Isolation and Characterisation of Four Novel Bacteriophages Infecting Clinically Relevant PCR Ribotypes of *Clostridium difficile*

Michaella Whittle

Thesis Submitted to the University of Nottingham for the Degree of Doctor of Philosophy

September 2017

Declaration

The work presented in this thesis is my own work, unless otherwise stated. No part has been submitted to another institute of learning or the University of Nottingham for another degree.

Michaella Whittle

September 2017

Abstract

Clostridium difficile infection is the leading cause of antibiotic-associated diarrhoea in North America and Europe. The disease is most commonly treated with only a handful of antibiotics, amongst those vancomycin, a so-called last resort therapy. Growing concerns over antibiotic resistance have led to the desire for more targeted therapies. In this study, we report the isolation and characterisation of four novel bacteriophages, ΦCD08011, ΦCD2301, ΦCD418 and ΦCD1801, able to infect a range of clinically relevant *C. difficile* PCR ribotypes. The four phages belong to the *Myoviridae* family of tailed phages, established through the use of transmission electron microscopy and the identification of key genes within their genomes. Host range analysis of the four phages showed that phage ΦCD1801 had the broadest host range, able to infect and lyse all but one of the PCR ribotype 078 isolates tested. Further growth rate characterisation, rate of attachment and determination of burst size were also completed. Finally, SlpA, the main component of *C. difficile's* S layer, was confirmed as the bacterial receptor for this phage. The presence of *slpA* from the S layer cassette of PCR ribotype 078 *C. difficile* strains conferred ΦCD1801 binding to an otherwise resistant strain. These findings will be instrumental in the ability to expand *C. difficile* phage host range to allow more targeted phages for the treatment of *C. difficile* infections.

Acknowledgements

I would like to thank my supervisors Professor Nigel Minton and Dr. Sarah Kuehne for their help, guidance and support throughout this project and for giving me the opportunity to conduct this research. I would like to thank the Medical Research Council for funding the research project and Phico Therapeutics Ltd. for their support and more specialist advice on parts of the project. I would also like to thank the members of the Synthetic Biology Research Centre for their continued support, advice and for creating a wonderful environment both inside and outside of research, particularly two Erasmus students Arlen Lücke and Renske van Esveld. Finally, I would like to thank my family and friends who have supported me throughout this research.

Abbreviations

%	Percentage
° C	Degrees Celsius
µg	Micrograms
µl	Microlitres
µm	Micrometre
µM	Micromolar
5FC	5-fluorocytosine
Abi	Abortive Infection
ACE	Allelic-Coupled Exchange
aTc	Anhydrotetracycline
BHI	Brain Heart Infusion
BHIs	Brain Heart Infusion Supplemented
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
CC	*Clostridium difficile* Selective Supplement
CDI	*Clostridium difficile* Infection
CDRN	*Clostridium difficile* Ribotyping Network
cfu	Colonies Forming Units
Cm	Chloramphenicol
cm	Centimetres
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
dH_20	Deionised Water
DNA	Deoxyribonucleic Acid
ECDC	European Centre for Disease Protection and Control
EDTA	Ethylenediaminetetraacetic Acid
EIBMV	Eliava Institute of Bacteriophage, Microbiology and Virology
ELISA	Enzyme Linked Immunosorbent Assay
EMBOSS	European Molecular Biology Open Software Suite
EOP	Efficiency of Plating
Erm	Erythromycin
EtOH	Ethanol
FDA	Food and Drug Administration
FMT	Faecal Microbiota Transplantation
FOA	Fluoroorotic Acid
g	Grams
g	Gravity
GBDP	Genome-BLAST Distance Phylogeny

GC	Guanine-cytosine
GI	Gastrointestinal
GOI	Gene of Interest
HMW	High Molecular Weight
HPA	Health Protection Agency
ICTV	International Committee on Taxonomy of Viruses
IPTG	Isopropyl β-D-1-thiogalactopyranoside
k	Rate of Attachment Constant
kbp	Kilo Base Pairs
kg	Kilograms
l	Litres
LB	Lysogeny Broth
LMW	Low Molecular Weight
M	Mole
mg	Milligrams
min	Minutes
ml	Millilitres
mM	Millimolar
mm	Millimetres
MOI	Multiplicity of Infection
mRNA	Messenger RNA
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs
ng	Nanograms
nm	Nanometre
No.	Number
OD	Optical Density
ORF	Open Reading Frame
PaLoc	Pathogenicity Locus
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
pfu	Plaque Forming Units
PHASTER	Phage Search Tool – Enhanced Release
PMC	Pseudomembranous Colitis
RAM	Retrotransposition-activated Marker
RAST	Rapid Annotation using Subsystem Technology
RBS	Ribosome Binding Site
RNA	Ribonucleic Acid
rpm	Rotations Per Minute

rRNA	Ribosomal RNA
s	Seconds
SBRC	Synthetic Biology Research Centre
SDS	Sodium Dodecyl Sulphate
SNP	Single Nucleotide Polymorphism
SOC	Super Optimal Broth
TAE	Tris-acetate-EDTA
TEM	Transmission Electron Microscopy
Tm	Thiamphenicol
U	Units
UK	United Kingdom
UV	Ultraviolet
V	Volts
VICTOR	Virus Classification and Tree Building Online Resource
VRE	Vancomycin Resistant Enterococci

Figures

Tables

Table of Contents

Chapter One:

Introduction

1 Introduction

1.1 Clostridial Species

The genus *Clostridium* encompasses a variety of Gram-positive, anaerobic, spore-forming bacteria of medical and industrial importance. They are commonly found in soil and as part of the normal intestinal flora of animals and humans (Wells and Wilkins, 1996). Pathogenic members of the genus include *Clostridium difficile* and *Clostridium botulinum.* The former causes a spectrum of disease ranging from antibiotic-associated diarrhoea to pseudomembranous colitis (PMC), while the latter causes paralysis of motor nerves, which can often be fatal. Industrially relevant clostridia include *Clostridium autoethanogenum* and *Clostridium acetobutylicum* which are utilised in large scale fermentation for the production of ethanol and butanol, from waste single carbon gas (Abrini *et al.*, 1994) or plant-derived sugars (Jones and Woods, 1986), respectively.

1.2 *Clostridium difficile*

Originally described by Hall and O'Toole (1935), *C. difficile* is an anaerobic, Gram-positive, endospore-forming, rod-shaped bacterium. In recent years the definition of the genus *Clostridium* has been debated in the scientific community and specifically the relatedness of *C. difficile* to the type strain *Clostridium butyricum.* In 2013, a phylogenetic reorganisation was conducted and a new genus, *Peptoclostridium,* was proposed (Yutin and Galperin, 2013). This change would have substantial cost implications, particularly for healthcare sectors, and could cause confusion to the public with a change from the well-known term "C. diff" to "P. diff." With this in mind, a new genus was proposed

Clostridioides, with *Clostridioides difficile* as the type strain (Lawson *et al*., 2016). For the purpose of this study the organism in question will be referred to as *Clostridium difficile.*

It is a major nosocomial pathogen and forms part of the natural gut microbiota of between 40 and 50 percent of neonates but less than 3 percent of healthy adults (Viscidi *et al*., 1981). Despite the identification of *C. difficile* vegetative cells and toxins in the faeces of neonates, disease symptoms are not observed. It has been hypothesised that this is due to the absence of toxin receptors in the infant colon, however there is no hard evidence to support this case. In 1978, *C. difficile* was first implicated as the cause of pseudomembranous colitis (PMC), with every patient in the study showing high titres of *C. difficile* toxins in their faeces (George *et al*., 1978).

1.2.1 Epidemiology and Risk Factors of Clostridium difficile *Infection*

C. difficile infection (CDI) is the main cause of antibiotic-associated diarrhoea in North America and Europe and is particularly prevalent in hospitals and long-term care facilities. In these locations, bacterial transmission is facilitated by patient clustering, the high proportion of elderly patients and the likelihood patients are receiving antibiotic treatment (Bartlett, 2008).

In 2012, The European Centre for Disease Protection and Control (ECDC) estimated that CDI costs the European Union around €3000 million per annum with an estimated cost of €5,000 - €15,000 per case in England alone (Jones *et al*., 2013). In the United States in 2011, *C. difficile* was associated with

approximately 29,000 deaths and responsible for almost half a million infections (Lessa *et al.*, 2015). In 2017, Public Health England reported around 14,000 cases of CDI in England, covering both NHS and private patients. In 2015, the reported average cost of CDI case management was $42,316 US dollars and the CDI-attributable costs were $21,448 US dollars per case (Zhang *et al*., 2016).

Administration of broad spectrum antibiotics, specifically clindamycin, leads to dysbiosis of the natural gut microbiota and is the primary risk factor associated with CDI (Theriot *et al*., 2014). Dysbiosis of the gut microbiota leads to the creation of a niche in which *C. difficile* can survive and outcompete other bacterial species. In the hamster model of CDI, it has been noted that the number of *Enterobacteriaceae* increases while the number of *Lachnospiraceae* decreases post-antibiotic exposure, leaving the hamsters susceptible to infection by *C. difficile* (Reeves *et al.*, 2011). Other antibiotics have also been implicated as risk factors for CDI, including cephalosporins, likely because of their extensive use in healthcare, and fluoroquinolones (Muto *et al*., 2005). The use of immunosuppressants and proton pump inhibitors have also been implicated as risk factors for CDI but to a lesser extent (Kwok *et al*., 2012). Duration of stay, age and antibiotic exposure remain the biggest risk factors associated with hospital-acquired CDI (Bartlett and Gerding, 2008). Since the late 1980s, incidences of community-acquired CDI have increased. Worryingly these cases are less likely to be caused by antibiotic exposure and have been observed in a lower age group (Khanna *et al*., 2012).

The changing epidemiology of CDI can be, in part, attributed to the emergence of so called "hypervirulent" strains of *C. difficile*, namely PCR ribotype 027 and 078 strains, which are observed to produce higher levels of toxins than other *C. difficile* PCR ribotypes (Stabler *et al*., 2008). *C. difficile* PCR ribotype 027 strains were responsible for the large UK outbreak between October 2003 and June 2004 at Stoke Mandeville Hospital, Buckinghamshire (Healthcare Commission, 2006). In this outbreak, 19 deaths out of 174 cases were directly attributed to CDI. A further 19 deaths were recorded for a second outbreak at the same hospital, between October 2004 and June 2005 (Healthcare Commission, 2006). Since these outbreaks, the mandatory recording of CDI infections across Europe has tracked the spread of *C. difficile* PCR ribotype 027 from the UK across Europe (Freeman *et al*., 2010).

1.2.2 C. difficile *Pathogenesis and Transmission*

In the environment *C. difficile* survives as metabolically dormant endospores, not as vegetative cells. Spores are intrinsically resistant to a variety of chemical and physical agents, such as antibiotics, commonly used hospital bleach-free disinfectants, heat, ultra-violet (UV) light and desiccation (Setlow, 2014). The existence of *C. difficile* as spores means they can easily persist in the environment and are easily transferred between patients if correct hand hygiene protocols are not followed.

Upon ingestion, *C. difficile* spores enter the gastrointestinal (GI) tract of the host and can germinate into vegetative cells in the presence of bile salts, such as taurocholate (Paredes-Sabja *et al.*, 2014). As vegetative cells, *C. difficile*

produces two toxins, TcdA and TcdB, which are responsible for the characteristic diarrhoeal symptoms of CDI (Voth and Ballard, 2005). A third toxin, binary toxin, is also produced by some *C. difficile* strains and has been implicated in disease, although it is not known to what extent it is involved (Barbut *et al*., 2007).

The two large Clostridial toxins, TcdA and TcdB, have glycosyltransferase activity which targets host guanosine triphosphatases which regulate actin, specifically they inactivate Cdc42, Rho and Rac within target cells (Voth and Ballard, 2005). The actin cytoskeleton is responsible for cell movement across surfaces and is directly driven by actin polymerisation. Such structural processes which are dependent on actin polymerisation are regulated by Rho, Rac and Cdc42 (Tapon and Hall, 1997). The loss of actin regulation causes a loss of cell integrity meaning that cells which are intoxicated cannot carry out their regular functions within the host (Voth and Ballard, 2005). The genes, *tcdA* and *tcdB,* encoding these toxins are found within a 19.6 kbp pathogenicity island, known as PaLoc (Braun *et al*., 1996). This locus also encompasses the genes involved in the regulation of toxin gene expression and release of the toxins from the cell (Borriello, 1998). Other genes encode factors that contribute to virulence, with strains lacking flagella known to be limited in their ability to adhere to epithelial cells. Cell wall proteins, Cwp66 and S-layer protein A, and fibronectin binding protein A have also been linked with *C. difficile* adherence (Abt *et al*., 2016).

CDI can cause a wide spectrum of disease, ranging from mild diarrhoea to PMC and the potentially fatal toxic megacolon (Sayedy *et al*., 2010). Unlike in other

forms of colitis, PMC is restricted to the rectum and colon (Farooq *et al*., 2015). CDI is rare in infants and children and most commonly affects elderly hospitalised patients where symptoms of foul smelling, watery diarrhoea are accompanied by abdominal pain (Burnham and Carroll, 2013). Onset is usually abrupt and, in some cases, fever and elevated white blood cell count are observed. Disease progression can be monitored by the formation of white or yellow plaques attached to the intestinal mucosa and varying degrees of necrosis (Borriello, 1998).

1.2.3 Diagnostics and Current Treatments

The gold standard diagnostic for CDI is the use of tissue culture cytotoxicity assays to detect the presence of TcdA and TcdB (Musher *et al*. 2007). In this method, the patient stool samples are filtered then added to monolayers of fibroblasts. The cytopathic effect of the toxins is characterised by cell rounding and is used to confirm the presence of toxin, producing a highly sensitive and specific diagnostic (Planche and Wilcox, 2011). However, the relative expense and labour intensity means enzyme linked immunosorbent assays (ELISAs) are often preferred. Although the rapid results and ease of use are beneficial, ELISAs are less sensitive than cytotoxicity assays and can miss 5 to 10 percent of cases (Bartlett and Gerding, 2008). There are now a variety of commercially available ELISA kits (such as those from TECHLAB® and Abnova) for the detection of *C. difficile* in stool samples which give results in a few hours. The detection of glutamate dehydrogenase in patient faecal samples, using ELISA, has proven highly successful in the detection of *C. difficile*, with sensitivity of almost 100 % (Keessen *et al*., 2011). In *C. difficile,* glutamate dehydrogenase is encoded by

gluD and converts glutamate to α-ketoglutarate and ammonia. The enzyme can be detected in the supernatant of *C. difficile* cultures as well as in faecal samples (Girinathan *et al*., 2014). An ELISA detecting glutamate dehydrogenase can be completed within 45 minutes and is relatively inexpensive in comparison to other detection methods (Ticehurst *et al*., 2006). Using a combination of diagnostics to detect toxins and to detect glutamate dehydrogenase has proven highly effective and, additionally, allows the identification of asymptomatic carriers.

In the UK, the Heath Protection Agency (HPA) guidance advises using either metronidazole or vancomycin for the treatment of CDI (Wilcox *et al*., 2013). Metronidazole is the preferred first line therapy for mild to moderate disease, while vancomycin is preferred for severe cases. Vancomycin is a so called "last line of defence" antibiotic and along with concerns over growing resistance to this antibiotic, such as the emergence of vancomycin resistant enterococci (VRE), its use has been minimised in the treatment of CDI (Cetinkaya *et al*., 2000). Vancomycin has proved highly effective in the treatment of relapse patients, however, metronidazole is preferred as the cheaper, effective alternative. It is recommended that oral metronidazole and vancomycin should be administered for 10 days, at a dose of 500 mg and 125 mg respectively (Wilcox *et al*., 2013). Tapering the dose of vancomycin for recurrent infections has proven successful. Cure rates for the two drugs is very similar, ranging from 73 – 94 % for metronidazole and 84 to 94 percent for vancomycin (Drekonja *et al.*, 2011). In the US, metronidazole is not approved by the FDA for the treatment

of CDI and consequently the use of vancomycin, which is FDA approved for the treatment of CDI, is often favoured (Bartlett, 2006).

In 2011 another antibiotic, Fidaxomicin, was approved by the FDA for the treatment of CDI in adults (Medscape, 2011). It is a narrow spectrum agent making it ideal for treating CDI as it is unlikely to disrupt the natural gut microbiota of the individual to the same extent as other broader spectrum antibiotics. Although more expensive than metronidazole it is a highly effective treatment (Venugopal and Johnson, 2012).

One of the biggest concerns in the treatment of CDI is the high incidence of relapse, occurring in around 15-35 % of treated patients (Marsh *et al*., 2012). Antibiotic exposure has been proven to be essential in *C. difficile* colonisation and persistence within the gut (Owens *et al*., 2008) and hence antibiotic use has been implicated in the development of disease. It has been reported that the use of metronidazole to treat unrelated infections has led to disruption of the gut microbiota (Saginur *et al*., 1980), which requires caution to be displayed when using it as a therapy for CDI as it could exasperate the infection considering dysbiosis of the gut microbiota through antibiotic use is the biggest risk factor for CDI.

Strategies designed to alter the gut microbiota through removal of harmful bacterial species have been extensively studied and have led to the creation of probiotics for CDI treatment (Na and Kelly, 2011) and to the development of faecal microbiota transplantation (FMT, Bakken *et al*., 2011). Probiotics for CDI

have focused on using *Lactobacillus*, *Bifidobacterium* and the yeast *Saccharomyces boulardii*. Extensive meta-analysis focusing on these studies have shown that the use of probiotics containing these organisms can lead to a 60.5 % reduction (Lau and Chamberlain, 2016) in the incidence of *Clostridium difficile* associated disease (CDAD). FMT aims to confer protection to *C. difficile* colonisation by restoring the diversity of the gut microbiota. The therapy is usually targeted to patients who have previously not responded to antibiotic therapy (Kassam *et al*., 2013). The stigma associated with FMT causes a significant road block to its widespread use, nonetheless efficiencies of 90 percent have been reported (Youngster *et al*., 2014). The cost of screening donor samples for harmful bacteria and viruses may be outweighed by the relative low cost of the necessary transplantation material (Rao and Young, 2015).

An alternative avenue to CDI treatment is to interfere with spore germination as without vegetative cells the characteristic disease symptoms cannot occur. Howerton and colleagues (Howerton *et al*., 2013) investigated the possible use of CamSA, a bile salt analog, for the treatment of CDI in mice. Mice were administered a dose of *C. difficile* spores and varying concentrations of CamSA and it was observed that a dose of 50 mg/kg body weight could prevent mice developing CDI. This route of intervention provides a promising alternative to antibiotics for the treatment of CDI without disrupting the natural gut microbiota (Howerton *et al*., 2013).

1.2.4 C. difficile *Surface Layer*

Like many other bacteria, *C. difficile* vegetative cells are completely coated in a two-dimensional paracrystalline protein surface layer called the S layer (Kawata *et al*., 1984), described as a two-dimensional structure due to the arrangement of the glycoproteins as a monolayer. The S layer is encoded by the *slpA* gene of which the product, SlpA, is post-translationally cleaved to form a high molecular weight (HMW) and a low molecular weight (LMW) form (Calabi *et al*., 2001; Karjalainen *et al*., 2001). SlpA precursor cleavage is mediated by the cysteine proteinase, Cwp84 (Kirby *et al.*, 2009) and these two forms, HMW and LMW, associate to form the S layer. The SlpA precursor also contains a signal peptide for translocation across cell membranes. The HMW form is relatively conserved across strains and is derived from the C-terminal region of the precursor (Calabi *et al*., 2001). The LMW form is highly variable, smaller and derived from the N-terminal region of the precursor. It is thought that the LMW form is the main serotyping antigen, whilst the HMW form possesses amidase activity (Calabi *et al*., 2001). The HMW form also shares sequence homology with the *N*-acetylmuramoyl-L-alanine amidase from *Bacillus subtilis* indicating its amidase activity (Calabi *et al*., 2001). Comparisons of the HMW and LWM forms have shown that there is extensive sequence variation between *C. difficile* species (Calabi and Fairweather, 2002; Karjalainen *et al.*, 2002), particular in the LMW form which is thought to be externally located on the cell wall and facing the environment (Fagan *et al*., 2009).

In the genome of *C. difficile* 630, although it is thought to be the same across all *C. difficile* strains, the *slpA* gene is located within a 36.6 kb cell wall protein

(*cwp)* gene cluster (Sebaihia *et al*., 2006). In this genome there are 28 paralogs of *slpA*, which are transcribed in the same direction, suggesting these genes have related functions and coordinated regulation (Calabi and Fairweather, 2002). These paralogs all contain the Pfam motif, PF04122, a cell wall binding motif and some also contain a variable second domain (Sebaihia *et al*., 2006). One of these paralogues is a known heat shock inducible adhesin, Cwp66 (Karjalainen *et al*., 2001) and another is a cysteine protease, Cwp84 (Savariau-Lacomme *et al*., 2003). Within the *cwp* gene cluster, a 9.7 kb S layer cassette has been identified (Dingle *et al*., 2013) which encodes *slpA*, *cwp66* and *secA2* (Figure 1.1). Fagan and Fairweather (2011) identified two *secA* genes and showed that SecA2 is involved in the translocation of surface layer proteins and cell wall proteins.

Figure 1.1: ***C. difficile* S layer cassette**, *slpA* encodes the SlpA precursor protein which is post-translationally cleaved to produce high molecular weight and low molecular weight surface layer proteins, *secA2* encodes an accessory secretory system which is involved in secretion of surface layer and cell wall proteins, *cwp66* encodes an adhesin and *cwp2* is a cell wall protein. The unlabelled ORF encodes an uncharacterised protein and represents the ORF annotated as CD2790 in *C. difficile* 630. In *C. difficile*, the 9.7 kb S layer cassette is part of a larger cell wall protein gene cluster of 36.6 kb.

Bacterial S layers are known to have a role in adhesion, colonisation and nutrient uptake (Sára and Sleytr, 2000). Recently, it has been hypothesised that the *C. difficile* S layer could play a role as the receptor for bacteriophage infection, although there is currently no direct evidence to confirm this (Kirk *et al*., 2016).

In 2017 it was reported that the S layer of *C. difficile* was the receptor for Avidocin-CD (Kirk *et al*., 2017), a genetically modified R-type bacteriocin. This bacteriocin resembles a Myoviridiae phage tail and contains a receptor binding protein at the tip of the tail fibre which causes sheath contraction upon binding with the appropriate receptor on the bacterial cell surface. The gene encoding the receptor binding protein can be replaced with a homolog from another species or strain to retarget the bacteriocin's killing. Rob Fagan and colleagues isolated two Avidocin-CD spontaneous mutants resistant to killing due to point mutations in the *slpA* gene, suggesting that the receptor for the Avidocin-CD was the S layer. This was further confirmed with the production of a panel of modified Avidocin-CDs containing newly isolated phage receptor binding proteins and showed that killing could be retargeted in a S layer dependent manner (Kirk *et al*., 2017). Although this work was not conducted with a complete phage it is strongly suggestive that the *C. difficile* S layer is the phage receptor.

1.3 Bacteriophages

Bacteriophages (phages) were first discovered in 1915 by Frederick Twort and in 1917 by Félix d'Hérelle, when they independently observed clearing of bacterial growth through the formation of plaques (Duckworth, 1976). The name "bacteriophage" was coined by d'Hérelle with a literal translation of "bacteria eater." Phages are viruses which invade bacterial cells and disrupt host cell metabolism which leads to cell lysis (Sulakvelidze *et al*., 2001). They are ubiquitous within the environment and are found anywhere their host bacteria are found, and it is thought that for every one bacterial cell there are 10 phage

particles, making them the single most populous entity on Earth, with an estimated 10^{31} present (Sulakvelidze *et al*., 2001).

1.3.1 Bacteriophage Morphology and Classification

Phage particles consist of a proteinaceous capsid, or head, which contains the phage nucleic acid. Phage nucleic acid can be DNA or RNA, single or double stranded and their genomes can range from simple to highly complex, varying greatly in size between species (Kutter and Sulakvelidze, 2005). Usually, the capsid is created through multiple copies of the same protein (Aksyuk and Rossmann, 2011). Tailed phages also contain a tail which is responsible for bacterial cell recognition and triggering nucleic acid release (Chaturongakul and Ounjai, 2014). Most phage genomes contain a tail tape measure protein which dictates the length of the tail (Kutter and Sulakvelidze, 2005). Figure 1.2 shows a schematic diagram of the morphology of a typical tailed phage particle.

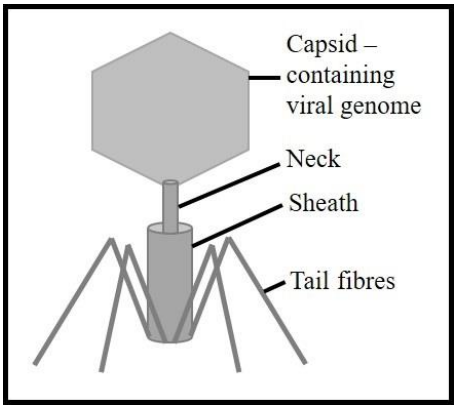


Figure 1.2: **Schematic representation of typical tailed phage particle belonging to the order *Caudovirales***, showing capsid containing nucleic acid and tail with tail fibres for recognition and attachment of bacterial cells.

The most well characterised phages belong to the order *Caudovirales*. These are double stranded DNA, tailed phages that are divided on tail length into three sub-families: *Myoviridae, Podoviridae* and *Siphoviridae*. *Siphoviridae* have long non-contractile tails, *Myoviridae* have contractile tails which are medium to long in length and *Podoviridae* have short tails (Kutter and Sulakvelidze, 2005). With the number of new phages discovered continuing to increase, the International Committee on Taxonomy of Viruses (ICTV) are constantly striving to classify all phages (Aksyuk and Rossmann, 2011).

1.3.2 Bacteriophage Life Cycles

Phages are unable to replicate without their bacterial host machinery and require cell lysis to spread (Koskella and Meaden, 2013), or adopt lysogeny as a replicative strategy where they are replicated as part of the bacterial genome. Phages can be either temperate or lytic, and while lytic phages can only follow the lytic life cycle, temperate phages have a choice to follow either the lytic or the lysogenic cycle (Kutter and Sulakvelidze, 2005). Figure 1.3 shows an outline of the two phage life cycles.

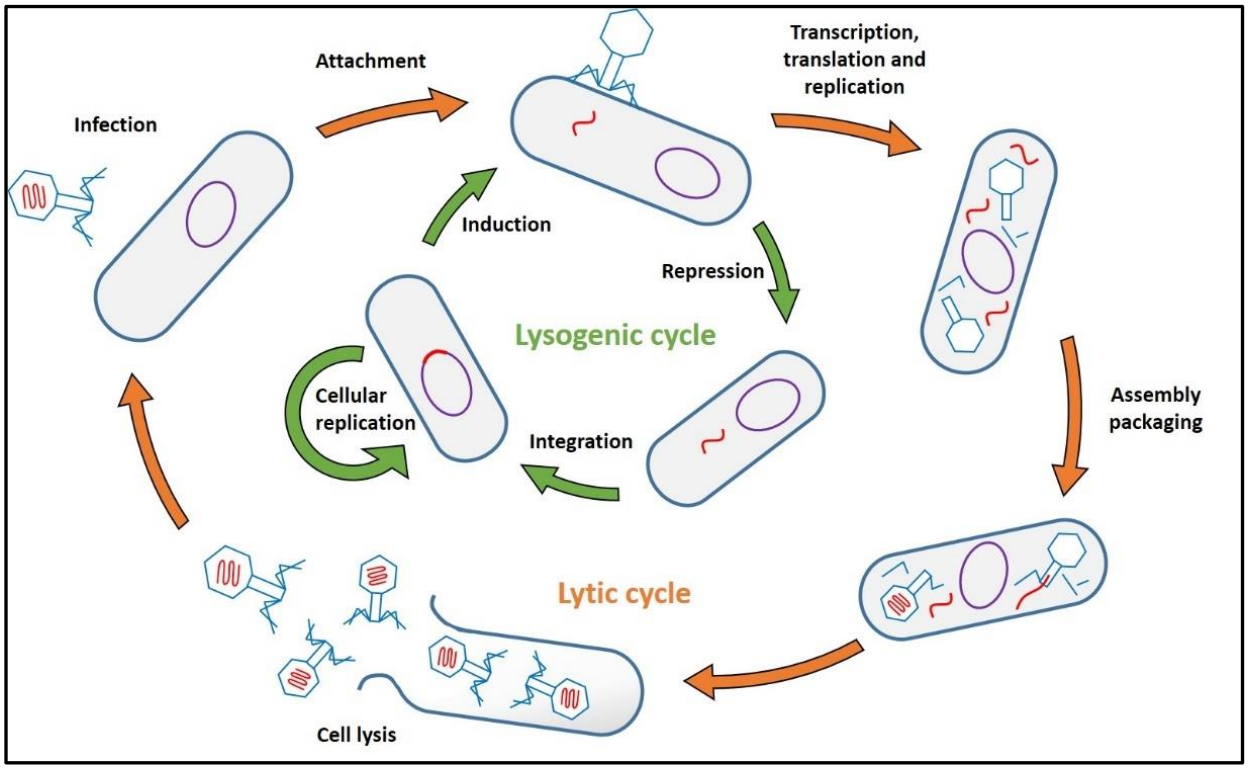


Figure 1.3: **The lytic and lysogenic life cycles of bacteriophage lambda**. In the lytic cycle phages infect and lyse their bacterial host, whereas in the lysogenic cycle phages can integrate into the host cell genome as a prophage. Image adapted from Campbell (2003).

Strictly lytic phages infect, lyse and kill their bacterial host cells whilst following the lytic life cycle. In this cycle, the host cell machinery is hijacked to bring about transcription of phage genes and the production of new viral particles (Campbell, 2003). Temperate phages, however, can utilise both the lytic and the lysogenic life cycles. In the lysogenic life cycle, phages can integrate their genomes into the bacterial host genome. In this state, the phage is known as a prophage and the bacterial cell containing the prophage is called a lysogen. Maintenance can be either as free circular DNA or as linear DNA within the host chromosome (Kutter and Sulakvelidze, 2005). Temperate phages can switch between lysogenic and lytic life cycles, in part depending on the conditions of

the environment (see section 1.3.3 for more information). Prophages can be induced to commence the lytic life cycle by exposure to UV irradiation or antibiotics, most commonly antibiotics that induce the SOS response in the host cell by causing DNA damage (Rokney *et al*., 2008). Prophages can also be induced using heat if the phage contains a temperature-sensitive repressor (Rokney *et al*., 2008).

In addition to the strictly lytic and lysogenic life cycles, a third life cycle has been documented called pseudo-lysogeny which is defined as the stalled development of a phage genome in the host cell. In this state neither lytic development nor lysogeny proceed. The phage genome is not degraded and therefore, phage development can restart when conditions are favourable. It is thought that pseudo-lysogeny plays an important role in phage survival in nature (Los and Wegrzyn, 2012).

Pseudo-lysogeny has been observed in *Pseudomonas aeruginosa* when cells are nutrient starved and is likely to occur due to a lack of cell energy for initiating either the lytic or lysogenic life cycles. The addition of nutrients to starved cells leads to the immediate commencement of phage particle production and subsequent host cell lysis (Ripp and Miller, 1998). Pseudo-lysogeny has also been confirmed in *Salmonella typhimurium*, where time lapse fluorescence microscopy was utilised to determine the location of temperate phage P22 DNA within the cell. It showed that, in some cells, the phage chromosome was not integrated into the host cell chromosome and was being maintained as free DNA (Cenens *et al*., 2013). Finally, pseudo-lysogeny has also been reported for

phages of *Bacillus subtilis* (Sonenshein, 2006) where it is associated with sporulation. In this case phage DNA is "captured" on formation of the endospore and maintained until germination of the spore when the lytic life cycle is induced (Sonenshein, 2006).

In both the lytic and lysogenic life cycles, the first step is phage adsorption to host cells, which occurs through chance encounters that allow attachment of phage tail fibres to bacterial cell surface proteins and can be reversible or irreversible (Baptista *et al*., 2008). Normally, a two-step attachment process is observed, such as in the *Lactococcus lactis* phages of the c2 species where initial binding of carbohydrates on the cell surface is reversible and a second irreversible binding is required to a predicted membrane-attached protein to ensure successful phage attachment (Tremblay *et al*., 2006). In *Caudovirales*, the next step is penetration, which involves the puncturing of the cell membrane and injection of viral DNA (Kutter and Sulakvelidze, 2005). At this point the phage either enters lysogeny or, those genes needed for the lytic cycle are transcribed and translated. These usually encode proteins which are responsible for replication of phage DNA including a phage-specific DNA polymerase. This is followed by the transcription and translation of late structural genes encoding the capsid and tail (Kutter and Sulakvelidze, 2005). Once phage particles are assembled they lyse the bacterial cell and can proceed to further infect neighbouring cells. In most phages cell lysis involves two phage encoded proteins, the endolysin and holin proteins (Young, 1992). The holin is required to create a hole in the cytoplasmic membrane to allow the endolysin to pass through. Phage endolysins do not usually contain a secretory signal sequence

and, therefore, the holin is required to allow them to escape the cytoplasm in order to disrupt the peptidoglycan rigidity of the cell and cause cell lysis (Young, 1992).

1.3.3 The Lysis Lysogeny Decision

Upon infection of a bacterial cell, phage DNA is injected into the cell and at this point the decision to proceed with lytic growth or to lysogenize the host is made. Although it is not known whether all phages follow this mechanism, the mechanisms and processes involved in the decision as to which route to follow have been best studied in *E. coli* phage λ and has been assumed as the mechanism for most phages. It involves the interplay between two regulatory proteins, CI and Cro (Ptashne, 2004). When in the prophage state, the only phage λ gene that is expressed encodes CI which, is both a positive and negative regulator of transcription, and is responsible for the maintenance of lysogeny, it is also called the λ repressor protein (Ptashne, 1967). The CI protein is found as dimers in the cell and is cleaved to monomers by RecA following phage genome exposure to DNA damaging agents. As a monomer, CI cannot bind as effectively to its operator and consequently transcription of the *cro* gene is triggered. The Cro protein is required for lytic growth (Svenningsen *et al*., 2005). These encoding genes (*cI* and *cro*) have their own promoters, which although next to each other on the chromosome do not overlap and are transcribed divergently. The genes are separated in the λ genome by an 80 bp gap which encompasses the promoter and operator regions of the two promoters, P_{RM} promoter in the case of the *cI* gene and P_R in the case of *cro.* There are three 17 bp operator sites, O_R1, O_R2 and O_R3, involved in the genetic switch. O_R1 overlaps P_R, O_R3 overlaps P_{RM} and

O_R2 overlaps both promoters. The CI and Cro proteins bind to these operator sites to regulate transcription of their associated genes (Ptashne, 2004). During lysogeny, the CI protein is bound to both O_R1 and O_R2, which prevents binding of the RNA polymerase to P_R and consequently prevents transcription of *cro* and late phage genes. Binding of CI at O_R2 also increases its own transcription by 10-fold in comparison to binding at O_R1 alone, by aiding the binding efficiency of the RNA polymerase at P_{RM} through protein to protein interactions (Ptashne, 2004). This means that once lysogeny is established and CI levels are high, CI promotes transcription of its own gene, resulting in the production of more CI and hence the lysogenic state is maintained.

For CI protein, the relative affinity for the operator sites is as follows $O_R1 > O_R2 > O_R3$, although Cro protein binds to the same operators as CI protein, it binds with opposite affinities where $O_R3 > O_R1, O_R2$ (Johnson *et al.*, 1978). These affinities ensure that during lysogeny, transcription of *cro* is essentially abolished and transcription of *cI* is maintained at a constant level. Cro and CI proteins contain three alpha helices forming a helix-turn-helix motif which is responsible for DNA binding. The recognition helix is alpha helix three and fits into the major groove of the operator DNA. The recognition helices of both Cro and CI proteins are conserved to begin with and then diverge. They use glutamine at position one to contact position two of the operator and serine at position two to contact position four of the operator. The other amino acids of the recognition helices enable the proteins to distinguish between operator sites, with Cro's recognition helix contacting position three of the operator, which is the preferred base found in O_R3. Whereas, the arm of the CI protein contacts

both positions six and eight of the operator and prefers those bases found in O_R1. In lytic growth Cro is bound to O_R3 preventing transcription of *cI* by stopping binding of RNA polymerase to P_{RM} (Ptashne, 2004). Initial transcription of *cro* occurs immediately after DNA damage where the inactivation of CI dimers allows the activation of the P_R promoter. As Cro protein binds to O_R3 with the highest affinity, the concentration of Cro protein required to turn off the transcription of *cI* is lower than that required of CI protein for the switching off of *cro* transcription (Johnson *et al*., 1978). The components of the genetic switch in phage λ are shown in Figure 1.4.

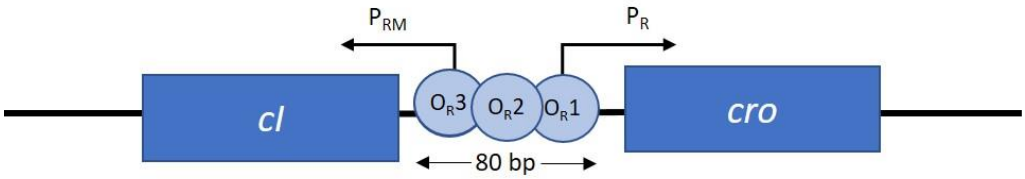


Figure 1.4: **Genetic regulation of *cI* and *cro* in *E. coli* phage λ**, Cro promotes lytic growth, *cI* encodes the CI protein, also called the λ repressor, and is expressed during lysogeny, binding of Cro and CI to operators regulates the expression of *cI* and *cro*. Each operator is 17 bp in length and overlap the promoters to control the expression of the two genes.

During lyosgeny at least 100 copies of the CI protein molecule are present within the cell, and as a consequence, the O_R3 operator is always occupied repressing *cro* transcription. Other unbound CI protein molecules bind to operators on incoming phages (the same phage or related phages) and prevent infection. The lysogenised cell is, therefore, immune to superinfection (Ptashne, 2004).

Prophages can be induced to enter the lytic cycle by exposure of the host cell to DNA damaging agents, such as UV radiation and antibiotics, usually mitomycin C and norfloxacin (Nale *et al*., 2012). In phage λ, exposure to UV irradiation activates the bacterial protein RecA (Patel *et al*., 2010), through its conversion into a highly active protease, which can cleave CI protein dimers into monomers (Pal and Chattopadhyaya, 2009). CI proteins are usually found as dimers in the cell, and as dimers they can bind strongly to the operators. In their monomer form, binding to operators is extremely weak, therefore cleavage of CI protein dimers into monomers by RecA means it effectively loses the ability to bind operators (Johnson *et al*., 1978). When the CI protein can no longer bind to O_R1 and O_R2, RNA polymerase is able to bind to P_R and begin transcription of *cro* and downstream genes. Cro can then bind to O_R3 and prevent further synthesis of the CI protein resulting in the commencement of the lytic life cycle (Ptashne, 2004).

In addition to the involvement of CI and Cro, two proteins, CII and CIII are also implicated in lysogeny control. In phage λ, the *cII* and *cIII* genes are located downstream of *cro* and are involved in the initial establishment of lysogeny. The CII protein is responsible for the activation of the P_I promoter for the phage encoded integrase gene, the product of which is required for the integration of the λ genome into the host cell chromosome (Schubert *et al*., 2007). The CII protein is also involved in the activation of the P_{RE} promoter which controls transcription of the *cI* gene and is therefore required for the initial establishment of lysogeny before the CI protein can autoregulate its own transcription (Fien *et al*., 1984). The CII protein is quickly degraded by host cell proteases, therefore

in order to stabilise CII, the CIII protein is required (Herman *et al*., 1997). Host cell proteases are highly abundant in rich media and it is thought that the rapid degradation of CII protein means that the lysogenic state cannot be established and therefore in these conditions lytic growth is more likely to ensue.

Although not known to occur in all phages, the decision for some phages to lysogenise their hosts has recently been determined to involve a small molecule communication system, called a Arbitrium system (Erez *et al*., 2017). This was demonstrated using phages of the SPbeta group which can infect *Bacillus subtilis*. Upon infection of the host cell, it was observed that phages would release a small communication peptide, six amino acids in length, into the growth medium, which was then sensed by progeny phages. When the concentration of the peptide was sufficiently high, progeny phages would choose to lysogenise their hosts instead of embarking on lytic growth (Erez *et al*., 2017). The Arbitrium system in the SPbeta group phages is encoded by three genes *aimP*, *aimR* and *aimX*. The product of *aimP* is the peptide of six amino acids in length, *aimR* encodes the intracellular peptide receptor and *aimX* encodes a negative regulator of lysogeny (Erez *et al*., 2017).

1.4 Bacteriophage Therapy

1.4.1 History of Bacteriophage Therapy

Ever since their first discovery, Félix d'Hérelle exploited phages as potential therapeutic agents, realising their potential to clear bacterial populations with low dosage and little side effects. Phages were initially used to treat dysentery in a 12 year old boy, who made a full recovery within three days after only one

dose of phage treatment (Sulakvelidze *et al*., 2001). Although there was great controversy over initial phage therapeutic studies, due to the lack of controls, the potential of phage therapy was rapidly recognised (Wittebole *et al*., 2014). This led to the commercialisation of phage products in France, d'Hérelle commercial laboratory which is now L'Oréal, and in the United States, the Eli Lilly. These companies produced phage products for the treatment of various human infections, such as wounds, abscesses, acute and chronic respiratory infections and vaginitis (Sulakvelidze *et al*., 2001), however the discovery and commercialisation of antibiotics halted research and production of phage therapeutics in the Western world. Interest and research continued in Eastern Europe and the former Soviet Union where several industries are still producing and manufacturing phage therapeutics. The Eliava Institute of Bacteriophage, Microbiology and Virology (EIBMV) in Georgia was founded in 1923 and remains one of the largest institutes in the world for the preparation of phage therapeutics. The therapeutics produced target a range of bacterial pathogens including *Staphylococcus* and *Pseudomonas* (Sulakvelidze *et al*., 2001). In recent years there has been a surge in companies aiming to get phage therapeutics to the market, including AmpliPhi Biosciences who currently have four products in the pipeline of which their most advanced, treating *Staphylococcus aureus* in chronic rhinosinusitis, has completed phase I clinical trials (http://www.ampliphibio.com/).

The overuse and misuse of antibiotics over many years has led to the growing spread of antibiotic resistant bacteria and an increased number of bacterial pathogens which are becoming untreatable with common antibiotics (Normark

and Normark, 2002), including the emergence of vancomycin-resistant enterococci (Cetinkaya *et al*., 2000), vancomycin being a so called drug of last resort. Although antibiotic resistance is not a new problem, it has become an increasing concern as the struggle to find new antibiotics becomes clearer. Further to these already known problems with antibiotic resistance, there are reports of resistance emerging to another antibiotic of last resort, colistin. Colistin has been heavily relied upon due to widespread carbapenem-resistant Enterobacteriaceae for the treatment of Gram-negative bacterial infections. However, its widespread use in agriculture and healthcare has led to the emergence of resistance. Initially colistin resistance could only be acquired through chromosomal mutations, which would be slow to disseminate throughout entire bacterial populations and limit its clinical importance. However, the recent discovery of an *E. coli* strain harbouring plasmid-bourne resistance is very concerning as it would provide a mechanism for the rapid spread of colistin resistance through bacterial populations (MacNair *et al*., 2018). The global issue of antibiotic resistance has led to a renewed interest in the potential of phages as therapeutics in the West and has led, in part, to the undertaking of this study to further the understanding of phages and their potential as therapeutic agents.

1.4.2 Advantage of Phage Therapy over Antibiotics

Phages have many attributes which make them highly desirable for therapy. Firstly, the highly specific nature of phages means that, unlike broad spectrum antibiotics, such as cephalosporins, the chance of disruption to the natural flora, specifically of the gut microbiota, is lowered (Manrique *et al*., 2017). Phages

usually only infect a certain species or strain of bacteria so are ideal for the treatment of pathogens such as *C. difficile*, where dysbiosis of the gut microbiota is a major risk factor for infection (Nale *et al*., 2016). Secondly, unlike antibiotics, phages can replicate at the site of infection so a lower initial treatment dose can be given, limiting the systemic dose and potentially reducing the chance of side effects (Miedzybrodzki *et al*., 2012). One drawback of phage therapy is the need to deliver them to the site of infection without them being disadvantaged by the harsh conditions within the human body, such as pH variations (Ly-Chatain, 2014). This is one reason why previously successful phage therapeutics have been topical treatments (Drilling *et al*., 2017). Finally, the emergence of phage resistant bacteria can be overcome more effectively than antibiotic resistance due to phages and bacteria co-evolving as the emergence of resistant bacteria drives natural evolution of phages to overcome this resistance. The relative cost and ease of isolating new phages, or the use of phage combinations can also minimise the impact of bacterial resistance to phage therapeutics (Golkar *et al*., 2014).

1.5 Bacteriophages of *Clostridium difficile*

1.5.1 C. difficile *and Its Associated Bacteriophages*

To date, August 2017, no strictly lytic phages infecting *C. difficile* have been reported in the literature. The first reported isolation of phages able to infect *C. difficile* was in 1983, when two phages were discovered after the induction of ten clinical isolates of *C. difficile* using Mitomycin C (Sell *et al*., 1983), for use in *C. difficile* typing. Other phages have also been isolated for *C. difficile* typing purposes but have not be characterised in detail (Hawkins *et al*., 1984; Bacon *et*

al., 1988). Since then, the difficulties in isolating and propagating *C. difficile* phages has been highlighted, such as their seemingly narrow spectrum and therefore the requirement for a large *C. difficile* strain library for their isolation (Hargreaves and Clokie, 2014). Even so, many temperate phages infecting *C. difficile* have been isolated and characterised to different degrees, ranging from use of transmission electron microscopy to host range analysis to full genome sequencing. Phages of *C. difficile* have been isolated through induction of *C. difficile* lysogenic strains using mitomycin C (Nagy and Foldes, 1991; Ramesh *et al*., 1999; Goh *et al*., 2005; Govind *et al*., 2006; Fortier and Moineau, 2007; Horgan *et al*., 2010; Nale *et al*., 2012), usually at a final concentration of 3 µg/ml. Another antibiotic, norfloxacin, has also been used to induce phages from lysogenic *C. difficile* strains (Shan *et al*., 2012). Interestingly, it was observed that when one *C. difficile* strain was exposed to mitomycin C and norfloxacin two different prophages were induced, one induced when mitomycin C was used and a second when norfloxacin was used. Lysogenic strains of *C. difficile* have also been induced successfully using UV radiation (Sekulovic *et al*., 2014). Initial attempts to isolate free phages from environmental samples and patient faecal samples were unsuccessful (Goh *et al*., 2005; Nale *et al*., 2012), however recent attempts have proven more fruitful, with free phages being isolated from patient stool (Meessen-Pinard *et al*., 2012) and soil samples (Rashid *et al.*, 2016) using sample enrichment techniques.

In all cases, only temperate phages have thus far been identified and it is postulated that the lack of strictly lytic phages of *C. difficile* is due to the presence of *C. difficile* as spores in the environment. Phages require actively growing cells

to maintain their populations through lytic growth, therefore integration and maintenance as prophage is advantageous as the prophages will be maintained through the sporulation process. Lytic phages would not be maintained through *C. difficile* sporulation and therefore in unfavourable conditions for vegetative growth, lytic growth would not be beneficial for the phage (Goh *et al.*, 2005). However, lytic phages of many endospore-forming bacteria have been isolated, such as those infecting *Clostridium perfringens* (Seal, 2013) and *Bacillus cereus* (Lee *et al*., 2011). It is apparent that large scale screening was required in these cases and could suggest that lytic phages of *C. difficile* are simply yet to be isolated and a larger number of potential host strains are required for screening. Pseudo-lysogeny has also been observed in phages of *Bacillus subtilis*, namely SP10, PBS1 and Φ29 (Sonenshein, 2006). In this case phage DNA is "captured" during the formation of the endospore and maintained as DNA in the cytoplasm. Upon germination of the endospore, lytic growth of the phage ensues. It appears that in this species the sigma factor for RNA polymerase for the transcription of phage lytic genes is only present in vegetative cells and hence, lytic growth cannot arise until after germination (Sonenshein, 2006).

