.3 Allelic exchange using codA as a negative selection marker

In 2012, Cartman *et al.*[80] were able to repair variant *tcdC* genes by using allelic exchange. This method utilises both positive and counter-selection markers to identify possible mutant clones. Previously, counter-selection markers have only been described in *Clostridium thermocellium* and *C. perfringens*, however, these markers have chromosomal homologues and therefore can only be used in mutant background strains[101,102]. To overcome the requirement to create a *C. difficile* strain with a mutant background the cytosine deaminase gene (*codA*) of *Escherichia coli* was utilised to create a heterologous counter-selection marker. Ordinarily, cytosine deaminase converts cytosine to uracil,

however, it is also able to convert the cytosine analogue 5-fluorocytosine (FC) into the toxic derivative 5-fluorouracil (FU). FU is phosphorylated by uracil phosphoribosyl transferase and after a number of subsequent steps, results in the misincorporation of fluorinated nucleotides into DNA and RNA resulting in toxicity[103]. Figure 1.5 describes the system in which the homologous recombination events occur and how to select for them. Complementation of the altered gene can be achieved through the same method, to distinguish between wildtype and complemented strains a single nucleotide variant (SNP) resulting in both a silent mutation and creating a new restriction site can be utilised.

Figure 1-6: Two-step allelic exchange. Consider a chromosomal allele to be divided into three regions [X, Y and Z]. (A) A recombinant allele, in which the sequence of [*Y*] differs from that in the chromosome [Y], is introduced on a plasmid vector (note, [*Y*] is omitted altogether if sequence is to be deleted from the chromosome). The plasmid vector is replication deficient and carries both a positive selection marker [+] (usually an antibiotic resistance marker) and a counter-selection marker [–]. (B) Positive selection enriches for single cross-over recombinant clones in which a homologous recombination event at either [X] or [Z] results in integration of the plasmid into the chromosome. (C) Subsequently, counter-selection selects for clones in which a second recombination event has occurred at either [X] or [Z] (resulting in plasmid excision from the chromosome) and which have lost the plasmid vector from the cell (due to its inherent instability). At the counter-selection stage, both wild-type revertants and double cross-over recombinant clones are isolated, as the second recombination event may occur in either the same or opposite region to the first (i.e., [X] or [Z]). (Taken from Cartman et al.[80] supplementary material with permission).



1.4 Aims of this project.

The overall aim of this thesis was to improve our fundamental knowledge of the molecular pathogenesis of *C. difficile*. This was accomplished through a multifaceted approach that involved investigating; the genetic differences between strains, improving DNA transfer, exploring how the toxin genes *tcdA* and *tcdB* are regulated and investigating what mechanisms are involved in patients who suffer recurrent CDI infections.

This thesis contains three results chapters:

- Chapter Three will investigate regulators of TcdA and TcdB production using both forward and reverse genetics. An R20291*Δtcd*R strain will be studied to ascertain the role of TcdR in toxin production. Random mutagenesis will also be utilised to identify putative genes involved in toxin production either directly or indirectly.
- Chapter Four will focus on the genetics of *C. difficile* strain VPI 10463, a high toxin producing strain. This will involve whole genome sequencing of this strain, comparing genotypes with other well studied strains to predict the reasons for increased toxin production. This chapter will also focus on improving DNA transfer in to strains VPI 10463 and R20291.
- Chapter Five will examine *C. difficile* isolates collected from patients with diagnosed CDI. These patient samples will be studied for co-infection of multiple *C. difficile* ribotypes, reduced antibiotic susceptibility and to identify SNP's that may incur a fitness benefit that could result in recurrence of infection.

Chapter Two Materials and Methods

2.1 Buffer, Solutions, Culture Media and Growth Conditions

2.1.1 Buffer and Solutions

The following list of solutions are stated at stock concentrations, further dilutions may have occurred for use.

10% SDS

100g sodium dodecyl sulphate
Final vol. 1 L dH_2O

3M NaAC

24.6g Sodium Acetate
Adjust to pH 5.2
Final vol. 100 mL. dH_2O

0.5M EDTA

186.1g EDTA disodium salt
Adjust to pH 8.0
Final vol. 1L dH_2O

10x TAE

48.4g of Tris
11.4 mL of glacial acetic acid
20 mL 0.5M EDTA
Final vol. 1 L dH_2O

0.4M NaOH

16g Sodium hydroxide
Final vol. 1 L dH_2O

20x SSC

175.3g Sodium chloride
88.2g Sodium citrate
Adjust to pH 7.0
Final vol. 1 L dH_2O

2x Wash solution

2x SSC
0.1% SDS
Final vol. 500 mL dH_2O

0.5x Wash solution

0.5x SSC
0.1% SDS
Final vol. 500 mL dH_2O

5x Maleic acid buffer

58.04 g Maleic acid
43.83 g Sodium Chloride
36g Sodium hydroxide
Adjust to pH 7.5
Final vol. 1 L dH_2O

1x Maleic acid buffer-T

1x Maleic acid buffer
0.3% Tween 20
Final vol. 500 mL dH_2O

1x Blocking buffer

10x Blocking buffer solution (Sigma-Aldrich)
1x Maleic acid buffer
Final vol. 60 mL dH_2O

Anti-DIG Ab Probe

30 mL 1x blocking buffer
3 µL anti-DIG Ab

1x Detection buffer

100mM Tris-HCl
100mM NaCl
Final vol. 30 mL

2.1.2 Culture Media

Media used to culture bacteria are listed below. Solid media plates were made with the addition of 1% w/v agar (Oxoid No. 1) unless otherwise stated. All media were sterilised by autoclaving at 121°C for 15 min at 100 pKa. Ingredients were sourced form Oxoid unless otherwise stated.

- Brain Heart Infusion Supplemented (BHIS); 37g brain heart infusion, 20g yeast extract and 1g L-cysteine in 1L dH_2O.
- Tryptose Yeast Extract medium (TY)[104]; 30g tryptose, 20g yeast extract and 1g thioglycolate in 1L dH_2O. When required 1g lichenan was added to the media prior to autoclaving.
- Luria-Bertani (LB)[105]; 10g Bacto tryptose (BD Biosciences), 5g yeast extract, 10g sodium chloride in 1L dH_2O. For solid media Oxoid No. 1 added at a concentration of 1.5%.
- Cycloserine Cefoxitin Egg Yolk (Lab M) (CCEY)[106]; 48g premixed CCEY in 1L dH_2O. Post autoclaving 40 mL (4%) egg yolk emulsion was added to the media.
- Wilkins – Chalgren Anaerobe Agar (WCAA); consisted of 43g premixed WCAA in 1L dH_2O. Post autoclaving 50 mL (5%) defibrinated horse blood added to the media.
- Cycloserine Cefoxitin Fructose Broth (CCFB)[107]; 40g proteose peptone, 5g sodium phosphate dibasic, 1g potassium phosphate monobasic, 2g sodium chloride, 0.1g magnesium sulphate, 6g fructose, 1g sodium taurocholate, 0.5g L-cysteine, 0.03g neutral red. For solid media Oxoid No. 1 added at a concentration of 2% and neutral red was omitted.
- Cycloserine Cefoxitin Mannitol Broth with Taurocholate and Lysozyme (CCMB-TAL)[108]; 40g proteose peptone, 5g sodium phosphate dibasic, 1g potassium phosphate monobasic, 2g sodium chloride, 0.1g magnesium sulphate, 6g mannitol, 1g sodium taurocholate, 0.5g L-cysteine, 0.5g Lysozyme,0.03g neutral red.
- SOC medium (Invitrogen) consists of 2% tryptone, 0.5% yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulphate, and 20 mM glucose.
- Clostridium difficile minimal media (CDMM)[97].
- ChromID *C. difficile* (bioMérieux); a chromogenic agar plate that enables *C. difficile* to be identified quickly from mixed cultures.
- Tryptone Soy Agar (Thermo Scientific) with 5% sheep blood.

2.1.3 Antibiotic Media Supplements

Antibiotics were prepared as recommended by Sambrook and Russell[109], filter sterilised and stored under the recommended conditions. Antibiotics were used in the following

working concentrations; Ampicillin 100 µg/mL, Chloramphenicol 25 µg/mL, Cycloserine 250 µg/mL, Cefoxitin 8 µg/mL, Erythromycin 10 µg/mL, Kanamycin 50 µg/mL, Lincomycin 20 µg/mL and Thiamphenicol 15 µg/mL.

2.1.4 Growth Conditions

C. difficile were routinely cultured in an anaerobic cabinet (Don Whitley, UK) containing CO_2:H_2:N_2 (80:10:10 v/v/v) at 37°C. Media were reduced in anaerobic conditions for a minimum of 4h or 8h for plates and broths, respectively, before use. *E. coli* strains were cultured in aerobic conditions at 37°C, broth cultures were shaken at 200 rpm.

2.2 Bacterial Strains and Plasmids

2.2.1 Bacterial Strains and Storage

Bacterial reference strains and constructed mutants are listed in Table 2-1. Two different R20291 strains were used in this study, NM-R20291 (Synthetic Biology Research Centre, University of Nottingham) and BW-R20291 (London school of Hygiene and Tropical Medicine). Briefly, NM-R20291 has improved growth kinetics when grown in either BHIS broth or minimal media containing glucose, fructose and mannitol. NM-R20291 demonstrates a reduced motility, produces more toxin, sporulation occurs later and can form biofilms better than BW-R20291. Full details given in Appendix One.

During this study a number of *C. difficile* strains were isolated from clinical stool samples, these strains are described in Chapter Five of this thesis.

C. difficile strains were stored at -80°C in 10% glycerol and BHIS broth stocks. *E. coli* were stored at -80°C in CryoBank tubes (Copan) as directed by the manufacturer.

Table 2-1: Bacterial strains used in this study.

Strains	Characteristics	Origin
E. coli		
TOP10	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lacX74 nupG recA1 araD139* Δ(*ara-leu*)*7697 galE15 galK16 rpsL*(Str^r) *endA1* λ$^-$	Invitrogen
CA434	*E. coli* HB101 [F$^-$ *mcrB mrr hsdS20*(r_B^- m_B^-) *recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Sm^r) *glnV44*λ$^-$] with plasmid R702	[105,110]
NEB express	*fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^S*)2 [dcm] R(zgb-210::Tn10--*Tet^S*) endA1 Δ(mcrC-mrr)114::IS10*	NEB
C. difficile		
630	Wild type, PCR ribotype 012	
NM-R20291*	Wild type, PCR ribotype 027 (Stoke Mandeville outbreak strain)	Jon Brazier (Anaerobe Reference Laboratory, Cardiff, United Kingdom)
BW-R20291*	Wild type, PCR ribotype 027 (Stoke Mandeville outbreak strain)	Brendan Wren (London School of Hygiene and Tropical Medicine, London, United Kingdom)
VPI 10463	Wild type, PCR ribotype 087	ACTC 43255
M120	Wild type, PCR ribotype 078, non-motile strain	Brendan Wren (London School of Hygiene and Tropical Medicine, London, United Kingdom)
CRG4603	BW-R20291::*hsdR*(*ermB*)	This study
CRG4605	VPI 10463:: *hsdR*(*ermB*)	This study
CRG2521	NM-R20291Δ*tcdR*	Clostridia Research Group (unpublished)
CRG3972	NM-R20291::*tcdR*(EcoRI)	Clostridia Research Group (unpublished)
RS-7	NM-R20291Δ*tcdB*::*licB*	Clostridia Research Group (unpublished)
BBM4	NM-R20291 *mariner*-transposon mutant, insertions in genes R20291_2908 & R20291_jag	This study
CRG4250	NM-R20291::*2908*(*ermB*)	This study
CRG4251	NM-R20291::*jag*(*ermB*)	This study

*For specific details on the differences between these strains see Appendix One.

2.2.2 Plasmids

All plasmids were stored at -20°C in dH_2O.

Table 2-2: Plasmids used in this study.

Plasmid	Description	Reference
pMTL82151	*E. coli – C. difficile* Shuttle Vector (pBP1, *catP*, ColE1, *traJ*)	[111]
pMTL83151	*E. coli – C. difficile* Shuttle Vector (pCB102, *catP*, ColE1, *traJ*)	[111]
pMTL84151	*E. coli – C. difficile* Shuttle Vector (pCD6, *catP*, ColE1, *traJ*)	[111]
pMTL85151	*E. coli – C. difficile* Shuttle Vector (pIM13, *catP*, ColE1, *traJ*)	[111]
pMTL84152	*E. coli – C. difficile* Shuttle Vector (pCD6, *catP*, ColE1, *traJ*, P_{thl})	[111]
pMTL007-E2::Cdi-R20291_2908-1042s::CT(ermB)	ClosTron plasmid with intron designed to insert into gene 2908 of R20291	This study
pMTL007-E2::Cdi-R20291_jag::96s::CT(ermB)	ClosTron plasmid with intron designed to insert into gene jag of R20291	This study
pMTL007-E2::Cdi-R-subunit-527s::CT(ermB)	ClosTron plasmid with intron designed to insert into gene *hsdR* of R20291and VPI 10463	This study
pMTL007-E2::Cdi-BI9_3942::CT(ermB)	ClosTron plasmid with intron designed to insert into gene *Ga0114281_1134* of VPI 10463	This study
pMTLSC7215	Vector for homologous recombination using *codA* as a negative selection marker. Gram positive replicon - pBP1	[80]
pBSK::ΔhsdR	Storage plasmid for Δ*hsdR* exchange cassette.	This study
pMTLSC7215::ΔhsdR	Allelic exchange plasmid for creation of Δ*hsdR* strains.	This study
pMTL84151::p2908	Complementation plasmid for R20291::*2908*(*ermB*)	This study
pMTL84152::2908	Over expression plasmid for R20291::*2908*(*ermB*)	This study
pMTL-SC0	*pMTL82151 containing mariner*-transposon. No promoter driving expression of the transposon.	[97]
pMTL-SC1	*pMTL82151 containing mariner*-transposon	[97]

2.3 Polymerase Chain Reaction (PCR)

For all PCR reactions annealing temperatures were calculated using the NEB Tm calculator (http://tmcalculator.neb.com/#!/) and extension times were estimated by predicted product length.

2.3.1 General Protocol for Colony PCR

Colony PCR was performed using OneTaq Quick Load 2X Master Mix (NEB) at a final volume of 25 µL containing 0.5µM of each primer. Three to four colonies were emulsified into 30 µL molecular grade water and heated to 94°C for 20 min, 1 µL was added to the reaction to serve as a template. PCR conditions were; 94°C for 30s, 30 cycles of 94°C for 15s, 45 - 68°C for 30s and 68°C for 1 min/Kb followed by a final extension step of 68°C for 5 min.

2.3.2 General Protocol for PCR with a Non-Proofreading DNA Polymerase

Non-proofreading PCR was performed at a final volume of 25 µL containing 0.5µM of each primer, 2.5 units of OneTaq (NEB) and 12.5 µL FailSafe PCR 2X PreMix E (Cambio). Template DNA was used at a concentration of 100 ng/µL. PCR conditions were; 94°C for 30s, 30 cycles of 94°C for 15s, 45 - 68°C for 30s and 68°C for 1 min/Kb followed by a final extension step of 68°C for 5 min.

2.3.3 General Protocol for PCR with a Proofreading DNA Polymerase

Proofreading PCR was performed at a final volume of 25 µL containing 0.5µM of each primer, 0.5 units of Q5 (NEB) and 12.5 µL FailSafe PCR 2X PreMix E (Cambio). Template DNA was used at a concentration of 100 ng/µL. PCR conditions were; 98°C for 30s, 30 cycles of 98°C for 10s, 50 - 72°C for 30s and 72°C for 30 s/Kb followed by a final extension step of 72°C for 2 min.

2.3.4 Inverse PCR

Genomic DNA at a concentration of 50 ng/µL was digested overnight by HindIII (NEB) and ligated at a concentration of 5 ng/µL as described above. PCR was performed using 2X KOD Hot Start PCR Master Mix (Novagen) at a final volume of 25 µL containing 0.5µM of each primer and 10 µL of ligated DNA template. PCR conditions were; 95°C for 2 min, 35 cycles of 95°C for 20s, 50°C for 10s and 70°C for 3 min 25s followed by a final extension step of 70°C for 10 min.

2.3.5 Primers

Primers were designed using Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and synthesised by Sigma-Aldridge. A full list of primers used in this study can be found in Table 2-3.

Table 2-3: Primers used in this study

Primer	Sequence	Reference
ClosTron		
EBS universal	CGAAATTAGAAACTTGCGTTCAGTAAAC	[68]
ErmRAM-F	ACGCGTTATATTGATAAAAATAATAATAGTGGG	[68]
ErmRAM-R	ACGCGTGCGACTCATAGAATTATTTCCTCCCG	[68]
EBS2	TGAACGCAAGTTTCTAATTTCGATTATATTTCGATA GAGGAAAGTGTCT	[68]
Sa1I-R1	CAGATTGTACAAATGTGGTGATAACAGATAAGTCG TAGAATCTAACTTACCTTTCTTTGT	[68]
M13F	TGTAAAACGACGGCCAGT	[68]
M13R	CAGGAAACAGCTATGACC	[68]
Chapter Three		
tcdR-Fs1	AAATTATCTTAAGAGAGGAGAAGTTTCTAAAATAT AAAAAGG	This study
tcdR-Rs2	GGTAAATTATTATTCTTTAGCTCTAATACTTCTGTA ACTAGG	This study
catP-INV-F1	TAAATCATTTTTAGCAGATTATGAAAGTGATACGC AACGGTATGG	[97]
catP-INV-R1	TATTGTATAGCTTGGTATCATCTCATCATATATCCC CAATTCCC	[97]
catP-INV-R2	TATTTGTGTGATATCCACTTTAACGGTCATGCTGTA GGTACAAGG	[97]
catP-F1	GGCAAGTGTTCAAGAAGTTATTAAGTCGGGAGTGC AGTCGAAGTGG	[97]
catP-R1	TGAAGTTAACTATTTATCAATTCCTGCAATTCGTTT ACAAAACGG	[97]
jag_F1	AAACTTGCAGCTTCTTTGCCTT	This study
oxa_R1	GCAGACTAAGTCAACAAAGGCAA	This study
R20291_2908-C_F1	TTTTTTGCGGCCGCTTTACTTTAATAAAATAAAAAT AGGTATAAAAA	This study
R20291_2908-C_R1	AAAAAACTCGAGTTATTTTATGGAAAATATTACTT CATTTAACTCTATAT	This study
R20291_2908-C_F2	TTTTTTGAATTCATGAAAGCAGGAGAAATTGAATT TCTTAGTTATTTAGA	This study
CDR20291_0447 F	AGAGCAACTAGAATTGGCGGT	This study
CDR20291_0447 R	GCTTGGTGCTGGCTCATGTA	This study
CDR20291_1065 F	AAGGGATTACCTCCTACACCAA	This study
CDR20291_1065 R	ACCAATTGCACATCCTACTTCT	This study
CDR20291_1086 F	CAAGTGGAAGCGCCAAACTT	This study
CDR20291_1086 R	CTCCACAATTTGGGCATTCCA	This study
CDR20291_1439 F	TCAAGCAATTGTAGATGGGTGGA	This study
CDR20291_1439 R	GTTCGGCGTACATCATCAACA	This study

CDR20291_0853 F	TGGAGGTAACACTTTCTTGTGT	This study
CDR20291_0853 R	GGTTGACTATATACCTCTCCTGCT	This study
CDR20291_2848 F	GCCACATGGATATGGAACTGGT	This study
CDR20291_2848 R	AGTTGTATCACCAGAAGCACG	This study
Chapter Four		
hsdR_Fs1	TTCCCAACCTGCAAGTTTA	This study
hsdR-Rs1	AGAAGACTTTAGAGCTTGCAGT	This study
BI93941_Fs2	CTCCATCAGCAGAAGTGCAT	This study
BI93943_Rs1	GGGGCTAGAACAAAACCCAA	This study
Chapter Five		
P3	CTGGGGTGAAGTCGTAACAAGG	[48]
P5	GCGCCCTTTGTAGCTTGACC	[48]

2.4 Preparation and Manipulation of DNA

2.4.1 Preparation of Chromosomal DNA

Chromosomal DNA was extracted from stationary phase cultures after centrifugation at 12,000 x *g*. Culture pellets were pre-treated with 180 µL 10 mg/mL lysozyme at 37°C for 30 min followed by 25 µL 10 mg/mL proteinase K solution, 85 µl dH_2O and 110 µl 10% (w/v) SDS solution incubated at 65°C for 30 min. For general PCR screening DNA was extracted using the GenElute Bacterial Genomic DNA Kit (Sigma Aldrich). Final elution of DNA was into 50 µL molecular grade water (Thermo Scientific) at 50°C. For high quality DNA, phenol chloroform extraction was utilised. Equal volumes of Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v/v) were added to pre-treated culture pellets and mixed by inversion. Liquid was transferred to phase lock tubes and centrifuged at 20,000 x *g* for 3 min. The top layer was transferred into a fresh phase lock tube and the process repeated a further two times. The final top layer was transferred into a 1.5 mL Eppendorf tube containing 40 µL 3M NaAC and 800 µL ice cold 100% EtOH, mixed gently by inversion and placed onto ice. DNA was transferred into 1 mL ice cold 70% EtOH by

5 mL glass pipette and then into a clean 1.5 mL Eppendorf tube for evaporation. DNA was rehydrated in 50 µL molecular grade water (Thermo Scientific).

2.4.2 Plasmid Extraction

Plasmids were extracted from 1 mL stationary phase cultures after centrifugation at 12,000 *x* g using the GenElute HP Plasmid Mini Prep Kit (Sigma Aldrich) as directed by the manufacturer. Final elution of plasmid was into 50 µL molecular grade water (Thermo Scientific) at 50°C.

2.4.3 Restriction Digest

Restriction digests were performed on ice with DNA at a concentration of 1µg, 1 unit of restriction endonuclease and 1x Cut Smart Reaction Buffer (50mM Potassium Acetate, 20mM Tris-Acetate, 10mM Magnesium Acetate and 100 µg/mL BSA) at a final volume of 25 µL. Reactions were incubated at 37°C for 2-3 h and heat inactivated at 65°C for 30 min. All enzymes were sourced from NEB and were stored according to the manufacturer's instructions at -20°C.

2.4.4 DNA ligation

DNA ligations were performed on ice at a molar ratio of 1:3 (vector to insert), 1 unit T4 ligase (NEB) and 1x T4 DNA ligase buffer (50mM Tris-HCl, 10mM Magnesium Chloride, 1mM dithiothreitol) at a final volume 25 µL. Exact concentrations were calculated using NEB BioCalculator (http://nebiocalculator.neb.com/#!/). Reactions were incubated at 16°C overnight and heat inactivated at 65°C for 30 min. Plasmids were dialysed into dH_2O using 0.025 µm millipore filters (Millipore Corporation) for 1 h before transformation.

2.4.5 DNA Analysis by Agarose Gel Electrophoresis

Electrophoresis for the visualisation of DNA samples was performed using 1% agarose gels (Sigma-Aldrich) in 1x TAE buffer. SyberSafe (Invitrogen) was added at a 1x

concentration. to allow for visualisation of DNA fragments. Prior to loading on to the gel, 6x loading dye (NEB) were added to DNA samples. For estimation of fragment size an appropriate ladder (NEB) containing 6x loading dye was included into each run. Gels were run at 100 V until appropriate migration of loading dye was observed. Visualisation of PCR products and image capture was carried out on Gel Doc XR system (BioRad).

2.4.6 DNA Purification from Agarose Gels

DNA fragments were extracted from agarose gels using a scalpel and purified using the Sigma Gel Extraction Kit according to the manufactures instructions. Final elution of fragments was into 50 µL molecular grade water (Thermo Scientific) at 50°C.

2.4.7 Cleaning of PCR product

PCR products were purified using the Sigma PCR Clean-up Kit according to the manufactures instructions. Final elution of fragments was into 50 µL molecular grade water (Thermo Scientific) at 50°C.

2.4.8 DNA Quantification

DNA was quantified either by micro spectrophotometry measuring absorbance at 260nm using the NanoDrop 2000 (Thermo Scientific) or by fluorometric quantification using the Qubit with the Qubit deDNA HS Assay Kit (Life Technologies).

2.5 Cloning

2.5.1 Preparation of Electro-competent E. coli

A 1:100 dilution of fresh overnight *E. coli* cells were added to 200 mL pre-warmed (37°C) LB broth. Seed cultures were grown at 37°C with 200 rpm shaking until mid-log phase (OD_{600} 0.5 – 1.0). Cells were then harvested by centrifugation at 4°C for 15 min at 4000 *x* g after cooling on ice for 30 min. Pellets were resuspended in 60 mL ice cold dH_2O and cells harvested as before. Pellets were resuspended in 10 mL 10% glycerol and harvested

as before. The final pellet was resuspended in 500 µL 10% glycerol, aliquot into 50 µL volumes and stored at -80°C.

2.5.2 Transformation of Bacteria via Electroporation

Electro-competent *E. coli* cells were thawed on ice before the addition of 1 – 10 ng plasmid DNA. Transformation of plasmid DNA was performed using a Bio Rad MicroPulser using a 0.2 cm electroporation cuvette, 2.5 kV voltage, 25 µF capacitance and 200 Ω resistance. After which 300 µL SOC medium (Sigma Aldrich) was added to the cuvette and the content aliquoted to a fresh Eppendorf tube and incubated at 37°C for 1 h at 200 rpm. The resulting mixture was then plated onto LB agar containing appropriate antibiotics.

2.5.3 Conjugation of Plasmid DNA into C. difficile

Conjugation of plasmid DNA into *C. difficile* was performed as previously described[112]. Briefly, 1 mL stationary phase donor cells were centrifuged at 4000 *x* g and pellets washed in PBS to remove trace antibiotics. Centrifugation of donor cells was repeated and resulting pellets were transferred to the ANO_2 chamber and resuspended in 200 µL of recipient cells taken from an overnight culture. The resulting mixture was inoculated onto BHIS agar in discrete spots and incubated at 37°C. For conjugation into *C. difficile* 630 suspensions were incubated for 4 – 8 h and for conjugation into *C. difficile* R20291 and VPI 10463 suspensions were incubated for a minimum of 18 h. Mating mixtures were then harvested into 500 µL PBS and inoculated onto BHIS agar containing antibiotics appropriate to counter select the donor strain. Resulting transconjugants were visible and ready for purification after 48 – 72 h. To calculate conjugation efficiency, the following calculation was applied:

$$\frac{total\ number\ of\ transconjugants}{total\ CFU\ C.difficile}$$

2.6 Genetic Manipulation of Chromosomal DNA

2.6.1 ClosTron Knock-Out Mutagenesis

Plasmids for the insertional inactivation of specific genes were designed using the Perutka method available free of charge at http://www.clostron.com[96]. Plasmids were constructed by DNA2.0. Plasmids were conjugated into recipient *C. difficile* strains as described above. Transconjugants selected for using BHIS agar containing cycloserine, cefoxitin and thiamphenicol were restreaked to purity on the same media. After which, pure isolates were inoculated onto BHIS agar containing cycloserine, cefoxitin and lincomycin to select for clones that have undergone splicing of the group I intron from the RAM to leave the functional *ermB* gene. Lincomycin resistant colonies were restreaked to purity and screened for correct insertion using EBS universal primer and a primer designed to flank the insertion site (Table 2-3). PCR products were also sequenced using the EBS universal primer to verify intron insertion to the correct location. Single intron insertion was confirmed by Southern blot using primers EBS2 and Intron SalI-R1 (Table 2-3) for probe synthesis.

2.6.2 Allelic Exchange using CodA as a Negative Selection Marker

Allelic exchange was performed as previously described[80] with slight modifications. Exchange cassettes flanked by restriction sites XhoI and SacI were constructed by Biomatik and provided on pBMH vectors (Table 2-3). Before cloning, exchange cassettes were amplified using universal M13 primers (Table 2-3) and resulting PCR products were cloned into pMTLSC7215 or pMTLSC7315 using XhoI and SacI for *C. difficile* R20291 and *C. difficile* VPI 10463 or *C. difficile* 630 respectively. Resultant plasmids were stored in *E. coli* TOP10 and sequence verified using primers M13_F and CodA_Seq_F (Table 2-3). After re-transformation into *E. coli* CA434, resulting *codA* plasmids and a vector only control (i.e. pMTLSC7215 or pMTLSC7315) were conjugated into donor strains as

described above. Transconjugants were selected for using BHIS agar containing cycloserine, cefoxitin and thiamphenicol. Resulting transconjugants were restreaked onto the same media and faster growing single colonies were selected as potential single crossovers and restreaked to purity. Single crossover events were screened for using a flanking chromosomal primer and a plasmid specific primer (Table 2-3) for each homology arm. Confirmed single crossover clones were then streaked on the non-selective medium (i.e. BHIS) and allowed to grow for up to four days. After which colonies were harvested into 500 µL PBS, serially diluted ($1x10^{-1}$ to $1x10^{-5}$) and plated onto *C. difficile* minimal media containing 5-fluorocytosine and allowed to incubate for 48h. Well defined single colonies were then patch plated onto BHIS agar and BHIS agar containing thiamphenicol. Colonies that were thiamphenicol sensitive were restreaked onto BHIS agar and then screened by chromosome flanking PCR to identify double crossover mutants. Confirmed clones were then stored as described above.

2.6.3 Random Mutagenesis using mariner-Transposon

Random mutagenesis was performed as previously described[97] with slight modifications. Plasmids pMTL-SC0 and pMTL-SC1 (Table 2-2) were transformed into *E. coli* Top10 and conjugated in to *C. difficile* RS-7 (Table 2-1) as described above. Transconjugants were selected for on BHIS agar containing cycloserine, cefoxitin and lincomycin. Transconjugant colonies were then restreaked onto TY agar containing cycloserine, cefoxitin and lincomycin. After five days incubation, growth was harvested into BHIS 10% glycerol until further use, this creates passage zero. Transconjugant stocks were cultured onto TY agar containing cycloserine, cefoxitin and lincomycin, one replicate for the pMTL-SC0 control and five replicates for the pMTL-SC1 transposons. After overnight incubation, growth was harvested in to 1 mL PBS and 10-fold serial dilutions to 10^{-8} were made. Subsequently 100 µL of each dilution was plated onto BHIS agar containing

cycloserine, cefoxitin and thiamphenicol, to select for transposon mutants and count mutant CFU, and 3 x 20 µL spots were inoculated onto ¼ BHIS agar for calculation of total CFU. To allow for the frequency of transposition to be calculated, total CFU and mutant CFU were enumerated at a dilution where approximately 50 and 100 colonies were present respectively. The calculation was as follows and performed on both control and mutant populations:

$$\frac{Mutant\ CFU/mL}{Total\ CFU/mL}$$

To ensure mutant diversity three passages of this procedure was performed. For creation of the next passage, colonies were harvested from the dilution plate below to the one used for the mutant CFU calculation and stored in BHIS 10% glycerol.

To confirm mutant diversity, 10 mutant colonies were taken from the dilution plate used to calculate mutant CFU and restreaked onto BHIS agar containing cycloserine, cefoxitin and thiamphenicol. DNA extractions were performed on resultant cultures. Inverse PCR was performed as described above to determine if double insertions were present that had occurred during transposition. Once all passages had been screened the most diverse mutant pool was selected by discerning the passage with a high transposition frequency with a low double insertion rate.

Toxin phenotypes were screened for using a plate lichenase reaction, 100 µL of stock mutant library was inoculated onto BHIS agar containing cycloserine, cefoxitin and lincomycin and incubated for 24 h. Colonies were then patch plated twice onto TY agar containing 0.2% lichenan, one for analysis at 24 h and one for 48 h. Plates were stained with 0.2% Congo Red (Sigma-Aldrich) solution and zones of clearance were examined. Interesting phenotypes were categorised as larger, smaller or no zone of clearance and

colonies were re-streaked onto BHIS agar containing cycloserine, cefoxitin and thiamphenicol for phenotype and PCR conformation. Phenotypes were confirmed by inoculating 20 µL volumes of overnight culture onto TY agar containing 0.2% lichenan and analysed as per the screening method. To verify the location of the transposition in interesting phenotypes, inverse PCR was performed as described above. Purified gel extracts were sequenced with primer catP-INV-R2 (Table 2-3). Single insertions were confirmed by Southern blot using catP-F1 and catP-R1 primers (Table 2-3) for probe synthesis. Toxin phenotype analysis was confirmed in the wild type by using the ClosTron group two intron gene knockout system.

2.6.4 Southern Blot

To generate a Southern Blot probe a digoxigenin-dUTP randomly labelled DNA product was produced using DIG-high prime (Sigma-Aldrich) and a standard Phusion PCR. In this work probes to detect *ermB* insertion of ClosTron mutants were created using EBS2 and Sa1I-R1 primers and probes to detect *catP* insertion of the *mariner*-transposon were created using CatP_F1/R1 (Table 2-3). The probe was labelled by denaturing DNA (ca. 1 ug) at 100°C and snap freezing on ice before adding 4 µL DIG-High Prime and incubating overnight at 4°C. The reaction is halted using 2 µL 0.2M EDTA and heating to 65°C or 10 min. The ladder probe is created using the same method except 1 µL λ DNA/HindIII marker (Promega) and 14 µL H_2O is used as a template.

DNA was isolated using phenol-chloroform extraction and quality checked on 1% (w/v) gel. Genomic and plasmid DNA at 2µg was digested with HindIII and/or EcoRI overnight in separate reactions. Control plasmids for each experiment were pMTL007C-E2 (ClosTron) or pMTL-SC1 (*mariner*-transposon). Digested DNA checked on 0.8% agarose gel prior to blotting procedures.

DNA transfer on to the Nitrocellulose membrane was achieved by laying blotting paper into 0.4M NaOH to create a wick. The gel was positioned on top of the wick followed by the Nitrocellulose membrane. This was followed by several blotting pads the first being pre-wet in 0.4M NaOH. To aid in the transfer a weight (ca. 500 g) was placed on top the blotting paper and was left for at least 3 h.

The nitrocellulose membrane is UV fixed for 2 min on each side and then washed in 2x SSC to remove excess 0.2M NaOH. The membrane was then placed into a hybridisation tube (DNA side in), 10 mL reconstituted DIG-Easy Hyb (Sigma-Aldrich) was added and incubated at 42°C for 1 h. Probes were prepared by boiling 20 µL labelled probe and 10 mL DIG-Easy Hyb for 10 min then snap freezing on ice to prevent reversion of ssDNA to dsDNA. The probe mixture was incubated on the membrane overnight at 42°C. Following this a number of washes were performed, 2x low stringency washes using 100 mL 2x wash solution for 5 min at room temperature then 2x high stringency washes using 100 mL 0.5x wash solution for 15 min at 68°C. The membrane was then equilibrated using 100 mL Maleic acid buffer-T for 1 min at room temperature followed by a blocking step using 25 mL 1x blocking buffer for 30 min at room temperature. The membrane was then probed using Anti-DIG Ab (Sigma-Aldrich) diluted 1:10,000 in 1x blocking solution (3 µL in 30 mL) by incubation for 30 min at room temperature. This was followed by 2x washed suing 100 mL Maleic acid buffer-T for 15 min at room temperature and then membrane equilibration in 25 mL 1x detection buffer for 2 min at room temperature.

To develop the Southern blot the membrane was placed onto an acetate sheet. 1.5 mL CDSP (Sigma-Aldrich) was pipetted directly on to the membrane and then a second acetate sheet placed on top. The edges sealed and then the membrane incubated at 37°C for 10 min to enhance enzyme reaction. The membrane was then transferred to the dark room for developing into photo film.

2.6.5 Complementation and overexpression plasmids.

Promoters were predicted for complementation plasmids using BPROM[113] if unknown. Cassettes were amplified using wildtype genomic DNA from the parental strain utilising primers that added flanking restriction sites, NotI and XhoI for complementation and EcoRI and XhoI for overexpression. Complementation plasmids cassettes were cloned directly into the *lacZ*α multiple cloning site of pMTL84151. Over expression plasmids were also cloned directly into the *lacZ*α multiple cloning site but upstream of the P_{thl} of pMTL84152.

2.7 DNA sequencing and Analysis Techniques

2.7.1 DNA Sequencing

DNA Sanger sequencing was performed by Source Bioscience, Nottingham, UK.

2.7.2 Illumina MiSeq and Single Molecule, Real-Time (SMRT) sequencing

High quality DNA at 10µg was prepared by phenol/chloroform extraction with an additional RNAse step. For Illumina MiSeq, DNA was sent to either the University of Nottingham DeepSeq facility, Dr. Alan McNally at Nottingham Trent University or MicrobesNG Facility at University of Birmingham for library preparation and sequencing. For SMRT sequencing DNA was sent to McGill University and Genome Québec, Canada.

2.7.3 Bioinformatics

General Sequence Analysis

Sanger sequence data were analysed using GENtle (Last accessed: 11 November 2017 http://gentle.magnusmanske.de/) and Basic Local Alignment Search Tool (BLAST) (Last accessed: 13 August 2017 http://www.ncbi.nlm.nih.gov). Multiple sequence alignments were performed using Clustal Omega (Last accessed: 12 March 2017 http://www.ebi.ac.uk/Tools/msa/clustalo/). Protein analysis was performed using InterPro (Last accessed: 27 March 2017 http://www.ebi.ac.uk/interpro/) and predicted interactions

by String (Last accessed: 27 March 2017 http://string-db.org/). Protein amino acid variations and their effects on biological function were predicted using PROVEAN (Last accessed: 27 March 2017 http://provean.jcvi.org/). Prophage regions within genomes were identified using PHASTER (Phage Search Tool Enhanced Release)[114]. Whole genome browsing and annotation were performed using Artemis[115].

VPI 10463 Genome Map Assembly

Illumina MiSeq reads obtained for the VPI 10463 genome were trimmed using Scythe (https://github.com/vsbuffalo/scythe) and quality trimmed using Sickle (https://github.com/najoshi/sickle) by DeepSeq, University of Nottingham. PacBio RSII sequencing reads were corrected using proovread[116], a hybrid correction pipeline and were assembled using Canu by DeepSeq. Mapping to the VPI 10463 genome sequence was performed using the Burrows-Wheeler Aligner (BWA)[117] version 0.7.17 using standard settings and the PacBio index flag. Output files were then sorted and indexed using SamTools[118] version 1.5. Visualisation of BAM files was performed using CLC genomics workbench version 8.5.3 using standard settings. Annotation of the VPI 10463 genome were performed by the Joint Genomes Institute Integrated Microbial Genomes Pipeline (Last accessed: 18 September 2017 http://jgi.doe.gov/). Restriction site generation was performed using SiteFind[119]. Appendix Three provides all code performed by the user.

***De novo* assembly of clinical *C. difficile* strains**

De novo assembly and mapping reads to reference of Illumina short reads were perfumed using CLC genomics workbench V8.5.3 using standard settings. Contigs were rearranged using progressiveMauve[120] with strains 630 and R20291 as templates. Rapid annotation followed using Subsystem Technology (RAST)[121] and the Seed[122]. Fixed Ploidy Variant Detection was performed using CLC genomics workbench V8.5.3 using settings described

in Appendix Two. Genome comparisons were performed using either progressiveMauve[120] or Artemis Comparison Tool (ACT)[123].

Whole genome sequences used in this study; CD630 (AM180355), R20291 (FN545816) and M120 (FN665653). Statistical analysis was performed using GraphPad Prism V6.

2.8 Characterisation of *Clostridium difficile*.

2.8.1 Growth curve analysis

Cell growth was recorded over a 48 h time period, starting from a lag growth phase after culture in TY broth. This was achieved by culturing glycerol strain stocks on to BHIS agar supplemented with cefoxitin and cycloserine. Three colonies where picked into 1 mL TY broth and incubated for 8h. For all experiments individual 1 mL starter cultures were used to ensure biological replicates rather than technical replicates. In fresh TY broth, serial dilutions (10^{-1} to 10^{-8}) were made and incubated for 16h. The most dilute inoculum, with visible growth, was then used for the start of the assay and diluted 1 in 100 in TY broth. Optical density (OD_{600}) or CFU was then measured at 0, 3, 6, 9, 12, 24 and 48h. To count CFU, cell suspensions were serially diluted and plated onto BHIS agar, allowed to incubate for 24 h before enumeration.

2.8.2 Toxin ELISA

Toxin supernatants were collected during growth curve analysis at points 0, 3, 6, 9, 12, 24 and 48h. Supernatants of 1 mL were centrifuged at 12,000 x *g* for 1 min, filter sterilised (0.2µm pore size) and stored at -20°C until required but for no longer than a month. Toxin levels were measured using *C. difficile* TOX A/B II ELISA (TechLab) as recommended by the manufacturer, using a toxin A/B standard. Clostridium difficile toxin A and B standards (The Native Antigen Company) were used at a starting concentration 125 ng/mL and serially diluted 1:2 until a concentration of 1.95 ng/mL was reached.

2.8.3 Cell culture and toxin neutralisation

Each well of a 96-well microtiter plate was seeded with 100 µL of Vero (African green monkey kidney) cell suspension at a density 2 x 10^5 cells/mL to create a cell monolayer. The cell culture media, Dulbecco's modified Eagles medium (DMEM), contained 1% streptomycin/penicillin (vol/vol) and 10% fetal calf serum (vol/vol). To allow the creation of monolayers, cells were incubated at 37°C 5% CO_2 for 48h. To the cell monolayer 20 µL of toxin supernatant was added, if required serial dilutions were made in PBS. Cultures were then incubated for 24 h (37°C 5% CO_2) before examination by phase contrast microscopy (Nickon Eclipse TS100). End point titre was expressed as "1/toxin endpoint titre" due to the inverse relationship between the endpoint titre and the amount of toxin in a sample. This was defined as the first dilution in the succession where cell morphology could not be differentiated from the negative controls. Toxin neutralisation assays were performed on 96-well microtiter that were identical to the cell culture assay using *C. difficile* Tox-B Test (TechLab). Cells showing no toxic effect should be identical to the negative controls. Pure toxin B and the control provided in the *C. difficile* Tox-B Test kit were used as positive and/or negative controls depending on which assay was being performed.

2.8.4 Quantitative Lichenase assay

Lichenase activity was assessed through the incubation of the supernatant with a substrate containing a known concentration of lichenan for a set period of time. Cell growth was recorded over a 48 h time period, starting from a lag growth phase after culture in TY broth. This was achieved by culturing glycerol strain stocks on to BHIS agar supplemented with cefoxitin and cycloserine. Three colonies where picked into 1 mL TY broth and incubated for 8h. For all experiments individual 1 mL starter cultures were used to ensure biological replicates rather than technical replicates. In fresh TY broth, serial dilutions (10^{-1} to 10^{-8})

were made and incubated for 16h. The most dilute inoculum, with visible growth, was then used for the start of the assay and diluted 1 in 100 in TY broth. Lichenase supernatants were collected at points 0, 3, 6, 9, 12, 24 and 48h. Supernatants of 1.5 mL were centrifuged at 6,000 x *g* for 1 min and stored at -20°C until required. To start the assay, 300 µL of supernatant was added to an Eppendorf containing 200 µL of TY with 0.1% lichenan. A control of 300 µL of TY was added to an Eppendorf containing 200 µL of TY + 0.1% lichenan. Test and controls were incubated at 50°C for 40 min. Reactions were halted by addition of 100 µL of 0.2M Sodium carbonate. After which 50 µL of 0.1% (w:v) Congo red solution was added to each reaction, tubes were inverted several times, and then add 200 µL of 2M NaCl, to stabilise the colour. Absorbance reading at 530nm (using 500 µL TY broth, 100 µL 0.2M Sodium carbonate, 50 µL Congo red solution, 200 µL 2M NaCl as a blank).

2.8.5 Motility assay

Motility of strains was assessed using standard BHIS agar plates with 10 mL 0.3% BHIS top agar. Single colonies were stab inoculated into the top agar using a sterile toothpick. Plates were incubated for 24 h and zones measured using a bidirectional method. Averages of these measurements were taken as the total distance travelled.

2.8.6 Transmission Electron Microscopy

Transmission electron microscopy was performed by Dr. E. Stevenson and Denise Creasy. Cells were collected at mid-exponential phase and absorbed onto a Formvar-coated copper grid for 5 min, the excess removed with blotting paper. Fixation was achieved using 1% glutaraldehyde for 1 min, the grid was then washed three times with sterile distilled water. Cells were negatively stained with Uranyl Acetate for 30 s and then allowed to air dry. A JOEL JEM1010 transmission electron microscope was used for the visualisation of cells operating at 80kV.

2.9 Clinical Isolation of *Clostridium difficile*

2.9.1 Isolation of Clostridium difficile *from stool samples*

For each patient sample three CCEY agar plates were reduced in anaerobic conditions for a minimum of four hours. Stool samples were homogenised 1:1 with PBS and then heat shocked at 80°C for 15 mins, after which samples were centrifuged for five mins at 1500 *x* g. Samples were transferred to the anaerobic chamber where 50 µL of supernatant was inoculated on to CCEY in triplicate; plates were incubated for 48 h. Individual colonies that have the appearance of *C. difficile* were picked by toothpick and inoculated into a 96 microtiter plate containing 200 µL BHIS broth, one per well and up to 20 per patient sample. Microtiter plates were sealed with breathable sterile film and incubated overnight in anaerobic conditions. A separate 96 well microtiter plate contained 180 µL PCR grade H_2O where a 1:10 dilution was made from the overnight broth cultures. A drop of glycerol was then added to the broth cultures and resealed using fresh breathable sterile film and stored at -80°C. The H_2O culture mix was also covered in breathable sterile film and stored at -20°C to be used later as a PCR template.

2.9.2 Ribotyping

Ribotyping of the clinical isolates were adapted from O'Neil *et al*[124]. and Walk *et al*.[125] The PCR template created in 2.10.1 was DNA extracted by heating at 95°C for 20 min after initially being defrosted. The PCR process was completed as in [124] and products were visualised using a Qiaexcel (Qiagen) using the Qiaexcel DNA High Resolution Kit. Bands were inspected by eye and one example of each banding pattern per sample was stored for future testing. Each isolate stored was sent to the CDRN for conformation and ribotype assignment.

2.9.3 Antibiotic Minimum Inhibitory Concentration (MIC) Assay

MIC were determined by Minimum Inhibitory Concentration Evaluator (MICE) strips as described by Baubet *et al.*[126] with slight modifications. Colonies of *C. difficile* were resuspended in PBS to a McFarland standard of 1. Cultures were swabbed onto Wilkins – Chalgren Anaerobe Agar with 5% horse blood in three directions and air dried for 15 min. MICE strips of Vancomycin and Metronidazole were applied and incubated for 24h in anaerobic conditions. MICs were determined to be where the zone of complete inhibition intercepted with the scale. CD630 was used as a control with each batch tested. Intermediate or resistant MICs were repeated twice more for confirmation.

2.10 Statistical Analysis

All data and statistical analysis were carried out using GraphPad Prism version 7.03 or lower. Individual comparisons were analysed using two tailed students T-test to determine if Group A has a statistically higher or lower value than Group B, a p value of <0.05 was deemed statistically significant result. Clinical data was analysed using Fishers exact test due to low sample numbers, a p value of <0.05 was deemed statistically significant result.

2.11 Ethics

The *C. difficile* isolation study in which the hamster faeces was collected was performed in strict accordance with the recommendations in the United Kingdom's Home Office Animals (Scientific Procedures) Act of 1986 which outlines the regulation of the use of laboratory animals for the use of animals in scientific procedures. Colleagues within the Synthetic Biology Research Group gathered faeces from uninfected hamsters.

Clostridium difficile isolates and clinical information was collected as part of a University of Nottingham Hospitals NHS Trust Service Improvement Project. Stool samples and

anonymised clinical information from patients with CDI were provided by clinicians in the NDDC Biomedical Research Centre.

Chapter Three Using Forward and Reverse Genetics to Assess Toxin Regulation.

3.1 Introduction

3.1.1 TcdR and its homologues

As briefly described in 1.2.2 TcdR, is a 22kDa protein that contains a C-terminal helix-turn-helix DNA-binding motif[75]. Moncrief *et al*.[127] first presented evidence that TcdR had a role in TcdA and TcdB regulation by activating *tcdA* and *tcdB* reporter fusions through expression of *tcdR, in trans*, in *E. coli*. Later Mani *et al*.[75,128] in two separate studies demonstrated similar findings in *C. perfringens* (as a surrogate host) and *C. difficile*. Furthermore, they were able to show biochemical and genetic evidence that TcdR acts as an σ-factor directing the RNA polymerase (RNAP) to the *tcdA* and *tcdB* promoters by binding directly to the RNAP core[128]. Although, the *tcdA* and *tcdB* promoters show similarity to each other, they do not resemble traditional σ^{70} promoter sites found in prokaryotes[104]. Additionally, TcdR is able to activate its own expression, consistent with the presence of two prospective TcdR-dependent promoters upstream of *tcdR*[75].

Due to a high degree of similarity of structure and function of TcdR and other σ-factors found in pathogenic Clostridia that regulate toxin expression, they have been assigned their own group, group V, of the σ^{70}- family[129]. The other σ-factors include TcsR of *C. sordelli* which transcribes both the lethal and haemorrhagic toxins[130], TetR that transcribes tetanus neurotoxin[131], BotR that transcribes botulinum neurotoxin[131], and the UviA and TpeR of *C. perfringens* which transcribe a bacteriocin and a cytotoxin, respectively[132]. These σ-factors share such similarity they are able to be interchanged and still function[129,130]. In comparison with TcdR the other σ-factors are auto-regulated and are induced by environmental factors [104,128,131,133,134].

3.1.2 PaLoc gene expression and regulation.

Nutritional signals play a role in the control of toxin expression including carbon sources and some amino acids. In complex medium the addition of glucose or other rapidly metabolisable carbon sources inhibit toxin production[104,135,136]. The effect of glucose on toxin regulation has been demonstrated in multiple *C. difficile* strains and shown to occur at the transcriptional level suggesting a general mechanism[75,104]. Since the majority of carbon sources that inhibit toxin expression are transported into the cell via the phosphoenolpyruvate-dependent carbohydrate phosphotransferase system (PTS), it can be inferred that carbon catabolite repression (CCR) is involved in toxin gene expression[104]. When more than one carbon source, such as glucose, is available to the cell, CCR promotes the use of this preferred resource. The requirements of both PTS and CcpA for toxin suppression have been demonstrated using *C. difficile* mutants deficient in these systems[137]. CcpA regulates fermentation, sugar uptake and amino acid metabolism; glucose dependent repression of toxin production occurs by direct binding of CcpA to the promoter regions of *tcdR, tcdA, tcdB* and *tcdC*[138]. It appears that glucose availability and toxin synthesis are controlled by a complex regulatory network where CcpA plays a central role and interacts with other regulators such as CodY[138].

CodY is a global transcriptional regulator that is involved in the adaptive response to nutrient concentrations in the environment[139]. Repression of all the PaLoc genes has been demonstrated when nutrients are plentiful[139]. In *Bacillus subtilis* CodY binds to co-factors, branched-chain amino acids and GTP, which increases its affinity for its binding targets[140,141], reinforcing the link between toxin suppression and nutrient / co-factor availability. Consequently, when nutrients and co-factors are limited CodY no longer represses genes associated with bacterial adaptation to starvation. In a generated *codY* mutant strain, higher levels of toxin were produced during exponential phase when

compared to the wild type strain. Derepression of PaLoc genes was also observed in that mutant strain[139]. CodY was found to bind directly to the *tcdR* promoter region[139] and it has been shown to control more than 140 different genes[142].

When added to culture media, proline and glycine have been shown to have an inhibitory effect on toxin synthesis[136]. The PrdR regulator responds to the addition of proline, represses toxin gene expression, glycine-reductase and activates the expression of proline-reductase. The latter are both enzymes involved in Stickland metabolism, a chemical reaction that requires the coupled oxidation and reduction of amino acids to organic acids[143]. Cysteine is also reported to have an inhibitory effect on toxin synthesis[144].

σ-factors other than TcdR have been implicated in toxin regulation. A *sigH* mutant, for instance, resulted in the over expression of *tcdA*, *tcdB* and *tcdR*[145]. No SigH promoters have been located upstream of any of the PaLoc genes suggesting that there is an indirect mechanism of toxin repression[145]. The *sigH* mutant could not sporulate, but could still produce toxins. This implies there is no link between these two events, which is further validated by there being no transcriptional control via four other sporulation specific σ-factors (SigE, SigF, SigG and SigK)[146]. However, there has been evidence of strain specific toxin regulation by Spo0A the master regulator of sporulation. Inactivation of *spo0A* in some RT027 strains resulted in suppressed toxin expression[147,148], conflicting results were found for 630Δ*erm*[147–151] and marginal to no effect on RT078 strains[147]. Again, no Spo0A promoters have been identified within the PaLoc suggesting an indirect mechanism. This area requires further investigation to identify the regulators that are under the control of both SigH and / or Spo0A[145,149]. RstA is an additional regulator that has a negative effect on toxin synthesis by repressing SigD, a flagellar-specific σ-factor that directly controls *tcdR* transcription[152].

Evidence has been presented indicating that both intra- and inter-species quorum sensing has an effect on toxin synthesis. All sequenced strains of *C. difficile* to date have an incomplete *agr*-locus that contains *agrBD*, together these genes produce and export the autoinducing peptides (AIP)[153,154]. A full *agr*-locus (*agrACBD*) has been described in strain R20291 and other clinical strains and the two loci have been termed Agr1 and Agr2 respectively[154,155]. An *agrA* mutant in R20291 demonstrated diminished toxin production and was less able to colonise mice during infection[153,156]. Suggesting that the Agr2 system has a role in toxin production for lineages where it is present. Similarly, the peptide produced by the Agr1 system has been shown to induce toxin synthesis[156]. A second class of quorum sensing molecules, autoinducer-2 (AI-2) is present in several bacterial species and its synthesis is dependent on the LuxS enzyme[157,158]. *C. difficile* has a quorum sensing system that can detect this molecule and when added to culture medium during early-log phase *tcdA*, *tcdB* and *tcdE* are upregulated[157] but not during stationary phase[158].

Cyclic di-guanosyl-5'monophosphate (c-di-GMP) is a signalling molecule in second messenger bacterial systems that control flagella motility[159]. High levels of c-di-GMP were shown to repress *tcdA*, *tcdB*, *tcdR* and *sigD*[159]. Further to this, *tcdA*, *tcdB* and *tcdR* transcripts were reduced in a *sigD* mutant[160]. In the same study it was established that SigD directs the RNAP core enzyme to the *tcdR* promoter but not the *tcdA* or *tcdB* promoters. These data suggest that c-di-GMP has an indirect effect on toxin synthesis through SigD.