1.5.2 Genome Features and Morphology of C. difficile *Phages*

Phages able to infect *C. difficile* currently described in the literature belong to the *Caudovirales*, an order which can be further split into three families on the basis of tail type. Described *C. difficile* phages are confined to just two of these families, the *Myoviridae* and *Siphoviridae*. The first *C. difficile* phage genome to be sequenced was ΦCD119 (Govind *et al*., 2006) and since then the number of sequenced genomes has increased rapidly. Currently (August 2017), the NCBI

database contains 25 full *C. difficile* phage genomes, of which seven are classified as *Siphoviridae* and 18 as *Myoviridae* (Table 1.1). The published genomes range from 31,674 bp to 131,326 bp in size, have a GC content between 26.4 and 30.8 % and are all composed of double-stranded DNA.

Table 1.1: List of the *C. difficile* phages with published full genome sequences available on NCBI

Phage Name	**Classification**	**Genome Size (bp)**	**Linear or Circular Genome Map**	**GC Content (%)**	**Reference and/or NCBI Accession Number**
ΦCD119	*Myoviridae*	53,325	Linear	28.8	Govind *et al.* (2006) NC_007917.1
ΦC2	*Myoviridae*	56,538	Linear	28.7	Goh *et al.* (2007) NC_009231.1
ΦCD27	*Myoviridae*	50,930	Linear	29.3	Mayer *et al.* (2008) NC_011398.1
ΦMMP02	*Myoviridae*	48,396	Circular	29.6	Meessen-Pinard *et al.* (2012) NC_019421.1
ΦMMP04	*Myoviridae*	31,674	Circular	30.0	Meessen-Pinard *et al.* (2012) NC_019422.1
ΦCDHM19	*Myoviridae*	54,295	Circular	28.5	Hargreaves *et al.* (2014) NC_028996.1
ΦCDMH1	*Myoviridae*	54,279	Linear	28.4	Hargreaves *et al.* (2014b) NC_024144.1
ΦCDHM11	*Myoviridae*	32,000	Circular	30.5	Hargreaves and Clokie (2015) NC_029001.1
ΦCDIF1296T	*Myoviridae*	131,326	Circular	26.4	Wittmann *et al.* (2015) CP011970.1

ΦCDKM15	*Myoviridae*	50,606	Linear	29.0	Rashid *et al.* (2016) KX228400.1
ΦCDKM9	*Myoviridae*	49,822	Linear	29.0	Rashid *et al.* (2016) KX228399.1
ΦCD505	*Myoviridae*	49,316	Linear	29.4	NC_028764.1
ΦCD211	*Myoviridae*	131,326	Linear	26.4	NC_029048.1
ΦCD506	*Myoviridae*	33,274	Linear	29.6	NC_028838.1
ΦMMP01	*Myoviridae*	44,461	Linear	28.9	NC_028883.1
ΦCD481-1	*Myoviridae*	32,846	Linear	30.3	NC_028951.1
ΦMMP03	*Myoviridae*	52,261	Linear	28.9	NC_028959.1
ΦCDHM14	*Myoviridae*	32,651	Circular	29.5	LK985321.1
ΦCD6356	*Siphoviridae*	37,664	Linear	28.4	Horgan *et al.* (2010) NC_015262.1
ΦCD38-2	*Siphoviridae*	41,090	Linear	30.8	Sekulovic *et al.* (2011) NC_015568.1
ΦCD111	*Siphoviridae*	41,560	Linear	30.9	NC_028905.1
ΦCD146	*Siphoviridae*	41,507	Linear	30.7	NC_028958.1
ΦCD24-1	*Siphoviridae*	44,129	Linear	28.5	LN681534.1
ΦCDSH1	*Siphoviridae*	41,619	Linear	30.8	KU057941.1
ΦCDHM13	*Siphoviridae*	33,596	Circular	30.0	NC_029116.1

Phage genomes can be circularly permuted meaning that when the DNA is replicated as a long concatemer of DNA it is then packaged into the phage capsid until it is full, by a headful packaging mechanism. In this mechanism more than the full length of DNA is packaged into each capsid and the DNA concatemer is cut when the capsid is full. In this case, the genome in each phage capsid will have different "ends." Phage genomes packaged in this way can be represented by a circular genome map. Alternatively, the full length of phage DNA can be packaged into the phage capsid with each phage particle containing a genome with the same defined ends. These ends could be single stranded cohesive ends, short or long exact direct terminal repeats, terminal host DNA sequences or

covalently bound terminal proteins. In Table 1.1, genomes with suspected headful packaging methods are denoted by a circular genome map whereas phages with defined ends or unknown packaging mechanisms are denoted by a linear genome map.

In common with other phages, the genomes of *C. difficile* phages are organised into modules of genes with related functions. The identification of integrase genes in all published *C. difficile* phage genomes confirms that the phages can enter the lysogenic life cycle and are therefore temperate phages (Hargreaves and Clokie, 2014), although no work has been published to show whether these genes are active or functional.

The endolysin gene from *C. difficile* phage ΦCD27 has been purified and was shown to be able to lyse a diverse range of *C. difficile* strains (Mayer *et al.*, 2008), suggesting an alternative approach of using phage for the treatment of CDI. Interestingly, studying *C. difficile* phage genomes can help determine gene evolution in *C. difficile*. It has been noted that phage holin genes have high homology to the toxin accessory gene, *tcdE* (Goh *et al*., 2007), which may indicate that PaLoc has phage origins but raises concerns that phage therapy could aid in the transfer of toxin genes and the production of more hypervirulent *C. difficile* strains.

1.5.3 Impact of C. difficile *Lysogens*

Since all reported phages of *C. difficile* are temperate in nature, concerns have been raised over the safety of phages for therapy that could alter the phenotype

of host cells on integration into the chromosome, as the lysogenic cells would survive treatment. Sekulovic and colleagues (Sekulovic *et al*., 2011) reported an increase in the rate of accumulation and concentration of extracellular toxins in lysogens containing phage ΦCD38-2 in comparison to the parental non-lysogenic strain. Conversely, Govind and colleagues (Govind *et al*., 2009) reported a downregulation of PaLoc genes in strains lysogenised by phage ΦCD119 compared to parental non-lysogenic strains. It was hypothesised that this was due to the binding of a transcriptional regulator, encoded on the phage genome, to a promoter region upstream of *tcdR* in the host cell chromosome. In addition to this, Williams and colleagues showed that when *C. difficile* strain NCTC11209 was lysogenised with phage ΦCD27 the lysogenic strain produced considerably less combined toxin A and B than the wild-type strain. It was suggested that this could demonstrate cross-talk between the *C. difficile* PaLoc and mobile genetic elements (Williams *et al*., 2013). This evidence suggests that the concern for phage treatment creating more toxic *C. difficile* strains could be misplaced as it does not occur with all phages.

1.5.4 In Vitro and In Vivo Studies

In 1999, Ramesh and co-workers (Ramesh *et al*., 1999) studied the effectiveness of phage therapy in a hamster model of *C. difficile*-induced ileocecitis, using a phage induced from lysogenic *C. difficile* 135I. In this study, hamsters were administered clindamycin and subsequently challenged with *C. difficile*. The *C. difficile* strain 135I was selected based on its ability to cause disease and its susceptibility to phage infection. A proportion of the hamsters were then administered phage (10^8 pfu/ml) every 8 hours for 72 hours. Hamsters that were

not treated with phage died within 96 hours whereas hamsters that received phage survived. This presents the first confirmation that phage therapy could be an effective treatment for antibiotic-induced *C. difficile* disease (Ramesh *et al*., 1999).

To test the impact of phage therapy on the gut microbiome and its efficiency at removing *C. difficile,* Meader and co-workers (Meader *et al*., 2010) created a batch fermentation model to replicate the conditions within the colon, using phage ΦCD27 and the sensitive *C. difficile* 11204 strain. In comparison to when metronidazole was used in the model, there was no significant impact on the commensals, namely *Bifidobacterium* sp., *Lactobacillus* sp. and *Bacteroides* sp., in the model when phages were used and a significant reduction in CFU recovery of *C. difficile* was reported. This is promising for the prospect of phage therapy for the treatment of CDI (Meader *et al*., 2010).

Recently, a new model has been developed to assess the efficiency of phages in the treatment of CDI using the wax moth, *Galleria mellonella* (Nale *et al*., 2016b). This organism was selected over use of the more standard hamster model, as there are no ethics surrounding its use and it is cheap and easy to obtain. In this study, a combination of four *C. difficile* phages (ΦCDHM1, ΦCDHM2, ΦCDHM5 and ΦCDHM6) were successfully used to prevent colonisation of *C. difficile* CD105LC2 (PCR ribotype 014) in the model. It was shown that the phages could prevent colonisation when used alone and when used in conjunction with vancomycin. These results offer another promising step for phage therapy for the treatment of CDI (Nale *et al*., 2016b). It is important

to note in this case that these phages are temperate in nature and it does not appear that their temperate nature is preventing them from having the desired therapeutic effect.

1.6 Project Aims

The long-term goal of this project is to create a therapeutic that can be used to treat CDI, without the need for antibiotics, using either traditional phage therapy or phages as a delivery vehicle for alternative "killing" agents. This project was designed to allow fundamental research which will contribute to this long-term goal using four main objectives:

1. Isolation of *C. difficile* strains and phages, that can exert lytic activity on *C. difficile* clinically relevant isolates, from the Nottinghamshire environment. These *C. difficile* isolates will be characterised in terms of PCR ribotype and *C. difficile* phage propagating strains will be characterised in terms of growth profile.
2. Characterisation of isolated phages in terms of morphology and therapeutic attributes, including determination of host range.
3. Genotypic characterisation of isolated phages in terms of genome sequencing and identification of genes of interest, such as the genes encoding integrase enzymes.
4. Identification of the phage receptor protein on the *C. difficile* bacterial cell surface using genome analysis and genetic tools.

Chapter Two:

Methods and Materials

2 Methods and Materials

2.1 Bacterial and Plasmid Strains

2.1.1 Bacterial strains

The full list of the bacterial strains used in this study are in Table 2.1. The bacterial strains used for phage host range testing can be found in appendix 8.1.

Table 2.1: Complete list of *Clostridium difficile* strains used for isolation of bacteriophages from environmental samples

***C. difficile* Strain Name**	**PCR Ribotype**	**Location of Isolation (if known)**	**Source/Reference/ Holding Location**
30011	001	Nottingham	University of Nottingham, N. Minton
3401	001	Nottingham	University of Nottingham, N. Minton
08011	002	Nottingham	University of Nottingham, N. Minton
1101	002	Nottingham	University of Nottingham, N. Minton
630 WT	012	Zurich	University College London, Peter Mullany
2301	014	Nottingham	University of Nottingham, N. Minton
2601	014	Nottingham	University of Nottingham, N. Minton
0101	015	Nottingham	University of Nottingham, N. Minton
0301	015	Nottingham	University of Nottingham, N. Minton
522418	023	Nottingham	University of Nottingham, N. Minton
520227	023	Nottingham	University of Nottingham, N. Minton
520577	023	Nottingham	University of Nottingham,

			N. Minton
522682	023	Nottingham	University of Nottingham, N. Minton
DH478	027	Taunton	Anaerobic Reference Laboratory, Cardiff V. Hall
DH349	027	Cambridge	Anaerobic Reference Laboratory, Cardiff V. Hall
DH320	027	Newcastle	Anaerobic Reference Laboratory, Cardiff V. Hall
DH1396	027	Slough	Anaerobic Reference Laboratory, Cardiff V. Hall
DH1432	027	Barnet	Anaerobic Reference Laboratory, Cardiff V. Hall
DH361	027	Lewisham	Anaerobic Reference Laboratory, Cardiff V. Hall
DH131	027	Manchester	Anaerobic Reference Laboratory, Cardiff V. Hall
DH482	027	Oxford	Anaerobic Reference Laboratory, Cardiff V. Hall
DH1916	027	Torbay	Anaerobic Reference Laboratory, Cardiff V. Hall
R20291	027	Stoke Mandeville	Anaerobic Reference Laboratory, Cardiff V. Hall
2801	027	Nottingham	University of Nottingham, N. Minton
5101	027	Nottingham	University of Nottingham, N. Minton
9201	027	Nottingham	University of Nottingham,

			N. Minton
7009825	078	The Netherlands	Leiden University Medical Centre (LUMC), E. Kuijper
1801	078	Nottingham	University of Nottingham, N. Minton
3801	078	Nottingham	University of Nottingham, N. Minton

Table 2.2: Complete list of bacterial strains used throughout this study

Strain	**Description/PCR Ribotype**	**Source/Reference**
Escherichia coli TOP10	Plasmid storage strain	Invitrogen
Escherichia coli NEB5α	Plasmid storage strain	New England Biolabs (NEB)
Escherichia coli CA434	Conjugative transfer strain – *E. coli* HB101 carrying R702	University of Nottingham, N. Minton (Purdy *et al*., 2002)
Escherichia coli NEB SExpress	Conjugative transfer strain – *E. coli* NEB Express containing R702 R factor from *E. coli* CA434	University of Nottingham, N. Minton
Clostridium difficile 630Δ*erm*	*C. difficile* 630 erythromycin sensitive spontaneous mutant, PCR ribotype 012	University of Nottingham, N. Minton (Hussain *et al*., 2005)
Clostridium difficile CD2315	Phage receptor investigation, PCR ribotype 078	Isolation Location: Hungary, Leiden University Medical Centre (LUMC), E. Kuijper
Clostridium difficile 2016	Phage receptor investigation, PCR ribotype 078	Isolation Location: Ireland, Leiden University Medical Centre (LUMC), E. Kuijper
Clostridium difficile 7009825	Phage receptor investigation, PCR ribotype 078	Isolation Location: The Netherlands, Leiden University Medical Centre (LUMC),

		E. Kuijper
Clostridium difficile 31662	Phage receptor investigation, PCR ribotype 027	Isolation Location: The Netherlands, Leiden University Medical Centre (LUMC), E. Kuijper
Clostridium difficile SI1	*C. difficile* isolated from environmental sample, PCR ribotype 078	This study
Clostridium difficile SI2	*C. difficile* isolated from environmental sample, PCR ribotype 039	This study
Clostridium difficile SI3	*C. difficile* isolated from environmental sample, PCR ribotype 711	This study
Clostridium difficile SI5	*C. difficile* isolated from environmental sample, PCR ribotype 011	This study
Clostridium difficile SI8	*C. difficile* isolated from environmental sample, PCR ribotype 009	This study

2.1.2 Plasmids

Table 2.3: Complete list of the plasmids used throughout this study

Plasmid	**Description**	**Source/Reference**
pMTL82151	Modular plasmid containing pBP1, ColE1+*tra*, Cm^R/Tm^R	Heap *et al*., 2009
pMTL84151	Modular plasmid containing pCD6, ColE1+*tra*, Cm^R/Tm^R	Heap *et al*., 2009
pMTL84153	Modular plasmid containing pCD6, ColE1+*tra*, Cm^R/Tm^R, ferredoxin promoter (P_{fdx})	Heap *et al.*, 2009
pMTL82123	Modular plasmid containing pBP1, p15a+*tra*, Cm^R/Tm^R, P_{fdx}	Heap *et al.*, 2009
pMTLMW1	Modular plasmid containing gene encoding putative membrane protein from *C. difficile* 1801	This study

pMTLMW3	Modular plasmid containing gene encoding putative membrane protein from *C. difficile* 1801 under control of P_{fdx}	This study
pMTLMW4	Modular plasmid containing gene encoding putative membrane protein from *C. difficile* 1801 under control of P_{fdx}, with p15a+*tra*	This study
pJAK002	Modular plasmid containing ColE1+*tra*, pCD6, Cm^R/Tm^R, *slpA* from S layer cassette H2/6 under control of tetracycline inducible promoter (P_{tet})	University of Sheffield, R. Fagan (Fagan and Fairweather, 2011)
pJAK018	Modular plasmid containing ColE1+*tra*, pCD6, Cm^R/Tm^R, *slpA* from S layer cassette 6 under control of P_{tet}	University of Sheffield, R. Fagan (Fagan and Fairweather, 2011)
pJAK023	Modular plasmid containing ColE1+*tra*, pCD6, Cm^R/Tm^R, *slpA* from S layer cassette 2 under control of P_{tet}	University of Sheffield, R. Fagan (Fagan and Fairweather, 2011)
pMTL007C-E2	ClosTron expression vector containing directed Group II intron with Erm RAM, flanked by FRT sites, pCB102, ColE1, Cm^R/Tm^R	Heap *et al*., 2007
pMTL007C-E2::Cdi1801_ repressor_61/62a	ClosTron vector targeting Group II insertional knockout to the repressor protein of *C. difficile* 630Δ*erm*	This study
pMTL007C-E2::Cdi1801_ repressor_417/418s	ClosTron vector targeting Group II insertional knockout to the repressor protein of *C. difficile* 630Δ*erm*	This study

2.2 Oligonucleotide Primers

Table 2.4: Complete list of the oligonucleotide primers used throughout this study for PCR amplification and sequencing

Oligonucleotide	**Sequence (5' – 3')**	**Function**	**Nucleotide Position of Primer Bind**
D*spo0A*-FA1	GTCTAATGAAGA TGATATCGATATA TTAGGCATAGCTA AGGATGG	*C. difficile* specific primers to amplify *spo0A*	34 – 79 *spo*0A gene Multiple *C. difficile* strains
D*spo0A*-RS1	CTGAATTAGTTGG TTTTCCTTTAGTA TTGTGTACCGTAT ATCC	*C. difficile* specific primers to amplify *spo0A*	727 – 770 *spo*0A gene Multiple *C. difficile* strains
Ribotyping_P3	CTGGGGTGAAGT CGTAACAAGG	For PCR ribotyping of *C. difficile* strains	1447 – 1469 *16S rRNA* gene Multiple *C. difficile* strains
Ribotyping_P5	GCGCCCTTTGTAG CTTGACC	For PCR ribotyping of *C. difficile* strains	1 – 20 *23S rRNA* gene Multiple *C. difficile* strains
pBP1_R1	CTTCATTAAATGC CTTAGAATC	Sequencing primer annealing to Gram positive replicon in pMTL82151	484 – 505 Plasmid pMTL82151
ColE1+*tra*_F2	CCATCAAGAAGA GCGAC	Sequencing primers annealing to Gram negative replicon in pMTL82151	5183 – 5199 Plasmid pMTL82151
CD1801_gp14_F	GGCAGTTTTCAGA CAGATATATAC	Amplification of gene coding for putative phage protein from phage ΦCD1801	35208 – 35231 Phage ΦCD1801 (Unpublished genome sequence)
CD1801_gp14_R	CTCCCTACTCCAA ACTTCTC	Amplification of gene coding for putative phage protein from ΦCD1801	36052 – 36071 Phage ΦCD1801 (Unpublished genome sequence)

CD1801_gp20_F	CGTGTTAGTACAGAAGAACAAGC	Amplification of integrase gene from phage ΦCD1801	30795 – 30817 Phage ΦCD1801 (Unpublished genome sequence)
CD1801_gp20_R	GTCTGCGTTCAACCCTGC	Amplification of integrase gene from phage ΦCD1801	29554 – 29571 Phage ΦCD1801 (Unpublished genome sequence)
CD1801_gp46_F	CACTAGCGTACGGACAAG	Amplification of gene coding for phage capsid protein from phage ΦCD1801	6459 – 6476 Phage ΦCD1801 (Unpublished genome sequence)
CD1801_gp46_R	CCCACAGGTTTAGCATC	Amplification of gene coding for phage capsid protein from phage ΦCD1801	7367 – 7383 Phage ΦCD1801 (Unpublished genome sequence)
Membraneprotein_CD2315_F1	CAATACAGATATTACAATTGC	Amplification and sequencing primer to confirm sequence of gene coding for putative membrane protein in *C. difficile* CD2315	3201288 – 3201308 *C. difficile* 2315 (unpublished genome sequence)
Membraneprotein_CD2315_R1	GGAATGATAATAGCCTCAG	Amplification and sequencing primer to confirm sequence of gene coding for putative membrane protein in *C. difficile* CD2315	3201872 – 3201890 *C. difficile* 2315 (unpublished genome sequence)
CD1801PMP_F_*Sac*I	TTTTTTGAGCTCGCGTAATCAAATCATTGTA	Amplification of gene coding for putative membrane protein from *C. difficile* 1801	3201092 – 3201110 *C. difficile* 1801 (unpublished genome sequence)
CD1801PMP_R_*Xho*I	AAAAAACTCGAGGCATTAGGTATAACTTTAGGTTC	Amplification of gene coding for putative membrane protein from *C. difficile* 1801	3202198 – 3202220 *C. difficile* 1801 (unpublished genome sequence)

CD1801PMP_R_ *Nde*I	TTTTTTCATATGAGCAAAGTAAAATCTATATAC	Amplification of gene coding for putative membrane protein from *C. difficile* 1801	3201801 – 3201821 *C. difficile* 1801 (unpublished genome sequence)
CatP_F1	GATAAATAGTTAACTTCAGGTTTGTC	Sequencing primer annealing to the *catP* gene of pMTLMW4	5453 – 5478 Plasmid pMTLMW4
p15a_R1	CAGAGAACCTTCGAAAAAC	Screening primer for amplification of p15a + *tra* in pMTLMW4	5628 – 5646 Plasmid pMTLMW4
pCD6_F1	GGAGTTTGAACCAATATTGG	Screening primer for amplification of p15a + *tra* in pMTLMW4	2045 – 2064 Plasmid pMTLMW4
M13F	TGTAAAACGACGGCCAGT	Sequencing primers annealing to *lacZ* alpha fragment of pMTL84151	199 – 216 Plasmid pMTL84151
M13R	CAGGAAACAGCTATGACC	Sequencing primers annealing to terminator of pMTL84151	63 – 80 Plasmid pMTL84151
pRPF_ColE1_F1	GGATAACGCAGGAAAGAAC	Amplification of Gram negative replicon, ColE1, from pJAK002, pJAK018 and pJAK023	8491 – 8509 Plasmid pJAK002, pJAK018 and pJAK023
pRPF_ColE1_R1	GTTCCACTGAGCGTCAGACC	Amplification of Gram negative replicon, ColE1, from pJAK002, pJAK018 and pJAK023	9183 – 9202 Plasmid pJAK002, pJAK018 and pJAK023
pRPF_*tra*_F1	CGAAGTCGCTCTTCTTGATG	Amplification of Gram negative replicon, *tra*, from pJAK002, pJAK018 and pJAK023	6613 – 6632 Plasmid pJAK002, pJAK018 and pJAK023

pRPF_*tra*_R1	CTTTCCTTGGTGTATCCAACG	Amplification of Gram negative replicon, *tra*, from pJAK002, pJAK018 and pJAK023	7219 – 7232 Plasmid pJAK002, pJAK018 and pJAK023
pRPF_CD6_F1	CGGCTTAATTTGAATACTTTG	Amplification of Gram positive replicon, *orfB* and *repA*, from pJAK002, pJAK018 and pJAK023	414 – 424 Plasmid pJAK002, pJAK018 and pJAK023
pRPF_CD6_R1	CGGTCTAACTGATTAAGTATC	Amplification of Gram positive replicon, *orfB* and *repA*, from pJAK002, pJAK018 and pJAK023	2351 – 2371 Plasmid pJAK002, pJAK018 and pJAK023
*slpA*_RT002_F	CAGTTGTAATGTCTGCCG	Amplification of *slpA* gene from *C. difficile* 08011	47 – 64 *slpA* gene (Accession No. DQ060628)
*slpA*_RT002_R	GCTATTCCTTTACCAACTTGAG	Amplification of *slpA* gene from *C. difficile* 08011	2144 – 2165 *slpA* gene (Accession No. DQ060728)
*slpA*_RT023_F	GCTATGGCAGCTGTTACTG	Amplification of *slpA* gene from *C. difficile* 522418	132 – 150 *slpA* gene (Accession No. DQ060630)
*slpA*_RT023_R	GCAACCGATTGATCAGAAG	Amplification of *slpA* gene from *C. difficile* 522418	1816 – 1834 *slpA* gene (Accession No. DQ060630)
*slpA*_RT014_078_F	GCTATGGCTGCTGTTACAG	Amplification of *slpA* gene from *C. difficile* 2301 and *C. difficile* 1801	132 – 150 *slpA* gene (Accession No. DQ060643)

*slpA*_RT014_R	GCTATACCTTGACCAACTTG	Amplification of *slpA* gene from *C. difficile* 2301	6272 – 6291 *slpA* gene from S layer cassette 6, pJAK018
*slpA*_RT078_F	GCTAATATGATTGGTGCTGG	Amplification of *slpA* gene from *C. difficile* 1801	2127 – 2149 *slpA* gene (Accession No. DQ060643)
*slpA*_RT027_F	CCATCTTTATCTACTGGCTTAC	Amplification of *slpA* gene from *C. difficile* CD31662	3162367 – 3162388 *C. difficile* 31662 (unpublished genome sequence)
*slpA*_RT027_R	GTTTGCAGCAGAAGATATG	Amplification of *slpA* gene from *C. difficile* CD31662	3164355 – 3164373 *C. difficile* 31662 (unpublished genome sequence)
630repressorCT_F	CACCTCGTTGCTTATTTG	Screening primer for pMTL007C-E2::Cdi1801_repressor_61/62a	3383563 – 3383580 *C. difficile* 630Δ*erm* (Accession No. NZ_LN614756.1)
630repressorCT_R	GAAATTAAAGGAACTTCTCTAAC	Screening primer for pMTL007C-E2::Cdi1801_repressor_61/62a	3384235 – 3384257 *C. difficile* 630Δ*erm* (Accession No. NZ_LN614756.1)
CD08011_ParA1	CGTATTAAGGCTTGTACTC	Amplification of ΦCD08011 for genome closure	20781 – 20799 Phage ΦCD08011 (Unpublished genome sequence)
CD08011_pol1	GCAAGAGCTGTTCATCAG	Amplification of ΦCD08011 for genome closure	22208 – 22225 Phage ΦCD08011 (Unpublished genome sequence)
CD1801_portal1	GCTGATATAGCTAGAAGACAAG	Amplification of ΦCD1801 for genome closure	1469 – 1490 Phage ΦCD1801 (Unpublished genome sequence)
CD1801_portal2	GACTTAGTTGCTATATCTGC	Amplification of ΦCD1801 for genome closure	3297 – 3316

			Phage ΦCD1801 (Unpublished genome sequence)
CD418_phage prot1	GGTCACATATAATTCTATTTG	Amplification of ΦCD418 for genome closure	24970 – 24990 Phage ΦCD418 (Unpublished genome sequence)
CD418_collar1	CGATTGATGAAGTTGTTAG	Amplification of ΦCD418 for genome closure	22851 – 22869 Phage ΦCD418 (Unpublished genome sequence)
CD2301_ endolysin1	GCAGAAGTACATGCTCCAC	Amplification of ΦCD2301 for genome closure	26087 – 26105 Phage ΦCD2301 (unpublished genome sequence)
CD2301_cellwall 1	CTCAAATTGGTGGAAATGG	Amplification of ΦCD2301 for genome closure	23709 – 23727 Phage ΦCD2301 (unpublished genome sequence)
CD2301_end1	GCTTACTATTAATCCTATTATATTG	Amplification of ΦCD2301 for genome closure	24773 – 24797 Phage ΦCD2301 (unpublished genome sequence)
CD2301_holin1	CTTTAACATGCTGGTAAGG	Amplification of ΦCD2301 for genome closure	25681 – 25699 Phage ΦCD2301 (unpublished genome sequence)
*slpA*_motif_F1	GTATTTAATCTCTTTCCATCTGC	Amplification of *slpA* gene from *C. difficile* 051223046	3119949 – 3119971 *C. difficile* 051223046 (unpublished genome sequence)
*slpA*_motif_R1	GTAATTTTAATATAATGTTGGGAGG	Amplification of *slpA* gene from *C. difficile* 051223046	3120957 – 3120981 *C. difficile* 051223046 (unpublished genome sequence)
*RS14955*_F1	GCTATATTGTTATAAGCAGTC	Amplification of the putative ΦCD1801	3239913 – 3239933

		prophage integration site in *C. difficile* 1801L	*C. difficile* 1801 (unpublished genome sequence)
CD1801_*attP*_R1	GAATTAGTTAATGTTCAACAAGG	Amplification of the putative ΦCD1801 prophage integration site in *C. difficile* 1801L	29203 – 29225 Phage ΦCD1801 (unpublished genome sequence)
*RS14960*_R1	GAAGTTGATTTAAGTTCTGC	Amplification of the putative ΦCD1801 prophage integration site in *C. difficile* 1801L	3240368 – 3240387 *C. difficile* 1801 (unpublished genome sequence)
CD1801_*attP*_F1	GGAGAGTCAAATTATTGATAC	Amplification of the putative ΦCD1801 prophage integration site in *C. difficile* 1801L	29628 – 29648 Phage ΦCD1801 (unpublished genome sequence)

2.3 General Microbiology Techniques

2.3.1 Media and Buffers

The composition of the media and buffers used within this study are detailed in Table 2.5. The media and buffers were dissolved in distilled water (dH_20, up to 1 l or adjusted for the volume required) and sterilised by autoclaving at 121 °C for 20 minutes. If necessary, the pH was measured and adjusted prior to the volume being made up and prior to autoclaving. If required, antibiotics and/or other components were added to the media post sterilisation but prior to pouring of plates, media was cooled to 50 °C prior to the addition of any extra components. Media and buffer compounds were purchased from Oxoid, unless otherwise stated.

Table 2.5: List of the media and buffers used throughout this study

Medium/Buffer	**Constituents**	**Quantity (g/l)**
Lysogeny Broth (LB)	Tryptone	10.00
	Yeast Extract	5.00
	Sodium Chloride	10.00
LB agar	Tryptone	10.00
	Yeast Extract	5.00
	Sodium Chloride	10.00
	Agar Bacteriological No. 1	15.00
Brain Heart Infusion Supplemented (BHIs) Broth	BHI	37.00
	Yeast Extract	5.00
	L-Cysteine Hydrochloride Monohydrate (Fisher Scientific)	1.00
BHIs Agar	BHI Agar	47.00
	Yeast Extract	5.00
	L-Cysteine Hydrochloride Monohydrate	1.00
BHIs Top Agar (0.5 %)	BHI	37.00
	Yeast Extract	5.00
	L-Cysteine Hydrochloride Monohydrate	1.00
	Agar Bacteriological No. 1	5.00
BHIs Top Agar (0.2 %)	BHI	37.00
	Yeast Extract	5.00
	L-Cysteine Hydrochloride Monohydrate	1.00
	Agar Bacteriological	2.00
Brazier's Agar	Brazier's Agar	47.50
Phosphate Buffered Saline (PBS), pH 7.4	NaCl	8.00
	KCl	0.20
	$NaHPO_4$	1.44
	KH_2PO_4	0.24
Phage Buffer, pH 7.4	50 mM Tris-HCl	50 ml
	8 mM $MgSO_4.7H_2O$	2.00
	100 mM NaCl	5.80

2.3.2 Antibiotic Supplements

Antibiotics were prepared to a standard stock concentration and stored at the appropriate temperature for no longer than one month. Where required, antibiotics were sterilised using 0.22 µm membrane filters (Millipore). The stock concentrations and working concentrations of the antibiotic supplements used in this study are listed in Table 2.6.

Table 2.6: Stock and working concentrations of antibiotic supplements used throughout this study

Antibiotic	**Solvent**	**Stock Concentration (mg/ml)**	**Working Concentration for *E. coli* (µg/ml)**	**Working Concentration for *C. difficile* (µg/ml)**
Chloramphenicol (Cm)	100% EtOH	25	12.5 (broth) 25 (agar)	-
Thiamphenicol (Tm)	50% EtOH	15	-	15
Erythromycin (Erm)	100% EtOH	50	500	10

2.3.3 Aerobic Bacterial Strains

2.3.3.1 Growth Conditions

E. coli strains were grown in LB broth or agar overnight at 37 °C; liquid cultures were grown with agitation at 200 rotations per minute (rpm). Where appropriate, media was supplemented with the suitable antibiotic for selection of plasmids.

2.3.3.2 Strain Storage and Revival

E. coli strains, including those harbouring plasmid DNA, were stored at -80 °C on beads using Microbank™ Long Term Bacterial and Fungus Storage System (Pro-lab Diagnostics). When required a 5 ml LB broth culture containing the appropriate antibiotic was inoculated using one bead and incubated at 37 °C overnight with agitation (2.3.3.1).

2.3.3.3 Preparation of E. coli *Electrocompetent Cells*

E. coli TOP10 and CA434 liquid cultures (5 ml) were prepared in LB broth and incubated overnight at 37 °C under agitation. A 1 % inoculum of the overnight culture was used to inoculate 200 ml of pre-warmed LB broth. This was incubated at 37 °C under agitation for three hours, until optical density (OD) at 600 nm was 0.5-0.7. Cells were harvested by centrifugation at 3709 x *g* for 10 minutes at 4 °C. The cells were subsequently washed by re-suspension of the cell pellet in 20 ml ice cold dH_2O. Cells were harvested under the same conditions as before and the wash procedure was completed twice. The second wash was performed using 10 ml ice cold dH_2O. The final cell pellet was re-suspended in 1 ml sterilised 10 % glycerol and stored in 50 µl aliquots at -80 °C.

2.3.3.4 Transformation of E. coli *by Electroporation*

Electrocompetent *E. coli* cells (TOP10 and CA434) previously prepared, were thawed slowly on ice. Plasmid (100 ng) or dialysed ligation mixture (5 µl) was added to the thawed electrocompetent cells and gently mixed. This was then transferred to a pre-chilled electroporation cuvette (0.2 cm gap, Biorad) and pulsed with an electroporator (Biorad Gene Pulser). The electroporator

conditions were set to 200 Ω, 2.5 kV and 25 µF capacitance. Immediately, 400 µl of Super Optimal Catabolite (SOC) medium (Invitrogen) was added to the pulsed cells and transferred to a 1.5 ml microcentrifuge tube. The cells were allowed to recover for two hours at 37 °C with agitation before being plated (100 µl) onto LB agar supplemented with the appropriate antibiotic for the plasmid and incubated overnight at 37 °C or for 48 hours at 30 °C. Undiluted and a 10^{-1} dilution of the recovery was plated, the dilution was created in SOC medium.

2.3.3.5 Chemical Transformation of E. coli

E. coli chemically competent cells (*E. coli* NEB 5α) were purchased from NEB. *E. coli* NEB SExpress (strain NEB Express containing the R702 R-factor derived from CA434) was created by C. Woods (University of Nottingham, Minton group) and were proved as chemically competent cells as described in the thesis of C. Woods (University of Nottingham, thesis 2018). Chemical transformation was conducted as stated in the NEB protocol. Briefly, competent cells (50 µl) were thawed on ice for 10 minutes before 100 ng plasmid was added and gently mixed by flicking. The cells were incubated on ice for 30 minutes and then heat shocked at 42 °C for 30 seconds. The time and duration of this heat shock is essential for successful transformation. Immediately after heat shock, cells were incubated on ice for five minutes before 950 µl SOC medium was added and incubated at 37 °C with agitation for 1 hour. The recovered cells were plated (100 µl) onto LB agar plates supplemented with the appropriate antibiotic for the plasmid overnight at 37 °C.

2.3.3.6 Selection of Clones using Blue/White Screening

Blue/white screening was used to determine successful insertion, disrupting *lacZ*, into the modular plasmid vector pMTL82151. Plasmid DNA was transferred to *E. coli* TOP10 by electroporation as described 2.3.3.4. The recovered cells were plated onto LB agar supplemented with the antibiotic appropriate for plasmid selection, X-Gal (20 mg/ml, Thermo Scientific) and 100 mM IPTG solution (Thermo Scientific), X-Gal and IPTG were purchased already made to the correct concentrations for use and were stored as per the manufacturer's instructions. Plates were incubated at 37 °C for 24-48 hours until colonies were visible. Resulting white colonies indicated successful insertion within *lacZ* whereas blue colonies indicated no insertion.

2.3.4 Anaerobic Bacterial Strains

2.3.4.1 Growth Conditions

C. difficile strains were cultured on pre-reduced BHIs agar plates, supplemented with D-cycloserine (250 µg/ml) and 8 µg/ml cefoxitin (*C. difficile* selective supplement, Oxoid, (Heap *et al*. 2009)), hereafter this antibiotic supplemented media will be referred to as BHIsCC. Cultures were incubated overnight at 37 °C in an anaerobic cabinet (Don Whitley Scientific Limited, 80 % N_2, 10 % CO_2 and 10 % H_2). *C. difficile* strains were also grown in liquid BHIs broth with no antibiotic supplement overnight under anaerobic conditions. Where required, strains were sub-cultured using a 1 % inoculum of overnight liquid culture.

2.3.4.2 Strain Storage and Revival

C. difficile strains were stored at -80 °C by re-suspension of plate overnight growth, harvested using a sterile 10 µl loop, in 1 ml BHIs broth containing 20 % glycerol (Sigma). Strains were revived on BHIsCC agar plates from frozen stocks, using a sterile 10 µl loop, and allowed to grow as described in 2.3.4.1. Prior to inoculation of media with *C. difficile,* broth or agar plates were pre-reduced under anaerobic conditions for at least four hours.

2.3.4.3 Growth Curve Profiles

To measure the growth of *C. difficile* over time, 10 µl of a 1 ml BHIs overnight culture was used to inoculate fresh pre-reduced BHIs broth (190 µl) in a 96 well microtitre plate. The plate was sealed under anaerobic conditions and transferred to the Promega plate reader. The plate was incubated at 37 °C for the duration of the experiment and cells were agitated for a few seconds prior to each measurement. Cell density, at 600 nm, was measured every hour for 24 hours. Measurements were conducted in triplicate.

2.3.4.4 Transfer of DNA to C. difficile *via Conjugation*

Plasmid DNA was transferred from *E. coli* CA434 or *E. coli* SExpress (NEB Express + R702) into *C. difficile* using conjugation, using a method adapted from Purdy *et al.* (2002). A 5 ml overnight culture of *E. coli* donor strain, with the appropriate antibiotic for plasmid selection, was harvested by centrifugation at 1,377 x *g* for two minutes. The cell pellet was washed twice with PBS, centrifugation steps were carried out as before, before the *E. coli* cell pellet was re-suspended in 200 µl *C. difficile* overnight culture, grown in 1 ml BHIs broth.

The mating mix was then plated onto BHIs agar, containing no *C. difficile* selective supplement or other selection, in discrete 20 µl spots. The mating plate was incubated at 37 °C under anaerobic conditions for eight hours (*C. difficile* 630Δ*erm*) or 24 hours (all other *C. difficile* strains). Post incubation, the cell growth was collected, using a sterile 10 µl loop, and re-suspended in 500 µl pre-reduced PBS. The conjugal suspension was plated (100 µl), using a sterile spreader, across five BHIsCC plates supplemented with the appropriate antibiotic selection for the plasmid, usually thiamphenicol or erythromycin. The selection plates were incubated for 24 to 72 hours at 37 °C under anaerobic conditions, until colonies were apparent. Resulting transconjugant colonies were streaked to purity three times on the same selection plates.

2.3.4.5 Isolation of C. difficile *and bacteriophages from environmental samples*

Sewage samples were collected from Stoke Bardolph sewage treatment plant in Nottinghamshire (Severn Trent Water) on two occasions, September 2014 and August 2015. Samples, 1 l, were collected from post-anaerobic digester two sludge. Upon collection, samples were stored at 4 °C until required. An aliquot (50 ml) of the sewage sample was added to the dry components of BHIs broth (weights adjusted for 50 ml) with the addition of 0.1 % taurocholate (Sigma, catalogue number 91223) and 1 % $MgCl_2$ (Sigma, catalogue number M8266). The addition of taurocholate instead of actively growing *C. difficile* cultures into the sewage enrichment means the isolation of phages is reliant upon the *C. difficile* cells already in the sample acting as hosts. The sewage sample was enriched overnight at 37 °C under anaerobic conditions before an aliquot of

40 ml was removed, the remaining 10 ml was discarded. The sample debris and bacterial cells were collected by centrifugation at 10,000 x *g* for 10 minutes at 4 °C. The resulting supernatant was sterilised using a 0.22 µm filter membrane and stored at 4 °C. The presence of phages in the filtered supernatant was determined by plaque assay. Individual phage plaques were passaged to purity by picking an individual plaque into 900 µl BHIs, using a sterile pipette tip, and repeating the plaque assay. Plaques were purified three times.

The cell pellet produced in the centrifugation step was used for isolation of *C. difficile* based on the protocol outlined by (Kim *et al.*, 1981). Briefly, 300 µl of sewage pellet was mixed with 300 µl PBS by vortexing. Subsequently, the sample was heat shocked at 80 °C for 15 minutes prior to centrifugation at 1,500 x *g* for five minutes at room temperature. The sample (50 µl per plate) was plated onto three Brazier's agar plates supplemented with 250 µg/ml D-cycloserine, 8 µg/ml cefoxitin and 4 % egg yolk (Sigma) and incubated overnight at 37 °C under anaerobic conditions. Colonies produced were streaked to purity, three times, on BHIsCC agar plates.

2.3.4.6 Identification and Enumeration of Phages by Plaque Assay

Phages were enumerated using the double agar overlay plaque assay described by Clokie and Kropinski (2009), with some alterations as follows. A 1 % inoculum of *C. difficile* potential host isolates or phage propagating strains overnight culture was transferred to 20 ml fresh pre-reduced BHIs broth and incubated at 37 °C under anaerobic conditions for six hours, OD600nm 0.8-1.0. Subsequently, 1 ml of this was mixed with 200 µl of sewage enrichment lysate

or phage stock and 3 ml of 0.5 % BHIs top agar (molten and incubated at 50 °C for at least 30 minutes prior to use, containing no antibiotic supplement) and poured over 1 % BHIsCC agar plates. Top agar was added to the phage/bacterial mixture outside of the cabinet and plates were poured under aerobic conditions. Top agar was allowed to set for 10 minutes on the bench before plates were transferred to the anaerobic cabinet and incubated at 37 °C overnight. Post incubation, the presence of phages was determined by the observation of plaques and phage titres were enumerated. Where required, phages were diluted in BHIs broth prior to the plaque assay for more accurate titre enumeration. Phage titres were increased by harvesting an almost confluent lysis plate, containing 0.2 % BHIs top agar (with no antibiotic supplement), into BHIs broth and incubating at 37 °C under anaerobic conditions for one hour to allow the phages to diffuse out of the agar. The agar and bacterial cells were removed by centrifugation at 10,000 x g for 10 minutes at 4 °C and the resulting supernatant was sterilised using a 0.22 µm membrane filter.

2.3.4.7 Storage of Phage Stocks

Phages were stored as filtered sterilised supernatant in BHIs broth or phage buffer at 4 °C. Phage stock titres were checked by plaque assay every two months and new phage stocks were created (using 0.2 % BHIs top agar method described in 2.3.4.6) when required, usually every six months.

2.3.4.8 Isolation of C. difficile *Lysogens*

For long-term phage storage, lysogens of the *C. difficile* propagating strains (*C. difficile* 08011, 2301, 1801 and 522418) were isolated. Lysogens were

isolated using the spot of the lawn method described by Govind *et al*. (2009). Confluent growth of *C. difficile* in 0.5 % BHIs top agar was created as described in 2.3.4.6 (without the addition of phages) and placed into the anaerobic cabinet. Next, 20 µl of high titre (>10^8 pfu/ml) phage stock was spotted onto the solidified top agar containing bacteria, three spots were made and incubated overnight at 37 °C under anaerobic conditions. The plates were not inverted to ensure that the phage spots did not run into each other. The next day, five colonies growing within the zone of clearing were picked and re-streaked to purity three times on BHIsCC agar plates. Lysogens were confirmed by induction of prophages (described 2.3.4.9) and were screened for immunity to further phage infection, from the same phage, by plaque assay. To screen for immunity using plaque assay, the potentially lysogenic colonies were used as indicator strains with their corresponding phage.

2.3.4.9 Induction of Prophages using Mitomycin C

Prophages were induced from potentially lysogenic *C. difficile* colonies or strains using Mitomycin C (Fisher Scientific) at a final concentration of 3 µg/ml, as described by Sell *et al.* (1983) and Nale *et al.* (2012) with some alterations. A 20 ml *C. difficile* overnight culture in BHIs broth was induced by the addition of Mitomycin C to a final concentration of 3 µg/ml and incubated for a further 24 hours at 37 °C under anaerobic conditions. Post incubation, cells were pelleted by centrifugation at 3,400 x *g* for 10 minutes at 4 °C and the supernatant was sterilised by passage through a 0.22 µm membrane filter. Filtered induced lysates were stored at 4 °C and tested for the presence of phages by plaque assay.

2.3.4.10 Induction of Genes Under Control of a Tetracyline Inducible Promoter

C. difficile strains containing plasmids pJAK002, pJAK018 and pJAK023 were grown overnight at 37 °C under anaerobic conditions in BHIs broth. The next day, a 1 % inoculum of the overnight culture was used to inoculate 20 ml of pre-reduced BHIs broth and incubated anaerobically at 37 °C for four hours. Anhydrotetracyline (Sigma, 2 mg/ml, made in 100 % EtOH) was used to induce the cultures for one hour at 37 °C under anaerobic conditions at a final concentration of 500 ng/ml as described by Fagan and Fairweather (2011). After induction, the expression of the protein was assessed using binding assays (2.6.5) and plaque assays (2.3.4.6).

2.4 Molecular Techniques

2.4.1 Preparation of Genomic DNA from C. difficile

DNA was isolated from 1 ml BHIs broth overnight cultures of *C. difficile* strains using a GenElute™ Bacterial Genomic DNA Kit (Sigma) per the manufacturer's instructions. Briefly, 1 ml of *C. difficile* overnight culture was harvested by centrifugation and re-suspended in 200 µl lysozyme (10 mg/ml) for 30 minutes at 37 °C. The cells were lysed by the addition of 20 µl Proteinase K (20 mg/ml) and 200 µl lysis solution C, for 10 minutes at 55 °C. The resulting solution was passed through a column and DNA was precipitated using ethanol. DNA was eluted in 20-50 µl dH_20 and stored at -20 °C until required.

C. difficile genomic DNA for sequencing was isolated using a phenol/chloroform procedure. The cells from overnight cultures (3 ml BHIs

broth) of *C. difficile* strains were harvested by centrifugation at maximum speed for two minutes. The resulting pellet was re-suspended in 180 µl lysozyme (10 mg/ml in PBS, Sigma) and incubated for 30 minutes at 37 °C. RNase solution (4 µl, Sigma) was added and incubated at room temperature for 15 minutes. Next, 25 µl Proteinase K (20 mg/ml, Qiagen), 85 µl dH_2O and 110 µl 10 % SDS (Fisher Scientific) were added to the samples and mixed by inversion, before the samples were incubated at 65 °C for 30 minutes. A 400 µl aliquot of phenol:chloroform:isoamylalcohol (25:24:1, Sigma) was added to the samples and mixed by pipette aspiration before being transferred to phase-lock tubes (Quanta Biosciences), followed by centrifugation at maximum speed for three minutes. The top phase was removed and added to another phase-lock tube and the extraction was repeated twice. The final top phase of each sample was added to 40 µl 3 M sodium acetate (Sigma) and 800 µl 100 % ethanol and placed on ice to allow the DNA to precipitate. Next, the samples were incubated at -80 °C for 30 minutes before being centrifuged at maximum speed for 15 minutes. The supernatant was discarded and the pellet was washed, carefully, in 1 ml 70 % ethanol before being centrifuged for a further three minutes at maximum speed. The ethanol was removed carefully so as not to disrupt the pellet and the DNA pellet was left to air dry at room temperature for at least 45 minutes. Finally, the DNA pellet was re-suspended in 50 µl dH_2O and quantified using the Nanodrop™ Lite spectrophotometer (Thermo Scientific). Genomic DNA samples were examined for degradation and RNA contamination using agarose gel electrophoresis (as described 2.4.7).

2.4.2 Preparation of Phage Genomic DNA

Phage nucleic acid was extracted from pure high titre phage lysates (>10^9 pfu/ml) using a RBC Bioscience Viral Nucleic Acid Extraction kit (YVN100) per the manufacturer's instruction. Briefly, 200 µl of pure phage lysate was mixed with 400 µl of VB buffer and incubated for 10 minutes at room temperature. Ethanol was added to the solution and mixed before being passed through a column, twice, to precipitate the DNA. Centrifugation was carried out at 10,000 x g for one minute. The DNA was eluted in 50 µl RNase free water and stored at 4 °C until required.

Phage genomic DNA for sequencing was isolated using a modified phenol/chloroform method (Sambrook and Russell, 2001; Nale *et al.* 2016). A 2 ml aliquot of crude phage lysate (≥10^9 pfu/ml) was mixed with 25 µl $MgCl_2$ (1 M, Sigma), 0.8 µl DNase I (2000 U/ml, Thermo Fisher Scientific) and 20 µl RNAse A (10 mg/ml, Thermo Fisher Scientific) and incubated at room temperature for 30 minutes. Subsequently, 80 µl EDTA (0.5 M, Thermo Fisher Scientific), 5 µl Proteinase K (20 mg/ml, Qiagen) and 100 µl 10 % SDS (Fisher Scientific) were added to the phage-$MgCl_2$ mixture and incubated at 55 °C for one hour. The resulting liquid was aliquoted into four phase-lock tubes (Quanta Biosciences) and extracted three times with an equal volume (500 µl) of phenol:chloroform:isoamylalcohol (25:24:1, Sigma). A final extraction with an equal volume of chloroform (Sigma) was conducted before the DNA was precipitated using 1 ml 100% ethanol and 50 µl sodium acetate (3 M, Sigma) and incubated on ice for five minutes. All centrifugation steps were performed at 14,549 x *g* for five minutes. The precipitated DNA was centrifuged at

14,549 x *g* for 10 minutes and resulting pellet washed with 1 ml 70 % ethanol. The centrifugation step was repeated and the pellet air-dried before the DNA was dissolved in elution buffer (10 mM Tris, Qiagen) at 65 °C for 20 minutes. Eluted DNA was pooled, quantified using NanoDrop™ Lite spectrophotometer (Thermo Scientific) and stored at 4 °C before sequencing.

2.4.3 Preparation of Plasmid DNA from E. coli *and* C. difficile

Plasmid DNA from *E. coli* TOP10 was extracted from a 5 ml overnight culture, LB broth supplemented with the appropriate antibiotic for the plasmid, using GenElute™ Plasmid Miniprep Kit (Sigma) and Monarch® Plasmid Miniprep Kit (New England Biolabs, NEB) per the manufacturer's instructions. Briefly, 3 ml of the 5 ml overnight culture of *E. coli*, harbouring plasmid, was harvested by centrifugation. The cells were re-suspended and lysed for five minutes through the addition of lysis solution. The sample was subsequently neutralised and passed through a column. The plasmid DNA was precipitated using ethanol and eluted in 20-50 µl dH_20. Plasmid DNA was stored at -20 °C until required.

Plasmid DNA from *C. difficile* strains, harbouring plasmids, was isolated from 1 ml BHIs overnight cultures using the same kits. However, after the cells were harvested by centrifugation, the cell pellet was freeze-thawed three times at alternating temperatures of -80 °C and 37 °C for five minutes prior to continuation with the cell lysis step.

2.4.4 Restriction Enzyme Digests

Restriction enzymes and corresponding buffers were purchased from NEB and digests were performed in line with the manufacturer's instructions. Around 1 µg of plasmid DNA, PCR product or phage genomic DNA was digested with 1 µl of the corresponding enzyme and buffer. For double digests, 1 µl of each enzyme (1 unit of enzyme) was added to the reaction and the buffer with the highest percentage activity for both enzymes was selected for use, usually CutSmart® Buffer (1X final concentration). Reactions were made up to a total 20 µl volume with dH_2O. Digests were incubated at 37 °C for two hours or overnight to ensure complete digestion.

2.4.5 De-phosphorylation of Linear Plasmid DNA

Digested plasmid DNA was de-phoshorylated to prevent re-circularisation using Antarctic Phosphatase (NEB). A 1 µl (5 units) aliquot of Antarctic Phosphatase and 1 µl of its buffer (10X) were added to the 20 µl plasmid DNA digest reaction and incubated at 37 °C for 30 minutes.

2.4.6 Ligation of Plasmid DNA and Insert

Digested and purified plasmid DNA and PCR product or phage genomic DNA were mixed in a reaction to a concentration ratio 3:1 of insert to vector, using 50 ng of vector, with 1 µl T4 DNA ligase (Promega, concentration 1-3 U/µl) and 1 µl of its associated buffer (10X). If required, reactions were made up to a 10 µl total volume with dH_2O. Reactions were incubated on ice overnight. Prior to transformation, ligation mixtures were dialysed using 0.025 µm dialysis discs (Millipore) over dH_2O for 20 minutes at room temperature.

2.4.7 Agarose Gel Electrophoresis

A 50X stock of Tris-acetate-EDTA (TAE) buffer (BioRad) was diluted to 1X and used to create 1 % concentration (w/v) agarose (Sigma). SYBR™ Safe (Thermo Fisher Scientific) DNA stain (1:10,000 final concentration) was added prior to pouring. DNA fragments for visualisation were mixed with 6X loading buffer (NEB) and loaded onto the gel with an appropriate DNA ladder, usually Quick-Load® Purple 2-Log DNA ladder (NEB, 0.1-10.0 kb). DNA fragments were separated by electrophoresis, in TAE buffer, at 80-100 V for 45-60 minutes. DNA fragments were visualised using Gel Doc™ XR+ Gel Documentation System (BioRad) and images captured using Image Lab™ Software (BioRad).

2.4.8 DNA Extraction from Agarose Gel

DNA fragments, of the correct size, were excised from agarose gels and purified using a Gen Elute™ Gel Extraction Kit (Sigma) and Monarch® DNA Gel Extraction Kit (NEB), per the manufacturer's instructions. Briefly, the DNA fragment of interest was excised from the gel using a scalpel, under light from a transilluminator, and dissolved in three volumes of Gel Solubilization Solution or four volumes of Monarch Gel Dissolving Buffer at 50 °C for 10 minutes, or until the gel slice was completely dissolved. The resulting solution was passed through a column and DNA was precipitated using ethanol. Centrifugation steps were conducted at 10,000 x *g* for one minute. DNA was eluted in 6-20 µl dH_20 and stored at -20 °C until required. Purified DNA was quantified using the Nanodrop™ Lite spectrophotometer.

2.4.9 Amplification of DNA using Polymerase Chain Reaction (PCR)

Oligonucleotide primers were manually designed (2.5.1) and synthesised by Sigma. Primers were reconstituted to 100 µM in dH_2O and stored at -20 °C until required. PCR was performed using Q5® High-Fidelity DNA polymerase (NEB), One*Taq*® DNA polymerase (NEB) or Dream*Taq* Green PCR Master Mix (2X, Thermo Scientific) in accordance with the manufacturer's guidelines with reactions set up as in Table 2.7. The Dream*Taq* Green PCR Master Mix is a complete PCR master mix containing *Taq* polymerase, dNTPs, buffer and gel loading dye therefore, only the addition of DNA and primers is required. Reactions were performed in a Biometra TRIO Thermocycler using protocols described by the manufacturers and shown in Table 2.8, Table 2.9 and Table 2.10. Annealing temperatures were altered for each individual primer set and were determined using NEB online Tm calculator, for Q5® and One*Taq*® reactions, or as 5 °C below the melting temperatures of the primers for Dream*Taq*. For each PCR, a mastermix for the number of reactions required was created based on the volumes in Table 2.7.

Table 2.7: PCR reaction components for Q5®, One*Taq*® and Dream*Taq* used in this study

Reaction Components	**Volume of Component Added to 1 Reaction (µl)**		
	Q5®	One*Taq*®	Dream*Taq*
Template DNA	1 (<1 µg)	1 (<1 µg)	1 (<1 µg)
Forward Primer (10 µM)	1.25	0.5	1
Reverse Primer (10 µM)	1.25	0.5	1
dNTPs (10 mM. NEB)	0.5	0.5	-

Dream*Taq* Green PCR Master Mix (2X)	-	-	10
One*Taq*® Standard Reaction Buffer (5X)	-	5	-
One*Taq*® DNA Polymerase	-	0.125	-
Q5® Reaction Buffer	5	-	-
Q5® High-Fidelity DNA Polymerase	0.25	-	-
Nuclease Free Water (Fisher Scientific)	Up to 25 µl	Up to 25 µl	7

Table 2.8: PCR protocol used with Q5® High-Fidelity DNA Polymerase in this study

Cycle Step	**Temperature (°C)**	**Duration**	**Number of Cycles**
Initial Denaturation	98	30 s	1
Denaturation	98	10 s	30
Annealing	50-72	30 s	
Extension	72	30 s/kb	
Final Extension	72	2 min	1
Hold	15	∞	

Table 2.9: PCR Protocol used with One*Taq*® DNA Polymerase in this study

Cycle Step	**Temperature (°C)**	**Duration**	**Number of Cycles**
Initial Denaturation	94	30 s	1
Denaturation	94	30 s	30
Annealing	45-68	30 s	
Extension	68	1 min/kb	
Final Extension	68	5 min	1
Hold	15	∞	

Table 2.10: PCR protocol used with Dream*Taq* Green PCR Master Mix in this study

Cycle Step	Temperature (°C)	Duration	Number of Cycles
Initial Denaturation	94	10 min	1
Denaturation	94	30 s	30
Annealing	Tm-5	30 s	
Extension	72	1 min/kb	
Final Extension	72	5 min	1
Hold	15	∞	

2.4.10 Classification of C. difficile *Isolates by PCR Ribotyping*

C. difficile isolates from environmental samples were classified by PCR ribotype. Genomic DNA from each isolate was extracted as described before (2.4.1), or one colony from each isolate was re-suspended in 20 µl dH_2O. Samples were heated for 20 minutes at 95 °C before being used in a PCR reaction. PCR ribotyping reaction buffer was created by mixing 300 µl dH_2O with 500 µl of One*Taq*® Standard Reaction Buffer (1X, NEB) and 200 µl dNTPs (10 mM, NEB). One PCR reaction was set up per isolate using 20 µl of gDNA or re-suspended colony, 10 µl of PCR ribotyping reaction buffer, 0.25 µl of primers ribotyping_p3 and ribotyping_p5 (100 µM), 17.75 µl dH_2O, 1.5 µl magnesium chloride (1.5 mM, NEB) and 0.25 µl One*Taq*® DNA polymerase (0.75 U final concentration, NEB). Reactions were performed in a Biometra TRIO Thermocycler using the protocol shown in Table 2.11. After PCR, samples were heated at 75 °C for 45 minutes with the tube lids open until around 20 µl of each reaction remained. The tubes were transferred to the QiAxcel Capillary Electrophoresis system (Qiagen) and DNA fragments were detected using the GCK-5000 cartridge and the Ribotyping OL400 method. Resulting PCR ribotyping banding patterns were visualised and aligned with the ScreenGel

software (Qiagen). PCR ribotyping was also performed by the *Clostridium difficile* Ribotyping Network (CDRN) service (Reference Laboratory, Leeds General Infirmary) and PCR ribotypes were allocated based on similar banding patterns to known PCR ribotypes. New PCR ribotypes were allocated when bands did not match with known patterns.

Table 2.11: PCR protocol used for PCR ribotyping of *C. difficile* isolates in this study

Cycle Step	**Temperature (°C)**	**Duration**	**Number of Cycles**
Initial Denaturation	95	2 min	1
Denaturation	92	1 min	30
Annealing	55	1 min	
Extension	72	1 min, 30 s	
Second Denaturation	95	1 min	1
Second Annealing	55	45 s	1
Final Extension	72	5 min	1
Hold	15	∞	

2.4.11 Confirmation of C. difficile *using Latex Agglutination Assay*

To confirm *C. difficile* colonies were isolated from environmental samples, a latex agglutination assay was performed using a *C. difficile* Latex Agglutination Test Kit (Oxoid) per the manufacturer's instructions. In this kit, latex particles are coated with IgG antibodies specific for cell wall antigens of *C. difficile.* Suspected *C. difficile* colonies were grown overnight on BHIsCC agar at 37 °C under anaerobic conditions, before three individual colonies from each isolate were selected and emulsified in a drop of saline. Next, a drop of *C. difficile* Latex Reagent was added to the colony/saline mixture and incubated at room temperature for a maximum of two minutes before observing for agglutination.

The presence of agglutination within two minutes confirms the presence of *C. difficile* colonies.

2.4.12 ClosTron Mutagenesis

2.4.12.1 Intron Design for pMTL007C-E2

The online Perutka algorithm (available at www.clostron.com) was used to select an appropriate intron insertion site within the gene of interest. Two sites (for one gene of interest) were selected, one with sense and one with anti-sense directionality, for intron insertion and both were selected due to their highest algorithmic score.

2.4.12.2 Synthesis of pMTL007C-E2 Plasmids

ClosTron plasmids, with the pMTL007C-E2 backbone, containing the intron targeting region for the gene of interest were synthesised and constructed by ATUM (formerly DNA2.0, USA).

2.4.12.3 Generation of ClosTron Mutants in C. difficile

Upon receipt, ClosTron plasmids were eluted from filter paper using the supplier's instructions to ensure maximum yield. Plasmids were immediately transformed into *E. coli* TOP10 for storage and plasmid propagating and *E. coli* CA434 for conjugation into *C. difficile* 630Δ*erm.* After conjugation into *C. difficile* and successful production of thiamphenicol resistant colonies, colonies were re-streaked to purity three times on selection plates. The overnight growth from one plate was harvested, using a sterile 10 µl loop, and plated onto

BHIsCC agar supplemented with erythromycin. Resulting erythromycin resistant colonies were screened for ClosTron insertion using PCR with primers flanking the expected insertion site in the gene of interest.

2.5 Bioinformatic Tools

2.5.1 Oligonucleotide Design

PCR primers were manually designed with sequence visualisation in Ugene Integrated Bioinformatic Tool (Uniprot). Designed primers were analysed for predicted secondary structures and primer dimer formation using OligoEvaluator™ provided for free online by Sigma. The melting temperatures of the primers were also determined using the same tool.

2.5.2 Data Analysis and Presentation

Numerical data was visualised using Microsoft Excel. GraphPad Prism and Microsoft Excel were used to create graphical data images and for statistical analysis.

2.5.3 Sequencing of PCR DNA Fragments and Plasmid DNA

Purified PCR fragments and plasmid DNA were sequenced using Source Bioscience Sanger Sequencing Service and sequences were analysed using Ugene Integrated Bioinformatics Tool. Alignment of nucleotide and amino acid sequences was conducted using The European Molecular Biology Open Software Suite (EMBOSS) pairwise alignment tool (Rice *et al.,* 2000).

2.5.4 Sequencing and Annotation of Phage and Bacterial Genomes

2.5.4.1 Preparation of DNA for Sequencing

Genomic DNA of *C. difficile* and phages was isolated using the phenol/chloroform procedures outlined in 2.4.1 and 2.4.2. DNA integrity was checked by agarose gel electrophoresis and quantified using the Nanodrop Lite spectrophotometer (Fisher Scientific).

2.5.4.2 Preparation of Sequencing Libraries

Sequencing libraries were prepared by DeepSeq (University of Nottingham) using the Illumina TrueSeq PCR Free Library Preparation Kit (Illumina). Genomic DNA (2 µg) was fragmented using Cavaris S2 sonicator to 400-600 base pairs (bp) and subsequently purified using Agencourt Ampure XP beads (Beckman Coulter) with a DNA to bead ratio of 1 to 1.6. Sonicated DNA fragments were end repaired for 30 minutes at 30 °C and selected for the desired fragment size using Ampure XP beads. Y adaptors for Illumina sequencing were ligated for 10 minutes at 30 °C and then incubated for five minutes at 70 °C. To remove un-ligated adaptors, two further rounds of purification using Ampure XP beads were performed with a DNA to bead ratio of 1 to 1. The purified adapter ligated DNA was quantified using qPCR (Kapa Biosystems).

2.5.4.3 Sequencing of Prepared DNA Libraries

The prepared DNA libraries were loaded onto the MiSeq sequencing platform in accordance with the manufacturer's instructions, by DeepSeq (University of

Nottingham). Sequencing was completed using Illumina MiSeq 500 bp V2 SBS chemistry.

2.5.4.4 Assembly, Analysis and Annotation of Bacterial and Phage Genomes

For phage genomes, raw sequencing reads were assembled into a single contig using the "*de novo* assembly" function within CLC Genomics Workbench 8.5.1 (Qiagen) using the default parameters. Artemis (Rutherford *et al.*, 2000) was used to identify putative open reading frames (ORFs) and manual annotation of the genome was completed using BLASTp (NCBI), UniProt and Pfam (The European Bioinformatics Institute) to assign putative protein functions. ORFs were manually trimmed to the correct start codon based on the presence of ribosome binding sites and promoter sequences in Artemis. Bacterial genomes were assembled by alignment of the raw sequencing reads to a reference genome from the same PCR ribotype using the "Map to Reference" function within CLC Genomics Workbench 8.5.1 under default parameters. Single nucleotide polymorphisms (SNPs) were analysed using the "Quality-based Variant Detection" tool in the same programme, using the fixed ploidy detection with a ploidy of 1. For the programme to call SNPs, the SNP must be present in the forward and reverse strand and appear in at least 35 % of the validated reads. Assembled genomes were annotated using the Rapid Annotation using Subsystem Technology (RAST). Bacterial genomes were viewed using Artemis or Ugene.

2.5.5 Identification of Prophage Regions within Bacterial Genomes

The presence of prophage regions in bacterial genomes was determined using the online tool PHASTER (Arndt *et al.*, 2016).

2.5.6 Phylogenomic Analysis

Phylogenomic trees were created using the Virus Classification and Tree Building Online Resource, VICTOR. Pairwise comparisons of the nucleotide sequences of 25 published *C. difficile* genomes, plus the four phages isolated in this study, were conducted using the Genome-BLAST Distance Phylogeny (GBDP) method (Meier-Kolthoff *et al*., 2013) under settings recommended for prokaryotic viruses (Meier-Kolthoff and Goeker, 2017). Three different algorithms were used for the phylogenomic analysis D0, D4 and D6. The formulas D0 and D6 compare complete genome sequences at the nucleotide level (D0) and at the amino acid level (D6), whereas the formula D4 is best utilised when incomplete genomes are being compared. For each of the formulas, D0, D4 and D6, intergenomic distances (100 replicates each) were used to deduce a minimum evolution tree with branch support via FastME, including subtree pruning and regrafting (SPR) postprocessing which removes a subtree from the initial tree and then reinserts the subtree by dividing any remaining branches in the initial tree (Lefort *et al*., 2015). The trees were visualised using FigTree (Rambaut, 2006) and rooted at the midpoint (Farris, 1972). The OPTSIL programme (Göker *et al*., 2009) was used to estimate taxon boundaries at the family, genus and species level, using the recommended clustering thresholds (Meier-Kolthoff and Goeker, 2017) and an F value (fraction of links required for cluster fusion) of 0.5 was used, which allowed average linkage clustering and

was selected as this value has been shown to produce optimal clustering results (Meier-Kolthoff *et al*., 2014).

2.5.7 Protein Modelling

The 3D structure of a candidate receptor protein was predicted using the SWISS-MODEL online tool, available at https://swissmodel.expasy.org, using structure homology modelling which relies on evolutionary structures to produce structural models of the protein of interest. The SWISS-MODEL searches two databases, BLAST and HHblits, to identify protein templates to obtain alignments. The use of two databases ensures good alignments at both high and low sequence identities. The position of amino acid substitutions in comparison to the sequence from other *C. difficile* strains was also predicted using this tool (Biasini *et al*., 2014).

2.6 Characterisation of Bacteriophages

2.6.1 Transmission Electron Microscopy

Prior to observation using transmission electron microscopy (TEM), the isolated phage lysates, at titre $>10^9$ pfu/ml, were precipitated by ammonium acetate as described by Fortier and Moineau (2007). Briefly, 1.5 ml phage lysate was harvested by centrifugation at 21,000 x *g* for 75 minutes. Immediately following centrifugation, 1.4 ml of the supernatant was very carefully removed, ensuring the bottom of the tube was not touched, and discarded. A 1 ml aliquot of ammonium acetate (0.1 M, Sigma) was added to the remaining 100 µl of phage lysate. Centrifugation and the wash step were repeated twice and the precipitated

phage particles were stored at 4 °C. A 5 µl volume of the precipitated phage was added to a 200 mesh Formvar carbon coated copper grid (Agar Scientific) and left for 30 seconds. The excess liquid was wicked off using Whatman paper. Next 10 µl of 2 % uranyl acetate (Sigma) was added to the grid, to negatively stain the phages, for 30 seconds. The excess liquid was removed by blotting with Whatman paper and dried for a few minutes before the phage particles were visualised using TEM.

2.6.2 Determining the Rate of Phage Attachment to Bacterial Cells

The rate of phage attachment was determined using the method of Clokie and Kropinski (2009). A 1 % inoculum of *C. difficile* propagating strain overnight culture was grown for four hours at 37 °C under anaerobic conditions until the OD at 600 nm reached 0.6-0.8 (~10^8 cfu/ml). A 9 ml volume of this culture was infected to a multiplicity of infection (MOI) of 0.0001 using 1 ml of phage stock (~10^5 pfu/ml) before samples (50 µl) were taken at five-minute intervals over 60 minutes, post infection. The 50 µl sample was diluted in 950 µl BHIs broth and stored on ice until it was required. Samples were centrifugated at 10,000 x *g* for 10 minutes and filter sterilised using a 0.22 µm membrane filter to ensure the removal of phages bound to cells. The number of free phages at each time point was enumerated using plaque assays. The rate of attachment constant (k) was determined in accordance with the formula given by Clokie and Kropinski (2009), where $k = (2.3/Bt) \log (Po/P)$, where B is the number of bacterial cells and *t* is the time interval for the phage titre to increase from Po (the original phage titre) to P (the final phage titre).

2.6.3 One Step Growth Curve for Determining Phage Burst Size

A one step growth curve of the phages was used to determine the burst size using a protocol adapted from Sekulovic *et al.* (2011). The *C. difficile* propagating strains were grown for four hours at 37 °C under anaerobic conditions, until the OD at 600 nm was 0.6-0.8 (~10^8 cfu/ml), from a 1 % inoculum of a 20 ml BHIs overnight culture. An aliquot of 2 ml of this culture was infected with the corresponding phage stock (~10^8 pfu/ml) to obtain an MOI of 1. The phage-bacterial mix was incubated for 15 minutes at 37 °C under anaerobic conditions to allow phage adsorption. Cultures were immediately centrifuged at 1,377 x *g* and washed three times using 1 ml BHIs broth to remove any non-adsorbed phages. The number of unbound phages (washed mixtures) was enumerated using plaque assays. The washed phage-bacterial mixture was incubated at 37 °C under anaerobic conditions for 120 minutes, with samples (50 µl) taken at intervals over this period. The 50 µl sample was diluted in 950 µl BHIs broth and stored on ice until it was required. Samples were centrifugated at 10,000 x g for 10 minutes and filter sterilised using a 0.22 µm membrane filter to ensure the removal of phages bound to cells. The number of free phages at each time point was enumerated using plaque assays. The burst size was calculated in accordance with Sekulovic *et al.* (2011), where burst size is equal to (Final Phage Titre – Initial Phage Titre)/Initial Phage Titre, where the initial titre is equal to the Infection Titre – The Number of Unbound Phages Post Adsorption and recorded as pfu per infected cell.

2.6.4 Determining the Host Range of Phages

The standard plaque assay described in 2.3.4.6, was used to determine the host range of the isolated phages testing against other strains from the same PCR ribotype as the phage propagating strain, using ~10^9 pfu/ml. More wide-scale host range testing was conducted using the "spot on the lawn" technique described in section 2.3.4.8, with the inclusion of all phage dilutions for each phage stock. Serial dilutions were made in BHIs broth down to 10^{-8}, usually equating to ~1 pfu/ml. A phage cocktail, containing four phages at titre ~10^8 pfu/ml, was used for host range testing when the "spot on the lawn" technique was used. After overnight incubation at 37 °C under anaerobic conditions, the presence of plaques within the phage spot could be observed and counted. Efficiency of plating (EOP) was determined for each indicator strain by comparison of the phage titre using the propagating strain against the phage titre using the indicator strain.

2.6.5 Determination of Phage to Bacterial Cell Binding

A 1 % inoculum of BHIs overnight culture of *C. difficile* strains was transferred to 20 ml pre-reduced BHIs broth and incubated for four hours at 37 °C under anaerobic conditions. Strains requiring induction with anhydrotetracycline were induced for one hour (at 37 °C under anaerobic conditions) prior to the binding assay. The 20 ml cultures were harvested by centrifugation at 10,000 x *g* for 10 minutes at room temperature and the resulting cell pellet re-suspended in 500 µl pre-reduced BHIs broth. An aliquot of 100 µl of harvested cells was removed and centrifuged again at 3,637 x *g* for two minutes before the supernatant was removed and discarded and 10 µl of phage (10^4 pfu/ml), corresponding to the

correct propagating strain, was added. This was incubated for 15 minutes under anaerobic conditions to allow the phage to bind to the bacterial cells, before being re-suspended in 1 ml BHIs broth. A final centrifugation step was completed at 14,549 x *g* for two minutes to remove the bacterial cells. The titre of phage in the supernatant that had not bound to cells was enumerated using plaque assay. A reduction in phage titre from the infection phage titre is indicative of phage binding.

Chapter Three:

Isolation of *Clostridium difficile* and Bacteriophages from Environmental Samples

3 Isolation of *Clostridium difficile* and Bacteriophages from Environmental Samples

3.1 Introduction

The emergence of antimicrobial resistance, and in the case of CDI additional concerns over relapse, has led to renewed interest in the field of phage therapy. For phage therapy to be successful, libraries of phages able to infect pathogens of interest are required. Consequently, the isolation of novel phages is compulsory in advancing the field.

Initial attempts to isolate free phages of *C. difficile* from environmental samples were unsuccessful (Goh *et al*., 2005; Nale *et al.,* 2012), likely because the selection of the indicator strain for their detection was inappropriate. This is likely explained by the fact that those phages isolated to date have proven to possess a very restricted host range, and as a consequence are able to infect a very narrow spectrum of *C. difficile* strains. Nonetheless, recent attempts to isolate free phages from patient faecal samples (Meessen-Pinard *et al.,* 2012) and soil (Rashid *et al*., 2016) have been more successful, proving that the isolation of free phages of *C. difficile* from the environment is possible. Currently, however, all previously isolated phages of *C. difficile* are temperate and, as a consequence of their narrow host range, require the use of combinations of different phage to successfully clear bacterial cultures, without regrowth of lysogenic bacteria (Nale *et al*., 2016). The use of phage combinations means that a larger and more comprehensive library of phages will be required if phage therapy is to be successful in the treatment of CDI. Consequently, there is a need

to isolate a broad range of novel phages which will be able, in combination, to infect and lyse all major clinical strains of *C. difficile.*

The isolation of *C. difficile* from soil and sewage samples (Kim *et al.*, 1981) and sediment and water samples (Hargreaves *et al*., 2013) has previously proven successful and their use as appropriate host indicator strains has allowed the isolation of phages from environmental samples and the isolation of novel phages through induction of prophage in lysogenic *C. difficile* strains (Hargreaves *et al*., 2013). Thus, by using *C. difficile* strains as the indicator host that are isolated from the same environmental sample being "mined" for the presence of phage, the chances that the indicator strain is susceptible to infection by phage particles is increased.

Ribotyping is a typing method for *C. difficile* based on the presence of multiple copies of the rRNA operon in the *C. difficile* chromosome. PCR is used to amplify the intergenic spacer region located between the 16S and the 23S rRNA genes, using primers specific for each gene but close to the intergenic region. As the operon is present multiple times in the chromosome, amplification produces DNA fragments of varying sizes which create a characteristic banding pattern when separated using gel electrophoresis (Bidet *et al*., 1999). Each known PCR ribotype therefore by definition has a unique specific banding pattern. Identification of the PCR ribotype of a new isolate may, therefore, be determined by comparison of banding patterns with strains of known PCR ribotype. PCR ribotyping is used to study the changing epidemiology of *C. difficile* strains and the *Clostridium difficile* Ribotyping Network (CDRN) collate data to determine

the most prevalent PCR ribotypes in England. Between 2008 and 2015, the CDRN reported a decrease in the prevalence of PCR ribotypes 001, 027 and 106 in England with emergent PCR ribotypes 002, 005, 014/020, 015, 023 and 078 increasing in prevalence over the same period (Public Health England, 2016).

3.2 Results

3.2.1 Optimising Conditions for Phage Isolation

Initial attempts (two attempts, January and March 2014) to isolate phage from sewage samples were unsuccessful which necessitated the optimisation of the isolation protocol. Prior to optimisation, *C. difficile* overnight cultures were added to the sewage sample to enrich the culture and allowed to grow overnight. The same *C. difficile* strains were subsequently used as the indicator host strains but no phages were isolated. The method was then altered to include the addition of the known *C. difficile* germinant, taurocholate (see section 2.3.4.5), to act as a sample enrichment agent, instead of spiking the sample with growing *C. difficile* cultures. This method proved successful in isolating phages and was therefore adopted for all subsequent phage isolation experiments (see section 2.3.4.5 for complete method). Sample collection from post-anaerobic digester material was chosen due to previous success in the group isolating phages from samples taken from this location (T. Bailey and N. Minton, University of Nottingham, personal communication), additionally a sample taken from pre-anaerobic digester material was proven unsuccessful in the isolation of phages.

3.2.2 Optimising Plaque Assays for Use with C. difficile

To be able to detect the presence of phages, confluent growth of *C. difficile* was required within a top agar layer. Plaque assays were conducted using the standard double layer agar method described in 2.3.4.6, however, the conditions for *C. difficile* growth first had to be optimised to ensure plaque formation was visible. Preliminary experiments were conducted to determine the most appropriate method for obtaining confluent growth of *C. difficile* within the top agar overlay. *C. difficile* 630Δ*erm* and *C. difficile* 522418 strains were selected for this because the first is a common laboratory strain used by most researchers in our laboratory and the second was a strain that was going to be used in the isolation of phages from environmental samples. Cells used to inoculate the molten top agar were taken from overnight cultures of *C. difficile 630Δerm* and 522418 together with sub-cultures taken at 4, 6 and 8 hours. It was determined that confluent growth was not produced using an overnight culture or 4-hour culture. However, when both *C. difficile* strains were grown for 6 or 8 hours, from a 1 % inoculum sub-culture, confluent growth was reliably obtained. At the same time, the volume of *C. difficile* culture required for the production of confluent growth was tested, using 100 µl, 500 µl and 1 ml of *C. difficile* culture from an overnight, 4, 6 or 8-hour sub-culture. It was determined that confluent growth was not produced when 100 µl or 500 µl of culture from any time point was used however, when 1 ml of culture from a 6 or 8-hour sub-culture was used, confluent growth was reliably produced. This method (as described in section 2.3.4.6) for confluent growth production, using 1 ml of 6-hour *C. difficile* subcultures, was adopted for all subsequent plaque assays using *C. difficile.* Prior to using a novel *C. difficile* strain for the first time in a phage plaque assay, the

ability for that strain to produce confluent growth after a 6-hour sub-culture was determined. This ensured that strains unable to reliably produce confluent growth were not selected for isolation of phages or host range analysis. This was completed using the standard double layer agar method described (2.3.4.6) but without the addition of phages.

3.2.3 Isolation of Phages From Environmental Samples

Post-anaerobic digester samples (n=2) were collected from Stoke Bardolph Sewage Treatment Plant in Nottinghamshire, U.K., on two separate occasions (October 2014 and August 2015). A total of ten potential host strains (n=5 environmental isolates, n=5 clinical isolates of which n=1 PCR ribotype 027 and n=4 PCR ribotype 023) were used in phage isolation assays with the first sample. A single phage was isolated using *C. difficile* 522418, a PCR ribotype 023 clinical isolate. Twenty *C. difficile* clinical isolates (n=2 PCR ribotype 001, n=2 PCR ribotype 002, n=2 PCR ribotype 014, n=2 PCR ribotype 015, n=10 PCR ribotype 027 and n=2 PCR ribotype 078) were used as indicator strains in phage isolation assays with the second sample. These PCR ribotypes were selected as they were the most commonly occurring PCR ribotypes, aside from PCR ribotype 023, in the research group collection of clinical isolates. From this sample, three phages were isolated using *C. difficile* 08011 (PCR ribotype 002), *C. difficile* 2301 (PCR ribotype 014) and *C. difficile* 1801 (PCR ribotype 078), all of which are clinical isolates. Phages were named after their *C. difficile* isolation strain: ΦCD418, ΦCD08011, ΦCD2301 and ΦCD1801. *C. difficile* strains 522418, 08011, 2301 and 1801 will be referred to as the phage propagating strains from here on in.

In each case plaques were small (~1 mm diameter) and hazy (Figure 3.1). Phages were purified three times by single plaque picking and high titre stocks were created by harvesting the top layer from plates showing confluent lysis. In each case a titre of at least 10^8 pfu/ml was obtained. The effect of maintaining sterilised phage stocks in BHIs broth at 4 °C was investigated and shown to provide a means for reasonable long-term storage. Thus, the titre of phage ΦCD08011 following storage at 4 °C for six months was reduced by only two orders of magnitude, to ~10^6 pfu/ml, whereas the titre of the other three phages was reduced to ~10^7 pfu/ml.

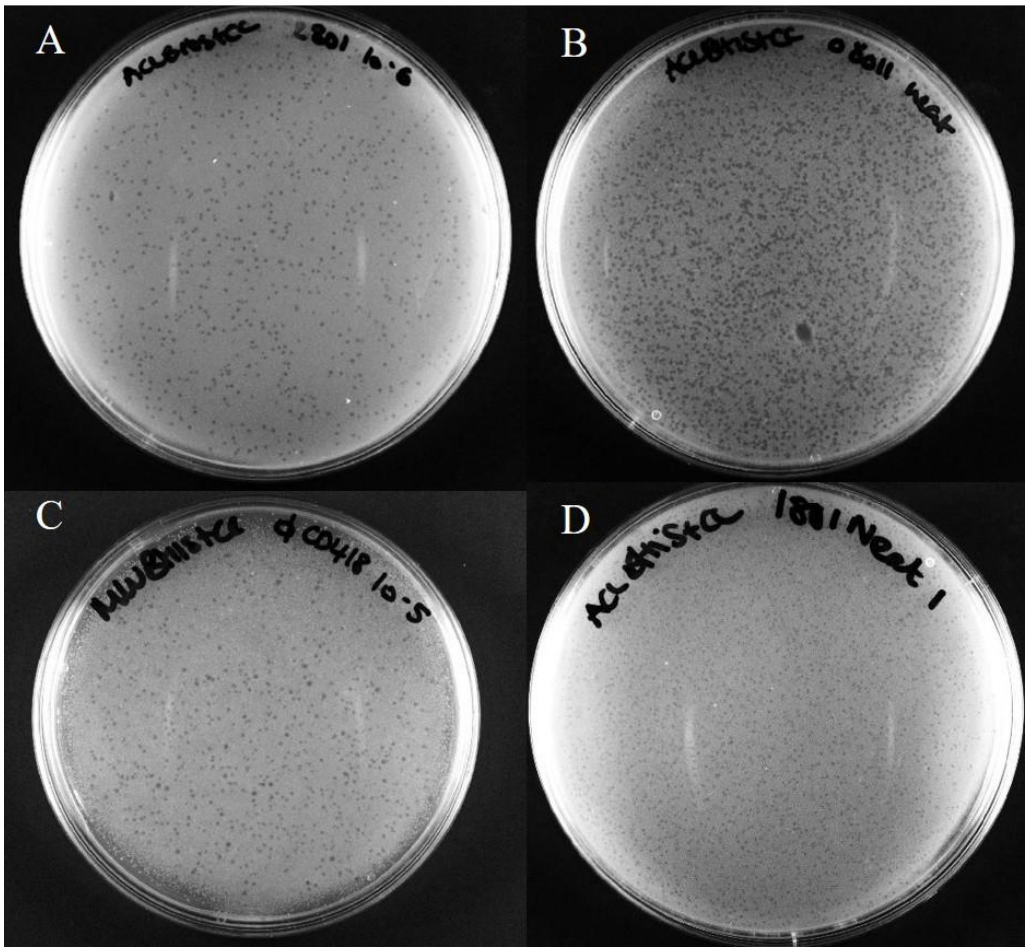


Figure 3.1: **Plaque formation of the four isolation phages**; A: ΦCD2301, showing sterilised phage stock diluted to ~10^4 pfu/ml, B: ΦCD08011, showing sterilised phage stock ~10^8 pfu/ml, C: ΦCD418, showing sterilised phage stock diluted to ~10^3 pfu/ml, D: ΦCD1801, showing sterilised phage stock ~10^8 pfu/ml. For each phage plaques appear hazy, suggesting the isolation of temperate phages, and small ~1 mm diameter. Phages were purified and stored as sterilised high titre phage stocks.

3.2.4 Random Phage Fragment Cloning

Genomic DNA of the four isolated phages was extracted using RBC Bioscience Viral Nucleic Acid Extraction kit (YVN100) from stocks with titre ~10^8 pfu/ml. Phage DNA was fragmented by digestion with *Hin*dIII restriction enzyme for two hours at 37 °C, resulting in the production of many small DNA fragments, however the DNA was too fragmented for discrete visualisation of bands using

gel electrophoresis. The plasmid, pMTL82151, was linearised using the same restriction enzyme, purified and ligated with the randomly digested phage DNA. Blue/white screening was used to determine successful insertion of phage DNA fragments into the plasmid backbone. Five white colonies, containing recombinant plasmids, were selected for each phage and plasmid DNA was isolated. Two of the random phage DNA inserts were sequenced for each phage using Sanger Sequencing and the pBP1_R1 and ColE1_traJF2 sequencing primers. Nucleotide BLAST (BLASTn) searches of the sequences obtained confirmed that phage had been isolated in all cases and that each phage was likely to be different to one another (Table 3.1).

Table 3.1: Nucleotide BLAST results of random isolated phage fragments

Phage	**Size of Insert (bp)**	**BLASTn Results**	**Gene Product**
ΦCD418	743	93 % identity to fragment of gp13 gene from *C. difficile* phages ΦCD27 and ΦCD505	Hypothetical protein
	155	99 % identity to fragment of *pgm1* gene from *C. difficile* 630, CD196 and R20291	Phosphoglucomutase
ΦCD08011	27	100 % identity to fragment of D864_gp14 gene *C. difficile* phage ΦMMP04	Tail tape measure protein
	258	99 % identity to fragment of D864_gp20 gene of *C. difficile* phage ΦMMP04	Base plate J-like assembly protein
ΦCD2301	86	100 % identity to fragment of phiMMP01_20003 and phiMMP04_20003 genes of *C. difficile* phages ΦMMP01 and ΦCDMMP03 respectively	Phage portal protein, SPP1 family Putative phage portal protein
	85	100 % identity to fragment of phiCDHM11_gp9, phiCDHM13_gp9 and phiCDHM14_gp9 genes of *C. difficile* phages ΦCDHM11, ΦCDHM13 and ΦCDHM14 respectively	Putative tail sheath protein
ΦCD1801	109	100 % identity to fragment of phiCDHM19_gp8 gene of *C. difficile* phage ΦCDHM19	Putative scaffold protein
	109	100 % identity to fragment CDBPCV119_gp08 gene of *C. difficile* phage ΦCD119	Putative scaffold protein

Isolated phage genomic DNA was digested using *Hin*dIII and cloned into the modular plasmid, pMTL82151. Randomly cloned fragments were sequenced and nucleotide BLAST search was used to ensure phages had been isolated.

3.2.5 Isolation and Confirmation of C. difficile *Lysogens*

For long-term storage, lysogenic versions of the four *C. difficile* propagating strains were created, containing each phage. High titre phage stocks (>10^8 pfu/ml) were spotted three times onto the surface of solidified top agar containing the propagating strain for each phage. After incubation overnight, zones of clearing were visible in the confluent growth of the propagating strains. For each phage, resistant colonies were picked from within the zone of clearing and re-streaked to purity three times to remove any phage contamination. Five of these potential lysogens of each strain were selected for further investigation. An overnight culture of each potential lysogen was induced using mitomycin C at a final concentration of 3 µg/ml, and following overnight growth, the supernatants were collected and shown to harbour phage particles through plaque assays on the propagating strains. Additionally, each lysogen was shown to be immune to infection by their respective phage, as no plaques were observed in plaque assays. A representative lysogen of each phage was stored at -80 °C, as a glycerol stock, and shown to remain able to produce infectious phage particles on induction after two years of storage. Induction of the different lysogenic strains with mitomycin C produced phage at various titres, ranging from ~10^2 to ~10^{10} pfu/ml. Phage ΦCD418 induced to a much higher titre than the other three isolated phages, ~10^{10} pfu/ml (Table 3.2). Lysogens were named after the *C. difficile* propagating strain followed by L for lysogen; *C. difficile* 522418L, *C. difficile* 08011L, *C. difficile* 2301L and *C. difficile* 1801L. Only the site of phage ΦCD1801 integration into *C. difficile* 1801 genome was analysed by genome sequencing (see 5.3.5).

Table 3.2: Isolation and confirmation of *C. difficile* lysogens

***C. difficile* Lysogenic Strain**	**Phage Induction Titre (pfu/ml)**	**Lysogenic Immunity to Phage Infection**	
		Phage	Immune/Sensitive
CD522418L	7.75×10^{10}	ΦCD418	Immune
CD08011L	6.00×10^{2}	ΦCD08011	Immune
CD2301L	8.45×10^{2}	ΦCD2301	Immune
CD1801L	1.40×10^{6}	ΦCD1801	Immune

Lysogens were induced to produce phage particles, using a 3 µg/ml final concentration of mitomycin C, overnight. Induction titres were enumerated using plaque assays with infection of the corresponding phage isolation strain. Lysogens were screened for immunity to future phage infection using plaque assays and the respective phage.

3.2.6 Isolation of C. difficile *from Environmental Samples*

C. difficile was isolated from the 'post-anaerobic digester two' sample collected in October 2014 from which, phage ΦCD418 was isolated. Sample was plated onto Brazier's agar, then eight potential *C. difficile* colonies were selected for re-streaking onto *C. difficile* specific media (BHIsCC), all of which grew successfully. Suspected *C. difficile* colonies were selected based on colony morphology, slightly yellow, round or spread edged colonies, and ability to grow overnight on *C. difficile* selective media. Colonies were re-streaked to purity three times and each isolate used as a potential host strain for phage isolation, however this method did not yield any phages, i.e. these isolates were not appropriate phage host strains. The potential of these isolates to harbour prophages was also investigated, but all attempts to isolate phages through induction with mitomycin C, 3 µg/ml, were unsuccessful. A panel of five *C. difficile* strains (n=5 clinical isolates of which n=1 PCR ribotype 027 and n=4

PCR ribotype 023) were used as potential host strains for the observation of plaquing with the induced lysates of the environmental isolates.

The environmental isolates of *C. difficile* were named after the environmental sample, SI for sewage isolate, and numbered in order of colonies isolated, for example *C. difficile* SI001-SI008.

3.2.7 Confirmation of C. difficile *using PCR and Latex Agglutination Assay*

The identity of the isolated strains as *C. difficile* was confirmed using One*Taq* PCR to amplify the *spo0A* gene which is conserved across *C. difficile* strains, using D*spo0A*-FA1 and D*spo0A*-RS1 primers. A single colony from each of the eight isolates was re-suspended in 10 µl dH_20 and 1 µl of this was used as the template for PCR, with an annealing temperature of 64 °C. For seven of the eight isolates, *spo0A* was successfully amplified (Figure 3.2) and confirmed by Sanger Sequencing, using the same primers as for amplification. It was determined that isolate SI006 was not *C. difficile* due to the absence of *spo0A* amplification. However, this is not shown in Figure 3.2.

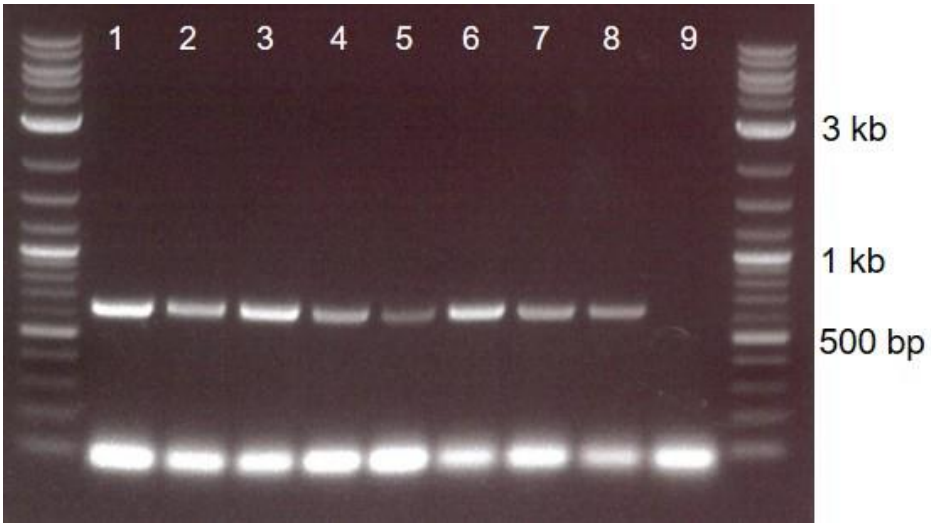


Figure 3.2: **Amplification of *spo0A* from environmental isolates to confirm the presence of *C. difficile*, producing a 686 bp product.** Lane 1: *C. difficile* SI001, Lane 2: *C. difficile* SI002, Lane 3: *C. difficile* SI003, Lane 4: *C. difficile* SI004, Lane 5: *C. difficile* SI005, Lane 6: *C. difficile* SI007, Lane 7: *C. difficile* SI008, Lane 8: *C. difficile* 522418, positive control for *spo0A* gene, Lane 9: No template negative control. Amplification of *spo0A* in *C. difficile* SI006 was not observed (not shown).

The latex agglutination assay was used, in triplicate, to further confirm the identity of each isolate as *C. difficile.* In this assay, the presence of agglutination confirms the presence of *C. difficile* and was observed for each of the seven isolates.

3.2.8 PCR ribotyping of C. difficile *Environmental Isolates*

The PCR ribotyping profile of each *C. difficile* environmental isolate showed that *C. difficile* SI001, SI003, SI005 and SI008 had unique banding patterns. *C. difficile* SI002, SI004 and SI007 showed identical banding patterns (Figure 3.3) and therefore only SI002 was selected for further analysis. The PCR ribotype number for the five isolates was assigned by the *C. difficile* Ribotyping

Network (CDRN), Leeds. This confirmed that the five isolates belonged to five different PCR ribotypes. Interestingly, one isolate, *C. difficile* SI003, belonged to a novel PCR ribotype that has not previously been mentioned in the literature, and was assigned the new PCR ribotype number of 711, and one isolate belonged to the hypervirulent PCR ribotype 078. The other three isolates SI002, SI005 and SI008 belonged to PCR ribotypes 039, 011 and 009, respectively.

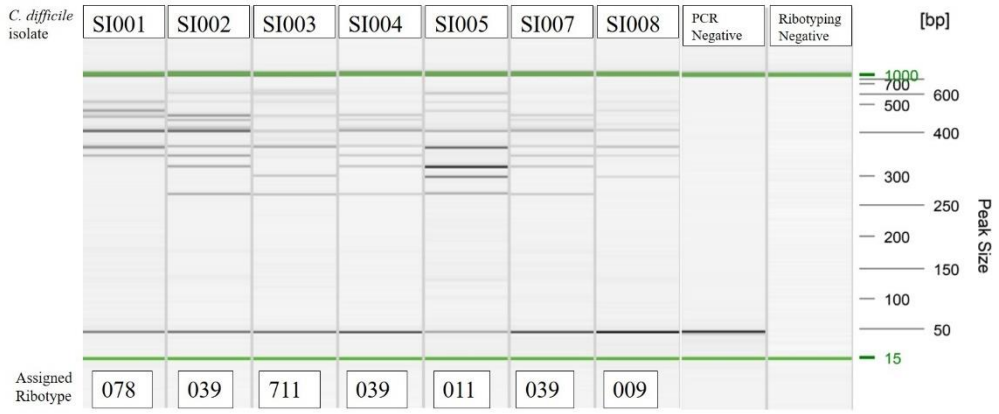


Figure 3.3: **PCR ribotyping profile of *C. difficile* environmental isolates and designated PCR ribotype number as determined by CDRN**. The PCR ribotyping profile of *C. difficile* SI001, SI003, SI005 and SI008 was unique in comparison to the other isolates, whereas the profile of *C. difficile* SI002, SI004 and SI007 appeared the same for each isolate. A variety of PCR ribotypes were isolated including the new PCR ribotype 711 and the hypervirulent PCR ribotype 078.