The flagellar proteins FliC and FliD (the flagellin and capping protein respectively) also have an effect on toxin production. Increased expression of all PaLoc genes except *tcdC* was described in a *fliC* mutant in 630Δ*erm*[161] and then in a later study *tcdA* transcripts were shown to be increased in both *fliC* and *fliD* mutants[162]. These data have been confirmed in cell culture and in some *in vivo* studies[163]. However, in a CD0240 mutant, a gene involved in the glycosylation of the flagellum, toxin levels were not significantly altered[161]. This

mutation resulted in a non-motile phenotype but unglycosylated FliC subunits were still expressed on the cell surface[164] suggesting it is the loss of flagellin rather than a non-functional flagellum that affects toxin production[165]. In R20291 there was no difference in the cytotoxicity between the *fliC* and *fliD* mutants and wild-type[162]. This could be attributed to the differences in R20291 strains as described in Appendix One. The described study was performed in NM-R20291 which only produces a single flagellum when compared to BW-R20291 which is peritrichously flagellated as is 630. Identical studies in BW-R20291 may reveal different results.

Bacteriophages have been identified in a number of different *C. difficile* genomes[166]. Lysogens derived from isolates after infection with phage ΦCD119 and ΦCD27 demonstrated a repressed toxin expression when compared to the parental strains[167,168]. Conversely, lysogenisation of a RT027 strain by ΦCD38-2 resulted in an increased toxin synthesis phenotype with toxin titres higher in culture supernatants, but not the cytosol when compared to the wild-type strain[169]. This has been attributed to an increased expression of all PaLoc genes; however the expression of *tcdE* was much stronger. The mechanism behind this is as yet unknown[169]. In another study *tcdE*-like genes were identified in *C. difficile* strains containing the prophages ΦC2, ΦC6 and ΦC8[170]. Increased toxin titres were observed in some of these lysogens even though gene expression was not[170]. The mechanism behind this was not identified in the study, but may be attributed to the presence of the *tcdE*-like genes.

LexA is the SOS response master regulator involved in DNA repair. A *lexA* mutant was found to produce more TcdA in the presence of sub-inhibitory concentrations of levofloxacin[171]. Similarly, an *mfd* (transcription-repair coupling factor) mutant also increased toxin expression at the transcriptional level[172].

3.1.3 The use of transposon libraries to identify novel TcdA and TcdB regulators.

Using reverse genetics to fully understand the complex mechanisms of PaLoc gene expression would be difficult, time consuming and expensive. Another method is to create random mutant libraries using transposable elements and then to screen for altered phenotypes. The use of random mutagenesis has been described in *C. difficile* to find regulators of sporulation and / or auxotrophs[97]. We aimed to use this system with some modification to identify altered toxin phenotypes. However, traditional methods for measuring toxin titres i.e. ELISA and / or cell culture are laborious and relatively expensive. To overcome this, a reporter strain (RS-7) was constructed which substituted *tcdB* in NM-R20291 with the lichenase gene *licB* of *Clostridium thermocellum* (Dr. S. Cartman unpublished). Lichenase hydrolyses the (1-4) bond after a (1-3) bond of ß-(1-3)(1-4)-glucans (lichenan) and shows high specificity for this substrate[173,174]. Congo red preferentially binds to contiguous polysaccharides containing both ß-(1-3)-glucans and ß-(1-4)-glucans, but if the lichenase has depolymerised the polysaccharide this binding does not occur and no colour change is observed[173,174]. Therefore, when media containing lichenan (and no other reducing sugars) is stained with 0.2% Congo Red solution, the media will acquire a deep red colouration unless the lichenan has been degraded by lichenase. As the lichenase is excreted out of the cell and into the media a zone of clearance around discrete colonies is identifiable after straining, resulting in BI-7 producing a phenotype that is rapidly and economically identifiable.

3.1.4 Aims

- Use a TcdR null mutant in NM-R20291 to assess its role in toxin synthesis.
- Identify and characterise novel toxin regulators using a random transposon mutant library created in a host containing a *licB* reporter system.

3.2 Results

3.2.1 The functional inactivation of tcdR *in strain NM-R20291*

To further understand the role of TcdR in regards to the regulation of *tcdA* and *tcdB* two isogenic strains, NM-R20291::Δ*tcdR* and NM-R20291::*tcdR*(EcoR1), were constructed via allelic exchange utilising *codA* as a negative selection marker[80] (strains created by Dr. S Cartman & Ms. M Kelly). NM-R20291::Δ*tcdR* was constructed using two homology arms at ~500bp directed either side of *tcdR* which also incorporated the first 9bp and last 9bp of *tcdR* in the left homology arm and right homology arm respectively. This resulted in a 537bp in-frame deletion in *tcdR* of NM-R20291 (Figure 3-1A) and a deletion of codons 4 through 181 of TcdR. NM-R20291::*tcdR*(*Eco*R1) was constructed by complementing the NM-R20291::Δ*tcdR* strain with a *tcdR* sequence that had a silent mutation at position 295 (T to C) which resulted in an *Eco*R1 restriction site (Figure 3-2).

Before phenotypic analysis of the isogenic strains was performed, PCR and sequencing was used to check the genotype for each strain, resulting in products of 1500bp for NM-R20291 and NM-R20291::*tcdR*(*Eco*R1) and 1000bp for NM-R20291::Δ*tcdR* (Figure 3-1B). To further check the genotype for each strain the PCR products were digested with *Eco*R1, resulting in products of 1500bp for NM-R20291, 1000bp for NM-R20291::Δ*tcdR* and 750bp for NM-R20291::*tcdR*(*Eco*R1) (Figure 3-1C).

3.2.2 Phenotypic analysis of the NM-R20291 tcdR *null mutant.*

To measure the effect of the *tcdR* genotype on growth and toxin production each strain was grown in TY medium and samples taken over a 48h time course in triplicate. There were no measurable differences in growth between the strains (Figure 3-3A). Toxin supernatants were measured by *C. difficile* Tox A/B II ELISA (TechLab), detection limit 1.0 ng/mL, and

toxin concentrations were determined using a standard curve (TcdA & TcdB mixture at 125 ng/mL – 1.95 ng/mL).

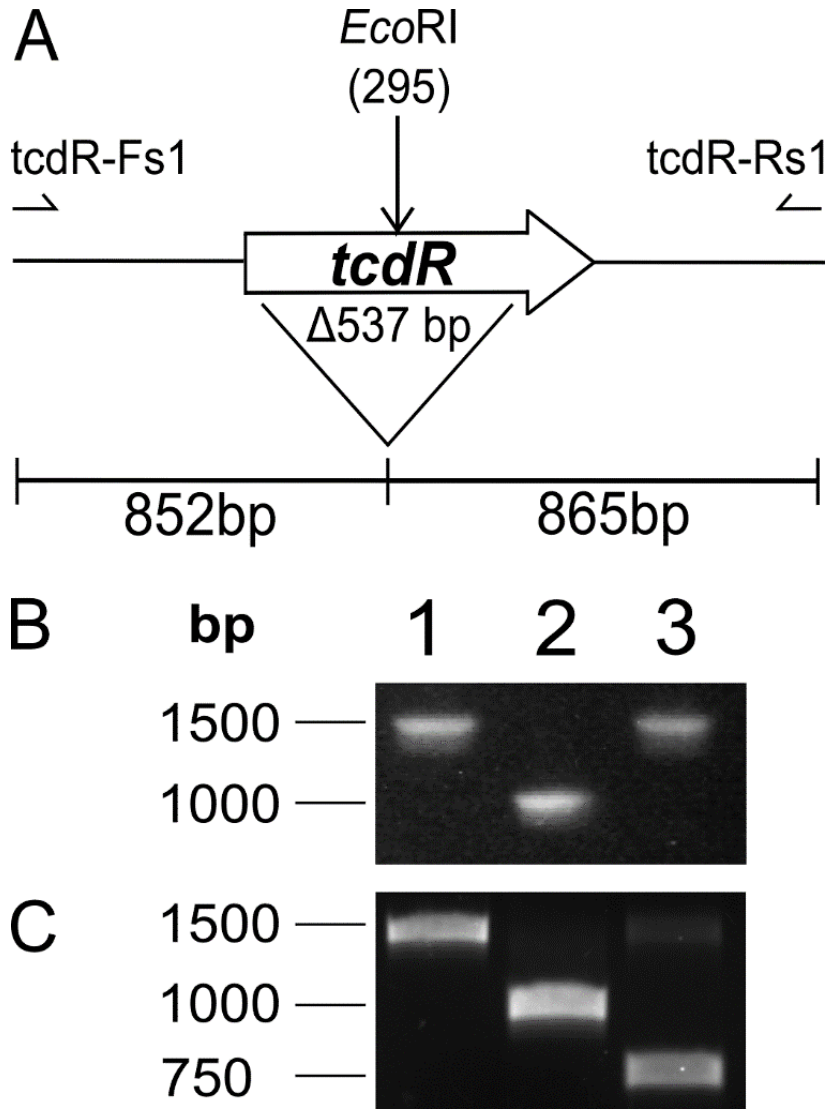


Figure 3-1: PCR analysis of recombinant NM-R20291 strains. (A) Schematic of the *tcdR* locus in wild-type NM-R20291. The half-arrows indicate annealing regions of primers tcdR-F1 and tcdR-R1. The triangle below the arrow depicting the *tcdR* ORF indicates the deletion made in NM-R20291::Δ*tcdR*. The approximate location of the EcoR1 site introduced into NM-R20291::(*tcdR*[EcoRI]) is shown above the ORF. (B) PCR analysis was performed with primers tcdR-F1 and tcdR-R1. (C) The subsequent PCR products after EcoR1 digestion. Lanes: 1: Wild-type NM-R20291; 2, NM-R20291::(Δ*tcdR*); 3, NM-R20291::(*tcdR*[EcoRI])

```
tcdR            ATGCAAAAGTCTTTTTATGAATTAATTGTTTTAGCAAGAAATAACTCAGTAGATGATTTG 60
tcdR(EcoRI)     ATGCAAAAGTCTTTTTATGAATTAATTGTTTTAGCAAGAAATAACTCAGTAGATGATTTG 60
ΔtcdR           ATGCAAAAG---------------------------------------------------- 9

tcdR            CAAGAAATTTTATTTATGTTTAAGCCATTAGTAAAAAAACTTAGTAGAGTTTTACATTAT 120
tcdR(EcoRI)     CAAGAAATTTTATTTATGTTTAAGCCATTAGTAAAAAAACTTAGTAGAGTTTTACATTAT 120
ΔtcdR           ------------------------------------------------------------

tcdR            GAAGAGGGAGAAACAGATTTAATAATATTTTTTATTGAATTAATAAAAAATATTAAATTA 180
tcdR(EcoRI)     GAAGAGGGAGAAACAGATTTAATAATATTTTTTATTGAATTAATAAAAAATATTAAATTA 180
ΔtcdR           ------------------------------------------------------------

tcdR            AGTAGCTTTTCAGAAAAAAGCGATGCTATTATAGTCAAATATATTCATAAATCATTACTG 240
tcdR(EcoRI)     AGTAGCTTTTCAGAAAAAAGCGATGCTATTATAGTCAAATATATTCATAAATCATTACTG 240
ΔtcdR           ------------------------------------------------------------

tcdR            AATAAGACTTTTGAGTTGTCTAGAAGATATTCTAAAATGAAGTTTAATTTTGTAGAATTT 300
tcdR(EcoRI)     AATAAGACTTTTGAGTTGTCTAGAAGATATTCTAAAATGAAGTTTAATTTTGTAGAATTC 300
ΔtcdR           ------------------------------------------------------------

tcdR            GATGAAAATATCTTAAATATGAAAAATAATTATCAAAGTAAGTCTGTTTTTGAGGAAGAT 360
tcdR(EcoRI)     GATGAAAATATCTTAAATATGAAAAATAATTATCAAAGTAAGTCTGTTTTTGAGGAAGAT 360
ΔtcdR           ------------------------------------------------------------

tcdR            ATTTGTTTTTTCGAATATATTTTGAAAGAATTATCTGGTATTCAAAGAAAAGTTATTTTT 420
tcdR(EcoRI)     ATTTGTTTTTTCGAATATATTTTGAAAGAATTATCTGGTATTCAAAGAAAAGTTATTTTT 420
ΔtcdR           ------------------------------------------------------------

tcdR            TATAAATATTTAAAAGGATATTCTGATAGAGAAATATCAGTGAAATTAAAAATATCTAGA 480
tcdR(EcoRI)     TATAAATATTTAAAAGGATATTCTGATAGAGAAATATCAGTGAAATTAAAAATATCTAGA 480
ΔtcdR           ------------------------------------------------------------

tcdR            CAAGCTGTTAATAAGGCTAAAAATAGAGCATTTAAAAAAATAAAAAAAGACTATGAAAAT 540
tcdR(EcoRI)     CAAGCTGTTAATAAGGCTAAAAATAGAGCATTTAAAAAAATAAAAAAAGACTATGAAAAT 540
ΔtcdR           ------------------------------------------------------------

tcdR            TATTTTAACTTGTAA 555
tcdR(EcoRI)     TATTTTAACTTGTAA 555
ΔtcdR           ------AACTTGTAA 18
```

Figure 3-2: Nucleotide sequence of *tcdR* of NM-R20291, NM-R20291::Δ*tcdR* and NM-R20291::*tcdR*(*Eco*R1). Strain NM-R20291::Δ*tcdR* has a 537bp deletion leaving the first and last 9bp. Strain NM-R20291::*tcdR*(*Eco*R1) has a complemented sequence with a single nucleotide substitution at position 300 (T to C) resulting in a silent mutation and the addition of an EcoR1 restriction site (highlighted in red).

Toxin production for both NM-R20291 and NM-R20291::*tcdR*(*Eco*R1) followed the expected temporal pattern. Accordingly, toxin levels steadily increased throughout the exponential growth phase and peaked during the stationary growth phase (Figure 3-3B). Interestingly, toxin production in NM-R20291::Δ*tcdR* also followed the same temporal pattern, although toxin production was reduced by an order of 1,000-10,000-fold at 48 h. Differences in toxin production were statistically significant from 12 h by unpaired T test (Figure 3-3B).

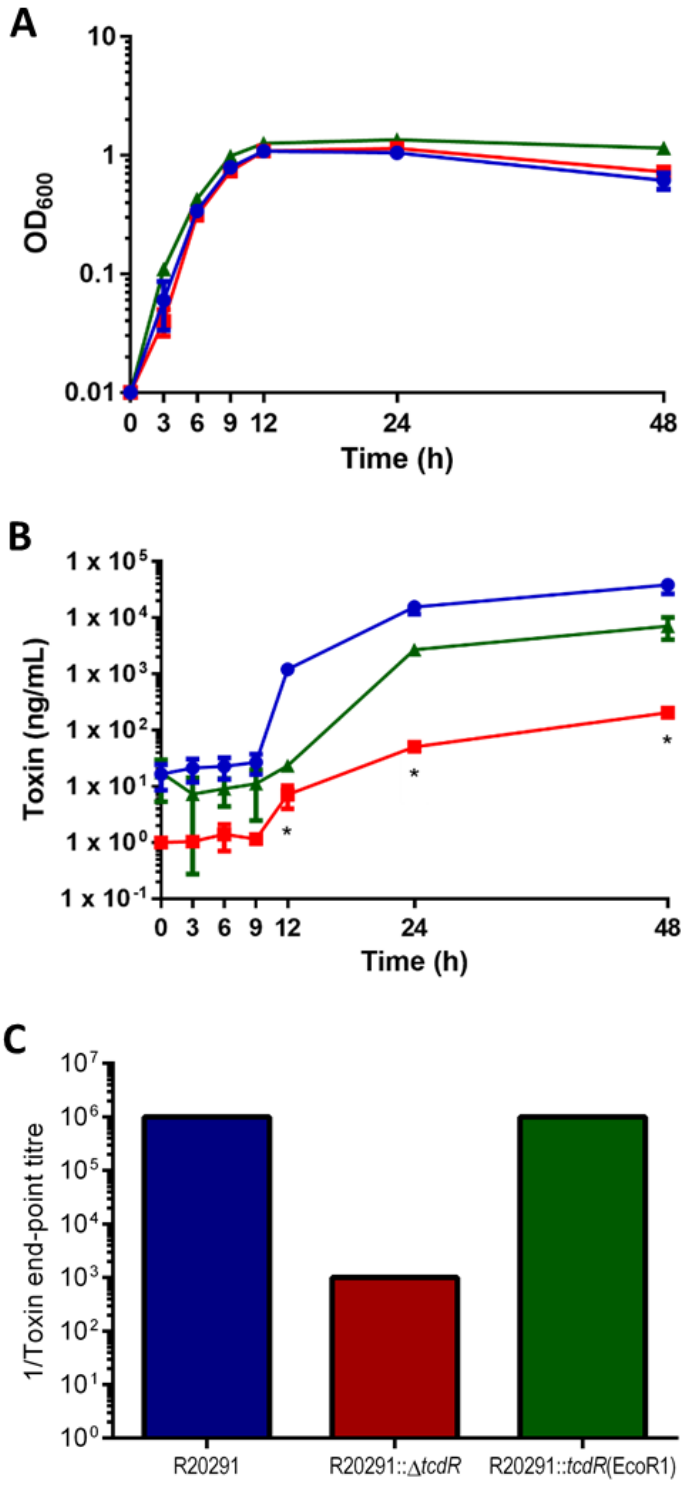


Figure 3-3: Cell growth and toxin production by recombinant *C. difficile* NM-R20291 strains in TY medium for 48h. Figure legend; (●) NM-R20291; (■) NM-R20291::Δ*tcdR*; (▲) NM-R20291::*tcdR*(*Eco*RI). (A) Growth was measured at OD_{600}. (B) Toxin concentration determined by *C. difficile* Tox A/B II ELISA using a standard curve of known Toxin A and B concentrations. (C) Cell culture using Vero cells for Toxin B at 48h. Data represent the mean ± SEM (n=3). * indicates statistical significance (p = <0.005) by unpaired T-test.

To further check the rate of toxin production, cell cytotoxicity assays were performed using Vero (African green monkey kidney) cell monolayers. Vero cells (detection limit 25pg/mL, toxin B) were subjected to titrated toxin supernatants taken at 48h. End-point titre was defined when cell morphology was indistinguishable from the negative controls. The cell cytotoxicity assay confirmed the results obtained by ELISA at 48h. Thus, toxin production by the null mutant strain NM-R20291::Δ*tcdR* was not eliminated but reduced by 100-1,000 times which was observed with both the wild-type NM-R20291 and the complemented NM-R20291::*tcdR*(*Eco*R1) strains (Figure 3-3C). Toxin neutralisation assays using *C. difficile* Tox-B Test (TechLab) in parallel to the cell cytotoxicity assays were performed to ensure that all of the observed cytotoxic effect was specific to the toxin and not due to another artefact. All toxin activity was successfully neutralised (data not shown).

3.2.3 Random Transposon Mutagenesis to Identify Altered Toxin Phenotypes

As discussed in 3.1.3 random mutagenesis has an advantage over reverse genetic analysis as it does not require a hypothesis or target genes of interest that will produce a measurable phenotype. It does require a robust screening method so that multiple mutants can be assessed in parallel for altered phenotypes. To overcome this a reporter strain was created, RS-7, in which the *tcdB* gene has been replaced with *licB* to allow for rapid phenotype identification using a lichenase plate assay.

3.2.4 Random mutant screening.

Random mutant libraries were produced using RS-7 as a host as described in 2.6.3 by Miss. B Boyle and after screening of passage 0 – 2 via PCR and sequencing it was deemed passage 1 was the most diverse. A number of mutants with altered phenotypes were identified using the lichenase plate assay (Table 3-1).

Table 3-1 List of random transposon mutants created in RS-7 that produce an altered phenotype.

ORF interrupted	Assigned function
R20291_0774	Cell surface protein
R20291_1065	Putative O-methyltransferase
R20291_1439	Putative uncharacterised protein
R20291_0853	ABC transporter
R20291_1086	Conserved hypothetical protein
R20291_2848	Putative glycosyl hydrolase
R20291_2908	putative uncharacterized protein
Intergenic spacer region between*:	
R20291_3537	SpoIIIJ-associated protein (*jag*)
R20291_3538	putative sporulation membrane protein

*This insertion was later termed BBM4.

3.2.5 Random mutant BBM4.

The most notable phenotype seen was in mutant BBM4, which was isolated initially by Miss. B Boyle. Initial phenotype screening using a tooth pick to stab inoculate TY agar containing 1% lichenan showed no zone of clearance of lichenan at 24 h and a reduced zone of clearance at 48 h when compared to RS-7 after staining with Congo red (data not available). To confirm this phenotype 20 µL of overnight culture was inoculated on TY agar containing 1% lichenan and cultures were incubated for 24 h and 48 h. This confirmed the phenotype previously seen where there was no zone of clearance of lichenan at 24 h (Figure 3-4A) and a reduced zone of clearance at 48 h (Figure 3-4B) when compared to RS-7 after staining with Congo red. To locate the insertion site of the transposon inverse PCR was employed, repeated sequencing of inverse PCR products after HindIII digestion showed a mixed signal indicating a double insertion site. To confirm the double insertion Southern blot was employed to detect the transposon containing the *catP* gene. Digestion with HindIII and MfeI in separate reactions (Figure 3-5C) showed a single band and a double band respectively. Estimated band sizes in the Southern Blot cannot be calculated due to overloading of ladder and product.

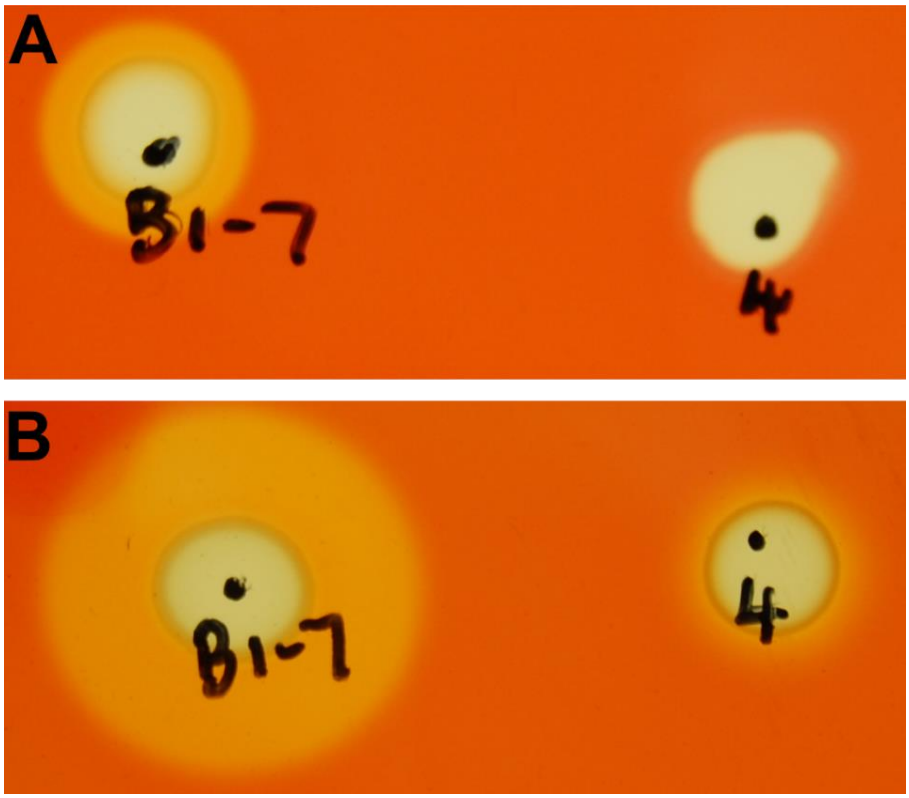
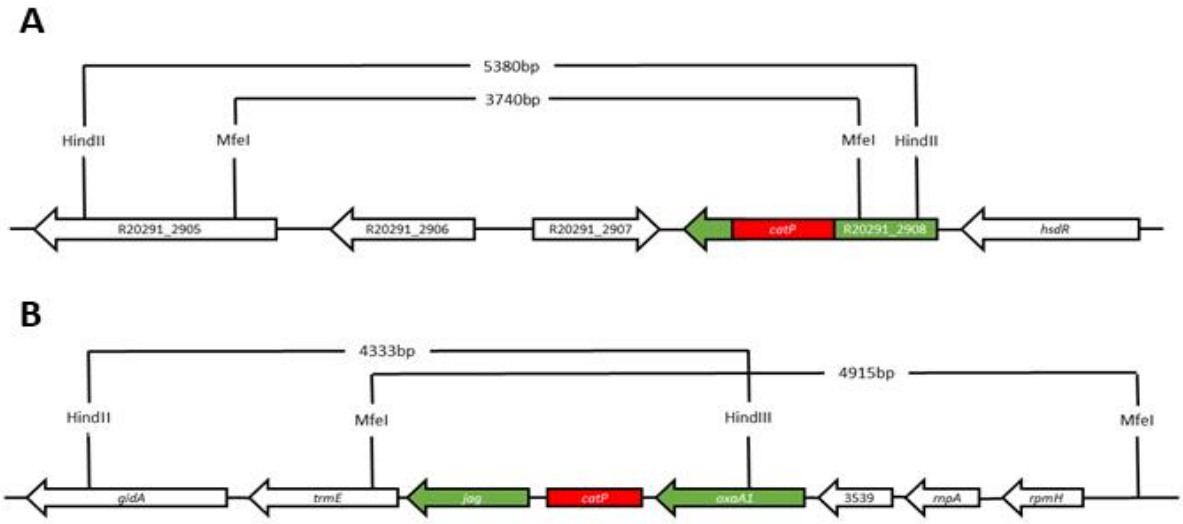


Figure 3-4: Phenotypic screening of transposon mutant BBM4 on TY agar containing 1% lichenan. Zones of clearance visualised by the addition of Congo red after incubation for (A) 24 h and (B) 48 h. Figure legend; "BI-7" is the RS-7 control strain and "4" is the BBM4 transposon mutant.

To identify the location of the transposons in BBM4 inverse PCR on *Mfe*I digested products was performed and sent for Sanger sequencing. This revealed that the genes of interest were located in *R20291_2908* at base pair 3,441,942 and in the intergenic spacer region of *jag* (R20291_3537) and *oxaA1* (R20291_3538) at base pair 4,189,071. To confirm the Southern blot findings a schematic to identify product length for the insertion at R20291_2908 and the *jag*/*oxaA1* region was created by using Artemis. Using this software restriction sites were identified and estimated product lengths calculated for R20291_2908 (Figure 3-5A) and *jag*/*oxaA1* (Figure 3-5B). This identified that the two products from the HindIII digestion should be 5380bp and 4333bp for R20291_2908 and *jag*/*oxaA1*respectivly. They are different sizes and should be distinguishable on a 0.8% agarose gel. It has not been investigated further as to why this band cannot been seen, however it could be due to a nutation in the restriction

site or an annotation error in the sequence map. This could be investigated by Sanger sequencing the region. The two products from the MfeI digestion should be 3740bp and 4915bp for R20291_2908 and *jag*/*oxaA1*respectivly and can clearly be seen as separate bands on the Southern Blot.

ClosTron plasmids were designed to interrupt the R20291_2908 and R20291_*jag*. It was decided to target *jag* rather than *oxaA1* as *jag* is located in the downstream location. ClosTron mutants were created resulting in NM-R20291_2908(*ermB*) and NM-R20291::*jag*(*ermB*) as described in 2.6.1. ClosTron insertion was screened for by junction PCR (**Figure** 3-6C) with a schematic representation for NM-R20291_2908(*ermB*) (**Figure** 3-6A) and NM-R20291::*jag*(*ermB*) (**Figure** 3-6B). Southern Blot confirmed the single insertion of ClosTron group II introns (Figure 3-7C) with the expected product size (Figure 3-7A and B). To measure the effect of the insertional inactivation of genes CDR20291_2908 and *jag* on growth and toxin production each strain was grown in TY medium and samples taken over a 48h time course. There was no significant difference between growth for any of the strains (Figure 3-8A). NM-R20291::2908(*ermB*) displayed a small initial increase in toxin production until 9 h which was then surpassed by both NM-R20291 and NM-R20291::*jag*(*ermB*) (Figure 3-8B). A significant difference in toxin concentration between NM-R20291 and NM-R20291::2908(*ermB*) was seen at 48h when analysed by unpaired T-test ($p = <0.05$) (Figure 3-8C).



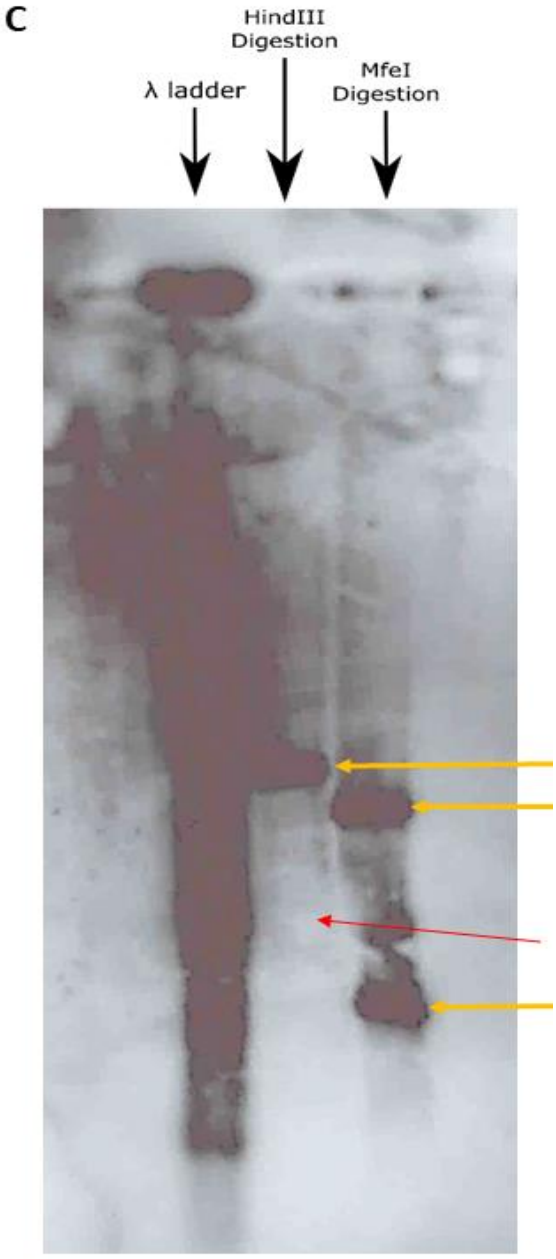


Figure 3-5: Schematic diagrams and Southern blot of transposon insertion into BBM4. Figure legend; Red box indicates insertion of the *catP* transposon at an AT site; green arrows indicate the predicted affected genes due to transposition. (A) Schematic of transposon insertion into R20291_2908, location of HindIII and MfeI digestion sites and estimated Southern blot band sizes. (B) Schematic of transposon insertion into the intergenic spacer region of *jag* and *oxaA1*, location of HindIII and MfeI digestion sites and estimated Southern blot band sizes. (C) One band observed from Southern blot analysis after HindIII digestion. Two bands observed from Southern blot analysis after MfeI digestion due to double insertion. All indicated by orange arrows. Red arrow indicates the approximate location of the missing HindIII digestion product. *catP* probe was generated by PCR of pMTL-SC1 using primers catP-F1 and catP-R1. Calculation of estimated band sizes have not been performed due to the overloading of ladder product. 0.8% gel.

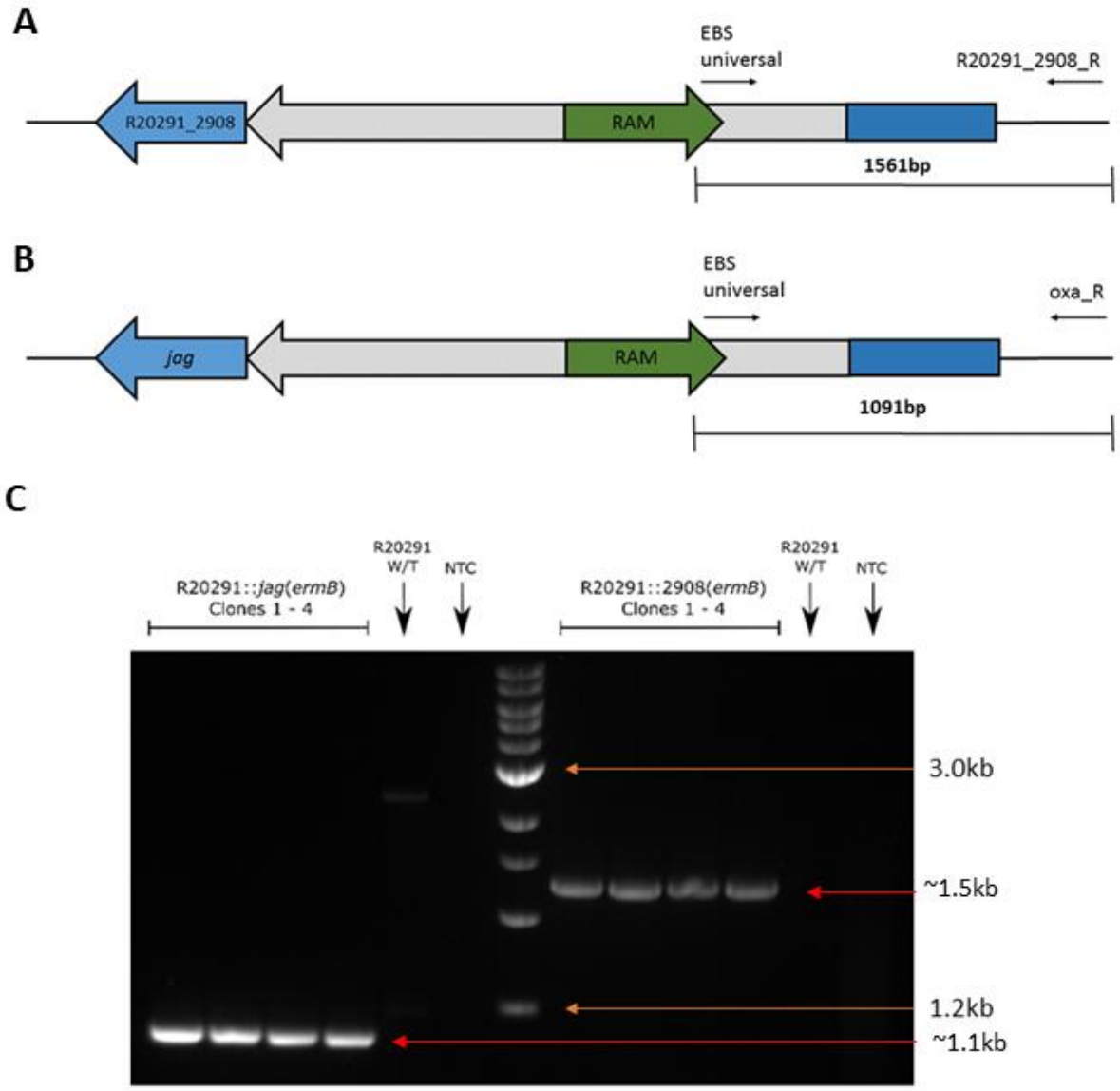


Figure 3-6: PCR screening of ClosTron mutants NM-R20291::*jag*(*ermB*) and NM-R20291::*2908*(*ermB*). Figure legend, Blue arrow represents the interrupted gene of interest, Grey arrow indicates the Group II intron, Green arrow indicates RAM containing the lincomycin resistance gene ermB (A) A schematic representation of the ClosTron insertion into NM-R20291_2908 showing primer binding sites for junction PCR and expected product sizes. (B) A schematic representation of the ClosTron insertion into NM-R20291_jag showing primer binding sites for junction PCR and expected product sizes. Schematics not to scale. (C) Junction PCR of four clones of each ClosTron mutant showing expected band sizes. NM-R20291 wildtype (W/T) used as negative control and a no template control (NTC) to assess for contamination. Product sizes measured using 2-Log Ladder.

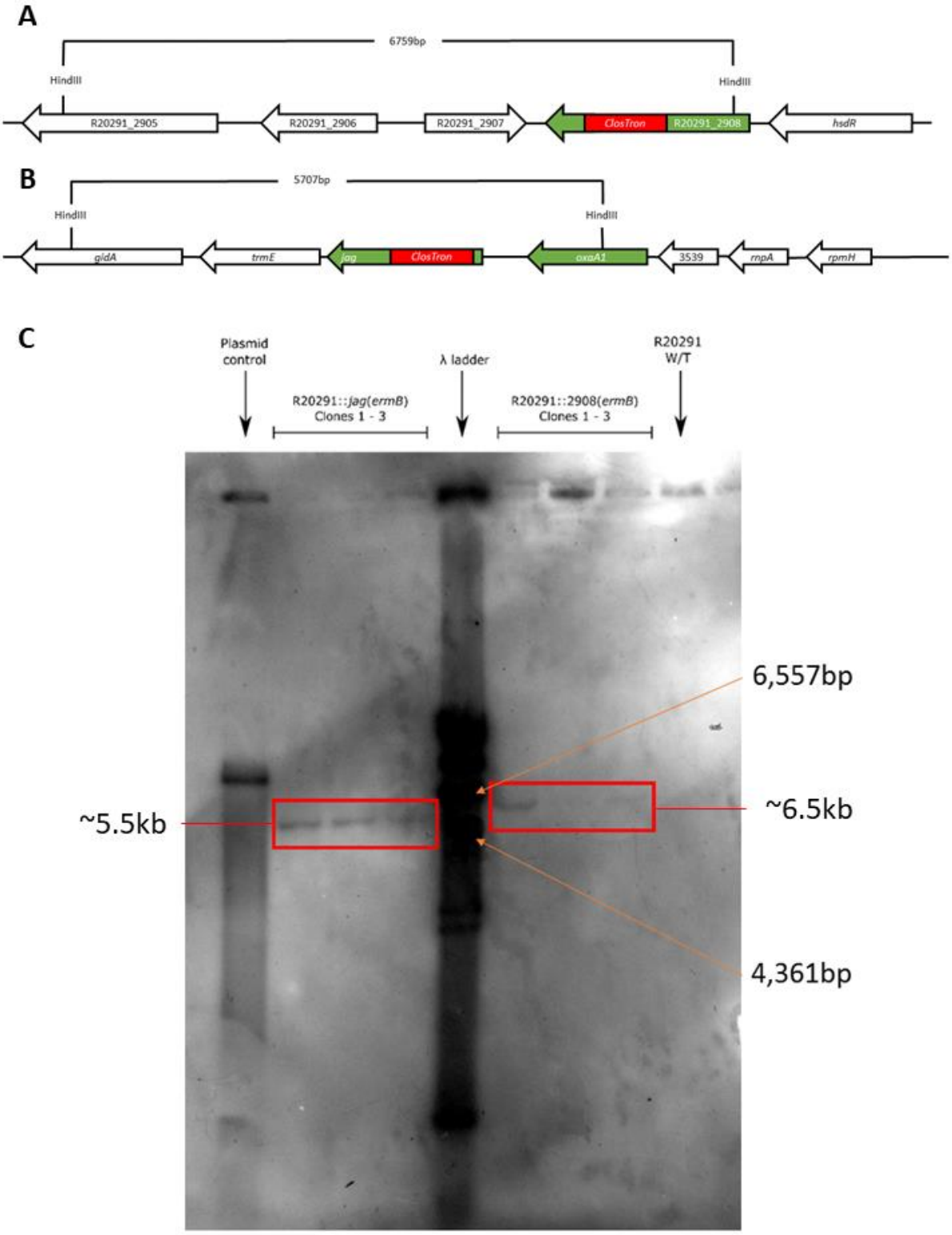


Figure 3-7: Southern blot analysis of ClosTron mutants NM-R20291_2908(*ermB*) and NM-R20291::*jag*(*ermB*). Figure legend; Red box indicates insertion of the group II intron, green arrows indicate genes of interest. (A) Schematic of ClosTron insertion into NM-R20291_2908, location of HindIII digestion sites and estimated Southern blot band sizes. (B) Schematic of ClosTron insertion into *jag*, location of HindIII digestion sites and estimated Southern blot band sizes. (C) Southern blot analysis using intron probe generated using primers EBS2 and IntronSalR. Lambda/HindIII ladder used for the estimation of band sizes. Single bands

identified for NM-R20291_2908(*ermB*) (~6.5kb) and NM-R20291::*jag*(*ermB*) (~5.5kb), highlighted in Red boxes. ClosTron plasmid used as positive control at ~6000bp and NM-R20291 wildtype (W/T) used as negative control.

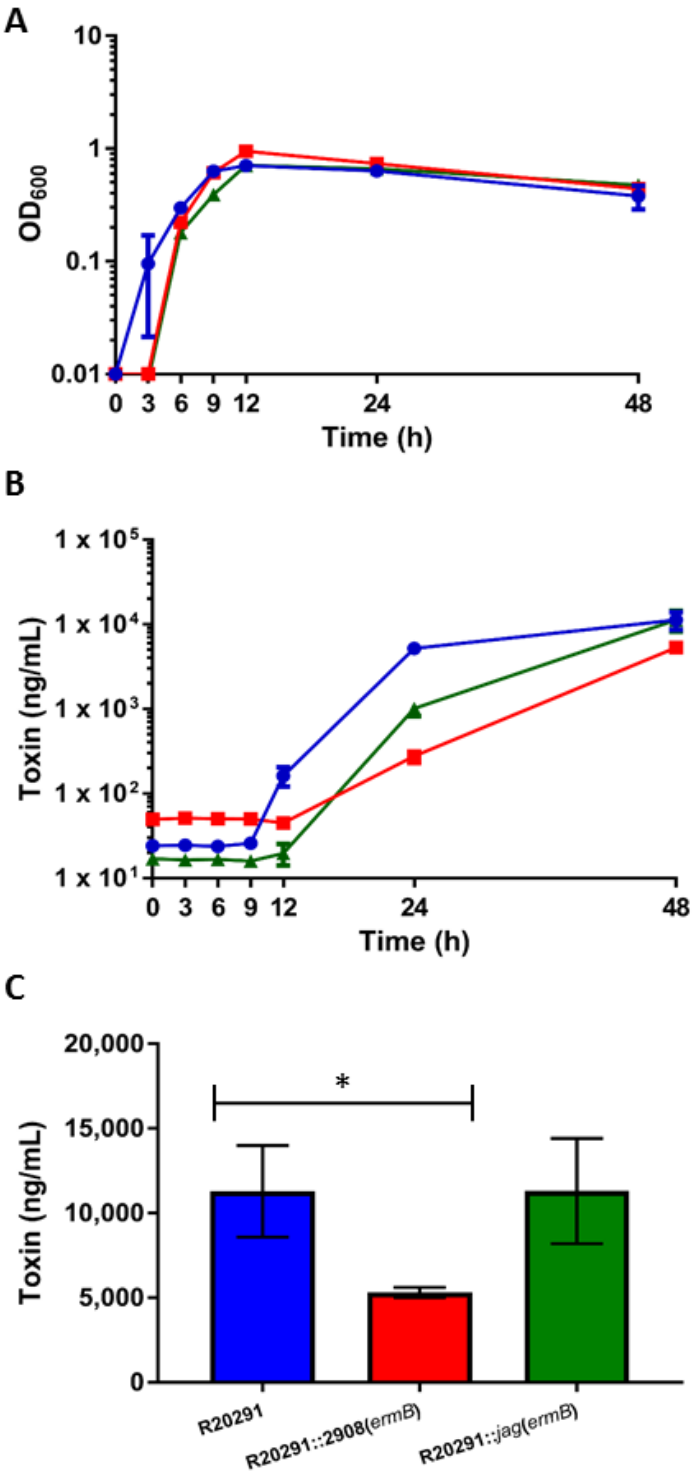


Figure 3-8: Cell growth and toxin production by recombinant *C. difficile* NM-R20291 strains in TY medium for 48h. Figure legend; (●) NM-R20291; (■) NM-R20291::2908(*ermB*); (▲) NM-R20291::*jag*(*ermB*). (A) Growth measured at OD_{600}. (B) Toxin production determined by *C. difficile* Tox A/B II ELISA using a standard curve of known Toxin A and B concentrations. (C) Toxin concentration at 48h. Data represent the mean ± SEM (n=3). * indicates statistical significance (p = <0.05) by unpaired T-test.

To restore functional activity of CDR20291_2908 both complementation and overexpression plasmids were constructed. Prediction of the native promoter region was performed using BPROM (Last accessed 12/03/17, http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb). A schematic representation is shown in (Figure 3-9). For the complementation plasmid, the native gene and predicted promoter were PCR amplified from NM-R20291 using primers, which introduced NotI (5'), and XhoI (3') restriction sites. The resulting product was cloned into plasmid pMTL84151. The overexpression plasmid was constructed using only the native gene using primers to introduce EcoR1 (5') and XhoI (3') restriction sites. The product was cloned in to pMTL84152; this plasmid contains a thialase promoter (P_{thl}) from *Clostridium acetobutylicum*, which is constitutive. Resulting plasmids pMTL84151::p2908 and pMTL84152::2908 were Sanger sequenced to ensure sequence accuracy.

The initial strategy for testing was to transform empty vectors into both NM-R20291 and NM-R20291::2908(*ermB*) and test these alongside the complementation, overexpression plasmids. Strains containing plasmids were grown in TY broth containing thiamphenicol (15mg/mL) to ensure retention of the plasmid. The control strains containing empty vectors were shown to have an inhibited growth pattern through OD measurements (data not available) and were unable to produce toxin comparable to their counterparts (Figure 3-10A). To overcome this, we decided to test all the strains in TY without thiamphenicol and after the 48h time point plate a 10 µL volume the remaining culture onto BHIS agar containing thiamphenicol (15mg/mL) to assess plasmid retention. This method will not account for any cells that have lost their plasmids throughout the assay resulting in altered phenotype populations, therefore is not 100% accurate. The control strains were found to lose the plasmids during the assay, the complementation, overexpression plasmids were retained in strain NM-R20291::2908(*ermB*).

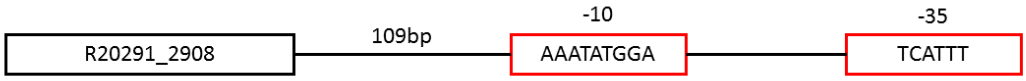


Figure 3-9: Schematic of predicted promoter region of gene R20291_2908. Prediction performed by BPROM and identified a -35 and -10 promoter region downstream of R20291_2908, which is located on the antisense strand. A sequence of 109 non-coding bp are located between the -10 and the start codon of the gene.

The assay was repeated on two separate occasions, in the both attempts the parental strain NM-R20291 was found to produce the same amount of toxin as NM-R20291::2908(*ermB*) over the 48 h time course. In these assays the complementation and overexpression strains were shown to produce equal (Figure 3-10C) or increased **(**Figure 3-10E) levels of toxin compared to both NM-R20291 and NM-R20291::2908(*ermB*) despite similar growth patterns (Figure 3-10B and D) over the 48 h time course.

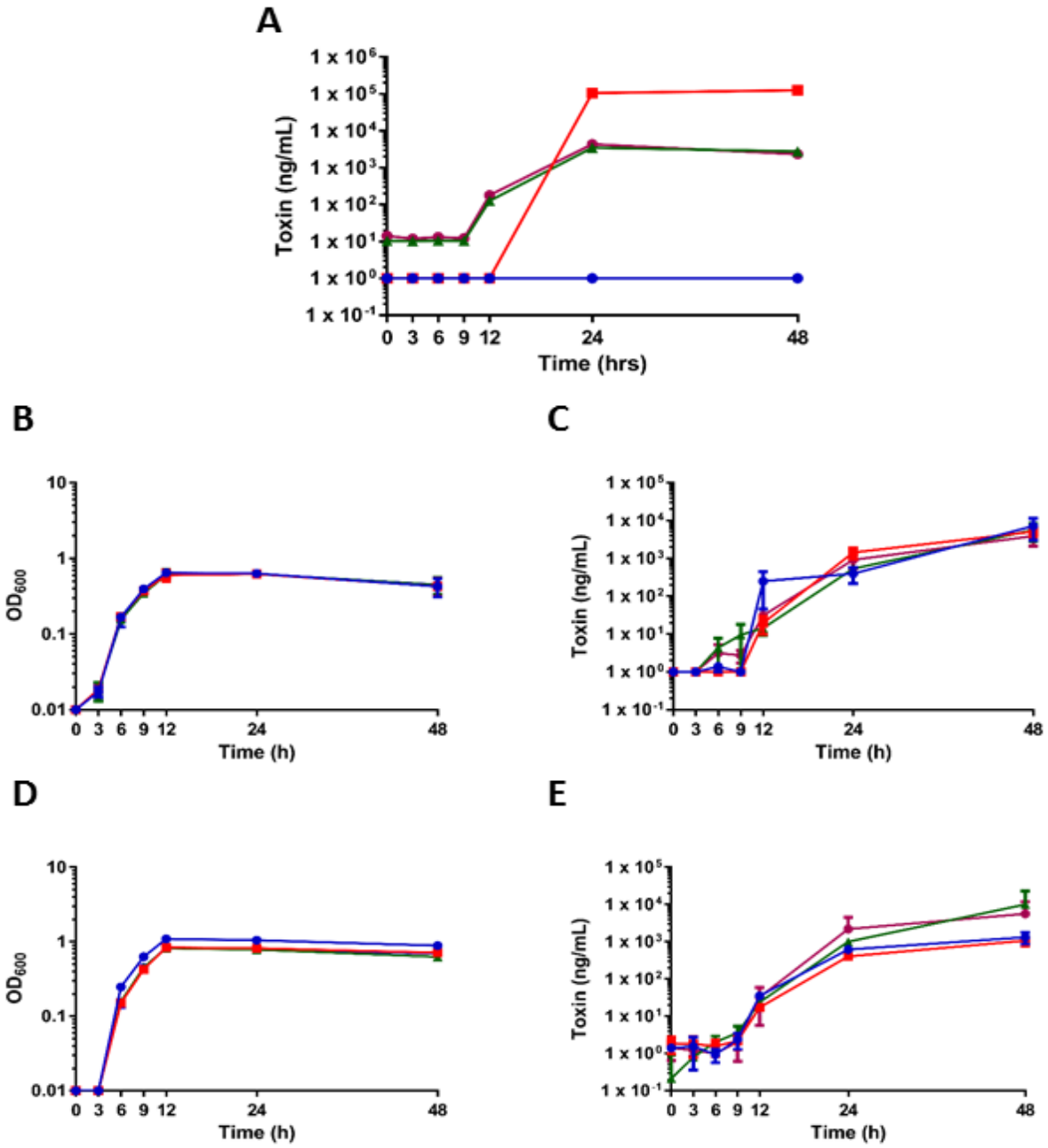


Figure 3-10: Cell growth and toxin production by recombinant *C. difficile* NM-R20291 strains in TY medium for 48h. Toxin concentrations determined using *C. difficile* Tox A/B II ELISA with a standard curve of known Toxin A and B concentrations. Figure legend; (●) NM-R20291 containing pMTL-84151; (■) NM-R20291::2908(*ermB*) containing pMTL-84151; (▲) NM-R20291::2908(*ermB*) containing pMTL-84151::2908; (●) NM-R20291::2908(*ermB*) containing pMTL-84152::2908. (A) Toxin concentrations from cultures grown in TY plus thiamphenicol. (B) Growth in TY only measured at OD_{600} experiment one (C) Toxin concentration from cultures grown in TY only experiment one. (D) Growth in TY only measured at OD_{600} experiment two (C) Toxin concentration from cultures grown in TY only experiment two. Data represent the mean ± SEM (n=3).

3.2.6 Quantitative Lichenase Assay

The lichenase plate assay is a useful tool for screening but it does not provide quantitative data. To overcome this, a previously developed protocol (Dr. C. Humphreys, unpublished) used for the quantification of lichenase in a *Clostridium botulinum licB* reporter system was assessed. This method measures lichenase activity through incubation, for a set period of time, of a supernatant with a substrate containing a known concentration of lichenan. Lichenase activity is terminated by the addition of 0.2M NaOH, the remaining lichenan is measured by the addition of 0.1% Congo Red solution resulting in a similar deep red colouration as described for the plate assay. Supernatant / substrate mixtures are measured at A_{530} and compared to a control sample that contains lichenan but no lichenase. The reduction in A_{530} between the two samples is directly proportional to the rate of lichenase activity (ΔA_{530}).

Initial testing of the quantitative lichenase assay, as described in 2.8.5, for use in the *C. difficile licB* reporter system was performed by Dr. E. Stevenson (unpublished results). It was shown that the addition of NaOH before the addition of 0.1% Congo Red solution resulted in no measurable difference in A_{530} between samples that did and did not contain lichenase. Conversely, if 0.1% Congo Red solution was added first and allowed to incubate for a set time, then 0.2M NaOH added, a measurable difference in A_{530} was observed. However, this produces no defined endpoint of enzymatic activity therefore it is not a reliable method.

Reasons for this phenomenon are unclear and review of the literature showed no possible explanation but it was assumed that the NaOH played role. To overcome this substitute enzyme altering solutions were tested for stopping the lichenase reactions. These included 0.5M Sodium acetate, 0.5M Sodium carbonate and 2% Trizma base and were chosen because of their high pH. Different substrate/enzyme incubation times were also considered to improve ΔA_{530} (Figure 3-11).

The difference in ΔA_{530} increased with correlation to the enzyme incubation time and 0.5M Sodium carbonate produced a significantly increased difference in ΔA_{530} when compared to 0.5M Sodium acetate ($p = <0.001$) and 2% Trizma base ($p = <0.001$) after a 40 minute incubation (Figure 3-11).

A modified quantitative lichenase assay was trialled where a 40 minute enzyme incubation and 0.5M Sodium carbonate stop solution was utilised. A 48 h assay was performed to assess the new protocol. It was found that no measurable difference in A_{530} could be seen in supernatants taken at <48 h between samples containing lichenase and samples that did not contain lichenase (data not shown). Multiple attempts to obtain reproducible results using 48 h supernatants failed, at this point it was decided to explore other means to develop a quantitative assay.

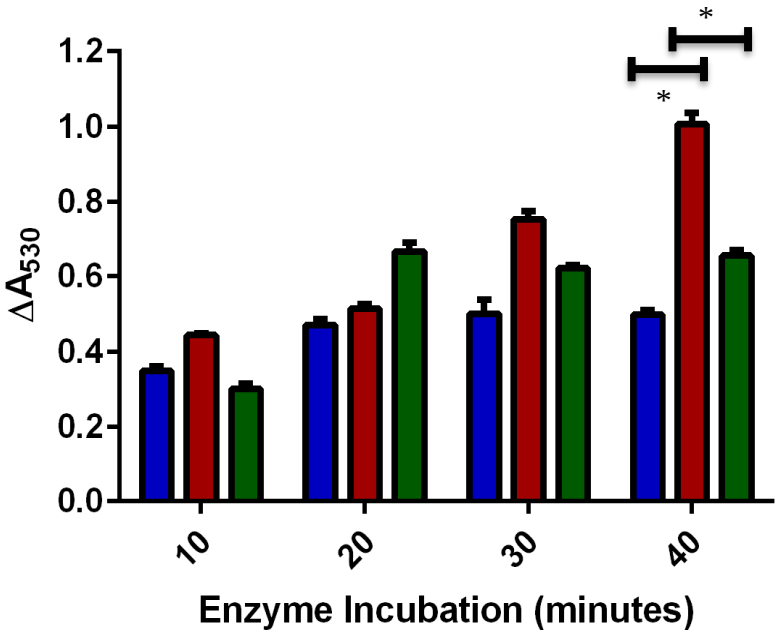


Figure 3-11: Lichenase enzyme activity over multiple incubation time points. Each bar indicates a different stop solution (■) 0.5M sodium acetate, (■) 0.5M sodium carbonate and (■) 2% Trizma base. Data represent the means of three independent experiments, error bars represent standard error. * indicates statistical significance (p = <0.001) by paired T-test.