3.2.9 Growth Curve Profile of Phage Isolation Strains

The four phage propagating strains were characterised in terms of growth profiles. Growth rates were observed over a 24-hour period through the hourly measurement of the culture using the Promega Spectrophotometer. Viable cell counts were also estimated at 4, 6, 8 and 24 hours by plating 20 µl of appropriately diluted culture on BHIsCC solidified media and enumerating the colonies formed after 24-hour anaerobic incubation at 37 °C. Each of the strains

share a common growth profile, exhibiting a short lag phase and reach stationary phase 5-6 hours after inoculation (Figure 3.4). The maximum growth rate (μ) and doubling time (g) were calculated for each strain and are shown in Table 3.3. The number of viable cells (cfu/ml) present at six, eight and 24 hours for each strain was very similar which is what would be expected as the calculated maximum growth rates for the four strains are very similar (Figure 3.5). Between four and six hours, cfu/ml appears to increase around two logs for *C. difficile* 1801, 08011 and 2301 which is likely to be an artefact of the method used, such as counting colonies, as the data was collected in biological triplicate but not in technical triplicate. For *C. difficile* 522418 a higher cfu/ml is reached by four hours and subsequently only less than one log increase in cfu/ml is seen by six hours (Figure 3.5). This is consistent with a calculated doubling time of 74 minutes for this strain.

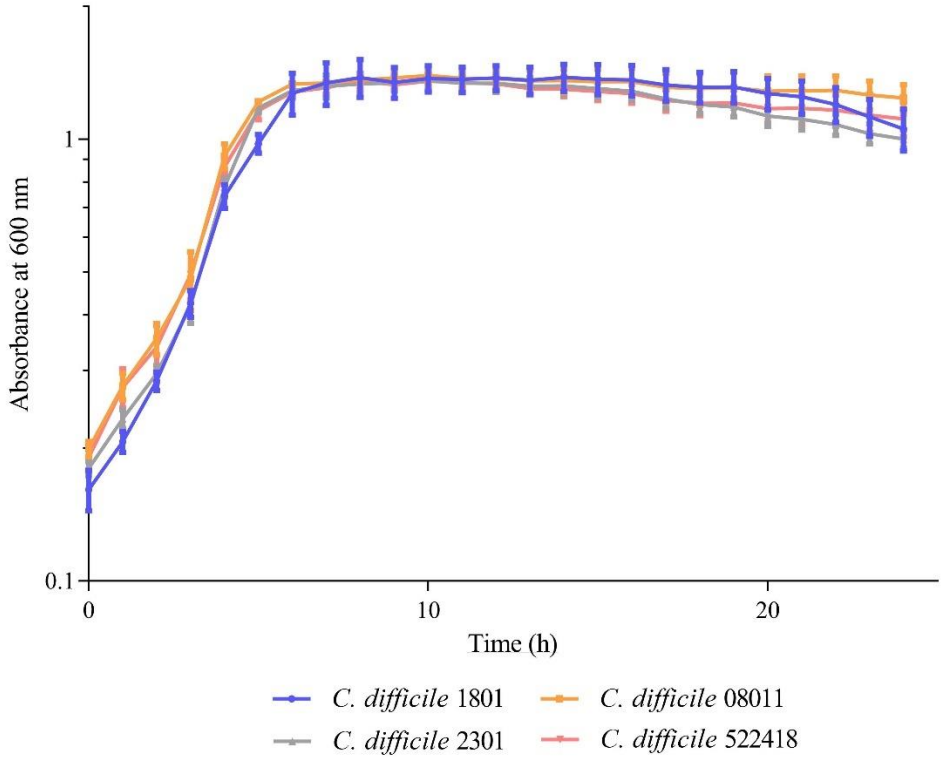


Figure 3.4: **Growth profile of the four *C. difficile* propagating strains**; *C. difficile* 522418, *C. difficile* 08011, *C. difficile* 2301 and *C. difficile* 1801. Although each strain belongs to a different PCR ribotype (023, 002, 014 and 078 respectively), the growth curve for each strain is very similar. Each strain reaches stationary phase after 5-6 hours and maintains stationary phase until the end of sampling. Optical density at 600 nm was measured, in biological triplicate, every hour for 24 hours using the Promega Spectrophotometer. Each point is the mean of biological triplicate data with standard error bars shown.

Table 3.3: Calculated maximum growth rates and doubling times of *C. difficile* propagating strains

***C. difficile* Strain**	**Maximum Growth Rate (μ, h^{-1})**	**Doubling Time (g, min^{-1})**
CD1801	0.55	74
CD08011	0.64	65
CD2301	0.62	67
CD522418	0.56	74

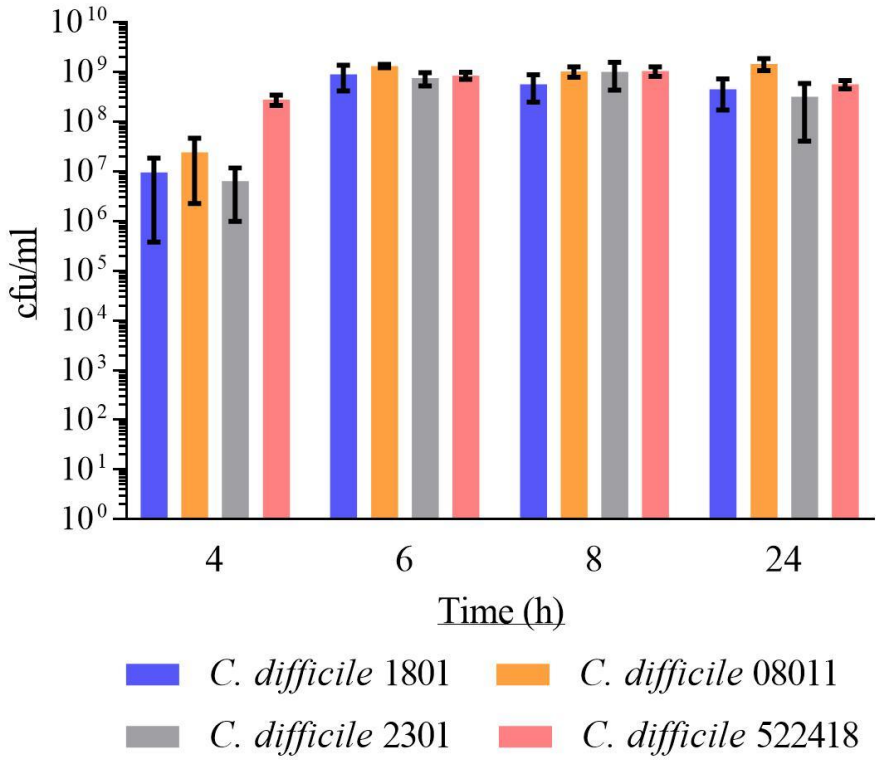


Figure 3.5: **Viable cell counts of *C. difficile* phage propagating strains**; *C. difficile* 522418, *C. difficile* 08011, *C. difficile* 2301 and *C. difficile* 08011. The number of cells present varies between the strains at 4 hours whereas at 6, 8 and 24 hours it is more comparable. Viable cells were measured as cfu/ml at 4, 6, 8 and 24 hours after 1 % inoculation from overnight cultures. Samples were measured in biological triplicate and the mean plotted with standard error bars shown.

3.3 Discussion

3.3.1 Isolation of C. difficile *Phages from Environmental Samples*

Previously published attempts to isolate phages infecting *C. difficile* from sewage, sludge or slurry samples have been unsuccessful (Goh *et al.*, 2005; Horgan *et al.*, 2010). Initial experiments suggested that the isolation of phages from the environment could be problematic with the selection of appropriate host

strains as the limiting factor. However, once hosts from a collection of local *C. difficile* patient faecal isolates were selected, phages were successfully recovered from every environmental sample tested in this study, likely due to the presence of these *C. difficile* isolates in the Nottinghamshire environment.

It is known that *C. difficile* can utilise carbon dioxide and hydrogen as its sole carbon source (Köpke *et al*., 2013) and it is likely that the process of anaerobic digestion promotes *C. difficile* outgrowth and hence the propagation of phages infecting *C. difficile*, such that post-anaerobic digester samples are successful in isolating phages. From this study, the addition of taurocholate into the sample enrichment proved to be vital in the successful isolation of phages. Addition of taurocholate has previously been shown to improve the yields of *C. difficile* vegetative cells from spores (Bliss *et al*., 1997) by bringing about the conversion of dormant spores into actively growing vegetative cells. In the present study, it was hypothesised that the addition of taurocholate would promote the germination of *C. difficile* spores present in the sample and their conversion into phage sensitive cells. Any free phages present in the sample would be able to infect the vegetative cells and propagate, enhancing the probability of isolating phages from the sample.

Plaques were hazy in appearance due to the presence of lysogenically immune colonies within the plaque. This is characteristic of temperate phage (Clokie and Kropinski, 2009). The plaques were very small, only measuring 1 mm in diameter, which is typical of previously reported *C. difficile* phages (Goh *et al*., 2005). The isolated phages appeared stable when stored at 4 °C in BHIs broth,

with only a one or two log pfu/ml reduction in titre over a 6-month period, contrary to reports of rapid decreases in *C. difficile* phage titres (Clokie *et al*., 2011).

The cloning of randomly digested fragments of the isolated phage DNA, and their subsequent nucleotide sequencing, was used to prove that phage particles had been isolated and that individual isolates were unlikely to be the same. A drawback of this method was that small inserts were preferentially cloned. Nonetheless, database BLAST searches of the nucleotide sequences obtained showed similarities to different previously characterised phages, suggesting that each of the phages isolated in this study were distinct from one another.

3.3.2 Isolation of C. difficile *Lysogenic Strains*

The long-term storage of temperate phages is most conveniently achieved through the isolation of a stable lysogen which can then be stored as either a glycerol stock at -80 °C or as a spore stock. In the past, stable lysogens of some *C. difficile* phages could not be isolated (Goh *et al*., 2005), however this did not prove to be the case in this study, and lysogens of each phage were obtained of appropriately sensitive strains. These lysogens were shown to release infectious phage particles following induction with mitomycin C and shown to be immune to further phage infection. Interestingly, each phage was induced to a different titre when the lysogenic strains were exposed to mitomycin C. The titre of free phages in the supernatant of induced CD522418L cultures showed that ΦCD418 could be induced to a titre of ~10^{10} pfu/ml. This is similar, although slightly higher, than the high induction titre obtained with *C. difficile* phage ΦMMP04,

which was reported to be ~10^9 pfu/ml (Meessen-Pinard *et al*., 2012). The isolation of stable lysogens, coupled with the hazy appearance of plaques, provides further evidence that the isolated phages are temperate. However, whilst apparently stable lysogens of each phage were generated, the site of integration of the various prophages into their respective host genomes was only determined for one of the four phages, ΦCD1801, due to resource constraints. Therefore, only one *C. difficile* propagating strain and lysogen pair was sequenced. Previously, the site of ΦCD27 integration was determined using an inverse PCR method (Williams *et al.*, 2013), showing that ΦCD27 could integrate at two *attB* sites in the *C. difficile* chromosome. Phage integration was observed within one ORF encoding a putative ATPase of an ABC transporter and a second ORF encoding a putative ATPase of the flagellin protein export apparatus (Williams *et al*., 2013). The location of phage ΦCD1801 integration is discussed in 5.3.5.

3.3.3 Isolation of C. difficile *from Environmental Samples*

C. difficile was successfully isolated from the same post-anaerobic digester sample as the phages. It was thought that these *C. difficile* isolates could represent potential sensitive indicator strains useful in phage detection or could themselves harbour prophage, however, no apparent release of phage by these strains was evident following treatment with mitomycin C when a panel of five clinically relevant *C. difficile* strains were used as isolation hosts. The use of a larger panel of potential host strains would be required to show if these environmental isolates harboured prophages which could be induced. Moreover, none of the isolates proved to be useful as sensitive indicator strains in detecting

phage in any of the environmental samples tested. Isolates were confirmed as *C. difficile* using the commercially available latex agglutination assay and PCR. The *spo0A* gene was selected for PCR amplification as it is conserved across *C. difficile* strains and *C. difficile spo0A* specific primers designed by Steve Cartman (N. Minton, University of Nottingham), were used.

Although PCR ribotyping is a useful molecular typing method for *C. difficile,* it can only group strains by the banding pattern produced from amplification of the intergenic spacer region between the 16S and 23S rRNA genes. It does not consider the full genome, unlike pulsed field gel electrophoresis or full genome analysis (Brazier, 2001). It is therefore unknown to what extent isolates differ within the ribotype. "In house" PCR ribotyping was useful for banding pattern comparison and showed that three isolates exhibited identical banding patterns and were therefore of the same PCR ribotype. The other isolates showed unique banding patterns. PCR ribotype numbers could not be assigned to each banding pattern in house (as a PCR ribotyping band pattern library is not available) and therefore, isolates were sent to the CDRN to be assigned PCR ribotype numbers.

The isolation of five different PCR ribotypes, including the new PCR ribotype 711, shows the diversity of *C. difficile* PCR ribotypes present in the environment. Clinically relevant PCR ribotype 078 was among those isolated, and has previously been isolated from sediment samples from a sewage pipe (Hargreaves *et al*., 2013) and waste water treatment plants (Romano *et al*., 2012). This links with the isolation of a PCR ribotype 078 infecting phage, ΦCD1801, as this PCR ribotype is common in sewage samples. PCR ribotypes 009 and 039

have also previously been isolated from waste water treatment plants (Romano *et al*., 2012), however phages for these PCR ribotypes were not isolated in this study. The rare PCR ribotype 011 was also isolated and has been previously isolated from patient faecal samples (Solomon *et al*., 2011). The novel PCR ribotype 711 was also isolated and requires characterisation in terms of sporulation behaviour and toxicity to determine whether it is of clinical relevance.

The growth profile and number of viable cells present were analysed for the four phage propagating strains, as the assays for phage characterisation require infection at specific growth phases or when a certain number of cells are present. It was determined that the four phage propagation strains have sufficiently similar growth profiles and therefore it is likely that phage characterisation assay conditions can be optimised for one bacteria/phage combination and will be applicable to the range of bacteria/phage combinations of interest.

3.4 Chapter Summary

The main outcomes of this chapter are:

- Four novel phages infecting clinically relevant PCR ribotypes of *C. difficile* have been isolated
- Five *C. difficile* isolates have been isolated from an environmental sample and classified by PCR ribotyping, including the identification of novel PCR ribotype 711
- Four phage propagation strains have been characterised in terms of viable cell counts and growth profile

Chapter Four:

Phenotypic Characterisation of Four Novel Bacteriophages Infecting Clinically Relevant PCR Ribotypes of *Clostridium difficile*

4 Phenotypic Characterisation of Four Novel Bacteriophages Infecting Clinically Relevant PCR Ribotypes of *Clostridium difficile*

4.1 Introduction

Many phages infecting *Clostridium difficile* have already been characterised to varying degrees. At the time of writing (August 2017), all published *C. difficile* phages have been characterised using electron microscopy to allow classification by particle morphology (Sell *et al*., 1983; Mahony *et al*., 1985; Goh *et al*., 2005; Govind *et al*., 2006; Mayer *et al*., 2008; Fortier and Moineau, 2007; Horgan *et al*., 2010; Sekulovic *et al*., 2011; Meessen-Pinard *et al*., 2012; Nale *et al*., 2012; Shan *et al*., 2012; Hargreaves *et al*., 2013; Sekulovic *et al*., 2014; Nale *et al*., 2016; Rashid *et al*., 2016), and host range analysed for many. Phages of *C. difficile* described in the literature belong to the order of *Caudovirales,* namely the two sub-families *Siphoviridae* and *Myoviridae.* The *Siphovirus* phages have long, non-contractile tails, whereas the *Myoviridae* phages have contractile tails.

At the time of writing, very few of these phages have been characterised in terms of burst size (Mahony *et al*., 1985; Goh *et al*., 2005; Sekulovic *et al*., 2011) and none in terms of rate of phage attachment, even though these are important attributes to consider for phage therapy. Phages with a high burst size, fast adsorption rate and broad host range are desired, as this makes for a more

effective treatment. Lower dosage is required when phages with a greater burst size are used for therapy.

Phages with broad host range for a particular bacterial species are often desired for therapeutics, but conversely phages with a host range that extends to include other species can lead to undesirable side effects, such as dysbiosis of the gut microbiota, a risk factor of CDI. The host range of *C. difficile* phages seems to vary considerably depending on the phage isolated with some having very narrow host ranges, only infecting a single strain, for example phage ΦCD506 (Sekulovic *et al*., 2014) and some having broader host ranges, where numerous isolates from many different PCR ribotypes can be infected, for example phage ΦCD38-2 (Sekulovic *et al*., 2011) and phage ΦCDHM3 (Nale *et al*., 2016). The reported host range of phages is dependent upon the panel of strains selected for use in host range analysis and the degree to which the strains in the panel have been characterised. A phage which appears to possess a broad host range to a panel of clinical isolates may be misrepresentative if the relatedness of the strains in question has not been assessed, as each isolate may be highly similar or identical. The variety of strains selected within the panel will also vary between reports and therefore will affect whether phages are deemed to have a "broad" or "narrow" host range. To date, there are no reports of any one phage that can infect all known *C. difficile* strains or PCR ribotypes.

Bacteria can be resistant to phage infection by many mechanisms, including through the activity of CRISPR-Cas systems and Abortive Infection (Abi) systems. In the genomes of *C. difficile*, CRISPR-Cas systems (Hargreaves *et al*.,

2014) and Abi systems (Goh *et al*., 2007) have been identified and are thought to contribute to *C. difficile* phage resistance. The most common form of bacterial phage resistance is through the lack of or a non-functioning receptor protein. This has not been reported in *C. difficile* as the phage receptor is as yet unknown.

In the previous chapter, four novel phages infecting clinically relevant *C. difficile* PCR ribotypes, 002, 014, 023 and 078, were isolated. In this chapter, these phages are characterised in terms of particle morphology, rate of phage attachment, burst size and host range.

4.2 Results

4.2.1 Optimising Assay for Determining the Rate of Phage Attachment

To determine the rate of phage attachment to the phage propagating strain, a method described by Clokie and Kropinski (2009) was optimised for use in *C. difficile*. For each phage, the phage propagating strain was grown overnight and an aliquot used to inoculate fresh media and incubated for around four hours such that the cells were in the exponential phase of growth. The number of viable cells at this time point (see 3.2.9) was used to determine the phage infection titre required to obtain the multiplicity of infection (MOI) needed. A low MOI was required to ensure that all phage particles could attach to bacterial cells and to reduce the likelihood that multiple phages could attach to the same bacterial cell. In order to obtain a low MOI, a low infection titre of phages was used, and this also allowed easier enumeration of free phages using plaque assays as less downstream dilutions were required. Preliminary experiments were used to determine the appropriate length of time for sample collection to allow

observation of the reduction in free phage titre, due to phage adsorption, and the beginning of the first burst cycle. In the original protocol a 10-minute sample period was used, however, preliminary experiments using *C. difficile* 1801 showed that a 60-minute period was required to see the desired cycles (Table 4.1). Preliminary experiments for the other three phages (Table 4.1) showed that a 60-minute sampling period was also required for phage ΦCD2301 and ΦCD418, as an increase in phage titre could be observed showing that adsorption had occurred. For phage ΦCD08011 only a 45-minute sampling period was required to see the desired phage cycles, with a high phage titre being observed after 45 minutes. It was determined that sample collection after this time point was unnecessary. As this data was collected in very preliminary experiments, data was not collected in biological triplicate.

Table 4.1: Preliminary data for optimising the rate of phage attachment assay

Time (min)	Titre of Free Phage (pfu/ml)			
	ΦCD08011	ΦCD2301	ΦCD418	ΦCD1801
0	37600000	720000	2900000	89000
5	2592000	516000	300000	57000
10	1320000	484000	600000	50000
15	744000	372000	400000	32000
20	224000	179000	20000	26000
25	50000	130000	56000	31000
30	25000	102000	30000	21000
35	61000	48000	24000	11000
40	600000	95000	1000	6000
45	10800000	720000	3000	1000
50	TNTC*	TNTC*	10000	22000
55	TNTC*	TNTC*	TNTC*	132000
60	TNTC*	TNTC*	TNTC*	535000
65	N/A**	N/A**	N/A**	1164000
70	N/A**	N/A**	N/A**	TNTC*
75	N/A**	N/A**	N/A**	TNTC*
80	N/A**	N/A**	N/A**	TNTC*

*Too many plaques to count on these plates, even with multiple dilutions
**Time points not tested.
Preliminary data was collected from one *C. difficile* culture per propagating strain and plaques were counted by the standard double agar technique. Downstream dilutions were required to produce countable plaques however, in some cases the number of plaques were still too many to count. This data was used to determine the time points phage adsorption could be observed and the subsequent increase in phage titre.

4.2.2 Optimising Single Step Growth Experiment

The single step growth experiment protocol was optimised based on the method described by Sekulovic *et al*. (2011). Preliminary experiments using each phage, along with the data collected for the rate of attachment (Table 4.1), showed that a sampling period of 120 minutes was required to observe the first burst cycle and the following phage titre plateau. The data for phage ΦCD2301 is shown in Figure 4.1, with samples being collected for 90 minutes. At 90 minutes the titre

of free phages seems to plateau however, this titre is not higher than the infection titre and consequently a negative burst size would be calculated for this data. For this assay, a higher MOI was selected, MOI of 1, compared to the MOI selected for the rate of attachment assay. An MOI of 1 aimed to ensure that every bacterial cell present in the sample would be infected by one phage particle. This would limit the impact of progeny phages infecting surrounding cells and would ensure that all progeny phages could be counted during plaque assay. The use of a high phage infection titre required multiple downstream dilutions to enumerate the phages during plaque assays, which in turn meant that although samples were collected for each phage at the same time, multiple separate plaque assays had to be carried out. Some sample degradation may have occurred during this time however, this was unavoidable.

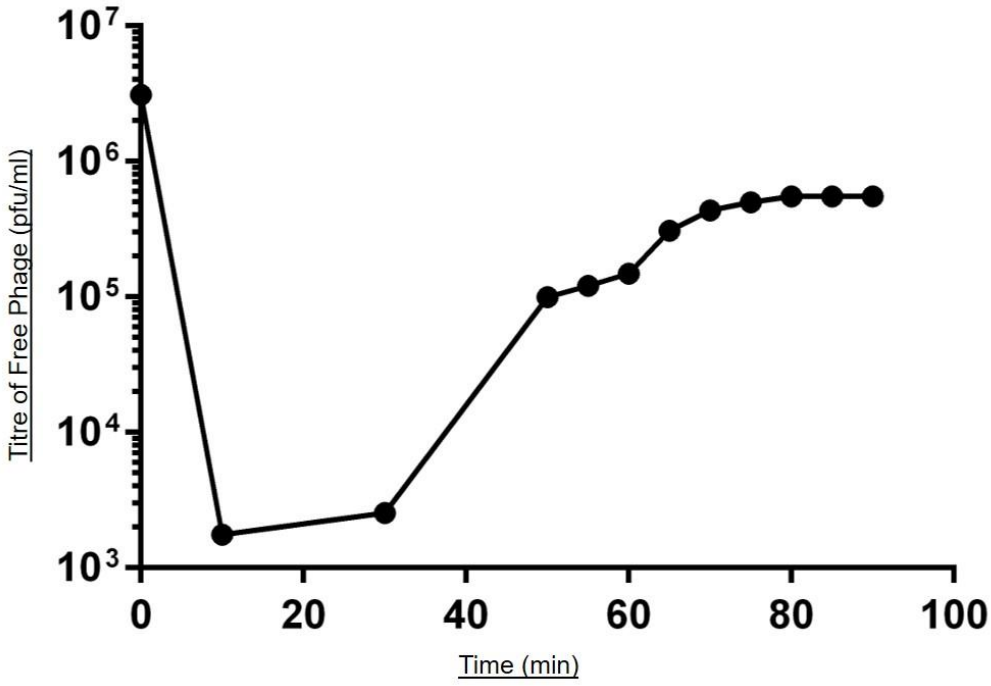


Figure 4.1: **Preliminary data for the determination of phage ΦCD2301 burst size**; the data was not collected in biological triplicate as it was a preliminary experiment to try to optimise the conditions for the burst size assay. In this assay samples were collected for a 90-minute period however, although the free phage titre is increasing, the final titre is not higher than the initial titre and therefore a negative burst size would be calculated. For this reason, the assay was optimised for a 120-minute sampling period.

In the preliminary experiments negative burst sizes were calculated for some of the phages (Figure 4.1), and it was hypothesised that this could be due to the impact of lysogeny as the phages are not strictly lytic. Consequently, the addition of mitomycin C during the assay was tested to try and induce prophages from lysogenised cells to determine whether this could overcome the issue of calculated negative burst sizes. Mitomycin C was added to a bacterial culture of *C. difficile* 2301 (final concentration 3 µg/ml) at the same time as the culture was infected with phage ΦCD2301. The preliminary data showed that this did not have the desired effect with the number of free phages at each time point being

lower than when mitomycin C was not added (Figure 4.2). A low MOI of 0.001 had been used in these assays, and fortunately the increase in the MOI along with the use of exponentially growing culture instead of stationary phase culture, seemed to overcome this problem, suggesting conditions for lytic growth were favoured (see section 4.2.6). Initially, 1 ml samples were taken during the assay however, reducing this to the smaller volume of 50 µl (the sample volume taken for the rate of attachment assays) seemed to also improve the assay conditions as the overall volume of the bacteria/phage mixture was not as impacted as when a large sample volume was taken (determined by comparison of Figure 4.2 and Figure 4.8).

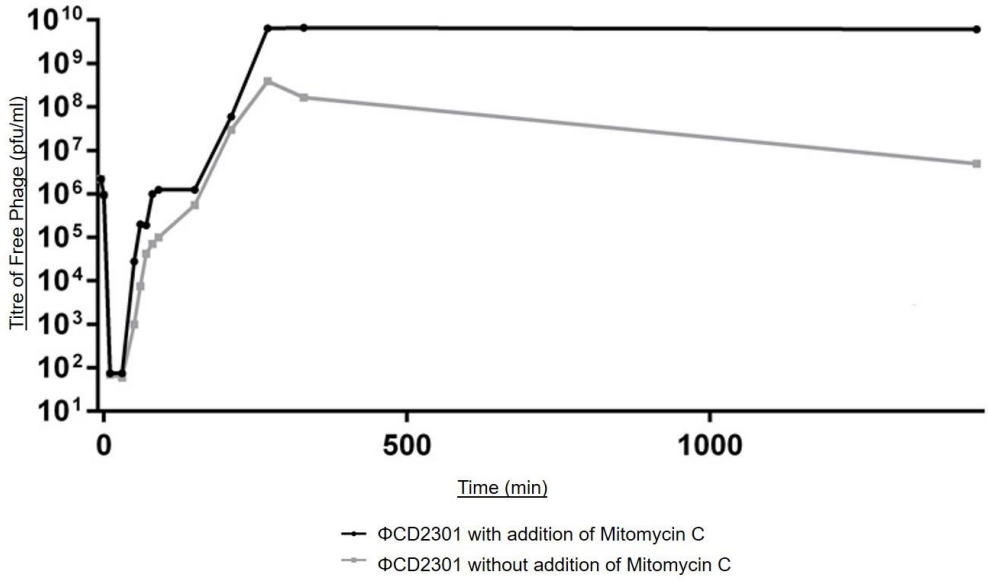


Figure 4.2: **Attempt to optimise burst size assay conditions through the addition of Mitomycin C**. Previous data showed that the final free phage titre for phage ΦCD2301 was lower than the initial phage titre and therefore resulted in a calculated negative burst size. It was hypothesised that this could be caused by the formation of lysogens due to the phage ΦCD2301 not being strictly lytic. The addition of mitomycin C seems to have reduced the number of free phage particles at each time point and therefore has not been successful in optimising the assay conditions to produce positive burst size calculations.

4.2.3 Determining Phage Host Range

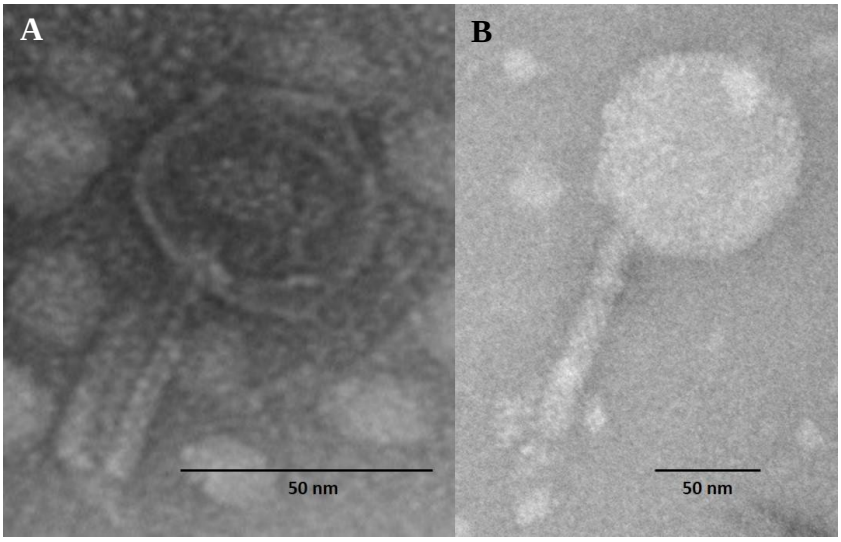
For the four phages, two different methods for determining phage host range were conducted. The first method used plaque assays with multiple phage dilutions to observe sensitivity of other clinical isolates from the same PCR ribotype as the phage propagating strains. In this case, *C. difficile* PCR ribotype 002, 014, 023 and 078 isolates were tested for sensitivity to phages ΦCD08011, ΦCD2301, ΦCD418 and ΦCD1801, respectively. For each isolate the efficiency of plating (EOP) was determined by dividing the phage titre produced by plaquing on the propagating strain with the phage titre produced when plaquing on the indicator strain, in this case the indicator strain is the strain being

examined for sensitivity. The second method relied on the development of a cocktail of the four phages to test a greater number of *C. difficile* isolates from a range of clinically relevant PCR ribotypes. The cocktail of four phages was developed by mixing 1 ml of 10^8 pfu/ml phage stock from each phage and a "spot on the lawn" technique was used to assess sensitivity. If an isolate was sensitive to the phage cocktail, plaque assays were used to determine which phage could cause infection and the EOP in comparison to the propagating strain was established.

4.2.4 Assessing Phage Particle Morphology using TEM

The particle morphology of the four phages was assessed through TEM analysis and allowed the classification of three of the phages as *Myoviridae* (Figure 4.4, Figure 4.5, Figure 4.6). The tail of phage ΦCD08011 appeared contracted under the TEM suggesting a membrane bleb contaminant in the sample. The capsid appears empty as it is not electron dense, further suggesting that the tail had contracted and the phage DNA released (Figure 4.3). Particle morphology could therefore not be determined as tail length ranged from 40 to 160 nm and capsid diameter ranged from 52 to 180 nm. TEM of this phage must be repeated to determine whether this is also *Myoviridae*. Phages ΦCD1801 and ΦCD2301 have tail lengths 137±10.8 nm and 106.4±17.4 nm respectively with similar capsid diameters, 58.5±5.5 nm and 50.4±5.3 nm respectively (Figure 4.4, Figure 4.5). The presence of a neck, connecting the capsid and tail, is visible for phages ΦCD08011, ΦCD2301 and ΦCD1801, suggesting *Myoviridae*. Phage ΦCD418 was classified as *Myoviridae* however, the appearance of potentially flexible tails on some of the particles could suggest *Siphoviridae* (Figure 4.6b). Unlike

the other three phages, the neck is not obviously visible however, the tail appears to be covered in a sheath and appears too short for *Siphoviridae,* hence should more likely be classified as *Myoviridae.*



Tail length: cannot be determined
Tail diameter: cannot be determined
Capsid diameter: cannot be determined

Figure 4.3: **ΦCD08011 phage particle under TEM analysis**. A: the tail of the particle appears to be contracted, measuring only 40 nm in length, and the phage capsid appears empty suggesting phage DNA release, B: the capsid of the particle appears larger than in image A, however, the tail is longer and appears intact. Tail and capsid measurements cannot be accurately determined due to tail contraction and empty phage capsids.

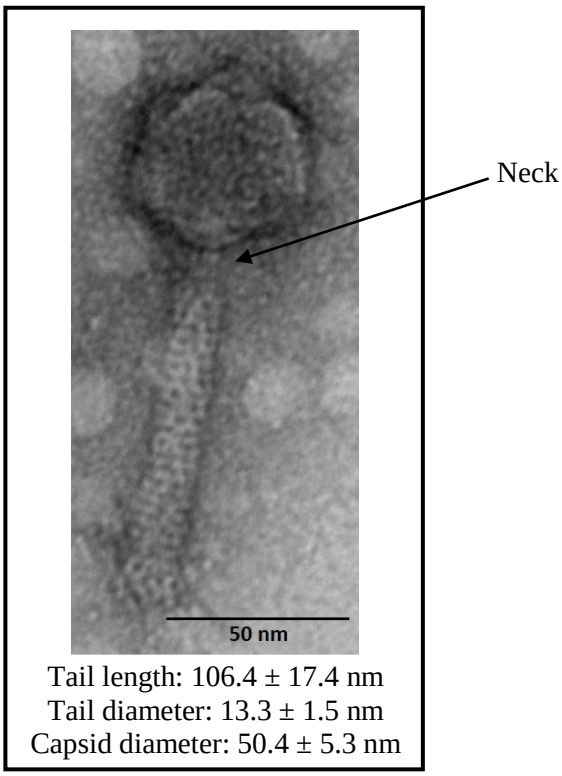


Figure 4.4: **ΦCD2301 phage particle under TEM analysis**; showing the phage belongs to the *Myoviridiae* sub-family of the order *Caudovirales.* The capsid is icosahedral and the tail contractile sheath is visible, as is the neck connecting the tail and capsid. Capsid and tail measurements are the mean values of five phage particles and the standard deviations are shown.

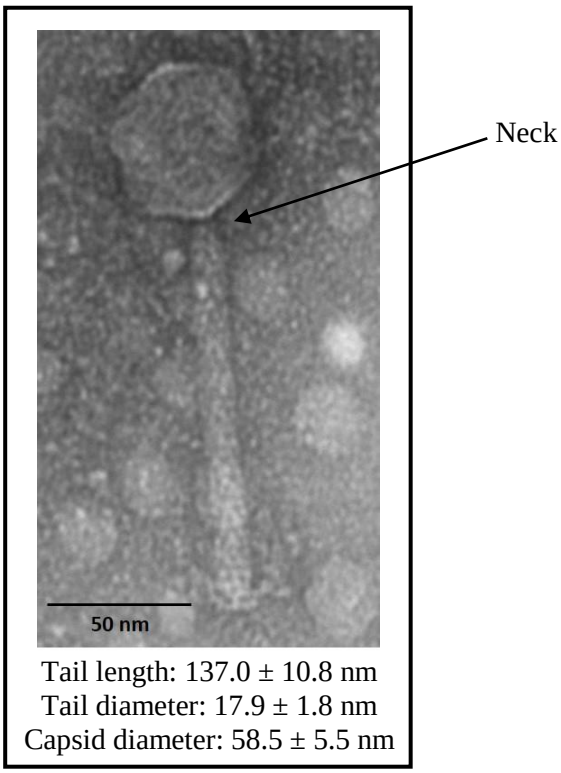
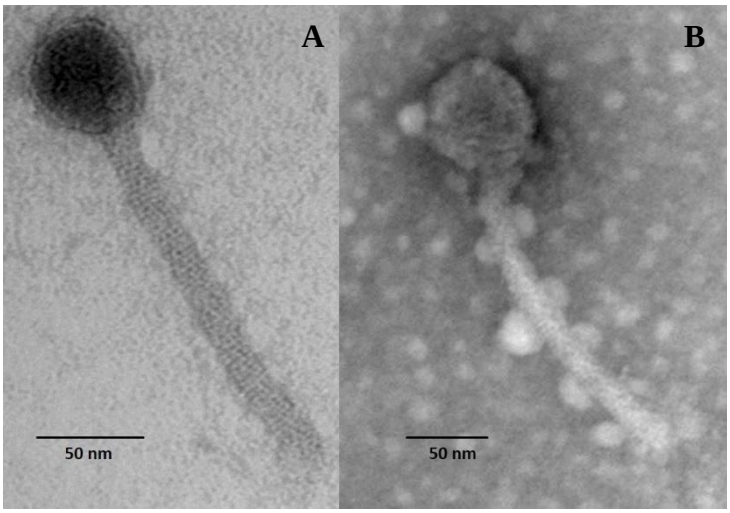


Figure 4.5: **ΦCD1801 phage particle under TEM;** showing the phage belongs to the *Myoviridiae* sub-family of the order *Caudovirales.* The capsid is icosahedral and the tail contractile sheath is visible. The neck connecting the capsid and tail is also visible. Capsid and tail measurements are the mean values of five phage particles and the standard deviations are shown.



Tail length: 204.0 ± 5.1 nm
Tail diameter: 16.9 ± 1.6 nm
Capsid diameter: 58.1 ± 5.4 nm

Figure 4.6: **Phage ΦCD418 particles visualised using TEM**; the phage has been classified as *Myoviridae*, due to its tail length, appearing too short for *Siphoviridae.* However, the tail of some particles appears to be flexible (B). The capsid appears isocahedral (A), and the neck is not clearly visible (therefore not annotated). Phage tail and capsid measurements are the mean values of five individual phage particles and the standard deviations are shown.

4.2.5 Measuring Rate of Phage Attachment

The rate of phage attachment to their respective propagating strains was determined for the four isolated phages, using a method based on Clokie and Kropinski (2009), but optimised for each of the phages. The rate of attachment constant (k) was determined for each phage by assessing the time point at which free phage titre was lowest before the titre began to increase, showing the start of the burst cycle (Figure 4.7). Some of the data points have large error bars (phage ΦCD08011 35 minutes and phage ΦCD2301 50 minutes) suggesting outlying points. This probably occurred when the plaques were counted and is most likely an artefact of the plaque assay rather than of the adsorption assay, as the samples were taken in biological triplicate but not counted in technical

triplicate. The rate constant was very similar between three of the phages, ΦCD08011, ΦCD2301 and ΦCD1801. Phage ΦCD418 appeared to have a faster rate of attachment than the other phages, with the rate constant, k, measuring 4.27 x 10^{-10} ml/min (Table 4.2). For all phages, complete adsorption had occurred between 30 and 45 minutes post infection.

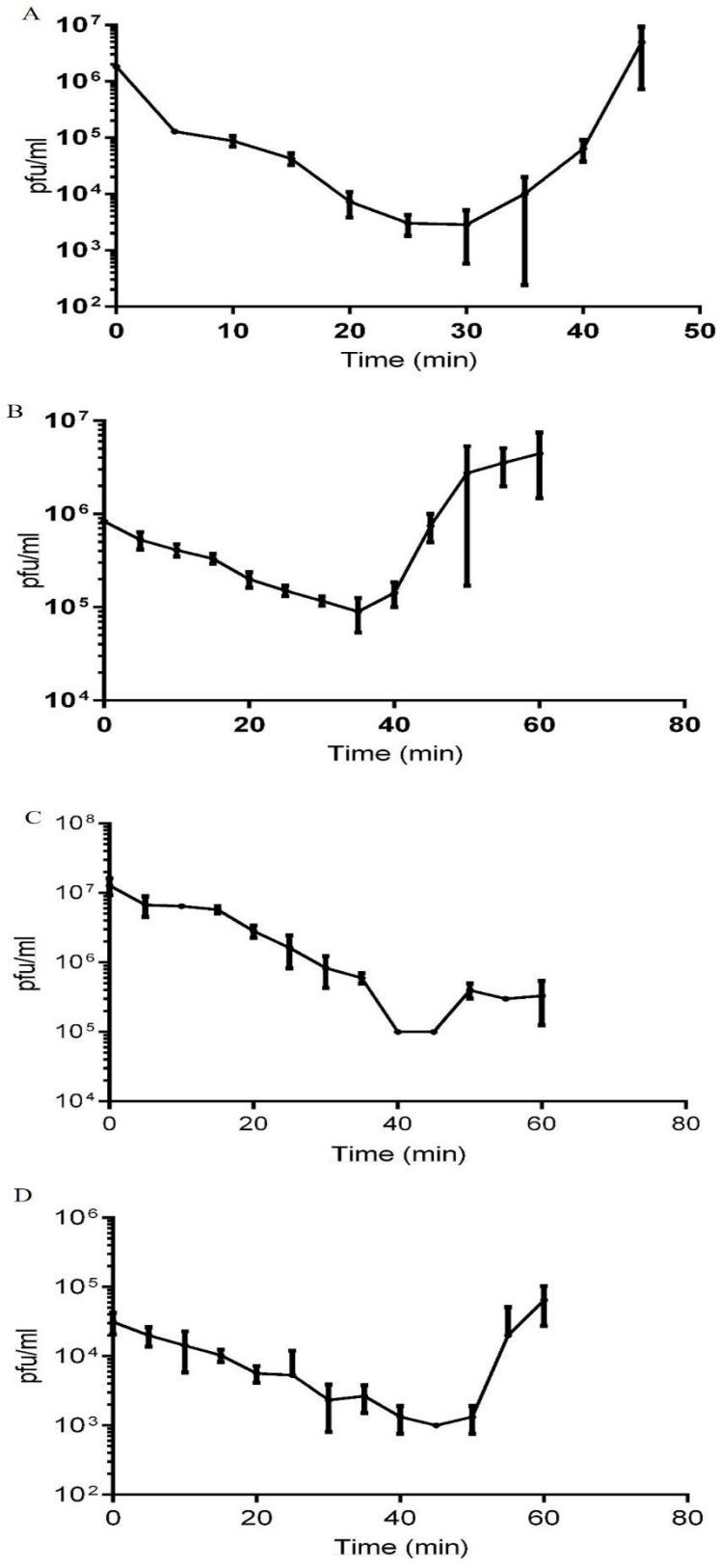


Figure 4.7: **Determining the rate of attachment of the four phages to their propagating host strains.** A: Phage ΦCD08011 attachment to *C. difficile*

08011, B: Phage ΦCD2301 attachment to *C. difficile* 2301, C: ΦCD418 attachment to *C. difficile* 522418, D: Phage ΦCD1801 attachment to *C. difficile* 1801. The rate of attachment constant (k) was determined for each phage over a 45-minute period for phage ΦCD08011 and a 60-minute period for phages ΦCD2301, ΦCD418 and ΦCD1801. Samples were taken, in biological triplicate, at five-minute intervals after infection and enumerated by plaque assay. Standard error bars are shown.

Table 4.2: Rate of attachment constant (k) for each of the four isolated phages

Phage	**Rate of Attachment Constant (k*, ml/min)**	**Time Interval for Phage Adsorption to Occur (min)**
ΦCD08011	5.84×10^{-12}	30
ΦCD2301	2.40×10^{-12}	35
ΦCD1801	1.88×10^{-12}	40
ΦCD418	4.27×10^{-10}	45

*Rate of Attachment Constant (k) = (2.3/B*t*)log(Po/P), where B is the number of bacterial cells, *t* is the time interval for adsorption to occur, Po is the original phage titre and P is the final phage titre after adsorption.

4.2.6 Single Step Growth Experiments

The phage burst size is defined as the yield of phage progeny that are spontaneously produced from an infected cell in a single cycle and is measured as an average burst of the population, using a single step growth experiment. Burst size was calculated as the (Final Phage Titre – Initial Phage Titre) / Initial Phage Titre, where the initial titre is the difference between the infection titre and the number of unbound phages after the 15-minute incubation which allows for adsorption. The point at which a titre plateau in free phages is reached is denoted as the end point of the first burst cycle. For the three phages, ΦCD08011,

ΦCD2301 and ΦCD1801, the end point was determined to be between 70 and 80 minutes (Figure 4.8a, b and d) after the bacteria and phage were mixed and washed. Whereas, for phage ΦCD418, the end of the first burst cycle did not take place until 120 minutes had elapsed (Figure 4.8c). For phages ΦCD08011, ΦCD1801 and ΦCD418, the burst sizes were similar, being 9±2, 8±6 and 13±1 pfu per infected cell, respectively (Table 4.3). The burst size of ΦCD2301 was considerably smaller, measuring 2 pfu per infected cell (Table 4.3). Burst sizes this small would not be indicative of natural bacteria phage dynamics and would not allow sustainment of the phage population. Although the methodology is sound, and the calculations correct for the data (pfu/ml) collected, the burst size assay is not efficient in estimating burst sizes that would be indicative of the "real world." The calculated burst sizes are also not in keeping with the data proposed graphically as the graphs show that at least a two-log increase in pfu/ml is observed over the time-period tested. It is a known problem amongst phage researchers that estimated burst sizes are not always representative of experimental data. Therefore, burst sizes cannot be accurately determined by this method.

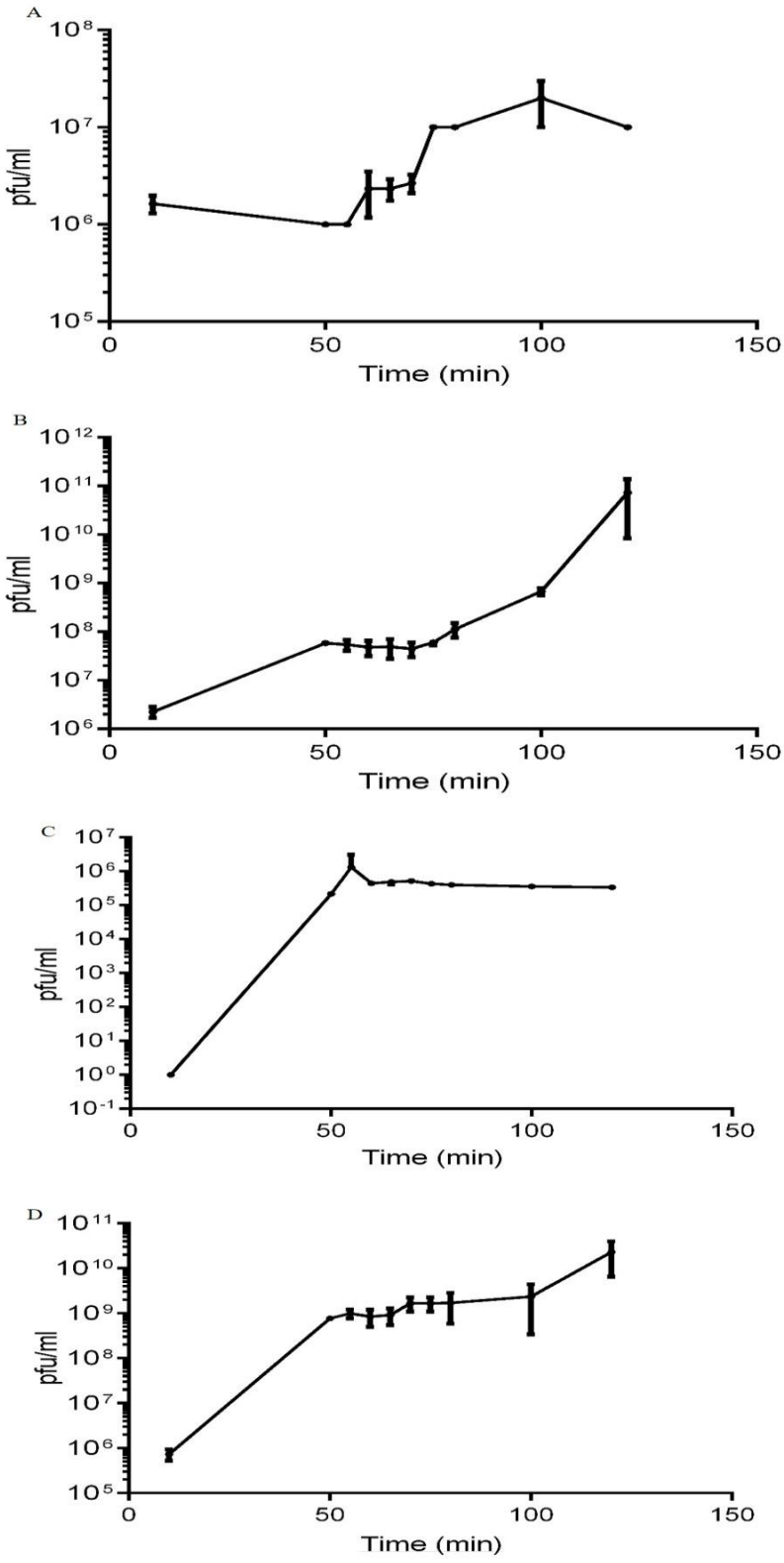


Figure 4.8: **Single step growth curves for the four isolated phages.** A: Phage ΦCD08011, burst size end point at 70 minutes, B: Phage ΦCD2301, burst size end point at 75 minutes, C: Phage ΦCD418, burst size end point at 120 minutes, D: Phage

ΦCD1801, burst size end point at 80 minutes. The end of the first burst cycle is denoted when free phage titre reaches a plateau. Propagating strain cultures were infected and phages allowed to attach to bacterial cells for 15 minutes prior to washing and then sampling. The number of free phages at each time point was enumerated by plaque assay and biological triplicate samples conducted for each phage. Standard error bars are shown.

Table 4.3: Burst Sizes for each of the four isolated phages

Phage	Time to End of First Burst Cycle (min)	Burst Size (pfu/infected cell*)
ΦCD08011	70	9±2
ΦCD2301	75	2±0
ΦCD1801	80	8±6
ΦCD418	120	13±1

*Burst size to nearest integer value and determined as (Final Titre-Initial Titre)/Initial Titre, where initial titre is Infection Titre – Titre of Unbound Phages post 15-minute incubation to allow attachment of phages to bacterial cells. Burst size was determined as an average for the population and conducted in triplicate for each phage/propagating strain combination. Standard deviations are shown. Calculated burst sizes are not consistent with the graphical data presented in Figure 4.8, as at least a two-log difference was observed experimentally.

4.2.7 Host Range Analysis of the Four Phages

The host range of the four phages was assessed using two methods. The first used the individual phages and tested for sensitivity using other *C. difficile* isolates from the same PCR ribotype as the propagating strains. The second method utilised a cocktail of four phages and a "spot on the lawn" technique to test a large number of isolates. Efficiency of plating (EOP) of sensitive strains was determined by plaque assay. In total, 162 *C. difficile* isolates, across nine different PCR ribotypes, (n=17 PCR ribotype 001, n=24 PCR ribotype 002, n=20 PCR ribotype 014, n=12 PCR ribotype 015, n=16 PCR ribotype 023, n=46 PCR ribotype 027, n=17 PCR ribotype 078, n=9 PCR ribotype 106, n=1 PCR ribotype

122, including propagating strains) were tested for sensitivity against the four phages. A full list of the isolates tested and their original isolation locations can be found in appendix 8.1. Of the 162 isolates tested, 43 isolates were sensitive to phage infection (Figure 4.9, Figure 4.10), to varying degrees of efficiency (see appendix 8.2 for the full list of EOPs).

Firstly, isolates from the same PCR ribotype as the propagating strain were tested for sensitivity. Phage ΦCD1801 had the broadest host range within isolates of PCR ribotype 078, infecting 94.1 % of isolates, which equated to 16 out of the 17 isolates tested being sensitive. Phage ΦCD08011 also showed a relatively broad host range, infecting 66.6 % of the isolates tested within PCR ribotype 002. Phages ΦCD2301 and ΦCD418 showed the narrowest host ranges, infecting 20 % and 25 % of isolates within PCR ribotype 014 and 023, respectively (Figure 4.9, Figure 4.10).

Further host range testing studied the sensitivity of isolates from the most common PCR ribotypes in a Nottinghamshire *C. difficile* collection, namely PCR ribotypes 001, 015, 027 and 106. This data showed that the phages had narrow host ranges with only three isolates being infected by any of the phages, notably no PCR ribotype 027 isolates were sensitive to phage infection (Figure 4.9, Figure 4.10). Interestingly, one PCR ribotype 106 isolate, CDDH183, was sensitive to infection by phage ΦCD2301, which usually infects PCR ribotype 014. CDDH183 was initially thought to be PCR ribotype 027, however it was confirmed as PCR ribotype 106 by the CDRN, Leeds. Of the 12 PCR ribotype 015 isolates tested, one, CD0101, was sensitive to infection by phage

ΦCD08011, which has a relatively broad infection spectrum for PCR ribotype 002. In both cases, the efficiency of phage infection was low, with EOP measuring less than 0.1 (appendix 8.2). Importantly each phage can only infect its own propagating strain.

C. difficile isolate	PCR Ribotype	Sensitivity to Phage Infection			
		ΦCD08011	ΦCD2301	ΦCD418	ΦCD1801
CD001-1	001				
CD001-2	001				
CD001-3	001				
CD001-4	001				
CD001-5	001				
CD001-6	001				
CD001-7	001				
NCTC 11205	001				
NCTC 11207	001				
NCTC 11209	001				
CD30011	001				
CD3401	001				
CD3901	001				
CD66011	001				
CD7501	001				
CD9601	001				
CDUK1	001				
CD1101	002				
CD1501	002				
CD2001	002				
CD2101	002				
CD4101	002				
CD52011	002				
CD6001	002				
CD83012	002				
CD9401	002				
CD08011	002				
CDSubtypeA2	002				
CD16839	002				
CD8083598	002				
CD8085053	002				
CD8092419	002				
CD9001966	002				
CD9003048	002				
CDMF081988	002				
CDTL178	002				
CD01013C17	002				
CD01026C3	002				
CD1036C4	002				
CD03008C4	002				
CD03008C2	002				

C. difficile isolate	PCR Ribotype	Sensitivity to Phage Infection			
		ΦCD08011	ΦCD2301	ΦCD418	ΦCD1801
CD2601	014				
CD4011	014				
CD5601	014				
CD6701	014				
CD77011	014				
CD8001	014				
CD88011	014				
CD2301	014				
CDserogroupH	014				
CDserogroupK	014				
CD8085054	014				
CDTL176	014				
CD01006C10	014				
CD01006C1	014				
CD01006C11	014				
CD01013C19	014				
CD11007C2	014				
CD11007C14	014				
CD11007C1	014				
CD11007C17	014				
CD0101	015				
CD0301	015				
CD1901	015				
CD3601	015				
CD4201	015				
CD4501	015				
CD4701	015				
CD5901	015				
CD6201	015				
CD7301	015				
CD8401	015				
CD8501	015				
CD5301	023				
CD5101	023				
CD7801	023				
CD7901	023				
CD2401	023				
CD1202	023				
CD1002	023				
CD3701	023				
CD520557	023				

Key:
- Resistant isolates
- EOP < 0.1
- EOP 0.1-1
- EOP ≥ 1

Figure 4.9: **Heat map of sensitivity of *C. difficile* isolates from PCR ribotypes 001, 002, 014, 015 and 023.** Each isolate was tested for sensitivity to the four phages, ΦCD08011, ΦCD2301, ΦCD1801 and ΦCD418. EOP was determined, for sensitive isolates, as the phage titre of propagating strain divided by the phage titre of the indicator strain. EOPs can be found in appendix 8.2.

C. difficile isolate	PCR Ribotype	Sensitivity to Phage Infection			
		ΦCD08011	ΦCD2301	ΦCD418	ΦCD1801
CD520227	023				
CD522682	023				
CD522418	023				
CD21	023				
CD305	023				
CD01015C1	023				
CD01015C14	023				
CDL1	027				
CD8156	027				
CDLibramont1	027				
CD5108111	027				
CD32219	027				
CDM246	027				
CD2191	027				
CD51556	027				
CD51557	027				
CD027Alexander	027				
CD60902	027				
CD26131	027				
CD26173	027				
CDC32	027				
CDC38	027				
CDM13042	027				
CDR20291	027				
CD81566	027				
CD60109015048	027				
CD31662	027				
CDDH478	027				
CDDH482	027				
CDDH361	027				
CDDH349	027				
CDDH85	027				
CDDH326	027				
CDDH131	027				
CDDH320	027				
CDDH1916	027				
CDDH1396	027				
CDDH1432	027				
CDDH1834	027				
CDDH1466	027				

C. difficile isolate	PCR Ribotype	Sensitivity to Phage Infection			
		ΦCD08011	ΦCD2301	ΦCD418	ΦCD1801
CDDH1329	027				
CDDH1751	027				
CDDH1342	027				
CDDH1858	027				
CDR24694	027				
CDR24988	027				
CDR23970	027				
CDR12087	027				
CDR20352	027				
CDR20298	027				
CD2801	027				
CD5101	027				
CD9201	027				
CD3801	078				
CD4401	078				
CD8601	078				
CD9001	078				
CD9501	078				
CD9701	078				
CD9801	078				
CD1801	078				
CDWilcox078	078				
CDType078	078				
CD2315	078				
CD2016	078				
CD7004578	078				
CD7009825	078				
CD7045389	078				
CDM120	078				
CD01027C1	078				
CDDH183	106				
CDR10432	106				
CDR12801	106				
CDR15347	106				
CDR108095	106				
CDR22079	106				
CDR23942	106				
CDR25469	106				
CDR27384	106				
CD051223046	122				

Key:
- Resistant isolates
- EOP < 0.1
- EOP 0.1-1
- EOP ≥ 1

Figure 4.10: **Heat map of sensitivity of *C. difficile* isolates from PCR ribotypes 023, 027, 078, 106 and 122.** Each isolate was tested for sensitivity to the four phages, ΦCD08011, ΦCD2301, ΦCD1801 and ΦCD418. EOP was determined, for sensitive isolates, as the phage titre of propagating strain divided by the phage titre of the indicator strain. EOPs can be found in appendix 8.2.

From the host range analysis, it was determined that two isolates could be infected by more than one phage. CD2001, PCR ribotype 002, was sensitive to infection by phages ΦCD08011 and ΦCD2301. Infection of CD2001 with phage ΦCD08011 was more efficient than infection of the propagating strain, CD08011, with an EOP measuring 1.08, whereas infection with ΦCD2301 was much less effective, with an EOP measuring less than 0.1 (Figure 4.9, Figure 4.10, appendix 8.2). The second isolate, CD051223046, was sensitive to infection by three phages, ΦCD08011, ΦCD2301 and ΦCD1801. Initially it was thought this isolate was a PCR ribotype 027 however, it was confirmed as PCR ribotype 122 by the CDRN, Leeds. In this case, the efficiency of phage infection varied between the three phages, with ΦCD08011 being the most effective, EOP 0.5 (appendix 8.2). Infection with ΦCD2301 produced disruption of growth with very small plaques being produced down to dilution, 10^{-2} from the phage stock, ~10^{6} pfu/ml, however, not to an enumerable number. Further dilutions of the phage stock showed no plaques and therefore, the EOP could not be determined. This was observed across all plates, in each plaque assay, when this isolate/phage combination was used. Infection with phage ΦCD1801 was observed however, the EOP remained very low, measuring 0.00024 (appendix 8.2).

It was postulated that CD051223046 could be a mixed population instead of pure culture to account for sensitivity to multiple phages. This was investigated by selecting 10 individual colonies and re-streaking these to purity three times on BHIsCC plates. Subsequently, each colony was used in a plaque assay to test for sensitivity to the three phages. In all cases the titre for each phage was the same

(Table 4.4), proving that CD051223046 is a pure culture that is sensitive to infection by three phages.

Table 4.4: Determining CD051223046 as a pure culture through phage infection

CD051223046 Colony Number	**Phage Titre (pfu/ml)**		
	ΦCD08011	ΦCD2301	ΦCD1801
1	2.95×10^8	N/A	2.10×10^6
2	3.10×10^8	N/A	2.95×10^6
3	2.45×10^8	N/A	2.20×10^6
4	3.05×10^8	N/A	3.00×10^6
5	2.90×10^8	N/A	1.75×10^6
6	3.40×10^8	N/A	1.85×10^6
7	3.30×10^8	N/A	2.15×10^6
8	2.55×10^8	N/A	2.05×10^6
9	2.75×10^8	N/A	1.40×10^6
10	3.55×10^8	N/A	1.50×10^6

Titres are not available for ΦCD2301 as no countable plaques were produced even with dilution of the phage stock (plaques were too small and too numerous). Titres for ΦCD08011 and ΦCD1801 are similar across the 10 colonies, suggesting a pure population. Phage titres were enumerated by plaque assay.

4.3 Discussion

4.3.1 Phage Particle Morphology

Phages ΦCD1801 and ΦCD2301 were classified as *Myoviridae*, due to their icosahedral capsids, contractile tails and visible necks. The exact classification of ΦCD08011 is problematic as the tail length and capsid diameter could not be accurately determined due to tail contraction and the observation of empty phage heads. This is likely due to the presence of membrane blebs in the sample triggering DNA release. Using a combination of the measurements across multiple phage preparations, it is likely that this phage can be classified as a small myovirus. Although due to the reasons stated above this cannot be

determined with certainty. Small myoviruses have a capsid diameter between 40 and 58 nm and tail lengths of ~106 nm (Hargreaves and Clokie, 2015). In the first TEM sample preparation, the capsid diameter was measured as 52 nm, which would fit with classification as a small myovirus, although the tail length is slightly longer at 120-160 nm. The capsid diameter and tail length of phage ΦCD2301, 50.4 nm and 106 nm respectively, indicates this is also a small myovirus.

Medium myoviruses are characterised as phages of *C. difficile* with capsid diameters of 60 to 70 nm and tail lengths of 110 to 130 nm (Hargreaves and Clokie, 2015). Phage ΦCD1801 has a tail length of 137 nm and a capsid diameter of 58.5 nm, demonstrating that it can be grouped into the medium myoviruses.

Unlike the other three phages, the neck connecting the capsid and tail was not visible for ΦCD418 and for some particles the tail appeared flexible. This suggests it belongs to the *Siphoviridae* family. Although the tail length and capsid diameter, 204 nm and 58 nm respectively, could suggest belonging to the long tailed myovirus group, it is not definite. Published *Siphoviridae* phages infecting *C. difficile*, are reported to possess capsid diameters of 46 to 90 nm and tail lengths ranging from 262 to 432 nm (Fortier and Moineau, 2007; Sekulovic *et al*., 2014). Shorter tail lengths of 193.4 nm have also been reported (Shan *et al*., 2012), further suggesting that ΦCD418 fits into this family. On closer observation of TEM images, the tail clearly shows it is covered with a sheath, this is indicative of *Myoviridae* and has not been documented for *Siphoviridae.* This suggests, along with the tail length, that phage ΦCD418 is most likely a

long tailed myovirus. Analysis of the genome of ΦCD418 will aim to confirm the exact classification of this phage (Chapter 5).

4.3.2 Single Step Growth Experiments and Rate of Attachment

Preliminary experiments to determine the burst size of the four isolated phages was problematic with negative burst sizes measured for some phages. This indicates that the infection titre is higher than the titre at the end of the first burst cycle. It was postulated this could be influenced by formation of lysogens as the phages are not strictly lytic. This was overcome by increasing the MOI used to 1, to try to ensure that every cell was being infected by phage. The addition of mitomycin C, to induce prophages, was unsuccessful and showed no impact on the final phage titres. The observed burst sizes are slightly lower, in the case of phage ΦCD2301, however within the ranges reported for *C. difficile* phages, ranging from 5 to 122 pfu per infected cell (Mahony *et al*., 1985; Goh *et al*., 2005; Sekulovic *et al*., 2011). In comparison to phages of other species, the burst sizes of *C. difficile* phages are comparatively low when one considers that some *E. coli* phages produce burst sizes greater than 1,000 pfu per infected cell (Delbrück, 1945). Burst size calculations are not in keeping with the data presented in the graphs and are not high enough to sustain a phage population in nature. It appears this method has not been appropriate in this situation for calculating phage burst sizes however, with further optimisation this method could still be viable for other phages.

The rate of phage attachment has not been widely reported in phages of *C. difficile.* It has, however, been shown that around 70 % of *C. difficile* phage

41 can absorb in 10 minutes (Mahony *et al*., 1985). In comparison to *E. coli* phage λ, the rate of attachment of the four isolated phages seems slightly slower, with phage λ reporting rates of 1.3×10^{-9} and 9.9×10^{-9} ml/min (Shao and Wang, 2008). The slower rate of attachment could be problematic for a phage therapeutic as the phage particles could be degraded before they reach the site of infection and treatment may not be as efficient.

4.3.3 Host Range Analysis of the Four Phages

The broadest host range phage, ΦCD1801, could infect 94.1 % (n=16/17) of PCR ribotype 078 isolates tested. Only one isolate, CD2315, was resistant to infection. Initially a second *C. difficile* isolate, CD31662, also showed resistance. However, CD31662 was later shown to belong to PCR ribotype 027 and not 078. CD2315 resistance to infection by phage ΦCD1801 could be due to the lack or dysfunction of the correct bacterial receptor binding protein. It could also be due to the presence of a prophage within the CD2315 genome creating lysogenic immunity to infection by phage ΦCD1801. Alternatively, resistance maybe a consequence of other resistance mechanisms, such as CRISPR-Cas systems or Abortive Infection (Abi) systems. The resistance of CD2315 to phage ΦCD1801 is further studied in Chapter 6.

Nale and colleagues studied the host range of 7 different *C. difficile* phages and showed they covered a relatively broad host range, infecting 78 % of isolates tested across 21 different PCR ribotypes. However, only 38 % of PCR ribotype 078 isolates tested were sensitive to infection and only one of the 7 phages were able to infect this PCR ribotype (Nale *et al*., 2016). Another study showed all 11

C. difficile PCR ribotype 078 isolates tested were resistant to infection from two different phages (Rashid *et al*., 2016). This suggests that phage ΦCD1801 is the first report of a broad-spectrum PCR ribotype 078 phage.

Host range analysis showed that all PCR ribotype 001 and 027 isolates were resistant to infection from all four phages. Prophage carriage is high in PCR ribotype 027 strains (Nale *et al*., 2012) and is thought to contribute to the diversity of these strains. It is quite likely that the high prophage carriage is causing the resistance observed to the four phages. However, this cannot be confirmed without full genome sequencing of a diverse set of PCR ribotype 027 isolates. Prophage carriage in *C. difficile* isolates from many different PCR ribotypes has been shown (Shan *et al.*, 2012) and is very likely to contribute to the narrow host ranges of the four phages.

Phage ΦCD08011 could infect 66.6 % (n=16/24) of the PCR ribotype 002 isolates and could also infect isolates from PCR ribotype 015 and 122. PCR ribotype 002 isolate, CD2001, was sensitive to phages ΦCD08011 and ΦCD2301, although with a very low efficiency for the latter. This suggests that phages ΦCD2301 and ΦCD08011 can attach to the same receptor protein and intrinsic resistance systems must be causing resistance of the other PCR ribotype 002 isolates to phage ΦCD2301. However, this is not conclusive as phage ΦCD08011 may be able to bind multiple receptors. Phages ΦCD2301 and ΦCD08011 can also infect CD051223046, as can phage ΦCD1801, again suggesting a common receptor, especially since establishing that CD051223046 is not a mixed population. The *C. difficile* phage receptor is further investigated

in Chapter 6. The narrowest host range was observed by phage ΦCD418 which could only infect one PCR ribotype 023 isolate (its propagating strain). It is thought that prophage carriage is high within this PCR ribotype (Emma Stevenson, University of Nottingham, personal communication) potentially explaining the narrow host range of this phage.

In conclusion, it would appear that phage ΦCD1801 would be more suitable for development as a phage therapeutic due to its broad-spectrum activity against *C. difficile* PCR ribotype 078 isolates. This phage does not have a particularly high burst size or fast rate of adsorption which may prove to be limitations for its use as a therapeutic, but it nevertheless has exciting potential in the therapeutics field.

4.4 Chapter Summary

The main outcomes of this chapter are:

- Particle morphology of the four isolated phages was analysed using TEM and suggested that the four phages belong to *Myoviridae* family of tailed phages
- Useful therapeutic attributes, burst size and rate of phage attachment, were determined for each of the four isolated phages
- The host range profile of each phage was determined against a panel of *C. difficile* strains, showing that phage ΦCD1801 possesses broad spectrum killing against *C. difficile* PCR ribotype 078 isolates

Chapter Five:

Genotypic Characterisation of Four Novel Bacteriophages Infecting Clinically Relevant PCR Ribotypes of *Clostridium difficile*

5 Genotypic Characterisation of Four Novel Bacteriophages Infecting Clinically Relevant PCR Ribotypes of *Clostridium difficile*

5.1 Introduction

In line with recent decreases in the cost of DNA sequencing, the number of published *C. difficile* phage genomes has significantly increased since the first genome, ΦCD119, was published in 2006 (Govind *et al*., 2006). At the time of writing (August 2017), 25 complete *C. difficile* phage genomes were available on the NCBI database (Table 1.1).

The genomes of all published *C. difficile* phages are composed of double stranded DNA in either linear or circular form. The availability of genome sizes, along with tail and capsid measurements, has led to improved phage classification, and allowed the subdivision of the myovirus group into small myoviruses, medium myoviruses and long tailed myoviruses (Hargreaves and Clokie, 2015). The presence of certain indicator genes has additionally allowed the classification of some myovirus phages into two genera, *phiCD119likevirus* and *phiMMP04likevirus* (Table 5.1). The genomes of phages within the genus *phiCD119likevirus* possess a three-gene cassette that encodes DnaD, a hypothetical protein and a single stranded DNA binding protein. They also contain a gene encoding for a cytosine C5 specific DNA methylase, whereas the genomes of phages belonging to the *phiMMP04likevirus* genus do not contain the methylase gene or the three-gene cassette. The genomes of

phiMMP04likevirus phages do, however, encode a ParA homologue. ParA is a chromosomal partitioning protein and the gene responsible for its production is usually located downstream of a lysis gene cassette, and a gene encoding for a Clp protease. Genome size is usually between 30 and 33 kbp in this genus, whereas phages belonging to *phiCD119likevirus* generally have larger genomes of around 45 kbp (Hargreaves and Clokie, 2015). The genomes of *C. difficile* phages belonging to *Siphoviridae* also contain genes encoding for a ParA homologue, as well as common endolysin and holin proteins.

Table 5.1: Classification of *C. difficile* by genome features

Myoviridae Genera	**Genome Features**	**Sequenced Phages in this Genus**
phiCD119likevirus	Cytosine-C5 specific DNA methylase 3 gene cassette encoding DnaD, a hypothetical protein and a single stranded DNA binding protein	ΦC2 ΦCDHM1 ΦCDHM19 ΦCD119 ΦMMP02 ΦCD27
phiMMP04likevirus	ParA homologue Clp protease	ΦCD506 ΦCD481-1 ΦMMP04 ΦCDHM1 ΦCDHM13 ΦCDHM14

The study of phage genomes is providing insight into phage evolution and has also provided some information of how certain phage interact with their environment. An example of this occurred in 2014, where *agr* (Accessory Gene Regulator) genes, normally associated with quorum sensing (QS), were identified in the genome of *C. difficile* phage ΦCDHM1 (Hargreaves *et al*.,

2014). QS was subsequently shown (Erez *et al*., 2017) to be involved in the decision to lysogenise by allowing communication between phages and their progeny not through classical N-acyl homoserine lactone or AI2-based quorum sensing systems but by another method that involves the secretion of a peptide of six amino acids in length. This was studied using the Φ3T phage of *Bacillus subtilis*, by showing that when conditioned media was used for a subsequent infection, bacterial cells appeared to be more protected from lysis. It was proven that the small communication peptide was released into the media, and when it had reached sufficient concentration it could be sensed, and progeny phages would preferentially lysogenise their hosts (Erez *et al*., 2017).

5.2 Methodology

5.2.1 Assembly and Annotation of Phage Genomes

CLC Genomics Workbench 8.5.1 was used to assemble the Illumina paired-end sequencing reads into a single contig in which the start and ends were chosen arbitrarily. Artemis was used to predict open reading frames (ORFs) of a minimum of 100 amino acid residues. Each was manually trimmed to the most likely start codon (either ATG, GTG, TTG or CTG) based on the presence of a potential ribosome binding site (RBS), conforming to the sequence AGGAGG or similar. Putative functions were assigned using BLASTp, UniProt and PFAM.

5.2.2 Determination of Linear or Circular Genome Maps

After sequencing, PCR was used to determine whether the genomes exist in a linear or circular form using primers flanking the contig end regions.

Amplification of contig ends would show that the phage genome is circularly permuted and packaged via headful packaging, whereas no amplification would suggest a phage with physical ends and an alternative packaging mechanism with a defined packaging cut site. Primers CD08011_ParA1 and CD08011_pol1 were used for amplification of the contig ends of phage ΦCD08011. For phage ΦCD1801, primers CD1801_portal1 and CD1801_portal2 were used and for phage ΦCD418 primers CD418_phageprot1 and CD418_collar1 were used to amplify the contig ends. PCR was conducted using two primer sets for phage ΦCD2301, the first primer set CD2301_endolysin1 and CD2301_cellwall were unsuccessful in amplification and subsequently primers CD2301_end1 and CD2301_holin1 were used. For each phage PCR was conducted using Q5® High-Fidelity DNA polymerase under conditions stated by the manufacturer (NEB) and annealing temperatures were determined using the NEB Tm calculator (50-55 °C). The regions amplified were sequenced using Sanger Sequencing (Source Bioscience) and used to close the linear contigs to produce circular genomes.

5.3 Results

5.3.1 ΦCD08011

Phage ΦCD08011 is a temperate phage isolated using the host strain *C. difficile* 08011 (PCR ribotype 002). Its genome is composed of double stranded DNA, of 31,394 bp in length with an average GC content of 29.81 % (Figure 5.1). A single contig was assembled, using CLC Genomics Workbench 8.5.1, from 1,656,422 reads with an average read length of 146.38 bp. Of the 35 ORFs predicted by Artemis, a putative product could be assigned to 23, based on protein homology.

PCR using primers flanking the contig ends showed that the genome was circularly permuted, and Sanger Sequencing confirmed that no bases were missing from the original sequencing data. The presence of an integrase gene (CD08011_gp33) indicates phage ΦCD08011 is temperate, in addition to experimental data showing the generation of ΦCD08011 lysogens. The presence of genes encoding a ParA homologue (CD08011_gp26) and a Clp protease (CD08011_gp3) suggests it belongs to the *phiMMP04likevirus* genus of small myoviruses. A full list of genes and their putative product can be found in Appendix 8.3.1.

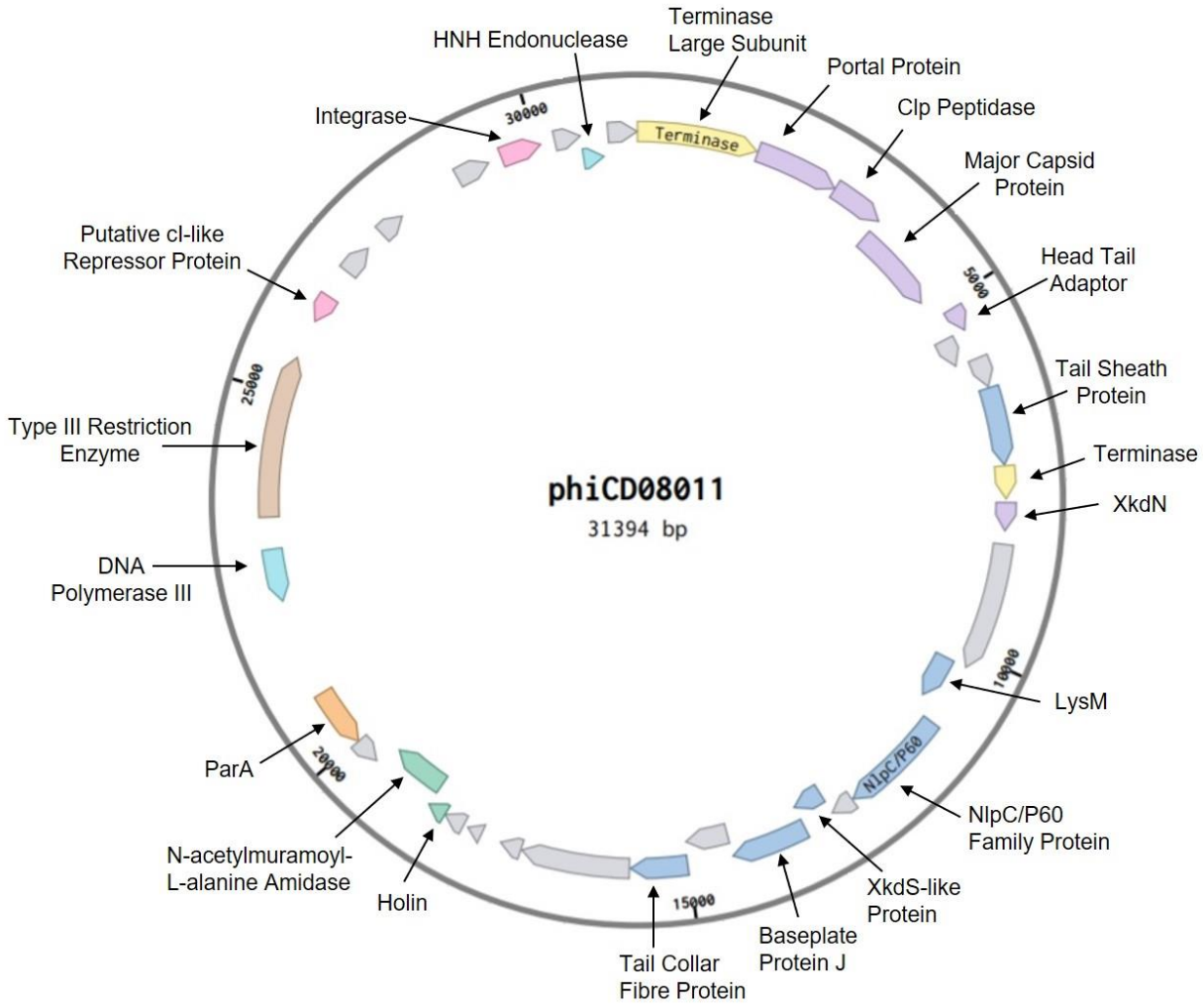


Figure 5.1: **Genome organisation of phage ΦCD08011**. The 31,394 bp genome of phage ΦCD08011 was assembled into a single contig using CLC Genomics Workbench 8.5.1 and was closed using PCR to amplify the contig ends, suggesting the phage is circularly permuted and packaged via headful packaging. The start point is assigned as the gene encoding the terminase protein, as is customary for phage genomes. The predicted CDSs are marked with arrows and colours indicate functional modules: DNA replication (turquoise), lysogeny control (pink), head structural components (purple), DNA packaging (light orange), tail structural components (blue) and cell lysis (green). CDSs with no assigned functionality are shown in grey and a gene encoding a type III restriction enzyme is shown in brown. The presence of a gene encoding a ParA (orange) and Clp protease (purple) suggests phage ΦCD08011 belongs to the *phiMMP04likevirus* genus of small myoviruses.

5.3.2 ΦCD2301

The genome of phage ΦCD2301 is also composed of double stranded DNA, of 38,695 bp in length with an average GC content of 29.47 % (Figure 5.2). It was assembled into a single contig from 1,759,148 reads with an average read length of 168.56 bp using CLC Genomics Workbench 8.5.1. A putative assignment of function based on protein similarity was possible in the case of 27 of the 39 predicted ORFs. PCR amplification to close the genome, using primers flanking the contig ends, was unsuccessful using two different primer sets (primer set one: CD2301_endolysin1 and CD2301_cellwall1, primer set two: CD2301_end1 and CD2301_holin1). This suggests that this phage genome has defined physical ends which is further supported by the raw sequencing data not showing any reads that overlapped at the contig ends. This means that this phage genome is likely packaged by an alternative mechanism, rather than headful packaging. The exact packaging cut sites have not been determined. The presence of an integrase gene (CD2301_gp37) indicates phage ΦCD2301 is temperate phage, in combination with experimental data showing the creation of a lysogen. It belongs to the *phiMMP04likevirus* genus of small myoviruses as its genome carries genes coding for a ParA homologue (CD2301_gp30) and a Clp protease (CD2301_gp3). The genome also contains two genes (CD2301_gp24 and CD2301_gp25) coding for Abortive Infection (Abi) System proteins, a form of bacteriophage resistance mechanism. However, it appears a SNP is present which creates a premature stop codon. This was determined by nucleotide sequence comparison with genes encoding for Abi proteins. The premature stop codon presumably renders this gene non-functional as there is no second RBS site present upstream of the second gene. Prior to further investigation the SNP

would need to be confirmed by Sanger Sequencing to ensure it is not an artefact of genome assembly. A full list of genes and their putative products can be found in Appendix 8.3.2.

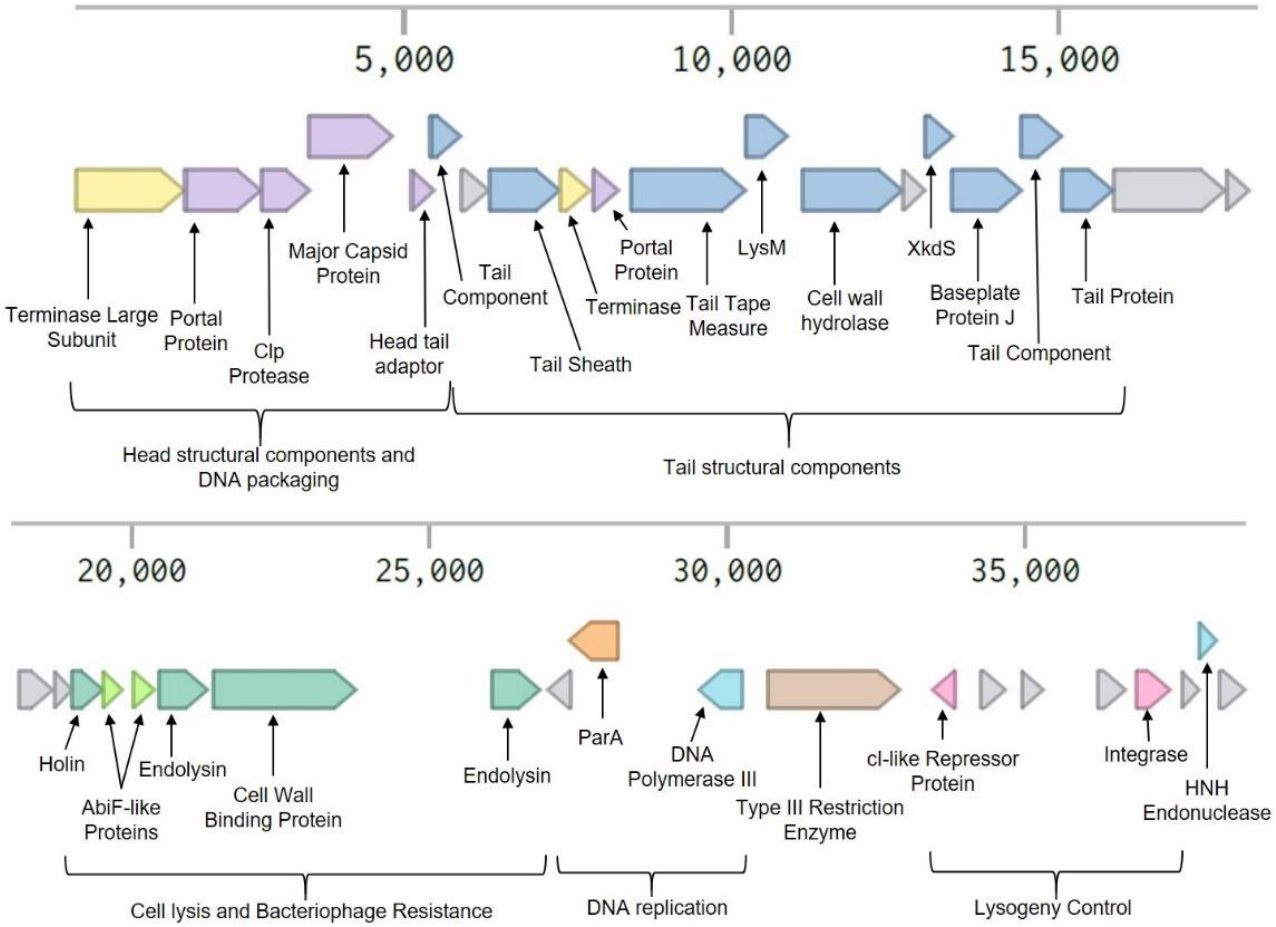


Figure 5.2: **Genome organisation of phage ΦCD2301**. The 38,695 bp genome of phage ΦCD2301 assembled into a single contig using CLC Genomics Workbench 8.5.1. PCR of the contig ends showed that the phage genome had physical ends and is therefore packaged by an alternative mechanism to headful packaging. The exact packaging cut sites have not been determined. The phage genome is represented as a linear genome map and the start has been reassigned to the start of the gene encoding the phage terminase, in the sense orientation, as is customary with phage genomes. The predicted CDSs are marked with arrows and colours indicate functional modules: DNA replication (turquoise), lysogeny control (pink), head structural components (purple), DNA packaging (light orange), tail structural components (blue) and cell lysis (green). CDSs with no assigned functionality are shown in grey and a gene encoding a type III restriction enzyme is shown in brown. A bacteriophage abortive infection system (Abi) is shown in lime green. The presence of a gene encoding ParA (orange) and Clp protease (purple) suggests ΦCD2301 belongs to the *phiMMP04likevirus* genus of small myoviruses.

5.3.3 ΦCD418

ΦCD418 is a double stranded DNA phage with a genome size of 53,311 bp and an average GC content of 29.06 % (Figure 5.3). A single contig was assembled, using CLC Genomics Workbench 8.5.1, from 3,513,940 reads with an average read length of 234.69 bp. Of the 58 identified ORFs, functions were putatively assigned to 35 coded proteins. The generation of a DNA product using PCR primers flanking the contig ends indicated that the genome was circularly permuted and could be represented as a circular genome map. Sanger Sequencing of the PCR amplified DNA fragment confirmed that no bases were missing from the original sequencing data. Phage ΦCD418 is also temperate, shown by experimental data creating lysogens and as it contains an integrase gene (CD418_gp36) in its genome. Unlike phages ΦCD08011 and ΦCD2301, the phage ΦCD418 genome does not contain genes encoding a ParA homologue or a Clp protease. It does, however, contain a putative phage DNA methylase gene (CD418_gp49) and the three-gene cassette present in phages of the *phiCD119likevirus* genus. The three-gene cassette contains genes encoding DnaD (CD418_gp42), a resolvase (CD418_gp43) and a single stranded DNA binding protein (CD418_gp44). The presence of these genes in its genome suggest phage ΦCD418 belongs to the *phiCD119likevirus* genus of medium myoviruses. The presence of a gene encoding a tail sheath protein (CD418_gp12) and the absence of a gene coding for a ParA homologue further confirm this phage as *Myoviridae,* not *Siphoviridae.* The genome of this phage also carries genes encoding a bacteriophage resistance protein (CD418_gp40) and an endolysin (CD418_gp26). A full list of genes and their putative products can be found in Appendix 8.3.3.

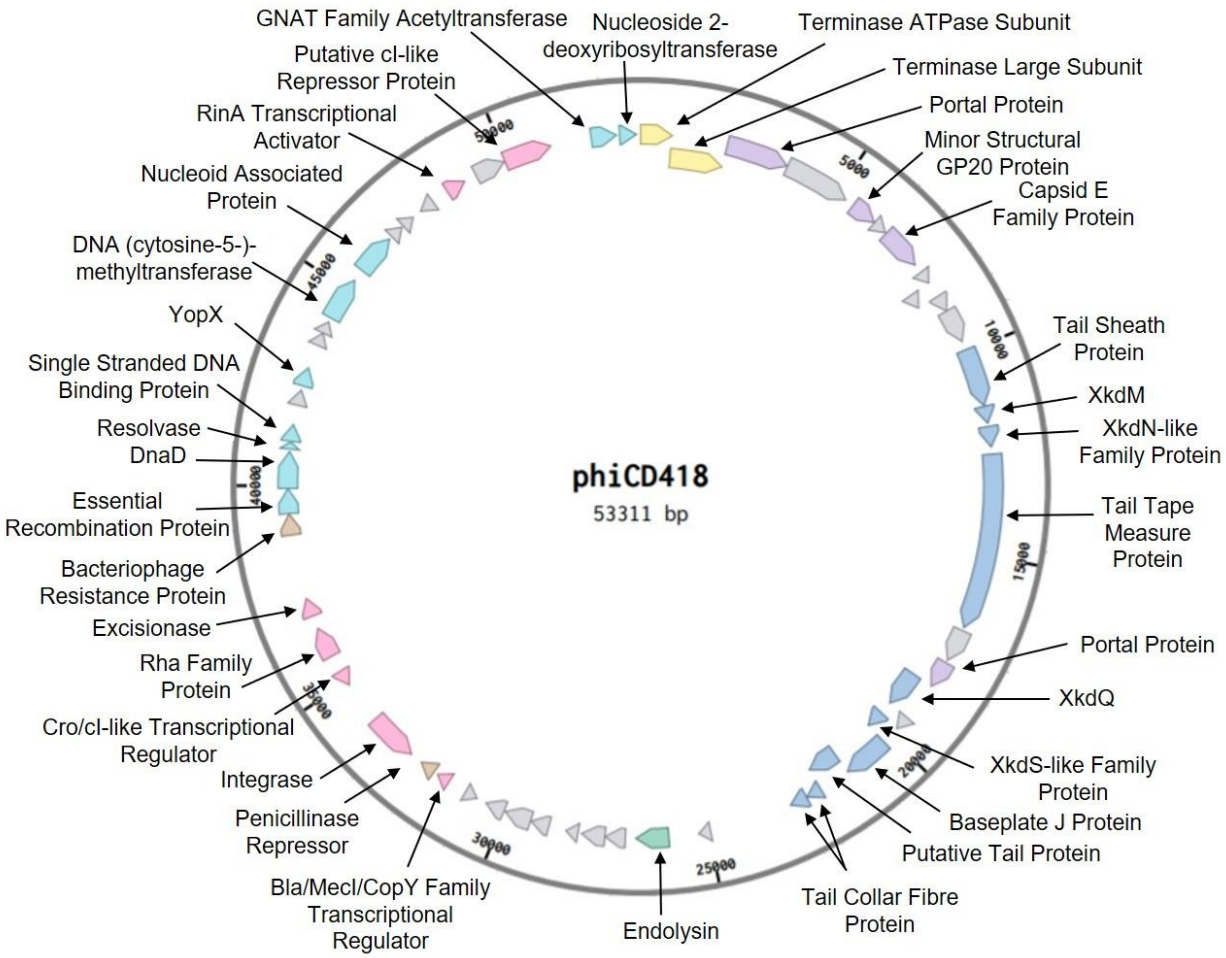


Figure 5.3: **Genome organisation of phage ΦCD418**. The 53,311 bp genome of phage ΦCD418 assembled into a single contig using CLC Genomics Workbench 8.5.1 and was closed using PCR to amplify the contig ends, suggesting that the phage genome is circularly permuted and packaged via headful packaging. The genome is reassigned to start at the gene encoding the terminase protein, as is customary for phage genomes. The predicted CDSs are marked with arrows and colours indicate functional modules: DNA replication (turquoise), lysogeny control (pink), head structural components (purple), DNA packaging (light orange), tail structural components (blue) and cell lysis (green). CDSs with no assigned functionality are shown in grey and a gene encoding a bacteriophage resistance protein is shown in brown. The endolysin gene is shown in green. The presence of a gene encoding a cytosine-5 methyltransferase and a three-gene cassette containing gene encoding DnaD, resolvase and a single stranded DNA binding protein (all shown in turquoise) suggest ΦCD418 belongs to the *phiCD119likevirus* genus of medium myoviruses.

5.3.4 ΦCD1801

ΦCD1801 is a double stranded DNA phage, with a genome size of 44,363 bp and an average GC content of 28.78 % (Figure 5.4). A single contig was assembled, using CLC Genomics Workbench 8.5.1, from 1,446,190 reads with an average read length of 221.49 bp. The contig was successfully closed by amplifying with primers complementary to the contig ends and showed that the genome was circularly permuted. Sanger Sequencing confirmed that no bases were missing from the original sequence data. Putative functions were assigned to 35 of the 50 predicted ORFs. Like all other *C. difficile* phages an integrase gene was identified in the genome of ΦCD1801 (CD1801_gp33), this combined with experimental data showing the creation of lysogens for this phage is proof that it is a temperate phage. This phage does not carry genes encoding either a methylase, a ParA homologue or a three-gene cassette containing *dnaD*. As a consequence, it does not fit into either genus, *phiCD119likevirus* or *phiMMP04likevirus.* The genome of phage ΦCD1801 does, however, appear to contain a gene encoding a cI-like repressor protein (CD1801_gp35) which shares 100 % sequence identity, at the amino acid level, with the gene encoding the cI-like repressor protein carried by *C. difficile* phage ΦCD27 (NCBI Accession No. NC_011398.1). Interestingly this phage also carries a bacteriocin synthesis gene (CD1801_gp48), which is located downstream of a gene encoding a transposase (CD1801_gp50). A full list of genes and their putative products can be found in Appendix 8.3.4.

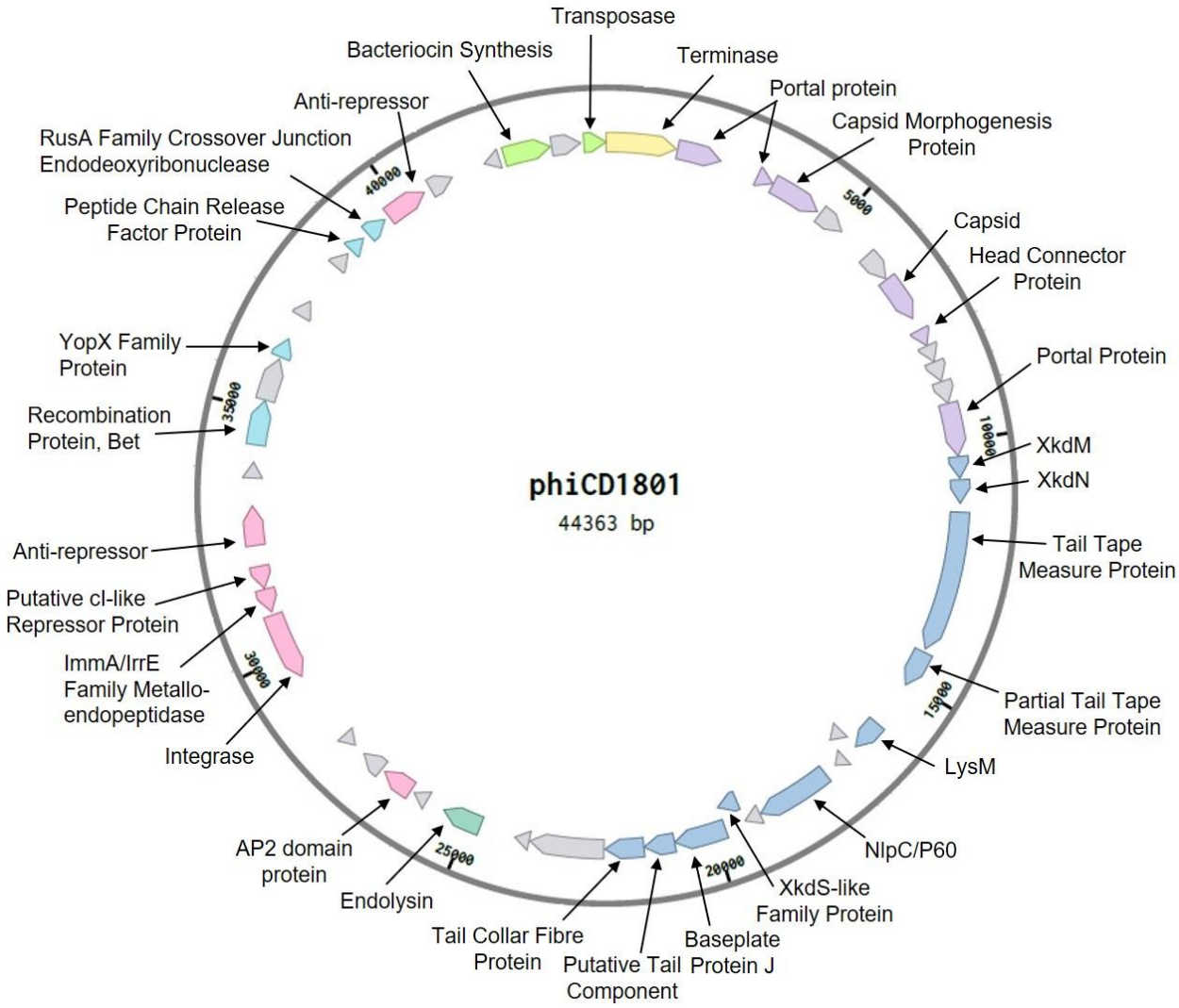


Figure 5.4: **Genome organisation of phage ΦCD1801**. The 44,363 bp genome of phage ΦCD1801 assembled into a single contig using CLC Genomics Workbench 8.5.1 and was closed using PCR to amplify the contig ends, suggesting that the phage genome is circularly permuted and packaged via headful packaging. The genome is reassigned to start at the gene encoding the terminase protein, as is customary for phage genomes. The predicted CDSs are marked with arrows and colours indicate functional modules: DNA replication (turquoise), lysogeny control (pink), head structural components (purple), DNA packaging (light orange), tail structural components (blue) and cell lysis (green). CDSs with no assigned functionality are shown in grey. A gene encoding a protein involved in bacteriocin synthesis and a gene encoding a transposase are shown in lime green. The gene encoding the cI-like repressor protein was identified by amino acid sequence identity to the gene encoding the cI-like repressor protein in the published *C. difficile* phage ΦCD27 genome (NCBI Accession No. NC_011398.1).