3.3 Discussion

TcdR is a known regulator of PaLoc gene expression. There have been studies looking at the role of TcdR in *C. difficile* toxin synthesis, these have involved using fusion reporter systems[75,128,129] or insertional gene inactivation[159], but there has yet to be an example of a *tcdR* null mutant. In this study we created a clean deletion of *tcdR* in strain NM-R20291 to overcome the drawbacks from using the previously described methods, such as, polar effect, non-native hosts and expression of genes from plasmids. The deletion was then complemented with *tcdR* in the same chromosomal location that contained a silent nucleotide substitution resulting in an *Eco*R1 restriction site as a watermark to distinguish from the parental strain.

The isogenic strains NM-R20291::Δt*cdR* and NM-R20291::*tcdR*(*Eco*R1) did not show any difference in growth kinetics when compared to NM-R20291. There was however a statistically significant difference in toxin concentration after 12 h in NM-R20291::Δ*tcdR* compared to the parental strain which was restored in the complemented strain NM-R20291::*tcdR*(*Eco*R1). At 48 h toxin concentration was reduced by an order of 1,000 to 10,000-fold and 1,000-fold when assessed by Toxin ELISA and cell cytotoxicity respectively. The variance in the estimates of toxin production obtained suggests that the ELISA lacks the sensitivity of cell-based cytotoxicity assay. Toxin neutralisation assays were used to ensure that the observed findings were due to TcdA and TcdB and not another artefact in the supernatant. This agrees with other published data which have shown that *tcdR* is a regulator of toxin but is not required to initiate expression of *tcdA* and *tcdB*[75,127,128].

Reverse genetic analysis as described above is suitable in situations where a gene of interest has been identified and a hypothesis is available. However, in *C. difficile* strain 630 there are 3,903 identified genes and of these 53% encode either putative, unknown or pseudo proteins[175], with similar figures seen in NM-R20291[154]. To identify novel genes for regulation of any system using a reverse genetic method would be inappropriate without first an idea of a

phenotype. Forward genetics provides a more convenient way to identify these novel regulators so long as there is a suitable screening method.

Random mutagenesis has been employed to identify altered sporulation, auxotrophic phenotypes and gene essentiality in *C. difficile*[97,176] but there is no described method for assessing altered toxin phenotypes. To overcome the laborious method of measuring toxin concentrations in toxin supernatants, i.e. cell culture or ELISA, a reporter strain was created. The reporter strain RS-7 was created previously in an NM-R20291 background where *tcdB* was replaced with *licB* of *C. thermocellum* a lichenase gene. This resulted in the creation of an easily identifiable phenotype when mutants with an altered phenotype were screened on agar plates containing lichenase.

Random mutant libraries were created and screened for diversity previously. From said libraries eight gene candidates were identified with altered lichenase patterns when compared to RS-7 suggesting either a direct or indirect link to toxin regulation (Table 3-1). One random mutant produced a striking phenotype in which no lichenase activity was detected at 24 h and reduced lichenase activity was identified at 48 h. This mutant designated BBM4 was found to have two transposon insertions within the genome in *R20291_2908* and in the intergenic spacer region of *jag* and *oxaA1*.

An InterPro search of *R20291_2908* revealed a hypothetical protein containing two functional domains, a ParB-like and HNH nuclease motif domain. ParB (IPR003115) is involved in chromosomal partition and a HNH nuclease domain is associated with the catalytic centre of homing endonucleases. Recent literature suggests this protein maybe involved in a programmed cell-death system and / or Type I restriction modification systems[177]. In *Bacillus subtilis* there are orthologues of *jag* (*spoIIIJ*) and *oxaA1* (*ygjG*) that have been shown to be involved in sporulation[178]. Inactivation of *spoIIIJ* in *B. subtilis* resulted in cells that are unable

to sporulate as SpoIIIJ is essential for a pre-spore specific σ-factor activation[178]. Both *spoIIIJ* and *ygjG* are linked with outer membrane biogenesis, including insertion of subunits of the cytochrome or oxidase and ATP synthase complexes[178].

ClosTron mutants were created to interrupt the genes where transposon insertion had occurred resulting in strains NM-R20291::*jag*(*ermB*) and NM-R20291::2908(*ermB*). The ClosTron insertional inactivation had no effect on growth for either of the mutant strains. Data at 48 h showed a significant difference in toxin concentration for NM-R20291::2908(*ermB*) but not NM-R20291::*jag*(*ermB*), suggesting this is the gene responsible for the lichenase phenotype. To confirm this complementation and overexpression plasmids were created containing the native NM-R20291_2908 promoter (pMTL84151::p2908) and a thialase promoter (pMTL84152::2908) respectively. As discussed earlier, results were inconsistent after the introduction of the complementation and overexpression plasmids even in identical experimental conditions. Unusual results included all four strains producing the same concentration of toxin and both NM-R20291 and NM-R20291::2908(*ermB*) producing reduced levels of toxin when compared to the complementation and overexpression strains. After consideration these data maybe due to during the initial growth and toxin assays, NM-R20291::*jag*(*ermB*) and NM-R20291::2908(*ermB*) were miss labelled and it was in fact NM-R20291::*jag*(*ermB*) that produced the reduced toxin phenotype. Therefore, the complementation would not produce an altered phenotype. The current literature would also suggest that the more likely candidate to produce an altered toxin phenotype would be the NM-R20291::*jag*(*ermB*) mutant as there are suggestions that sporulation and toxin production are linked[165].

Screening of transposon mutants could be improved by the implementation of a rapid semi-quantifiable assay. This can be achieved by measuring the relative zone of clearance on TY agar with 1% lichenan and subtracting the diameter of the colonial growth when using 20 µL

of overnight culture at 24, 48 and 72h. The average size of triplicate replications would be compared to the calculated ZOC produced by the control (RS-7). A "more than" "less than" result would be definitive as an effect on PaLoc regulation.

A liquid semi-quantitative assay was in part developed for use in *C. difficile*, through the adaptation of an assay proved to work within our group in *C. botulinum*. The early stages of this development used culture medium that did not contain lichenan. This substrate was added at the time of the enzymatic assay. This resulted in no measurable difference in A_{530} in supernatants taken at 48 h from transposon mutants and the control strain in a *C. difficile* host. Dr. E. Stevenson showed that the addition of NaOH appeared to have a role in this occurrence, demonstrated by a difference in A_{530} if NaOH is added after the colour change reaction. As this does not result in a defined endpoint we assessed different enzyme altering solutions (0.5M Sodium acetate, 0.5M Sodium carbonate and 2% Trizma base) to end the reaction along with different enzyme incubation periods (10, 20, 30 and 40 minutes).

Unsurprisingly, with supernatants taken from 48 h cultures, the difference in ΔA_{530} increased with correlation to the increased enzyme incubation time. 0.5M Sodium carbonate produced a significantly increased difference in ΔA_{530} when compared to 0.5M Sodium acetate and 2% Trizma base after a 40 minute incubation. A 48 h assay was performed to assess the new protocol over a set time period. No measurable difference in A_{530} could be seen in supernatants taken at <48 h and the results seen at 48 h were not reproducible over several attempts. Even though the original method was shown to work in *C. botulinum*. The reason for the differences in the quantitative assays between reporter strains created in a *C. difficile* and a *C. botulinum* host have not been investigated here but may be due to subtle differences in growth and toxin regulation. Although not completely comparable due to too many variables i.e. different growth media, *C. botulinum* is able to grow to a higher OD within exponential phase and produce ~10x more toxin within 24 h compared to *C. difficile* strain R20291[78,179]. This could

suggest that a high enough concentration of lichenase is not being produced for a reliable assay in *C. difficile*.

On refection, the main issue with the semi-quantitative liquid assay was most probably due to the length of time that the lichenase could act on the lichenan. In the plate assays this time-period was between 24 and 48 h whereas in the liquid assay the maximum time was 40 min. To see assess if this was the case the experiment should be repeated but using TY with 1% lichenan and halting the enzymatic reaction after the removal of 1 mL aliquot at each time-point. A solution of Congo red would be added after the enzymatic reaction for a set period of time and the absorbance measured. This data when compared to the control sample, a supernatant of liquid culture containing only the reporter strain RS-7 taken at the same time points, would be used to calculate the $\Delta A_{530.}$ This assay would require some optimisation.

Other screening methods have been considered, these would ideally be designed using a wild-type strain as the host for the mutant library generation. This is so that the method could be used in multiple *C. difficile* strain backgrounds without the need to create new reporter strains which can be laborious. One alternative is using the methods described by Darkoh *et al*.[180,181], where they utilise TcdA and TcdB ability to cleave a chromogenic substrate, p-nitrophenyl-β-D-glucopyranoside (PNPG), that has stereochemical characteristics similar to their natural substrate UDP-glucose. They have demonstrated that this method can work in both a plate assay[180] and a semi-quantitative assay[181] filling both requirements for a robust screening method.

3.4 Key outcomes

- Significant reduction of TcdA and TcdB concentration was observed in a *tcdR* null NM-R20291 strain after 12 h of growth.

- Altered toxin phenotypes can be identified using the RS-7 reporter strain and the lichenase plate assay.
- Either gene NM-R20291::*jag* or NM-R20291::2908 has an effect on toxin production.

3.5 Future work

- Repeat the original NM-R20291, NM-R20291::*jag*(*ermB*) and NM-R20291::2908(*ermB*) toxin assay to confirm which strain showed the reduced toxin phenotype.
- Perform complementation and over expression assays.
- Optimise the semi-quantitative assay using TY and 1% lichenan as the culture medium.
- Trial screening of random mutant libraries using p-nitrophenyl-β-D-glucopyranoside as a chromogenic substrate.
- Once these assays are optimised, more potential toxin regulators can be identified using these systems.

Chapter Four Whole Genome Analysis of *Clostridium difficile* strain VPI 10463

4.1 Introduction

4.1.1 Whole Genome Sequencing a Brief History

The first bacterial genome sequences were completed in 1995 by Fraser *et al.*[182] and Fleischemann *et al.*[183] in *Mycoplasma genitalium* and *Haemophilus influenza* respectively. These projects employed the first generation sequencing method Sanger sequencing to produce shotgun sequenced DNA libraries. The principle behind Sanger sequencing is being able to detect the termination of DNA chain elongation by the incorporation of didexoyNTPs (ddNTPs). Since its inception the process has been improved, originally four separate reaction tubes were required and sequence identification was performed using visualisation of the size of the bands after denaturing gel electrophoresis and autoradiography[184]. Fluorescently labelled ddNTPs improved this method by allowing visualisation after capillary electrophoresis via computer[185]. This method, although accurate and able to produce relatively long reads (~1000bp), is time consuming, laborious and expensive. It also has difficulty in defining long repetitive regions of DNA due to the read length produced.

In 2004 the 454 Pyrosequencer was the first second generation sequencing platform to become commercially available and has a 100-fold greater throughput compared to Sanger sequencing[186]. The increase is due to the parallelism in which multiple DNA strands can be sequenced at the same time. Emulsion PCR is utilised to amplify a single DNA template fixed to a bead by a DNA adapter, beads are then planted in to picolitre-volume wells, one per well. During the sequencing by synthesis only one type of dNTP is available at any one time. The polymerase activity is coupled to the activity of a luciferase by an ATP sulfurylase, the amount of light generated is proportional to the number of dNTPs incorporated at a time. After the reaction an apyrase degrades the ATP and dNTPs present, a different dNTP is added and this starts the process again. Disadvantages of this system compared to Sanger sequencing are the inability to resolve long homopolymer runs due to the saturation of the charge coupled device

(CCD) which is used to record the amount of light created and the reduced read length to ~400bp.

Illumina sequencing was the next second generation sequencing platform to be released. The method is similar to the 454 pyrosequencing in that it utilises sequencing by synthesis but adapter linkage is to a glass slide separated into eight lanes[187]. Fragmented DNA is bound to the slide and DNA clusters are formed during amplification resulting in approximately one million copies of the original fragment. The attachment of different adaptors allows several genomes to be sequenced at once (multiplexing). As with the 454 pyrosequencing method significantly shorter read lengths are achieved compared to Sanger sequencing[187].

The second generation sequencing platforms made whole genome sequencing more accessible as costs were reduced. However, these technologies are better suited to re-sequencing projects, where a reference genome is available, rather than for *de novo* assembly[188]. At an assembly level this is because the short reads are unable to accurately resolve repeat sequences and may misassemble sequences[189]. Regions of high or low coverage might signify a single polymorphic locus that has been classified as two distinct loci or a merged repetitive region respectively[190]. At the sequencing level errors can be substitutions, insertions or deletions but these should be overlooked in the consensus sequence after assembly.

More recently third generation sequencing platforms have become available, one example being Pacific Biosciences (PacBio) single molecule real time sequencing technology (SMRT). This method also uses a sequencing by synthesis approach, it utilises a polymerase held in a zero-mode waveguide which visualises the incorporation of a single fluorescently labelled nucleotide[191]. This technology has improved read length by 30 to 200 times compared to second generation platforms[192]. Due to this issues with *de novo* assemblies are reduced allowing resolution of repeat sequences, closing gaps and increasing quality over GC rich or

poor regions[192]. It should be kept in mind that individual reads do contain more errors than what is seen in short read sequences, but with sufficient coverage an accurate consensus sequence is achievable[192]. To further check the quality of the consensus sequence, Illumina short reads can be used for corrections.

Not all sequencing platforms have been discussed in this section; a comprehensive list can be seen in **Table** 4-1.

Table 4-1: Currently available next generation sequencing platforms. Taken from[193]

Machine (manufacturer)	Chemistry	Modal read length* (bases)	Run time	Gb per run	Current, approximate cost (US$)‡	Advantages	Disadvantages
High-end instruments							
454 GS FLX+ (Roche)	Pyrosequencing	700–800	23 hours	0.7	500,000	• Long read lengths	• Appreciable hands-on time • High reagent costs • High error rate in homopolymers
HiSeq 2000/2500 (Illumina)	Reversible terminator	2 × 100	11 days (regular mode) or 2 days (rapid run mode)§	600 (regular mode) or 120 (rapid run mode)§	750,000	• Cost-effectiveness • Steadily improving read lengths • Massive throughput • Minimal hands-on time	• Long run time • Short read lengths • HiSeq 2500 instrument upgrade not available at time of writing (available end 2012)
5500xl SOLiD (Life Technologies)	Ligation	75 + 35	8 days	150	350,000	• Low error rate • Massive throughput	• Very short read lengths • Long run times
PacBio *RS* (Pacific Biosciences)	Real-time sequencing	3,000 (maximum 15,000)	20 minutes	3 per day	750,000	• Simple sample preparation • Low reagent costs • Very long read lengths	• High error rate • Expensive system • Difficult installation
Bench-top instruments							
454 GS Junior (Roche)	Pyrosequencing	500	8 hours	0.035	100,000	• Long read lengths	• Appreciable hands-on time • High reagent costs • High error rate in homopolymers
Ion Personal Genome Machine (Life Technologies)	Proton detection	100 or 200	3 hours	0.01–0.1 (314 chip), 0.1–0.5 (316 chip) or up to 1 (318 chip)	80,000 (including OneTouch and server)	• Short run times • Appropriate throughput for microbial applications	• Appreciable hands-on time • High error rate in homopolymers
Ion Proton (Life Technologies)	Proton detection	Up to 200	2 hours	Up to 10 (Proton I chip) or up to 100 (Proton II chip)	145,000 + 75,000 for compulsory server	• Short run times • Flexible chip reagents	• Instrument not available at time of writing
MiSeq (Illumina)	Reversible terminator	2 × 150	27 hours	1.5	125,000	• Cost-effectiveness • Short run times • Appropriate throughput for microbial applications • Minimal hands-on time	• Read lengths too short for efficient assembly

*Average read length for a fragment-based run. ‡Approximate cost per machine plus additional instrumentation and service contract. See REF. 58. §Available only on the HiSeq 2500.

4.1.2 C. difficile strain VPI 10463

C. difficile strain VPI 10463 (reference strain ATCC 43255) was first isolated from an abdominal wound and was the first strain to have both TcdA and TcdB purified[194]. Although the PaLoc sequence in other strains had been partially characterised it was in VPI 10463 that the full 19.6Kb toxigenic element was identified[73]. This study also showed that there were five ORFs present within the element, later known as *tcdR*, *tcdB*, *tcdE*, *tcdA* and *tcdC*, although only *tcdR* had not been previously identified[195]. It was similarly confirmed that the PaLoc was located on the chromosome during this study[73]. A few years later a new toxinotyping scheme was developed for epidemiological reasons and VPI 10463 was used as the reference strain[49].

VPI 10463 has been utilised in a number of TcdA and TcdB regulatory studies namely due to its ability to produce high levels of these toxins *in vitro*[127,136,144,196,197] and is frequently used in *in vivo* models[198–200]. Despite this, even since the advances of molecular techniques, there has been very little to no genetic manipulation performed in this strain. Reasons for this remain unclear but it seems possible that there is a defence mechanism that is preventing DNA transfer into this strain and potentially other *C. difficile* strains, possibly a restriction modification system. To identify such mechanisms and for future genetic manipulations a high quality reference genome is required. At the beginning of this project there was an unpublished draft genome for VPI 10463 available through the NCBI submitted by the McGill University (GenBank CM000604.1). The whole genome shotgun sequence contained 22 contigs which were ordered and orientated against the complete 630 genome along with 32 contigs which could not be placed with certainty. We aimed to improve upon the genome sequence for use in future molecular studies in VPI 10463 as well as improve DNA transfer into this strain.

4.1.3 Experimental Aims

- Produce a high-quality reference genome for strain VPI 10463.
- Comparative analysis of strain VPI 10463 against 630 and BW-R20291.

- Identify putative genes that might be involved in TcdA and TcdB regulation in VPI 10463 that are not present in 630 and R20291.
- Identify and circumvent mechanisms that prevent efficient DNA transfer in strain VPI 10463.

4.2 Results

4.2.1 Phenotypic analysis of strain VPI 10463.

Although it has been shown previously that VPI 10463 can produce higher levels of toxin compared to 630, BW-R20291 and M120 (a low toxin producer), initially we wanted to check this phenotype. A 48 h growth curve and toxin assay was performed in TY broth over a 48h time course as described in 2.9.1 for this purpose. Toxins A and B were measured in the supernatants by *C. difficile* Tox A/B II ELISA (TechLab), detection limit 1.0 ng/mL, and toxin concentrations were determined using a standard curve (TcdA & TcdB mixture at 125 ng/mL – 1.95 ng/mL). The results confirmed previous findings and showed that although all the strains grew in a similar manner (Figure 4-1A) VPI 10463 produced 10-fold more toxin than both 630 and BW-R20291 and 100-fold more than M120 at the 48 h time point (Figure 4-1B).

4.2.2 Whole Genome sequencing

Whole genome sequencing was performed by both PacBio RSII and Illumina MiSeq technologies. Illumina MiSeq was performed by University of Nottingham DeepSeq Facility and raw reads were returned after filtering reads with low sequencing score as well as reads aligned to adaptor sequences. DeepSeq firstly trimmed raw reads against adaptors using Scythe (https://github.com/vsbuffalo/scythe). Then reads were quality trimmed using Sickle (https://github.com/najoshi/sickle). A total of 1,294,739 paired end reads were obtained with a sequence length of 35-251 and an average GC content of 28%. Pacbio RSII sequencing was performed at the McGill University and Genome Quebec using a sheared large insert library and one SMRT cell. Library and assembly statistics can be seen in Table 4-2.

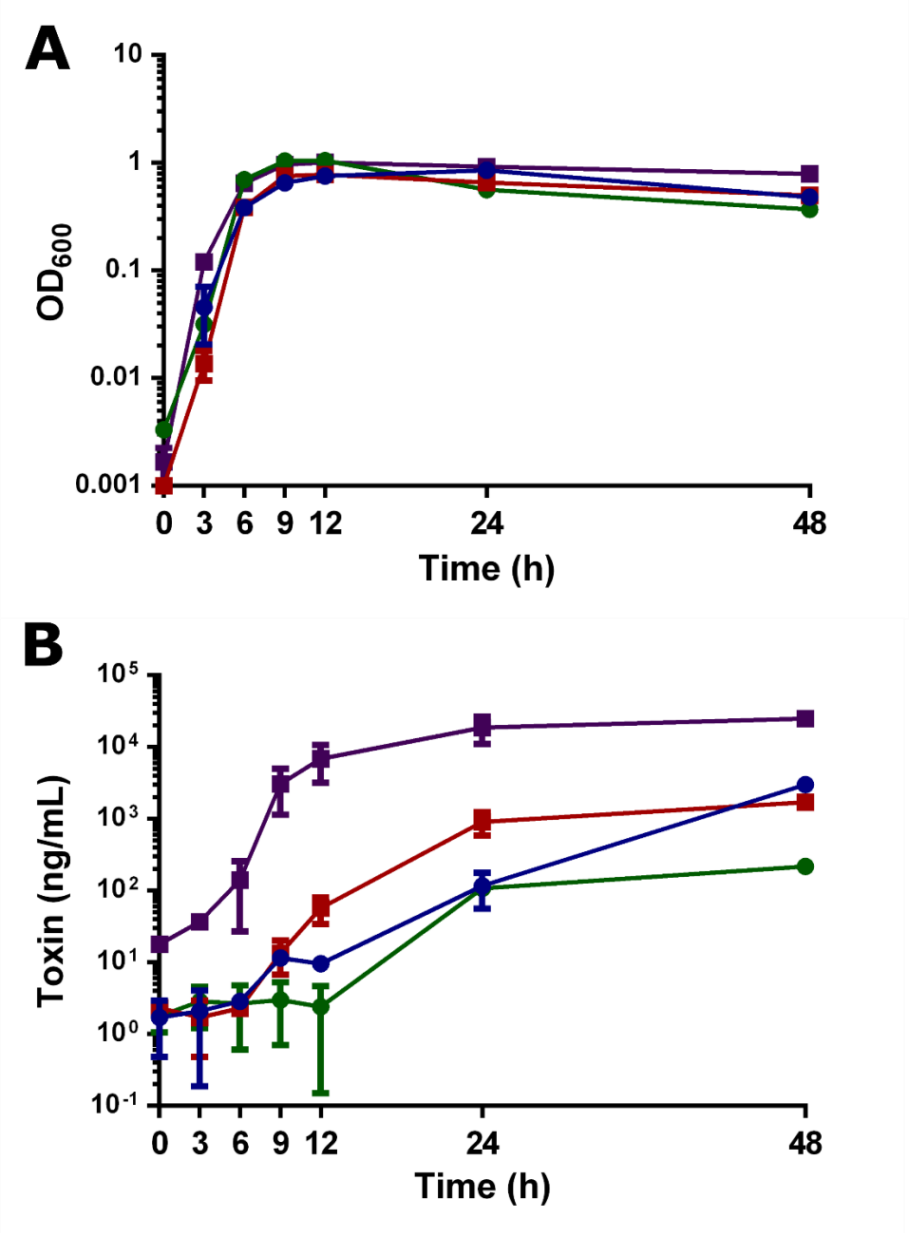


Figure 4-1: Cell growth and toxin production by *C. difficile* strains VPI 10463, 630, BW-R20291 and M120. Figure legend; (●) BW-R20291; (■) 630; (■) VPI 10463; (●) M120. (A) Growth was measured at OD_{600}. (B) Toxin concentration determined by *C. difficile* Tox A/B II ELISA using a standard curve of known Toxin A and B concentrations. Data represent the mean ± SEM (n=3).

As a full consensus sequence was not achieved through PacBio RSII sequencing alone, the generated raw long reads were corrected with the MiSeq paired end reads by DeepSeq, University of Nottingham using proovread[116], a hybrid correction pipeline. The resulting corrected reads were assembled using Canu, a fork of the Celera Assembler designed for high-noise single-molecule sequencing[201]. The resulting assembly produced six contigs, the largest at 4,319,225bp. The remaining five contigs were mapped back onto the largest contig using CLC genomics workbench V8.5.1 and therefore discarded. Analysis of the sequence showed complementary sequences at each end of the sequence comprising of 5224bp suggesting the contig is circular and encompassing of the whole genome. This complementary sequence was trimmed at one end of the sequence and the consensus sequence circularised.

Table 4-2: PacBio sequencing library and assembly statistics.

Library size	**15,000**
Raw average read length	10,888bp
All subreads	
Count	108,272
Mean	4,545
bp	492,096,256
Longest >3Kb	
Count	29,870
Mean	8,022
bp	239,617,136
Longest >7Kb	
Count	13,368
Mean	12,100
bp	161,752,800
Longest >12Kb	
Count	5,147
Mean	16,798
bp	864,459,304
Estimated coverage (X)	110
Total contigs	22
Minimum contig length	3142
Maximum contig length	2576293
N50	2576293

4.2.3 Annotation of the VPI 10463 genome

The corrected genome sequence was annotated using the Joint Genome Institute (JGI) Integrated Microbial Genomes & Microbiome Samples (IMG/MS) annotation pipeline MGAP v.4[202,203]. Genome statistics can be seen in Table 4-3. Functional Clusters of Orthologous Groups (COG) categories were predicted during the annotation process and can be seen in Table 4-4.

Table 4-3: General genome statistics of strain VPI 10463 as predicted by the JGI IMG annotation pipeline

	Number	% of total
DNA total number of bases	4,314,004	100
Number of coding bases	3,631,788	84.19
Number of G/C bases	1,243,772	28.83
Number of genes	4037	100
Protein coding genes	3887	96.28
Protein coding genes with function prediction	3111	77.06
RNA genes	150	3.72
rRNA genes	35	0.87
5s rRNA	11	0.27
16s rRNA	12	0.30
23s rRNA	12	0.30
tRNA genes	90	2.23

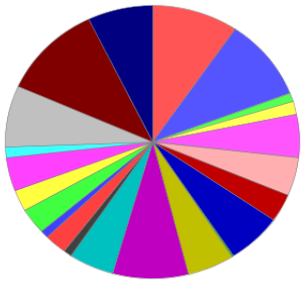
For ease of comparison, the sequence was then orientated with *dnaA* as the first CDS as is with strains 630 and R20291.

4.2.4 Comparative analysis of 630, R20291 and VPI 10463

To assess the overall similarity of VPI 10463 to 630 and R20291 a Mauve[120] alignment was performed using standard settings. Analysis shows there are five conserved regions within the genomes each depicted by a locally collinear block (Figure 4-2). Interestingly there appears to be a large inversion in VPI 10463 which is flanked by two lineage specific regions (positions 308054 – 358721, 3849698 – 3900424). The zoomed in images in Figure 4-2 show an ~21kb homologous sequence (blue dotted boxes at positions 317517 – 338086, 3866365 – 3887294) containing 37 CDS with functions ranging from iron (III) transport systems, FAD synthetase, Holliday junction resolvase, several phage related genes to multiple hypothetical proteins. The remaining sections of the lineage specific regions are highly similar but not homologues. It is interesting to note that within each lineage specific region there are three genes that are involved in recombination events; *xerD, dnaD* and *recT*. Although their potential involvement in this recombination event has not been experimentally demonstrated.

Table 4-4: **Predicted COG categories for *C. difficile* strain VPI 10463. Table shows number of genes and percentage total. The associated pie chart shows the breakdown of the functional categories.**

COG Categories	Gene Count	Percent
Amino acid transport and metabolism	1057	9%
Carbohydrate transport and metabolism	1078	10%
Cell cycle control, cell division, chromosome partitioning	116	1%
Cell motility	169	2%
Cell wall/membrane/envelope biogenesis	570	5%
Chromatin structure and dynamics	3	0%
Coenzyme transport and metabolism	502	5%
Cytoskeleton	4	0%
Defence mechanisms	367	3%
Energy production and conversion	674	6%
Extracellular structures	21	0%
Function unknown	573	5%
General function prediction only	925	8%
Inorganic ion transport and metabolism	553	5%
Intracellular trafficking, secretion, and vesicular transport	104	1%
Lipid transport and metabolism	277	2%
Mobilome: prophages, transposons	99	1%
Nucleotide transport and metabolism	341	3%
Posttranslational modification, protein turnover, chaperones	276	2%
Replication, recombination and repair	448	4%
Secondary metabolites biosynthesis, transport and catabolism	145	1%
Signal transduction mechanisms	808	7%
Transcription	1243	11%
Translation, ribosomal structure and biogenesis	795	7%



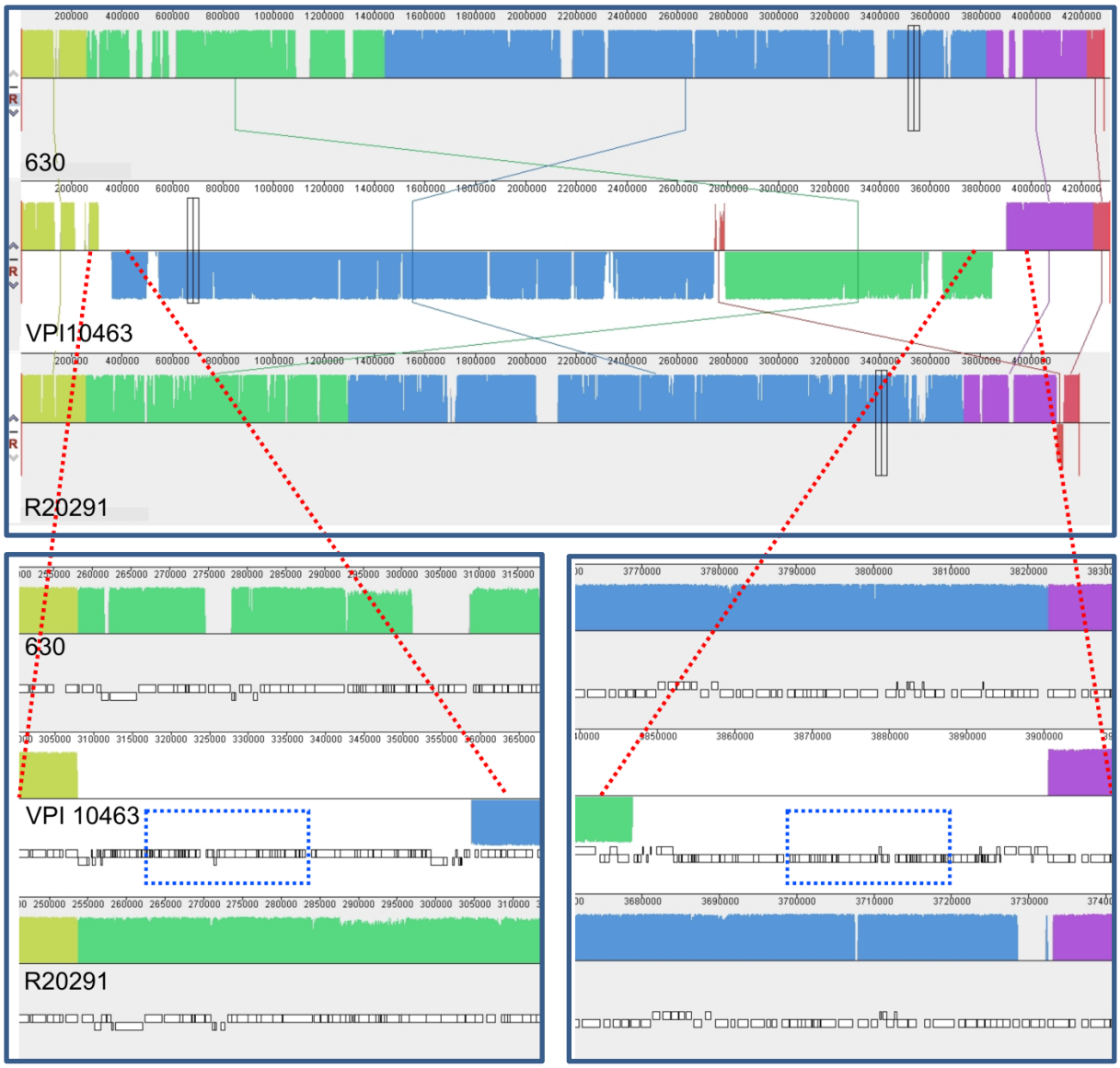


Figure 4-2: Mauve genome alignment of *C. difficile* strains. Top 630; middle VPI 10463 and bottom R20291. Alignment shows five conserved regions in each genome depicted by a different colour for each. VPI 10463 appears to have undergone a large recombination event depicted by the blue and green conserved regions having been shifted down. Red dotted lines indicate a zoomed in view of the region. The zoomed in view shows two unique regions that flank the inversion. The blue dotted boxes show two regions that are reverse complements of each other (positions 317517 – 338086, 3866365 – 3887294).

An inversion like this has not been previously described in *C. difficile*, to check these data the PacBio/Illumina corrected reads provided by DeepSeq were *de novo* assembled again in Canu in-house using the standard settings. The resulting consensus sequence was then compaired to the orignial genome sequence returned from DeepSeq (Figure 4-3). Because the new genome

mapping was shown to be almost identical to the annotated DeepSeq genome sequence all further work was performed on the sequence VPI 10463_DS. The corrected PacBio/Illumina sequences were mapped back to VPI 10463_DS genome sequence using the Burrows-Wheeler Aligner (BWA)[117] using standard settings and the PacBio index flag. The output files were then sorted and indexed using SamTools[118]. Commands used in this process can be found in Appendix Three.

Mapping data was analysed using CLC genomics workbench version 10.1.1. The mapping statistics showed an average coverage of 93x (minimum 21x & maximum 238x) no zero coverage regions and a total of 103,808 mapped reads. The read mapping over the two homologous and proposed inversion regions was assessed in more detail. Using CLC genomics workbench it was shown that sequence specific mapping occurred at regions spanning the beginning and end of the homologous regions as indicated by green (forward) and red (reverse) reads in the middle regions of Figure 4-4 A&B. This specifies that these mapped sequences are specific and only occur once in the whole genome and that the flanking regions are unique to each homologous region. The mapping also showed non-specific mapping as indicated by yellow reads in the middle of these homologous regions further confirming that these two 21kb sequences are the same but the flanking regions are unique (Figure 4-4).

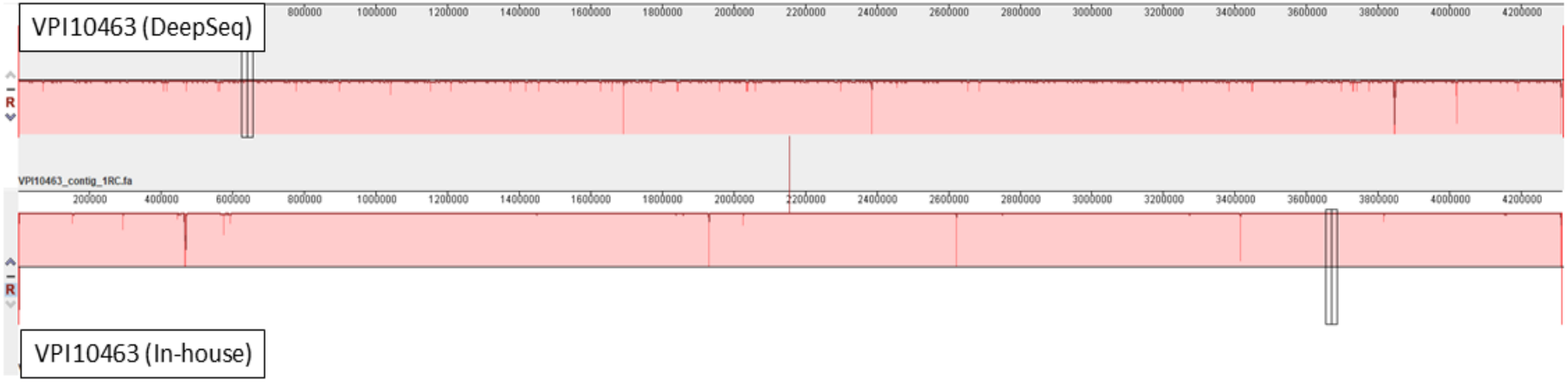


Figure 4-3: Genome comparison of VPI 10463 using Mauve. Top genome sequence was provided by DeepSeq at the University of Nottingham. The bottom genome sequence was recreated using the corrected PacBio/Illumina reads in Canu. The comparison produced one locally collinear block indicating that the two sequences are highly similar. The genome assembly produced in-house is presented in the opposite orientation to the genome sequence provided by DeepSeq which had been previously reverse complemented to begin at *dnaA* on the sense strand.

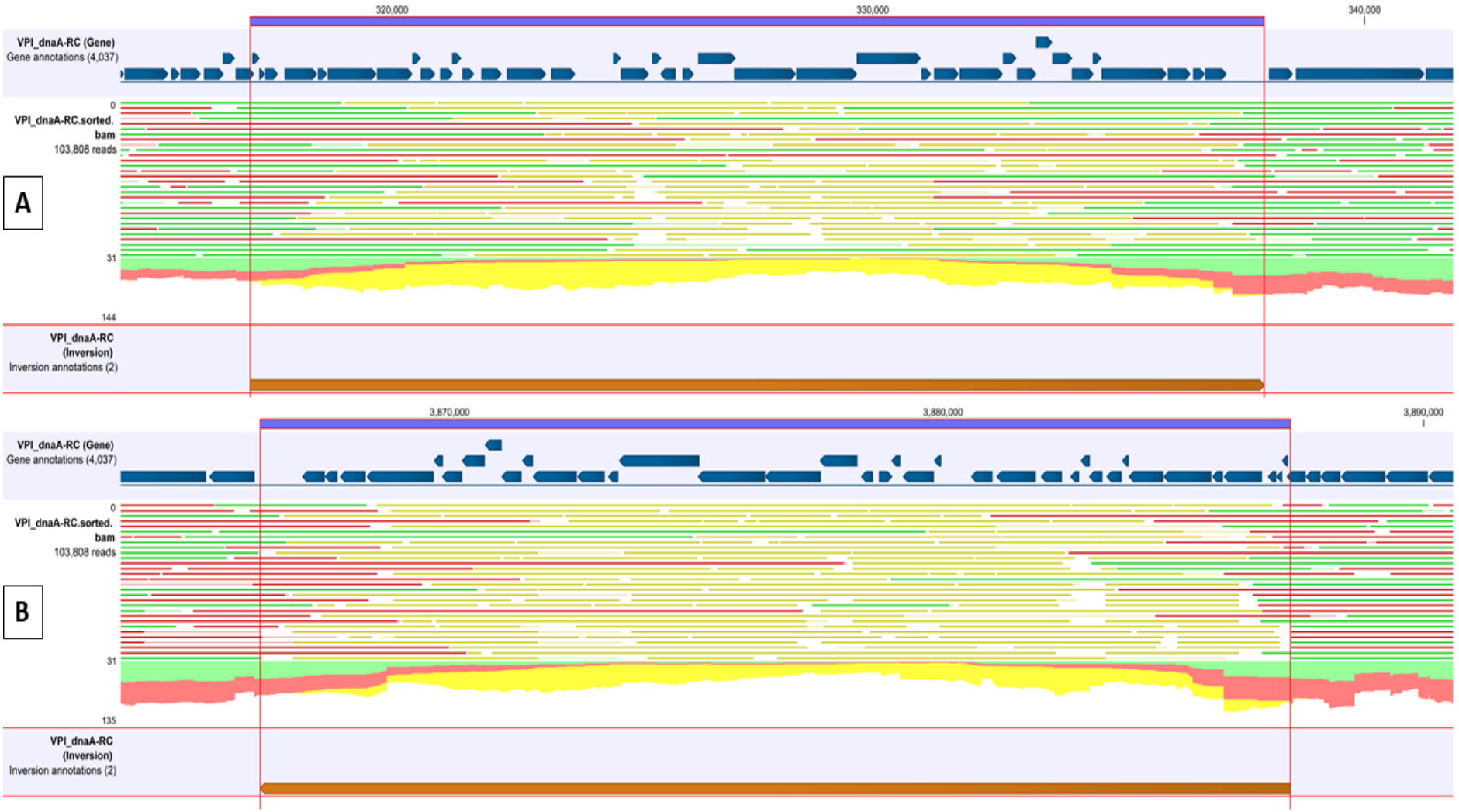


Figure 4-4: Mapping coverage over the two proposed inversion regions in VPI 10463. Figure legend (A) Region 1 of the homologous sequence at positions 308054 – 358721; (B) Region 2 of the homologous sequence at positions 3849698 – 3900424. Top section for A & B show the CDS location and orientation for the specific region being analysed, depicted by blue arrows. Middle section for A & B shows the sequence mapping over each

respective region; green mapping indicate alignment in the forward orientation; red mapping indicate alignment in the reverse orientation; yellow mapping indicates non-specific alignment signifying that these reads can map in more than one place in the sequence. These data show location specific mapping over the junction regions of the homologous sequencing indicating that these are unique locations within the genome sequence. Bottom region for A & B shows an orange arrow depicting the exact position of the homologous sequence in each location.

Using the Mauve export orthologues option with standard settings a list of putative orthologues was created. This showed the pan-genome in these strains consists of 4680 genes with 512 genes unique to strain VPI 10463 (Figure 4-5), of these 512 genes 284 are categorised as hypothetical proteins. The remainder of unique genes comprises of phage and transposon related genes, transcriptional regulators, recombinases and antibiotic resistance genes. A comprehensive list is available in Appendix Four.

During analysis, a second holin like gene was identified as a candidate gene which could potentially explain the increased toxin production in VPI 10463. There were also two putative type I restriction modification systems identified as candidates for the reduced ability to transfer DNA into this strain. These are discussed in more detail later in this chapter.

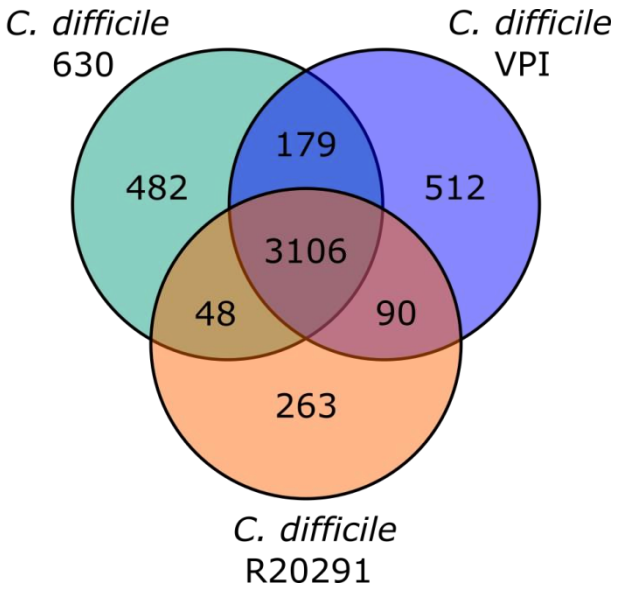


Figure 4-5: Predicted orthologues for *C. difficile* strains 630, VPI 10463 and R20291. The pan-genome comprises a total of 4680 genes with 512 being unique to VPI 10463.

4.2.5 Flagella operon analysis.

As discussed in 3.1.2 there has been evidence to suggest that the flagella operons play a role in the regulation of toxin production and so we looked at the three flagella operons at the sequence level. An ACT comparison revealed that VPI 10463 shows an 89% and 96% sequence identity

in the F1 and F3 operons respectively, but not in the F2 operon when compared to 630. There is a 98% sequence identity in all three operons when compared to R20291 (Figure 4-7). Comparison of the genes at sequence level revealed two potential mutations in FlgE and FliJ, which might explain the increased toxin phenotype. To check what affect the altered amino acid sequence would have on function both sequences were analysed using PROVEAN compared to the amino acid sequences of the same proteins found in R20291. The analysis indicated that all the changes were neutral in FlgE but a change in DNA sequence in FliJ resulted in a nonsense mutation (*fliJ*111del results in an I37X). As this change in DNA sequence occurs in a sequence of nine adenine residues (Figure 4-7) it was confirmed by Sanger sequencing to ensure it was not a sequencing and/or assembly error.

Figure 4-6: ACT comparison of the three flagella operons. Top 630; middle VPI 10463 and bottom R20291. Green box indicates F1 operon, red box indicates F2 operon and blue box indicates F3 operon. F1 and F3 operons are highly similar between all three strains. The F2 operon is divergent in both VPI 10463 and R20291 when compared to 630.

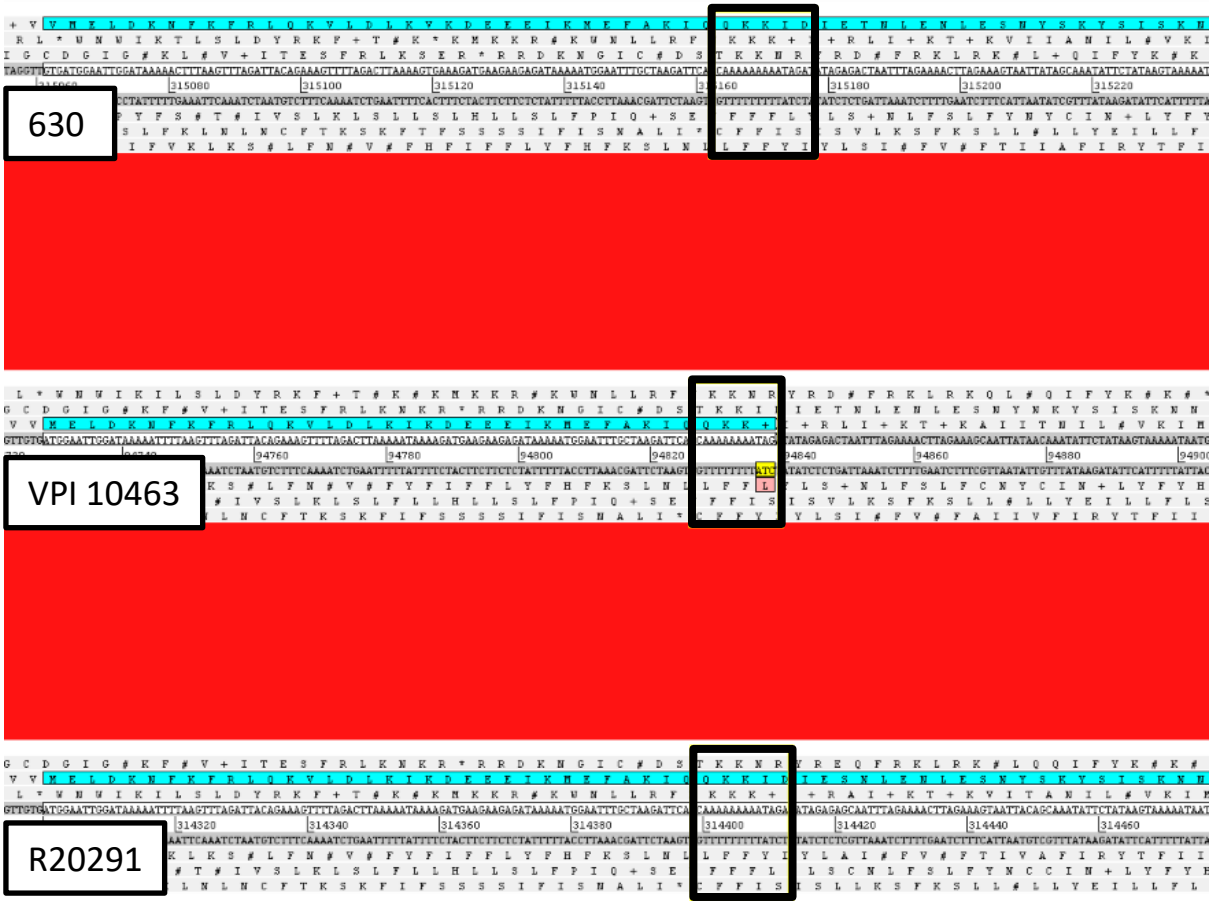


Figure 4-7: ACT comparison of *fliJ*. Top 630; middle VPI 10463 and bottom R20291. Black box indicates a sequence of nine adenine residues. In VPI 10463 fliJ111del results in an I37X and a truncation of the protein FliJ.

To evaluate if the nonsense mutation in *fliJ* affected the motility phenotype plate motility assays using 0.3% BHIS top agar were performed. This showed that both 630 and BW-R20291 were motile as previously described (Figure 4-8A & B), M120 and VPI 10463 had a reduced motility (Figure 4-8C & D). VPI 10463 showed a statistically significant reduced motility compared to these 630 and BW-R20291 (Figure 4-8E). To check if this phenotype was due to a lack of flagella we employed transmission electron microscopy which revealed that VPI 10463 is monotrichous, whereas 630 and BW-R20291 are both peritrichous as reported previously. M120 does not produce flagellar (Figure 4-9).

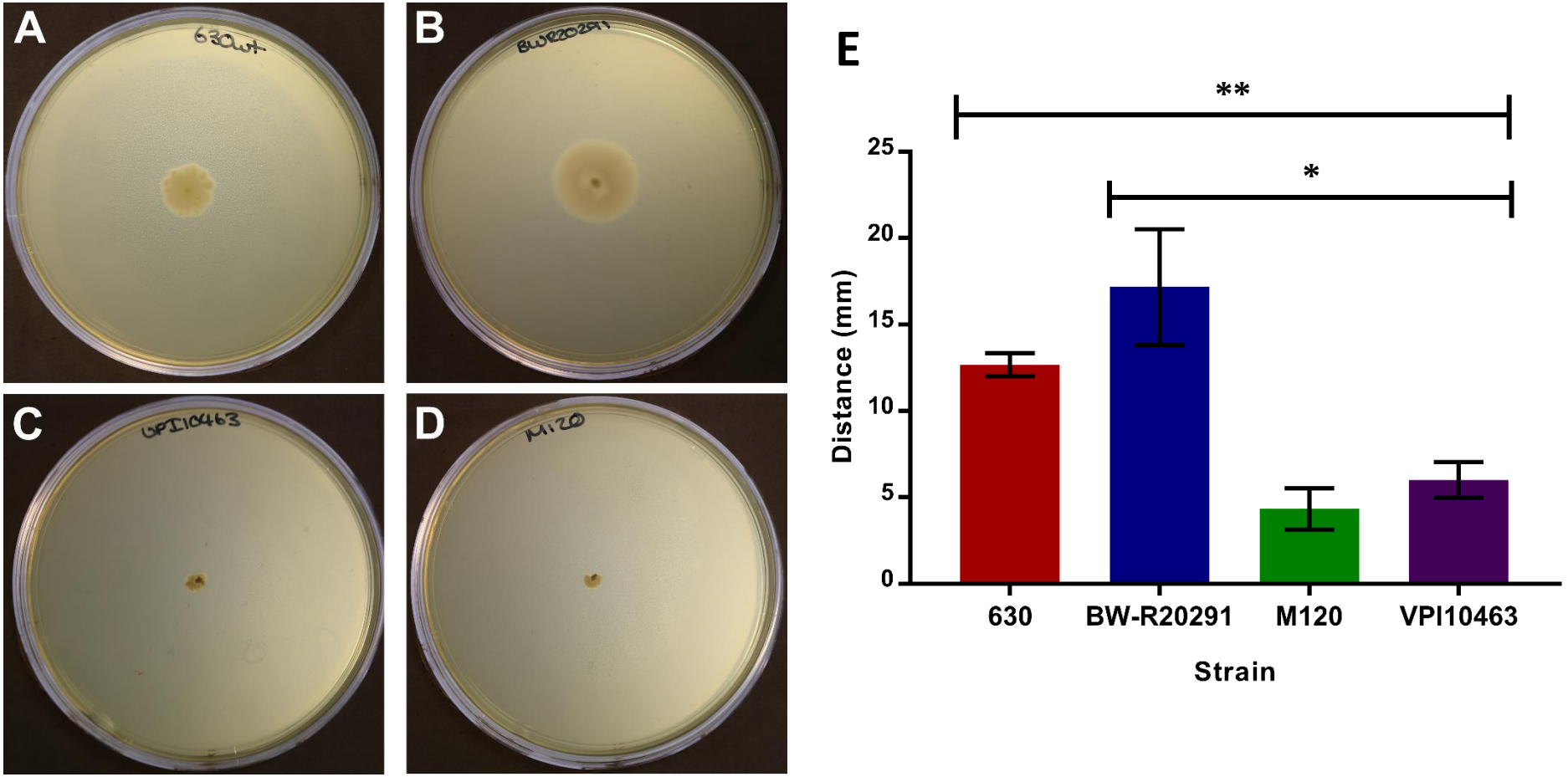


Figure 4-8: Motility assay for Clostridium difficile strains. Figure legend; (A) 630, (B) BW-R20291, (C) VPI 10463 and (D) M120 on 0.3% BHIS Top agar. Each image represents one example of triplicate experiments. Single colonies were stab inoculated into the top agar using a sterile toothpick and incubated at 37°C overnight. (E) Motility zones were measured in two directions and averaged after 24 h incubation. Bar chart data represent the mean ± SEM (n=3). ** indicates statistical significance (p = <0.005) by unpaired T-test between 630 and VPI 10463. * indicates statistical significance (p = 0.03) by unpaired T-test between BW-R20291 and VPI 10463.