5.3.5 Determining the Site of ΦCD1801 Integration in the Bacterial Chromosome

In order to determine the location, within the bacterial genome, of phage ΦCD1801 integration during lysogeny, genome sequencing was utilised. Genomic DNA of *C. difficile* 1801L (CD1801 containing ΦCD1801 prophage) was extracted using a phenol chloroform protocol and subjected to Illumina MiSeq paired-end sequencing. The resulting sequencing reads were assembled using CLC Genomics Workbench 8.5.1 by aligning the reads to the genome of phage ΦCD1801 and the genome of CD1801 (see 6.3.1 for the creation of this genome). This allowed the mapping of 1,144,945 reads with an average read length of 245 bp, using CLC Genomics Workbench 8.5.1. The potential integration site was then manually search for by locating the integrase gene of phage ΦCD1801. It is usual for the phage *attP* site to be immediately downstream of the integrase gene and hence, the integrase gene was searched for. At this point, single reads were studied to identify a read which contained both phage and bacterial DNA (Figure 5.5).

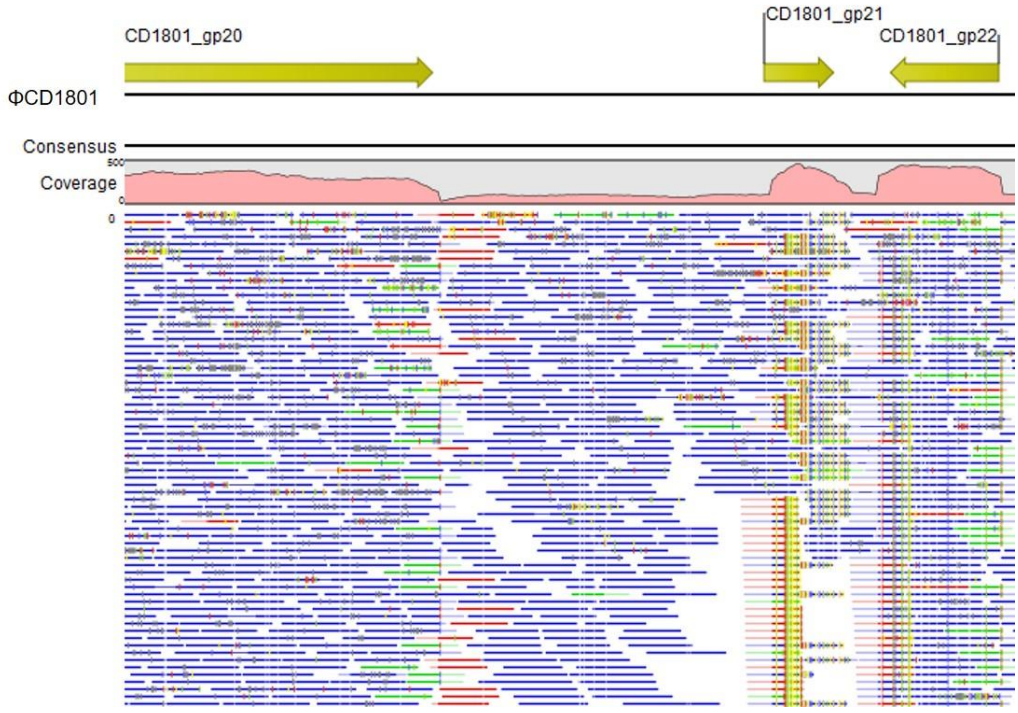


Figure 5.5: **Putative integration site for phage ΦCD1801 within the genome of *C. difficile* 1801.** This shows alignment of the phage integrase gene and the potential *attP* site immediately after this gene. The read coverage is comparatively low in this area in comparison to the integrase gene and other genes of the phage. Reads which contained phage DNA and bacterial DNA were identified (shown in green) and a putative integration site within the *CDM120_RS14955* gene in *C. difficile* 1801 was determined. This gene encodes a phosphonate ABC transporter ATP-binding protein.

The resulting read sequence was searched for in the bacterial genome to determine a putative integration site. It was determined this site was within gene *CDM120_RS14955* which encodes a phosphonate ABC transporter ATP-binding protein. The predicted CD1801L genome was created using Benchling, by inserting the phage ΦCD1801 genome at the predicted site of integration into the CD1801 genome. This genome was used as a reference to align the raw sequencing reads for a second time and showed that reads aligned across the predicted integration site, suggesting that this is the integration site. PCR primers were designed to amplify across this region, one primer within the phage

integrase gene and one within the bacterial chromosome. Two primer sets were designed to allow amplification of each end of the integrated phage and the bacterial DNA region adjacent to it. Primers RS14955_F1 and CD1801_*attP*_R1 were used to amplify one phage/bacterial DNA junction and a second primer set, RS14960_R1 and CD1801_*attP*_F1, was used to amplify the other junction. PCR was conducted using Q5® High-Fidelity DNA polymerase with annealing temperatures determined using the NEB Tm calculator. For both primer sets a PCR product of the expected size (431 bp) was produced (Figure 5.6). This was excised, purified and subjected to Sanger Sequencing (Figure 5.7). The PCR amplification and the sequencing confirmed the MiSeq data, that phage ΦCD1801 had integrated into the chromosome of *C. difficile* 1801 interrupting an ORF encoding a phosphonate ABC transporter ATP-binding protein. It should be noted that the exact *attP* and *attB* sites have not been determined by this method.

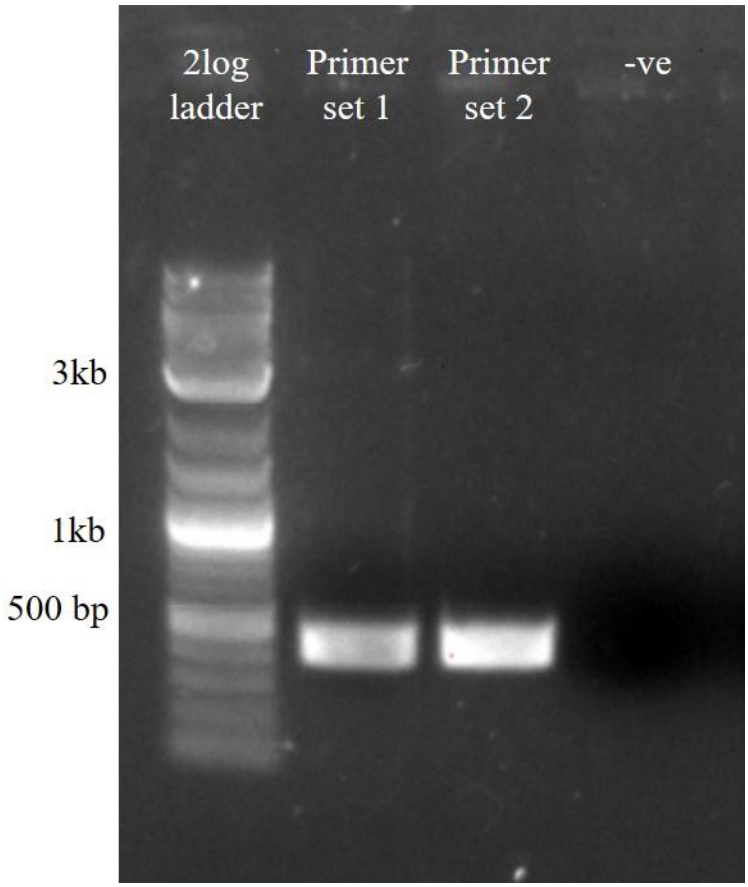


Figure 5.6: **PCR amplification of the phage bacterial junction.** The two primer sets were used to amplify a 431 bp sequence, with one primer designed in the phage genome and one designed in the bacterial chromosome. Successful amplification supports the location of the putative phage ΦCD1801 integration site in the chromosome of *C. difficile* 1801L. Lane 1: 2log ladder (0.1-10kb), lane 2: Amplification using primers RS14955_F1 and CD1801_*attP*_R1, lane 3: Amplification using primers RS14960_R1 and CD1801_*attP*_F1, lane 4: PCR negative control (no DNA).

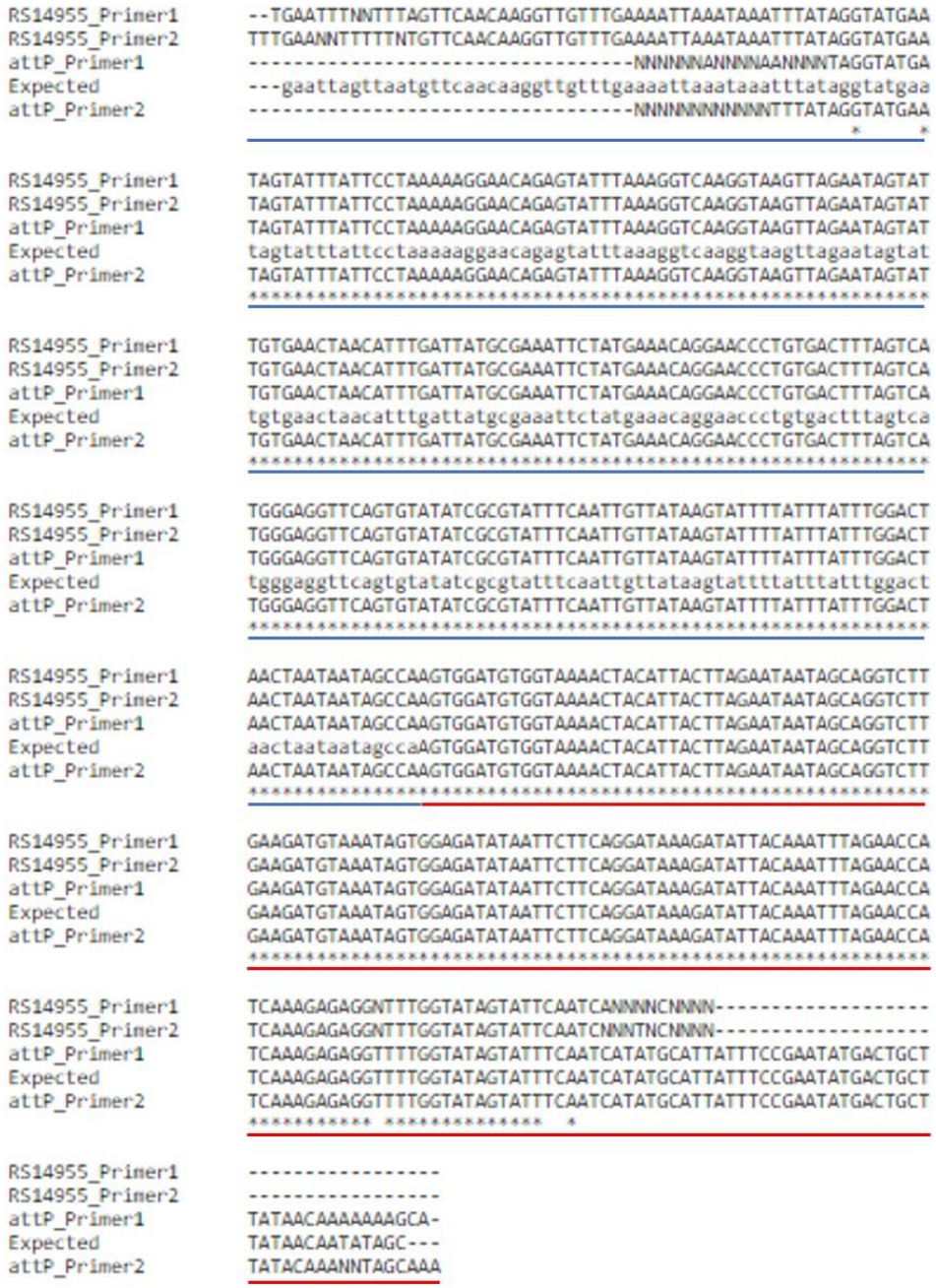


Figure 5.7: **Nucleotide alignment of putative phage integration site into the bacterial chromosome**. Sanger sequencing of the DNA amplified using primers flanking the putative site of phage ΦCD1801 into the chromosome of *C. difficile* 1801L. Alignment created using CLUSTAL Omega. Stars show where bases match. Phage DNA is represented in lower case letters and underlined in blue, bacterial DNA is represented in upper case letters and is underlined in red. The sequencing data matches the expected sequence and further confirms the putative phage integration site.

5.3.6 Integrase Gene Investigation

The presence of an integrase gene in a phage genome is characteristic of a temperate phage and is responsible for the integration of phage DNA into the host chromosome. An integrase is a site specific recombinase, of which there are two families which are recognised by their catalytic residue, either tyrosine or serine (Fogg *et al*., 2014). In order to assess the nature of the four integrases found in each of the phages a protein BLAST search was conducted to determine the different domains present in each protein (Figure 5.8). It was determined that the integrases present in phages ΦCD418, ΦCD2301 and ΦCD08011 belonged to the tyrosine integrase family, signified by the presence of protein domain similarity to XerC, a known tyrosine recombinase (Figure 5.8). A comparison of the amino acid sequences of the integrase genes from these three phages, using CLUSTAL Omega, showed that phages ΦCD2301 and ΦCD08011 contain almost identical integrase genes with only two amino acid changes, 95S>A and 123I>V, in phage ΦCD2301 in comparison to ΦCD08011 (Figure 5.9). Conversely, the integrase gene of ΦCD418 was highly dissimilar at the amino acid level to those of phages ΦCD08011 and ΦCD2301, despite also being identified as a tyrosine recombinase. Unlike the other three phages, phage ΦCD1801 contains a serine integrase determined by the presence of a protein domain that is predicted in serine recombinases. A zinc recombinase ribbon domain was also predicted which is commonly found in site-specific recombinases and is thought to be involved in DNA binding (Figure 5.8).

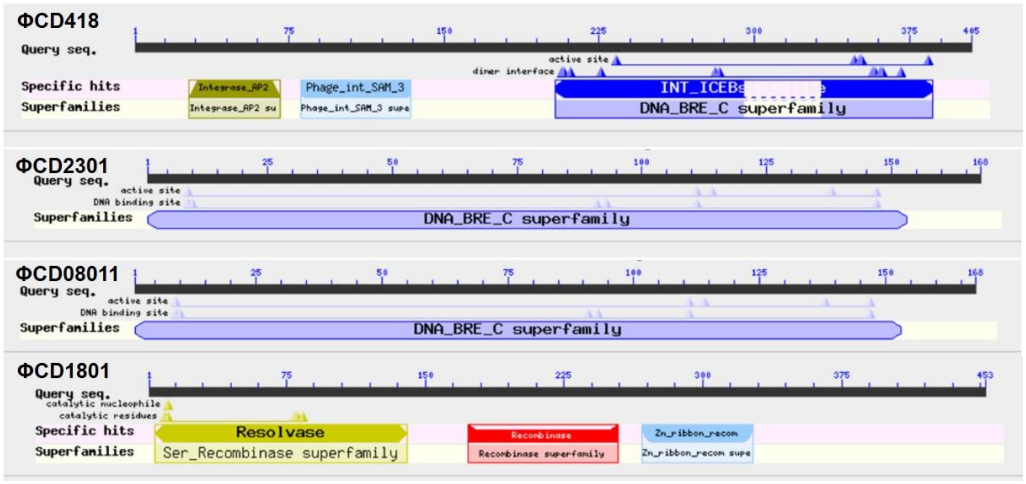


Figure 5.8: **Predicted conserved domains identified in the integrases of phages ΦCD418, ΦCD2301, ΦCD08011 and ΦCD1801.** The prediction of a conserved domain from XerC is consistent with tyrosine recombinases and is predicted in the integrases of phages ΦCD418, ΦCD2301 and ΦCD08011. The prediction of a serine recombinase conserved domain in the integrase of phage ΦCD1801 suggests this integrase belongs to the serine recombinase family.

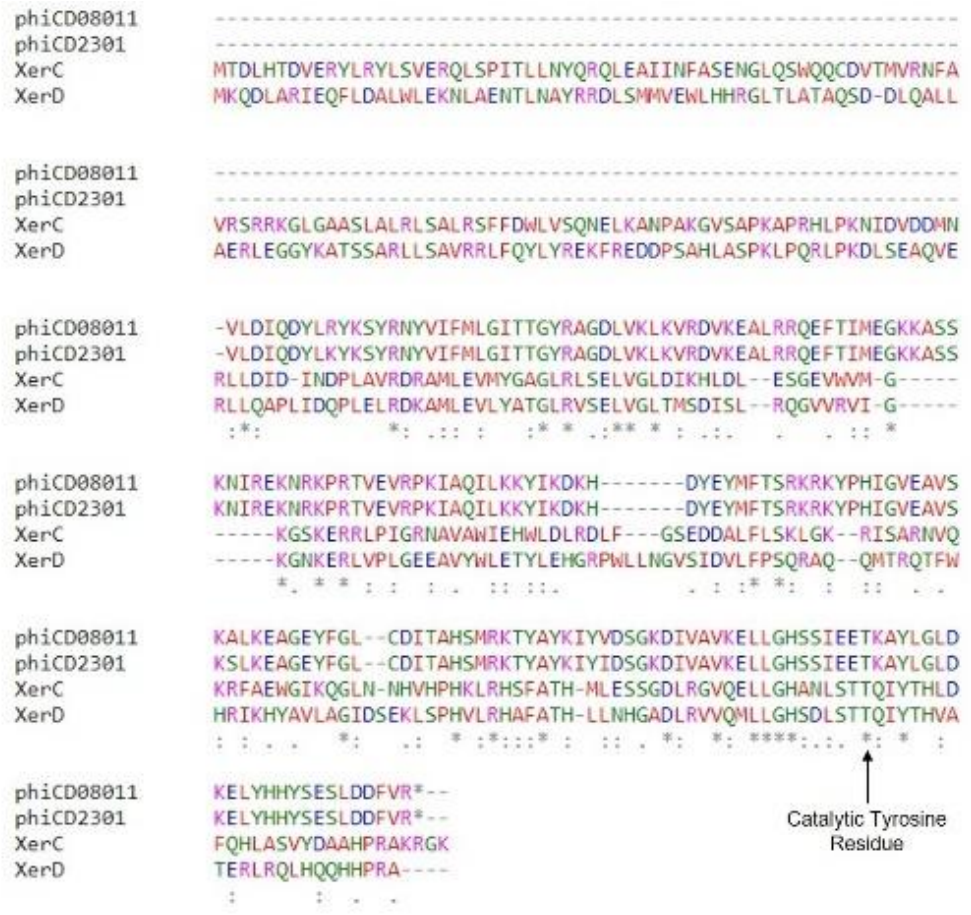


Figure 5.9: **CLUSTAL Omega alignment of the amino acid sequences from the integrase genes of phages ΦCD2301 and ΦCD08011 in comparison to other tyrosine recombinases.** Both ΦCD2301 and ΦCD08011 integrases belong to the tyrosine family of recombinases and are highly similar apart from two amino acid changes in ΦCD2301 integrase (95S>A and 123I>V) to that of the integrase from ΦCD08011. Shown here in comparison to the bacterial tyrosine recombinases XerC and XerD, the catalytically tyrosine residue can be identified. The integrases of phages ΦCD08011 and ΦCD2301 are shorter than the recombinases XerC and XerD and consist of only the catalytic domain of these.

The presence of a putative RBS and start codon, along with experimental data confirming the isolation of lysogens, indicates that the integrase gene in each of these four phages is active. Integrase genes have been identified in all the published *C. difficile* phage genomes however, whether these genes are active or

not is yet to be reported. Genome analysis of the published phage ΦCDKM9 suggests the presence of an integrase gene but a putative RBS site could not be identified (Figure 5.10), the presence of a putative promoter was identified using Soft Berry online promoter predictor. Phage ΦCDKM9 was selected for analysis as it is the most recently isolated *C. difficile* phage, isolated from an environmental sample, and its genome sequence is publicly available.

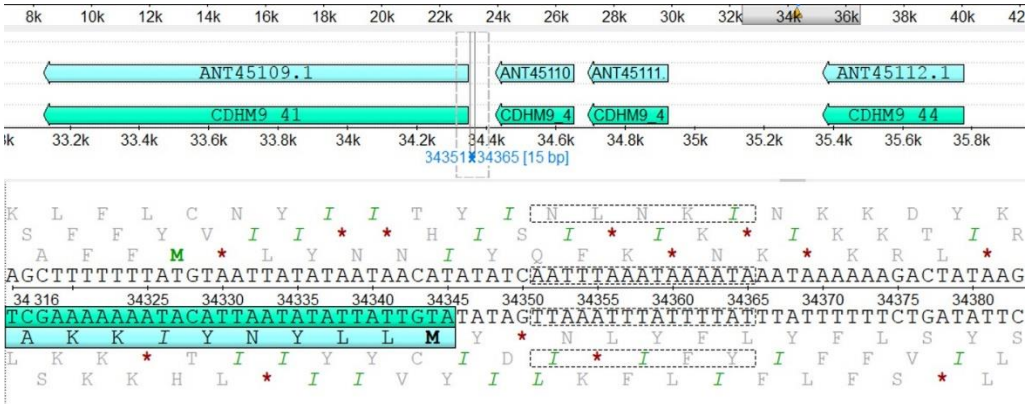


Figure 5.10: **Sequence upstream of the integrase gene from phage ΦCDKM9** (CDHM9_41) and the location of the expected RBS (dotted box), 5 to 15 nucleotides upstream of the start codon (ATG) for the gene. In this case no RBS, based on the broad consensus sequence AAGGAGG, can be identified in this region suggesting that the integrase gene of this phage is not being translated.

5.3.7 Phylogenomic Analysis

Phylogenomic trees were created, using the Virus Classification and Tree Building Online Resource, VICTOR, to compare the relatedness of the four phages, ΦCD08011, ΦCD2301, ΦCD418, and ΦCD1801, with the 25 published *C. difficile* phage genomes in the NCBI database. The *C. difficile* phage genomes used are listed in Table 1.1. The comparison was completed using the full nucleotide sequences of each phage. The phylogenomic GBDP trees were

inferred using three different formulas D0 (Figure 5.11) for full genomes at nucleotide level, D4 (Figure 5.12) for incomplete genomes and D6 (Figure 5.13) for full genomes at the amino acid level. An average support of 50 %, 29 % and 52 % was generated, respectively. OPTSIL clustering was used to create clusters at the species, genus and family level, using each of the three formulas, D0, D4 and D6. At the species level, 27, 26 and 24 clusters were created, respectively. At the genus level, 5, 12 and 4 clusters resulted, respectively and at the family level, 1, 7 and 1 clusters resulted, respectively. Phages ΦCD08011 and ΦCD2301 were most closely related to each other and to phage ΦMMP04, a small myovirus, using each of the three formulas. Phage ΦCD1801 clustered most closely with ΦCD119 and ΦCDHM19, both medium myoviruses, using each of the three formulas. Phage ΦCD418 clustered similarly using formulas D0 and D6, however slightly differently when the D4 formula was used. Using the D4 formula, it clustered with phages ΦC2, ΦMMP03, ΦCDKM15 and ΦCDMH1, although more distantly than these phages were related to each other. Using formulas D0 and D6, ΦCD418 clustered most closely with ΦCDKM15, ΦCD27, ΦCDKM9, ΦMMP02 and ΦCD505, a mixture of medium myoviruses and long tailed myoviruses. More generalised analysis of the phylogenomic trees show that phages belonging to *Siphoviridae* cluster together, and the two *Myoviridae* genus', *phiCD119likevirus* and *phiMMP04likevirus,* cluster with phages classified in their genus.

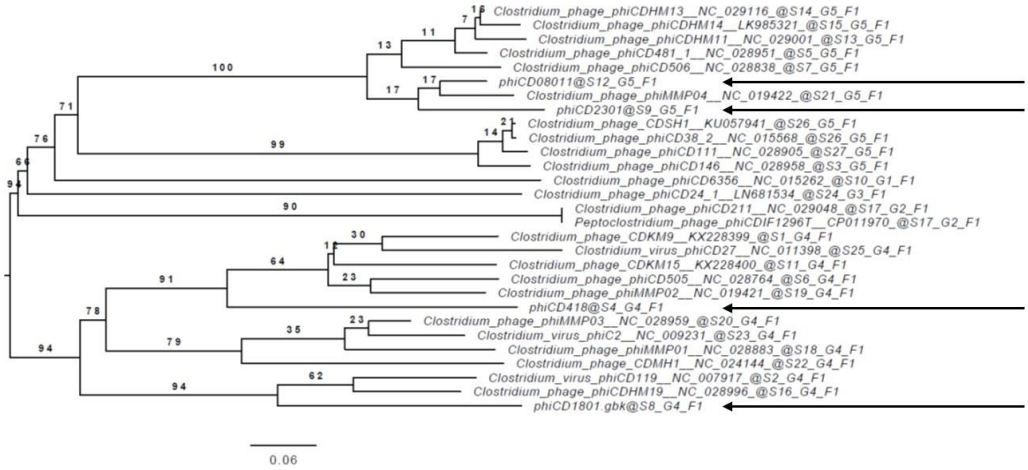


Figure 5.11: **Phylogenomic analysis of the four isolated phages created using VICTOR, D0 formula**. The D0 formula is used to compare full genomes at the nucleotide level. The four newly isolated phages are compared to 25 published phages, analysis compares the full nucleic acid sequences. Arrows show the position of the four isolated phages; ΦCD08011, ΦCD2301, ΦCD418 and ΦCD1801 from top.

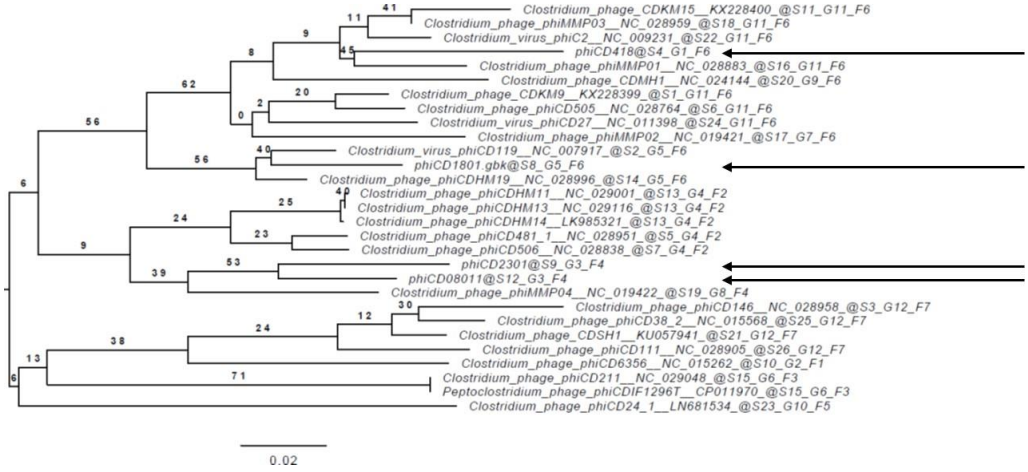


Figure 5.12: **Phylogenomic analysis of the four isolated phages created using VICTOR, D4 formula**. The D4 formula is optimal when any of the genomes being compared are incomplete. The four newly isolated phages are compared to 25 published phages, analysis compares the full nucleic acid sequences. Arrows show the position of the four isolated phages; ΦCD418, ΦCD1801, ΦCD2301 and ΦCD08011 from top.

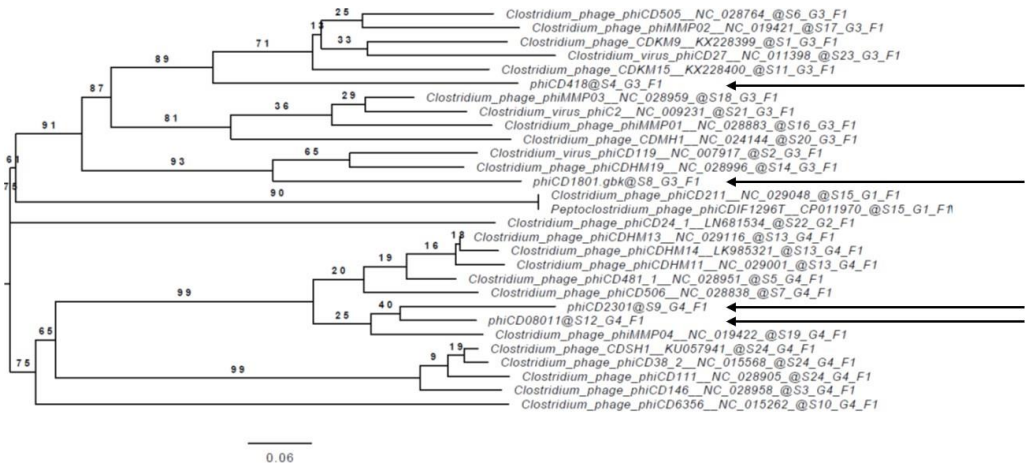


Figure 5.13: **Phylogenomic analysis of the four isolated phages created using VICTOR, D6 formula**. The D6 formula is used to compare full genomes at the amino acid level. The four newly isolated phages are compared to 25 published phages, analysis compares the full nucleic acid sequences. Arrows show the position of the four isolated phages; ΦCD418, ΦCD1801, ΦCD2301 and ΦCD08011 from top.

5.4 Discussion

Illumina MiSeq coverage was of sufficient quality, the average read length varied from between 146 and 234 bp, to allow the *de novo* assembly of each genome into a single contig, with no errors, using CLC Genomics Workbench 8.5.1. The use of PCR showed that three of phages had circularly permuted genomes, ΦCD08011, ΦCD418 and ΦCD1801, and one, ΦCD2301, had a genome with physical ends suggesting that this phage, in comparison to the other three phages has a different DNA packaging mechanism. Sanger sequencing data suggested that no bases were added in the formation of the circular genome and therefore, assembly of the genome had not missed any bases. ORF's were predicted by Artemis but trimmed manually to the predicted correct start codon by the identification of potential RBS and promoter sequences upstream of each gene. This was useful as it ensured that every gene was positioned with a RBS

and therefore are likely to be transcribed and translated, suggesting the protein products will be functional. The genomes were manually annotated as the use of computerised systems for automated annotation can be limited by published annotations. Manually, three different programmes were used for the prediction of putative gene products and allowed appropriate selection of the putative product by comparison of the top hits in each programme, potentially this could mean that the annotations are more extensive than an automated annotation pipeline may provide.

The presence of a putative integrase gene within the genome of each of the four phages indicates that each phage is temperate, which is further confirmed by the production of stable lysogens (3.2.5). The presence of putative RBS and promoter sequences upstream of the four integrase genes suggests that the genes will be transcribed and translated and the experimental data showing the creation of lysogens for each phage shows that the integrases are active. Currently, the genome of every published *C. difficile* phage is known to contain an integrase gene but it is not clear whether these are all expressed and active. The published phage ΦCDKM9 was selected for investigation into integrase genes in other *C. difficile* phages with surprising results. The annotated integrase gene (CDHM9_41) appeared to contain a putative promoter sequence, determined using the Soft Berry promoter predicter online tool however, a putative RBS could not be identified. This would signify that this integrase gene is transcribed but cannot be translated, suggesting that this gene is not active. The RBS prediction software utilised by Rashid and colleagues for the annotation and publication of this phage genome, indicated that a putative RBS was present but

only within the integrase gene and one was not identified upstream of the gene (Rashid *et al*., 2016 supplementary data). The presence of a RBS within the gene could suggest that a truncated version of the protein is produced which may still be active. Further analysis of the integrase from this phage would be warranted to determine its functionality however, initial results indicate that this integrase gene is not functional and could potentially suggest that this phage is truly lytic.

Phages ΦCD08011 and ΦCD2301 cluster with ΦMMP04, suggesting they are small myoviruses belonging to the *phiMMP04likevirus* genera. TEM measurements and the presence of genes encoding a Clp protease and a ParA homologue within their genomes further confirms this. TEM analysis of ΦCD418 phage particles suggested that the phage could be either a long tailed myovirus or belong to *Siphoviridae.* The presence of a gene encoding a tail sheath is indicative of *Myoviridae* and suggests that phage ΦCD418 belongs to this family. This is further confirmed by the presence of a cytosine-5 methyltransferase gene and the identification of a three-gene cassette encoding DnaD, a resolvase and a single stranded DNA binding protein. These genes are characteristic features of the genus *phiCD119likevirus* therefore, suggests phage ΦCD418 could be classified into this genus. However, this conclusion was not supported by phylogenomic analysis where it did not cluster with other *phiCD119likevirus* phages but was found to be more closely related to the long tailed myovirus phage ΦCD27. The identity of phage ΦCD418 as a long tailed myovirus was supported by TEM analysis.

Phage ΦCD1801 clusters with ΦCD119, however, it does not contain genes within its genome that are indicative of belonging to the *phiCD119likevirus* genus, nor does it contain any of the genes indicative of belonging to *phiMMP04likevirus* genera. It is likely that the diversity of *C. difficile* phages is so great that the publication of further phages could lead to the creation of a new genera to which phage ΦCD1801 may belong. Phylogenomic analysis of the four phages in comparison to 25 published phage genome sequences was conducted using the full nucleotide sequence of each phage which has previously been used for classification of *C. difficile* phages (Rashid *et al*., 2016), however, comparison between the terminase gene, endolysin gene and the gene encoding the major capsid protein could have been used for phylogenetic analysis (Hargreaves and Clokie, 2015).

Within the genome of phage ΦCD2301, two genes which together encode an abortive infection (Abi) protein have been identified. Abi systems are encoded by a single gene, which are usually found on plasmids, and are a form of bacterial resistance to phage infection (Stern and Sorek, 2012). The term Abi-mediated resistance is a general term for host mechanisms that interrupt phage development through disruption of transcription, replication or packaging (Chopin *et al*., 2005). This form of resistance is usually defined by programmed cell death, or cell "suicide" of the host bacterial cell. Although this gives no survival advantage to the infected bacterial cell, it does contain the phage infection and protects the rest of the bacterial population by preventing the spread of infectious phage particles (Stern and Sorek, 2012). Genes encoding Abi-like proteins have been identified in the genome of the *C. difficile* phage

ΦC2 and in phages of *Lactococcus lactis* (Stern and Sorek, 2012). In phage ΦCD2301, there are two ORFs encoding an Abi-like protein which is unusual as Abi systems are usually encoded by a single gene. In phage ΦC2, an *abiF* gene has been identified (Goh *et al*., 2007). This gene encodes AbiF which has been shown to interfere with phage DNA replication (Garvey *et al*., 1995). EMBOSS pairwise sequence alignment was used to compare the full length of the two genes and the non-coding DNA between them from phage ΦCD2301 with the *abiF* gene from phage ΦC2, showing that they were completely identical, at the nucleotide level, except for two nucleotide variants in ΦCD2301. One ORF from ΦCD2301 had similarity to the 5' end of *abiF* and the second ORF had similarity to the 3' end of *abiF*. The two nucleotide variants in ΦCD2301 create two stop codons, which is why the gene appears as two separate ORFs in this phage genome. These SNP's have not been verified by Sanger Sequencing and could be an artefact from the genome assembly and they would need to be confirmed to determine whether the Abi system in phage ΦCD2301 is active. Interestingly, the genes encoding the Abi-like proteins in phage ΦCD2301 and ΦC2 are located upstream of a gene encoding an endolysin protein and downstream of a gene encoding a holin protein. These two proteins are involved in cell lysis and suggests that the genes encoding the Abi-like proteins are located within the cell lysis module and could, therefore, be involved in cell lysis. The same gene, *abiF*, was also observed in the bacterial genomes of *C. difficile* R20291 (Accession No. FN545816) and DH196 (Accession No. FN538970) and are again located upstream of a gene encoding an endolysin and downstream of a gene encoding a holin. It seems that in *C. difficile* phages the location of the Abi protein encoding gene is different to that of phages of *L. lactis*, where they are located

within the lysogenic-conversion region of the phage genome. It is thought this gene is transcribed when the phages are in the lysogenic state and is, therefore, suspected to be involved in protecting against superinfection by disrupting phage multiplication of the incoming phage (Ventura *et al*., 2007).

The gene encoding the putative cI-like repressor protein in phage ΦCD1801 encodes a protein, 152 residues in length, and was first identified using protein BLAST to search for conserved domains. This revealed the presence of a helix-turn-helix domain which is commonly found in CI and Cro proteins. This conserved domain was located between residues 30 and 50 which is the same position as the same domain in the cI repressor protein of bacteriophage λ. The helix-turn-helix domain is capable of binding DNA and is usually located in proteins involved in gene expression, however the DNA binding ability of the protein in ΦCD1801 has not been experimentally proven. EMBOSS pairwise sequence alignment was conducted to compare the amino acid sequence of the gene encoding the putative cI-like repressor from phage ΦCD1801 with that of the published phage ΦCD27 (Accession No. NC_011298.1). This showed that the two genes were identical at the amino acid level and further confirmed the annotation of this gene as encoding the putative cI-like repressor. Finally, the gene encoding the putative cI-like repressor is located upstream of a gene annotated as encoding an anti-repressor which is orientated in the opposite direction. This arrangement is also seen in the genome of phage ΦCD27, suggesting the anti-repressor is a cro-like protein. The cI-like repressor and cro-like protein are involved in the genetic switch controlling the lysogeny-lysis decision in phage λ (refer to 1.3.3 for more detailed information on the system).

Although other mechanisms are known to exist in this circumstance it appears that the same mechanism as in phage λ is likely.

The identification of the cI-like repressor protein in phage ΦCD1801 is important in studying phage resistant and sensitive strains as it is involved in phage resistance to superinfection. Whilst in the prophage state the only gene to be expressed is *cI*, the product of which binds to the two operators, O_R1 and O_R2, and prevents the transcription of *cro* and the late phage genes, whilst simultaneously activating its own expression to maintain the latent state. Any cI-like repressor that is not bound to the operators is free to bind to the operators of incoming phage, therefore the transcription of *cro* and the gene encoding the integrase gene cannot occur. This results in the incoming phage being unable to integrate into the host chromosome or undertake the lytic cycle. This is referred to as lysogenic immunity and although phage binding to bacterial cells could still be observed, no plaques would be created as the lytic cycle cannot be conducted.

A putative integration location for phage ΦCD1801 into the chromosome of *C. difficile* 1801 was determined using genome sequencing however, the exact *attB* and *attP* sites were not precisely defined. It was suggested that the integration site, *attB,* could be within a gene encoding a phosphonate ABC transporter ATP-binding protein. The prophage integration location was confirmed by PCR amplification of the bacteria/phage DNA junctions and the subsequent Sanger Sequencing of the PCR product. The integration location within the bacterial chromosome of some *C. difficile* phages has already been confirmed, with phage ΦC2 integrating into *gntR*, in the chromosome of

C. difficile CD242 lysogen, whose product regulates the pathway of D-gluconate degradation (Goh *et al*., 2007). Phage ΦCD119 has been shown to integrate into a non-coding region of DNA in the *C. difficile* ΦCD119 F10 lysogen. The non-coding DNA lies downstream of a gene encoding a hypothetical protein and upstream of *gltP*, whose product is part of a proton/glutamate-aspartate symporter (Govind *et al*., 2006). Finally, two integration sites of phage ΦCD27 have been identified, one into a gene encoding a putative ATPase of the flagella export apparatus and the other a putative ATPase of an ABC transporter (Williams *et al*., 2013). It appears that *C. difficile* phages preferentially integrate into genes encoding ABC transporter proteins, and it is hypothesised that these proteins have homologs within the *C. difficile* genome which means that integration does not have a negative effect on the cell. Due to resource constraints the putative integration location could only be determined for one of the phages isolated in this study.

5.5 Chapter Summary

The main outcomes of this chapter are:

- The genome sequences of four phages, ΦCD08011, ΦCD2301, ΦCD418 and ΦCD1801, have been determined and annotated
- A gene encoding an integrase has been identified in each genome and confirmed the temperate nature of the four phages
- A gene encoding a cI-like repressor protein in phage ΦCD1801 has been identified
- The identification of genes within the phage genomes which aid in their classification; ΦCD08011 and ΦCD2301 as *phiMMP04likevirus*, ΦCD418 as a long tailed myovirus and ΦCD1801 as not belonging to any of the previously described myovirus genera
- A putative integration site of phage ΦCD1801 into the chromosome of *C. difficile* lysogen 1801L has been identified

Chapter Six:

Identification of the Bacterial Surface Receptor for Phage ΦCD1801

6 Identification of the Bacterial Surface Receptor for Phage ΦCD1801

6.1 Introduction

C. difficile cells are completely coated in a paracrystalline protein surface layer called the S layer which is encoded by a single gene, *slpA* (Calabi *et al*., 2001). The product, SlpA, is post-translationally cleaved to form a high molecular weight (HMW) and low molecular weight (LMW) form. The HMW form is highly conserved between *C. difficile* strains and the LMW form, which is externally located on the cell wall, is more variable and the main serotyping antigen (Calabi *et al*., 2001, Karjalainen *et al*., 2001). The *slpA* gene is located within a 9.7 kb S layer cassette which also contains *secA2*. Previously 13 different S layer cassettes have been identified in *C. difficile* (Dingle *et al*., 2013) including a hybrid cassette found in PCR ribotype 078 strains. The hybrid cassette, H2/6, comprises the *slpA* and *secA2* genes from S layer cassette 6 and the *cwp66* gene from S layer cassette 2 (Dingle *et al*., 2013). Although, the receptors for phages infecting *C. difficile* have yet to be identified the S layer has been proposed as the most likely candidate (Kirk *et al*., 2016). In 2017 a 14th S layer cassette type was identified (Kirk *et al*., 2017).

A database, PheReD, is available for all known phage receptors and catalogues the identity of known phage receptors in Gram-positive and Gram-negative bacteria (Bertozzi *et al*., 2016). In *Bacillus anthracis* and *Bacillus subtilis*, the cell wall proteins GamR (Davison *et al*., 2005) and YueB (São-José *et al*., 2004)

respectively have been shown to be the phage receptors. In *Lactococcus lactis* the rhamnose moieties in teichoic acids in the cell wall have been demonstrated to act as phage receptors (Monteville *et al*., 1994). In the Gram-negative bacteria, *Caulobacter crescentus,* it was determined that the S layer was required for infection of phage ΦCr30 (Edwards and Smit, 1991). In this organism, the S layer is encoded by a single gene, *rsaA*, and its inactivation through the insertion of a DNA cassette encoding an antibiotic resistance led to mutant cells that were no longer sensitive to phage infection (Edwards and Smit, 1991).

The identification of phage receptors in *C. difficile* is yet to be reported in the literature. Progress has been somewhat hampered by the presence of multiple phage resistance systems and the fact that phages are temperate in nature. Phage receptors are most easily identified through the isolation of variants of the susceptible indicator strain that have become resistant due to mutations in the receptor. Such a screening procedure is difficult to apply with temperate phages as such rare mutants are obscured by the large number of lysogens that arise and which are also resistant. Large scale screening for lysogens, potentially using PCR, would be required to eliminate these cells to allow truly resistant mutants to be selected. CRISPR-Cas systems are also common in *C. difficile* which would also hamper this search due to the presence of intrinsically resistant cells emerging in the population and, again, obscuring the rare truly resistant mutants (Hargreaves *et al*., 2013 and 2014).

In this chapter two *C. difficile* genetic tools will be utilised, ClosTron mutagenesis and the use of a tetracycline inducible gene promoter. The ClosTron

utilises insertional mutagenesis for the rapid production of gene knockouts (Heap *et al*., 2007), however, polar effects on downstream genes can sometimes be an issue. This usually results from the promoter that drives the antibiotic resistance gene within the intron, as this could potentially lead to differential expression of genes downstream of the promoter. The ClosTron utilises the Group II intron from *Lactococcus lactis* Ll.LtrB, as Ll-LtrB has the ability to lose the intron-encoded protein LtrA (Karberg *et al*., 2001). The Group II introns can be directed by retargeted sequences to ensure they insert into the gene of interest within the genome and limit insertion elsewhere. LtrA is encoded on the ClosTron plasmid which means it can be lost from the cells after insertion within the genome, ensuring the isolation of stable mutants (Heap *et al*., 2007). Insertional mutants can be selected on the basis of acquisition of erythromycin resistance which is activated upon integration into the chromosome using a retrotransposition-activated marker (RAM).

The use of the tetracycline system to control gene expression has been widely studied and is based around the *tetR* and *tetA* genes which are divergently orientated to one another (Orth *et al*., 2000). The product of *tetR,* TetR, can bind independently to two operators, $tetO_1$ and $tetO_2$, to inhibit *tetR* and *tetA* gene expression or to repress gene expression of *tetA* only, by inhibiting binding of RNA polymerase (Orth *et al*., 2000). When in the presence of tetracycline, a conformational change of TetR is caused by the formation of a magnesium and tetracycline complex. This conformational change reduces the binding affinity of TetR to $tetO_1$ and $tetO_2$ by nine-fold, allowing RNA polymerase to bind and expression of *tetR* and *tetA* (Orth *et al*., 1998). The product of *tetA,* TetA, is a

membrane spanning tetracycline antiporter which pumps the tetracycline back out of the cell, thus when TetA is functional the cell is resistant to the tetracycline antibiotic (Bertram and Hillen, 2008). The tetracycline inducible gene expression system has been proven in many Gram-negative and Gram-positive bacteria and its low basal rate expression and high induction efficiency makes it a very desirable tool for gene expression (Bertram and Hillen, 2008).

In 2011, this tool was adapted for use in *C. difficile* (Fagan and Fairweather, 2011) in order to study its accessory secretory systems. The tetracycline inducible promoter, P_{tet}, from *Staphylococcus aureus* (Corrigan and Foster, 2009) was utilised to drive the expression of codon optimised *gusA* and the P_{tetR} promoter used to drive expression of *tetR.* The basis of the system is to use two divergent promoters which have overlapping operator sequences whose expression is controlled by the binding of TetR (Orth *et al*., 2000). The *gusA* gene encodes β-glucuronidase, allowing quantification of the level of gene expression using an enzyme assay. In this system, the addition of tetracycline removes TetR repression to allow increased transcription of *tetR* and *gusA.* This work was carried out in *C. difficile* 630 which is tetracycline resistant due to the production of TetM, encoded by the *tetM* gene in its genome (Sebaihia *et al*., 2006). Although there is a lag phase between exposure to tetracycline and complete expression of *tetM,* a non-antibiotic analogue anhydrotetracycline (aTc), was utilised to minimise any growth effects in strains that are tetracycline sensitive (Ehrt *et al*., 2005). In the absence of aTc no detectable expression of *gusA* was observed, whereas induction using 500 ng/ml aTc resulted in expression that matched expression using a P_{cwp2} constitutive promoter (Fagan

and Fairweather, 2011). This tool will be exploited in this chapter to investigate the *C. difficile* phage receptor.

Phage ΦCD1801 was selected for the study of phage receptors as it has the broadest infection spectrum of the four phages isolated, infecting all but one of the *C. difficile* PCR ribotype 078 isolates tested.

6.2 Development of Methodology

6.2.1 SNP Analysis of C. difficile *PCR Ribotype 078 Sensitive and Resistant Strains*

Of the tested PCR ribotype 078 strains, only two (CD2315 and CD31662) were resistant to infection by phage ΦCD1801. One possible explanation for their insensitivity was that the ΦCD1801 receptor had mutated such that the phage could no longer bind. To explore this possibility the genomes sequences of these two *C. difficile* strains were compared to those of three sensitive strains (CD1801, CD2016 and CD7009825).

To derive their sequences, genomic DNA was prepared and subjected to Illumina paired-end sequencing on a MiSeq platform. The reads obtained were mapped to the available *C. difficile* M120 (Accession No. NC_017174), the reference *C. difficile* PCR ribotype 078 genome, using CLC Genomics Workbench 8.5.1. This allowed the derivation of a consensus sequence for each strain. Annotation of the derived sequences was accomplished by the direct transposition of the reference genome annotation data. Thereafter, a SNP analysis in comparison to the reference genome was conducted, using the CLC Genomics Workbench

8.5.1 fixed ploidy variant analysis function with a ploidy of 1, a required variant probability of 90 % and default parameters where the variant minimal frequency is 20 %. A second SNP analysis was conducted, using the same conditions, to compare each resistant strain to the three sensitive strains. This was completed by using the extracted consensus sequence for each resistant strain as a reference genome for assembly of the sensitive strains. Manual interpretation of the SNP analysis for each resistant strain in comparison to each sensitive strain was used to identify any potential phage receptor proteins, where a SNP was present in the resistant strains when compared to the sensitive strains. Potential receptor candidates were determined as any protein that could be on the cell surface, for example membrane proteins or cell wall proteins, however this was limited by the accuracy of the genome annotation.

6.2.2 Creation of Plasmid Containing CDM120_RS14800 *gene from* C. difficile *1801*

From the comparative sequencing analysis, a gene encoding a protein containing an amino acid change in the resistant strain was identified, *CDM120_RS14800*, annotated as a putative membrane protein (6.3.1). The gene was amplified from the genome of phage sensitive *C. difficile* 1801, to include the potential native promoter, using the CD1801PMP_F_*Sac*I and CD1801PMP_R_*Xho*I primers, Q5® High-Fidelity DNA Polymerase and an annealing temperature of 68 °C. The primers were designed to include recognition sites for the restriction enzymes, *Sac*I and *Xho*I. The modular plasmid, pMTL84151, was selected as the cloning vehicle due to its ability to replicate in both *E. coli* and *C. difficile* and to remain stable within *C. difficile* over many generations (Heap *et al*., 2009). It

comprises the ColE1+*tra* high copy number Gram-negative replicon, the Gram-positive replication region of *C. difficile* CD6 (*orfB* and *repA*) and a *catP* antibiotic resistance marker. The vector was linearised by cleavage with *Sac*I and *Xho*I and the large DNA backbone fragment gel purified. PCR product encompassing the CD1801 gene *CDM120_RS14800* was digested with the same enzymes and gel purified. The plasmid backbone and the PCR product were ligated together to create plasmid pMTLMW1 (Figure 6.1) and propagated in *E. coli* Top10. The plasmid was transformed into the conjugal donor strain *E. coli* CA434 and the resultant transformant independently conjugated with the two recipient strains *C. difficile* CD2315 and CD31662.

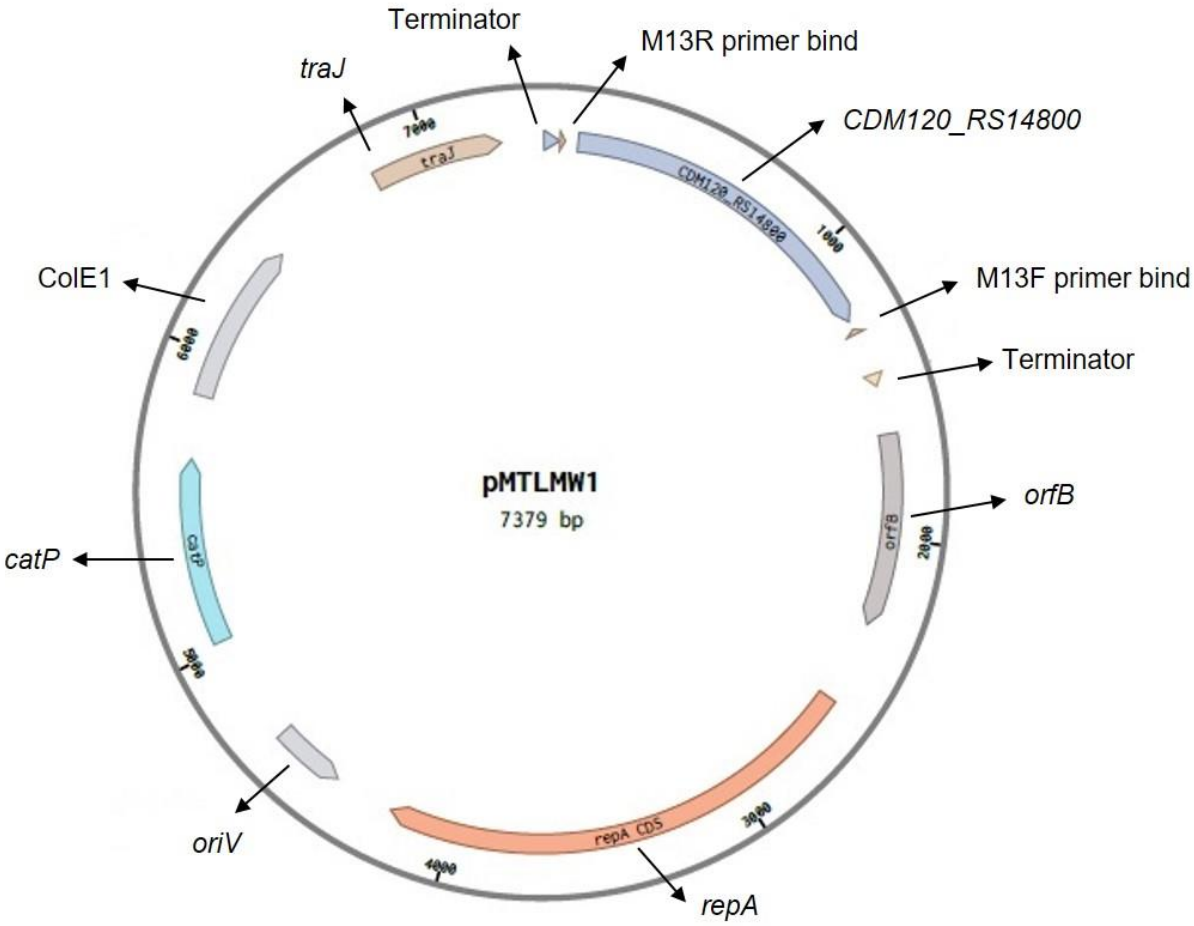


Figure 6.1: **Plasmid pMTLMW1**. The plasmid is based on pMTL84151, containing ColE1 + *tra* Gram-negative replicon, CD6 (*orfB* and *repA*) Gram-positive replicon, *catP* antibiotic resistance marker and *CDM120_RS14800* gene from *C. difficile* 1801, cloned within the multiple cloning site using *Sac*I and *Xho*I sites.

6.2.3 Creation of an Overexpression Plasmid Containing M120_RS14800 *gene from* C. difficile *1801*

The *CDM120_RS14800* gene was cloned under the control of the promoter (P_{fdx}) of the *C. sporogenes* ferredoxin gene (*fdx*), including its RBS. This is a strong promoter and, therefore, should lead to overexpression of the *M120_RS14800* gene. The *CDM120_RS14800* gene was amplified from *C. difficile* 1801, without its native promoter using primers CD1801PMP_F_*Sac*I and CD1801PMP_R_*Nde*I, Q5® High-Fidelity DNA Polymerase and an annealing temperature of 61 °C. The two primers introduced a *Sac*I restriction recognition site at the 3' end of the gene and a *Nde*I site at the start of the gene, respectively. In the latter case, the ATG of the CATATG restriction recognition site equated to the ATG start codon of the gene. Following digestion of the PCR amplified DNA with *Nde*I and *Sac*I, the cleaved fragment was cloned between the equivalent sites of the cloning vector pMTL84153. This plasmid contains the P_{fdx} promoter, the CD6 Gram-positive replication region (*orfB* and *repA*), the ColE1+*tra* Gram-negative replicon and a *catP* antibiotic resistance marker. The recombinant plasmid obtained was designated, pMTLMW3 (Figure 6.2). The presence of the *CDM120_RS14800* gene from *C. difficile* 1801 in plasmid pMTLMW3 was confirmed using Sanger sequencing with primers M13F and M13R. Growth of *E. coli* Top10 carrying this plasmid was undertaken at 30 °C, as growth at 37 °C was unsuccessful. It was postulated that the overexpression of the *CDM120_RS14800* gene was lethal to *E. coli* cells when they were fast growing at the higher temperature. The reduction in growth temperature overcame this limitation. Attempts to transform pMTLMW3 into *E. coli* CA434, however, met with no success, even when cells were incubated at room

temperature or 30 °C. One explanation was that over production of the cell membrane protein, encoded by *CDM120_RS14800*, a consequence of high copy number on the plasmid was detrimental to the cell. To overcome this, the copy number of the vector was reduced by replacing the ColE1 (around 600 copies per cell, Chambers *et al*., 1988) with that of a lower copy number (around 20 copies per cell, Chang and Cohen, 1978) p15a. The Gram-negative replicon of pMTL82123 (carries the p15a replicon) was excised, following digestion with *Sbf*1 and *Pfm*1, and the released DNA fragment gel purified. The ColE1+*tra* replicon was removed from the pMTLMW3 plasmid using the same restriction sites and the new Gram-negative replicon was cloned in its place, creating plasmid pMTLMW4 (Figure 6.3). The authenticity of the plasmid was confirmed by PCR amplifying the Gram-negative replicon region, using primers p15a_R1 and pCD6_F1, and subjecting it to Sanger Sequencing. The plasmid was propagated in *E. coli* Top10 overnight at 30 °C, extracted and transformed into *E. coli* CA434 and *E. coli* NEB SExpress (NEB Express carrying the conjugative R702 R-factor, C. Woods, thesis, University of Nottingham). The plasmid was transformed into *E. coli* NEB SExpress because it has been shown to improve conjugation efficiency with *C. difficile* DH1916 (P. Ingle, thesis, University of Nottingham), as the *E. coli* strain has a dcm- and dam+ methylation profile, this means it can evade any type IV restriction modification systems present in the host. *E. coli* CA434 was used as the donor strain for *C. difficile* CD31662. The strains obtained were then used as donor strains in conjugation with the recipient strains *C. difficile* CD31662 and CDDH1916.

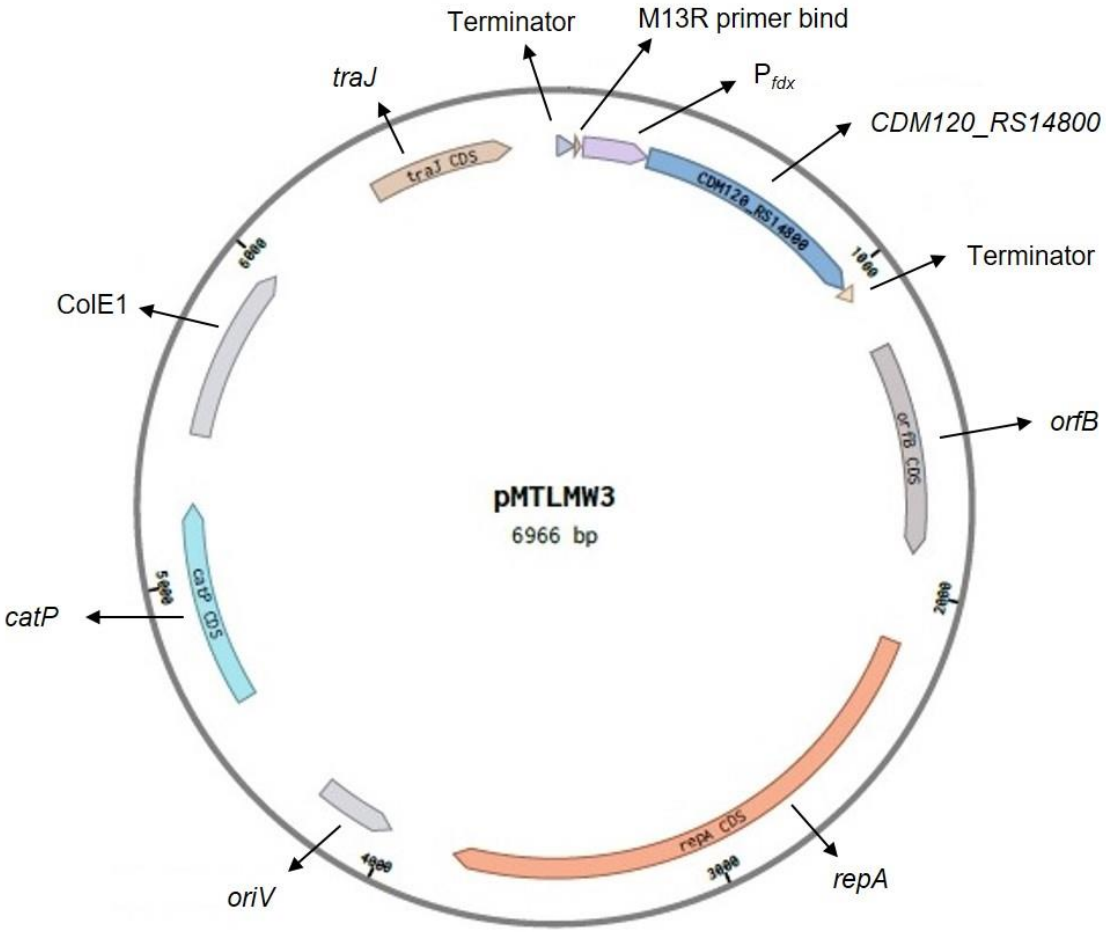


Figure 6.2: **Plasmid pMTLMW3.** This vector is based on the modular plasmid pMTL84153, containing ColE1 + *tra* Gram-negative replicon, CD6 (*orfB* and *repA*) Gram-positive replicon, *catP* antibiotic resistance marker *CDM120_RS14800* gene from *C. difficile* 1801 under control of P_{fdx} of the *fdx* gene from *C. sporogenes*, cloned within the multiple cloning site using *Sac*I and *Nde*I sites.

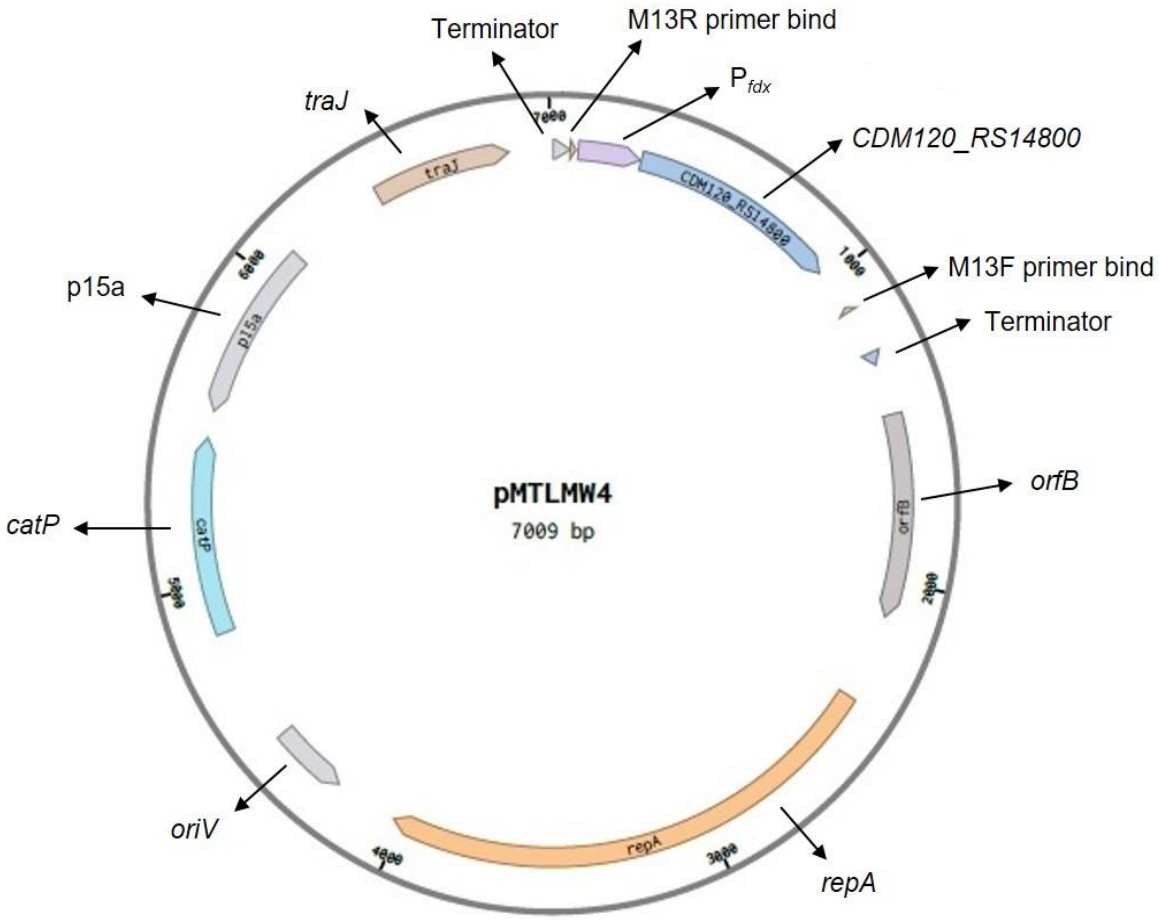


Figure 6.3: **Plasmid pMTLMW4**. This was created using pMTLMW3 with the replacement of a high copy Gram-negative replicon for a low copy one, p15a + *tra*, also containing the CD6 (*orfB* and *repA*) Gram-positive replicon, *catP* antibiotic resistance marker and *CM120_RS14800* gene from *C. difficile* 1801 under control of P_{fdx} of the *fdx* gene from *C. sporogenes*, cloned within the multiple cloning site using *Sac*I and *Nde*I sites.

6.2.4 SlpA Plasmids

Plasmids containing the *slpA* gene from three different S layer cassettes were obtained from Rob Fagan at the University of Sheffield. Plasmid pJAK002 contained the *slpA* from S layer cassette H2/6 (Figure 6.4), commonly found in *C. difficile* PCR ribotype 078 strains. Plasmid pJAK018 contained the *slpA* from S layer cassette 6 (Figure 6.5) and plasmid pJAK023 contained the *slpA* from S

layer cassette 2 (Figure 6.6). The three plasmids were propagated in *E. coli* Top10 and CA434, with growth at room temperature or 30 °C. Noticeably, however, compared to *E. coli* CA434, plasmid preparations of the *E. coli* Top10 clones resulted in comparatively low yields of plasmid DNA. Attempts to transfer the plasmids from *E. coli* CA434 into *C. difficile* 630Δ*erm* by conjugation were, however, unsuccessful.

To investigate the low yield of plasmid from *E. coli* Top10 and the failure to demonstrate conjugative plasmid transfer from *E. coli* CA434, a number of tests were performed to authenticate the functional integrity of the plasmids. Thus, the regions of the plasmids encompassing the Gram-negative and Gram-positive replicons were PCR amplified both before and after transformation into *E. coli* Top10. After the plasmid had been transformed into *E. coli* Top10, the CD6 Gram-positive replicon (*orfB* and *repA*) could no longer be amplified. The Gram-negative replicon was amplified in both cases and, following Sanger Sequencing, showed no SNPs in comparison to the expected sequence. Therefore, the plasmids which had originally been obtained had lost the Gram-positive replicon when they were introduced into *E. coli* Top10. To overcome this issue, three *C. difficile* 630Δ*erm* strains, each containing one of the three plasmids, were obtained from Rob Fagan at the University of Sheffield. The plasmids were extracted from *C. difficile* using three rounds of freeze-thaw prior to the miniprep kit and chemically transformed into *E. coli* DH5α and *E. coli* NEB SExpress (NEB Express carrying R702 R-factor), with growth at 37 °C. Subsequently, the plasmids were successfully conjugated into *C. difficile* DH1916 and CD31662.

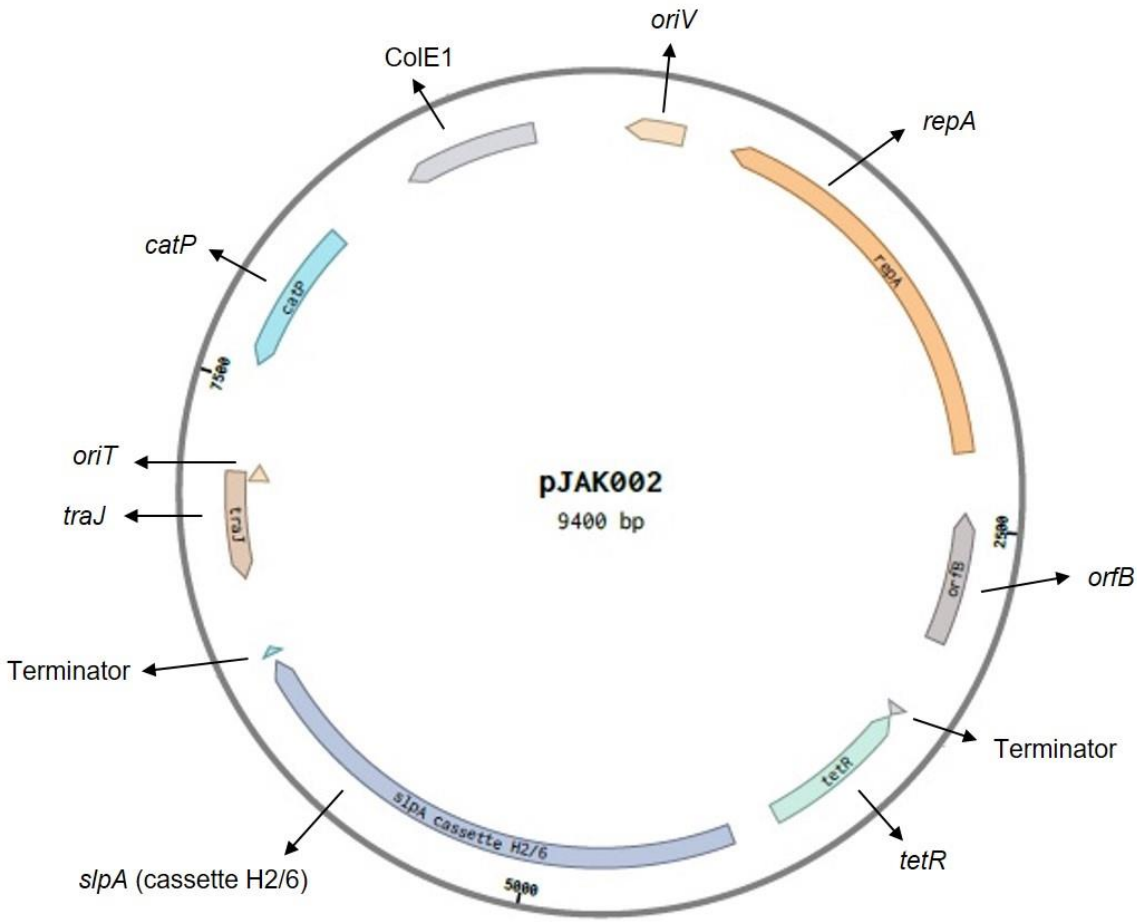


Figure 6.4: **Plasmid pJAK002**. The *slpA* gene from S layer cassette H2/6 has been cloned between *Sac*I and *Bam*HI sites into a plasmid based on pRPF185, *slpA* is under control of the tetracycline inducible promoter and is induced using 500 ng/ml anhydrotetracyline. Plasmids obtained from Rob Fagan at the University of Sheffield, in *C. difficile* 630Δ*erm*. Plasmids contain ColE1 + *tra* Gram-negative replicon, CD6 (*orfB* and *repA*) Gram-positive replicon and the *catP* antibiotic resistance marker for chloramphenicol and thiamphenicol resistance.

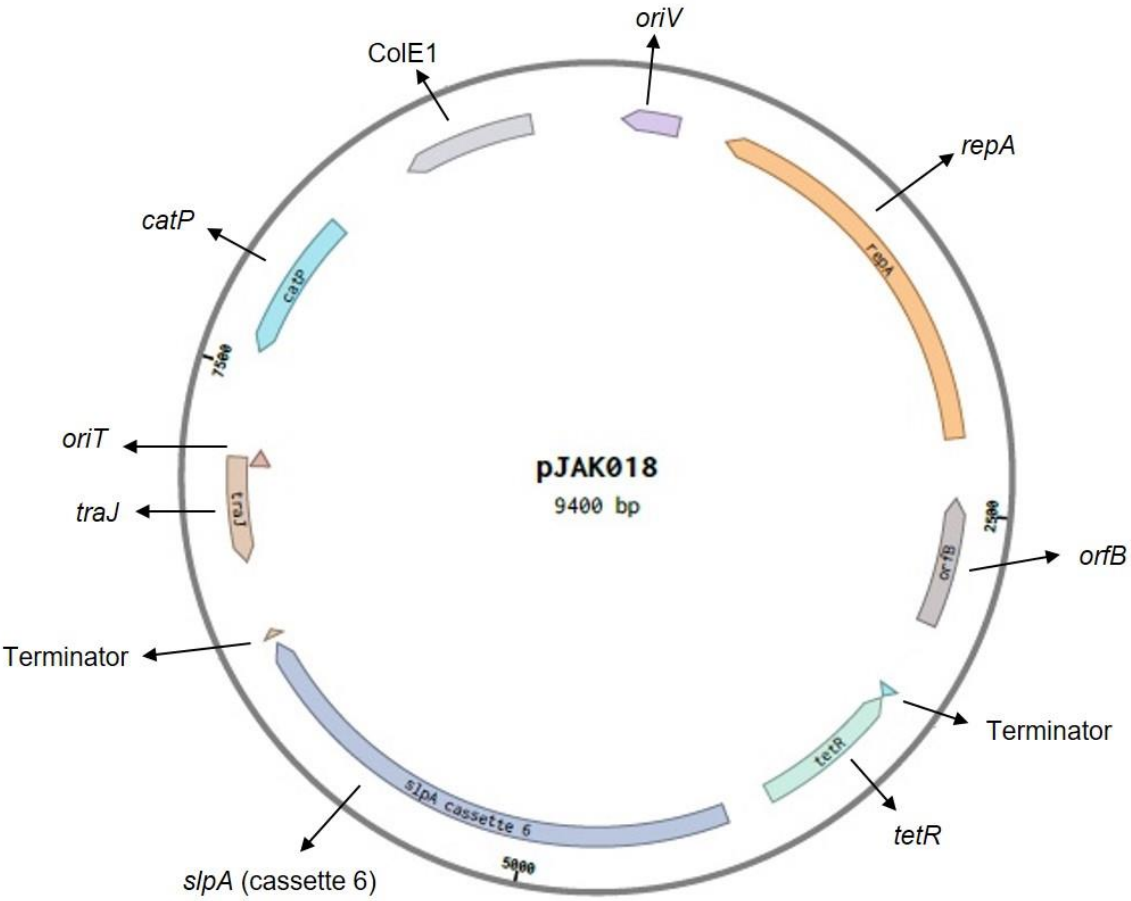


Figure 6.5: **Plasmid pJAK018**. The *slpA* gene from S layer cassette 6 has been cloned between *Sac*I and *Bam*HI sites into a plasmid based on pRPF185, *slpA* is under control of the tetracycline inducible promoter and is induced using 500 ng/ml anhydrotetracyline. Plasmids obtained from Rob Fagan at the University of Sheffield, in *C. difficile* 630Δ*erm*. Plasmids contain ColE1 + *tra* Gram-negative replicon, CD6 (*orfB* and *repA*) Gram-positive replicon and the *catP* antibiotic resistance marker for chloramphenicol and thiamphenicol resistance.

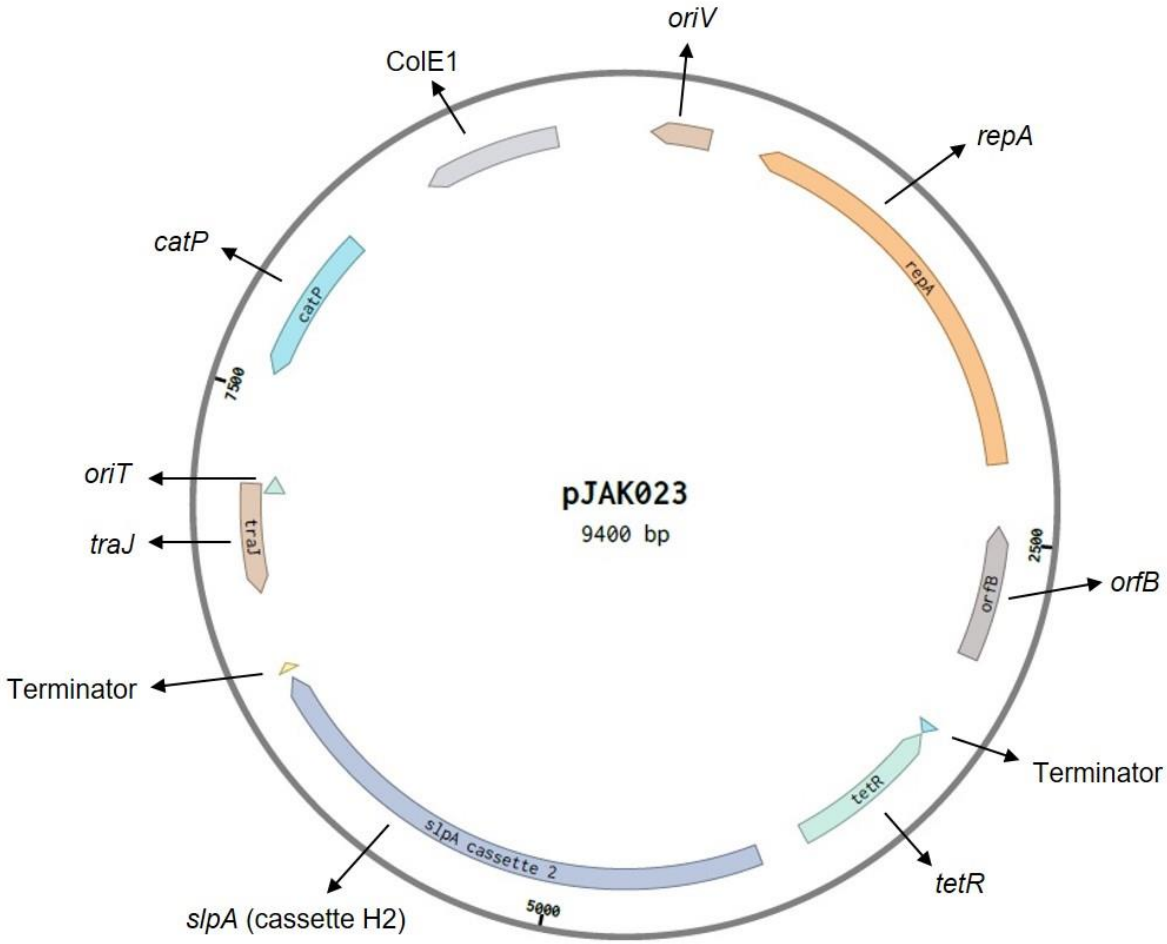


Figure 6.6: **Plasmid pJAK023**. The *slpA* gene from S layer cassette 2 has been cloned between *Sac*I and *Bam*HI sites into a plasmid based on pRPF185, *slpA* is under control of the tetracycline inducible promoter and is induced using 500 ng/ml anhydrotetracyline. Plasmids obtained from Rob Fagan at the University of Sheffield, in *C. difficile* 630Δ*erm*. Plasmids contain ColE1 + *tra* Gram-negative replicon, CD6 (*orfB* and *repA*) Gram-positive replicon and the *catP* antibiotic resistance marker for chloramphenicol and thiamphenicol resistance.

6.2.5 Investigating slpA *in other* C. difficile *strains*

Currently, there are 14 known S layer cassettes in *C. difficile*, these and the PCR ribotypes they correspond to can be found in Appendix 8.4. In order to determine if the *slpA* genes present in the four phage propagating strains conformed to the

published sequences for *slpA* genes of these PCR ribotypes, PCR primers were designed. Primers were also designed to amplify *slpA* from S layer cassette 4, from PCR ribotype 027 strains, to further determine the PCR ribotype of CD31662, CDDH183 and CD051223046. Primers were designed for the *slpA* gene from PCR ribotype 002 strains, S layer cassette 9, to amplify from *C. difficile* 08011 genomic DNA. For *C. difficile* 522418, PCR ribotype 023, the *slpA* gene from S layer cassette 11 was used for primer design. For *C. difficile* 2301, PCR ribotype 014, S layer cassette 6 was used for primer design and for *C. difficile* 1801, PCR ribotype 078, S layer cassette H2/6 was used. The primers used are listed in Table 2.4. PCR was conducted using Q5® High-Fidelity DNA Polymerase and an annealing temperate of 56 °C. To further assess the identity of the *slpA* gene from CD051223046, a PCR ribotype 122 strain, genomic DNA was extracted using a phenol chloroform protocol and subjected to Illumina MiSeq paired-end sequencing. The resulting sequencing reads were assembled to two reference genomes using CLC Genomics Workbench 8.5.1 and the *slpA* gene identified in the consensus sequence. The *slpA* gene was located by searching for a 46 bp nucleotide sequence (Figure 6.7) that is conserved across all *C. difficile* strains which lies upstream of the *slpA* gene. Once the *slpA* gene had been located in the consensus sequence of CD051223046, primers were designed to amplify the complete *slpA* gene and Sanger Sequencing was used to confirm the exact nucleotide sequence.

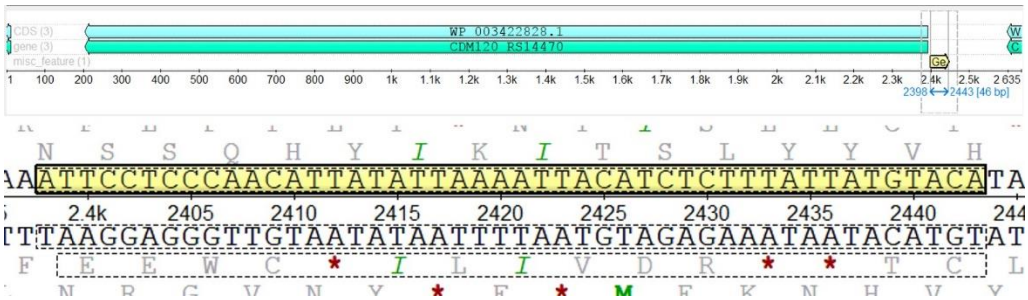


Figure 6.7: **Conserved 46 bp sequence upstream of *slpA* in all *C. difficile* strains**. This sequence is conserved in the genomes of all *C. difficile* strains and can be used to identify the *slpA* gene. It is a 46 bp sequence and lies upstream of *slpA* in the S layer cassette. Shown here is the *slpA* gene, annotated as CDM120_RS14470, in the genome of *C. difficile* 1801.

6.3 Results

6.3.1 Identification of a Phage Receptor Candidate – Putative Membrane Protein

One route to the identification of a phage receptor is to compare the genome of a resistant strain to that of sensitive strains and search for any gene that is altered such that the amino acid sequence of the encoding gene product has changed. Moreover, in order to act as a receptor, the encoded product should be localised to the cell surface. In such a search, it is important to focus on those proteins predicted to have an external cellular location that is accessible to the phage.

In the case of phage ΦCD1801 it was notable that with two exceptions, strains CD2315 and CD31662, all of the *C. difficile* PCR ribotype 078 isolates tested could be infected. It later transpired (4.3.3 and 6.3.1) that strain CD31662 was in fact a PCR ribotype 027 strain. Efforts were, therefore, focussed on this phage and the genome sequences of three sensitive PCR ribotype 078 strains (CD1801, CD2016 and CD7009825) determined, together with the two resistant strains

(CD2016 and CD31662, later shown to be a PCR ribotype 027 strain). Chromosomal DNA was prepared from all five strains and subjected to Illumina paired-end sequencing on a MiSeq platform. The sequence reads obtained were mapped to the annotated genome (Accession No. NC_017174) of PCR ribotype 078 reference strain, M120, using CLC Genomics Workbench 8.5.1. The results of the assembly are shown in Table 6.1. *C. difficile* CD31662 was subsequently assembled to the PCR ribotype 027 reference genome, R20291 (Accession No. FN545816). This allowed 99.66 % of the sequencing reads to be mapped, a percentage considerably higher than when the sequencing reads were mapped to *C. difficile* M120, where only 87.72 % of the reads were mapped.

Table 6.1: CLC Genomics Workbench 8.5.1 Assembly Data of Resistant and Sensitive *C. difficile* PCR Ribotype 078 Strains in Comparison to the Reference Genome, *C. difficile* M120

***C. difficile* strain**	**Number of mapped reads**	**Percentage of mapped reads (%)**
CD1801	1,677,232	93.56
CD2315	1,224,227	90.64
CD2016	1,703,357	97.54
CD7009825	1,729,837	97.60
CD31662	1,496,033	87.72

The number and percentage of mapped reads as determined by CLC Genomics Workbench 8.5.1, when sequencing data for five sensitive and resistant *C. difficile* isolates from PCR ribotype 078 were assembled to the PCR ribotype 078 reference genome, M120. The strains CD31662 was later determined to belong to PCR ribotype 027.

The number of SNPs relative to the M120 reference genome for each strain was determined using CLC fixed ploidy variant analysis, with a ploidy of 1, using default parameters (Table 6.2). Significantly, the number of SNPs for CD31662 was considerably higher than for the other four strains. This observation was the

first suggestion that CD31662 may not be a PCR ribotype 078 strain. The genomes were not closed and were studied for SNP analysis only. A second SNP analysis, using the same conditions, was conducted to compare the two resistant strains with the three sensitive strains (Table 6.2). In this case, the consensus sequence for each resistant strain was extracted from the assembly to the reference genome and used as the reference genome for assembly of the sensitive strains.

Table 6.2: SNP Analysis of Sensitive and Resistant *C. difficile* PCR Ribotype 078 Isolates

***C. difficile* Isolate**	**No. of SNPs compared to M120 Reference Strain**	**No. of SNPs compared to CD2315 Resistant Isolate**	**No. of SNPs compared to CD31662 Resistant Isolate**
CD1801	1,200	864	116,079
CD2016	263	804	115,247
CD7009825	259	801	115,209
CD2315	1,257	-	115,819
CD31662	116,309	115,819	-

The number of SNPs was determined using the fixed ploidy variant analysis function in CLC Genomics Workbench 8.5.1, using the standard parameters. The number of SNPs was determined for three isolates sensitive to phage ΦCD1801 and two resistant isolates in comparison to *C. difficile* M120 reference genome. A second SNP analysis compared the number of SNPs in the sensitive isolates in comparison to each resistant isolate.

As the number of SNPs was quite high for each strain, over 800, the SNPs were manually studied to identify any that could potentially reside within a gene encoding a surface associated protein (i.e. cell wall or cell membrane proteins), where a SNP was seen in the resistant strains that was not present in the sensitive strains. One protein, a putative membrane protein (CDM120_RS14800), was identified that had a SNP which caused an amino acid change in CD2315

(131C>T, Thr44Ile), and in CD31662 (395G>T, Ser132Ile). Thus, while the SNP was not in the same position in *CDM120_RS14800* homologs of the two strains, the same gene was affected. The SNP causing the amino acid change in *CDM120_RS14800* in CD2315 was confirmed by PCR amplifying the region around the SNP and subjecting the amplified DNA fragment to Sanger Sequencing. This confirmed that the SNP was real and not an artefact of genome assembly.

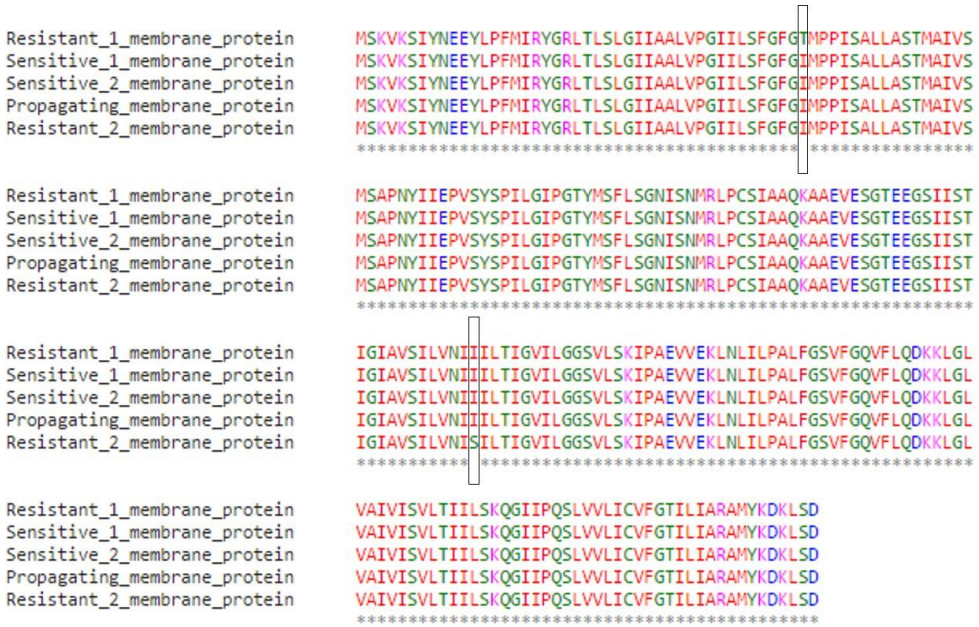


Figure 6.8: **Amino Acid Pairwise Sequence Alignment of translated sequence of gene *CDM120_RS14800* in *C. difficile* PCR ribotype 078 isolates.** *CDM120_RS14800* annotated as putative membrane protein. Three sensitive and two resistant isolates (to phage ΦCD1801 infection) are compared for amino acid changes in the *CDM120_RS14800* gene encoding a putative membrane protein, identified as a potential phage receptor. Amino acid changes Thr44Ile in resistant isolate 1 (CD2315) and Ser132Ile in resistant isolate 2 (CD31662) were observed in comparison to the three sensitive isolates (CD1801, CD2016 and CD7009825) and are shown in the black boxes.

A BLASTp search of the putative membrane protein, encoded by gene *CDM120_RS14800,* shows 92 % sequence identity across a 99 % query coverage to a putative membrane protein from *C. difficile* Y171 (Accession No. EQ126259.1, E value 8e-139). The putative membrane protein, encoded by gene *CDM120_RS14800*, was modelled using the SWISS-MODEL online software to predict the 3D structure of the protein and the location of the amino acid substitution (Figure 6.9). Whilst the model could not predict the structure of the entire protein, it did show three alpha helices for one domain of the protein that are likely to be membrane spanning. The amino acid residue of interest (amino acid that are different between the sensitive and resistant PCR ribotype 078 isolates) appears to be located in a region between the membrane spanning alpha helices. This could suggest that it would be either on the outside or on the inside of the cell. If it were to be on the outside of the cell it could be a credible candidate as a phage receptor.

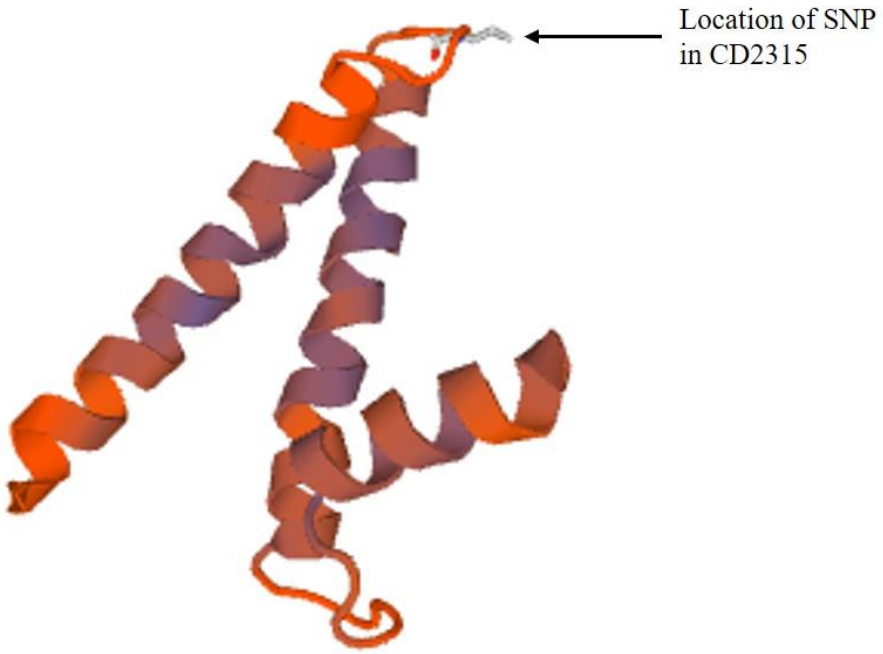


Figure 6.9: **SWISS-MODEL of putative membrane protein encoded by *CDM120_RS14800***; model predicting the structure of the putative membrane protein and showing the location of the SNP identified in *C. difficile* resistant isolate, CD2315, 131C>T. The software could only predict the structure of half of the protein and suggested that this half of the protein was made up of three alpha helices, which could be membrane spanning. The model suggested the location of the SNP in CD2315 (indicated by a block arrow) to be in an area linking two membrane spanning alpha helices, suggesting this location is either surface exposed on the outside of the cell or on the inside of the cell. If this area is exposed on the outside of the cell, this putative membrane protein could be a real phage receptor candidate protein with this substitution between resistant and sensitive isolates being the cause of phage resistance.

6.3.2 Attempt to Create Sensitive CD2315 through In Trans Expression of CDM120_RS14800 *from* C. difficile *1801*

The *CDM120_RS14800* gene from *C. difficile* 1801 was amplified and cloned into a vector under control of its native promoter and the resultant plasmid,

pMTLMW1, introduced into *C. difficile* CD2315. A binding assay was used to determine whether binding of phage ΦCD1801 to CD2315 was promoted by the presence of the introduced gene encoding the putative membrane protein. Whilst, no binding was, as expected, observed in the absence of the plasmid, it could not be determined whether binding took place in the presence of the introduced putative membrane protein (Table 6.3) because it would seem that strain CD2315 carries a prophage which was spontaneously released during the binding assay. The spontaneous prophage induction caused an increase in phage titre that may have masked the phage ΦCD1801 binding effect. The presence of a prophage in CD2315 that is active against *C. difficile* 1801 would account for the observed increase in phage titre. In order to detect whether binding was occurring in the presence of the introduced putative membrane protein, plasmid pMTLMW1 was subsequently transferred to strain CD31662, a known resistant strain. Binding of phage ΦCD1801 to CD31662 was not observed despite the presence of the putative membrane protein from CD1801. It is likely that the identified ORF is not the phage receptor however, the lack of binding could be due to some other mechanism and not necessarily that the identified ORF is not the receptor.

Table 6.3: Analysis of binding of phage ΦCD1801 to *C. difficile* strains in the presence and absence of putative membrane protein from *C. difficile* 1801

***C. difficile* Strain**	**Titre of Free Phages (pfu/ml)**	**Phage Binding Observed?**
CD1801	63	Yes
CD2315	2.0×10^6	No
CD31662	2.8×10^4	No
CD2315:pMTLMW1	2.5×10^6	No
CD31662:pMTLMW1	4.0×10^4	No
Phage Only Control	2.5×10^4	No

Phage binding has occurred when the free phage titre is lower than the phage titre for the phage only control, where no bacterial strain has been used. The strain/phage mixture was allowed to adsorb for 15 minutes prior to measurement of free phage. The number of free phages reported is the average titre from three experiments.