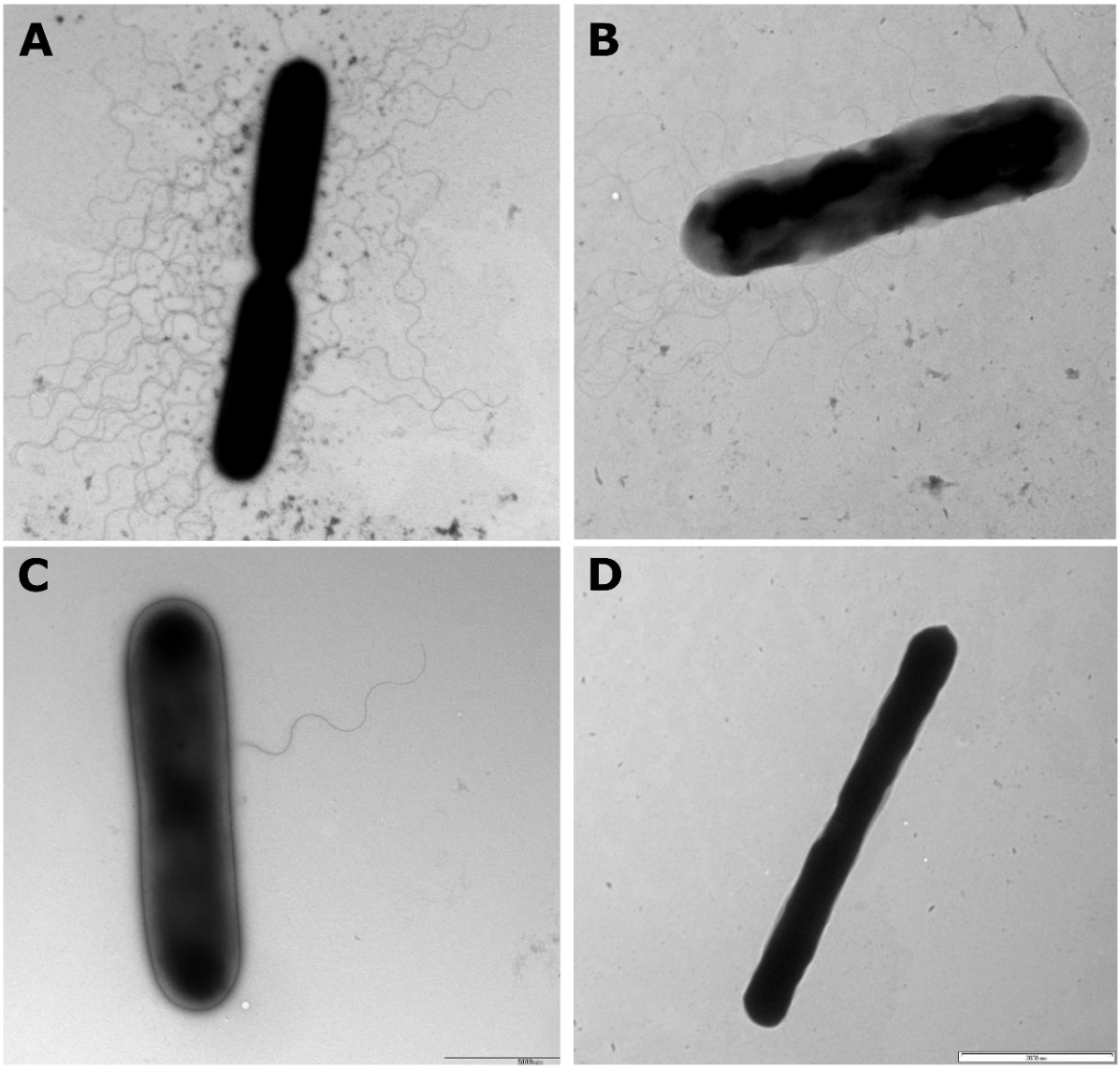


Figure 4-9: Transmission electron micrograph of *Clostridium difficile* strains. Figure legend (A) 630, (B) BW-R20291, (C) VPI 10463 and (D) M120. Images reveal strains 630 and BW-R20291 are peritrichous, M120 have no flagella and VPI 10463 are monotrichous.

4.2.6 Second Holin-like gene

The unexpected second holin-like gene (*Ga0114281_1134*) was studied in more detail. A BLAST search showed that this gene shows 100% query coverage and identity to genes identified in two *C. difficile* strains, two *C. difficile* plasmids and one *C. difficile* phage with 100% query coverage and 94% identity (Table 4-5).

Table 4-5: BLAST comparison of gene *Ga0114281_1134*

Description	Query cover	E value	Identity	Accession
***C. difficile* ATCC 9689 = DSM 1296, complete genome**	100%	0.0	100%	CP011968.1
***C. difficile* ATCC 9689 = DSM 1296 plasmid, complete sequence**	100%	0.0	100%	CP011968.1
***C. difficile* BI9 chromosome**	100%	0.0	100%	FN668944.1
***C. difficile* BI1 plasmid pCDBI1, complete sequences**	100%	0.0	100%	FN668942.1
***C. difficile* phage phiCDHM19, complete genome**	100%	0.0	94%	JX145342.1

An InterPro search of the amino acid sequence for this gene concluded that the protein belongs to the phage holin four superfamily along with TcdE. At the sequence level, when comparing *Ga0114281_1134* to *C. difficile* genome sequences, there is an 86% identity of this sequence to *tcdE* of 630.

To assess if this second holin-like gene has an effect on toxin levels found in the supernatant we attempted to make a ClosTron insertional mutant within the gene. During the ClosTron design stages using the design tool (www.clostron.com) the highest algorithm score was 4.969 where a score of over seven is recommended. We decided to proceed further with the low scoring target sequence but even though we were able to successfully conjugate pMTL007-E2::Cdi-BI9_3942::CT(*ermB*) into VPI 10463 no lincomycin resistant colonies were recovered suggesting intron insertion was not achieved.

4.2.7 Restriction modification system

We hypothesised that the two putative type I restriction modification systems (RMS) were the mechanism preventing DNA transfer into VPI 10463. During the comparative genetic analysis we also identified that R20291 carried a putative type I RMS which showed 99% homology to the *hsdR* gene (*Ga0114281_112960* in VPI 10463 and *CDR20291_2909* in R20291) and 98% homology to the *hsdM* gene (*Ga0114281_112965* in VPI 10463 and *CDR20291_2912* in R20291) (Figure 4-10) but no homology was found for the *hsdS* gene (*Ga0114281_112964* in VPI 10463 and *CDR20291_2911* in R20291) (Figure 4-10), suggesting a different nucleotide target site for these type I RMSs. The second type I RMS in VPI 10463 does not show homology to the first type I RMS in VPI 10463 nor is there a corresponding RMS in the genome sequence for R20291 (Figure 4-11).

As PacBio sequencing technology is capable of identifying methylation patterns based on statistical analysis of the polymerase kinetics during sequencing[204,205] we were able to infer the recognition motifs for the two VPI 10463 type I restriction modification systems. This was achieved by submitting the motif summery to the Restriction Enzyme Database (REBASE)[206] for analysis, also confirming the presence of two type I RMS in VPI 10463. Using the same database we were able to retrieve the recognition motif for the type I RMS in R20291 from a previously published data set[154].

Using CLC genomics workbench (V8.5.1) the three recognition motifs were compared to the modular plasmid maps for pMTL8*151 (* represents 2, 3, 4 or 5[111]) to identify cleavage sites. In all plasmids there was one recognition site for the R20291 type I RMS located in the *catP* gene and one recognition site for the first VPI 10463 type I RMS which was also located in the *catP* gene (Figure 4-12). The second type I RMS in VPI 10463 has three recognition sites in all the plasmids located in *lacZα*, *catP* and *traJ* with an additional site in pMTL84151 in *orfB* (Figure 4-12).

Figure 4-10: Comparative analysis of two type I restriction modification systems identified in VPI 10463 and R20291. Top sequence is R20291 bottom sequence is VPI 10463. Red boxes indicate *hsdR* genes and blue boxes indicate both *hsdS* then *hsdM* gene. The *hsdR* genes show 99% homology and the *hsdM* genes show 98% homology. The *hsdS* genes are not homologous suggesting different recognition sites.

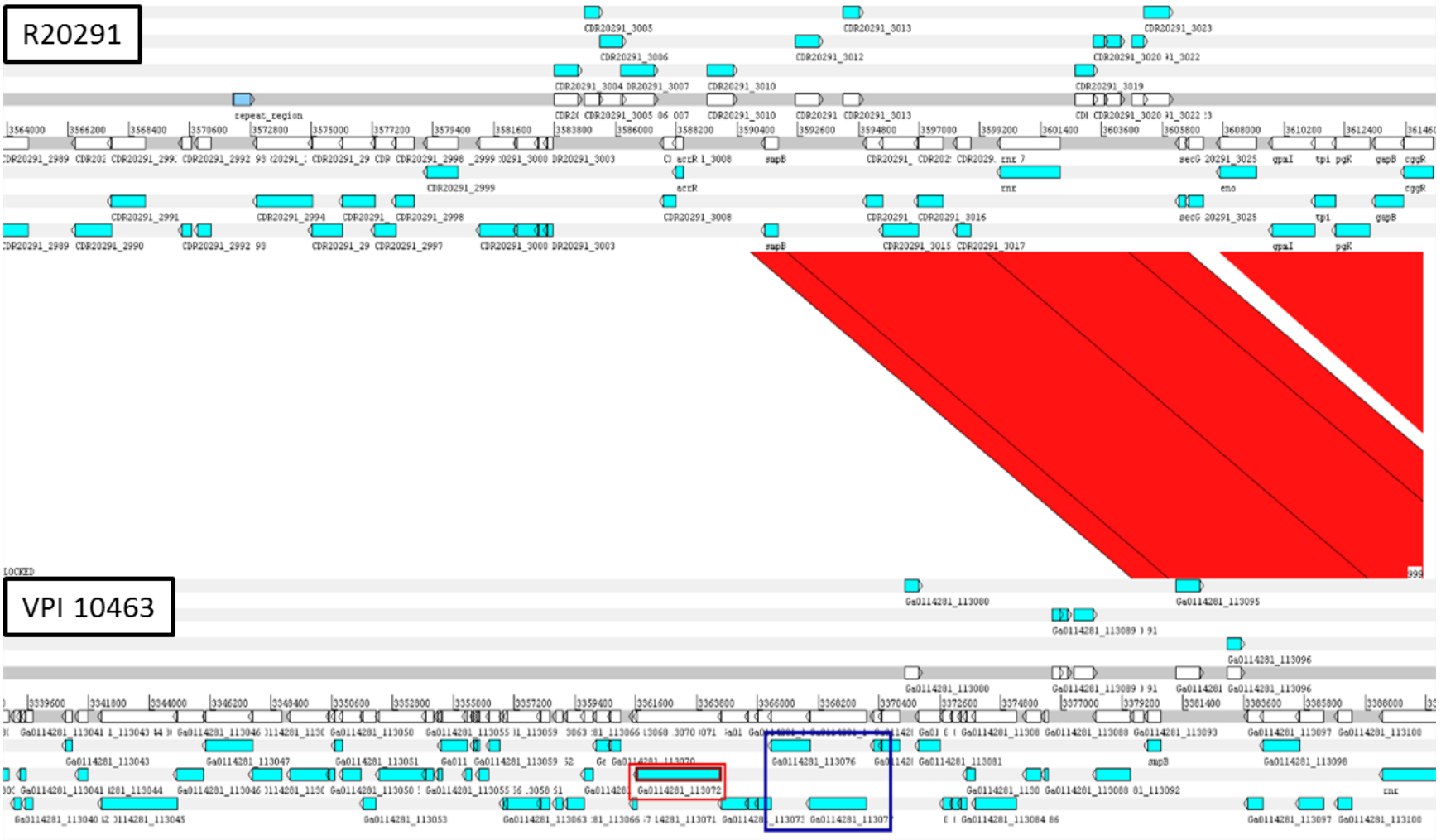


Figure 4-11: Comparative analysis of VPI 10463 and R20291. Top sequence R20291 bottom sequence VPI 10463. In the VPI 10463 sequence; red box indicates *hsdR* gene and blue box indicates both *hsdS* then *hsdM* gene. These boxes indicate the second type I restriction modification system which does not have a homologue in R20291.

To test the hypothesis, we created ClosTron mutants in the restriction modification subunit (*hsdR*) of the type I RMS in BW-R20291 and the *hsdR* is the first VPI 10463 type I RMS (from herein known as *hsdR1*) as the same ClosTron plasmid could be used in both strains (pMTL007-E2::Cdi-R-subunit-527s::CT(*ermB*)). ClosTron mutants were created as described in 2.6.1 over a period of eight months. Due to the restriction modification system, there were issues around the initial conjugation of the plasmid, over 50 individual attempts were made. Numerous reconfigurations of the protocol included altering the donor / recipient volumes, age of starter cultures, OD of starter cultures and the time in which the donor and recipient cultures were mixed before plating onto non-selective agar. The latter appeared to be the most effective variable to change, this was increased to 1 hour rather than direct plating. Once transconjugants were isolated they were restreacked on BHIS agar containing antibiotics appropriate to counter select the plasmid and glycerol stocks were preserved. From there integration events were selected for using lincomycin. There were many apparent mutants that showed the ability to grow on lincomycin but did not show the appropriate genotype. Eventually after screening 30 individual colonies for both BW-R20291 and VPI 10463, from 6 different attempted recombination events, three potential mutants were isolated. A schematic representation of the ClosTron insertion and the junction PCR products can be seen in Figure 4-13A. The insertions were confirmed by flanking PCR (Figure 4-13B) and junction PCR (Figure 4-13Figure 4-13C) using primers hsdR_Fs1, hsdR_Rs1 and EBS universal (Table 2-3). The junction PCR products were sent for Sanger Sequencing to further confirm the location of insertion. Single insertion of the retargeted group II introns was confirmed by Southern blot using a probe designed to target the RAM and gDNA digestion with HindIII (Figure 4-14), VPI 10463::*hsdR1*(*ermB*) clone 2 was shown to have two insertion sites after digestion.

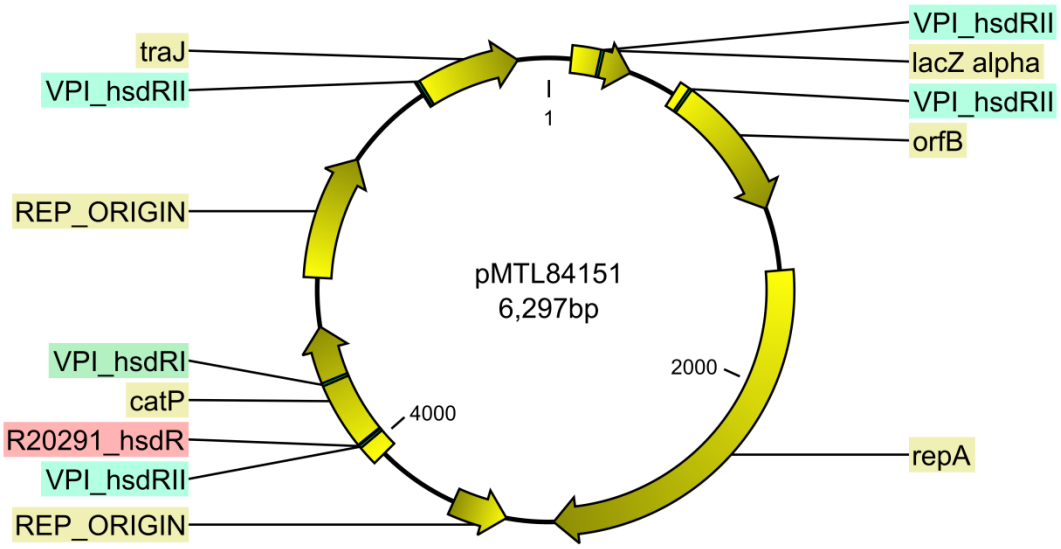


Figure 4-12: Schematic of pMTL84151 showing cleavage sites for each type I restriction modification system. Figure legend; yellow boxes indicated plasmid gene names; red boxes indicate cleavage site of the R20291 type I restriction modification system; green boxes indicate cleavage site of the first type I restriction modification system in VPI 10463; blue boxes indicate cleavage site of the second I restriction modification system in VPI 10463.

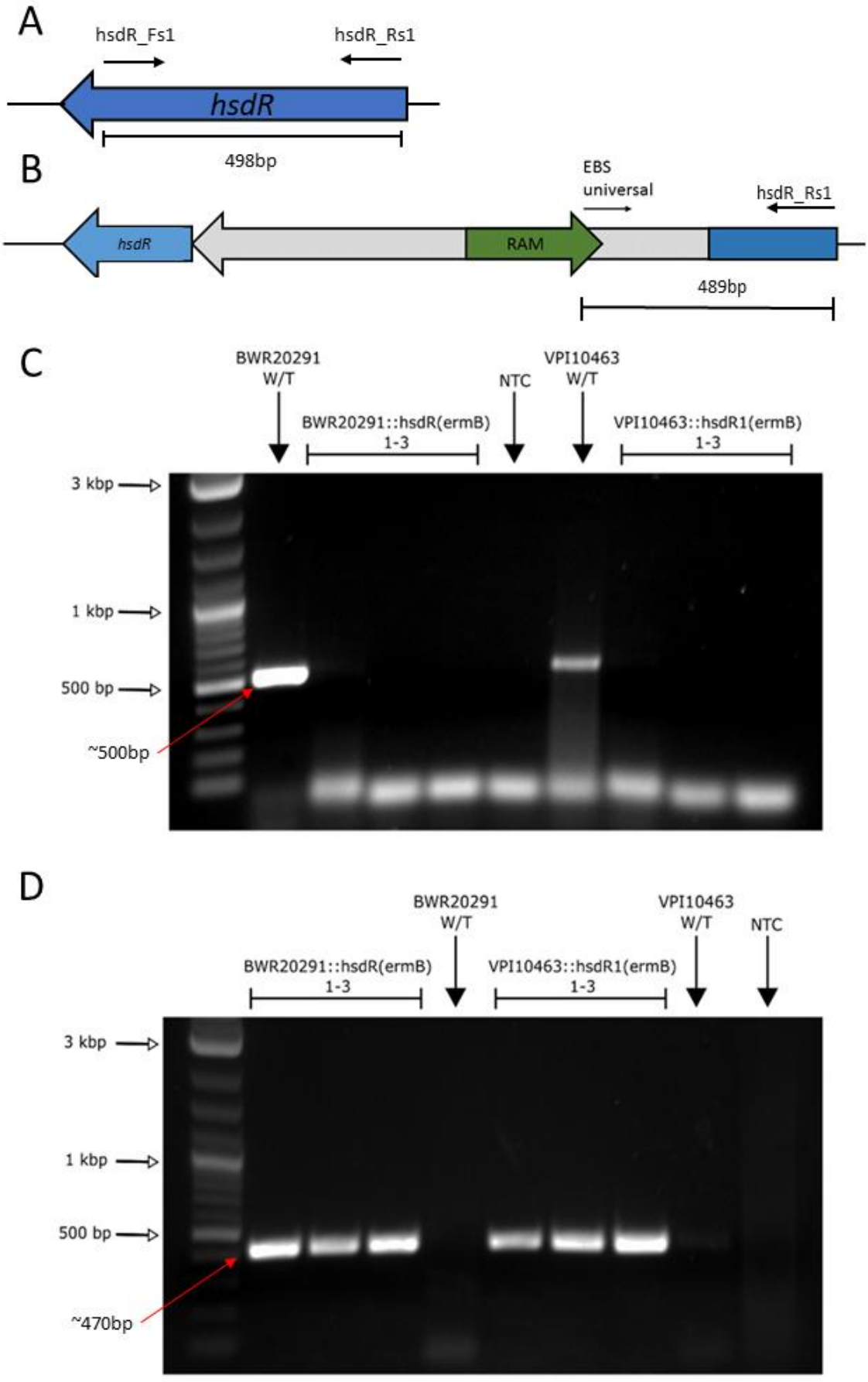

A
hsdR_Fs1
hsdR_Rs1
hsdR
498bp
B
EBS
universal
hsdR_Rs1
hsdR
RAM
489bp
C
BWR20291
W/T
NTC
VPI10463
W/T
BWR20291::hsdR(ermB)
1-3
VPI10463::hsdR1(ermB)
1-3
3 kbp
1 kbp
500 bp
~500bp
D
BWR20291
W/T
VPI10463
W/T
NTC
BWR20291::hsdR(ermB)
1-3
VPI10463::hsdR1(ermB)
1-3
3 kbp
1 kbp
500 bp
~470bp

Figure 4-13: PCR screening of lincomycin resistant colonies for the integration of the retargeted group II intron into the *hsdR* and *hsdR1* genes of BW-R20291 and VPI 10463 respectively. (A) A schematic representation of the *hsdR* in BW-R20291 and *hsdR1* in VPI 10463 showing primer binding sites for wildtype PCR and expected product sizes. (B) A schematic representation of the ClosTron insertion into *hsdR* in BW-R20291 and *hsdR1* in VPI 10463. showing primer binding sites for junction PCR and expected product sizes. (C) PCR screen using primer hsdR_Fs1 and hsdR_Rs1 which flank the intron insertion site ~500bp bands can only be seen in the BW-R20291 and VPI 10463 wildtype strains. (B) PCR screening of integrants using intron-exon junction primers hsdR_Rs1 and EBS universal, ~470bp products can only be seen in the integrants. NTC = No template control.

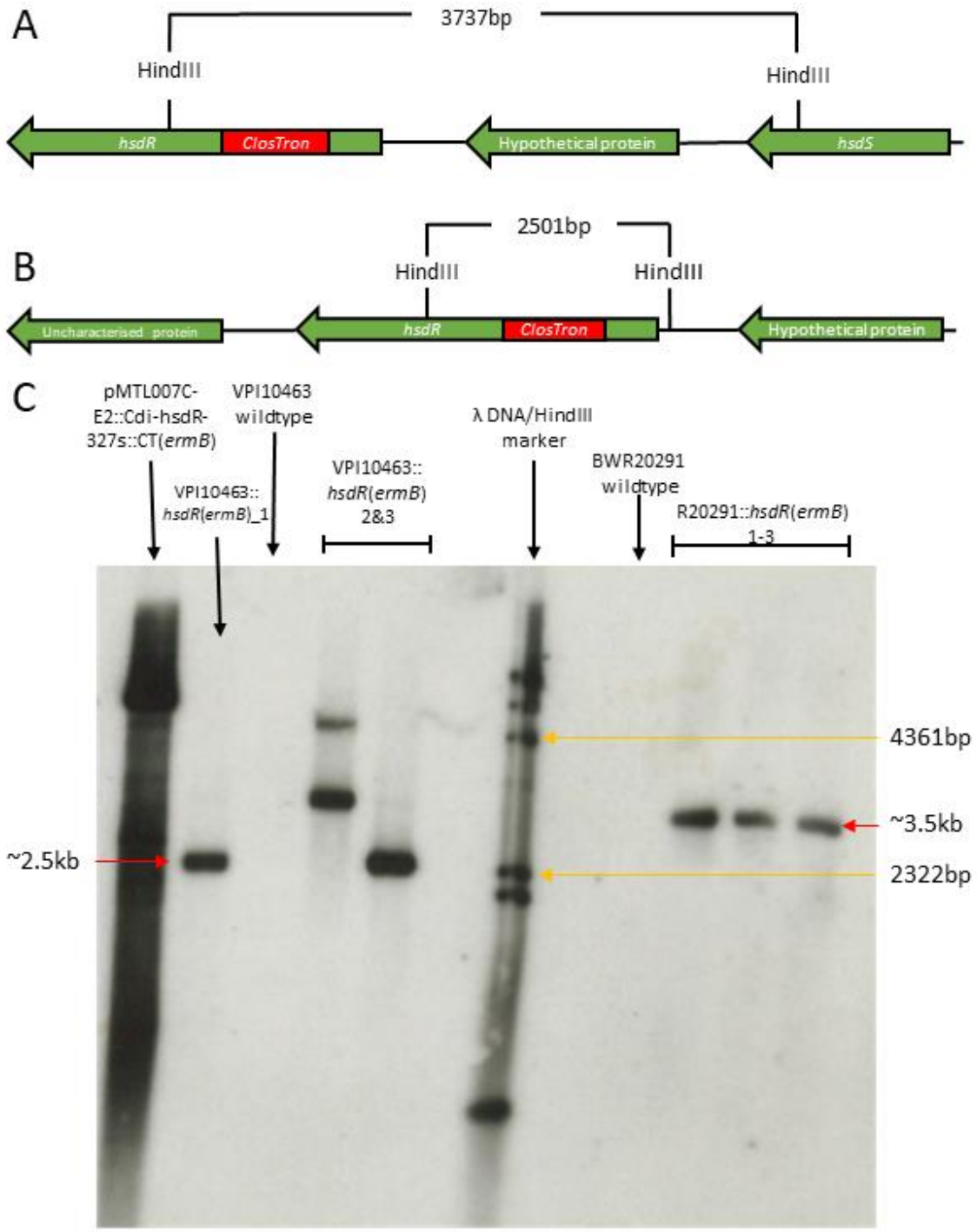


Figure 4-14: Schematic diagrams and Southern blot to confirm ClosTron insertion in the *hsdR* gene of BW-R20291 and the *hsdR1* gene of VPI 10463. Figure legend; Red box indicates insertion of the class II intron; green arrows indicate effected genes due to insertion. (A) Schematic representation of the class II intron at position 327 of *hsdR* in BW-R20291. Locations of the HindIII restriction sites result in a product of 3737bp. (B) Schematic representation of the class II intron at position 327 of *hsdR1* in VPI 10463. Locations of the HindIII restriction sites result in a product of 2501bp. (C) Southern blot analysis using intron probe generated using primers EBS2 and IntronSalR. Lambda/HindIII ladder used for the estimation of band sizes. Single bands identified for BW-R20291_*hsdR*(*ermB*) (~3.5kb) and single bands identified for VPI 10463_*hsdR*(*ermB*) clones 1 and 3 (~2.5kb). Double insertion identified in VPI 10463_*hsdR*(*ermB*) clone 2, this was excluded from further analysis. Plasmid

pMTL007C-E2::Cdi-327-::CT(*ermB*) at ~6000bp as the positive control. BW-R20291 and VPI 10463 wildtype strains used as negative controls. Products visualised on a 0.8% gel.

After confirmation of single integration, conjugation frequencies were calculated as described in 2.5.3. *C. difficile* strain 630 was used as a control as there is no type I RMS described in this strain. Each of the modular plasmids was conjugated into each strain in three separate experiments to ensure results were not due to an artefact in the media. Conjugation frequency is defined as number of transconjugants per recipient cell.

When compared to 630 using the modular plasmids[111] analysis showed in both VPI 10463 and BW-R20291 wild type strains that conjugation frequency was reduced by 100 – 1000 fold or that conjugation occurred at a level below the detection limit of the assay. Conjugation frequencies in the VPI 10463::*hsdR1*(*ermB*) mutant were more efficient than what is established in 630 by ~10-fold for all of the modular plasmids (Table 4-6). The conjugation frequencies of VPI 10463::*hsdR1*(*ermB*) compared to the parental strain were statistically significant for plasmids pMTL82151 (*P*= <0.0001) and pMTL84151 (*P*= <0.0005) by paired T-test. Transconjugants were detectable in VPI 10463::*hsdR1*(*ermB*) pMTL83151 and pMTL85151.

Conjugation frequencies in the BW-R20291::*hsdR1*(*ermB*) mutant were comparable to what is observed in 630 (Table 4-7). The conjugation frequencies of BW-R20291::*hsdR1*(*ermB*) compared to the parental were statistically significant for plasmids pMTL82151 (*P*= <0.0001) and pMTL84151 (*P*= <0.0001). Transconjugants were detectable in pMTL83151 but not in pMTL85151. These data suggest that an insertional inactivation of the restriction subunit (*hsdR)* genes in the type I RMS improved conjugation frequencies into these strains.

Table 4-6: Mean conjugation frequencies of *C. difficile* strains 630, VPI 10463 and the ClosTron mutant VPI 10463::*hsdR1*(*ermB*) (*n*=3). *P* value calculated by paired T-test.

	630	VPI 10462	VPI 10463::*hsdR1* (*ermB*)	*P*
pMTL82151	$6.18x10^{-6}$	$4.84x10^{-8}$	$7.05x10^{-5}$	<0.0001
pMTL83151	$8.93x10^{-7}$	-	$4.04x10^{-6}$	
pMTL84151	$4.36x10^{-6}$	$9.67x10^{-9}$	$4.79x10^{-5}$	<0.0005
pMTL85151	$2.98x10^{-8}$	-	$4.73x10^{-8}$	

Table 4-7 Mean Conjugation frequencies of *C. difficile* strains 630, BW-R20291 and the ClosTron mutant BW-R20291::*hsdR*(*ermB*) (*n*=3). *P* value calculated by paired T-test.

	630	BWR20291	BWR20291::*hsdR* (*ermB*)	*P*
pMTL82151	$6.18x10^{-6}$	$6.79x10^{-9}$	$1.32x10^{-6}$	<0.0001
pMTL83151	$8.93x10^{-7}$	$4.94x10^{-9}$	$7.88x10^{-8}$	
pMTL84151	$4.36x10^{-6}$	$1.42x10^{-8}$	$1.58x10^{-6}$	<0.0001
pMTL85151	$2.98x10^{-8}$	-	-	

4.3 Discussion

It has been demonstrated in numerous studies that drawing conclusions on any metabolic regulation from just one strain can lead to inaccurate inferences. This is likely due to the wide genomic diversity that is displayed by *C. difficile*[166]. The aim of these analyses was to provide an accurate genome sequence for a high toxin producing strain VPI 10463, to look for unique genes that may contribute to this phenotype and to improve the ability of DNA transfer into this strain.

Firstly, we looked at the toxin phenotypes of four different *C. difficile* strains, 630, BW-R20291, M120 and VPI 10463. These strains were chosen due to their specific characteristics. Derivatives of strain 630 are the most commonly studied due to the relative ease of DNA transfer into the strain[111]. Strain R20291 is a clinically relevant strain and a so-called hyper-virulent strain of the 027 PCR-ribotype[207]. Strain M120 (PCR-ribotype

078) is a non-motile low toxin producing strain which was chosen as a control[154]. Although there have been studies showing the relative toxin production of these strains they have all been at specific time points and not over a set time course. This initial phenotypic study was to confirm the increased toxin production previously shown in VPI 10463[78]. Growth curve analysis showed that over a 48-hour period in TY broth each of the four strains grew in a similar temporal pattern. Toxin supernatants taken at 48 hours showed that M120 produced 10-fold less toxin than 630 and BW-R20291 and 100-fold less toxin than VPI 10463 which had produced approximately 10,000 ng/mL. Despite the dissimilar quantities of toxin being produced by each strain, they all followed the same temporal pattern.

To investigate possible genetic factors that may give rise to this increased toxin phenotype we employed whole genome sequencing of VPI 10463. This was performed using two different technologies to compliment the weaknesses from using only one. The genomic DNA of VPI 10463 was sequenced using PacBio RSII and Illumina MiSeq. The Illumina short reads were used to correct mistakes made in the long reads produced by the PacBio RSII using a hybrid approach as described by Hackl *et al*[116]. Annotation and comparison of the genome to 630 and R20291 revealed a number of interesting differences.

Using progressiveMauve it could be seen that there was a large scale genomic rearrangement present in VPI 10463 when compared to 630 and R20291. Closer investigation of this inversion showed that when the reads were mapped back to the genome the coverage was sufficient to suggest that this was not a mistake in the *de novo* assembly process. The reorganisation appears to occur within two lineage specific regions which are almost equidistant from *dnaA* and the predicted *oric*[208]. These regions contain two ~21kb inverted repeats containing highly conserved head and tail phage genes[209]. A BLAST search of this sequence identified a high similarity to a number of *C. difficile* phage's. This may suggest two separate phage integration events as although the other genes within the

lineage specific regions are phage related they do not show homology to each other. This type of large scale recombination has not been described in *C. difficile* previously but it has been in other bacterial species including other low G+C members of the *Firmicutes* [210–213]. In fact, it is now believed these large scale recombination events are required for bacterial evolution, and that genomes have a high plasticity[214–216]. Although the mechanism behind this genome rearrangement has not been investigated in this study there has been evidence to show that large scale recombination occurs in IS elements, rRNA operons and prophage regions[210,212,217,218]. One study showed spontaneous recombination after repeated subculture between prophage regions in *E. coli* O157:H7 strain EDL993[217]. Many of these studies suggest either a RecA recombination event has occurred or that of a phage mediated integrase, both of which are found in VPI 10463, and that these events are likely to occur during chromosomal replication[210,211,213,217].

Although an interesting feature the inversion does not appear to explain why VPI 10463 is able to produce more toxin than other strains. Flagella are known to be important for motility, colonisation, to enhance growth, for survival and as discussed in 3.1.2 flagellar proteins have been linked to toxin regulation. The flagella operons for *C. difficile* are known to comprise of three operons; F1 late stage flagellar genes; F2, flagellar glycosylation genes; F3, early stage flagellar genes[161]. Early microarray analysis showed that there was high diversity in these three operons when 630 was compared to other *C. difficile* strains[215,219]. With regards to the strains used in this analysis, previous microarray data has shown that the F2 region in 630 consists of four genes which in R20291 and M120 have been replaced by six different genes[220]. The F3 operon is completely missing in the M120 strain rendering it non-motile[219]. However, the F3 operons between 630 and R20291 show high levels of sequence identity and the F1 operons show high levels of sequence identity between all three strains.

To assess if the VPI 10463 flagella operons were divergent from either 630 or R20291 we performed an ACT comparison of the three strains. From this it was shown that the F1 and F3 operons in VPI 10463 had high levels of sequence identity to both 630 and R20291. VPI 10463 also had high levels of sequence identity to the F2 operon in R20291. The only major difference between all three strains was a nonsense mutation that was introduced by a deletion of an adenine at position 108 in *fliJ* resulting in a 36-amino-acid truncated protein. FliJ is a 147-amino-acid protein that has been shown to be a general chaperone for the delivery of rod/hook flagella proteins to the export gate[221]. In *Salmonella* spontaneous mutants in *fliJ* have been shown to result in "leaky" motility phenotypes suggesting that it is required for efficient flagella substrate export but flagella synthesis can occur in its absence[222]. To test this hypothesis, we conducted motility assays on all three strains using M120 as a non-motile control. Motility assays showed that VPI 10463 had a reduced motility phenotype when compared to 630 and R20291 but was not completely inhibited as was seen in the Salmonella *fliJ* mutants[222]. Due to this finding transmission electron microscopy was performed to assess the flagella structures. TEM revealed that VPI 10463 produced a single flagellum (monotrichous) whereas 630 and BW-R20291 produced multiple flagella (peritrichous), this also supports the findings seen in the Salmonella f*liJ* mutants.

The current understanding is that mutations in the late stage flagella operon F1 increase toxin production and the inverse is true for the early stage flagella operon F3[161,162]. If this holds true for all genes in the F3 operon, then it would be expected that a mutation in *fliJ* would decrease toxin production. We propose that the loss of flagellin on the cell surface as seen in VPI 10463 results, at least in part, in the increased toxin produced by this strain as this would fit into previous observations[161–165].

Whether this influences toxin regulation has not been experimentally examined in this study however this would be suitable for future work by creating mutations in *fliJ* in additional strains of *C. difficile* and measuring resulting toxin levels.

A holins primary functions are to form pores in cytoplasmic membranes of bacterial cells for the release of phage particles after replication within the host genome. Although it is now becoming more apparent that holins may have a wide variety of functions in phage free prokaryotic cells[223]. These include bacterial gene transfer and biofilm formation[224,225]. As discussed in 1.2.2, it is believed that the holin like protein TcdE is responsible for the transportation of TcdA and TcdB out of the bacterial cell. Analysis of the unique genes found in VPI 10463 identified a second holin-like gene *Ga0114281_1134* that shared an 86% sequence identity to *tcdE* of 630. A BLAST search revealed that this gene was also present in *C. difficile* strain DSM1296 chromosome and plasmid, the BI9 chromosome and the BI1 plasmid. It was also present in the ΦCDMM19 genome. We hypothesised that this extra holin like gene was responsible for the high levels of toxin present in culture supernatant by allowing TcdA and TcdB to pass through the cell wall more freely as similar findings have been observed before in C. difficile strains containing holin-like genes[170]. To test this hypothesis, we attempted to construct an insertional mutation of *Ga0114281_1134* using ClosTron technology, however this was unsuccessful due to not being able to identify a suitable integration site for the group II intron. If improved DNA transfer into VPI 10463 is achieved it would be possible to make a clean deletion of *Ga0114281_1134* to assess its function, if any, in toxin release. It would also be interesting to compare the TcdA and TcdB supernatant levels of the other strains in which *Ga0114281_1134* homologues have been identified.

During the analysis two type I RMS were identified in VPI 10463 and one in R20291. They were hypothesised to be the cause of low conjugation frequencies into these strains when

compared to 630 which does not have any type I RMSs. Using CLC genomics workbench the predicted cleavage points were identified on our most commonly used plasmids, the modular plasmid series[111]. This identified different recognitions sites in *catP* for the type I RMS in R20291 and the first type I RMS in VPI 10463. There were also three recognition sites for the second type I RMS in VPI 10463 located in *lacZα*, *catP* and *traJ* with an additional site in pMTL84151 in *orfB*. We were able to functionally inactivate the *hsdR* in BW-R20291 and *hsdR1* in VPI 10463 using ClosTron technology. This would enable the methylation of DNA to still occur but inhibit the restriction of unmethylated DNA at the recognition sites. Using this method, we were able to show that DNA transfer was improved to levels similar to those found in a strain without a type I RMS namely 630 when analysing conjugation frequencies.

This work further validates that restriction modification systems are a barrier for DNA transfer into *C. difficile*[112,226] which has also been described in other bacterial species[227–229]. There have been alternative methods for evading these systems described in the literature. One such method is termed plasmid artificial methylation (PAM) in which plasmid DNA is pre-methylated in the *E. coli* host. This is achieved by cloning the specificity and modification subunits of the RMS either into the *E. coli* genome or onto a plasmid where they are heterologously expressed, for a number of different bacterial species an improvement of DNA transfer has been observed[227,230,231]. The very nature of this method gives it a significant limitation as well, for each strain that would be studied a new *E. coli* donor would have to be created and to achieve this the RMS would have to be described for the recipient strain beforehand. Although as PacBio SMRT sequencing becomes more widely used this will become easier.

A less complex method which was only recently described in *C. difficile* is the use of heat-shock of the recipient cells before conjugation and/or transformation[232–234]. The premise is

that the heat-shock temporarily inactivates the host defence systems, including any RMS, and allows DNA transfer into the cell. Although, efficiency is variable and maybe dependent on the type of RMS present within the recipient genome[232] and modified protocols maybe required for different strains of the same species[234].

The inception of CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 (CRISPR-associated proteins) genome editing offers a new method of evading RMS by functionally inactivating the *hsdR*. CRISPR are part of the bacterial immune system, short palindromic repeats are interspersed with spacer sequences which are homologous to bacteriophage DNA[235]. These sequences are flanked by Cas associated genes and together they are able to defeat invading bacteriophages[236]. When the CRISPR and spacer sequences are transcribed this produces crRNA which acts as a template for the Cas proteins and tracrRNA to find and destroy foreign DNA. Recent advances in the technology now make it possible to introduce a synthetic CRISPR-Cas system into a host cell on a plasmid which contains a synthetic crRNA and tracrRNA termed gRNA (guide RNA)[236]. This system then introduces a double strand break which is repaired by endogenous non-homologous end joining (NHEJ) or homology-directed repair pathways[235]. There have only been a small number of publications citing the use of this technology in prokaryotes, possibly due to the inability of bacteria to efficiently repair these breaks caused by the CRISPR-Cas9 system[235]. However, this method has recently been shown to work in *Clostridium pasteurianum* using the CRISPR-Cas9 but also by co-opting native Cas proteins[237]. The rational for using the native system was due to low transformation frequencies of the Cas9 plasmids into *Clostridium*, however, this might be overcome by using a heat-shock method as described earlier. Once the mutation has been created in *hsdR* there should be no need for the heat-shock protocol.

4.4 Key outcomes

- A high quality genome map for VPI 10463.
- A large scale genome rearrangement was identified in VPI 10463, the first described in *C. difficile*.
- Identification of two putative genes that potentially have a role in the increased levels of toxin produced by VPI 10463; *Ga0114281_1134* and a nonsense mutation in *fliJ*.
- VPI 10463 has low motility and is monotrichous.
- Circumvention of the type I RMS in R20291 and one of the type I RMS in VPI 10463 increases genetic transfer from an *E. coli* donor.

4.5 Future work

- A more in-depth comparative analysis of the VPI 10463 genome compared to other clinically relevant strains.
- Creation and characterisation of FliJ mutants in *C. difficile* strains to assess its role, if any, in the regulation of toxin production.
- Restoration of FliJ in VPI 10463 to measure a reduction, if any, of toxin production.
- Create a clean deletion of the holin-like gene *Ga0114281_1134* in VPI 10463 and assess if toxin supernatant levels are reduced.
- Explore different methods for creating mutations in the *hsdR* genes to improve DNA transfer.

Chapter Five Analysis of co-infection, antibiotic resistance and isolate evolution in recurrence of *Clostridium difficile* infection.

5.1 Introduction

5.1.1 Clinical diagnosis of Clostridium difficile

Historically, the diagnosis of *Clostridium difficile* infection (CDI) were performed using a Cytotoxin Neutralisation (CTN) assay which detects TcdA and TcdB in the supernatants of patients' faeces and was classed as the gold standard until recently[238]. Although CTN is a sensitive assay it is time consuming, taking on average 24 – 48 h to complete. When an enzyme linked immunosorbent assay (ELISA) method was developed to detect (initially only) TcdA in faecal samples many clinical laboratories discontinued the use of CTN[239]. In recent years the method of diagnosis has changed due to the low sensitivity of using just a TcdA/B ELISA alone and currently a twostep assay is performed[240]. Firstly, either an ELISA for glutamate dehydrogenase (GDH) or a nucleic acid amplification test (NAAT) for *tcdB*, to screen samples for the presence of *C. difficile*[240]. If this test is positive, then an ELISA for TcdA/B is performed to confirm clinical infection. If the GDH/NAAT are negative, symptoms are caused by another aetiology. If the GDH/NAAT are positive and the EIA is negative, then toxigenic *C. difficile* may be present and the patient is a potential excretor[240].

5.1.2 Epidemiology and hypervirulence of Clostridium difficile

Hypervirulent strains of *C. difficile* have been implicated in more severe disease states and have recently been reported to have become epidemic in many countries[241]. The most renowned offender is classified as RT027, PFGE type NAP1 and REA type BI (027/B1/NAP1) and belongs to toxinotype III[242]. It has been suggested that this strain gets its hypervirulence, resulting in a more severe disease, not only from producing an additional toxin; binary toxin but also having a variant *tcdC* repressor gene, producing higher levels of toxin A and toxin B and an increased sporulation rate[243]. Although these findings are becoming more controversial[244–247]. In the UK, during 1990-2003, cases of 027/NAP1/B1

isolates were relatively infrequent with RT001 and RT106 being the most prominent[248]. During 2005-2007 there was a change in the ribotype distribution with RT001, RT027 and RT106 all sharing approximately 25% each of the total incidences[249]. In 2007/08 RT027 became the most common at 55.3% of all isolates, but since 2008/09 the prevalence has steadily been falling although in 2010/11 it was still the most frequently isolated at 12.4%[42]. Other ribotypes recently associated with hypervirulence are RT078, RT056 and RT018[250,251].

In the *Clostridium difficile* ribotype network (CDRN) report for 2013 – 15 the shift in ribotype distribution discussed above can be seen over the six years the service has been mandatory (Figure 5-1)[252].

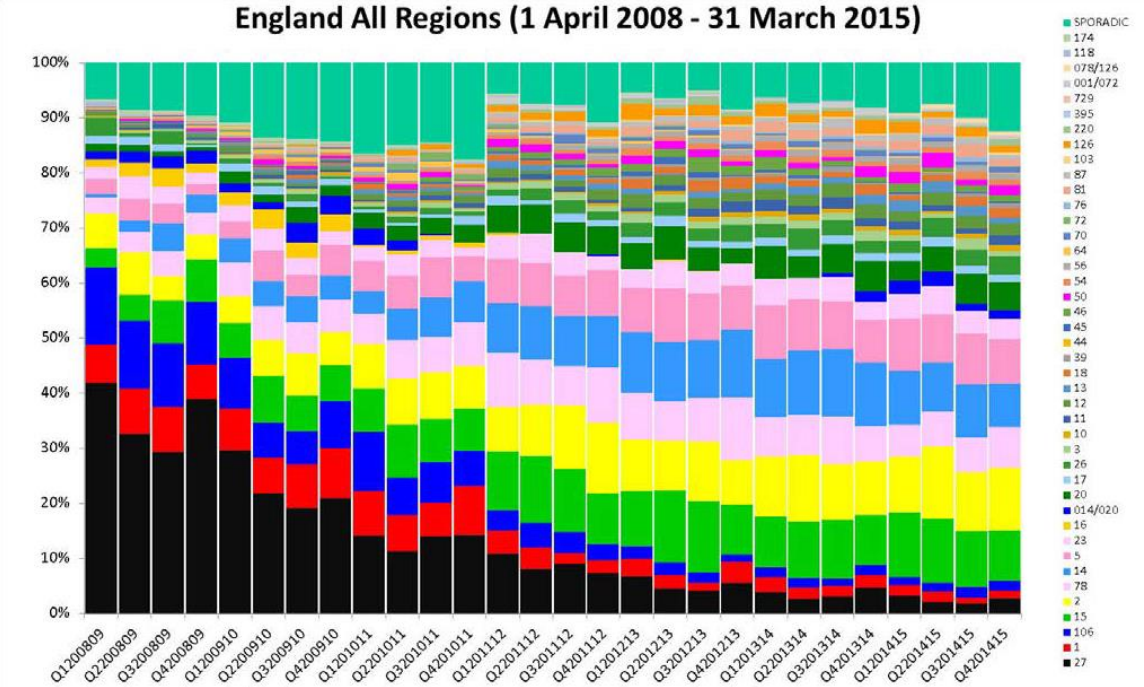


Figure 5-1: Ribotype distribution over a six-year period (2009 – 15) for all regions of England. Different coloured blocks indicate a different ribotype which has a prevalence of over 2%, other ribotypes are classed as sporadic. Image taken from[252].

Although the incidence of RT027 has decreased new "emergent" ribotypes are increasing including RT078, RT002, RT005, RT014/020, RT023 and RT015[252]. RT078 is prevalent in the Netherlands and has caused outbreaks in Northern Ireland[251,253].

In the East Midlands the incidence of RT027 has decreased although there was a slight rise in the last quarter of 2014 – 15 (Figure 5-2). The most isolated ribotypes within the East Midlands, from the latest published figures, are RT027, RT015, RT002, RT078, RT014, RT005, RT023, RT020, RT017 and RT026 (Figure 5-2).

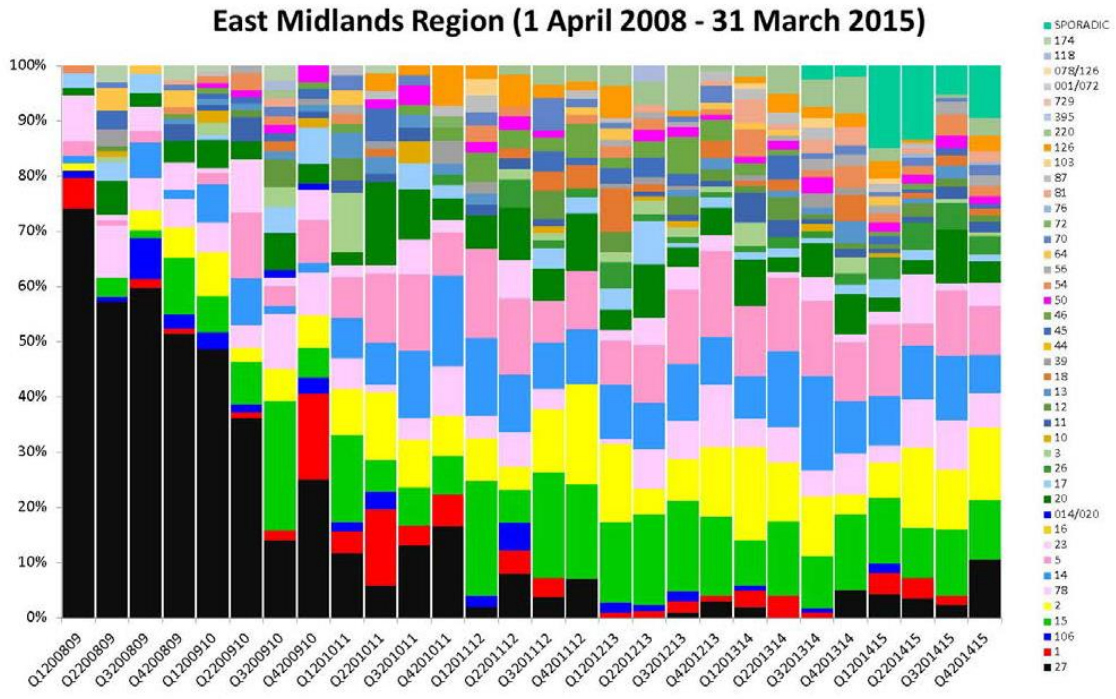


Figure 5-2: Ribotype distribution over a six-year period (2009 – 15) for the East Midlands region. Different coloured blocks indicate a different ribotype which has a prevalence of over 2%, other ribotypes are classed as sporadic. Image taken from[252]

5.1.3 Hospital acquired vs. community acquired Clostridium difficile infection

Traditionally CDI was seen as a nosocomial infection but over the past decade there has been an increase of reports of CDI from the community[254]. This patient demographic have not usually had any exposure to antibiotics, no history of recent hospital stays and they are usually younger and healthier[255]. Community acquired (CA-CDI) generally refers to patients with CDI who show symptoms within the first 48 h of admission to hospital or after four weeks following discharge[256]. Hospital acquired CDI (HA-CDI) is usually defined as presentation of symptoms after 48 h of admission and within four weeks of discharge[256]. However, this narrow classification does not take into account patients who

have had out-patient exposure or residents of long term care facilities[257]. The differences of definitions for these patient cohorts and a lower incidence of testing for CDI in the community[258] may not fully reflect the true number of CA-CDA cases. Initially, there appeared to be a small range of epidemic ribotypes that were associated with CA-CDI[16,259] but a more recent study has shown similar ribotype distributions between both CA- and HA-CDI[258].

5.1.4 Recurrence of Clostridium difficile infection

Recurrent CDI is defined as the return of symptoms after successful treatment, which occurs in 19 – 35% of patients with CDI, these figures increase after every subsequent episode[28–30]. Recurrent infection can be classed as either relapse or re-infection where infection is caused by the same strain as the initial infection or where a different strain is the causative agent of infection respectively[260]. Within the clinical setting, if a patient has had a second independent episode within 8 weeks of successful treatment it would be classed as a recurrence episode[261]. It is currently estimated that relapse and re-infection are approximately equal in incidence[262] but these figures may change as whole genome sequencing becomes more commonplace.

Phenotypic characteristics i.e. toxin production, sporulation and germination were shown not to be a factor in recurrent infection in one study[263] but another has found that increase in germination efficiency was seen more in strains from recurrence cases than non-recurrence cases[264]. However, both studies only used small sample sizes therefore more investigation is required before a definite conclusion can be made.

Host risk factors include advancing age, a high Horn index score which clinically assesses the severity of underling illness/co-morbidities[29] and low anti-toxin antibody responses to *C. difficile* colonisation and initial symptomatic infection increase risk of infection[265]. C-

reactive protein (CRP) is a known marker of inflammation with reference ranges of 10 – 50 mg/L as an indication of viral infection and >100 mg/L for bacterial infection[266]. CRP is a general measure of inflammation and an increase in levels could be caused by diseases such as coronary artery disease, myocardial infarction or cancer[266] and so is generally used as a guide of infection in conjunction with other pathology assays. Although, Eyre *et al*. described CRP to be an indication of first recurrence if a level of ≥85 mg/L was observed when used in their scoring matrix[267]. Other risk factors involve use of antibiotics and / or proton pump inhibitors after treatment of CDI[260] and the use of fluoroquinolones[268]. The nature of common typing practices i.e. ribotyping and MLST, are not discriminatory enough and may lead to miss-classification of relapse and reinfection[269].

5.1.5 Co-infection of Clostridium difficile strains

Currently there are only a few studies looking at the effect of co-infection with two or more strains of *C. difficile* and their effect on relapse and / or epidemiology. It currently stands that 7.7% – 13% of CDI cases are co-infected with multiple strains[270–272], but there is no definitive answer for this question. The isolation of single colonies for epidemiology can result in the under representation of transmission events, although this method is the most cost effective, relapse / recurrent infections maybe miss-classified[270,273].

5.1.6 Emerging resistance to Metronidazole and Vancomycin

Accounts of resistance to first line treatments of *C. difficile* have been increasingly reported over the past twenty years[126,274–276]. With routine susceptibility testing not performed on clinical isolates the exact level of resistance within the community and hospital reservoirs remains unknown. A recent study, where 953 isolates from 22 European countries were assessed for their resistance to Vancomycin and Metronidazole, showed that 96.84% and 97.82% of isolates were sensitive respectively. In the latest CDRN report the analysis of

200 isolates showed four of these displayed increased MICs to metronidazole (n= 3 RT027, n= 1 RT106) and one showed increased MIC to vancomycin (RT001)[252].

5.1.7 Isolation of Clostridium difficile from stool samples

The isolation of *C. difficile* from stool samples is known to be difficult, with historical recovery rates of 87.5% recorded in 2008 – 09 from the CDRN[252]. This issue most likely arises due to low numbers of spores / vegetative cells following antibiotic treatment of the patient or small sample availability. A number of studies focusing on the isolation of *C. difficile* from stool samples[277–280] or the environment[281] have focused on using expensive commercial agar that may not be affordable in the research and potentially the clinical setting. These studies focused only on recovery of *C. difficile* from clinical samples and not the semi-quantification of a known concentration of viable spores which may also be of importance in the research setting.

5.1.8 Nottingham University Hospitals NHS Trust

Nottingham University Hospitals NHS Trust (NUH) is a large multi-centre acute teaching trust which provides services for 3 – 4 million people from across the region[282]. In this study, we identified the first 100 patients without an episode of *C. difficile* within the previous six months from the period 1st November 2013 until 30th November 2014 from both the hospital and community settings. These patients were followed up for at least three months to monitor recurrence episodes. Clinical information was gathered at the time of infection to assess co-factors that may indicate recurrent infection including age, sex, site of acquisition, initial CDI treatment, use of PPI and other antibiotics at the time of infection and CRP. Where possible multiple colonies of *C. difficile* were isolated to evaluate co-infection rates.

5.1.9 Experimental aims

- Develop a cost effective and reliable method of isolating *C. difficile* from stool samples.
- Measure the level of co-infection and recurrent CDI within the NUH NHS region.
- Evaluate clinical information and perform MIC levels on all isolates to identify a role in recurrence episodes.
- Use whole genome sequencing on isolates from recurrence episodes and their index samples to assess recurrence / relapse and the microevolution of strains.

5.2 Results

5.2.1 Development of a Clostridium difficile culture method

Initially we aimed to develop a method of culturing *C. difficile* from stool samples where low numbers of spores / vegetative cells may be present. The first step was to identify which of two broths (CCFB and CCMB-TAL) performed better as an enrichment step. Hamster faeces which contained either *C. difficile* 630 (samples 1069 and 1076) or NM-R20291 (samples 1547 and 1557) were used at this stage. 0.5 g of faeces was added to 1 mL PBS, homogenised and heat shocked for 10 min at 80°C. After which samples were centrifuged (4000 *x*g for 1 min) and 100 µL supernatant was added to 5 mL CCFB and CCMB-TAL and incubated for 24 – 96 h at 37°C in anaerobic conditions. Both broth recipes contained 1% Neutral Red as a pH indicator. Growth was defined by turbidity and a pH change altering the colour of the broth from red to yellow. After a positive indication 50 µL broth were subcultured onto CCFA using a four-quadrant streak method then further incubated for 24 – 48 h. Recovery of *C. difficile* was enumerated semi-quantitatively (0 = no growth. 1 = growth in the 1st quadrant, 2 = growth in the 2nd quadrant, 3 = growth in the 3rd quadrant, 4 = growth in the 4th quadrant). Each sample was examined in triplicate. Recovery of *C. difficile* from two different broth compositions is represented in Table 5-1.

Table 5-1: Recovery of *C. difficile* from two broth compositions.

	CCFB		**CCMB-TAL**	
Sample	Time (h)	Recovery[a]	Time (h)	Recovery[a]
1069	36	1	48[b]	3
1076	36	1	48[b]	3
1547	24	1/3[c]	36	1
1557	24	1	24/36[d]	1/3[e]

a Recovery semi-quantified as 0 = no growth. 1 = 1st quadrant, 2 = 2nd quadrant, 3 = 3rd quadrant, 4 = 4th quadrant.

b Growth was only detected in one of the three broths tested.

c Recovery was semi-quantified as one in two broths and three in one broth.

d Positive broth recorded as 1 at 24h and two at 36h.

e Recovery was semi-quantified as three in two broths and one in one broth.

When trialling the different broth compositions, growth was detected in all CCFB samples. For samples 1069 and 1076 in CCMB-TAL growth was only detected in one of the triplicate broths for each (Table 5-1). In the samples where growth was detected in CCMB-TAL it was always 24 h after it had been detected in the CCFB (Table 5-1). The semi-quantification of positive broths revealed that CCMB-TAL gave a superior recovery rate over CCFB (Table 5-1). This observation appears to confirm previous studies' findings that the presence of lysozyme increases the recovery rate of *C. difficile*[283]. To assess this, we added lysozyme to CCFB and repeated the experiment, however, this addition did not improve recovery (data not shown). For this reason, CCFB was chosen for the enrichment step.

For the assessment of the solid medium, serial dilutions of 630Δ*erm* spore stocks were prepared in PBS homogenised non-infected mouse faeces (0.5 g in 5 mL) to give final concentrations of 10^1 to 10^4 spores / mL. Before use prepared samples were heat shocked and centrifuged as previously described. For each dilution samples were either directly plated onto solid media (CCFA, CCEY, ChromID C. difficile and TSA with 5% sheep's blood) or first subjected to a broth enrichment step using CCFB. An aliquot of 100 µL of heat shocked supernatant was added to CCFB and incubated for up to 120 h. When a positive broth was indicated or 120 h had passed 50 µL broth was inoculated onto all four solid agar or CCFB (to ensure a negative result) respectively using a four quadrant streak method. 50 µL heat shocked samples were also directly plated on to CCFA, CCEY, ChromID C. difficile and TSA with 5% sheep's blood using a four-quadrant streak method. Solid agar was allowed to incubate for up to 72 h in both cases. As before recovery of *C. difficile* was enumerated semi-quantitatively and examined in triplicate.

The enrichment in CCFB resulted in no positive indication for samples containing 10^1 to 10^3 spores and after plating onto CCFA no growth was observed after 72 h. Samples that

contained 10^4 spores gave a positive reaction in CCFB and were subsequently plated onto all four solid media. Growth was detected up to and including the 4th quadrant on CCEY and ChromID C. difficile after 24 h, on CCFA and TSA (5% sheep blood) after 48 h.

TSA with 5% sheep blood performed the poorest of the four solid media, when direct plating was employed, with no spores being recovered after 72 h incubation (Table 5-2). The most sensitive of the media trialled was CCEY agar supplemented with egg yolk, with 10^2 spores being detected in one of the replicates in the first quadrant after 72 h incubation (Table 5-2). In samples containing 10^4 spores, colonies large enough for manipulation were also observed in the 1st quadrant of CCEY for all three replicates after 48 h (Table 5-2). Recovery was only seen in one replicate after 72 h incubation with 10^4 spores in CCFA with growth being detected in the first quadrant (Table 5-2). ChromID C. difficile delivered the best recovery rate of the four agars in this study with growth being detected in the 2nd quadrant in one replicate but it was not as sensitive as CCEY agar being able to recover 10^4 spores (Table 5-2).

Table 5-2: Recovery of *C. difficile* from spiked non-infected mouse faeces using a direct plating method.

	CCFA		**CCEY**		**ChromID *C. difficile***		**TSA (5% sheep blood)**	
	Time (h)	Recovery[a]	Time (h)	Recovery[a]	Time (h)	Recovery[a]	Time (h)	Recovery[a]
10^4	72	0/1[b]	48	1	48	1/2[d]	72	0
10^3	72	0	72	0/1[c]	72	0	72	0
10^2	72	0	72	0/1[b]	72	0	72	0
10^1	72	0	72	0	72	0	72	0

a Recovery semi-quantified as 0 = no growth. 1 = 1st quadrant, 2 = 2nd quadrant, 3 = 3rd quadrant, 4 = 4th quadrant.

b Recovery was semi-quantified as 0 on two plates and 1 on one plate.

c Recovery was semi-quantified as 1 on two plates and 0 on one plate.

d Recovery was semi-quantified as 1 on two plates and 2 on one plate.

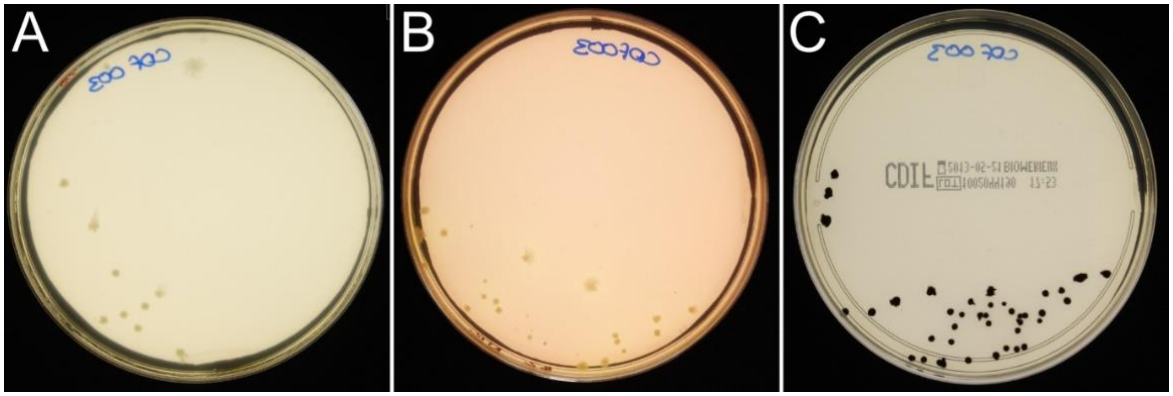


Figure 5-3: An example of comparison plates when identifying *C. difficile* from the best performing media. Each plate was inoculated with an identical clinical sample. (A) CCEY; (B) CCEY with 1% Neutral Red; (C) ChromID *C. difficile*.

The ability to identify *C. difficile* in mixed faecal flora is important to prevent laborious identification steps. The two best performing solid media were trialled using four different clinical samples (kindly provided by Dr Steve Michell, University of Exeter). After emulsification in PBS at a 1:1 ratio samples were heat shocked and centrifuged as before and 50 µL supernatant was plated on either ChromID C. difficile and CCEY agar, with and without 1% Neutral Red (to act as a pH indicator). CCEY, with and without 1% Neutral Red, gave a typical flat grey colony appearance (Figure 5-3A) although plates containing 1% Neutral Red produced a slight yellow shade (Figure 5-3B). The addition of the 1% Neutral Red did not give an adequate colour change to be considered an indicator of *C. difficile*. The ChromID agar, C. difficile generated the expected black colonies due to the chromogen integrated into the media (Figure 5-3C). No other faecal flora was detected in any of the samples.

Due to these data, it was decided to use a 1:1 ratio where possible of stool to PBS. Stool PBS mixtures were homogenised, heat shocked and then centrifuged. 50 µL sample supernatants were plated onto CCEY in triplicate to cultivate *C. difficile* for isolation and typing.

5.2.2 Isolation of **Clostridium difficile** *strains from patients within the University of Nottingham Hospitals Trust*

During the period 1st November 2013 to 30th September 2014 patients who were clinically diagnosed with CDI were considered for this study if they had not had an episode of CDI within the previous six months. Patients were followed up for at least three months after their index episode of CDI to identify any recurrence events. During this period, there were 235 reported incidents of CDI combining both community / hospital acquisition and recurrence episodes.

Isolation of *C. difficile* from stool was performed as described above. Briefly stool samples were mixed in PBS 1:1 dilution, heat shocked (10 min at 80°C), centrifuged and 50 μL supernatant was plated using a four-quadrant streak, plates were incubated for 48 h. After initial incubation, plates were inspected for the presence of *C. difficile* and up to 20 individual isolates per sample were picked and resuspended in 200 μL BHIS broth in a 96 well microtiter plate and allowed to incubate overnight. After which, a 1:10 dilution of cell suspension was made in PCR grade H_2O in a separate 96 well microtiter plate to act as a ribotyping template. One drop of glycerol was added to each well of the 96 well microtiter plate containing the *C. difficile* suspensions and stored at -80°C until further use. The PCR template 96 well microtiter plate was stored a -20°C.

If *C. difficile* was not recovered from the index sample the patient was disregarded from the study. This occurred on five occasions resulting in a recovery rate of 95.9% (n=122).

5.2.3 Statistical analysis of patient demographics and recurrence cases.

The data presented in Table 5-3 represents the clinical information collected at the index sample, these data were then analysed further to assess if the patient went onto suffer a relapse episode. The patient demographics recorded at this time were sex, age, location of

CDI acquisition, co-infection, initial CDI therapy, if additional antibiotics given during CDI therapy, regular prescriptions of proton pump inhibitors during or after CDI therapy and CRP levels at the time of stool collection (Table 5-3). Statistical analysis of these data in relation to the likelihood of a recurrence episode was performed using Fisher's exact test. These data show that 70% of patients in this study were ≥ 65 years of age and 59% of cases were female. None of these demographics were deemed statistically significant, however, of these it is interesting to note that 10/11 patients who relapsed were ≥ 65 years of age. Acquisition of CDI occurred in hospital for 88% of patients, of the 12% that acquired their CDI in the community two patient's relapsed. Co-infection occurred in 12% patients and although not statistically significant only one suffered relapse. Initial CDI therapy included metronidazole, vancomycin, fidaxomicin, a combination of the previous antibiotics, no treatment or unknown. Statistical analysis of metronidazole or vancomycin when compared to fidaxomicin showed no statistical significance (Table 5-3) but it is interesting to note that none of the patients who received fidaxomicin suffered a relapse episode (n=5). The use of additional antibiotics during CDI therapy were recorded in 37% of patients, 44% were not given additional antibiotics and in 19% it was unknown if they had received additional antibiotics. There was no statistical significance seen between the use of additional antibiotics or not with regards to relapse. A regular prescription of proton pump inhibitors during or after CDI therapy was recorded in 22% of patients of which 4 (18%) relapsed, 59% of patents did not receive proton pump inhibitors during or after CDI therapy of which 5 (8%) relapsed. Although not statistically significant a higher percentage of patients relapsed if they were receiving proton pump inhibitors. CRP levels were found to be >85mg/L in 40% of patients, <85mg/L in 36% of patients and untested in 24% patients.

5.2.4 Ribotype analysis of **Clostridium difficile** *strains*

Twenty individual isolates of *C. difficile* were isolated from 75% of stool samples within this study (range 3 – 20). All isolates were identified as *C. difficile* except for two; patient 05 had a mixture of *Clostridium glycocellium* and *C. difficile* RT005 and patient 98 had a mixture of *Clostridium butyricum* and *C. difficile* RT078. Co-infecting strains of *C. difficile* were found in 12% of samples and no combinations of ribotypes were detected more than once. Of the 13 patients who were co-infected with multiple strains of *C. difficile* only two suffered recurrence episodes. Patient 10 had RT062 and RT023 in their index sample and RT023 in their recurrence sample and patient 39 had RT001 in their index sample and RT001 and RT070 in their recurrence sample. Although it cannot be said for certain that both ribotypes were not present in all samples. A total of 41 different ribotypes were identified in this study with the most prevalent shown in Figure 5-4.

Table 5-3: Patient demographic statistics taken at the index episode of *Clostridium difficile* infection.

		Future Recurrent CDI		
	All (n=100)	Yes (n=11)	No (n=89)	*P* Value[b]
Female	59	6	53	0.7561
Male	41	5	36	
Age <65 y	30	1	29	0.1658
Age ≥65 y	70	10	60	
Hospital Acquired CDI	88	9	79	0.6176
Community Acquired CDI	12	2	10	
Not co-infected	89	11	78	0.6051
Co-infected	11	0	11	
Initial CDI therapy				
Metronidazole	35	5	30	>0.999[d]
Vancomycin	35	3	32	>0.999[d]
Fidaxomicin	5	0	5	
Multiple[a]	9	3	6	
None	6	0	6	
Unknown	10	0	10	
Use of antibiotics other than CDI therapy during the treatment course				
Yes	37	2	35	0.1700
No	44	7	37	
Unknown[c]	19	2	17	
Regular prescription of proton pump inhibitors during or after CDI therapy				
Yes	22	4	18	0.2450
No	59	5	54	
Unknown[c]	19	2	17	
CRP >85mg/L				
Yes	40	3	37	0.2937
No	36	6	30	
Unknown[c]	24	2	22	

[a]Multiple antibiotics were used for initial CDI treatment were vancomycin / metronidazole (n=7) and vancomycin / fidaxomicin (n=2)
[b]*P* value calculated by Fishers exact test.
[c]Unknown results were excluded from statistical calculations.
[d]When analysed against Fidaxomicin.

5.2.5 Minimum inhibitory concentration analysis

The minimum inhibitory concentrations for all *C. difficile* isolates were determined by minimum inhibitory concentrations evaluators (MICE). Antibiotics assayed were vancomycin and metronidazole, resistance for both antibiotics is defined as an MIC >2 mg/L. Each isolate was tested once unless an MIC of resistance was recorded after which the result was confirmed in triplicate. All isolates in this study were sensitive to both antibiotics (MIC data shown in Appendix Five). The MIC^{90} was calculated as 1 mg/L for

both metronidazole and vancomycin. The MIC^{50} was calculated as 1 mg/L for vancomycin and 0.5 mg/L for metronidazole.

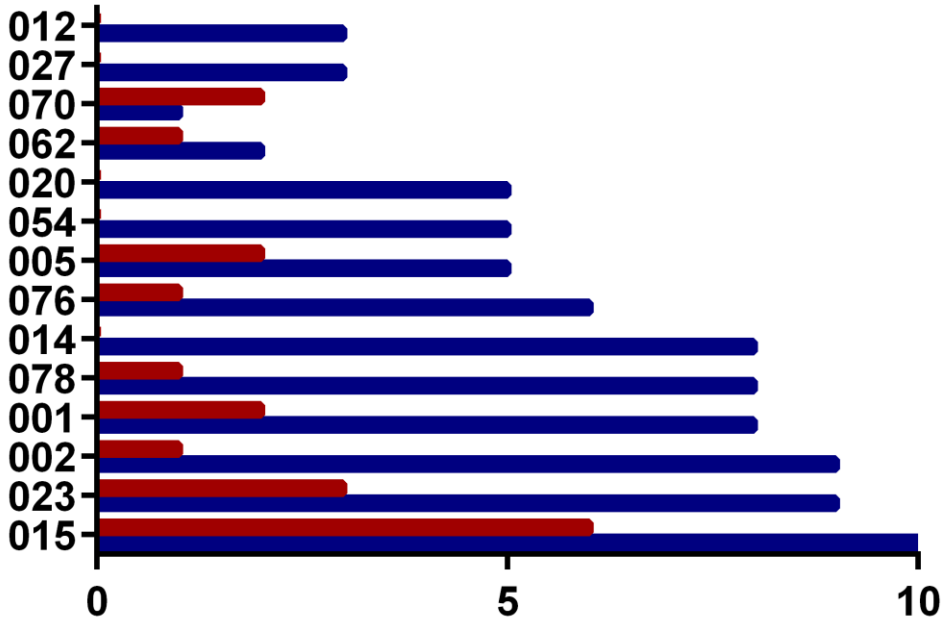


Figure 5-4: The 14 most prevalent ribotypes isolated in this study. Figure legend; blue bars indicate Hospital Acquired CDI and red bars indicate Community Acquired CDI.

5.2.6 Whole genome sequencing, global alignment and single nucleotide polymorphism analysis of Clostridium difficile strains from patients with recurrent infection

C. difficile strains isolated from patients with recurrent infection were whole genome sequenced. To serve as a control, one isolate of a corresponding ribotype isolated from a non-relapse patient was also sequenced (Table 5-5). The hypothesis being that the isolates of the same ribotype isolated from both relapse and non-relapse patients should be genetically very similar and therefore differing genes and SNPs would be more indicative of providing a fitness benefit.

Illumina MiSeq paired end short reads were adapter trimmed by the providing centres using their own methods. Returned reads were quality checked using FastQC[284] revealing

increased noise at both 5' and 3' ends. Reads were further trimmed using CLC genomics workbench V8.5.3 (CLC) to remove this noise and filtered to remove broken pairs and reads <50bp. Reads were then quality checked a second time using FastQC revealing an improvement in overall quality. Figure 5-5 gives an example of FastQC output from the Illumina MiSeq reads received from isolate 01-01 before and after trimming. Basic statistics for MiSeq paired end reads before and after trimming are presented in Table 5-4.

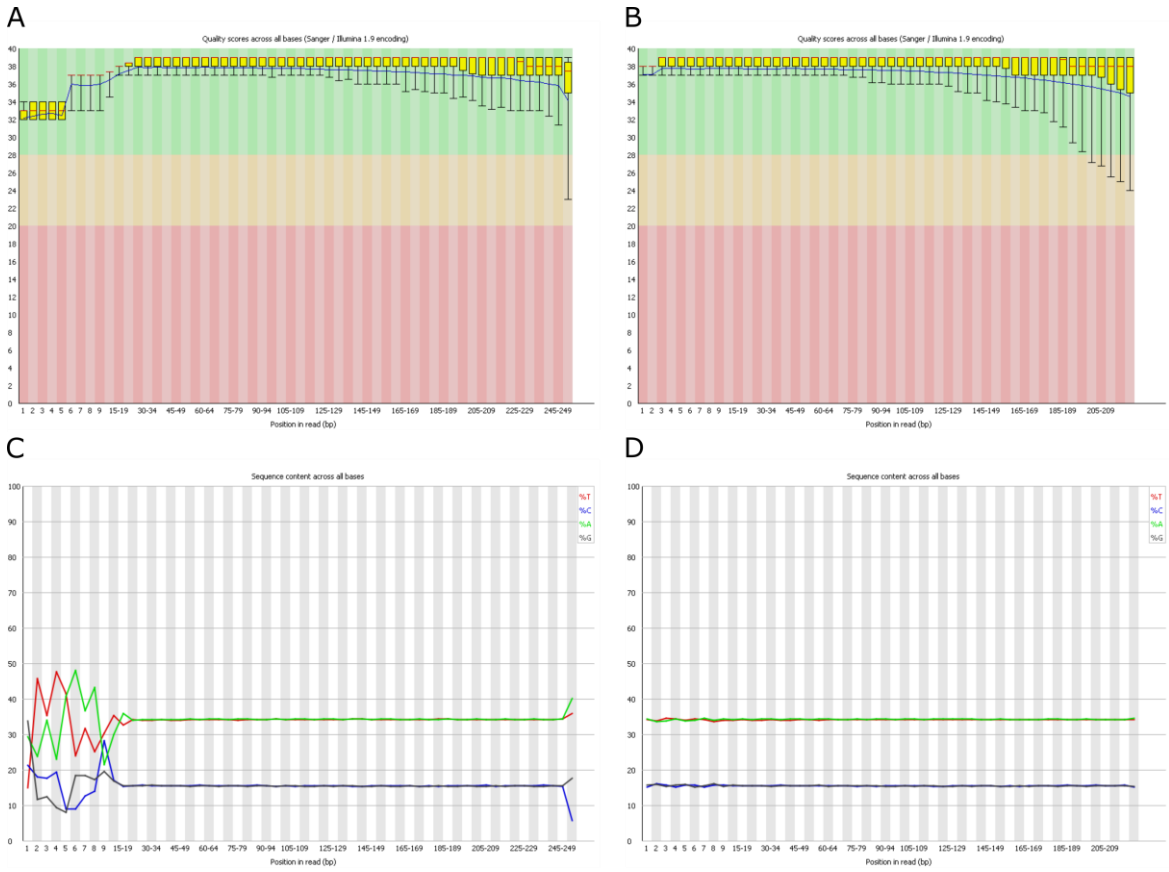


Figure 5-5: FastQC analysis of Illumina MiSeq reads from 01-01. Quality Scores across all bases (A) before and (B) after trimming. Sequence content across all bases (C) before and (D) after trimming. Showing a general improvement in quality score for bases and reduced noise at the 5' and 3' ends of the reads.

Table 5-4: Statistics for the Illumina MiSeq reads after in-house filtering and trimming using CLC Genomics workbench V8.5.3

Strain numbers	Sequencing Centre[a]	Ave. reads	Ave. reads after trimming	Ave. length	Ave. length after trimming
34-01	NTU	920,240	911,644	240.1	211.8
09-01	NTU	1,859,432	1,814,837	217.7	191.6
01-01	NTU	1,219,978	1,207,351	238.8	210.6
06-01	UoB	1,875,458	1,768,633	213.7	193.1
69-01	NTU	1,122,342	1,101,270	226.7	199.9
57-01	UoB	1,340,932	1,199,476	193.3	179.2
96-01	UoB	4,314,276	4,033,025	211.3	192
96-02	NTU	1,104,892	1,094,587	239.5	221.2
96-03	NTU	1,298,572	1,283,042	236.6	208.7
96-04	NTU	1,129,934	1,115,979	237.8	210
39-01	NTU	1,629,804	1,596,732	220.7	194
39-02(1)	NTU	1,885,308	1,831,138	215..	189.6
05-01	NTU	909,882	900,816	232.3	204
05-02	UoB	3,180,140	2,966,539	209.6	190.5
19-01	NTU	1,463,294	1,442,126	231	203.5
19-02	NTU	944,136	929,608	238.1	210.9
42-01	NTU	1,230,402	1,219,387	235.6	207.2
42-02	NTU	1,578,852	1,555,007	227.4	200.5
45-01	NTU	1,051,788	1,043,009	238,5	210
45-02	NTU	1,359,826	1,338,818	230.8	203.5
85-01	NTU	1,528,166	1,509,706	235.9	208
85-02	NTU	1,466,500	1,036,585	240.1	211.9
10-01(2)	NTU	1,019,070	1,004,758	231.8	204.3
10-02	NTU	980,884	968,102	232.2	204.5
12-01	NTU	1,227,136	1,210,203	229.1	201.5
12-02	NTU	1,129,512	1,114,657	232.2	204.5
55-01	NTU	1,475,640	1,451,743	229.6	202
55-02	NTU	1,328,590	1,309,460	230.5	203
61-01	UoB	1,858,536	1,724,740	206.4	187.8
61-02	NTU	1,042,430	1,028,998	235.7	207.9
61-03	UoB	2,471,980	2,255,276	200.3	184
61-04	UoB	1,711,572	1,437,313	174.1	166.2
61-05	UoB	1,854,494	1,739,055	210.9	191

[a]UoB – University of Birmingham; NTU – Nottingham Trent University.

Sequence reads for the control strains and index samples for each of the relapse patients were assembled *de novo* in CLC genomics workbench using standard settings. Assembled contigs for each genome were aligned to the CD630 reference genome map using Mauve's[120] move contigs feature. Rearranged contigs were then annotated using the RAST pipeline[121,122] and exported as one consensus .gbk file. Further analysis of the relapsing strain genomes was performed in two ways. Firstly, by using progressiveMauve to inspect the genome of the index strains for genetic regions not identified in CD630, R20291 and the control reference genome. Secondly paired end reads of the relapsing strains were mapped to the index strain to identify single nucleotide variants (SNPs) as described in 2.7.3. Only SNPs that met specific read criteria were included; a minimum coverage of 20, count of 10, frequency of 90% and Phred score >30. SNPs were manually validated in Artemis[115], any located within 10 bases of a contig end were excluded.

Each index strain from the relapse patient was compared to the control strain as described above to assess their relatedness by mapping the paired-end reads from each index sample to the corresponding control for their ribotype (Table 5-5). These data are collected under the assumption that the more SNP differences between strains the more unrelated they are[285]. High numbers of SNP within the same ribotype can be attributed to dense SNP clusters resulting from homologous recombination events[285]. This has not been investigated here due to time limitations but it may explain the high SNP differences seen between the control strains for each tested ribotype and the test strains. The largest SNP difference of all the isolates studied were the two RT001 strains isolated from the index sample of PT96 (relapsing strain) and PT34 (RT001 control strain) with 19,740 SNPs (Table 5-5). The smallest SNP difference were seen in two RT054 strains isolated from the index sample of PT55 (relapsing strain) and PT69 (RT054 control strain) with 22 SNPs (Table 5-5). These data suggest there were no transmission events between these patients.

However, this does not rule out transmission occurring between patients that were not included in the whole genome study. The number of SNPs between index strain and relapsing strains to be classed as recurrent infection is ≤2[261]; this indicates that patients PT05, PT39 and PT45 all relapsed with a different strain of the same Ribotype (Table 5-5).

Table 5-5: List of number of SNPs compared to a control strain and the index strain of all isolates taken from relapse patients.

Patient	Ribotype	Control Strain isolated from patient	No. of SNPs compared to control strain	No. of SNPs compared to index strain at first relapse	No. of relapse episodes
PT39	001	PT34	1,462	4	1
PT96	001	PT34	19,740	1 – 2*	3
PT05	005	PT09	906	79	1
PT19	015	PT01	5,660	1	1
PT42	015	PT01	5,589	0	1
PT45	015	PT01	552	6	1
PT85	015	PT01	5,509	0	1
PT10	023	PT06	171	0	1
PT12	023	PT06	688	0	1
PT55	054	PT69	22	0	1
PT61	076	PT57	189	0	4

*A second SNP was identified in the isolate from the isolate from the second relapse that was not present in the first or third.

For each patient, clinical information was gathered (where possible) at the same time as their standard of care relapse stool sample was taken. This combined with SNP, vancomycin/metronidazole MIC and the identification of unique strain genetic regions will be used to assess the likelihood of contributing to relapse. Non-synonymous SNP changes will be analysed using PROVEAN (Protein Variation Effect Analyzer)[286], this web based software utilises pairwise sequence alignments to assess homology between the original and the altered sequence. A change in alignment score is interpreted as the impact of the

variation on protein function. Briefly, the software clusters BLAST hits with an e-value cut-off of 0.1, the top 30 clusters are used for the delta alignment score and these scores are then averaged within and across clusters to generate the final score. If this score is <-2.5 the variation is predicted deleterious. PHASTER (Phage Search Tool Enhanced Release)[114] was used to look for prophage regions within the genomes when unique locations were believed to contain phage related genes. This web based software performs database comparisons using both GLIMMER (ORF prediction) and BLAST (protein identification) to locate prophage regions (using a phage-specific sequence database) resulting in annotated reports.

Table 5-6: Relapse patient demographics, *Clostridium difficile* ribotype and length of infection.

Patient sample	Sex	Age	Ribotype	Date of sample	No of days between episodes
39-01	Male	93	001	23-03-14	
39-02(1)			001	19-05-14	57
96-01	Female	55	001	01-08-14	
96-02			001	20-08-14	19
96-03			001	25-09-14	36
96-04			001	03-11-14	39
					94[a]
05-01	Female	90	005	05-11-13	
05-02			005	19-12-13	44
19-01	Female	79	015	08-01-14	
19-02			015	15-04-14	97
42-01	Female	96	015	30-03-14	
42-02			015	25-04-14	26
45-01	Male	95	015	03-04-14	
45-02			015	29-06-14	87
85-01	Male	66	015	16-07-14	
85-02			015	11-08-14	26
10-01(2)	Female	69	023	10-12-13	
10-02			023	03-03-14	83
12-01	Male	85	023	13-12-13	
12-02			023	20-01-14	38
55-01	Male	79	054	23-05-14	
55-02			054	22-06-14	30
61-01	Female	69	076	03-06-14	
61-02			076	30-06-14	27
61-03			076	19-07-14	19
61-04			076	07-09-14	50
61-05			076	03-12-14	87
					183[a]

[a]Total number of days between taken samples.

5.2.7.1 RT001

RT001 was the causative agent of CDI in two patients who suffered recurrent CDI.

Patient 39; a male 93 years of age had one relapse episode 57 days after initial diagnosis (Table 5-6). The index episode for this patient was determined to be hospital acquired and the relapse was classified as occurring in the community (Table 5-7). The recurrence sample was shown to have two distinct ribotypes; RT001 and RT070 both were sensitive to vancomycin and metronidazole *in vitro*. There were no recorded additional antibiotic treatments except for the CDI therapy which was vancomycin and fidaxomicin. The proton pump inhibitor omeprazole was prescribed during CDI therapy for the index infection, this is a risk factor for relapse. The CRP level was not raised to a level indicative of a risk of relapse. Only isolate 39-02(1) (RT001) was chosen for further genetic analysis.

Table 5-7: Clinical information for patient 39 gathered at each sample collection.

Sample Date	**23/03/14**	**19/05/14**
Hospital Acquired CDI Community Acquired CDI	Hospital	Community
Co-infected	No	Yes (RT070)
CDI Treatment	Vancomycin then Fidaxomicin	Fidaxomicin
Use of antibiotics other than CDI therapy during the treatment course	None	None
Regular prescription of proton pump inhibitors during or after CDI therapy	Omeprazole	None
CRP >85mg/L	81mg/L	Unknown
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	1 mg/L	RT001 – 1 mg/L RT070 – 0.5 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	0.25 mg/L	RT001 – 0.5 mg/L RT070 – 0.25 mg/L

Genome analysis revealed four SNPs unique to sample 39-02(1) when compared to the control strain 34-01 (Table 5-5), one ambiguous and three non-synonymous (NS). The ambiguous SNP is located 28bp upstream of a hypothetical protein. The first NS SNP is in *PT3902_478* a diguanylate cyclase domain protein where an 841C>T a transition substitution results in a His281Tyr. The second NS SNP is in *PT3902_1837* a putative ABC transporter permease protein where an 719T>G a transversion substitution results in a Val240Gly. The third NS SNP is in *PT3902_2026* a hypothetical protein where a 140A>G a transition substitution results in Tyr47Cys. All of which were shown to be a tolerated substitution by PROVEAN suggesting these SNPs are not providing a fitness benefit for this strain.

Whole genome alignment of 39-01 to 630, R20291 and 34-01 showed two unique phage regions in 39-01 at 62.8kb (position 1,088,587) and 107.5kb (position 3,398,363) in length, analysis by PHASTER showed they resemble *C. difficile* bacteriophages ΦMMP02 and ΦCD211 respectively. These regions include phage related genes, hypothetical proteins, transcriptional regulators, spore related proteins and transposases. A third unique region shares 99% identity and 100% coverage with *Tn6218* (accession number HG002386) located in RT001 strain Ox746b presented in a study by Dingle *et al* [287]. The insertion has occurred ~6kb upstream of the flagella operon F3 at position 294,535 in strain PT39-01 which contains a MATE family of multi drug resistance (MDR) efflux pumps (Figure 5-6).

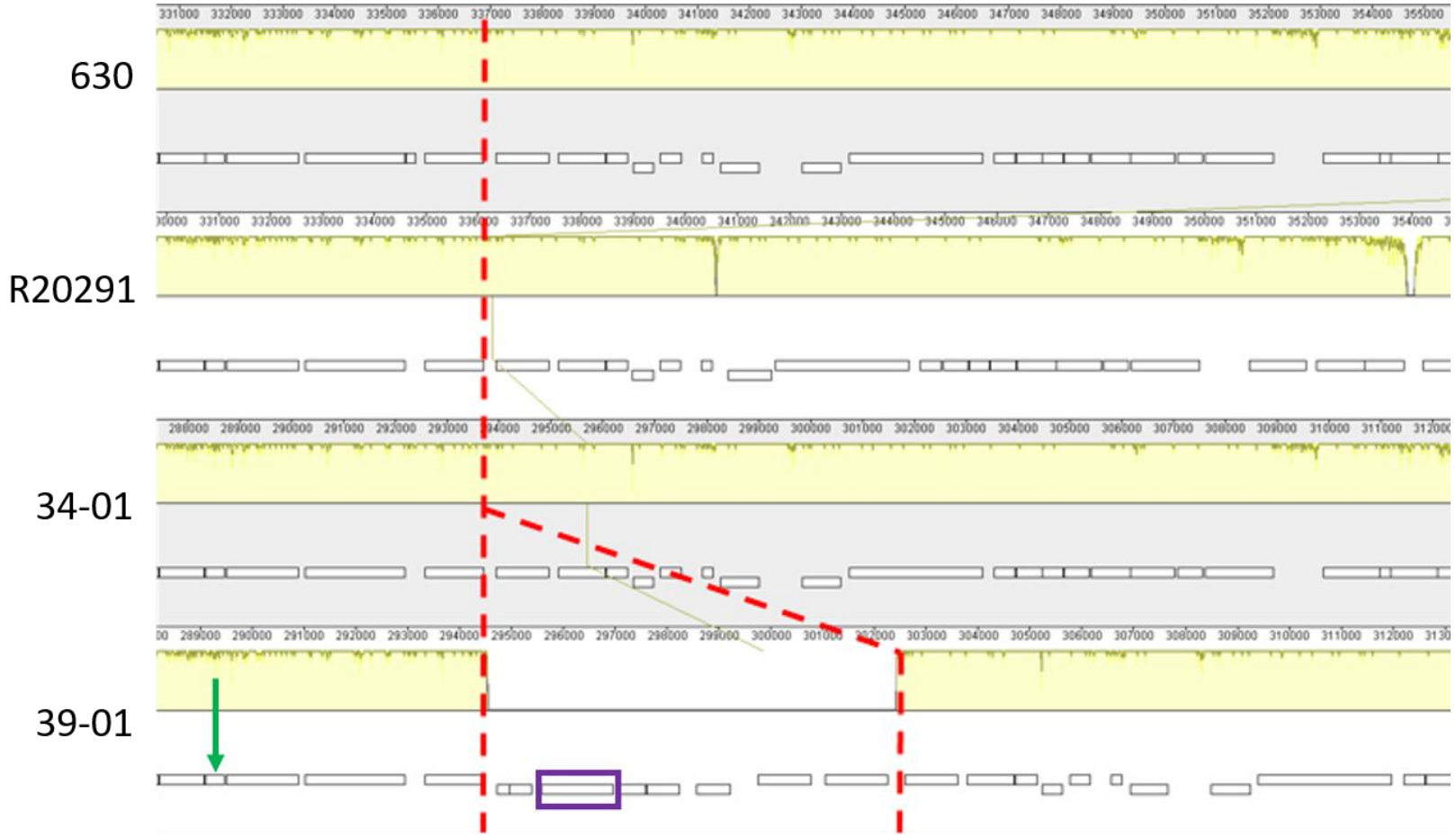


Figure 5-6: A Mauve alignment between *Clostridium difficile* strains 630, R20291, 34-01 (RT001 control strain) and 39-01 the index strain from the relapsing patient PT39. The red checked lines indicate an insertion of *Tn6218* containing MATE family of multi drug resistance (MDR) efflux pumps (purple box) at position 294,535 in 39-01 (7,879bp in length). The green arrow indicates the last gene in the F3 flagella operon ~6kb downstream of the insertion site. This region appears highly conserved between 630, R20291 and 34-01 but the insertion. This insertion is located within a 240,846bp contig reducing the chances of an assembly error.

Patient 96; a female 55 years of age had four distinct episodes of CDI totalling a period of 94 days (Table 5-6). The index case was clinically determined to be hospital acquired along with the first two relapse episodes, the final recorded relapse was classified as occurring in the community (Table 5-8). The clinical information gathered at the time of each CDI episode showed no risk factor for recurrence of CDI and it appears that the patient only received treatment for her infection after identification of the index episode and not for the relapse episodes. *In vitro* MIC analysis of *C. difficile* strains showed sensitivity to both metronidazole and vancomycin.

When comparing strain 96-01 to 34-01 (RT001 control strain), analysis revealed one SNP C>T a transition substitution present in all recurrence strains located in a 58bp intergenic region between *tRNA* and *rRNA* genes (Table 5-5). There was also one NS SNP found in gene *PT9603_3091*, a hypoxanthine-guanine phosphoribosyl transferase, where a 425T>C a transition substitution results in Val142Ala. This was shown to be a tolerated substitution by PROVEAN.

Table 5-8: Clinical information for patient 96 gathered at each sample collection.

Sample Date	**01/08/14**	**19/08/14**	**25/09/14**	**03/11/14**
Hospital Acquired CDI Community Acquired CDI	Hospital	Hospital	Hospital	Community
Co-infected	No	No	No	No
CDI Treatment	Metronidazole	None	None	None
Use of antibiotics other than CDI therapy during the treatment course	None	None	None	None
Regular prescription of proton pump inhibitors during or after CDI therapy	None	None	None	None
CRP >85mg/L	11 mg/L	10 mg/L	32 mg/L	Unknown
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	1 mg/L	0.5 mg/L	0.5 mg/L	1 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	1 mg/L	0.5 mg/L	0.5 mg/L	0.5 mg/L

Whole genome alignment of 96-01 to 630, R20291 and 34-01 showed one unique phage region of ~33kb, analysis by PHASTER showed it resembles *C. difficile* bacteriophage ΦMMP02 although it is incomplete. When compared to 630, R20291 and 34-01 there is evidence of a unique 25kb *Tn916*-like element containing hypothetical proteins, conjugation proteins, membrane proteins, IAA acetyltransferase and β-ketoadipate enol-lactone hydrolase (Figure 5-7). The *Tn6218* element identified in 39-01 was similarly found in this strain (position 3,829,473), it is however in a different location (Figure 5-8). This may be due to an assembly error that occurred during *de novo* process however both regions are located within longer contigs (240,846bp contig in PT39 and 124,615 contig in PT96). This transposon element has been shown to be located in different regions within different ribotypes but this has yet to be seen within the same ribotype[287]. Sanger sequencing or long read sequencing technologies could be used to identify if these are the true location or due to miss assembly.

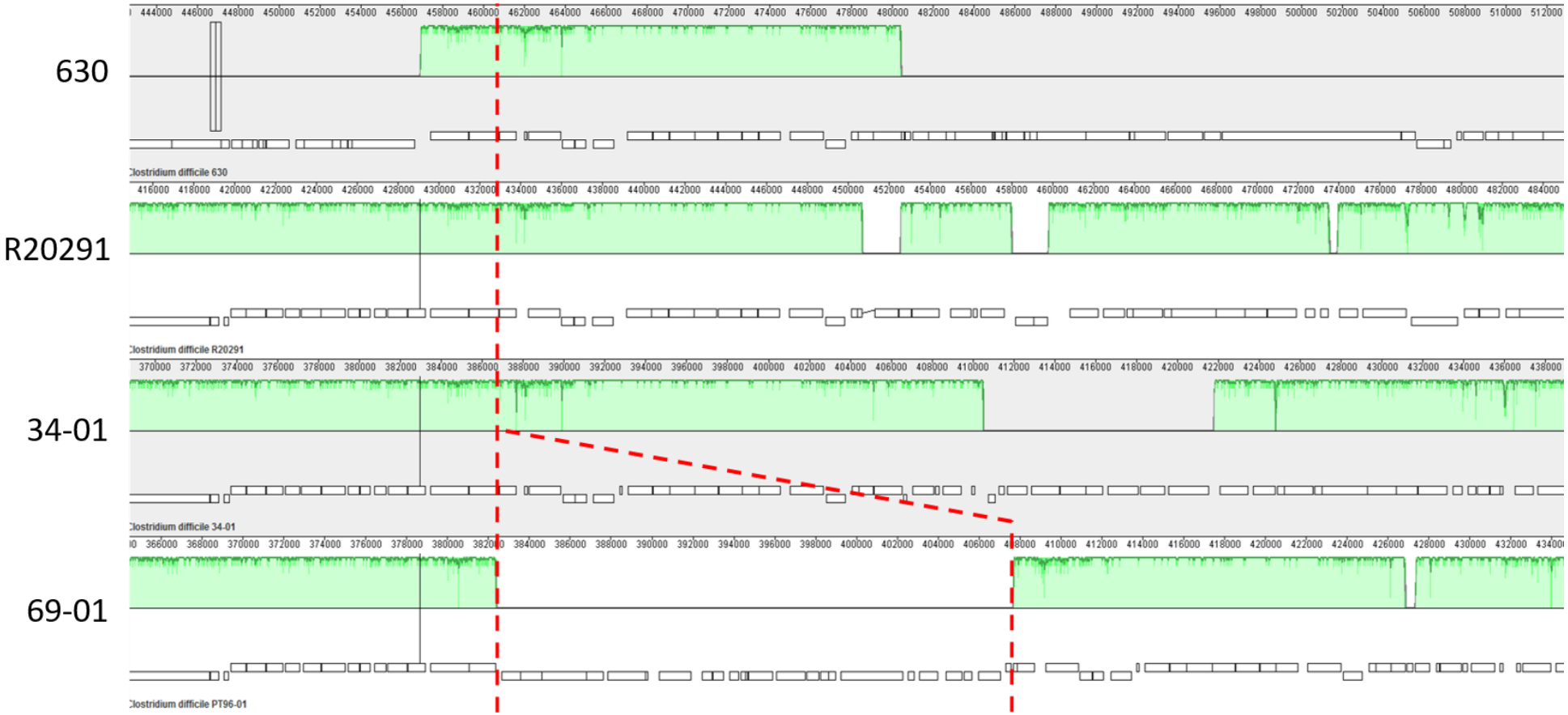


Figure 5-7: A Mauve alignment between *Clostridium difficile* strains 630, R20291, 34-01 (RT001 control strain) and 96-01 the index strain from the relapsing patient PT96. The red checked lines indicate a unique insertion of *Tn916*-like element containing hypothetical proteins, conjugation proteins, membrane proteins, IAA acetyltransferase and β-ketoadipate enol-lactone hydrolase at position 382,468 in 96-01 (25,243bp in length). This insertion comprises of one contig which does not span any locally collinear blocks so misassembly cannot be discounted. This region appears to have a lot of variability between the four strains which can be identified by the lack of pairwise locally collinear blocks (green regions).

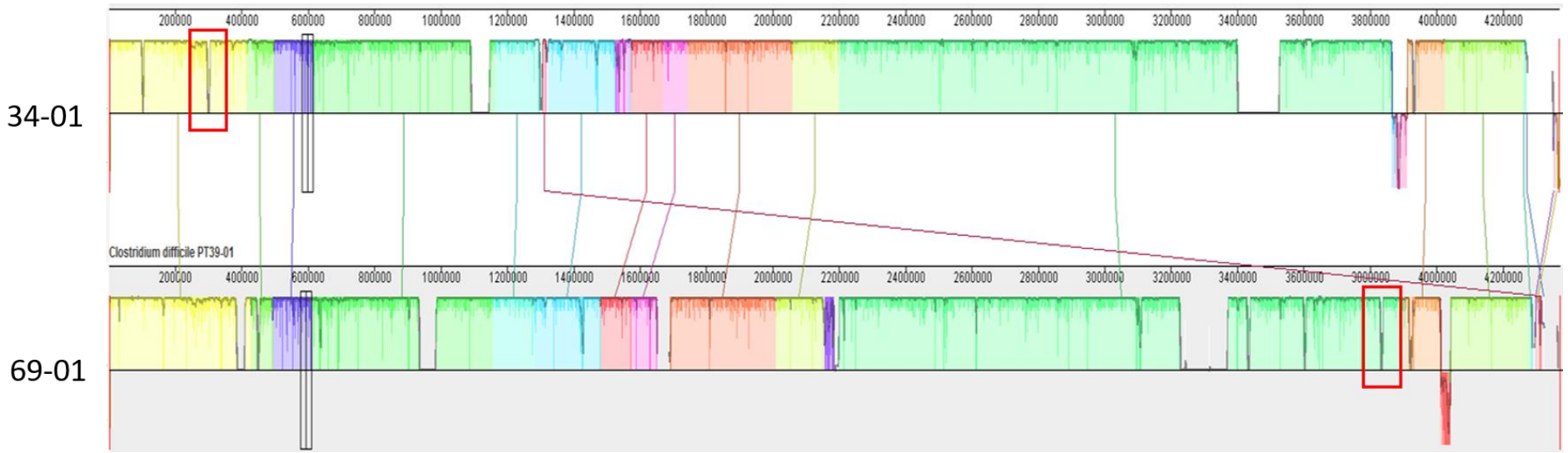


Figure 5-8: A Mauve alignment of *Clostridium difficile* strains 39-01 and 96-01 both RT001. Red boxes indicate the locations of a *Tn6218* element, previously identified in Ox746b[287]. Both are located in different regions (294,535 in 39-01 and 3,829,473 in 96-01) but within much larger contigs (240,846bp contig in 39-01 and 124,615 contig in 96-01) reducing the likelihood of misalignment. Similar locally collinear blocks between strains are linked by matching coloured lines.

5.2.7.2 RT005

RT005 was the causative agent of CDI in one patient who suffered recurrent CDI.

Patient 05; a female 90 years of age who had a relapse episode 44 days after initial diagnosis (Table 5-6). Both index and relapse episodes were categorised as community acquired infections (Table 5-9). Due to this limited clinical information was available only treatment notes were recorded; these were metronidazole for initial CDI therapy and vancomycin during the relapse episode. *In vitro* MIC analysis of *C. difficile* strains showed sensitivity to both metronidazole and vancomycin (Table 5-9).

Table 5-9: Clinical information for patient 05 gathered at each sample collection.

Sample Date	**05/11/13**	**19/12/13**
Hospital Acquired CDI Community Acquired CDI	Community	Community
Co-infected	No	No
Initial CDI Treatment	Metronidazole	Vancomycin
Use of antibiotics other than CDI therapy during the treatment course	Unknown	Unknown
Regular prescription of proton pump inhibitors during or after CDI therapy	Unknown	Unknown
CRP >85mg/L	Unknown	Unknown
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	0.25 mg/L	0.25 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	1 mg/L	1 mg/L

Mapping of Illimina MiSeq reads from 05-02 (relapse strain) to 05-01 (index strain) revealed 88 differences; 79 SNPs unique to 05-02 of which 56 are ambiguous (in a non-coding region), 4 synonymous, 19 non-synonymous, two multiple nucleotide variants

(MNV) and seven indels. Further analysis of SNPs were not performed as this is indicative of reinfection rather than relapse.

Whole genome alignment of 05-01 to 630, R20291 and 09-01 (the RT005 control strain) revealed two interesting regions, the first located at position 400,234 is ~23kb and contains hypothetical proteins, transcriptional regulators, thiamine biosynthesis protein, lantibiotic transport permease protein, lantibiotic transport ATP-binding protein and a putative *Tn916* transcriptional regulator (Figure 5-9).

A second unique region to PT05 index strain 05-01 contained ten hypothetical proteins and a macrolide-efflux protein (Figure 5-10). Although there was no record of a macrolide class of antibiotic being prescribed in this case, fidaxomicin may not have been effective if used (Table 5-9).

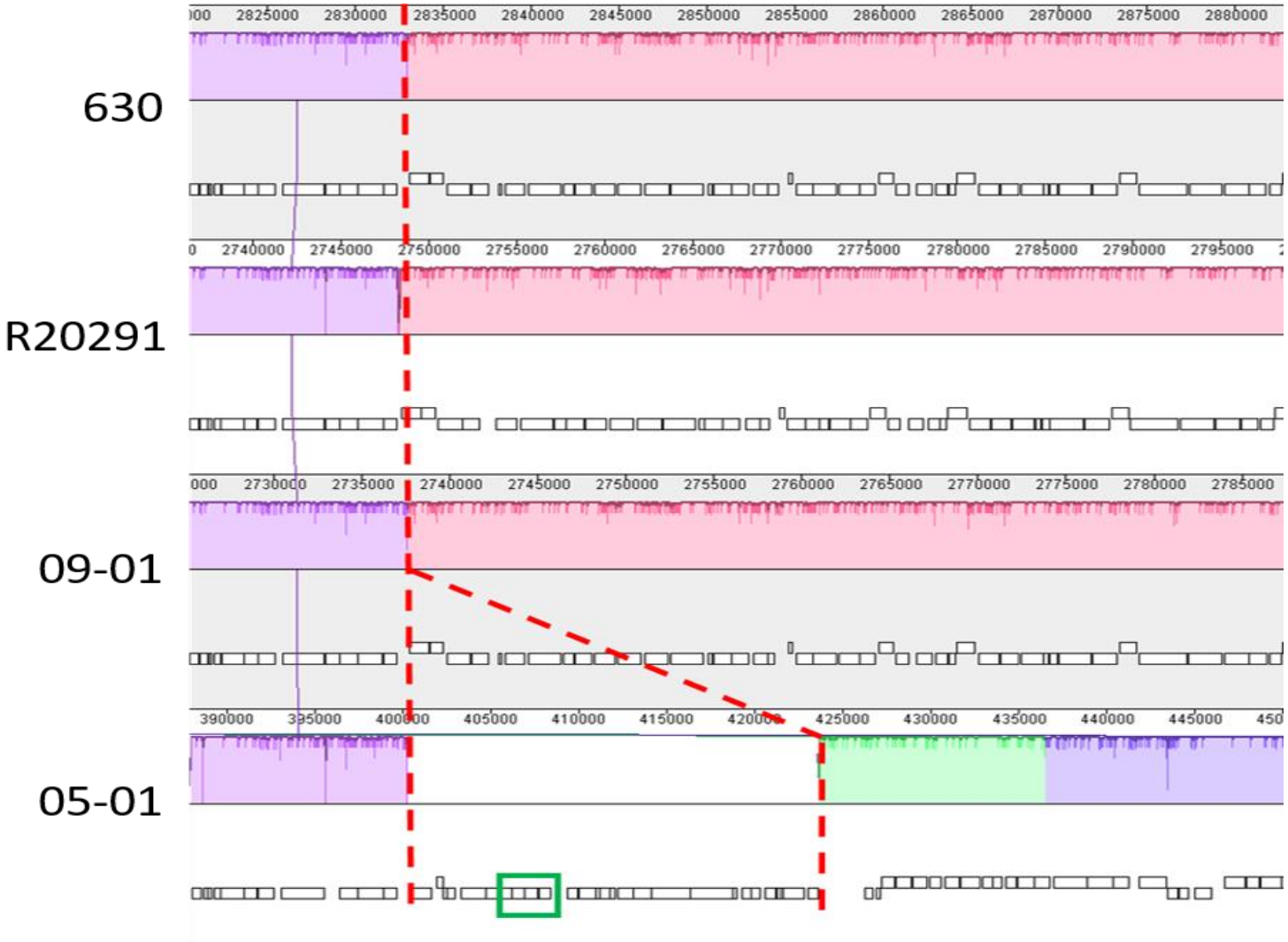


Figure 5-9: A Mauve alignment between *Clostridium difficile* strains 630, R20291, 09-01 (RT005 control strain) and 05-01 the index strain from the relapsing patient PT96. The red checked lines indicate a unique insertion of *Tn916*-like element at position 400,234 in 05-01 (23,433bp in length), green box indicates lantibiotic related proteins. The insertion is located within a contig of 34,012bp which overlaps the purple locally collinear blocks suggesting an accurate assembly. Similar locally collinear blocks between strains are linked by matching coloured lines.

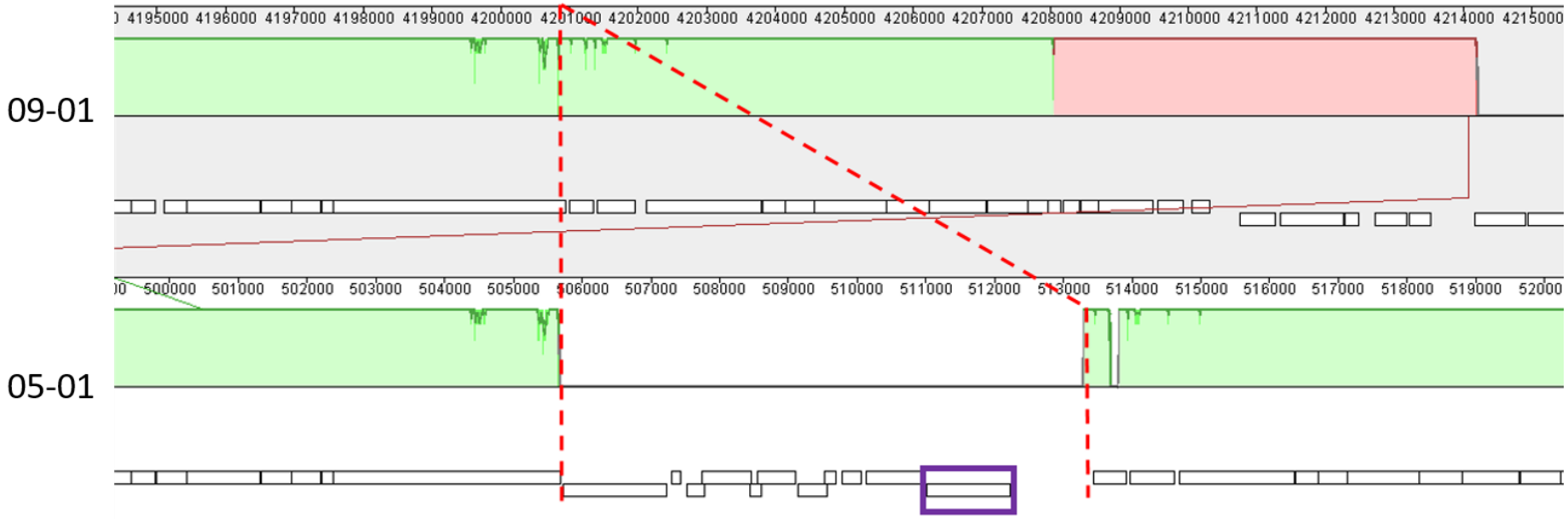


Figure 5-10: A Mauve alignment of *Clostridium difficile* strains 09-01 and 05-01 both RT005. The green locally collinear blocks shown were not found in 630 or R20291. The Red dashed lines indicate a unique genomic region to PT05 at position 50,5757 in 05-01 (7,531bp in length); Purple box indicates the location of a macrolide-efflux protein. It is possible that there have been issues with the assembly of the contigs in these genomes as the location of the green locally collinear blocks begin at either end of the genomes (position 4,174,169 in 09-01 and 480,408 in 05-01). However, this unique region is within the longer contig of 50,577bp suggesting that it is present within this region.

5.2.7.3 RT015

RT015 was the causative agent of CDI in four patients who suffered recurrent CDI.

Patient 19; a female 79 years of age had one recurrence episode 97 days after initial diagnosis (Table 5-6). The index episode for this patient was determined to be hospital acquired and the relapse was classified as occurring in the community (Table 5-10). PT19 index case was initially treated with vancomycin but there is a record of Tazocin and co-amoxiclav also having been prescribed at around the time of initial symptoms of CDI, a risk factor for relapse. Although in this case CRP was not elevated. No clinical information was available for collection during the relapse episode. *In vitro* MIC analysis of *C. difficile* strains showed sensitivity to both metronidazole and vancomycin.

Genome analysis revealed one ambiguous SNP unique to 19-02 when compared to 01-01 at position 519206 with a C>A a transversion substitution (Table 5-5). This is located ~300 bases upstream of a predicted amidohydrolase on the anti-sense strand and ~30bp downstream of a quaternary ammonium compound-resistance protein. Whole genome alignment revealed that all four strains were very similar except for a ~56kb region unique to 19-01 which contains hypothetical proteins, peptidoglycan bound protein, a putative *Tn916* regulator, iron acquisition yersiniabactin synthesis enzyme, various transport and transcriptional proteins (Figure 5-11).

Table 5-10: Clinical information for patient 19 gathered at each sample collection.

Sample Date	**08/01/14**	**15/04/14**
Hospital Acquired CDI Community Acquired CDI	Hospital	Community
Co-infected	No	No
Initial CDI Treatment	Vancomycin	Metronidazole
Use of antibiotics other than CDI therapy during the treatment course	Recent Tazocin and Co-amoxiclav	Unknown
Regular prescription of proton pump inhibitors during or after CDI therapy	No	Unknown
CRP >85mg/L	14 mg/L	Unknown
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	0.25 mg/L	0.5 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	1 mg/L	0.5 mg/L

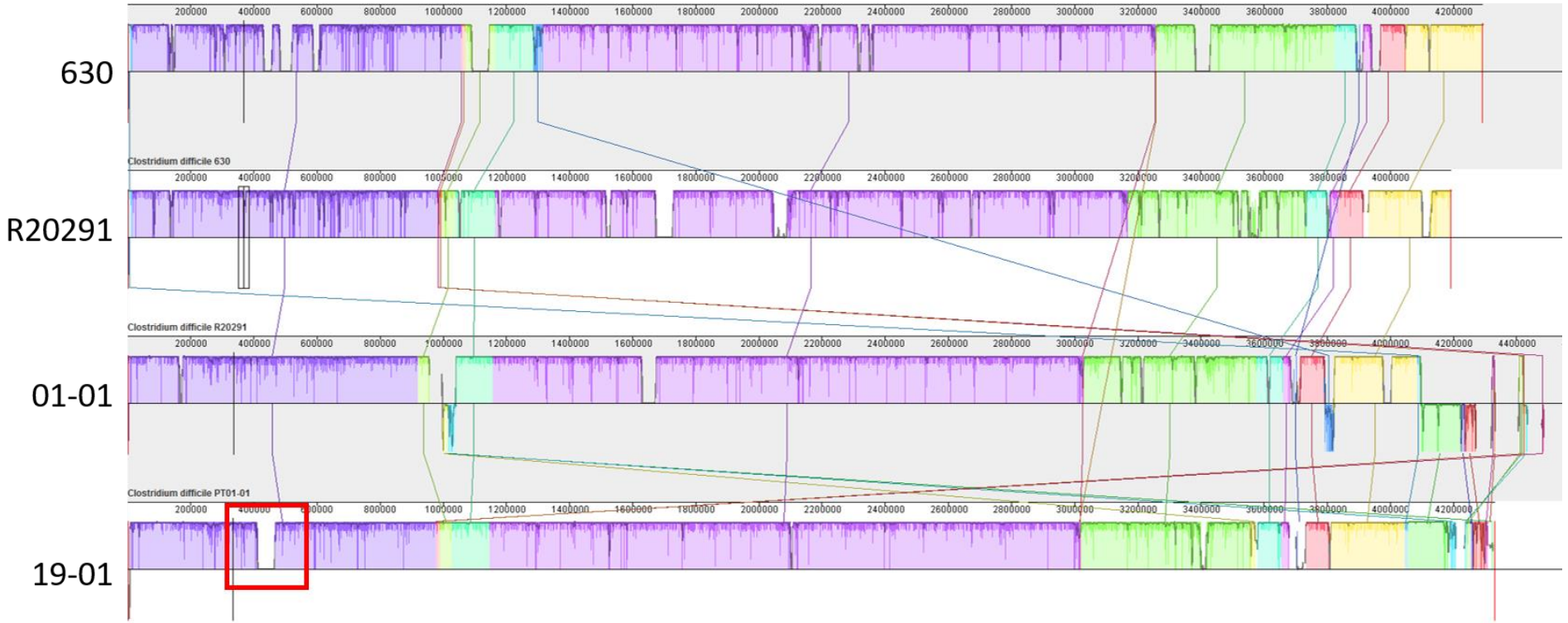


Figure 5-11: A Mauve alignment of *Clostridium difficile* strains 630, R2091, 01-01 (RT015 control strain) and 19-01 the index strain from the relapsing patient PT19. Red box indicates a unique insertion which contains hypothetical proteins, peptidoglycan bound protein, a putative *Tn916* regulator, iron acquisition yersiniabactin synthesis enzyme, various transport and transcriptional proteins at position 408,567 in 19-01 (56,834bp in length). However, this unique region is within the longer contig of 231,116bp suggesting that it is present within this region. The comparison of the genomes in general shows they are very similar as defined by the large locally collinear blocks of the same colour. Similar locally collinear blocks between strains are linked by matching coloured lines.

Patient 42, a female 96 years of age had one recurrent episode 26 days after initial diagnosis (Table 5-6). The index episode for this patient was determined to be hospital acquired and the relapse was classified as occurring in the community (Table 5-11). From the acquired data, there were no indications of risk factors for relapse episodes occurring. The patient's initial CDI therapy was both metronidazole and vancomycin which the *C. difficile* strain was determined to be sensitive to *in vitro*. During the relapse episode fidaxomicin was given as CDI therapy. There was no recorded third episode within the study period.

Table 5-11: Clinical information for patient 42 gathered at each sample collection.

Sample Date	**30/03/14**	**25/04/14**
Hospital Acquired CDI Community Acquired CDI	Hospital	Community
Co-infected	No	No
Initial CDI Treatment	Metronidazole then Vancomycin	Fidaxomicin
Use of antibiotics other than CDI therapy during the treatment course	No	No
Regular prescription of proton pump inhibitors during or after CDI therapy	No	No
CRP >85mg/L	69 mg/L	Unknown
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	1 mg/L	1 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	0.25 mg/L	0.5 mg/L

There were no SNPs identified between 42-01 and 42-02 (Table 5-5). Whole genome alignment of 42-01 revealed a unique *Tn619*-like region that contains two teicoplanin resistance genes and one tetracycline resistance protein (Figure 5-12). Lastly a *Tn6218*-like region containing hypothetical proteins, regulatory proteins and Multi antimicrobial

extrusion protein MATE family of MDR efflux pumps is located at position 3,003,729. This appears like the *Tn6218* elements found in RT001 strains.

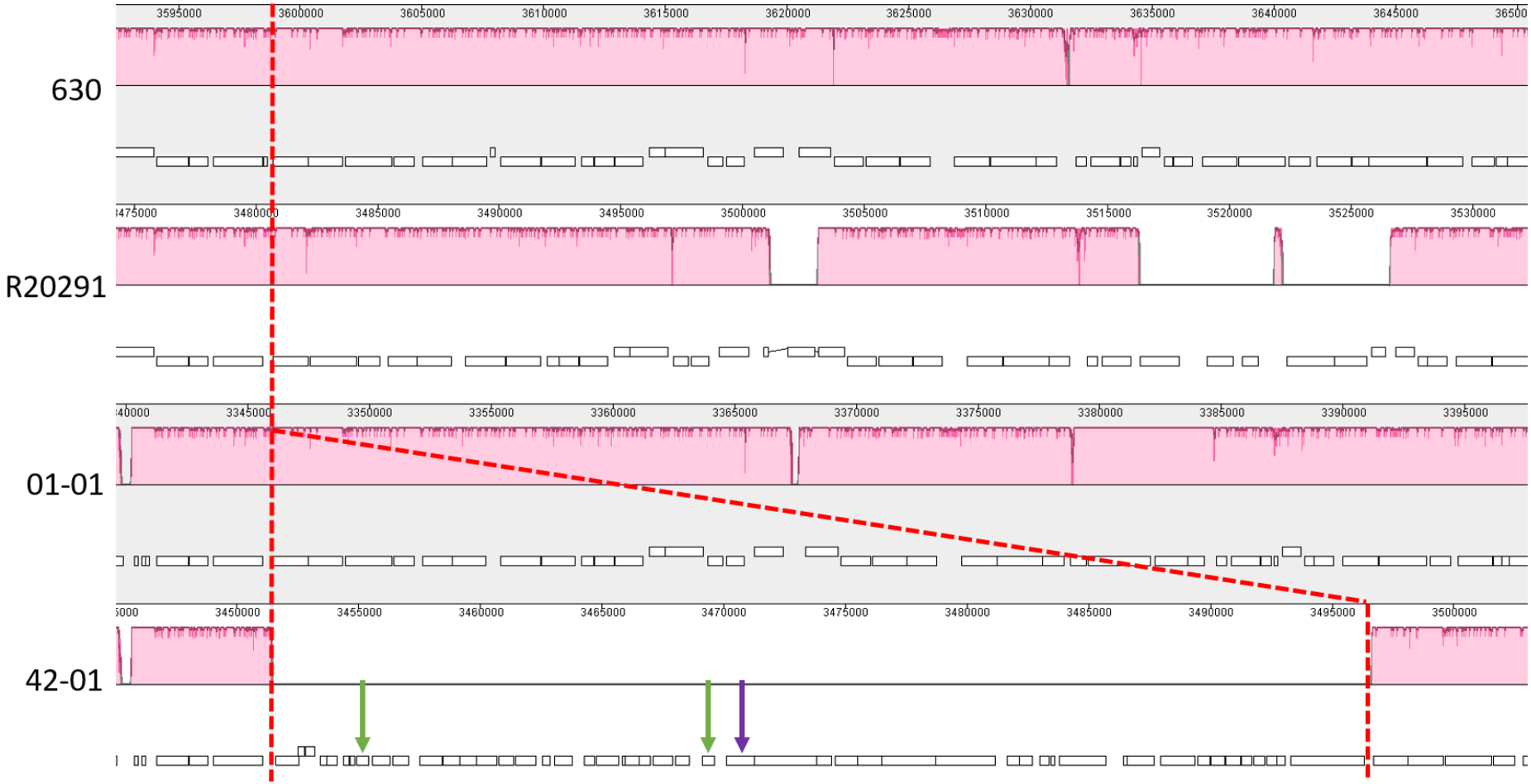


Figure 5-12: A Mauve alignment between *Clostridium difficile* strains 630, R20291, 01-01 (RT015 control strain) and 42-01 the index strain from the relapsing patient PT42. The red checked lines indicate a unique insertion of *Tn916*-like element, green arrows indicate teicoplanin resistance genes and purple arrow indicates tetracycline resistance gene at position 3,451,408 in 42-01 (43,592bp in length). The unique region is within the longer contig of 66,668bp suggesting that it is present within this region.

Patient 45, a male 95 years of age had one recurrent episode 87 days after initial diagnosis (Table 5-6). The index episode was clinically indicated as occurring in the community and the relapse episode as hospital acquired (Table 5-12). The index episode was treated with metronidazole which the *C. difficile* strain was determined to be sensitive to *in vitro*. During the relapse episode fidaxomicin was given as CDI therapy. There was no recorded third episode within the study period. *In vitro* MIC analysis of *C. difficile* strains showed sensitivity to vancomycin. At the time of the relapse episode the patient was receiving the proton pump inhibitory lansoprazole and had an elevated CRP (90 mg/L) which are both risk factors for relapse.

Table 5-12: Clinical information for patient 45 gathered at each sample collection.

Sample Date	**03/04/14**	**29/06/14**
Hospital Acquired CDI Community Acquired CDI	Community	Hospital
Co-infected	No	No
Initial CDI Treatment	Metronidazole	Fidaxomicin
Use of antibiotics other than CDI therapy during the treatment course	Unknown	No
Regular prescription of proton pump inhibitors during or after CDI therapy	Unknown	Lansoprazole
CRP >85mg/L	Unknown	90 mg/L
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	0.5 mg/L	0.5 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	0.5 mg/L	0.5 mg/L

Mapping of Illimina MiSeq reads from 45-02 to 45-01 revealed six SNPs unique to sample 45-02 (Table 5-5), one ambiguous and five NS. The ambiguous SNP is located 143bp downstream of *PT45_398* a hypothetical protein and 300bp downstream *PT45_399* a

partial integrase. The first NS SNP is in *PT45_2600* a 2,4-dienoyl-CoA reductase (NADPH) where a 389G>A a transition substitution results in Gly130Asp. This is deemed a tolerated substitution by PROVEAN. The second NS SNP is in *PT45_1933* a RNA methyltransferase, TrmA family were an 585G>T a transversion substitution results in a Leu195Phe. The third NS SNP is in *PT45_434* a putative aldo/keto reductase were an 585G>A a transition substitution results in Gly180Arg. The fourth NS SNP is in *PT45_3493* a hypothetical protein where a 12A>C a transversion substitution results in Arg4Ser. The sixth NS SNP is in *PT45_ 2351* a Stage V sporulation protein D where a 1442A>T a transversion substitution results in Asp481Val. Analyses of these substitutions were all determined as deleterious by PROVEAN.

Whole genome alignment of 45-01 only revealed one unique region ~31kb that contains a lantibiotic transport permease protein, lantibiotic transport ATP-binding protein and a putative *Tn916* transcriptional regulator.

Figure 5-13: A Mauve alignment between *Clostridium difficile* strains 630, R20291, 01-01 (RT015 control strain) and 45-01 the index strain from the relapsing patient PT45. The red checked lines indicate a unique insertion of *Tn916*-like element at position 409,131 in 45-01 (31,243bp in length), green box indicates lantibiotic related proteins. This insertion comprises of three contigs so misassembly cannot be discounted however the longest of the three contigs does span over the locally collinear block upstream. This insertion requires further investigation to confirm as unique to this strain.

Patent 85, a male 66 years of age suffered one recurrent episode 26 days after initial diagnosis (Table 5-6). The index episode for this patient was clinically determined to be hospital acquired and the relapse was classified as occurring in the community (Table 5-13). The index episode was treated with metronidazole which the *C. difficile* strain was determined to be sensitive to *in vitro*. There was no record of treatment was given as CDI therapy during the relapse episode. During the treatment period for the index episode both Tazocin and omeprazole were given, both count as risk factors for relapse. The patient also had a CRP level of 104 mg/L another risk factor for relapse. It is not known if any other treatments were being given during the relapse episode.

Table 5-13: Clinical information for patient 85 gathered at each sample collection.

Sample Date	**16/07/14**	**11/08/14**
Hospital Acquired CDI Community Acquired CDI	Hospital	Community
Co-infected	No	No
Initial CDI Treatment	Metronidazole	Unknown
Use of antibiotics other than CDI therapy during the treatment course	Tazocin	Unknown
Regular prescription of proton pump inhibitors during or after CDI therapy	Omeprazole	Unknown
CRP >85mg/L	104 mg/L	Unknown
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	1 mg/L	0.5 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	0.5 mg/L	1 mg/L

Genome analysis revealed no SNPs unique to sample 85-02 compared to the index strain (Table 5-5). Whole genome alignment revealed that all four strains were very similar except for a ~56kb region unique to 85-01 which contains hypothetical proteins, peptidoglycan bound protein, a putative *Tn916* regulator, iron acquisition yersiniabactin synthesis enzyme, various transport and transcriptional proteins (Figure 5-14). This sequence shows similarity to the *Tn916* region found in 19-01.

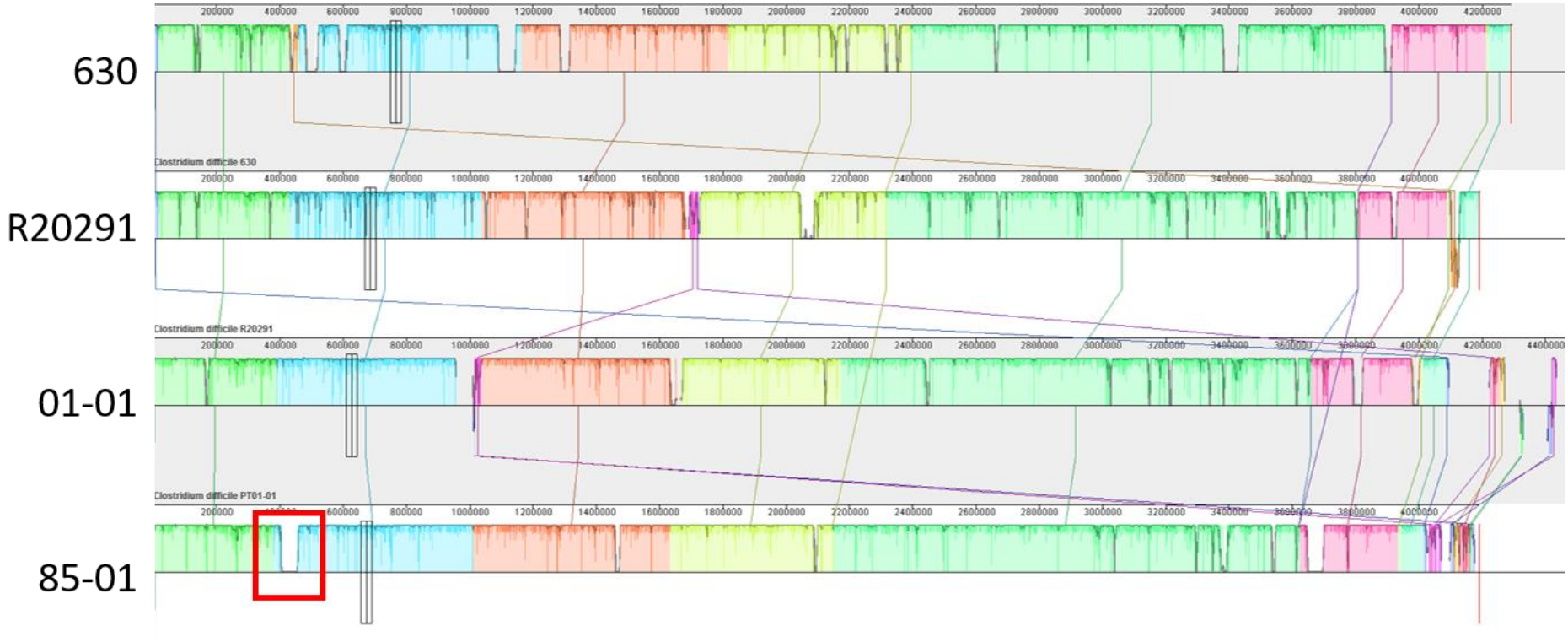


Figure 5-14: A Mauve alignment between *Clostridium difficile* strains 630, R2091, 01-01 (RT015 control strain) and 85-01 the index strain from the relapsing patient PT85 Red box indicates a unique insertion containing hypothetical proteins, peptidoglycan bound protein, a putative *Tn916* regulator, iron acquisition yersiniabactin synthesis enzyme, various transport and transcriptional proteins at position 401,317 in 85-01 (55,153bp in length. This insertion comprises of two contigs however both the upstream and the downstream regions span locally collinear blocks. Similar locally collinear blocks between strains are linked by matching coloured lines.

5.2.7.4 RT023

RT023 was the causative agent of CDI in two patients who suffered recurrent infection.

Patient 10, a female 69 years of age had one recurrent episode 83 days after initial diagnosis (Table 5-6). Both index and relapse episodes were classed clinically as hospital acquired (Table 5-14). The index sample was shown to have two distinct ribotypes; RT023 and RT062 both were sensitive to vancomycin and metronidazole *in vitro*. The relapse isolate was also sensitive to vancomycin and metronidazole *in vitro.* The patient had multiple risk factors for relapse such as taking meropenem, they also had an elevated CRP of 367 mg/L. During their relapse episode both vancomycin and metronidazole were given for CDI therapy. No other antibiotics or proton pump inhibitors were given at this time. There was no recorded third episode of relapse during the study period even though the patient had a raised CRP of 244 mg/L.

Genome analysis revealed no SNP unique to sample 10-02 (Table 5-5). Whole genome alignment did not reveal any unique genetic regions when compared to 630, R20291 and 06-01. There was evidence of the insertion previously described within the PaLoc in both RT023 strains[287].

Table 5-14: Clinical information for patient 10 gathered at each sample collection.

Sample Date	**10/12/13**	**03/03/14**
Hospital Acquired CDI Community Acquired CDI	Hospital	Hospital
Co-infected	Yes (RT062)	No
Initial CDI Treatment	Vancomycin	Vancomycin and Metronidazole
Use of antibiotics other than CDI therapy during the treatment course	Meropenem	None
Regular prescription of proton pump inhibitors during or after CDI therapy	No	None
CRP >85mg/L	367 mg/L	244 mg/L
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	RT023 – 1 mg/L RT062 – 0.25 mg/L	0.25 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	RT023 – 1 mg/L RT062 – 1 mg/L	1 mg/L

Patient 12, a male 85 years of age had one recurrent episode 38 days after initial infection (Table 5-6). Both index and relapse episodes were classed clinically as hospital acquired (Table 5-15). No other antibiotics were given during CDI therapy in either the index or the relapse cases. The proton pump inhibitor omeprazole was given during both CDI episodes. The patient did not have a CRP level indicative of relapse during the index episode of CDI but it was raised during the relapse at 171 mg/L. Both strains were shown to be sensitive to vancomycin and metronidazole *in vitro.*

No SNP were identified during analysis of the relapse strain (Table 5-5). As with strain 10-01, 12-01 showed high similarity to the control strain 06-01.

Table 5-15: Clinical information for patient 12 gathered at each sample collection.

Sample Date	13/12/13	20/01/14
Hospital Acquired CDI Community Acquired CDI	Hospital	Hospital
Co-infected	No	No
Initial CDI Treatment	Metronidazole	Vancomycin
Use of antibiotics other than CDI therapy during the treatment course	None	None
Regular prescription of proton pump inhibitors during or after CDI therapy	Omeprazole	Omeprazole
CRP >85 mg/L	28 mg/L	171 mg/L
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	0.25 mg/L	0.12 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	1 mg/L	1 mg/L

5.2.7.5 RT054

RT054 was the causative agent of CDI in one patient who suffered recurrent infection.

Patient 55, a male 77 years of age had one recurrent episode 30 days after initial diagnosis (Table 5-6). Both index and relapse episodes were classed clinically as hospital acquired (Table 5-16). Vancomycin was prescribed as the initial CDI therapy, the strains isolated at both index and relapse episodes were deemed sensitive to this antibiotic *in vitro* as was metronidazole. During the index episode, there was no recorded administration of either alternative antibiotics or proton pump inhibitors and the CRP level was unknown. All three of these factors were unknown during the relapse episode.

Table 5-16: Clinical information for patient 55 gathered at each sample collection.

Sample Date	**23/05/14**	**22/06/14**
Hospital Acquired CDI Community Acquired CDI	Hospital	Hospital
Co-infected	No	No
Initial CDI Treatment	Vancomycin	None
Use of antibiotics other than CDI therapy during the treatment course	None	Unknown
Regular prescription of proton pump inhibitors during or after CDI therapy	None	Unknown
CRP >85 mg/L	Unknown	Unknown
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	0.5 mg/L	1 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	0.25 mg/L	0.5 mg/L

No SNP were identified during analysis when comparing 55-02 to 55-01 (Table 5-5). Whole genome analysis between 630, R20291, 69-01 and 55-01 showed two unique regions. One with a *Tn916*-like element containing ParA/B, a recombinase, threonine dehydrogenase and hypothetical proteins (Figure 5-15). The second region contains phage related genes, hypothetical proteins, a lipoprotein, TrsK and TrsE-like genes and resembles ΦMMP03 according to PHASTER.

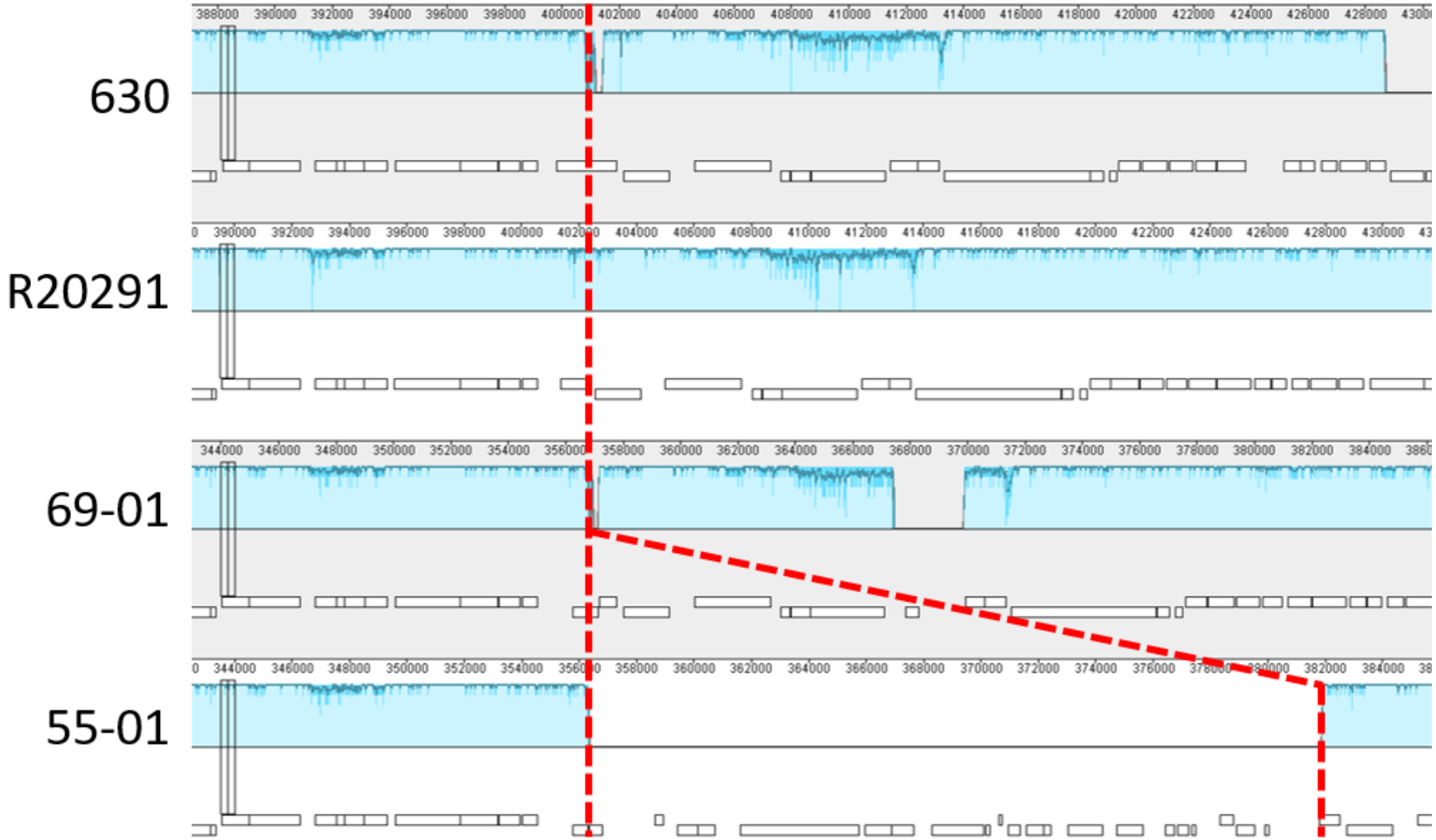


Figure 5-15: A Mauve alignment between *Clostridium difficile* strains 630, R20291, 69-01 (RT054 control strain) and 55-01 the index strain from the relapsing patient PT55. The red checked lines indicate a unique insertion of a *Tn916*-like element containing ParA/B, a recombinase, threonine dehydrogenase and hypothetical proteins at position 356,396 in 55-01 (25,415bp in length). This insertion comprises of nine contigs so misassembly cannot be discounted.

5.2.7.6 RT076

RT076 was the causative agent of CDI in one patient who suffered recurrent infection.

Patient 61, a male 69 years of age had five distinct episodes of CDI totalling a period of 183 days (Table 5-6). The first four episodes were classed clinically as hospital acquired with the fifth occurring in the community (Table 5-17). During the index episode, the patient was treated with both metronidazole and vancomycin which the isolates were shown to be sensitivity to *in vitro*. The patient was also prescribed omeprazole during this time and had a raised CRP of 228 mg/L, both are risk factors for relapse. During the second and third episodes, the patient was treated with fidaxomicin but this was unsuccessful in preventing future relapses. It was not recorded if the patient was receiving any other treatments at this time or if his CRP level was raised. For the fourth and fifth episodes, the patient received vancomycin alone as CDI therapy. At the fourth episode, the patient again had a raised CRP level of 169 mg/L. There was not a sixth relapse during the study period.

SNP analysis of all four recurrence strains revealed no differences between the strains (Table 5-5). Whole genome analysis of 61-01 compared to 630, R20291 and 57-01 (RT076 control strain) showed a single unique region containing a *Tn916*-like element, region that shares 90% coverage with QCD23M63[288] and contains a mobile element protein, hypothetical proteins, ABC transporter Vex, DltR, sensor histidine kinase and transcriptional regulators (Figure 5-16).

Table 5-17: Clinical information for patient 61 gathered at each sample collection.

Sample Date	03/06/14	30/06/14	19/07/14	07/09/14	03/12/14
Hospital Acquired CDI Community Acquired CDI	Hospital	Hospital	Hospital	Hospital	Community
Co-infected	No	No	No	No	No
Initial CDI Treatment	Metronidazole Vancomycin	Fidaxomicin	Fidaxomicin	Vancomycin	Vancomycin
Use of antibiotics other than CDI therapy during the treatment course	None	Unknown	Unknown	None	Unknown
Regular prescription of proton pump inhibitors during or after CDI therapy	Omeprazole	Unknown	Unknown	None	Unknown
CRP >85 mg/L	228 mg/L	Unknown	Unknown	169 mg/L	Unknown
Vancomycin minimum inhibitory concentration (≤2 mg/L = sensitive)	1 mg/L	0.5 mg/L	1 mg/L	1 mg/L	1 mg/L
Metronidazole minimum inhibitory concentration (≤2 mg/L = sensitive)	0.5 mg/L	1 mg/L	0.5 mg/L	0.5 mg/L	0.5 mg/L

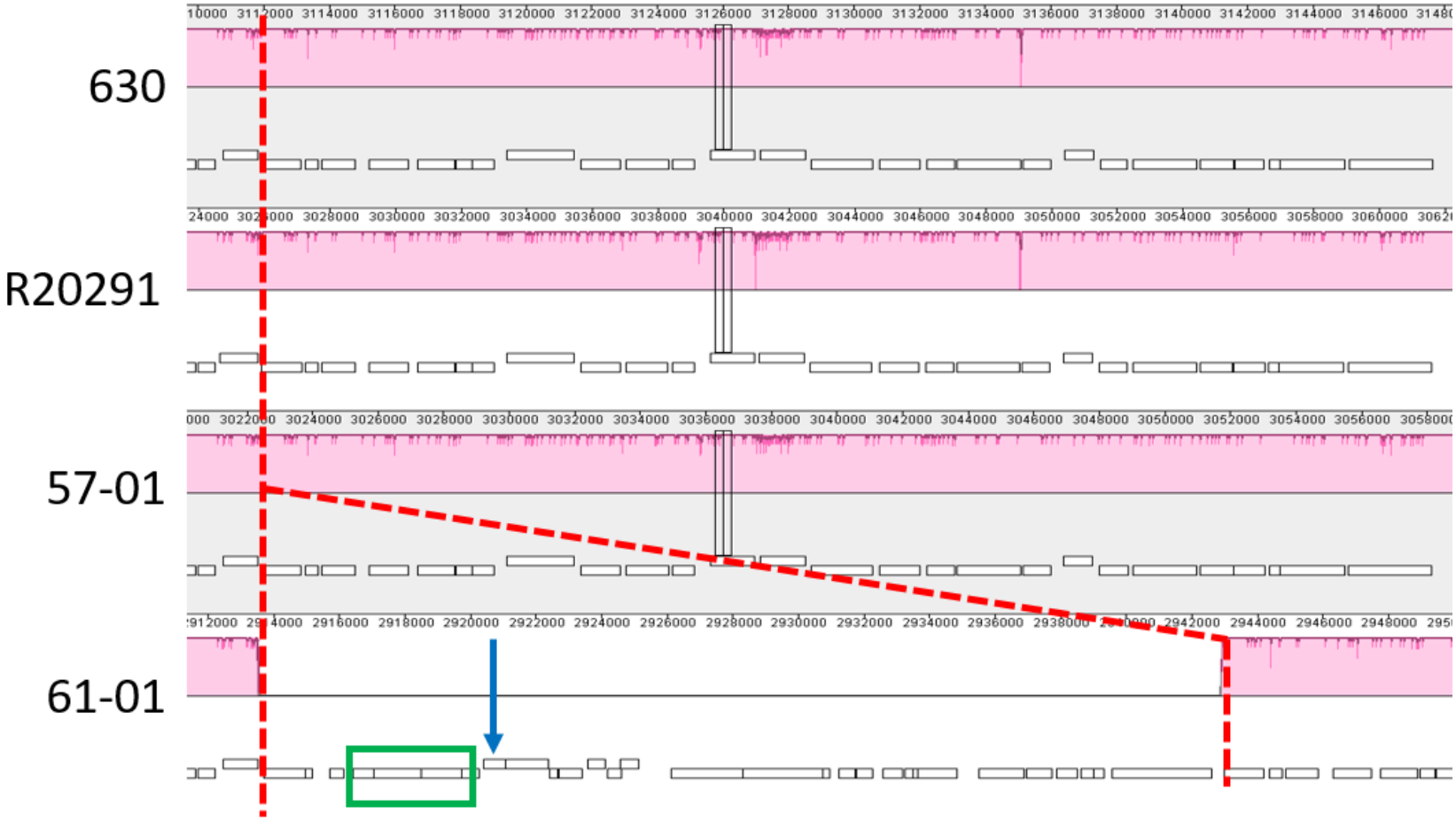


Figure 5-16: A Mauve alignment between *Clostridium difficile* strains 630, R20291, 57-01 (RT075 control strain) and 61-01 the index strain from the relapsing patient PT61. The red checked lines indicate a unique insertion of a *Tn916*-like element containing a region that shares 90% coverage with QCD23M63[288] and contains a mobile element protein, hypothetical proteins, ABC transporter Vex, DltR, sensor histidine kinase and transcriptional regulators at position 2,913,503 in 61-01 (29,363bp in length). Green box indicates Vex related genes and a blue arrow indicates the DltR gene. This insertion comprises of two contigs however both the upstream and the downstream regions span locally collinear blocks.

5.3 Discussion

The aim of this study was to identify if risk factors, co-infection and / or host evolution of strains are associated with recurrent CDI. To achieve this, we studied a cohort of patients from Nottingham University Hospitals NHS Trust between November 2013 and November 2014 whom were suffering from their first episode of CDI within the past six months.

There are many studies investigating the isolation of *C. difficile* from both stool[277–280] and the environment[281] however, they often rely on expensive commercially made chromogenic media which is often not affordable in the research setting. As one of our interests was the rate of co-infection and its effect on recurrent CDI we wanted to develop a cost effective and robust method of *C. difficile* isolation of multiple colonies in small stool samples.

We first looked at a broth enrichment step which revealed that CCFB could recover spores from faecal samples of hamsters infected with either CD630 or R20291. CCMB-TAL was only able to recover spores from faeces containing spores from R20291 and this was after an additional 24 h incubation, when compared to CCFB, in all cases. Whereas spores from both R20291 and CD630 were recoverable in CCFB after an overnight incubation. These findings suggest that fructose is a preferable carbon source over mannitol. CCMB-TAL had an improved recovery rate over CCFB, which appears to correlate with previous findings where lysozyme increased the recovery rate of *C. difficile*[283]. We assessed if the addition of lysozyme would enhance recovery in CCFB when tested against the faecal samples previously used. In our case, there was no observed difference. However, this sample size was small and may not be reliable. CCFB was the chosen broth for future studies as sensitivity rather than recovery is more important in this case.

The next stage of the *C. difficile* culture comparison study was to assess which of four solid media (CCFA, CCEY, ChromID C. difficile and TSA (5% sheep blood)) exhibited superior

performance. This was assessed in two steps, firstly with a broth enrichment step followed by direct plating onto the solid media. Secondly, plating stool supernatants directly onto the solid media. Our findings were that only broths containing 10^4 spore/faecal mixtures gave a positive reaction when a broth enrichment step was utilised. A recently published study stated that up to 10 CFU/mL *C. difficile* can be recovered after broth enrichment using CCMB-TAL[289], although not directly comparable as different broths were utilised, our findings are contradictory. It appears that Hink *et al*. may not have taken into account the number of spores within the spiked stool sample. The medium used for calculating the number of CFU/mL used for spiking did not contain germinant therefore only vegetative cells would be enumerated. However, the medium used in the experimental procedure contains taurocholate allowing spores to germinate and altering the CFU/mL in each sample by an undetermined amount. Therefore, the 10 CFU/mL value may be overestimated.

The direct plating of stools onto the four chosen media showed that CCEY was the most sensitive being able to recover up to 10^2 spores but ChromID C. difficile had the best recovery rate. In this study sensitivity is more important than recovery rate for recovering *C. difficile* from stool samples, especially in samples from patients where antibiotic treatment for CDI may have commenced. Consequently, we suggest a method of direct plating onto CCEY without a broth enrichment step. Also by removing the broth enrichment step we eliminate the risk of selecting out a dominant ribotype through growth competition.

Lastly we looked at using Neutral Red as an indicator to aid identification of growth of *C. difficile* on solid media. We incorporated 1% Neutral Red into the recipe for CCEY and compared it to ChromID C. difficile. For this we wanted to use clinical samples as even though the addition of cycloserine and cefoxitin is used to inhibit growth of faecal flora

there can be breakthrough of other organisms. It is clear that the black pigmentation produced by ChromID C. difficile makes the identification of *C. difficile* easy. It has been reported that not all *C. difficile* strains yield this pigment. One study found that 3% and 1% of samples did not produce the expected black pigment after 24 h and 48 h, respectively[280] and this should be taken into account. We found that no other faecal flora was cultivated from these samples and that the addition of the 1% neutral red did not give a necessary colour change to be used as an indicator of *C. difficile* growth.

Overall direct plating of faecal supernatants onto CCEY provided a 95.9% recovery rate in this study when applied to the clinical samples, comparable to that of the CDRN[252].

Selection of patients into the study was based on them not having had a previous episode of CDI within the previous six months and that we were able to isolate *C. difficile* from the stool samples provided to the clinical microbiology department as part of their standard of care treatment. The patient cohort was 59% female (59/100), 70% were 65 or over (70/100) and 15% were classified as community acquired cases of CDI (15/100). There are reports of CA-CDI being more severe and affecting patients with no risk factors[256,290,291]. Fifteen of the index strains were acquired in the community and of these five went on to relapse which agrees with other reports that CDI in no longer merely a nosocomial infection. Although there were no statistically significant findings when analysing the clinical information from the patient cohort there are some interesting results. In this cohort almost all (10/11) patients who relapsed were ≥65 years of age and no one who received fidaxomicin for treatment of their index episode of CDI went on to relapse (5/5).

There might have been a statistical significance for the risk factors i.e. additional antibiotics, use of PPI and raised CRP, if we were able to collect full clinical information. However, the number of unknowns in these categories may have affected these data.

The nature of this study was not one of epidemiological value due to a lack of patient location data, however, from the ribotype data gathered from the index episodes of CDI it can be seen that RT015 was the most prevalent within the hospital and community. According to CDRN data from December 2013 – December 2014 RT015 was the most prevalent throughout the East Midlands[252]. Other ribotypes that are seen as emergent throughout the UK are RT002, RT078, RT014, RT005 and RT020 which were seen in 12.21%, 7.63%, 6.87%, 6.11%, 5.34% and 3.82% of cases respectively. Ribotypes that were seen more frequently but not yet classed as emergent were RT023 and RT076 with 9.16% and 5.34% of cases respectively. There has been much debate on whether lineage determines host specific outcome but as increasing evidence becomes available it appears likely that biomarkers and clinical scores are more accurate methods of prediction[245,263,292–296].

In two patient samples, colonies originally thought to be *C. difficile* were revealed to be *C. butyricum* (PT98) and *C. glycolicum* (PT05). These findings are interesting in themselves as although *C. butyricum* are known as a common commensal of the human and animal gut, subtype E1 can produce botulinum neurotoxin (BoNT). This is due to a chromosomal insertion of an operon that has high similarity to BoNT carried by group II type E toxin-producing *C. botulinum*[297]. *C. butyricum* is also implicated in necrotizing enterocolitis in infants[297] and antibiotic associated diarrhoea[298]. *C. glycolicum* is also a common commensal of the human and animal gut although compared to *C. butyricum* there are few clinical cases demonstrating pathogenicity[299]. To clarify if these organisms contributed to the clinical presentations in these cases further analysis of strains and a clinical history would be required but was not performed.

Co-infection has recently been a topic of conversation within the *C. difficile* community. A few recent studies have looked into the rates of co-infection although they are mainly to

assess the levels for epidemiological reasons[271–273,300]. Only one study has looked at co-infection and its relationship with relapse[270], they found that 9% of co-infected patients relapsed, although the study number was small (25 relapse patients).

In our study cohort 13 out of 100 patients were found to be co-infected with two different ribotypes. This value may be under representative due to the relatively insensitive nature of ribotyping and that only 20 colonies were used for downstream analysis. The level of co-infection might have been higher if either more colonies were studies or another typing method was used i.e. MLVA or whole genome sequencing[271]. This is supported by the number of SNPs seen in this study between isolates of the same ribotype (Table 5-5). Two patients who suffered relapse episodes were also co-infected in either their index sample (PT10) or their relapse sample (PT39). The method of screening used in our study is likely not sensitive enough to conclusively state that both ribotypes were not present in all samples for each patient. Mixed colonisation may be due to two or more separate transmission events or possibly from a carrier whom is also colonised with more than one lineage which are transmitted simultaneously[301]. This has been described in patients colonised with *Staphylococcus aureus* using VNTR[302] and either scenario cannot be ruled out in these cases. In this study co-infection was not statistically significant as a predictor of recurrent CDI although this might change if a more sensitive method of typing and / or more colonies were analysed from initial culture. Therefore, this should be considered in future projects which employ more sensitive typing methods.

Genome evolution can be a result of point mutation and is usually due to a beneficial modification such as adapting to environmental changes or avoiding host immune system. These mutations rates have been described in *Helicobacter pylori* as high as 30 mutations per year per genome[303]. In *C. difficile* this rate has been estimated to be ~2 mutations per year per genome[304]. These differences may be attributed to genome length and efficiency

of mismatch repair systems[301]. On a larger scale horizontal gene transfer (HGT) speeds up diversification through the acquisition of multiple genes at one time. Genomic plasticity is a widely accepted concept within the *C. difficile* genetics with examples of transposons, bacteriophages and plasmids being described[154,207,215,288,305]. Phase variation, although not directly linked to genome evolution may be responsible for adaptation to a specific environment i.e. repression of surface proteins to avoid detection by the host immune system in *Neisseria meningitides*[306]. We used whole genome and SNP analysis to see if any *C. difficile* isolates from our patient cohort that suffered relapse had evolved during infection or if any HGT had occurred that provided a fitness benefit.

There have been a small number of studies using whole genome sequencing to study recurrence episodes in CDI. However, these were used to assess transmission of strains between patients and differentiation between patients[34,261,269] and only one looked further in to the genetic locations of SNPs to speculate about increased fitness benefit[261].

SNPs were analysed in strains isolated from the 11 patients suffering recurrence episodes to assess the level of within host evolution. Unique genomic regions were also assessed, when compared to CD630, R20291 and the control strain of the same ribotype, to identify genes that may provide a fitness benefit. A common definition of 8 weeks (56 days) is generally used to associate a recurrence episode[261]. In two patients PT85 (RT015) and PT12 (RT023) there were no SNPs detected between the index and the recurrence suggesting relapse in both cases. The length of time between index and relapse was 26 and 38 days respectively. Whole genome comparison of RT015 from PT85 showed a ~15kb region containing a number of genes but most interestingly a putative pilin which although not demonstrated experimentally, could provide a fitness benefit for adherence within the gut[307]. Whole genome comparison of RT023 from PT12 revealed a unique phage region with similarity to ΦMMP02 although the genes within this region do not appear to confer

a fitness benefit. SNP analysis of RT001 from PT39 had four tolerated SNPs and two unique phage regions with similarity to ΦMMP03 and ΦCD211. As with RT023 from PT12 these regions do not appear to provide a fitness benefit.

There were three patients who suffered one relapse more than 80 days after their index episode. *C. difficile* ribotypes from PT19 and PT45 were RT015 and the ribotype from PT10 was RT023. Both the *C. difficile* strains isolated from PT19 and PT10 had one ambiguous SNP and no unique genetic regions that appear to give a fitness benefit, these data do not provide an insight as to why the strains were able to persist in these patients for extended periods of time. In the recurrence strain for PT45 there were six SNPs in total, one ambiguous, one tolerated substitution and four deleterious substitutions. These were located in a RNA methyltransferase, a putative aldo/keto reductase, a hypothetical protein and a stage V sporulation protein. It would only be speculation to suggest that these SNPs had an effect on recurrence or length of carriage. However, the most interesting fact about these three patients is that they appear to have been colonised with the same strain of *C. difficile* for many weeks. Previously, Lawley *et al*. demonstrated that in immunocompetent mice colonisation of *C. difficile* can occur and in this state a low level of spores are shed into the environment, and after treatment with antibiotics these mice became super shedders[93]. Although only supposition it would be interesting to look into these patients in more detail to see if any transmission events occurred during this time to other patients within this study. We have grouped PT45 in this category, however, studies suggest that 2 – 10 SNP differences between genomes should be classed as intermediate and therefor indistinguishable between the same strain and a highly related strain[308]. Meaning that this patient could have acquired a new strain.

PT96s (RT001) length of carriage lasted 94 days with four distinct episodes of CDI. There was one ambiguous SNP present in all four relapse samples and a tolerated SNP in a

hypoxanthine-guanine phosphoribosyl transferase found in PT96-03, which likely demonstrates heterogeneity of the population. PT42 (RT015) had a 26-day length of carriage and no SNP were present in the recurrence genome. The interesting observation in these two strains is there are multi antimicrobial extrusion protein MATE family of MDR efflux pump genes found in both that are unique to each strain. Again one can only hypothesise that these genes may have a role in an increasing fitness benefit, but these genes are known to transport antibiotics out of bacterial cells in other organisms[309] and although these isolates were found to be vancomycin and metronidazole sensitive *in vitro* this may not be the case *in vivo*. There have also been studies identifying phase variation in both Staphylococci[310] and *Neisseria gonorrhoae*[311] where phase variation has been attributed to resistance to antimicrobials when under selective pressure.

PT59 (RT076) length of carriage lasted 183 days with five distinct episodes of CDI. No SNPs were detected in any of the relapse isolates suggesting the same strain of *C. difficile* was present throughout the whole period. Whole genome analysis showed a *Tn6073*-like region which contains three Vex-like ABC transporter genes. These genes along with a Pep_{27} (cell – cell signalling peptide) and VncR/S (subunits of a two-component signal transduction system) in have been demonstrated in *Streptococcus pneumoniae* to infer vancomycin tolerance[312]. This patient was initially treated with vancomycin and metronidazole, then with fidaxomicin during their second and third episodes and finally with vancomycin during their fourth and fifth episodes. So these genes alone would not likely explain the multiple recurrence episodes, it is however a curious correlation.

Lastly PT05 (RT005) is the only patient identified as a re-infection through whole genome sequencing. Even though both strains were of the same ribotype there were 61 SNPs identified between the index and the recurrence strain which is deemed an indicator of re-infection[34]. This is an interesting finding which suggests that low sensitive typing methods

are most likely under estimating the number of re-infections and labelling them relapse episodes[34,261].

There are some limitations to this study; the patient reference sequences may be deemed low quality due to the nature of *de novo* assembly of short reads. This would be improved by using a third-generation sequencing technology such as SMRT or MinION sequencing. The method used in this study may not have given enough resolution to identify the locations of the unique regions and these may have been "dumped" at the end of the sequence during the move contigs step when using Mauve. Also, we are not able to determine if the unique genes found in the whole genome alignment are not present in all the strains found in the non-relapse patients sharing the same ribotype as these were not analysed. Using ribotyping as a method of assessing co-infection might have underestimated the level of co-infection within our study cohort. It is not possible to state if the patients with apparent long-term carriage were indeed colonised for the whole period or if they re-acquired the same strain from an environmental source at a later date. This could be resolved by collecting follow up stool samples after successful treatment in future studies.

Although there are some limitations to this study we have identified a cost effective and sensitive method of isolating *C. difficile* from stool samples. RT015 was the dominant circulating ribotype at the time but these data suggest that they may not be highly related at all and therefore questions are raised on how epidemiologically relevant this typing method is. The sequencing data has also revealed that many of the relapse patients were potentially colonised for longer periods of time than the standard 8 weeks of classification for a relapse[261]. This suggests two things; some of the *C. difficile* strains analysed in this study are not under short term selection pressure and these data suggests the commonly used classification of relapse should be redefined. We also identified a number of genes

and SNPs that appear unique to the relapse strains which may infer a fitness benefit. However, this would require verifying experimentally before any conclusions could be made.

5.4 Key Outcomes

- A cost effective and sensitive method for the isolation of *Clostridium difficile* from stool samples.
- CA-CDI is a significant risk factor for recurrent episodes of infection.
- Co-infection was not a risk factor for recurrent CDI in this study cohort.
- Carriage of the same *C. difficile* strain was seen for periods of time longer than the standard 8 weeks of classification for a relapse.
- A number of genes and SNPs were identified as potential candidates for increased fitness benefit within relapsing strains.

5.5 Future Work

- Collect stool samples from non-relapse cases and assess the length of carriage over time.
- Use whole genome sequencing on the non-relapse strains to confirm the uniqueness of genes found in strains from relapse cases of the same ribotype.
- Use a gut model to assess phase variation of strains while under selective pressure.

Chapter Six Concluding Remarks

The aim of the first part of this thesis was to try and improve our knowledge of TcdA and TcdB regulation within *C. difficile*. The role of TcdR in this process has been studied previously[75,128] but our group are the first to create a clean deletion of this gene. Earlier work as relied on fusion reporter systems and insertional gene inactivation, but we hoped that our study would more reflect the natural environment. We were able to show that NM-R20291Δ*tcdR* had significantly less TcdA and TcdB in supernatants at 12, 24 and 48 h when compared to the wild-type and complemented strain. These data further confirm the previous findings of Mani *et al*. in demonstrating that TcdR is a positive regulator of TcdA and TcdB transcription.

It has been shown previously that over 50% of the CDS in *C. difficile* 630 have unknown functions[175]. This hinders the understanding of regulatory networks when using reverse genetic analysis. The use of forward genetic analysis allows us to randomly look for altered phenotypes and then identify candidate genes that are involved in these networks either directly or indirectly. This is reliant on a robust and rapid screening method. To overcome the laborious nature of measuring toxin in supernatants of *C. difficile* cultures a "reporter" strain was developed in which *tcdB* was replaced with *licB* a lichenase gene. The lichenase breaks down lichenan which could be utilised in a plate assay and was believed to be comparative to the amount of toxin produced. A random mutant library was produced using the *mariner*-transposon system which revealed a number of interesting phenotypes upon screening. The most striking resulted from a double insertion where two genes were interrupted, R20291::*2908* and R20291::*jag*. R20291::2908 is thought to be involved in a programmed cell-death system and / or Type I restriction modification systems[177] and R20291::*jag* in sporulation[178]. ClosTron mutagenesis was employed to disrupt each of these genes individually to confirm the phenotypes seen in the transposon mutants. Initially measurements of toxin in supernatants revealed that NM-R20291::2908(*ermB*) had reduced

concentrations when compared to wild-type. However, these finding were not reproducible during complementation studies which was attributed to user error. Current literature would also suggest that a gene involved in sporulation would more likely alter toxin production when disrupted. Unfortunately, time constraints did not allow resolution of this issue.

Even though through using the plate screening method we were able to identify altered phenotypes we wanted to produce a semi-quantitative method. This involved using culture supernatants containing lichenase and liquid media containing a known concentration of lichenan. This system had been used previously in *C. botulinum* within our group but we found it unreproducible in *C. difficile*. We surmised that this is due to differences in both growth and toxin regulation between the two strains.

The second aim of this study was to produce a complete genome map for the high toxin producing strain VPI 10463, to investigate possible genetic factors that may support this phenotype. Using PacBio RSII and Illumina MiSeq technologies we were able to produce a complete circular genome for VPI 10463. Analysis of the genome revealed a large inversion that occurs in two lineage specific regions, the largest described within *C. difficile* strains to date. This was confirmed through the mapping of the long PacBio reads back to the newly created genome map for VPI 10463 which showed sufficient coverage over the inversion sites to rule out assembly error. Analysis of VPI 10463s orthologues revealed a candidate gene for the increased toxin production seen in this strain. A putative second holin like gene which is not present in either CD630 or R20291 was hypothesised to allow more TcdA and TcdB out of the cell. We aimed to test this experimentally using ClosTron mutagenesis but were unable to get transposition of the intron into the target gene.

A number of studies have shown there to be a link between toxin regulation and the inactivation of numerous flagella related genes[161,162]. Analysis of the flagella operons in VPI 10463 revealed a nonsense mutation in *fliJ* resulting in a truncated protein. This mutation has resulted in a monotrichous phenotype and a reduced zone of motility. The effect of this mutation hasn't been studied experimentally here but it may play a part in the increased toxin production seen in this strain.

One of the issues hindering basic research of *C. difficile* is the low level of DNA transfer achievable within clinically relevant strains. This means that most published studies where genetic manipulation occurs are focused on derivatives of CD630. Recently there are examples of how making inferences from one strain to encompass the whole genus may result in inaccuracies[313]. This being said we wanted to improve DNA transfer into VPI 10463 and potentially other strains, so a broader knowledge can be gained from genetic studies. Strains VPI 10463 and R20291 were found to have two and one type I restriction modification systems respectively. ClosTron mutants were created to interrupt one of the restriction subunits (*hsdR*) in each strain. We demonstrated that this insertional inactivation increased conjugation frequency in both strains.

The last body of work presented here involved a study of patients within the University of Nottingham Hospitals NHS Trust. The first part of this study was to develop an inexpensive and reliable method to isolate *C. difficile* from stool samples. It was determined that direct plating of faecal supernatants onto CCEY after heat shock was sufficient to isolate *C. difficile* in 95.9% of cases.

The overall aim of this study was to identify if risk factors, co-infection and / or host evolution of strains are associated with recurrent CDI. The study included 100 patients suffering an episode of CDI if they had not had an episode within the previous six months.

Both co-infection and recurrence was found to be at 11% within this cohort although co-infection was shown not a risk factor for recurrence. None of the clinical information gathered for these patients were found to be statistically significant, although 90.9% of patients who relapsed were ≥65 years of age and if the patient received fidaxomicin for treatment of their index episode they did not suffer recurrence.

Ribotype analysis of the infecting strains showed that the pattern observed in this patient cohort followed the trend reported by the CDRN for the East Midlands. Interestingly, two isolates initially believed to be *C. difficile* were later found to be *C. butyricum* (PT98) and *C. glycolicum* (PT05). Their involvement in the infection of these patients was not experimentally demonstrated here but has been shown to be pathogenic previously[297,299].

Whole genome sequencing of the isolates recovered from patients with recurrent CDI showed potential SNPs and unique genomic regions that could infer a fitness benefit in these strains. Unique genes include a unique putative pilin which could increase adherence of this strain in the gut[307], multi antimicrobial extrusion protein MATE family of MDR efflux pump genes which have been shown to increase resistance to antimicrobials when under selective pressure[309] and Vex-like ABC transporter genes shown to infer vancomycin tolerance in *Streptococcus pneumoniae*[312].

Two patients had more than one recurrence episode, PT96 was 94 days with four episodes and PT59 was 183 days with five episodes and were found to have a MATE family of MDR efflux pump and a Vex-like ABC transporter gene respectively.

The most common definition of time to recurrences is eight weeks (56 days), however in this study three patients relapsed strain >80 days after their index episode two of these patients with a genetically indistinct (one SNP difference) strain. This suggests that there may be colonisation within the gut for much longer than we originally believed.

Only one patient who had a recurrence episode was shown to have re-infection through a different strain. The index strain from PT05 was shown to be RT005, this patients recurrence strain was also shown to be RT005 however, there were 61 SNP identified between the two strains. This number of SNP has been defined as indicative of a re-infection[34]. These findings show that routine typing methods may falsely define a recurrence episode as relapse and could mistakenly predict transmission events.

1. Collins MD, Lawson PA, Willems A, Cordoba JJ, Fernandez-Garayzabal J, Garcia P, et al. The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. International Journal of Systematic Bacteriology. 1994;44(4):812–26.

2. Carroll KC, Bartlett JG. Biology of *Clostridium difficile*: implications for epidemiology and diagnosis. Annual Review of Microbiology. 2011;65:501–21.

3. Tracy BP, Jones SW, Fast AG, Indurthi DC, Papoutsakis ET. Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. Current Opinion in Biotechnology. 2012;23(3):364–81.

4. Wei MQ, Mengesha A, Good D, Anné J. Bacterial targeted tumour therapy-dawn of a new era. Cancer Letters. 2008;259(1):16–27.

5. Brynestad S, Granum PE. *Clostridium perfringens* and foodborne infections. International Journal of Medical Microbiology. 2002;74(3):195–202.

6. Brazier JS. *Clostridium difficile*: from obscurity to superbug. British Journal of Biomedical Science. 2008;65(1):39–44.

7. Jacobs A, Barnard K, Fishel R, Gradon JD. Extracolonic manifestations of *Clostridium difficile* infections. Presentation of 2 cases and review of the literature. Medicine. 2001;80(2):88–101.

8. Hall IC, O'Toole E. Intestinal flora in new-born infants: with a description of a new pathogenic anaerobe, *Bacillus difficilis*. Archives of Pediatrics & Adolescent Medicine. 1935;49(2):390.

9. Larson HE, Price AB, Honour P, Borriello SP. *Clostridium difficile* and the aetiology of pseudomembranous colitis. Lancet. 1978;1(8073):1063–6.

10. Johnson S, Samore MH, Farrow KA, Killgore GE, Tenover FC, Lyras D, et al. Epidemics of diarrhea caused by a clindamycin-resistant strain of *Clostridium difficile* in four hospitals. The New England Journal of Medicine. 1999;341(22):1645–51.

11. Loo VG, Bourgault AM, Poirier L, Lamothe F, Michaud S, Turgeon N, et al. Host and pathogen factors for *Clostridium difficile* infection and colonization. The New England Journal of Medicine. 2011;365(18):1693–703.

12. Wiegand P, Nathwani D, Wilcox M, 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;81(1):1–14.

13. Stevens V, Dumyati G, Fine LS, Fisher SG, van Wijngaarden E. Cumulative antibiotic exposures over time and the risk of *Clostridium difficile* infection. Clinical Infectious Diseases. 2011;53(1):42–8.

14. Blondeau JM. What have we learned about antimicrobial use and the risks for *Clostridium difficile*-associated diarrhoea? Journal of Antimicrobial Chemotherapy. 2009;63(2):238–42.

15. Pépin J, Saheb N, Coulombe MA, Alary ME, Corriveau MP, Authier S, et al. Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-associated diarrhea: a cohort study during an epidemic in Quebec. Clinical Infectious Diseases. 2005;41(9):1254–60.

16. Freeman J, Bauer MP, Baines SD, Corver J, Fawley WN, Goorhuis B, et al. The changing epidemiology of *Clostridium difficile* infections. Clinical Microbiology Reviews. 2010;23(3):529–49.

17. Hensgens MP, Goorhuis A, Dekkers OM, Kuijper EJ. Time interval of increased risk for *Clostridium difficile* infection after exposure to antibiotics. Journal of Antimicrobial Chemotherapy. 2012;67(3):742–8.

18. Garg S, Mirza YR, Girotra M, Kumar V, Yoselevitz S, Segon A, et al. Epidemiology of *Clostridium difficile*-associated disease (CDAD): A shift from hospital-acquired infection to long-term care facility-based infection. Digestive Diseases and Sciences. 2013;58(12):3407–12.

19. Bavishi C, Dupont HL. Systematic review: the use of proton pump inhibitors and increased susceptibility to enteric infection. Alimentary Pharmacology and Therapeutics. 2011;34:1269–81.

20. McCollum DL, Rodriguez JM. Detection, treatment and prevention of *Clostridium difficile* infection. Clinical Gastroenterology and Hepatology. 2012;10(6):581–92.

21. Riggs MM, Sethi AK, Zabarsky TF, Eckstein EC, Jump RLP, Donskey CJ. Asymptomatic Carriers Are a Potential Source for Transmission of Epidemic and Nonepidemic *Clostridium difficile* Strains among Long-Term Care Facility Residents. Clinical Infectious Diseases. 2007;45(8):992–8.

22. Heeg D, Burns DA, Cartman ST, Minton NP. Spores of *Clostridium difficile* clinical isolates display a diverse germination response to bile salts. PLoS ONE. 2012;7(2):e32381.

23. Nerandzic MM, Pultz MJ, Donskey CJ. Examination of potential mechanisms to explain the association between proton pump inhibitors and *Clostridium difficile* infection. Antimicrobial Agents and Chemotherapy. 2009;53(10):4133–7.

24. Wilson KH, Sheagren JN, Freter R. Population dynamics of ingested *Clostridium difficile* in the gastrointestinal tract of the Syrian hamster. Journal of Infectious Diseases. 1985;151(2):355–61.

25. Barbut F, Decré D, Lalande V, Burghoffer B, Noussair L, Gigandon A, et al. Clinical features of *Clostridium difficile*-associated diarrhoea due to binary toxin (actin-specific ADP-ribosyltransferase)-producing strains. Journal of Medical

Microbiology. 2005;54(2):181–5.

26. Department of Health. *Clostridium difficile* infection: How to deal with the problem. London; DoH; December 2008;

27. Cohen SH, Gerding DN, Johnson S, Kelly CP, Loo VG, McDonald LC, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). Infection Control and Hospital Epidemiology. 2010;31(5):431–55.

28. Johnson S. Recurrent *Clostridium difficile* infection: A review of risk factors, treatments, and outcomes. Journal of Infection. 2009;58(6):403–10.

29. Kelly CP. Can we identify patients at high risk of recurrent *Clostridium difficile* infection? Clinical Microbiology and Infection. 2012;18 (Suppl. 6):21–7.

30. Mattila E, Uusitalo-Seppälä R, Wuorela M, Lehtola L, Nurmi H, Ristikankare M, et al. Fecal transplantation, through colonoscopy, is effective therapy for recurrent *Clostridium difficile* infection. Gastroenterology. 2012;142(3):490–6.

31. Wilcox MH, Hawkey P, Patel B, Planche T, Stone S. Updated guidance on the management and treatment of *Clostridium difficile* infection. London, UK: Public Health England. 2013;

32. HICPAC. Recommendations for preventing the spread of vancomycin resistance. Recommendations of the Hospital Infection Control Practices Advisory Committee (HICPAC). MMWR Recommendation Report. 1995;44:1–13.

33. Gerding DN. Metronidazole for *Clostridium difficile*-associated disease: is it okay for Mom? Clinical Infectious Diseases. 2005;40(11):1598–600.

34. Eyre DW, Babakhani F, Griffiths D, Seddon J, Elias CDO, Gorbach SL, et al. Whole-genome sequencing demonstrates that fidaxomicin is superior to vancomycin for preventing reinfection and relapse of infection with *Clostridium difficile*. Journal of Infectious Diseases. 2014;209(9):1446–51.

35. Chilton C, Crowther G, Freeman J, Todhunter S, Nicholson S, Longshaw C, et al. Successful treatment of simulated *Clostridium difficile* infection in a human gut model by fidaxomicin first line and after vancomycin or metronidazole failure. Journal of Antimicrobial Chemotherapy. 2014;69(2):451–62.

36. Babakhani F, Bouillaut L, Sears P, Sims C, Gomez A, Sonenshein AL. Fidaxomicin inhibits toxin production in *Clostridium difficile*. Journal of Antimicrobial Chemotherapy. 2013;68(3):515–22.

37. Garey KW, Ghantoji SS, Shah DN, Habib M, Arora V, Jiang Z-D, et al. A randomized, double-blind, placebo-controlled pilot study to assess the ability of rifaximin to prevent recurrent diarrhoea in patients with *Clostridium difficile*

infection. Journal of Antimicrobial Chemotherapy. 2011;66(12):2850–5.

38. Johnson S, Schriever C, Galang M, Kelly CP, Gerding DN. Interruption of recurrent *Clostridium difficile*-associated diarrhea episodes by serial therapy with vancomycin and rifaximin. Clinical Infectious Diseases. 2007;44(6):846–8.

39. Carman RJ, Boone JH, Grover H, Wickham KN, Chen L. In vivo selection of rifamycin-resistant *Clostridium difficile* during rifaximin therapy. Antimicrobial Agents and Chemotherapy. 2012;56(11):6019–20.

40. O'Connor JR, Galang MA, Sambol SP, Hecht DW, Vedantam G, Gerding DN, et al. Rifampin and rifaximin resistance in clinical isolates of *Clostridium difficile*. Antimicrobial Agents and Chemotherapy. 2008;52(8):2813–7.

41. Public Health England. Infection Report. Health Protection Report: Weekly Report. London; PHE; June 2015; 2015.

42. Health Protection Agency. *Clostridium difficile* Ribotyping Network (CDRN) for England and Nothern Ireland annual report 2010/11. London; HPA; 2011; 2011.

43. Department of Health. Changes to the mandatory healthcare associated infection surveillance system for *Clostridium difficile* associated diarrhoea from April 2007. London: DoH; April 2007.;

44. Killgore G, Thompson A, Johnson S, Brazier J, Kuijper E, Pepin J, et al. Comparison of seven techniques for typing international epidemic strains of *Clostridium difficile*: restriction endonuclease analysis, pulsed-field gel electrophoresis, PCR-ribotyping, multilocus sequence typing, multilocus variable-number tandem-repeat analysis, amplified fragment length polymorphism, and surface layer protein A gene sequence typing. Journal of Clinical Microbiology. 2008;46(2):431–7.

45. Health Protection Agency. *Clostridium difficile* Ribotyping Network (CDRN) for England and Nothern Ireland 2011-13 Report. London: HPA; 2014; 2014.

46. Wei HL, Kao CW, Wei SH, Tzen JTC, Chiou CS. Comparison of PCR ribotyping and multilocus variable-number tandem-repeat analysis (MLVA) for improved detection of *Clostridium difficile*. BMC Microbiology. 2011;1(11):217.

47. Dawson LF, Valiente E, Wren BW. *Clostridium difficile*-a continually evolving and problematic pathogen. Infection, Genetics and Evolution. 2009;9(6):1410–7.

48. Stubbs SL, Brazier JS, O'Neill GL, Duerden BI. PCR targeted to the 16S-23S rRNA gene intergenic spacer region of *Clostridium difficile* and construction of a library consisting of 116 different PCR ribotypes. Journal of Clinical Microbiology. 1999;37(2):461–3.

49. Rupnik M, Avesani V, Janc M, von Eichel-Streiber C, Delmée M. A Novel Toxinotyping Scheme and Correlation of Toxinotypes with Serogroups of *Clostridium difficile* Isolates. Journal of Clinical Microbiology. 1998;36(8):2240–

7.

50. Rupnik M, Janezic S. An Update on *Clostridium difficile* Toxinotyping. Journal of Clinical Microbiology. 2016;54(1):13–8.

51. Elliott B, Squire MM, Thean S, Chang BJ, Brazier JS, Rupnik M, et al. New types of toxin A-negative, toxin B-positive strains among clinical isolates of *Clostridium difficile* in Australia. Journal of Medical Microbiology. 2011;60(8):1108–11.

52. Huber CA, Foster NF, Riley TV, Paterson DL. Challenges for standardization of *Clostridium difficile* typing methods. Journal of Clinical Microbiology. Am Soc Microbiol; 2013;51(9):2810–4.

53. Li W, Raoult D, Fournier P-E. Bacterial strain typing in the genomic era. FEMS Microbiology Reviews. Wiley Online Library; 2009;33(5):892–916.

54. Gerding DN, Meyer T, Lee C, Cohen SH, Murthy UK, Poirier A, et al. Administration of spores of nontoxigenic *Clostridium difficile* strain M3 for prevention of recurrent C. difficile infection: a randomized clinical trial. Journal of the American Medical Association. 2015;313(17):1719–27.

55. 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):243–50.

56. Pruitt RN, Lacy DB. Toward a structural understanding of *Clostridium difficile* toxins A and B. Frontiers in Cellular and Infection Microbiology. 2012;2:28.

57. 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):375–82.

58. Greco A, Ho JG, Lin S-J, Palcic MM, Rupnik M, Ng KK. Carbohydrate recognition by *Clostridium difficile* toxin A. Nature Structural & Molecular Biology. 2006;13(5):460–1.

59. Tucker KD, Wilkins TD. Toxin A of *Clostridium difficile* binds to the human carbohydrate antigens I, X, and Y. Infection and Immunity. 1991;59(1):73–8.

60. Yuan P, Zhang H, Cai C, Zhu S, Zhou Y, Yang X, et al. Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for *Clostridium difficile* toxin B. Cell Research. 2015;25(2):157–68.

61. Pruitt RN, Chambers MG, Ng KK-S, Ohi MD, Lacy DB. Structural organization of the functional domains of *Clostridium difficile* toxins A and B. Proceedings of the National Academy of Sciences. 2010;107(30):13467–72.

62. Pruitt RN, Chagot B, Cover M, Chazin WJ, Spiller B, Lacy DB. Structure-function analysis of inositol hexakisphosphate-induced autoprocessing in *Clostridium*

difficile toxin A. Journal of Biological Chemistry. 2009;284(33):21934–40.

63. Lyerly DM, Saum KE, MacDonald DK, Wilkins TD. Effects of *Clostridium difficile* toxins given intragastrically to animals. Infection and Immunity. 1985;47(2):349–52.

64. Voth DE, Ballard JD. *Clostridium difficile* toxins: mechanism of action and role in disease. Clinical Microbiology Reviews. 2005;18(2):247–63.

65. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, et al. Toxin B is essential for virulence of *Clostridium difficile*. Nature. 2009;458(7242):1176–9.

66. Kuehne SA, Cartman ST, Minton NP. Both, toxin A and toxin B, are important in *Clostridium difficile* infection. Gut Microbes. 2011;2(4):252–5.

67. Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP. The role of toxin A and toxin B in *Clostridium difficile* infection. Nature. 2010;467(7316):711–3.

68. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. The ClosTron: A universal gene knock-out system for the genus Clostridium. Journal of Microbiological Methods. 2007;70(3):452–64.

69. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, et al. The ClosTron: Mutagenesis in Clostridium refined and streamlined. Journal of Microbiological Methods. 2010;80(1):49–55.

70. Kuehne SA, Collery MM, Kelly ML, Cartman ST, Cockayne A, Minton NP. Importance of toxin A, toxin B, and CDT in virulence of an epidemic *Clostridium difficile* strain. Journal of Infectious Diseases. 2014;209(1):83–6.

71. Carter GP, Chakravorty A, Pham Nguyen TA, Mileto S, Schreiber F, Li L, et al. Defining the Roles of TcdA and TcdB in Localized Gastrointestinal Disease, Systemic Organ Damage, and the Host Response during *Clostridium difficile* Infections. mBio. 2015;6(3):e00551.

72. Monot M, Eckert C, Lemire A, Hamiot A, Dubois T, Tessier C, et al. *Clostridium difficile*: New Insights into the Evolution of the Pathogenicity Locus. Scientific Reports. 2015;5:15023.

73. Hammond GA, Johnson JL. The toxigenic element of *Clostridium difficile* strain VPI 10463. Microbial Pathogenesis. 1995;19(4):203–13.

74. O'Connor JR, Johnson S, Gerding DN. *Clostridium difficile* infection caused by the epidemic BI/NAP1/027 strain. Gastroenterology. 2009;136(6):1913–24.

75. Mani N, Lyras D, Barroso L, Howarth P, Wilkins T, Rood JI, et al. Environmental response and autoregulation of *Clostridium difficile* TxeR, a sigma factor for toxin

gene expression. Journal of Bacteriology. 2002;184(21):5971–8.

76. Bakker D, Smits WK, Kuijper EJ, Corver J. TcdC does not significantly repress toxin expression in *Clostridium difficile* 630ΔErm. PLoS ONE. 2012;7(8):e43247.

77. Matamouros S, England P, Dupuy B. *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. Molecular Microbiology. 2007;64(5):1274–88.

78. Vohra P, Poxton IR. Comparison of toxin and spore production in clinically relevant strains of *Clostridium difficile*. Microbiology. 2011;157(5):1343–53.

79. Carter GP, Douce GR, Govind R, Howarth PM, Mackin KE, Spencer J, et al. The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of *Clostridium difficile*. PLoS Pathogens. 2011;7(10):e1002317.

80. Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. Applied and Environmental Microbiology. 2012;78(13):4683–90.

81. Van Leeuwen HC, Bakker D, Steindel P, Kuijper EJ, Corver J. *Clostridium difficile* TcdC protein binds four-stranded G-quadruplex structures. Nucleic Acids Research. 2013;

82. Govind R, Dupuy B. Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. PLoS Pathogens. 2012;8(6):e1002727.

83. Tan KS, Wee BY, Song KP. Evidence for holin function of tcdE gene in the pathogenicity of *Clostridium difficile*. Journal of Medical Microbiology. 2001;50(7):613–9.

84. Olling A, Seehase S, Minton NP, Tatge H, Schröter S, Kohlscheen S, et al. Release of TcdA and TcdB from *Clostridium difficile* cdi 630 is not affected by functional inactivation of the tcdE gene. Microbial Pathogenesis. 2012;52(1):92–100.

85. Govind R, Fitzwater L, Nichols R. Observations on the Role of TcdE Isoforms in *Clostridium difficile* Toxin Secretion. Journal of Bacteriology. 2015;197(15):2600–9.

86. Rupnik M, Grabnar M, Geric B. Binary toxin producing *Clostridium difficile* strains. Anaerobe. 2003;9(6):289–94.

87. Stewart DB, Berg A, Hegarty J. Predicting recurrence of *Clostridium difficile* colitis using bacterial virulence factors: binary toxin is the key. Journal of Gastrointestinal Surgery. 2013;17(1):118–25.

88. Schwan C, Kruppke AS, Nölke T, Schumacher L, Koch-Nolte F, Kudryashev M, et al. *Clostridium difficile* toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence. Proceedings of the National

Academy of Sciences. 2014;111(6):2313–8.

89. Wang S, Rustandi RR, Lancaster C, Hong LG, Thiriot DS, Xie J, et al. Toxicity assessment of *Clostridium difficile* toxins in rodent models and protection of vaccination. Vaccine. 2016;34(10):1319–23.

90. Geric B, Carman RJ, Rupnik M, Genheimer CW, Sambol SP, Lyerly DM, et al. Binary toxin-producing, large clostridial toxin-negative *Clostridium difficile* strains are enterotoxic but do not cause disease in hamsters. Journal of Infectious Diseases. 2006;193(8):1143–50.

91. Janoir C. Virulence factors of *Clostridium difficile* and their role during infection. Anaerobe. 2016;37:13–24.

92. Lawley TD, Clare S, Deakin LJ, Goulding D, Yen JL, Raisen C, et al. Use of purified *Clostridium difficile* spores to facilitate evaluation of health care disinfection regimens. Applied and Environmental Microbiology. 2010;76(20):6895–900.

93. Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, et al. Antibiotic treatment of *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. Infection and Immunity [Internet]. 2009;77(9):3661–9. Available from: t

94. Falkow S. Molecular Koch's postulates applied to microbial pathogenicity. Review of Infectious Diseases. 1988;10(Supplement 2):S274–S276.

95. O'Connor JR, Lyras D, Farrow KA, Adams V, Powell DR, Hinds J, et al. Construction and analysis of chromosomal *Clostridium difficile* mutants. Molecular Microbiology. 2006;61(5):1335–51.

96. Perutka J, Wang W, Goerlitz D, Lambowitz AM. Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. Journal of Molecular Biology. 2004;336(2):421–39.

97. Cartman ST, Minton NP. A mariner-based transposon system for in vivo random mutagenesis of *Clostridium difficile*. Applied and Environmental Microbiology. 2010;76(4):1103–9.

98. Wang H, Smith MCM, Mullany P. The conjugative transposon Tn5397 has a strong preference for integration into its *Clostridium difficile* target site. Journal of Bacteriology. 2006;188(13):4871–8.

99. Hussain HA, Roberts AP, Mullany P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. Journal of Medical Microbiology. 2005;54(2):137–41.

100. Lampe DJ, Grant TE, Robertson HM. Factors affecting transposition of the Himar1 mariner transposon in vitro. Genetics. 1998;149(1):179–87.

101. Nariya H, Miyata S, Suzuki M, Tamai E, Okabe A. Development and application of a method for counterselectable in-frame deletion in *Clostridium perfringens*. Applied and Environmental Microbiology. 2011;77(4):1375–82.

102. Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, et al. Development of pyrF-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant. Applied and Environmental Microbiology. 2010;76(19):6591–9.

103. Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. Nature Reviews Cancer. 2003;3(5):330–8.

104. Dupuy B, Sonenshein AL. Regulated transcription of *Clostridium difficile* toxin genes. Molecular Microbiology. 1998;27(1):107–20.

105. Williams DR, Young DI, Young M. Conjugative plasmid transfer from Escherichia coli to *Clostridium acetobutylicum*. Journal of General Microbiology. 1990;136(5):819–26.

106. Brazier JS. Role of the laboratory in investigations of *Clostridium difficile* diarrhoea. Clinical Infectious Diseases. 1993;16(Supplement 4):S228–S233.

107. Clabots CR, Bettin KM, Peterson LR, Gerding D. Evaluation of cycloserine-cefoxitin-fructose agar and cycloserine-cefoxitin-fructose broth for recovery of *Clostridium difficile* from environmental sites. Journal of Clinical Microbiology. 1991;29(11):2633–5.

108. Anaerobe Systems. Cycloserine Cefoxitin Mannitol Broth with Taurocholate and Lysozyme (CCMB-TAL). 2013 [cited 2014 Jan 6]; Available from: http://www.anaerobesystems.com/Home/pras-tubed-media/cycloserine-cefoxitin-mannitol-broth-with-taurocholate-and-lysozyme-ccmb-tal

109. Sambrook J, Russell D. Molecular Cloning: A Laboratory Manual. The 3rd ed., . Cold Spring Harbor, NY; 2001.

110. Hedges RW, Jacob AE. Transposition of ampicillin resistance from RP4 to other replicons. Molecular and General Genetics. 1974;132(1):31–40.

111. Heap JT, Pennington OJ, Cartman ST, Minton NP. A modular system for Clostridium shuttle plasmids. Journal of Microbiological Methods. 2009;78(1):79–85.

112. Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, et al. Conjugative transfer of clostridial shuttle vectors from Escherichia coli to *Clostridium difficile* through circumvention of the restriction barrier. Molecular Microbiology. 2002;46(2):439–52.

113. Solovyev V, Salamov A. In book: Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies, Chapter: Automatic Annotation of Microbial Genomes and Metagenomic Sequences. 2011. p. 61–78.

114. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Research. 2016;44(W1):W16–21.

115. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, et al. Artemis: sequence visualization and annotation. Bioinformatics. 2000;16(10):944–5.

116. Hackl T, Hedrich R, Schultz J, Förster F. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics. 2014;30(21):3004–11.

117. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

118. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

119. Evans PM, Liu C. SiteFind: a software tool for introducing a restriction site as a marker for successful site-directed mutagenesis. BMC Molecular Biology. 2005;6:22.

120. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 2010;5(6):e11147.

121. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.

122. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Research. 2014;42(Database issue):D206–14.

123. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J. ACT: the Artemis comparison tool. Bioinformatics. 2005;21(16):3422–3.

124. O'Neill G, Ogunsola F, Brazier J, Duerden B. Modification of a PCR Ribotyping Method for Application as a Routine Typing Scheme for *Clostridium difficile*. Anaerobe. 1996;2(4):205–9.

125. Walk ST, Micic D, Jain R, Lo ES, Trivedi I, Liu EW, et al. *Clostridium difficile* ribotype does not predict severe infection. Clinical Infectious Diseases. 2012;55(12):1661–8.

126. Barbut F, Decré D, Burghoffer B, Lesage D, Delisle F, Lalande V, et al. Antimicrobial susceptibilities and serogroups of clinical strains of *Clostridium*

difficile isolated in France in 1991 and 1997. Antimicrobial Agents and Chemotherapy. 1999;43(11):2607–11.

127. Moncrief JS, Barroso LA, Wilkins TD. Positive regulation of *Clostridium difficile* toxins. Infection and Immunity. 1997;65(3):1105–8.

128. Mani N, Dupuy B. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. Proceedings of the National Academy of Sciences. 2001;98(10):5844–9.

129. Dupuy B, Raffestin S, Matamouros S, Mani N, Popoff MR, Sonenshein AL. Regulation of toxin and bacteriocin gene expression in Clostridium by interchangeable RNA polymerase sigma factors. Molecular Microbiology. 2006;60(4):1044–57.

130. Reddy ARS, Girinathan BP, Zapotocny R, Govind R. Identification and Characterization of *Clostridium sordellii* Toxin Gene Regulator. Journal of Bacteriology. 2013;195(18):4246–54.

131. Raffestin S, Dupuy B, Marvaud JC, Popoff MR. BotR/A and TetR are alternative RNA polymerase sigma factors controlling the expression of the neurotoxin and associated protein genes in *Clostridium botulinum* type A and *Clostridium tetani*. Molecular Microbiology. 2005;55(1):235–49.

132. Dupuy B, Matamouros S. Regulation of toxin and bacteriocin synthesis in Clostridium species by a new subgroup of RNA polymerase sigma-factors. Research in Microbiology. 2006;157(3):201–5.

133. Carter GP, Larcombe S, Li L, Jayawardena D, Awad MM, Songer JG, et al. Expression of the large clostridial toxins is controlled by conserved regulatory mechanisms. International Journal of Medical Microbiology. 2014;

134. Karlsson S, Dupuy B, Mukherjee K, Norin E, Burman LG, Åkerlund T. Expression of *Clostridium difficile* toxins A and B and their sigma factor TcdD is controlled by temperature. Infection and Immunity. 2003;71(4):1784–93.

135. Neumann-Schaal M, Hofmann JD, Will SE, Schomburg D. Time-resolved amino acid uptake of *Clostridium difficile* 630Δerm and concomitant fermentation product and toxin formation. BMC Microbiology. 2015;15:281.

136. Karlsson S, Burman LG, Åkerlund T. Suppression of toxin production in *Clostridium difficile* VPI 10463 by amino acids. Microbiology. 1999;145(7):1683–93.

137. Antunes A, Martin-Verstraete I, Dupuy B. CcpA-mediated repression of *Clostridium difficile* toxin gene expression. Molecular Microbiology. 2011;79(4):882–99.

138. Antunes A, Camiade E, Monot M, Courtois E, Barbut F, Sernova NV, et al. Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium*

difficile. Nucleic Acids Research. 2012;40(21):10701–18.

139. Dineen SS, Villapakkam AC, Nordman JT, Sonenshein AL. Repression of *Clostridium difficile* toxin gene expression by CodY. Molecular Microbiology. 2007;66(1):206–19.

140. Shivers RP, Sonenshein AL. Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. Molecular Microbiology. 2004;53(2):599–611.

141. Ratnayake-Lecamwasam M, Serror P, Wong KW, Sonenshein AL. *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. Genes & Development. 2001;15(9):1093–103.

142. Dineen SS, McBride SM, Sonenshein AL. Integration of metabolism and virulence by *Clostridium difficile* CodY. Journal of Bacteriology. 2010;192(20):5350–62.

143. Bouillaut L, Self WT, Sonenshein AL. Proline-dependent regulation of *Clostridium difficile* Stickland metabolism. Journal of Bacteriology. 2013;195(4):844–54.

144. Karlsson S, Lindberg A, Norin E, Burman LG, Åkerlund T. Toxins, butyric acid, and other short-chain fatty acids are coordinately expressed and down-regulated by cysteine in *Clostridium difficile*. Infection and Immunity. 2000;68(10):5881–8.

145. Saujet L, Monot M, Dupuy B, Soutourina O, Martin-Verstraete I. The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*. Journal of Bacteriology. 2011;193(13):3186–96.

146. Fimlaid KA, Bond JP, Schutz KC, Putnam EE, Leung JM, Lawley TD, et al. Global analysis of the sporulation pathway of *Clostridium difficile*. PLoS Genetics. 2013;9(8):e1003660.

147. Mackin KE, Carter GP, Howarth P, Rood JI, Lyras D. Spo0A differentially regulates toxin production in evolutionarily diverse strains of *Clostridium difficile*. PLoS ONE. 2013;8(11):e79666.

148. Deakin LJ, Clare S, Fagan RP, Dawson LF, Pickard DJ, West MR, et al. The *Clostridium difficile* spo0A gene is a persistence and transmission factor. Infection and Immunity. 2012;80(8):2704–11.

149. Pettit LJ, Browne HP, Yu L, Smits WK, Fagan RP, Barquist L, et al. Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. BMC Genomics. 2014;15:160.

150. Rosenbusch KE, Bakker D, Kuijper EJ, Smits WK. *Clostridium 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.

151. Underwood S, Guan S, Vijayasubhash V, Baines SD, Graham L, Lewis RJ, et al. Characterization of the sporulation initiation pathway of *Clostridium difficile* and its role in toxin production. Journal of Bacteriology. 2009;191(23):7296–305.

152. Edwards AN, Tamayo R, McBride SM. A novel regulator controls *Clostridium difficile* sporulation, motility and toxin production. Molecular Microbiology. 2016;100(6):954–71.

153. Martin MJ, Clare S, Goulding D, Faulds-Pain A, Barquist L, Browne HP, et al. The agr locus regulates virulence and colonization genes in *Clostridium difficile* 027. Journal of Bacteriology. 2013;195(16):3672–81.

154. Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, et al. 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.

155. Marsden GL, Davis IJ, Wright VJ, Sebaihia M, Kuijper EJ, Minton NP. Array comparative hybridisation reveals a high degree of similarity between UK and European clinical isolates of hypervirulent *Clostridium difficile*. BMC Genomics. 2010;11:389.

156. Darkoh C, DuPont HL, Norris SJ, Kaplan HB. Toxin synthesis by *Clostridium difficile* is regulated through quorum signaling. mBio. 2015;6(2):e02569.

157. Lee ASY, Song KP. LuxS/autoinducer-2 quorum sensing molecule regulates transcriptional virulence gene expression in *Clostridium difficile*. Biochemical and Biophysical Research Communications. 2005;335(3):659–66.

158. Carter GP, Purdy D, Williams P, Minton NP. Quorum sensing in *Clostridium difficile*: analysis of a luxS-type signalling system. Journal of Medical Microbiology. 2005;54(Pt 2):119–27.

159. McKee RW, Mangalea MR, Purcell EB, Borchardt EK, Tamayo R. The second messenger cyclic Di-GMP regulates *Clostridium difficile* toxin production by controlling expression of sigD. Journal of Bacteriology. 2013;195(22):5174–85.

160. El Meouche I, Peltier J, Monot M, Soutourina O, Pestel-Caron M, Dupuy B, et al. Characterization of the SigD regulon of C. difficile and its positive control of toxin production through the regulation of tcdR. PLoS ONE. 2013;8(12):e83748.

161. Aubry A, Hussack G, Chen W, KuoLee R, Twine SM, Fulton KM, et al. Modulation of toxin production by the flagellar regulon in *Clostridium difficile*. Infection and Immunity. 2012;80(10):3521–32.

162. Baban ST, Kuehne SA, Barketi-Klai A, Cartman ST, Kelly ML, Hardie KR, et al. The role of flagella in *Clostridium difficile* pathogenesis: Comparison between a non-epidemic and an epidemic strain. PLoS ONE. 2013;8(9):e73026.

163. Dingle TC, Mulvey GL, Armstrong GD. Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. Infection and Immunity. 2011;79(10):4061–7.

164. Twine SM, Reid CW, Aubry A, McMullin DR, Fulton KM, Austin J, et al. Motility and flagellar glycosylation in *Clostridium difficile*. Journal of Bacteriology. 2009;191(22):7050–62.

165. Martin-Verstraete I, Peltier J, Dupuy B. The Regulatory Networks That Control *Clostridium difficile* Toxin Synthesis. Toxins. 2016;8(5).

166. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and Evolution in the Genome of *Clostridium difficile* . Clinical Microbiology Reviews. 2015;28(3):721–41.

167. Williams R, Meader E, Mayer M, Narbad A, Roberts AP, Mullany P. Determination of the attP and attB sites of phage CD27 from *Clostridium difficile* NCTC 12727. Journal of Medical Microbiology. 2013;62(Pt 9):1439–43.

168. Govind R, Vediyappan G, Rolfe RD, Dupuy B, Fralick JA. Bacteriophage-mediated toxin gene regulation in *Clostridium difficile*. Journal of Virology. 2009;83(23):12037–45.

169. Sekulovic O, Meessen-Pinard M, Fortier L-C. Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. Journal of Bacteriology. 2011;193(11):2726–34.

170. Goh S, Chang BJ, Riley TV. Effect of phage infection on toxin production by *Clostridium difficile*. Journal of Medical Microbiology. 2005;54(2):129–35.

171. Walter BM, Cartman ST, Minton NP, Butala M, Rupnik M. The SOS Response Master Regulator LexA Is Associated with Sporulation, Motility and Biofilm Formation in *Clostridium difficile* . PLoS ONE. 2015;10(12):e0144763.

172. Willing SE, Richards EJ, Sempere L, Dale AG, Cutting SM, Fairweather NF. Increased toxin expression in a *Clostridium difficile* mfd mutant. BMC Microbiology. 2015;15:280.

173. Tosh SM. The research legacy of Peter J. Wood. Bioactive Carbohydrates and Dietary Fibre. 2013;2(2):170–80.

174. Fry SC, Nesselrode BHWA, Miller JG, Mewburn BR. Mixed-linkage (1->3,1->4)-beta-D-glucan is a major hemicellulose of Equisetum (horsetail) cell walls. The New Phytologist. 2008;179(1):104–15.

175. Monot M, Boursaux-Eude C, Thibonnier M, Vallenet D, Moszer I, Medigue C, et al. Reannotation of the genome sequence of *Clostridium difficile* strain 630. Journal of Medical Microbiology. 2011;60(8):1193–9.

176. Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, et al. High-Throughput Analysis of Gene Essentiality and Sporulation in *Clostridium difficile* . mBio. 2015;6(2):e02383–14.

177. Makarova KS, Wolf YI, Koonin EV. Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Research. 2013;gkt157.

178. Saller MJ, Fusetti F, Driessen AJ. *Bacillus subtilis* SpoIIIJ and YqjG function in membrane protein biogenesis. Journal of Bacteriology. 2009;191(21):6749–57.

179. Bradshaw M, Dineen SS, Maks ND, Johnson EA. Regulation of neurotoxin complex expression in *Clostridium botulinum* strains 62A, Hall A-hyper, and NCTC 2916. Anaerobe. 2004;10(6):321–33.

180. Darkoh C, Dupont HL, Kaplan HB. Novel one-step method for detection and isolation of active-toxin-producing *Clostridium difficile* strains directly from stool samples. Journal of Clinical Microbiology. 2011;49(12):4219–24.

181. Darkoh C, Kaplan HB, Dupont HL. Harnessing the glucosyltransferase activities of *Clostridium difficile* for functional studies of toxins A and B. Journal of Clinical Microbiology. 2011;49(8):2933–41.

182. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, et al. The minimal gene complement of *Mycoplasma genitalium*. Science. 1995;270(5235):397–404.

183. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. 1995;269(5223):496–512.

184. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology. 1975;94(3):441–8.

185. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. Fluorescence detection in automated DNA sequence analysis. Nature. 1985;321(6071):674–9.

186. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376–80.

187. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53–9.

188. Shendure J, Ji H. Next-generation DNA sequencing. Nature Biotechnology. 2008;26(10):1135–45.

189. Hunt M, Newbold C, Berriman M, Otto TD. A comprehensive evaluation of assembly scaffolding tools. Genome Biology. 2014;15(3):R42.

190. Baker M. De novo genome assembly: what every biologist should know. Nature Methods. 2012;9(4):333–7.

191. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323(5910):133–8.

192. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biology. 2013;14(7):405.

193. Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M, Penn CW, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nature Reviews Microbiology. 2012;10(9):599–606.

194. Sullivan NM, Pellett S, Wilkins T. Purification and characterization of toxins A and B of *Clostridium difficile*. Infection and Immunity. 1982;35(3):1032–40.

195. Von Eichel-Streiber C, Laufenberg-Feldmann R, Sartingen S, Schulze J, Sauerborn M. Comparative sequence analysis of the *Clostridium difficile* toxins A and B. Molecular and General Genetics. 1992;233(1-2):260–8.

196. Lyerly DM, Wilkins TD. Commercial latex test for *Clostridium difficile* toxin A does not detect toxin A. Journal of Clinical Microbiology. 1986;23(3):622–3.

197. Ikeda D, Karasawa T, Yamakawa K, Tanaka R, Namiki M, Nakamura S. Effect of isoleucine on toxin production by *Clostridium difficile* in a defined medium. Zentralblatt für Bakteriologie. 1998;287(4):375–86.

198. Trindade BC, Theriot CM, Leslie JL, Carlson PE, Bergin IL, Peters-Golden M, et al. *Clostridium difficile*-induced colitis in mice is independent of leukotrienes. Anaerobe. 2014;30:90–8.

199. Anosova NG, Cole LE, Li L, Zhang J, Brown AM, Mundle S, et al. A Combination of Three Fully Human Toxin A- and Toxin B-Specific Monoclonal Antibodies Protects against Challenge with Highly Virulent Epidemic Strains of *Clostridium difficile* in the Hamster Model. Clinical and Vaccine Immunology. 2015;22(7):711–25.

200. McDermott AJ, Falkowski NR, McDonald RA, Pandit CR, Young VB, Huffnagle GB. Interleukin-23 (IL-23), independent of IL-17 and IL-22, drives neutrophil recruitment and innate inflammation during *Clostridium difficile* colitis in mice. Immunology. 2016;147(1):114–24.

201. Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nature Biotechnology. 2015;33(10):1109.

202. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, et al. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v. 4). Standards in Genomic Sciences. 2015;10(1):1.

203. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Research. 2012;40(Database issue):D115–22.

204. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, et al. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Research. 2012;40(4):e29.

205. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature Methods. 2010;7(6):461–5.

206. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—restriction enzymes and DNA methyltransferases. Nucleic Acids Research. 2005;33(suppl 1):D230–D232.

207. Stabler RA, Valiente E, Dawson LF, He M, Parkhill J, Wren BW. In-depth genetic analysis of *Clostridium difficile* PCR-ribotype 027 strains reveals high genome fluidity including point mutations and inversions. Gut Microbes. 2010;1(4):269–76.

208. Briggs GS, Smits WK, Soultanas P. Chromosomal replication initiation machinery of low-G+ C-content Firmicutes. Journal of Bacteriology. 2012;194(19):5162–70.

209. Kristensen DM, Cai X, Mushegian A. Evolutionarily conserved orthologous families in phages are relatively rare in their prokaryotic hosts. Journal of Bacteriology. 2011;193(8):1806–14.

210. Nakagawa I, Kurokawa K, Yamashita A, Nakata M, Tomiyasu Y, Okahashi N, et al. Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. Genome Research. 2003;13(6A):1042–55.

211. Tillier ER, Collins RA. Genome rearrangement by replication-directed translocation. Nature Genetics. 2000;26(2):195–7.

212. Daveran-Mingot ML, Campo N, Ritzenthaler P, Le Bourgeois P. A natural large chromosomal inversion in *Lactococcus lactis* is mediated by homologous recombination between two insertion sequences. Journal of Bacteriology. 1998;180(18):4834–42.

213. Segall AM, Roth JR. Approaches to half-tetrad analysis in bacteria: recombination between repeated, inverse-order chromosomal sequences. Genetics. 1994;136(1):27–39.

214. Mullany P, Allan E, Roberts AP. Mobile genetic elements in *Clostridium difficile* and their role in genome function. Research in Microbiology. 2015;166(4):361–7.

215. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, et al. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nature Genetics. 2006;38(7):779–86.

216. Campo N, Dias MJ, Daveran-Mingot M-L, Ritzenthaler P, Le Bourgeois P. Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. Molecular Microbiology. 2004;51(2):511–22.

217. Iguchi A, Iyoda S, Terajima J, Watanabe H, Osawa R. Spontaneous recombination between homologous prophage regions causes large-scale inversions within the *Escherichia coli* O157:H7 chromosome. Gene. 2006;372:199–207.

218. Van Sluys MA, de Oliveira MC, Monteiro-Vitorello CB, Miyaki CY, Furlan LR, Camargo LEA, et al. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. Journal of Bacteriology. 2003;185(3):1018–26.

219. Stabler RA, Gerding DN, Songer JG, Drudy D, Brazier JS, Trinh HT, et al. Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. Journal of Bacteriology. 2006;188(20):7297–305.

220. Faulds-Pain A, Twine SM, Vinogradov E, Strong PC, Dell A, Buckley AM, et al. The post-translational modification of the *Clostridium difficile* flagellin affects motility, cell surface properties and virulence. Molecular Microbiology. 2014;94(2):272–89.

221. Ibuki T, Imada K, Minamino T, Kato T, Miyata T, Namba K. Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. Nature Structural & Molecular Biology. 2011;18(3):277–82.

222. Minamino T, Chu R, Yamaguchi S, Macnab RM. Role of FliJ in Flagellar Protein Export in Salmonella. Journal of Bacteriology. 2000;182(15):4207–15.

223. Saier MH, Reddy BL. Holins in bacteria, eukaryotes, and archaea: multifunctional xenologues with potential biotechnological and biomedical applications. Journal of Bacteriology. 2015;197(1):7–17.

224. Lang AS, Zhaxybayeva O, Beatty JT. Gene transfer agents: phage-like elements of genetic exchange. Nature Reviews Microbiology. 2012;10(7):472–82.

225. Ranjit DK, Endres JL, Bayles KW. *Staphylococcus aureus* CidA and LrgA proteins exhibit holin-like properties. Journal of Bacteriology. 2011;193(10):2468–76.

226. Herbert M, O'Keeffe TA, Purdy D, Elmore M, Minton NP. Gene transfer into *Clostridium difficile* CD630 and characterisation of its methylase genes. FEMS

Microbiology Letters. 2003;229(1):103–10.

227. Monk IR, Tree JJ, Howden BP, Stinear TP, Foster TJ. Complete Bypass of Restriction Systems for Major *Staphylococcus aureus* Lineages. mBio. 2015;6(3):e00308–15.

228. Lesiak JM, Liebl W, Ehrenreich A. Development of an in vivo methylation system for the solventogen *Clostridium saccharobutylicum* NCP 262 and analysis of two endonuclease mutants. Journal of Biotechnology. 2014;188:97–9.

229. Murray NE. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiology and Molecular Biology Reviews. 2000;64(2):412–34.

230. Zhang G, Wang W, Deng A, Sun Z, Zhang Y, Liang Y, et al. A mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria. PLoS Genetics. 2012;8(9):e1002987.

231. O'Connell Motherway M, O'Driscoll J, Fitzgerald GF, Van Sinderen D. Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in *Bifidobacterium breve* UCC2003. Microbial Biotechnology. 2009;2(3):321–32.

232. Gonzalez P, Samuel J, Epstein DL, Borrás T. Increasing DNA transfer efficiency by temporary inactivation of host restriction. BioTechniques. 1999;26(5):892–900.

233. Zeng X, Ardeshna D, Lin J. Heat Shock-Enhanced Conjugation Efficiency in Standard *Campylobacter jejuni* Strains. Applied and Environmental Microbiology. 2015;81(13):4546–52.

234. Kirk JA, Fagan RP. Heat shock increases conjugation efficiency in *Clostridium difficile*. Anaerobe. 2016;42:1–5.

235. Luo ML, Leenay RT, Beisel CL. Current and future prospects for CRISPR-based tools in bacteria. Biotechnology and Bioengineering. 2016;113(5):930–43.

236. Haeussler M, Concordet J-P. Genome Editing with CRISPR-Cas9: Can It Get Any Better? Journal of Genetics and Genomics . 2016;43(5):239–50.

237. Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. Scientific Reports. 2016;6:25666.

238. Bartlett JG. *Clostridium difficile*: progress and challenges. Annals of the New York Academy of Sciences. 2010;1213:62–9.

239. Schmidt ML, Gilligan PH. *Clostridium difficile* testing algorithms: what is practical and feasible? Anaerobe. 2009;15(6):270–3.

240. Department of Health. Updated guidance on the diagnosis and reporting of *Clostridium difficile*. London; DoH; December 2012; 2012.

241. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, et al. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. Nature Genetics. 2012;43(1):109–13.

242. Cartman ST, Heap JT, Kuehne SA, Cockayne A, Minton NP. The emergence of "hypervirulence" in *Clostridium difficile*. International Journal of Medical Microbiology. 2010;300(6):387–95.

243. Deneve C, Janoir C, Poilane I, Fantinato C, Collignon A. New trends in *Clostridium difficile* virulence and pathogenesis. International Journal of Antimicrobial Agents. 2009;33:S24–S28.

244. Burns DA, Heeg D, Cartman ST, Minton NP. Reconsidering the sporulation characteristics of hypervirulent *Clostridium difficile* BI/NAP1/027. PLoS ONE. 2011;6(9):e24894.

245. Aitken SL, Alam MJ, Khaleduzzuman M, Walk ST, Musick WL, Pham VP, et al. In the Endemic Setting, *Clostridium difficile* Ribotype 027 Is Virulent But Not Hypervirulent. Infection Control and Hospital Epidemiology. 2015;1–6.

246. Carlson PE, Walk ST, Bourgis AET, Liu MW, Kopliku F, Lo E, et al. The relationship between phenotype, ribotype, and clinical disease in human Clostridium difficile isolates. Anaerobe. 2013;24:109–16.

247. Barbut F, Rupnik M. 027, 078, and others: going beyond the numbers (and away from the hypervirulence). Clinical Infectious Diseases. 2012;55(12):1669–72.

248. John R, Brazier J. Antimicrobial susceptibility of polymerase chain reaction ribotypes of *Clostridium difficile* commonly isolated from symptomatic hospital patients in the UK. Journal of Hospital Infection. 2005;61(1):11–4.

249. Brazier J, Patel B, Pearson A, others. Distribution of *Clostridium difficile* PCR ribotype 027 in British hospitals. Euro Surveillance. 2007;12(4):E070426.

250. Bauer MP, Notermans DW, van Benthem BHB, Brazier JS, Wilcox MH, Rupnik M, et al. *Clostridium difficile* infection in Europe: a hospital-based survey. Lancet. 2011;377(9759):63–73.

251. Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, et al. Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clinical Infectious Diseases. 2008;47(9):1162–70.

252. Public Health England. *Clostridium difficile* Ribotyping Networj (CDRN) for England and Nothern Ireland, Biennial Report (2013-2015). London; PHE; 2016; 2016.

253. Bakker D, Corver J, Harmanus C, Goorhuis A, Keessen EC, Fawley WN, et al. Relatedness of human and animal *Clostridium difficile* PCR ribotype 078 isolates determined on the basis of multilocus variable-number tandem-repeat analysis and tetracycline resistance. Journal of Clinical Microbiology. 2010;48(10):3744–9.

254. Chitnis AS, Holzbauer SM, Belflower RM, Winston LG, Bamberg WM, Lyons C, et al. Epidemiology of community-associated *Clostridium difficile* infection, 2009 through 2011. JAMA Internal Medicine. 2013;173(14):1359–67.

255. Wilcox M, Mooney L, Bendall R, Settle C, Fawley W. A case-control study of community-associated *Clostridium difficile* infection. Journal of Antimicrobial Chemotherapy. 2008;62(2):388–96.

256. Fellmeth G, Yarlagadda S, Iyer S. Epidemiology of community-onset *Clostridium difficile* infection in a community in the South of England. Journal of Infection and Public Health. 2010;3(3):118–23.

257. Martin JSH, Monaghan TM, Wilcox MH. *Clostridium difficile* infection: epidemiology, diagnosis and understanding transmission. Nature reviews Gastroenterology & hepatology. 2016;13(4):206–16.

258. Fawley W, Davies K, Morris T, Parnell P, Howe R, Wilcox M. Surveillance and outbreak report: Enhanced surveillance of *Clostridium difficile* infection occurring outside hospital, England, 2011 to 2013. Euro Surveillance. 2016;21(29).

259. Wilcox M, Shetty N, Fawley W, Shemko M, Coen P, Birtles A, et al. Changing epidemiology of *Clostridium difficile* infection following the introduction of a national ribotyping-based surveillance scheme in England. Clinical Infectious Diseases. 2012;614–21.

260. Shields K, Araujo-Castillo RV, Theethira TG, Alonso CD, Kelly CP. Recurrent *Clostridium difficile* infection: From colonization to cure. Anaerobe. 2015;34:59–73.

261. Mac Aogáin M, Moloney G, Kilkenny S, Kelleher M, Kelleghan M, Boyle B, et al. Whole-genome sequencing improves discrimination of relapse from reinfection and identifies transmission events among patients with recurrent *Clostridium difficile* infections. Journal of Hospital Infection. 2015;90(2):108–16.

262. Barbut F, Richard A, Hamadi K, Chomette V, Burghoffer B, Petit J-C. Epidemiology of recurrences or reinfections of *Clostridium difficile*-associated diarrhea. Journal of Clinical Microbiology. 2000;38(6):2386–8.

263. Plaza-Garrido Á, Miranda-Cárdenas C, Castro-Córdova P, Olguin-Araneda V, Cofré-Araneda G, Hernández-Rocha C, et al. Outcome of relapsing *Clostridium difficile* infections do not correlate with virulence-, spore- and vegetative cell-associated phenotypes. Anaerobe. 2015;36:30–8.

264. Oka K, Osaki T, Hanawa T, Kurata S, Okazaki M, Manzoku T, et al. Molecular and microbiological characterization of *Clostridium difficile* isolates from single,

relapse, and reinfection cases. Journal of Clinical Microbiology. 2012;50(3):915–21.

265. Kyne L, Warny M, Qamar A, Kelly CP. Association between antibody response to toxin A and protection against recurrent *Clostridium difficile* diarrhoea. Lancet. 2001;357(9251):189–93.

266. Póvoa P, Coelho L, Almeida E, Fernandes A, Mealha R, Moreira P, et al. C-reactive protein as a marker of infection in critically ill patients. Clinical Microbiology and Infection. 2005;11(2):101–8.

267. Eyre DW, Walker AS, Wyllie D, Dingle KE, Griffiths D, Finney J, et al. Predictors of first recurrence of *Clostridium difficile* infection: implications for initial management. Clinical Infectious Diseases. 2012;55(suppl 2):S77–S87.

268. Hebert C, Du H, Peterson LR, Robicsek A. Electronic health record-based detection of risk factors for *Clostridium difficile* infection relapse. Infection Control and Hospital Epidemiology. 2013;34(4):407–14.

269. Kumar N, Miyajima F, He M, Roberts P, Swale A, Ellison L, et al. Genome-Based Infection Tracking Reveals Dynamics of *Clostridium difficile* Transmission and Disease Recurrence. Clinical Infectious Diseases. 2016;62(6):746–52.

270. Sun J, Mc Millen T, Babady NE, Kamboj M. Role of Coinfecting Strains in Recurrent *Clostridium difficile* Infection. Infection Control and Hospital Epidemiology. 2016;1–4.

271. Behroozian AA, Chludzinski JP, Lo ES, Ewing SA, Waslawski S, Newton DW, et al. Detection of Mixed Populations of *Clostridium difficile* from Symptomatic Patients Using Capillary-Based Polymerase Chain Reaction Ribotyping. Infection Control and Hospital Epidemiology. 2013;34(9):961.

272. Van den Berg RJ, Ameen HA, Furusawa T, Claas EC, van der Vorm ER, Kuijper EJ. Coexistence of multiple PCR-ribotype strains of *Clostridium difficile* in faecal samples limits epidemiological studies. Journal of Medical Microbiology. 2005;54(2):173–9.

273. Eyre DW, Cule ML, Griffiths D, Crook DW, Peto TEA, Walker AS, et al. Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in *Clostridium difficile* transmission. PLoS Computational Biology. 2013;9(5):e1003059.

274. Moura I, Spigaglia P, Barbanti F, Mastrantonio P. Analysis of metronidazole susceptibility in different *Clostridium difficile* PCR ribotypes. Journal of Antimicrobial Chemotherapy. 2013;68(2):362–5.

275. Baines SD, O'Connor R, Freeman J, Fawley WN, Harmanus C, Mastrantonio P, et al. Emergence of reduced susceptibility to metronidazole in *Clostridium difficile*. Journal of Antimicrobial Chemotherapy. 2008;62(5):1046–52.

276. Freeman J, Stott J, Baines SD, Fawley WN, Wilcox MH. Surveillance for resistance to metronidazole and vancomycin in genotypically distinct and UK epidemic *Clostridium difficile* isolates in a large teaching hospital. Journal of Antimicrobial Chemotherapy. 2005;56(5):988–9.

277. Eckert C, Burghoffer B, Lalande V, Barbut F. Evaluation of the Chromogenic Agar ChromID *Clostridium difficile*. Journal of Clinical Microbiology. 2013;51(3):1002–4.

278. Carson KC, Boseiwaqa LV, Thean SK, Foster NF, Riley TV. Isolation of *Clostridium difficile* from faecal specimens-a comparison of ChromID C. difficile agar and cycloserine cefoxitin fructose agar. Journal of Medical Microbiology. 2013;62:1423–727.

279. Tyrrell KL, Citron DM, Leoncio ES, Merriam CV, Goldstein EJ. Evaluation of Cycloserine-Cefoxitin Fructose Agar (CCFA), CCFA with Horse Blood and Taurocholate, and Cycloserine-Cefoxitin Mannitol Broth with Taurocholate and Lysozyme for Recovery of *Clostridium difficile* Isolates from Faecal Samples. Journal of Clinical Microbiology. 2013;51(9):3094–6.

280. Perry JD, Asir K, Halimi D, Orenga S, Dale J, Payne M, et al. Evaluation of a chromogenic culture medium for isolation of *Clostridium difficile* within 24 hours. Journal of Clinical Microbiology. 2010;48(11):3852–8.

281. Hill K, Collins J, Wilson L, Perry J, Gould F. Comparison of two selective media for the recovery of *Clostridium difficile* from environmental surfaces. Journal of Hospital Infection. 2012;83:164–6.

282. Nottingham University Hospitals NHS Trust, [Internet]. [cited 2016 Sep 13]. Available from: https://www.nuh.nhs.uk/

283. Wilcox M, Fawley W, Parnell P. Value of lysozyme agar incorporation and alkaline thioglycollate exposure for the environmental recovery of *Clostridium difficile*. Journal of Hospital Infection. 2000;44(1):65–9.

284. Andrews S. FastQC: a quality control tool for high throughput sequence data [Internet]. 2010 [cited 2016 Nov 21]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

285. He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, et al. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. Proceedings of the National Academy of Sciences. 2010;107(16):7527–32.

286. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745–7.

287. Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW, Stoesser N, et al. Evolutionary history of the *Clostridium difficile* pathogenicity locus. Genome Biology and Evolution. 2014;6(1):36–52.

288. Brouwer MSM, Warburton PJ, Roberts AP, 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.

289. Hink T, Burnham C-AD, Dubberke ER. A systematic evaluation of methods to optimize culture-based recovery of *Clostridium difficile* from stool specimens. Anaerobe. 2013;19:39–43.

290. Richardson C, Kim P, Lee C, Bersenas A, Weese JS. Comparison of *Clostridium difficile* isolates from individuals with recurrent and single episode of infection. Anaerobe. 2015;33:105–8.

291. Ogielska M, Lanotte P, Le Brun C, Valentin AS, Garot D, Tellier A-C, et al. Emergence of community-acquired *Clostridium difficile* infection: the experience of a French hospital and review of the literature. International Journal of Infectious Diseases. 2015;37:36–41.

292. Gerding DN, Johnson S. Does infection with specific *Clostridium difficile* strains or clades influence clinical outcome? Clinical Infectious Diseases. 2013;133–6.

293. 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):4040–6.

294. Petrella LA, Sambol SP, Cheknis A, Nagaro K, Kean Y, Sears PS, et al. Decreased cure and increased recurrence rates for *Clostridium difficile* infection caused by the epidemic *C. difficile* BI strain. Clinical Infectious Diseases. 2012;55(3):351–7.

295. Gupta SB, Mehta V, Dubberke ER, Zhao X, Dorr MB, Guris D, et al. Antibodies to Toxin B Are Protective Against *Clostridium difficile* Infection Recurrence. Clinical Infectious Diseases. 2016;63(6):730–4.

296. Butt E, Foster JA, Keedwell E, Bell JE, Titball RW, Bhangu A, et al. Derivation and validation of a simple, accurate and robust prediction rule for risk of mortality in patients with *Clostridium difficile* infection. BMC Infectious Diseases. 2013;13(1):316.

297. Cassir N, Benamar S, La Scola B. *Clostridium butyricum*: from beneficial to a new emerging pathogen. Clinical Microbiology and Infection. 2016;22(1):37–45.

298. Kwok JSL, Ip M, Chan T-F, Lam W-Y, Tsui SKW. Draft Genome Sequence of *Clostridium butyricum* Strain NOR 33234, Isolated from an Elderly Patient with Diarrhea. Genome Announcements. 2014;2(6).

299. Cai D, Sorokin V, Lutwick L, Liu W, Dalal S, Sandhu K, et al. *C. glycolicum* as the sole cause of bacteremia in a patient with acute cholecystitis. Annals of clinical and laboratory science. 2012;42(2):162–4.

300. Hell M, Permoser M, Chmelizek G, Kern J, Maass M, Huhulescu S, et al. *Clostridium difficile* infection: monoclonal or polyclonal genesis? Infection. 2011;39(5):461–5.

301. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nature Reviews Microbiology. 2016;14:150–62.

302. Votintseva AA, Miller RR, Fung R, Knox K, Godwin H, Peto TEA, et al. Multiple-strain colonization in nasal carriers of *Staphylococcus aureus*. Journal of Clinical Microbiology. 2014;52(4):1192–200.

303. Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, et al. *Helicobacter pylori* genome evolution during human infection. Proceedings of the National Academy of Sciences. 2011;108(12):5033–8.

304. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of *Clostridium difficile* infection identified on whole-genome sequencing. The New England Journal of Medicine. 2013;369(13):1195–205.

305. Corver J, Bakker D, Brouwer MS, Harmanus C, Hensgens MP, Roberts AP, et al. Analysis of a *Clostridium difficile* PCR ribotype 078 100 kilobase island reveals the presence of a novel transposon, Tn6164. BMC Microbiology. 2012;12(1):130.

306. Alamro M, Bidmos FA, Chan H, Oldfield NJ, Newton E, Bai X, et al. Phase variation mediates reductions in expression of surface proteins during persistent meningococcal carriage. Infection and Immunity. 2014;82(6):2472–84.

307. Purcell EB, McKee RW, Bordeleau E, Burrus V, Tamayo R. Regulation of type IV pili contributes to surface behaviors of historical and epidemic strains of *Clostridium difficile*. Journal of bacteriology. 2016;198(3):565–77.

308. Didelot X, Eyre DW, Cule M, Ip CLC, Ansari MA, Griffiths D, et al. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. Genome Biology. 2012;13(12):R118.

309. Sun J, Deng Z, Yan A. Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. Biochemical and biophysical research communications. 2014;453(2):254–67.

310. Christensen GD, Baddour LM, Madison BM, Parisi JT, Abraham SN, Hasty DL, et al. Colonial morphology of staphylococci on Memphis agar: phase variation of slime production, resistance to beta-lactam antibiotics, and virulence. The Journal of Infectious Diseases. 1990;161(6):1153–69.

311. Kandler JL, Joseph SJ, Balthazar JT, Dhulipala V, Read TD, Jerse AE, et al. Phase-variable expression of lptA modulates the resistance of *Neisseria gonorrhoeae* to cationic antimicrobial peptides. Antimicrobial Agents and Chemotherapy. 2014;58(7):4230–3.

312. Haas W, Sublett J, Kaushal D, Tuomanen EI. Revising the role of the pneumococcal vex-vncRS locus in vancomycin tolerance. Journal of Bacteriology. 2004;186(24):8463–71.

313. Collery MM, Kuehne SA, McBride SM, Kelly ML, Monot M, Cockayne A, et al. What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of *Clostridium difficile* strain 630 and derivatives. Virulence. 2017;8(6):767–81.

Appendix One - Genetic differences in two R20291 strains resulting in different phenotypes.

These data were kindly provided by Jorge Montfort-Gardeazabal and form part of his Ph.D. thesis.

A number of SNPs have been identified in two different stock collections of *C. difficile* strain R20291 compared to the published sequence. These strains are one from our laboratory in the Synthetic Biology Research Centre, University of Nottingham (NM-R20291) and one from the London school of Hygiene and Tropical medicine laboratory (BW-R20291). Table A1-0-1 shows a list of SNPs that are either found in both strains (yellow) compared to the reference strain sequence or SNPs only found in NM-R20291 (green). Yellow SNPs may be due to mistakes in the reference sequence. SNPs found in NM-R20291 include a SNP in both *rsbW* (anti-sigma-B factor) and *vncR* (a two-component response regulator) resulting in missense mutations. Deletions in *rbsK* (a ribokinase belonging to the pfkB family) and *R20291_2456* (putative two-component sensor histidine kinase) resulting in transcription of incomplete proteins.

It is believed that the SNPs in NM-R20291 have led to a number of phenotypic changes as BW-R20291 is genetically more identical to the ancestral strain. The phenotypic changes identified in NM-R20291 include improved growth kinetics when grown in either BHIS broth or minimal media containing glucose, fructose and mannitol. NM-R20291 demonstrated a reduced motility and was found to be monotrichous by TEM compared to BW-R20291, which is peritrichously flagellated. NM-R20291 produces more toxin, sporulation occurs later and is able to form biofilms better than BW-R20291 (data from Jorge Montfort-Gardeazabal Ph.D. thesis, unpublished).

Table A1-0-1: Single nucleotide polymorphisms found in two R20291 strains compared to the reference sequence. Yellow indicates SNPs identified in both strains and green indicates SNPs identified in NM-R20291.

Reference Position	Type	Length	Reference	Allele	Frequency	Gene	Amino acid change
9694	SNP	1	G	T	99.34211	rsbW	Gly82Val
132924	Insertion	1	-	A	100		
132939	SNP	1	G	T	100	CDR20291_t40	
132955	SNP	1	C	A	100	CDR20291_t40	
132958	SNP	1	T	G	100	CDR20291_t40	
132959	SNP	1	T	C	100	CDR20291_t40	
143465	Insertion	1	-	A	97.56098	similar to 16S ribosomal RNA	
206403	Insertion	1	-	A	97.22222		
358264	Deletion	1	A	-	98.07692	rbsK	Met57fs
581484	Insertion	1	-	A	94.94382		
581491	Insertion	1	-	A	97.74011		
581498	Insertion	1	-	A	94.47514		
1564437	Deletion	1	A	-	96.79487		
1568676	SNP	1	C	A	100	CDR20291_1323	Gln138Lys
1578174	Deletion	1	T	-	96.32353		
1578209	Insertion	1	-	A	94.28571		
1592813	SNP	1	A	T	100		
1864421	Insertion	1	-	T	97.82609	CDR20291_1576	
1899604	Deletion	1	A	-	94.57364		
2077305	Deletion	1	C	-	100		
2120669	SNP	1	A	G	99.43182	vncR	Asp202Gly
2235743	Deletion	1	T	-	99.29078	CDR20291_1913	Lys81fs
2262066	Insertion	1	-	A	94.81481		
2264191	Deletion	1	T	-	98.78049		
2298116	Insertion	1	-	T	95.41985		
2361948	SNP	1	C	A	100		
2361961	Insertion	1	-	A	97.91667		
2367948	Insertion	1	-	T	97.94521		
2578164	Deletion	1	T	-	99.35484		
2674749	Deletion	1	T	-	99.35484		
2680792	Insertion	1	-	T	96.20253		
2772184	Deletion	1	T	-	97.59615	CDR20291_2368	
2881470	Deletion	1	T	-	98.21429	CDR20291_2456	Leu434fs
3077994	Deletion	1	A	-	98.77301		
3162105	Deletion	1	T	-	96.7033		
3361923	Deletion	1	A	-	97.00599		

Appendix Two – Bioinformatics Pipeline

In-house Whole Genome Analysis Pipeline

De novo sequence assembly -

$ canu -assemble -p *name* -d *output_directory* genomeSize=4.3m -pacbio-corrected *seq*.subreads.corrected.fastq

Options Evoked

canu [-correct | -trim | -assemble | -trim-assemble]
-p <assembly-prefix>
-d <assembly-directory>
genomeSize=<number>[g|m|k]
[other-options] [-pacbio-raw | -pacbio-corrected | -nanopore-raw | -nanopore-corrected] *fastq

Mapping sequence reads to *De novo* sequence assembly -

Indexing a reference file

$ bwa index *reference*.fasta

Aligning reads to reference

$ bwa mem -x pacbio -t 24 *reference*.fasta raw_subreads.fasta >output.sam

Options Evoked

-x Sequence data type
-t Number of threads
>output.sam output is SAM

Generating a Bam File

$ samtools view -S -b -o *file*.bam *file*.sam

Options Evoked

-b output is BAM
-S input is SAM
-o output file name

Sorting the Bam File

$ samtools sort *file*.bam *file*_sorted

Indexing the Bam file

$ samtools index *file*_sorted.bam

Appendix Three – Settings used for Fixed Ploidy Variant Detection

Table A3-0-1: Settings utilised for fixed ploidy variant detection in CLC genomics workbench V8.5.3

Fixed Ploidy Variant Detection	
Ploidy	1
Required variant probability (%)	90
Ignore positions with coverage above	100000
Minimum coverage	20
Minimum count	10
Minimum frequency (%)	90
Restrict calling to target regions	False
Ignore broken pairs	True
Ignore non-specific matches	Reads
Minimum read length	20
Base quality filter	True
Neighborhood radius	10
Minimum central quality	30
Minimum neighborhood quality	25
Read direction filter	True
Direction frequency (%)	5
Relative read direction filter	True
Significance (%)	1
Read position filter	True
Significance (%)	1
Remove pyro-error variants	False
In homopolymer regions with minimum length	3
With frequency below	0.8

Appendix Four – List of genes unique to *Clostridium difficile* strain VPI 10463

Table A4-0-1: List of genes unique to *C. difficile* strain VPI 10463 after comparison to 630 and R20291

Gene Name	Product
Ga0114281_113664	chromosomal replication initiator protein DnaA
Ga0114281_113570	putative transposase
Ga0114281_113569	putative transposase
Ga0114281_113566	glycoside/pentoside/hexuronide:cation symporter, GPH family
Ga0114281_113565	glycoside/pentoside/hexuronide:cation symporter, GPH family
Ga0114281_113456	aspartate aminotransferase
Ga0114281_113455	ribonucleoside-triphosphate reductase class III catalytic subunit
Ga0114281_113454	anaerobic ribonucleoside-triphosphate reductase activating protein
Ga0114281_113396	hypothetical protein
Ga0114281_113382	ParB/RepB/Spo0J family partition protein
Ga0114281_113381	hypothetical protein
Ga0114281_113380	hypothetical protein
Ga0114281_113379	Antirestriction protein (ArdA)
Ga0114281_113378	hypothetical protein
Ga0114281_113377	Protein of unknown function (DUF3801)
Ga0114281_113376	type IV secretion system protein VirD4
Ga0114281_113375	cAMP-binding domain of CRP or a regulatory subunit of cAMP-dependent protein kinases
Ga0114281_113374	hypothetical protein
Ga0114281_113373	RNA polymerase sigma factor, sigma-70 family
Ga0114281_113372	RNA polymerase sigma-70 factor, ECF subfamily
Ga0114281_113371	plasmid mobilization system relaxase
Ga0114281_113370	Cysteine-rich VLP
Ga0114281_113369	phage replisome organizer, putative, N-terminal region
Ga0114281_113368	phage DNA replication protein (predicted replicative helicase loader)
Ga0114281_113367	Site-specific DNA recombinase
Ga0114281_113366	hypothetical protein
Ga0114281_113365	stage V sporulation protein G
Ga0114281_113364	hypothetical protein
Ga0114281_113363	PrgI family protein
Ga0114281_113362	Type IV secretory pathway, VirB4 component
Ga0114281_113361	NlpC/P60 family protein
Ga0114281_113360	protein of unknown function (DUF4366)
Ga0114281_113359	hypothetical protein
Ga0114281_113358	hypothetical protein
Ga0114281_113357	hypothetical protein
Ga0114281_113356	Relaxase/Mobilisation nuclease domain-containing protein
Ga0114281_113355	DNA-binding transcriptional regulator, XRE-family HTH domain
Ga0114281_113354	His Kinase A (phospho-acceptor) domain-containing protein
Ga0114281_113353	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
Ga0114281_113352	bacitracin transport system ATP-binding protein
Ga0114281_113351	ABC-2 type transport system permease protein/bacitracin transport system permease protein

Ga0114281_113350	ABC-2 type transport system permease protein
Ga0114281_113349	conserved hypothetical protein TIGR01655
Ga0114281_113348	hypothetical protein
Ga0114281_113347	Site-specific DNA recombinase
Ga0114281_113344	hypothetical protein
Ga0114281_113342	putative ABC transport system ATP-binding protein
Ga0114281_113341	ABC-type transport system, involved in lipoprotein release, permease component
Ga0114281_113340	hypothetical protein
Ga0114281_113339	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
Ga0114281_113338	Signal transduction histidine kinase
Ga0114281_113337	putative ABC transport system ATP-binding protein
Ga0114281_113336	putative ABC transport system permease protein
Ga0114281_113335	hypothetical protein
Ga0114281_113298	Site-specific recombinase XerD
Ga0114281_113297	Helix-turn-helix domain-containing protein
Ga0114281_113296	hypothetical protein
Ga0114281_113295	DNA-binding transcriptional regulator, XRE-family HTH domain
Ga0114281_113294	putative transcriptional regulator
Ga0114281_113293	hypothetical protein
Ga0114281_113292	Arc-like DNA binding domain-containing protein
Ga0114281_113291	hypothetical protein
Ga0114281_113290	anti-repressor protein
Ga0114281_113289	hypothetical protein
Ga0114281_113288	DeoR-like helix-turn-helix domain-containing protein
Ga0114281_113287	hypothetical protein
Ga0114281_113286	hypothetical protein
Ga0114281_113285	virus Gp157
Ga0114281_113284	ERF superfamily protein
Ga0114281_113283	DnaD and phage-associated domain-containing protein
Ga0114281_113282	hypothetical protein
Ga0114281_113281	single-strand DNA-binding protein
Ga0114281_113280	phage uncharacterized protein TIGR01671
Ga0114281_113279	hypothetical protein
Ga0114281_113278	hypothetical protein
Ga0114281_113277	iron(III) transport system ATP-binding protein
Ga0114281_113276	hypothetical protein
Ga0114281_113275	hypothetical protein
Ga0114281_113274	hypothetical protein
Ga0114281_113273	hypothetical protein
Ga0114281_113272	hypothetical protein
Ga0114281_113271	3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS reductase)/FAD synthetase
Ga0114281_113270	hypothetical protein
Ga0114281_113269	hypothetical protein
Ga0114281_113268	hypothetical protein

Ga0114281_113267	hypothetical protein
Ga0114281_113266	hypothetical protein
Ga0114281_113265	Holliday junction resolvase RusA (prophage-encoded endonuclease)
Ga0114281_113264	ORF6N domain-containing protein
Ga0114281_113263	hypothetical protein
Ga0114281_113262	iron(III) transport system ATP-binding protein
Ga0114281_113261	hypothetical protein
Ga0114281_113260	iron(III) transport system ATP-binding protein
Ga0114281_113259	hypothetical protein
Ga0114281_113258	hypothetical protein
Ga0114281_113257	Uncharacterized protein YjcR
Ga0114281_113256	phage terminase large subunit
Ga0114281_113255	phage portal protein, SPP1 family
Ga0114281_113254	phage putative head morphogenesis protein, SPP1 gp7 family
Ga0114281_113253	hypothetical protein
Ga0114281_113252	Phage minor structural protein GP20
Ga0114281_113251	hypothetical protein
Ga0114281_113250	hypothetical protein
Ga0114281_113249	hypothetical protein
Ga0114281_113248	hypothetical protein
Ga0114281_113247	Bacteriophage HK97-gp10, putative tail-component
Ga0114281_113246	hypothetical protein
Ga0114281_113245	hypothetical protein
Ga0114281_113244	Phage tail sheath protein
Ga0114281_113243	Phage tail tube protein
Ga0114281_113242	hypothetical protein
Ga0114281_113241	Phage XkdN-like tail assembly chaperone protein, TAC
Ga0114281_113240	hypothetical protein
Ga0114281_113239	tape measure domain-containing protein
Ga0114281_113238	LysM domain-containing protein
Ga0114281_113237	NlpC/P60 family protein
Ga0114281_113236	Protein of unknown function (DUF2577)
Ga0114281_113235	Protein of unknown function (DUF2634)
Ga0114281_113234	Uncharacterized phage protein gp47/JayE
Ga0114281_113233	hypothetical protein (DUF2313)
Ga0114281_113232	Phage tail-collar fibre protein
Ga0114281_113231	Glycine rich protein
Ga0114281_113230	hypothetical protein
Ga0114281_113229	hypothetical protein
Ga0114281_113228	hypothetical protein
Ga0114281_113227	hypothetical protein
Ga0114281_113226	Haemolysin XhlA
Ga0114281_113225	Phage holin family Hol44, holin superfamily V
Ga0114281_113224	N-acetylmuramoyl-L-alanine amidase

Ga0114281_113223	hypothetical protein
Ga0114281_113222	hypothetical protein
Ga0114281_113221	hypothetical protein
Ga0114281_113220	Zn-dependent peptidase ImmA, M78 family
Ga0114281_113219	hypothetical protein
Ga0114281_113218	hypothetical protein
Ga0114281_113217	hypothetical protein
Ga0114281_113215	Predicted transcriptional regulator
Ga0114281_113214	Predicted transcriptional regulator
Ga0114281_113092	hypothetical protein
Ga0114281_113091	Phage integrase family protein
Ga0114281_113090	hypothetical protein
Ga0114281_113089	Helix-turn-helix
Ga0114281_113088	hypothetical protein
Ga0114281_113087	hypothetical protein
Ga0114281_113086	hypothetical protein
Ga0114281_113085	hypothetical protein
Ga0114281_113084	hypothetical protein
Ga0114281_113083	hypothetical protein
Ga0114281_113082	toxin secretion/phage lysis holin
Ga0114281_113081	N-acetylmuramoyl-L-alanine amidase
Ga0114281_113080	hypothetical protein
Ga0114281_113079	hypothetical protein
Ga0114281_113078	hypothetical protein
Ga0114281_113077	type I restriction enzyme M protein
Ga0114281_113076	type I restriction enzyme, S subunit
Ga0114281_113075	hypothetical protein
Ga0114281_113074	hypothetical protein
Ga0114281_113073	hypothetical protein
Ga0114281_113072	type I restriction enzyme, R subunit
Ga0114281_113071	hypothetical protein
Ga0114281_113070	hypothetical protein
Ga0114281_113069	DNA repair protein RadC
Ga0114281_113068	protein of unknown function (DUF960)
Ga0114281_113067	hypothetical protein
Ga0114281_113066	hypothetical protein
Ga0114281_113065	hypothetical protein
Ga0114281_113064	hypothetical protein
Ga0114281_113063	hypothetical protein
Ga0114281_113062	hypothetical protein
Ga0114281_113061	hypothetical protein
Ga0114281_113060	hypothetical protein
Ga0114281_113059	hypothetical protein
Ga0114281_113058	hypothetical protein

Ga0114281_113057	hypothetical protein
Ga0114281_113056	hypothetical protein
Ga0114281_113055	hypothetical protein
Ga0114281_113054	hypothetical protein
Ga0114281_113053	hypothetical protein
Ga0114281_113052	hypothetical protein
Ga0114281_113051	hypothetical protein
Ga0114281_113050	hypothetical protein
Ga0114281_113049	site-specific DNA recombinase
Ga0114281_113048	protein of unknown function (DUF4297)
Ga0114281_113047	hypothetical protein
Ga0114281_113046	hypothetical protein
Ga0114281_113045	hypothetical protein
Ga0114281_113044	transcriptional regulator, LytTR family
Ga0114281_113043	hypothetical protein
Ga0114281_113042	hypothetical protein
Ga0114281_113041	hypothetical protein
Ga0114281_113040	Protein of unknown function (DUF3139)
Ga0114281_113039	putative ABC transport system permease protein
Ga0114281_113038	putative ABC transport system ATP-binding protein
Ga0114281_113029	iron(III) transport system permease protein
Ga0114281_112987	6-phospho-beta-glucosidase
Ga0114281_112964	type I restriction enzyme, S subunit
Ga0114281_112963	hypothetical protein
Ga0114281_112962	Predicted nucleotidyltransferase
Ga0114281_112779	Ala-tRNA(Pro) deacylase
Ga0114281_112708	hypothetical protein
Ga0114281_112447	hypothetical protein
Ga0114281_112446	Putative cell wall binding repeat 2
Ga0114281_112445	hypothetical protein
Ga0114281_112443	hypothetical protein
Ga0114281_112328	hydrophobic/amphiphilic exporter-1, HAE1 family
Ga0114281_112308	Putative peptidoglycan binding domain-containing protein
Ga0114281_112307	transcriptional repressor, CopY family
Ga0114281_112306	Signal transducer regulating beta-lactamase production, contains metallopeptidase domain
Ga0114281_112305	Putative peptidoglycan binding domain-containing protein
Ga0114281_112229	Response regulator receiver domain-containing protein
Ga0114281_112226	ABC transporter
Ga0114281_112214	putative transposase
Ga0114281_112213	hypothetical protein
Ga0114281_112212	hypothetical protein
Ga0114281_112211	hypothetical protein
Ga0114281_112210	hypothetical protein
Ga0114281_112209	hypothetical protein

Ga0114281_112208	BRO family, N-terminal domain
Ga0114281_112207	hypothetical protein
Ga0114281_112206	hypothetical protein
Ga0114281_112205	hypothetical protein
Ga0114281_112204	hypothetical protein
Ga0114281_112203	hypothetical protein
Ga0114281_112202	hypothetical protein
Ga0114281_112201	putative transcriptional regulator
Ga0114281_112200	Predicted transcriptional regulator
Ga0114281_112121	hypothetical protein
Ga0114281_112077	Sodium:dicarboxylate symporter family protein
Ga0114281_112032	hypothetical protein
Ga0114281_112031	hypothetical protein
Ga0114281_112006	MatE protein
Ga0114281_112005	putative efflux protein, MATE family
Ga0114281_111904	transcriptional regulator, TetR family
Ga0114281_111903	Pimeloyl-ACP methyl ester carboxylesterase
Ga0114281_111902	hypothetical protein
Ga0114281_111901	hypothetical protein
Ga0114281_111900	DNA-binding transcriptional regulator, MarR family
Ga0114281_111899	putative efflux protein, MATE family
Ga0114281_111898	hypothetical protein
Ga0114281_111897	hypothetical protein
Ga0114281_111875	hypothetical protein
Ga0114281_111869	hypothetical protein
Ga0114281_111771	hypothetical protein
Ga0114281_111770	ABC-2 type transport system ATP-binding protein
Ga0114281_111769	hypothetical protein
Ga0114281_111768	RNA polymerase sigma-70 factor, ECF subfamily
Ga0114281_111767	transcriptional regulator, PadR family
Ga0114281_111766	protein of unknown function (DUF303)
Ga0114281_111765	hypothetical protein
Ga0114281_111764	transcriptional regulator, AraC family
Ga0114281_111763	Site-specific DNA recombinase
Ga0114281_111760	conserved repeat domain-containing protein
Ga0114281_111715	Site-specific recombinase XerD
Ga0114281_111714	Site-specific recombinase XerD
Ga0114281_111713	hypothetical protein
Ga0114281_111569	Transcriptional regulatory protein, C terminal
Ga0114281_111568	Restriction endonuclease
Ga0114281_111567	hypothetical protein
Ga0114281_111566	hypothetical protein
Ga0114281_111565	dGTPase
Ga0114281_111564	N-acetylmuramoyl-L-alanine amidase

Ga0114281_111563	N-acetylmuramoyl-L-alanine amidase
Ga0114281_111562	toxin secretion/phage lysis holin
Ga0114281_111561	hypothetical protein
Ga0114281_111560	hypothetical protein
Ga0114281_111559	hypothetical protein
Ga0114281_111558	phage uncharacterized protein, XkdX family
Ga0114281_111557	hypothetical protein
Ga0114281_111499	hypothetical protein
Ga0114281_111453	Lysophospholipase, alpha-beta hydrolase superfamily
Ga0114281_111449	hypothetical protein
Ga0114281_111448	hypothetical protein
Ga0114281_111447	Putative cell wall binding repeat 2
Ga0114281_111443	Haemolysin XhlA
Ga0114281_111442	hypothetical protein
Ga0114281_111441	hypothetical protein
Ga0114281_111440	hypothetical protein
Ga0114281_111439	hypothetical protein
Ga0114281_111438	hypothetical protein
Ga0114281_111437	hypothetical protein
Ga0114281_111436	hypothetical protein
Ga0114281_111435	hypothetical protein
Ga0114281_111432	Glycine rich protein
Ga0114281_111431	Phage tail-collar fibre protein
Ga0114281_111430	hypothetical protein (DUF2313)
Ga0114281_111429	Uncharacterized phage protein gp47/JayE
Ga0114281_111428	Protein of unknown function (DUF2634)
Ga0114281_111427	Protein of unknown function (DUF2577)
Ga0114281_111426	NlpC/P60 family protein
Ga0114281_111424	tape measure domain-containing protein
Ga0114281_111423	hypothetical protein
Ga0114281_111422	hypothetical protein
Ga0114281_111418	hypothetical protein
Ga0114281_111416	Bacteriophage HK97-gp10, putative tail-component
Ga0114281_111415	hypothetical protein
Ga0114281_111414	hypothetical protein
Ga0114281_111413	hypothetical protein
Ga0114281_111410	hypothetical protein
Ga0114281_111409	phage putative head morphogenesis protein, SPP1 gp7 family
Ga0114281_111408	phage portal protein, SPP1 family
Ga0114281_111407	phage terminase large subunit
Ga0114281_111406	Uncharacterized protein YjcR
Ga0114281_111405	hypothetical protein
Ga0114281_111404	hypothetical protein
Ga0114281_111403	protein of unknown function (DUF4868)

Ga0114281_111402	KTSC domain-containing protein
Ga0114281_111401	hypothetical protein
Ga0114281_111400	hypothetical protein
Ga0114281_111397	VRR-NUC domain-containing protein
Ga0114281_111396	hypothetical protein
Ga0114281_111391	hypothetical protein
Ga0114281_111388	hypothetical protein
Ga0114281_111387	Helix-turn-helix domain-containing protein
Ga0114281_111386	Cro/C1-type HTH DNA-binding domain-containing protein
Ga0114281_111385	hypothetical protein
Ga0114281_111384	hypothetical protein
Ga0114281_111383	Phage antirepressor protein YoqD, KilAC domain
Ga0114281_111382	hypothetical protein
Ga0114281_111381	hypothetical protein
Ga0114281_111380	hypothetical protein
Ga0114281_111379	Arc-like DNA binding domain-containing protein
Ga0114281_111378	phage regulatory protein, rha family
Ga0114281_111377	putative transcriptional regulator
Ga0114281_111376	Transcriptional regulator, contains XRE-family HTH domain
Ga0114281_111375	SIR2-like domain-containing protein
Ga0114281_111267	putative ABC transport system permease protein
Ga0114281_111266	putative ABC transport system permease protein
Ga0114281_111262	hypothetical protein
Ga0114281_111261	GntR family transcriptional regulator
Ga0114281_111121	PAS domain S-box-containing protein/diguanylate cyclase (GGDEF) domain-containing protein
Ga0114281_111024	hypothetical protein
Ga0114281_111020	hypothetical protein
Ga0114281_111018	transposon Tn916 excisionase
Ga0114281_111017	Helix-turn-helix domain-containing protein
Ga0114281_111016	RNA polymerase sigma factor, sigma-70 family
Ga0114281_111015	hypothetical protein
Ga0114281_111014	protein of unknown function (DUF4177)
Ga0114281_111013	Glycopeptide antibiotics resistance protein
Ga0114281_111012	Cyclopropane fatty-acyl-phospholipid synthase
Ga0114281_111011	transcriptional regulator, TetR family
Ga0114281_111010	hypothetical protein
Ga0114281_111003	Conjugative transposon protein TcpC
Ga0114281_111002	NlpC/P60 family protein
Ga0114281_111001	AraC family transcriptional regulator
Ga0114281_111000	AraC family transcriptional regulator
Ga0114281_11999	ribosomal-protein-alanine N-acetyltransferase
Ga0114281_11998	tRNA-Thr(GGU) m(6)t(6)A37 methyltransferase TsaA
Ga0114281_11997	DNA-binding transcriptional regulator, MerR family
Ga0114281_11996	Conjugative transposon protein TcpC

Ga0114281_11995	Glycopeptide antibiotics resistance protein
Ga0114281_11994	MFS transporter, DHA1 family, multidrug resistance protein
Ga0114281_11993	pyruvate, water dikinase
Ga0114281_11992	transcriptional regulator, TetR family
Ga0114281_11987	TcpE family protein
Ga0114281_11985	Antirestriction protein (ArdA)
Ga0114281_11984	hypothetical protein
Ga0114281_11983	Acetyltransferase (GNAT) domain-containing protein
Ga0114281_11981	Protein of unknown function (DUF3789)
Ga0114281_11976	hypothetical protein
Ga0114281_11975	Protein of unknown function (DUF3788)
Ga0114281_11974	Catechol 2,3-dioxygenase
Ga0114281_11973	protein of unknown function (DUF961)
Ga0114281_11967	Helix-turn-helix
Ga0114281_11966	Peptidase propeptide and YPEB domain-containing protein
Ga0114281_11965	Putative cell wall binding repeat 2
Ga0114281_11964	hypothetical protein
Ga0114281_11539	Acetyltransferase (isoleucine patch superfamily)
Ga0114281_11538	DNA-binding transcriptional regulator, LysR family
Ga0114281_11457	hypothetical protein
Ga0114281_11435	hypothetical protein
Ga0114281_11399	hypothetical protein
Ga0114281_11394	hypothetical protein
Ga0114281_11329	Putative cell wall binding repeat 2
Ga0114281_11328	hypothetical protein
Ga0114281_11327	Fic family protein
Ga0114281_11309	ABC-2 type transport system ATP-binding protein
Ga0114281_11308	hypothetical protein
Ga0114281_11300	putative transposase
Ga0114281_11291	regulatory protein, gntR family
Ga0114281_11290	diguanylate cyclase (GGDEF) domain-containing protein
Ga0114281_11289	hypothetical protein
Ga0114281_11288	His Kinase A (phospho-acceptor) domain-containing protein
Ga0114281_11273	Site-specific DNA recombinase
Ga0114281_11272	hypothetical protein
Ga0114281_11271	Helix-turn-helix domain-containing protein
Ga0114281_11270	Sigma-70, region 4
Ga0114281_11269	ATP-binding cassette, subfamily B
Ga0114281_11268	ATP-binding cassette, subfamily B
Ga0114281_11267	AraC-type DNA-binding protein
Ga0114281_11266	energy-coupling factor transport system ATP-binding protein
Ga0114281_11265	energy-coupling factor transport system permease protein
Ga0114281_11264	energy-coupling factor transport system substrate-specific component
Ga0114281_11263	thiazolinyl imide reductase

Ga0114281_11262	hypothetical protein
Ga0114281_11261	pyochelin synthetase
Ga0114281_11260	yersiniabactin nonribosomal peptide synthetase
Ga0114281_11259	yersiniabactin salicyl-AMP ligase
Ga0114281_11258	4'-phosphopantetheinyl transferase
Ga0114281_11257	Surfactin synthase thioesterase subunit
Ga0114281_11256	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase
Ga0114281_11255	salicylate synthetase/yersiniabactin salicyl-AMP ligase
Ga0114281_11254	iron (metal) dependent repressor, DtxR family
Ga0114281_11253	Conjugative transposon protein TcpC
Ga0114281_11252	Cell wall-associated hydrolase, NlpC family
Ga0114281_11251	hypothetical protein
Ga0114281_11250	AAA-like domain-containing protein
Ga0114281_11249	TcpE family protein
Ga0114281_11248	hypothetical protein
Ga0114281_11247	hypothetical protein
Ga0114281_11246	hypothetical protein
Ga0114281_11245	Antirestriction protein
Ga0114281_11244	hypothetical protein
Ga0114281_11243	hypothetical protein
Ga0114281_11242	hypothetical protein
Ga0114281_11241	phage replication initiation protein
Ga0114281_11240	FtsK/SpoIIIE family protein
Ga0114281_11239	protein of unknown function (DUF961)
Ga0114281_11238	protein of unknown function (DUF961)
Ga0114281_11237	LPXTG-motif cell wall anchor domain-containing protein
Ga0114281_11236	TatD DNase family protein
Ga0114281_11235	hypothetical protein
Ga0114281_11234	hypothetical protein
Ga0114281_11233	KAP family P-loop domain-containing protein
Ga0114281_11228	hypothetical protein
Ga0114281_11105	flagellar FliJ protein
Ga0114281_11104	flagellar FliJ protein
Ga0114281_1194	WxcM-like, C-terminal
Ga0114281_1144	Predicted transcriptional regulator
Ga0114281_1143	Predicted transcriptional regulator
Ga0114281_1141	hypothetical protein
Ga0114281_1140	hypothetical protein
Ga0114281_1139	Zn-dependent peptidase ImmA, M78 family
Ga0114281_1138	hypothetical protein
Ga0114281_1137	hypothetical protein
Ga0114281_1136	hypothetical protein
Ga0114281_1135	N-acetylmuramoyl-L-alanine amidase
Ga0114281_1134	holin, TcdE family

Ga0114281_1133	hypothetical protein
Ga0114281_1132	hypothetical protein
Ga0114281_1131	hypothetical protein
Ga0114281_1130	Glycine rich protein
Ga0114281_1129	Phage tail-collar fibre protein
Ga0114281_1128	hypothetical protein (DUF2313)
Ga0114281_1127	Uncharacterized phage protein gp47/JayE
Ga0114281_1126	Protein of unknown function (DUF2634)
Ga0114281_1125	Protein of unknown function (DUF2577)
Ga0114281_1124	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase
Ga0114281_1123	LysM domain-containing protein
Ga0114281_1122	tape measure domain-containing protein
Ga0114281_1121	protein of unknown function (DUF4428)
Ga0114281_1120	Phage XkdN-like tail assembly chaperone protein, TAC
Ga0114281_1119	hypothetical protein
Ga0114281_1118	Phage tail tube protein
Ga0114281_1117	Phage tail sheath protein
Ga0114281_1116	hypothetical protein
Ga0114281_1115	hypothetical protein
Ga0114281_1114	Bacteriophage HK97-gp10, putative tail-component
Ga0114281_1113	hypothetical protein
Ga0114281_1112	hypothetical protein
Ga0114281_1111	hypothetical protein
Ga0114281_1110	hypothetical protein
Ga0114281_119	Phage minor structural protein GP20
Ga0114281_118	hypothetical protein
Ga0114281_117	phage putative head morphogenesis protein, SPP1 gp7 family
Ga0114281_116	phage portal protein, SPP1 family
Ga0114281_115	phage terminase large subunit
Ga0114281_114	Uncharacterized protein YjcR
Ga0114281_113	hypothetical protein
Ga0114281_112	hypothetical protein
Ga0114281_111	iron(III) transport system ATP-binding protein
Ga0114281_114103	hypothetical protein
Ga0114281_114102	iron(III) transport system ATP-binding protein
Ga0114281_114101	hypothetical protein
Ga0114281_114100	ORF6N domain-containing protein
Ga0114281_114099	Holliday junction resolvase RusA (prophage-encoded endonuclease)
Ga0114281_114098	hypothetical protein
Ga0114281_114097	hypothetical protein
Ga0114281_114096	hypothetical protein
Ga0114281_114095	hypothetical protein
Ga0114281_114094	hypothetical protein
Ga0114281_114093	3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS reductase)/FAD synthetase

Ga0114281_114092	hypothetical protein
Ga0114281_114091	hypothetical protein
Ga0114281_114090	hypothetical protein
Ga0114281_114089	hypothetical protein
Ga0114281_114088	hypothetical protein
Ga0114281_114087	iron(III) transport system ATP-binding protein
Ga0114281_114086	hypothetical protein
Ga0114281_114085	hypothetical protein
Ga0114281_114084	phage uncharacterized protein TIGR01671
Ga0114281_114083	DnaD and phage-associated domain-containing protein
Ga0114281_114082	recombination protein RecT
Ga0114281_114081	putative phage-type endonuclease
Ga0114281_114080	hypothetical protein
Ga0114281_114079	hypothetical protein
Ga0114281_114078	Arc-like DNA binding domain-containing protein
Ga0114281_114077	Arc-like DNA binding domain-containing protein
Ga0114281_114076	hypothetical protein
Ga0114281_114075	DNA binding domain-containing protein, excisionase family
Ga0114281_114074	ORF6N domain-containing protein
Ga0114281_114073	putative transcriptional regulator
Ga0114281_114072	Helix-turn-helix
Ga0114281_114071	hypothetical protein
Ga0114281_114070	SAP domain-containing protein
Ga0114281_114069	T5orf172 domain-containing protein
Ga0114281_114068	hypothetical protein
Ga0114281_114067	Predicted nuclease of the RNAse H fold, HicB family
Ga0114281_114066	Site-specific recombinase XerD
Ga0114281_113834	iron(III) transport system ATP-binding protein
Ga0114281_113726	PAS domain-containing protein
Ga0114281_113719	CAAX protease self-immunity
Ga0114281_113718	DNA-binding transcriptional regulator, MerR family

Appendix Five – Table of Clinical Information for Patient Cohort

Table A5-0-1: Clinical information gathered for all patients included in the recurrence study.

Study Number	Registered Inpatient / Community	Relapse sample	Ribotype	Metronidazole MIC[a] (mg/L)	Vancomycin MIC[a] (mg/L)	Treatment	Antibiotic use	PPI	CRP[b] (Normal range 1-10 mg/L)
01-01	Hospital Acquired	No	015	0.5	0.5	Vancomycin	Yes	Yes	131
02-01	Community Acquired	No	070	1	0.5	Vancomycin	Yes	No	124
03-01	Hospital Acquired	No	015	0.5	0.5	Vancomycin	Yes	No	49
04-01(1)	Hospital Acquired	No	031	0.5	0.25	Metronidazole	Yes	Yes	111
04-01(2)	Hospital Acquired	No	224	1	0.25				
05-01	Community Acquired	No	005	1	0.25	Metronidazole	Unknown	Unknown	Unknown
05-02	Community Acquired	Yes	005	1	0.25	Vancomycin	Unknown	Unknown	Unknown
06-01	Community Acquired	No	023	1	0.12	Metronidazole	Unknown	Unknown	Unknown
07-01	Hospital Acquired	No	023	0.5	0.12	Vancomycin	Yes	No	253
08-01(1)	Hospital Acquired	No	002	1	0.5	Metronidazole	No	No	43
08-01(2)	Hospital Acquired	No	062	1	0.5				
09-01	Hospital Acquired	No	005	1	0.5	Vancomycin	No	No	22
10-01(1)	Hospital Acquired	No	062	1	0.5	Vancomycin	Yes	No	367
10-01(2)	Hospital Acquired	No	023	1	0.25				
10-02	Hospital Acquired	Yes	023	0.25	1	Vancomycin	No	No	244
11-01	Hospital Acquired	No	002	0.5	0.5	Vancomycin	No	No	26
12-01	Hospital Acquired	No	023	1	0.25	Metronidazole	No	Yes	28
12-02	Hospital Acquired	Yes	023	0.12	1	Vancomycin	No	Yes	171
13-01	Hospital Acquired	No	013	0.5	0.5	Vancomycin	Yes	No	79
14-01	Hospital Acquired	No	020	0.5	0.5	Vancomycin	Yes	Yes	135
15-01	Hospital Acquired	No	002	1	0.5	Metronidazole	No	Yes	95
16-01	Hospital Acquired	No	017	0.5	0.5	Unknown	Yes	No	Unknown
17-01	Hospital Acquired	No	220	1	1	Vancomycin	No	No	198
18-01	Hospital Acquired	No	078	0.5	0.25	Metronidazole	Yes	Yes	267
19-01	Hospital Acquired	No	015	1	0.25	Vancomycin	No	No	14

Study Number	Registered Inpatient / Community	Relapse sample	Ribotype	Metronidazole MIC[a] (mg/L)	Vancomycin MIC[a] (mg/L)	Treatment	Antibiotic use	PPI	CRP[b] (Normal range 1-10 mg/L)
19-02	Community Acquired	Yes	015	0.5	0.5	Metronidazole	Unknown	Unknown	Unknown
20-01	Hospital Acquired	No	002	0.25	0.5	Metronidazole	No	No	12
21-01	Hospital Acquired	No	002	0.25	0.5	Multiple	Yes	No	171
22-01(1)	Hospital Acquired	No	032	0.25	0.5	Metronidazole	Yes	Yes	Unknown
22-01(2)	Hospital Acquired	No	050	0.25	1				
23-01	Hospital Acquired	No	014	0.5	0.5	None	Yes	No	125
24-01	Hospital Acquired	No	023	0.12	0.5	Vancomycin	Yes	No	31
25-01	Hospital Acquired	No	056	0.5	1	Metronidazole	Yes	Yes	13
26-01	Hospital Acquired	No	014	0.5	0.5	Vancomycin	No	No	25
27-01	Hospital Acquired	No	013	0.5	1	Vancomycin	No	No	118
28-01	Hospital Acquired	No	027	1	1	Vancomycin	No	No	Unknown
29-01	Hospital Acquired	No	365	0.5	1	Vancomycin	No	Yes	Unknown
30-01(1)	Hospital Acquired	No	001	1	0.5	Multiple	No	No	109
30-01(2)	Hospital Acquired	No	005	0.25	1				
31-01	Hospital Acquired	No	296	0.5	0.5	Vancomycin	No	No	389
32-01	Hospital Acquired	No	181	0.25	0.5	Metronidazole	No	No	Unknown
33-01	Hospital Acquired	No	003	0.5	0.5	Metronidazole	No	Yes	181
34-01	Hospital Acquired	No	001	0.5	1	None	Unknown	Unknown	155
35-01	Hospital Acquired	No	070	0.5	1	Vancomycin	Yes	No	272
36-01	Community Acquired	No	015	0.25	0.5	Metronidazole	Yes	No	Unknown
37-01	Community Acquired	No	023	0.25	1	Metronidazole	Unknown	Unknown	Unknown
38-01	Hospital Acquired	No	078	0.5	2	Vancomycin	Yes	No	367
39-01	Hospital Acquired	No	001	0.25	1	Multiple	No	Yes	81
39-02(1)	Community Acquired	Yes	001	0.5	1	Fidaxomicin			
39-02(2)	Community Acquired	Yes	070	0.25	0.5				
40-01(1)	Hospital Acquired	No	014	0.5	1	Vancomycin	Yes	Yes	163

Study Number	Registered Inpatient / Community	Relapse sample	Ribotype	Metronidazole MIC[a] (mg/L)	Vancomycin MIC[a] (mg/L)	Treatment	Antibiotic use	PPI	CRP[b] (Normal range 1-10 mg/L)
40-01(2)	Hospital Acquired	No	035	0.5	1				
41-01	Hospital Acquired	No	002	0.5	1	Multiple	No	No	37
42-01	Hospital Acquired	No	015	0.25	1	Multiple	No	No	69
42-02	Community Acquired	Yes	015	0.5	1	Fidaxomicin	Unknown	Unknown	Unknown
43-01	Community Acquired	No	029	0.5	0.5	Metronidazole	Unknown	Unknown	Unknown
44-01	Hospital Acquired	No	078	0.5	0.5	Vancomycin	Yes	Yes	88
45-01	Community Acquired	No	015	0.5	0.5	Metronidazole	Unknown	Unknown	Unknown
45-02	Hospital Acquired	Yes	015	0.5	0.5	Fidaxomicin	No	Yes	90
46-01	Hospital Acquired	No	479	0.5	0.5	Unknown	Unknown	Yes	68
47-01	Community Acquired	No	015	0.25	1	Metronidazole	No	No	93
48-01	Hospital Acquired	No	005	0.5	1	Metronidazole	No	No	12
49-01	Hospital Acquired	No	021	0.5	1	Vancomycin	Yes	No	120
50-01	Hospital Acquired	No	018	0.5	1	None	Yes	No	113
51-01	Hospital Acquired	No	027	1	1	Metronidazole	Yes	No	Unknown
52-01(1)	Hospital Acquired	No	002	0.5	1	Fidaxomicin	Unknown	Unknown	Unknown
52-01(2)	Hospital Acquired	No	005	0.25	1				
53-01	Community Acquired	No	023	0.12	1	Unknown	Unknown	Unknown	Unknown
54-01	Hospital Acquired	No	023	0.25	1	Multiple	Yes	No	5
55-01	Hospital Acquired	No	054	0.25	0.5	Vancomycin	No	No	Unknown
55-02	Hospital Acquired	Yes	054	0.5	1	None	Unknown	Unknown	Unknown
56-01	Hospital Acquired	No	014	0.5	0.5	Vancomycin	No	No	233
57-01	Hospital Acquired	No	076	0.5	1	Metronidazole	No	No	5
58-01	Hospital Acquired	No	076	0.5	1	Vancomycin	Yes	No	266
59-01	Hospital Acquired	No	015	0.5	0.5	Vancomycin	No	No	55
60-01	Hospital Acquired	No	002	0.5	1	Multiple	No	Yes	275
61-01	Hospital Acquired	No	076	0.5	1	Multiple	No	Yes	228

Study Number	Registered Inpatient / Community	Relapse sample	Ribotype	Metronidazole MIC[a] (mg/L)	Vancomycin MIC[a] (mg/L)	Treatment	Antibiotic use	PPI	CRP[b] (Normal range 1-10 mg/L)
61-02	Hospital Acquired	Yes	076	1	0.5	Fidaxomicin	Unknown	Unknown	Unknown
61-03	Hospital Acquired	Yes	076	0.5	1	Fidaxomicin	Unknown	Unknown	Unknown
61-04	Hospital Acquired	Yes	076	0.5	1	Vancomycin	No	No	169
61-05	Community Acquired	Yes	076	0.5	1	Vancomycin	No	No	Unknown
62-01	Hospital Acquired	No	015	0.25	1	Vancomycin	No	Yes	144
63-01(1)	Hospital Acquired	No	012	0.5	1	Fidaxomicin	Yes	No	161
63-01(2)	Hospital Acquired	No	020	0.5	0.5				
64-01	Hospital Acquired	No	255	0.5	0.5	Vancomycin	No	No	25
65-01	Hospital Acquired	No	050	0.25	1	Vancomycin	Yes	No	23
66-01(1)	Hospital Acquired	No	001	0.5	1	Unknown	No	No	Unknown
66-01(2)	Hospital Acquired	No	026	0.25	0.5				
67-01	Hospital Acquired	No	014	0.5	0.5	Unknown	Unknown	Unknown	Unknown
68-01	Hospital Acquired	No	007	1	0.5	Fidaxomicin	Unknown	Unknown	Unknown
69-01	Hospital Acquired	No	054	0.5	1	Multiple	No	No	308
70-01	Hospital Acquired	No	005	0.25	0.5	Fidaxomicin	Yes	No	12
71-01	Hospital Acquired	No	011	1	0.5	Vancomycin	No	No	143
72-01	Hospital Acquired	No	220	0.25	1	Metronidazole	Yes	No	19
73-01	Hospital Acquired	No	015	0.5	1	None	No	No	18
74-01	Hospital Acquired	No	054	1	1	Metronidazole	No	No	167
75-01	Hospital Acquired	No	001	0.5	0.5	Metronidazole	No	No	124
76-01(1)	Hospital Acquired	No	012	0.5	1	Metronidazole	Yes	Yes	5
76-01(2)	Hospital Acquired	No	116	1	0.5				
77-01(1)	Hospital Acquired	No	014	0.5	1	Metronidazole	No	No	113
77-01(2)	Hospital Acquired	No	009	0.5	0.5				
78-01	Hospital Acquired	No	023	0.5	1	None	Yes	No	23
79-01	Hospital Acquired	No	023	0.5	1	None	Unknown	Unknown	286

Study Number	Registered Inpatient / Community	Relapse sample	Ribotype	Metronidazole MIC[a] (mg/L)	Vancomycin MIC[a] (mg/L)	Treatment	Antibiotic use	PPI	CRP[b] (Normal range 1-10 mg/L)
80-01	Hospital Acquired	No	014	0.5	1	Unknown	No	No	Unknown
81-01	Hospital Acquired	No	020	0.5	1	Metronidazole	No	No	114
81-02	Hospital Acquired	Yes	020	0.5	1	Fidaxomicin	No	No	43
82-01	Hospital Acquired	No	020	1	0.5	Unknown			
83-01(1)	Hospital Acquired	No	026	0.5	1	Metronidazole	Yes	Yes	62
83-01(2)	Hospital Acquired	No	002	0.5	1				
84-01	Hospital Acquired	No	015	0.5	1	Vancomycin	Yes	No	16
85-01	Hospital Acquired	No	015	0.5	1	Metronidazole	Yes	Yes	104
85-02	Community Acquired	Yes	015	1	0.5	Unknown	Unknown	Unknown	Unknown
86-01	Hospital Acquired	No	078	0.5	1	Vancomycin	No	No	80
87-01	Hospital Acquired	No	351	1	1	Metronidazole	No	Yes	284
88-01(1)	Hospital Acquired	No	014	1	1	Vancomycin	Yes	Yes	56
88-01(2)	Hospital Acquired	No	054	1	0.5				
89-01	Hospital Acquired	No	012	0.5	1	Metronidazole	No	No	74
90-01	Hospital Acquired	No	078	0.5	0.5	Metronidazole	Yes	No	Unknown
91-01	Hospital Acquired	No	087	0.5	1	Metronidazole	No	No	15
92-01	Hospital Acquired	No	027	1	1	Unknown	Unknown	Unknown	258
93-01	Hospital Acquired	No	078	0.5	0.5	Fidaxomicin	Yes	Unknown	161
94-01	Community Acquired	No	002	0.5	1	Vancomycin	Unknown	Unknown	Unknown
95-01	Hospital Acquired	No	078	0.25	1	Unknown	Unknown	Unknown	60
96-01	Hospital Acquired	No	001	1	1	Metronidazole	No	No	11
96-02	Hospital Acquired	Yes	001	0.5	0.5	None	No	No	10
96-03	Hospital Acquired	Yes	001	0.5	0.5	None	No	No	32
96-04	Community Acquired	Yes	001	0.5	1	None	Unknown	Unknown	Unknown
97-01	Hospital Acquired	No	078	0.5	0.5	Unknown	Unknown	Unknown	Unknown
98-01	Community Acquired	No	078	0.25	0.5	Vancomycin	Unknown	Unknown	Unknown

Study Number	Registered Inpatient / Community	Relapse sample	Ribotype	Metronidazole MIC[a] (mg/L)	Vancomycin MIC[a] (mg/L)	Treatment	Antibiotic use	PPI	CRP[b] (Normal range 1-10 mg/L)
99-01	Community Acquired	No	062	0.5	0.5	Metronidazole	Unknown	Unknown	Unknown
100-01	Hospital Acquired	No	064	1	1	Metronidazole	No	No	23

[a]MIC – Minimum Inhibitory Concentration, values of >2 mg/L indicate the strain is resistant.
[b]CRP – C-reactive protein, normal range is between 1-10 mg/L.