6.3.3 Identification of a Prophage in C. difficile *2315*

An increase in the number of free phages during the burst assay experiments appeared to show the spontaneous release of prophage, as the number of phages at the end of the experiment was higher than the infection titre (Table 6.3). Firstly, PCR was used to identify whether the suspected prophage contained in the genome of *C. difficile* 2315 was the same as phage ΦCD1801. Three primer sets were designed to amplify the integrase gene and genes encoding a putative phage protein and the capsid protein. Primers CD1801_*gp14*_F and CD1801_*gp14*_R were used to amplify the gene encoding the putative phage protein. Primers CD1801_gp20_F and CD1801_gp20_R were used to amplify the integrase gene and primers CD1801_gp46_F and CD1801_gp46_R were used to amplify the gene coding for the phage capsid protein. PCR was conducted using Q5® High-Fidelity DNA Polymerase and an annealing temperature of 60 °C. Amplification of the gene encoding the putative phage protein was observed for both phage ΦCD1801 and from *C. difficile* 2315 DNA

however, no amplification of the gene encoding the capsid protein or the integrase gene was observed with *C. difficile* 2315. The amplification of the three genes from genomic DNA of phage ΦCD1801 and only the amplification of one of the genes from *C. difficile* 2315 genomic DNA suggests that the prophage present in this strain is not the same phage as ΦCD1801.

The presence of a prophage within strain CD2315 was further investigated using the PHASTER online tool. This programme identifies phage regions within bacterial genomes, or raw sequencing data. The programme identified three prophage regions of 17,690, 27,298 and 10,460 bp in size. BLAST analysis of the predicted prophage region showed the presence of a gene encoding for a cI-like repressor protein with 100 % identity, at the amino acid level, to the cI-like repressor protein in the published genome sequence of *C. difficile* phage ΦCD27 (Mayer *et al.*, 2008). The gene encoding the cI-like repressor protein is also identical, at the amino acid level, to the predicted gene coding for the cI-like repressor protein (CD1801_gp35) in phage ΦCD1801, as determined using EMBOSS pairwise sequence alignment. This evidence suggests that the prophage within strain CD2315 is causing lysogenic immunity which is preventing infection by phage ΦCD1801.

The prophage was induced from an overnight culture of CD2315 using mitomycin C, final concentration 3 µg/ml, and was observed to be able to infect and lyse *C. difficile* 1801. Therefore, the study of phage binding cannot be observed when using this strain as the titre of phage increases likely due to the spontaneous release of prophage from this strain.

6.3.4 Attempt to Create Sensitive CDDH1916 and CD31662 Strains Through the Over Expression of CDM120_RS14800 *from* C. difficile *1801*

Plasmid pMTLMW4, containing the *CDM120_RS14800* gene under control of the P_{fdx} promoter, was expressed in strains CDDH1916 and CD31662, two PCR ribotype 027 strains resistant to infection from phage ΦCD1801. Binding of phage ΦCD1801 was assayed with each strain, with and without the presence of plasmid pMTLMW4 (Table 6.4). Binding was not observed in the presence or in the absence of the putative membrane protein. Although there is no experimental data to show whether the protein was being expressed, there was no impact on *C. difficile* growth or colony morphology. In combination this suggests the putative membrane protein is not the receptor for phage ΦCD1801.

Table 6.4: Analysis of binding of phage ΦCD1801 to *C. difficile* strains in the presence and absence of putative membrane protein from *C. difficile* 1801, under control of the P_{fdx} promoter

***C. difficile* Strain**	**Titre of Free Phages (pfu/ml)**	**Phage Binding Observed?**
CD1801	5	Yes
CDDH1916	1.8×10^5	No
CD31662	1.5×10^5	No
CDDH1916:pMTLMW4	2.2×10^5	No
CD31662:pMTLMW4	2.2×10^5	No
Phage Only Control	2.8×10^5	No

Phage binding has occurred when the free phage titre is lower than the phage titre for the phage only control, where no bacterial strain has been used. The strain/phage mixture was allowed to adsorb for 15 minutes prior to measurement of free phage. The number of free phages after this time is reported.

6.3.5 SlpA as the Receptor for Phage ΦCD1801

The *C. difficile* S layer was recently proposed as a phage receptor candidate by Fagan and colleagues (Kirk *et al*., 2016). Three plasmids, pJAK002, pJAK018 and pJAK023, containing *slpA* from cassette H2/6, 6 and 2, respectively, were obtained from Rob Fagan at the University of Sheffield. In each plasmid, *slpA* was under control of the tetracycline inducible promoter. *C. difficile* 630Δ*erm* strains carrying each plasmid were treated with 500 ng/ml anhydrotetracyline to induce expression of the *slpA* cassettes. Phage ΦCD1801 was unable to bind to *C. difficile* 630Δ*erm* (Table 6.5), however, in the presence of pJAK002, *slpA* from cassette H2/6, binding was observed, to the same degree as binding to the propagating strain (Table 6.5). S layer cassette H2/6 is found in PCR ribotype 078 strains. Binding was also observed in the presence of *slpA* from cassette 6 (pJAK018), although it was not observed in the presence of *slpA* from cassette 2 (pJAK023, Table 6.5). This data suggests that ΦCD1801 could bind and infect PCR ribotype 014 strains however, host range analysis showed that it was unable to infect *C. difficile* 2301 (PCR ribotype 014). This is likely explained by the fact that PCR ribotype 014 strains have more than one S layer cassette type within their ribotype (Dingle *et al*., 2013).

The plasmid containing *slpA* from cassette H2/6 was then introduced into the strains CDDH1916 and CD31662. Binding of phage ΦCD1801 was observed when *slpA* was expressed from the plasmid, in both strains, and binding was not observed when the plasmid was absent (Table 6.5). However, binding was not as efficient in these two strains as when *C. difficile* 630Δ*erm* was expressing

slpA or to the propagating strain, strongly suggesting that SlpA is the receptor for phage ΦCD1801.

Table 6.5: Analysis of binding of phage ΦCD1801 to *C. difficile* strains in the presence and absence of *slpA* from three different S layer cassettes, under control of a tetracycline inducible promoter

***C. difficile* Strain**	**Titre of Free Phages (pfu/ml)**	**Phage Binding Observed?**
CD1801	152	Yes
CD630Δ*erm*	4.3×10^4	No
CD31662	5.0×10^4	No
CDDH1916	6.0×10^4	No
CD630:pJAK002	78	Yes
CD630:pJAK018	20	Yes
CD630:pJAK023	5.5×10^4	No
CDDH1916:pJAK002	2.4×10^3	Yes
CD31662:pJAK002	4.0×10^3	Yes
Phage Only Control	7.5×10^4	No

Phage binding has occurred when the free phage titre is lower than the phage titre for the phage only control, where no bacterial strain has been used. The presence of plasmid caused a slower growth rate than non-plasmid containing strains and therefore any reduction in titre is documented as binding. The strain/phage mixture was allowed to adsorb for 15 minutes prior to measurement of free phage. The number of free phages reported is an average of triplicate data. pJAK002 contains *slpA* from cassette H2/6, pJAK018 contains *slpA* from cassette 6 and pJAK023 contains *slpA* from cassette 2.

6.3.6 Attempt to Create a Strain Capable of Full Phage Infection

C. difficile 630Δ*erm* contains a prophage, ΦC2, with a cI-like repressor protein that has 100 % identity, at the amino acid and nucleotide level, to the cI-like repressor protein in phage ΦCD1801. It was postulated that this strain was lysogenically immune to infection from phage ΦCD1801. To show full phage infection in the presence of the candidate receptor protein in this strain, as opposed to only observing binding, retargeted ClosTron plasmids were designed to insertionally knock out *CD630_29500*, the gene encoding the cI-like repressor

protein, in prophage ΦC2 by intron retargeting. The Perutka algorithm was used to predict a target site and the top two hits were selected for synthesis, one on the anti-sense strand and one on the sense strand, between bases 61 and 62 and 417 and 418 respectively (Figure 6.10). The retargeted ClosTron plasmids were synthesised and subsequently propagated in *E. coli* Top10. The two plasmids were then transformed into *E. coli* CA434 and successfully conjugated into *C. difficile* 630Δ*erm.* Thiamphenicol resistant colonies were produced showing the successful transfer of plasmid DNA however, no erythromycin resistant colonies were subsequently obtained. This was repeated and after 72 hours incubation erythromycin resistant colonies were eventually observable. Colony PCR was conducted to determine whether the ClosTron had inserted at the correct locus, using primers, 630repressorCT_F and 630repressorCT_R, flanking the insertion target site. PCR was conducted using Dream*Taq*, with annealing temperature 5 °C lower than the melting temperature of the primers. PCR showed that all erythromycin resistant colonies had no insertion at the correct locus. This suggests that the insertional knockout at this locus was lethal to the cells, with only cells with insertion at another locus surviving due to low level insertion of RAM at an off-target location. Therefore, the creation of a strain where full phage infection can be observed was not successful.

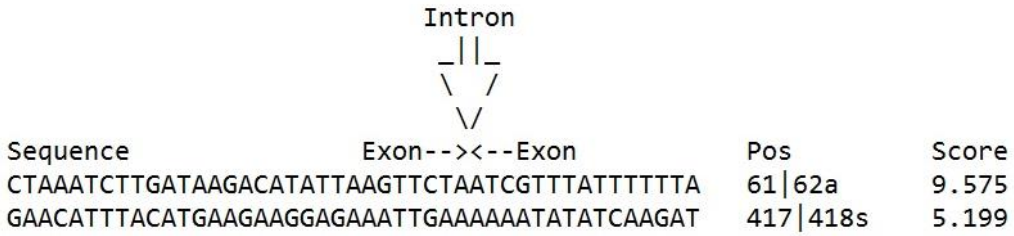


Figure 6.10: **Perutka Algorithm Target Site Identification Results**; The Perutka algorithm identified potential target sites for the ClosTron intron insertion in the *CD630_29500* gene coding for the cI-like repressor protein sequence from *C. difficile* 630Δ*erm* prophage, ΦC2. The target sites are ranked by score, target site 61/62a gave the highest score, 417/418s gave the second highest score. These two sites were selected as one is the sense and one is the anti-sense orientation. The selection of two target sites increases the likelihood of successful insertion.

6.3.7 Investigating slpA *in other* C. difficile *strains*

The sequence identity of *slpA* in other *C. difficile* strains was investigated to determine which SlpA proteins were capable of binding the four isolated phages. For the three propagating strains, CD08011, CD522418 and CD1801, PCR ribotype 002, 023 and 078 respectively, amplification of the *slpA* gene from the published S layer cassette for their PCR ribotype was observed. However, no amplification of *slpA* from S layer cassette 6 was observed using genomic DNA from CD2301 (Table 6.6, Table 6.7).

Table 6.6: Suspected S layer cassette and subsequent amplification of *slpA* for the four *C. difficile* phage propagating strains

***C. difficile* propagating strain**	**PCR ribotype**	**Suspected S layer cassette**	**Amplification of *slpA* observed**
CD08011	002	9	Yes
CD2301	014	6	No
CD522418	023	11	Yes
CD1801	078	H2/6	Yes

Primers were designed for *slpA* from the four S layer cassettes. S layer cassettes for each PCR ribotype were chosen based on information detailed in (Dingle *et al*., 2013). PCR primers were designed using S layer cassette sequence information detailed in (Eidhin *et al*., 2006).

To further confirm the identity of CD31662 as a PCR ribotype 027 strains, previously thought to be PCR ribotype 078, primers were designed to amplify the *slpA* gene from S layer cassette 4. This gene was successfully amplified and the sequence confirmed by Sanger Sequencing. This further confirms that strain CD31662 is a PCR ribotype 027. *C. difficile* R20291, a known PCR ribotype 027 strain, genomic DNA was used as a positive control for the amplification of *slpA* from S layer cassette 4. Host range analysis of the four phages showed the sensitivity of *C. difficile* strains CDDH183 and CD051223046. Both strains were thought to be PCR ribotype 027 however, no amplification of the *slpA* gene from S layer cassette 4 could be observed. It has subsequently been determined that these strains are PCR ribotype 106 and 122, respectively, by the CDRN.

The identity of the *slpA* gene in CD051223046 is important as the phage host range analysis showed that this strain was sensitive to infection by three phages, ΦCD1801, ΦCD08011 and ΦCD2301. In order to investigate this genomic DNA from CD051223046 was extracted and subjected to Illumina MiSeq paired end

sequencing. The resulting 1,269,774 reads were assembled in CLC Genomics Workbench 8.5.1 to two reference genomes, *C. difficile* M120 (Accession No. NC_017174) a PCR ribotype 078 strain and *C. difficile* R20291 (Accession No. FN545816) a PCR ribotype 027 strain (Table 6.7). These two were chosen because a PCR ribotype 122 reference genome is not currently publicly available. The consensus sequence of the mapping against both reference strains was used to produce two draft genomes of CD051223046. A fixed ploidy variant analysis using CLC Genomics Workbench 8.5.1 and a ploidy of 1 was carried out to compare CD051223046 to each of the reference strains (Table 6.7) suggested that CD051223046 was more similar to R20291 due to the higher percentage of mapped reads and lower number of variants.

Table 6.7: Genomic analysis of sequencing reads mapping of CD051223046 against two *C. difficile* reference genomes and fixed ploidy variant analysis

Genomics output	***C. difficile* reference genome for genome assembly**	
	M120	R20291
Number of mapped reads	1,132,093	1,219,676
Percentage of reads mapped (%)	89.16	96.05
Number of variants*	112,107	20,306

*Variants determined using CLC Genomics Workbench 8.5.1 fixed ploidy variant analysis with a ploidy of 1 and default parameters, in comparison to the reference genome. A higher percentage of reads mapped to *C. difficile* R20291.

Although CD051223046 reads were mapped to two different reference genomes the full *slpA* was not mapped in either case. This suggests that part of the gene is relatively conserved with the two reference genomes and could be mapped but a large section of the middle of the gene is dissimilar to the *slpA* present in either reference genome such that it could not be accurately mapped (Figure 6.11). To

further investigate the nucleotide sequence of the *slpA* gene (containing N's where no reads were mapped) from CD051223046 was aligned to the *slpA* from CD08011 and CD1801 using EMBOSS pairwise sequence alignment (see appendix 8.5). CD08011 and CD1801 were chosen for *slpA* comparison as these are the propagating strains for phages ΦCD08011 and ΦCD1801, which can infect CD051223046. The sequence alignment showed that there was greater similarity between the *slpA* genes between the bases 967 and 2181 therefore, a primer in this region was designed for the amplification of *slpA* from CD051223046. A second primer was designed in the conserved sequence motif (Figure 6.7) just upstream of the *slpA* gene.

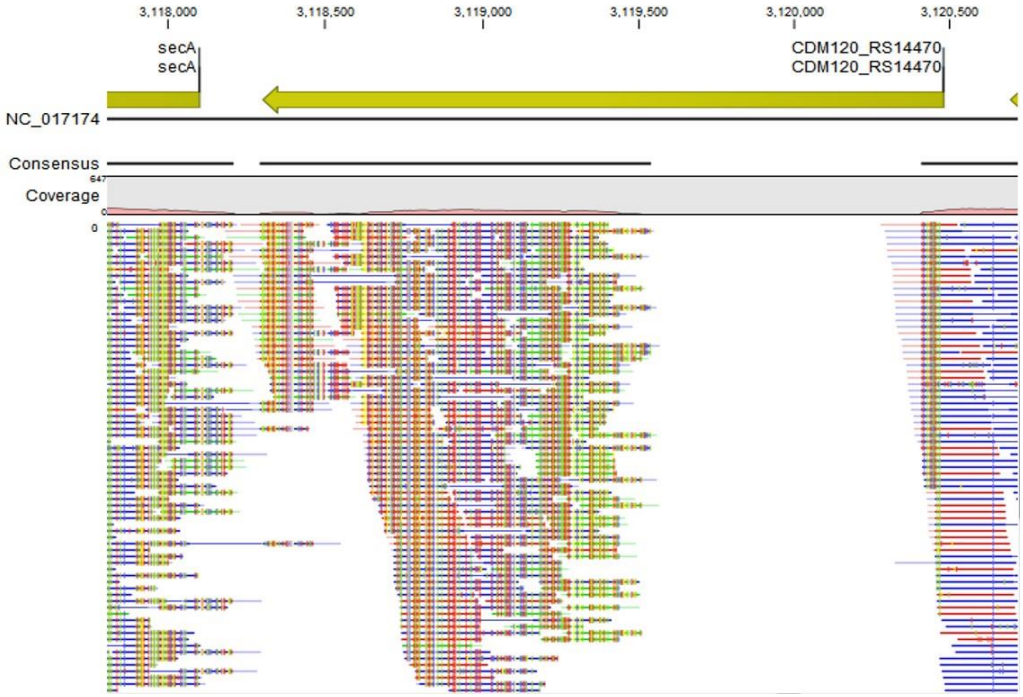


Figure 6.11: **Assembly of *C. difficile* CD051223046 *slpA* gene to the reference genome *C. difficile* M120 using CLC Genomics Workbench 8.5.1.** The *slpA* gene (annotated as CDM120_RS14470 in the reference genome *C. difficile* M120, NC_017174) of CD051223046 only partially maps to the reference genome suggesting that the sequences are very different.

The *slpA* gene from CD051223046 was amplified using primers *slpA*_motif_F1 and *slpA*_motif_R1, an annealing temperature of 55 °C and Q5® DNA polymerase. A DNA band of the expected size, 1033 bp, was produced, purified and subjected to Sanger Sequencing. A nucleotide BLAST of the resulting sequencing data was conducted to determine if the *slpA* gene was similar to any published sequences. The top hit showed similarity to the *slpA* gene from *C. difficile* isolate HR02 (Table 6.8) however, the full genome sequence or the PCR ribotype of this strain is unknown. The BLAST search also showed the amount of sequence identity to the *slpA* gene from the common laboratory strain *C. difficile* 630Δ*erm*, showing a very low percentage coverage of this gene. This suggests that the *slpA* genes are highly variable. Further work is required to determine the S layer cassette present in this strain.

Table 6.8: Nucleotide BLAST results of CD051223046 *slpA* gene sequence

BLASTn Hit	Percentage Query Coverage (%)	Percentage Identity (%)	E Value
Clostridium difficile slpA gene for S layer protein, partial CDS, isolate HR02	89	99	0.00
Clostridium difficile 630Δ*erm*, *slpA* gene	22	85	3e-60

The *slpA* gene was amplified from CD051223046 genomic DNA using Q5® DNA polymerase and subjected to Sanger Sequencing. A nucleotide BLAST was conducted on the resulting sequence.

6.4 Discussion

The successful identification of SlpA as the receptor for phage ΦCD1801 represents the first experimental evidence of a phage receptor in *C. difficile*. The use of a tetracycline inducible promoter, for the controlled expression of *slpA*,

sought to ensure that *slpA* was only expressed in *C. difficile* as it was thought to be likely toxic to *E. coli* and hence, the use of the native promoter could have limited the success of cloning. However, cloning issues were still observed with the Gram-positive replicon being interrupted after the plasmids had been transferred into *E. coli* Top10. A switch in the *E. coli* plasmid propagation strain to *E. coli* DH5α seemed to overcome this problem and successful plasmid propagation could be observed. Induction was conducted using anhydrotetracycline (aTc), a tetracycline homologue as the use of tetracycline causes disruption of growth in *C. difficile* which is eliminated through use of the homologue. Full induction is at around 1 µg/ml aTc but is sufficient with 500 ng/ml with incubation for 1 hour. Phage ΦCD1801 binding was observed in the presence of *slpA* from S layer cassette H2/6 and cassette 6. S layer cassette H2/6, found in PCR ribotype 078 strains, is a hybrid cassette evolved from cassette 2 and cassette 6 by recombination, containing *slpA* and *secA2* from cassette 6 and *cwp66* from cassette 2. As binding is observed in the presence of *slpA* from cassette H2/6 and cassette 6, it is likely these cassettes share the same or similar *slpA*s. It appears that binding in the presence of cassette 6 is more efficient than the propagating strain, this could be due to the size of cell pellet used in the binding assay. Strains were incubated for the same length of time, to the same optical density, however, the cell pellet sizes varied between strains, suggesting that in some cases more cells were present. If more cells are present then the number of phages able to adsorb would be higher and a lower final titre would be observed, this may not necessarily show that binding is more efficient.

The gene *CDM120_RS14800* encoding a putative membrane protein was identified through comparative SNP analysis between sensitive and resistant PCR ribotype 078 *C. difficile* strains. It was proven unlikely to be the phage receptor as the plasmid-based expression and subsequent overexpression of *CDM120_RS14800* in resistant strains did not lead to sensitivity. It was likely that strain CD2315 was resistant to phage infection from ΦCD1801 due to the presence of a prophage in its genome which contained a gene coding for an identical cI-like repressor protein as phage ΦCD1801, as determined by PHASTER. The use of PHASTER and PCR amplification of phage ΦCD1801 specific genes suggested that the CD2315 prophage was not the same phage as ΦCD1801. The amplification of a gene encoding a putative phage protein was observed in CD2315 however, the integrase gene and the gene encoding the capsid protein could not be amplified. It is not surprising that the integrase gene could not be amplified in CD2315 as the site for phage integration into the chromosome is likely to be located just after the integrase gene and may have affected primer binding. The lack of amplification of the gene encoding the capsid protein in CD2315 is enough to suggest that the prophage and phage ΦCD1801 are not the same.

The tentative identification of the putative membrane protein, encoded by *CDM120_RS14800*, as a potential phage receptor in CD2315 was supported by the fact that the other strain (CD31662) resistant to phage ΦCD1801 also contained a SNP in this gene. Although the position of the non-synonymous SNPs of the two strains are not identical, it could be the case that both prevent binding of the phage. The confirmation of CD31662 as a PCR ribotype 027 strain

rather than a PCR ribotype 078 meant that the SNP analysis comparison could only compare one resistant strain with three sensitive strains. This suggested that perhaps the putative membrane protein was a coincidental misdirection. As the binding assays could not show whether binding of phage ΦCD1801 to strain CD2315 was occurring it could not be determined whether two forms of resistance were present in this strain, namely, resistance due to a mutated receptor protein and lysogenic immunity. The spontaneous release of this prophage and its ability to infect CD1801, the phage ΦCD1801 propagating strain, meant that an increase in phage titre from the infection titre was observed in binding assays. Therefore, this strain could not be used to express the putative membrane protein. Comparative SNP analysis would have been more beneficial to the identification of the phage receptor if resistant mutants of the phage propagating strain had been created, using continuous cultures of phage infection. A SNP analysis of resistant mutants could then be directly compared to the propagating strain and would likely produce a more manageable number of SNPs to analyse. This method was not conducted as the production of resistant mutants would be hampered by the emergence of lysogenically immune colonies, as phage ΦCD1801 is not strictly lytic, this would involve large scale screening to select colonies that were true resistant mutants and not lysogens.

Initial cloning of the putative membrane protein and expression in *E. coli* Top10 was problematic, when the gene was under control of the P_{fdx} promoter, during the creation of pMTLMW3. Cloning genes under control of the P_{fdx} promoter has resulted in half of the gene being excised or SNPs appearing in the promoter region when incubations were above room temperature (M. Zygouropoulou,

University of Nottingham, personal communication). The replacement of the high copy number Gram-negative replicon ColE1 with the low copy number replicon, p15a, overcame this problem and allowed growth at 37 °C without excision or SNPs appearing. This was therefore adopted and plasmid pMTLMW4 was created.

A strain that is capable of full infection is desirable for the confirmation of the phage receptor, as the use of binding assays can only confirm that phage binding can occur. *C. difficile* 630Δ*erm* was selected for expression of *slpA* from other S layer cassettes due to its high conjugal efficiency allowing quick and efficient transfer of large plasmids. As its genome sequence is known it was confirmed that it did not contain the same *slpA* sequence as S layer cassette H2/6 and was therefore an appropriate host for binding assays. Unfortunately, due to the presence of the same cI-like repressor protein in prophage ΦC2 which is present in *C. difficile* 630Δ*erm*, this strain could not be used to show full infection and the production of plaques. In order to overcome this, the cI-like repressor protein in the prophage was targeted for ClosTron mutagenesis. The aim was to knockout the cI-like repressor protein, then introduce the plasmid containing *slpA* from cassette H2/6 to show full phage infection. This was unsuccessful and colonies containing the ClosTron intron insertion at the correct locus could not be isolated. It is likely this is because the knockout is lethal. The cI-like repressor protein is involved in the maintenance of prophage lysogeny by blocking the operator which allows transcription of lytic genes and late phage genes. It is postulated that when the cI-like repressor protein is knocked out, the prophage state can no longer be maintained, and lytic growth ensues.

Although the phage receptor for *C. difficile* has been identified for phage ΦCD1801, it is not yet known whether the three other phages have the S layer as the receptor. PCR was used to successfully amplify the *slpA* gene from CD08011 and CD522418 meaning it is likely that SlpA from S layer cassette 9 is the phage receptor for ΦCD08011 and the SlpA from S layer cassette 11 is the phage receptor for ΦCD418. Amplification of the *slpA* gene from S layer cassette 6 from the genomic DNA of CD2301 was unsuccessful suggesting that this SlpA is not the phage receptor for phage ΦCD2301. S layer cassettes 10 and 12 have been reported in *C. difficile* PCR ribotype 014 strains (Dingle *et al*., 2013) and further investigation is required to determine whether the *slpA* gene from either of these cassettes is present in CD2301 and hence, which could be the phage receptor for ΦCD2301.

Host range analysis of the four phages showed that *C. difficile* 051223046 was sensitive to infection from phages ΦCD1801, ΦCD08011 and less efficiently ΦCD2301. Initially it was thought that CD051223046 belonged to PCR ribotype 027 however, amplification of *slpA* using primers specific for S layer cassette 4 was unsuccessful. The PCR ribotype was subsequently determined by the CDRN as PCR ribotype 122. In order to determine the sequence of the *slpA* gene present in this strain, Illumina MiSeq paired end sequencing was conducted, and the resulting reads assembled by mapping to two reference genomes. The full sequence of the *slpA* gene was not mapped in either case however, areas of sequence similarity could be identified and primers were designed to amplify the gene. The *slpA* gene was amplified and the sequence determined. It was shown

to be most similar to the *slpA* gene present in *C. difficile* HR02 however, the S layer cassette present could not be determined. The identification of the S layer as the phage receptor for *C. difficile* could also lead to the identification of the receptor binding proteins in the phage genome.

In the long term, the confirmation of the S layer as the phage receptor in *C. difficile* could lead to the production of a more targeted phage cocktail. Instead of a large phage cocktail containing many phages each targeting different PCR ribotypes, theoretically a phage cocktail targeting each S layer cassette variant could be designed. At present, this would require a cocktail comprising of 14 phages, one phage that is able to target each of the known S layer cassettes. Whether this cocktail could target all *C. difficile* strains would require further investigation as the presence of CRISPR, Abi and restriction modification systems, as well as the widespread presence of prophages in *C. difficile* could remain problematic.

6.5 Chapter Summary

The main outcomes of this chapter are:

- Identification of a potential phage receptor protein through comparative SNP analysis between phage resistant and sensitive strains
- Identification and confirmation of the *C. difficile* S layer as the receptor for phage ΦCD1801 binding

Chapter Seven:

General Discussion

7 General Discussion

7.1 Addressing the Objectives of the Study

The objectives set out at the start of this study were as follows:

- To isolate novel phages, able to infect clinically relevant PCR ribotypes of *C. difficile,* from environmental samples
- To characterise these phages in terms of therapeutic attributes, such as rate of phage to cell attachment and burst size, and to study their host range specificities
- To genotypically characterise these phages using full genome sequencing and to identify key genes, such as those for lysogeny
- To identify potential phage receptor candidates on the *C. difficile* cell surface and to confirm the correct receptor by showing the ability for phage binding in the presence of the phage receptor in otherwise phage resistant *C. difficile* strains

The work detailed in this thesis shows the successful isolation of four novel phages infecting different clinically relevant PCR ribotypes of *C. difficile* (PCR ribotypes 002, 014, 023 and 078). The production of an optimised protocol for *C. difficile* phage isolation has allowed the isolation of multiple phages on different occasions, proving a robust methodology. This can be taken forward for isolation of more phages and the targeted isolation of phages for a particular *C. difficile* PCR ribotype or strain. The four phages, ΦCD08011, ΦCD2301, ΦCD418 and ΦCD1801, were characterised in terms of therapeutic attributes, including average burst size of the population and the rate of phage to cell

attachment. Host range analysis was completed for each phage using a large panel of *C. difficile* isolates from clinically relevant PCR ribotypes, including the hypervirulent PCR ribotype 027. The complete genome sequence for each phage has been detailed and key genes, such as the integrase, endolysin and cI-like repressor genes, have been identified where possible using amino acid sequence homology with published *C. difficile* phages. The isolation of these phages and their subsequent comprehensive characterisation will allow further investigation into phage-host interactions in this organism, as well as furthering our understanding for the potential of phage therapy in the treatment of *C. difficile* infection.

This study has shown the importance of traditional characterisation techniques, such as TEM, in addition to whole genome sequencing for the description of newly isolated phages. Although whole genome sequencing can provide much information about the morphology and classification of a phage, for example through the identification of key genes that are indicative of a certain classification, accurate classification cannot occur without TEM. In this study phage ΦCD418 would have been classified as *Siphoviridae* if TEM analysis had not been used in conjunction with genome sequencing data that indicated the presence of genes associated with *Myoviridae.* For future studies the use of TEM and whole genome sequencing should be utilised in conjunction for accurate classification of phages.

Finally, the *C. difficile* S layer has been identified as the cell surface receptor for phage ΦCD1801. This was confirmed through the expression of *slpA* from S

layer cassette H2/6, found in PCR ribotype 078 strains, in the otherwise resistant strain, *C. difficile* 630Δ*erm.* The *slpA* gene was expressed under the control of a tetracycline inducible promoter and it was shown that upon induction phage ΦCD1801 could bind to *C. difficile* 630Δ*erm.* The identification of the *C. difficile* phage receptor will allow identification of the phage tail proteins involved in binding and could contribute to the ability to broaden phage host ranges using synthetic methods. It can also allow the development of a multi-phage cocktail with a broad host range, potentially one containing 14 phages targeting each of the known *C. difficile* S layer cassettes.

7.2 Future Work

The priority investigation to further this body of work would be the identification of the phage tail proteins involved in phage to bacterial cell binding. This could be completed using bacterial to phage hybridisation assays. The workload could be significantly decreased through the use of a robotic system in the screening for protein to protein interactions, this would mean a higher number of candidate proteins could be screened.

The discovery of the phage proteins involved in receptor binding could allow the creation of a synthetic phage which could target a broad spectrum of *C. difficile* strains. The premise being that one phage could contain a range of receptor binding proteins and hence could bind to multiple *C. difficile* S layers. Currently, there are 14 known S layer cassettes. The use of the isolation protocol optimised in this thesis, could allow the rapid isolation of a phage for each cassette. The subsequent investigation of each phage to ensure the S layer is the receptor in

each case and the identification of phage proteins involved in this binding would be required. Ultimately this will lead to the ability to create a phage cocktail or a synthetic phage with broad host range.

The other avenue to explore from this work is the creation of a strictly lytic phage able to infect *C. difficile.* The identification of genes involved in lysogeny control through genome sequencing can allow targeted mutagenesis with the aim of removing the ability for the phages to lysogenise their hosts. This is likely to be hampered by the screening methods available. The production of a lytic phage will be lethal to the host bacterium and the subsequent screening for lytic phage is therefore problematic. This could be overcome by the expression of the genes required to repress the lytic cycle on a plasmid under control of an inducible promoter. In this case the cI-like repressor gene on the plasmid could be codon optimised so that only the gene within the phage is targeted for mutagenesis. This could be accomplished using ClosTron technology or CRISPR, both of which have been developed for use in *C. difficile* within the group.

Finally, the prospect of using phages as a delivery vehicle for delivering "killing" agents could be investigated. This would move away from traditional phage therapy and use the phages purely as a delivery vehicle. In this case, the isolation or creation of a phage with a broad spectrum is essential. There are multiple industrial companies working to develop this kind of phage therapy, with the delivery of lethal small acid soluble proteins (SASPs, Phico Therapeutics Ltd.) or the potential for the delivery of a CRISPR-Cas system (SNIPR Biome). Other than host range, one of the biggest limitations of tradition phage therapy in the

treatment of CDI is the multiple resistance systems present in *C. difficile* strains, such as the widespread presence of prophage and CRISPR systems. Using phages as a delivery vehicle will allow bacterial "killing" in two forms, firstly killing solely through lytic phage infection and secondly killing through delivery of lethal genes. This completely removes the limitation of bacterial lysogeny, as any lysogens would be killed by the killing agent and therefore strictly lytic phages would not be required. This also overcomes the problems with other resistance present in *C. difficile,* the only limitation envisaged with this type of phage therapy is the requirement for a broad host range phage.

Chapter Eight:

Appendix

8 Appendix

8.1 Full list of *C. difficile* isolates used in the host range testing

C. difficile Isolate	PCR Ribotype	Isolation Location
CD001-1	001	The Netherlands
CD001-2	001	The Netherlands
CD001-3	001	The Netherlands
CD001-4	001	The Netherlands
CD001-5	001	The Netherlands
CD001-6	001	The Netherlands
CD001-7	001	The Netherlands
NCTC 11205	001	Sheffield, UK
NCTC 11207	001	Sheffield, UK
NCTC 11209	001	Sheffield, UK
CD30011	001	Nottinghamshire, UK
CD3401	001	Nottinghamshire, UK
CD3901	001	Nottinghamshire, UK
CD66011	001	Nottinghamshire, UK
CD7501	001	Nottinghamshire, UK
CD9601	001	Nottinghamshire, UK
CDUK1	001	Stoke Mandeville, UK
CD1101	002	Nottinghamshire, UK
CD1501	002	Nottinghamshire, UK
CD2001	002	Nottinghamshire, UK
CD2101	002	Nottinghamshire, UK
CD4101	002	Nottinghamshire, UK
CD52011	002	Nottinghamshire, UK
CD6001	002	Nottinghamshire, UK
CD83012	002	Nottinghamshire, UK
CD9401	002	Nottinghamshire, UK
CD08011	002	Nottinghamshire, UK
CDSubtypeA2	002	The Netherlands
CD16839	002	Hungary
CD8083598	002	The Netherlands
CD8085053	002	The Netherlands
CD8092419	002	The Netherlands
CD9001966	002	The Netherlands

CD9003048	002	The Netherlands
CDMF081988	002	Ireland
CDTL178	002	Unknown
CD01013C17	002	Nottinghamshire, UK
CD01026C3	002	Nottinghamshire, UK
CD1036C4	002	Nottinghamshire, UK
CD03008C4	002	Nottinghamshire, UK
CD03008C2	002	Nottinghamshire, UK
CD2601	014	Nottinghamshire, UK
CD4011	014	Nottinghamshire, UK
CD5601	014	Nottinghamshire, UK
CD6701	014	Nottinghamshire, UK
CD77011	014	Nottinghamshire, UK
CD8001	014	Nottinghamshire, UK
CD88011	014	Nottinghamshire, UK
CD2301	014	Nottinghamshire, UK
CDserogroupH	014	The Netherlands
CDserogroupK	014	The Netherlands
CD8085054	014	The Netherlands
CDTL176	014	Unknown
CD01006C10	014	Nottinghamshire, UK
CD01006C1	014	Nottinghamshire, UK
CD01006C11	014	Nottinghamshire, UK
CD01013C19	014	Nottinghamshire, UK
CD11007C2	014	Nottinghamshire, UK
CD11007C14	014	Nottinghamshire, UK
CD11007C1	014	Nottinghamshire, UK
CD11007C17	014	Nottinghamshire, UK
CD0101	015	Nottinghamshire, UK
CD0301	015	Nottinghamshire, UK
CD1901	015	Nottinghamshire, UK
CD3601	015	Nottinghamshire, UK
CD4201	015	Nottinghamshire, UK
CD4501	015	Nottinghamshire, UK
CD4701	015	Nottinghamshire, UK
CD5901	015	Nottinghamshire, UK
CD6201	015	Nottinghamshire, UK
CD7301	015	Nottinghamshire, UK

CD8401	015	Nottinghamshire, UK
CD8501	015	Nottinghamshire, UK
CD5301	023	Nottinghamshire, UK
CD5101	023	Nottinghamshire, UK
CD7801	023	Nottinghamshire, UK
CD7901	023	Nottinghamshire, UK
CD2401	023	Nottinghamshire, UK
CD1202	023	Nottinghamshire, UK
CD1002	023	Nottinghamshire, UK
CD3701	023	Nottinghamshire, UK
CD520557	023	Nottinghamshire, UK
CD520227	023	Nottinghamshire, UK
CD522682	023	Nottinghamshire, UK
CD522418	023	Nottinghamshire, UK
CD21	023	Unknown
CD305	023	Unknown
CD01015C1	023	Nottinghamshire, UK
CD01015C14	023	Nottinghamshire, UK
CDL1	027	Unknown
CD8156	027	France
CDLibramont1	027	Belgium
CD5108111	027	The Netherlands
CD32219	027	Luxembourg
CDM246	027	Ireland
CD2191	027	Ireland
CD51556	027	Germany
CD51557	027	Germany
CD027Alexander	027	Austria
CD60902	027	Switzerland
CD26131	027	Finland
CD26173	027	Finland
CDC32	027	USA
CDC38	027	USA
CDM13042	027	Canada
CDR20291	027	Stoke Mandeville, UK
CD81566	027	France
CD60109015048	027	The Netherlands
CD31662	027	The Netherlands

CDDH478	027	Taunton, UK
CDDH482	027	Oxford, UK
CDDH361	027	Lewisham, UK
CDDH349	027	Cambridge, UK
CDDH85	027	Birmingham, UK
CDDH326	027	Sheffield, UK
CDDH131	027	Manchester, UK
CDDH320	027	Newcastle, UK
CDDH1916	027	Torbay, UK
CDDH1396	027	Slough, UK
CDDH1432	027	Barnett, UK
CDDH1834	027	Ipswich, UK
CDDH1466	027	Northampton, UK
CDDH1329	027	Coventry, UK
CDDH1751	027	Bradford, UK
CDDH1342	027	Macclesfield, UK
CDDH1858	027	Sunderland, UK
CDR24694	027	Haverfordwest, UK
CDR24988	027	Antrim, UK
CDR23970	027	Dublin, Ireland
CDR12087	027	Europe
CDR20352	027	Canada
CDR20298	027	USA
CD2801	027	Nottinghamshire, UK
CD5101	027	Nottinghamshire, UK
CD9201	027	Nottinghamshire, UK
CD3801	078	Nottinghamshire, UK
CD4401	078	Nottinghamshire, UK
CD8601	078	Nottinghamshire, UK
CD9001	078	Nottinghamshire, UK
CD9501	078	Nottinghamshire, UK
CD9701	078	Nottinghamshire, UK
CD9801	078	Nottinghamshire, UK
CD1801	078	Nottinghamshire, UK
CDWilcox078	078	Leeds, UK
CDType078	078	Leeds, UK
CD2315	078	Hungary
CD2016	078	Ireland

CD7004578	078	The Netherlands
CD7009825	078	The Netherlands
CD7045389	078	The Netherlands
CDM120	078	Unknown, UK
CD01027C1	078	Nottinghamshire, UK
CDDH183	106	Leicester, UK
CDR10432	106	Birmingham, UK
CDR12801	106	Bristol, UK
CDR15347	106	London, UK
CDR108095	106	Poole, UK
CDR22079	106	Sheffield, UK
CDR23942	106	Cambridge, UK
CDR25469	106	Antrim, UK
CDR27384	106	Merthyr Tydfil, UK
CD051223046	122	Belgium

8.2 Efficiency of Plating (EOP) of Sensitive *C. difficile* Isolates

C. difficile isolate	PCR Ribotype	Infecting Phage	Efficiency of Plating (EOP)
CD1101	002	ΦCD08011	0.91
CD1501	002	ΦCD08011	0.51
CD2001	002	ΦCD08011	1.08
CD2001	002	ΦCD2301	0.000000074
CD4101	002	ΦCD08011	0.12
CD52011	002	ΦCD08011	0.32
CD6001	002	ΦCD08011	0.02
CD9401	002	ΦCD08011	0.45
CD08011	002	ΦCD08011	Propagating Strain
CDSubtypeA2	002	ΦCD08011	0.87
CD16839	002	ΦCD08011	2.72
CD8083598	002	ΦCD08011	0.91
CD8092419	002	ΦCD08011	0.62
CD9003048	002	ΦCD08011	1.18
CD01013C17	002	ΦCD08011	0.81
CD01026C3	002	ΦCD08011	0.43
CD1036C4	002	ΦCD08011	1.34
CD6701	014	ΦCD2301	0.00003
CD8001	014	ΦCD2301	0.0000001
CD88011	014	ΦCD2301	0.00002
CD2301	014	ΦCD2301	Propagating Strain
CD0101	015	ΦCD08011	0.02
CD1002	023	ΦCD418	1.00
CD522418	023	ΦCD418	Propagating Strain
CD01015C1	023	ΦCD418	0.43
CD01015C14	023	ΦCD418	0.05
CD3801	078	ΦCD1801	0.003
CD4401	078	ΦCD1801	0.01
CD8601	078	ΦCD1801	0.67
CD9001	078	ΦCD1801	1.00
CD9501	078	ΦCD1801	0.63
CD9701	078	ΦCD1801	0.81

CD9801	078	ΦCD1801	0.53
CD1801	078	ΦCD1801	Propagating Strain
CDWilcox078	078	ΦCD1801	1.67
CDType078	078	ΦCD1801	1.63
CD2016	078	ΦCD1801	0.63
CD7004578	078	ΦCD1801	0.92
CD7009825	078	ΦCD1801	0.21
CD7045389	078	ΦCD1801	0.38
CDM120	078	ΦCD1801	0.001
CD01027C1	078	ΦCD1801	0.09
CDDH183	106	ΦCD2301	0.0000049
CD051223046	122	ΦCD1801	0.00024
CD051223046	122	ΦCD08011	0.5
CD051223046	122	ΦCD2301	N/A*

*No countable plaques produced at any dilution, however repeatable and clearly infection.

8.3 Phage Genome Annotation

8.3.1 ΦCD08011

Locus Tag	Putative Product/Function	Organism (Top BLAST Hit)	BLASTp Query Coverage (%)	BLASTp Identity (%)	BLASTp Expectation (E) value	Accession No. of Top BLAST Hit
CD08011_gp1	Terminase	*C. difficile*	99	98	0	WP_074176751.1
CD08011_gp2	Phage portal protein	*C. difficile*	99	98	0	WP_016056378.1
CD08011_gp3	Clp peptidase	*C. difficile*	99	99	3e-176	WP_074176749.1
CD08011_gp4	Phage major capsid protein	*C. difficile*	99	99	0	WP_074176748.1
CD08011_gp5	Head-tail adaptor protein	*C. difficile*	99	99	3e-73	WP_077723222.1
CD08011_gp6	Phage protein	*C. difficile*	99	100	1e-97	WP_077723222.1
CD08011_gp7	Hypothetical protein	*C. difficile*	99	99	2e-89	WP_074183858.1
CD08011_gp8	Tail sheath protein	*C. difficile*	99	99	0	WP_077723219.1
CD08011_gp9	Terminase	*C. difficile*	99	99	2e-101	WP_077709508.1
CD08011_gp10	XkdN related, phage portal protein	*Clostridium phage phiMMP04*	94	100	3e-81	YP_006990566.1
CD08011_gp11	Hypothetical protein	*C. difficile*	99	99	0	WP_032544517.1
CD08011_gp12	Peptidoglycan binding protein LysM	*C. difficile*	99	95	1e-147	WP_074176742.1
CD08011_gp13	Cell wall hydrolase, NlpC/P60 family protein	*C. difficile*	99	96	0	WP_021418367.1

CD08011_gp14	DUF2577 domain containing protein, unknown function	Multispecies [Clostridiales]	99	98	1e-70	WP_003438447.1
CD08011_gp15	XkdS conserved domain protein, lysozyme like	*C. difficile*	99	94	9e-92	WP_074068728.1
CD08011_gp16	Baseplate protein J	*C. difficile*	99	99	0	WP_077723212.1
CD08011_gp17	Hypothetical protein	*C. difficile*	99	98	2e-143	WP_021418316.1
CD08011_gp18	Phage tail-collar fibre family protein	*C. difficile*	99	93	1e-170	WP_021382766.1
CD08011_gp19	Hypothetical phage protein	*C. difficile*	98	100	1e-61	WP_074124370.1
CD08011_gp20	Phage protein	*C. difficile*	98	98	6e-61	WP_075052697.1
CD08011_gp21	Hypothetical protein	*C. difficile*	98	100	4e-41	WP_074145407.1
CD08011_gp22	Hypothetical protein	*C. difficile*	98	100	6e-63	WP_021365229.1
CD08011_gp23	Holin	*C. difficile*	98	100	2e-52	WP_074137214.1
CD08011_gp24	N-acetylmuramoyl-L-alanine amidase, endolysin	*C. difficile*	99	96	0	WP_077710840.1
CD08011_gp25	Hypothetical protein, unknown function	*C. difficile*	99	100	3e-76	WP_0777267664.1
CD08011_gp26	ParA homolog, chromosome partitioning protein	*C. difficile*	99	99	0	WP_016056403.1
CD08011_gp27	DNA polymerase III epsilon subunit	*C. difficile*	100	99	2e-170	WP_021404582.1
CD08011_gp28	Type III restriction enzyme	*C. difficile*	99	84	0	SJV93868.1
CD08011_gp29	XRE transcriptional regulator, putative cI-like repressor protein	*C. difficile*	99	99	3e-83	WP_076639258.1
CD08011_gp30	Hypothetical protein	*C. difficile*	99	96	5e-93	WP_016056413.1
CD08011_gp31	Hypothetical protein	*C. difficile*	99	99	4e-77	WP_074081164.1
CD08011_gp32	Hypothetical protein	*C. difficile*	99	99	4e-102	WP_016056421.1

CD08011_gp33	Integrase	*C. difficile*	99	99	1e-131	WP_074081161.1
CD08011_gp34	Hypothetical protein	*C. difficile*	79	98	1e-56	WP_074081160.1
CD08011_gp35	HNH endonuclease	*C. difficile*	98	91	1e-58	WP_022619370.1
CD08011_gp36	Hypothetical protein	*C. difficile*	99	100	5e-79	WP_077726766.1

8.3.2 ΦCD2301

Locus Tag	**Putative Product/Function**	**Organism (Top BLAST Hit)**	**BLASTp Query Coverage (%)**	**BLASTp Identity (%)**	**BLASTp Expectation (E) value**	**Accession No. of Top BLAST Hit**
CD2301_gp1	Terminase large subunit	*C. difficile*	99	98	0	WP_021414852.1
CD2301_gp2	Phage portal protein	*C. difficile*	99	96	0	WP_021414849.1
CD2301_gp3	Clp protease ClpP	*C. difficile*	99	94	5e-164	WP_074183863.1
CD2301_gp4	Phage major capsid protein	*C. difficile*	99	97	0	WP_003429615.1
CD2301_gp5	Head-tail adaptor protein	*C. difficile*	99	100	9e-74	WP_016056382.1
CD2301_gp6	Hypothetical protein, putative tail component	*C. difficile*	89	97	2e-95	WP_076639269.1
CD2301_gp7	Hypothetical protein	*C. difficile*	99	97	9e-88	WP_022619361.1
CD2301_gp8	Putative tail sheath protein	*C. difficile*	99	99	0	WP_103951332.1
CD2301_gp9	Terminase	*C. difficile*	99	99	2e-101	WP_077709508.1
CD2301_gp10	Portal protein, XkdN-like protein	*C. difficile*	99	98	6e-85	WP_074176744.1
CD2301_gp11	Tail tape measure protein	*C. difficile*	99	94	0	WP_077745824.1
CD2301_gp12	LysM	*C. difficile*	99	97	2e-150	WP_021425765.1
CD2301_gp13	Cell wall hydrolase	*C. difficile*	99	99	0	WP_074183854.1
CD2301_gp14	DUF2577 domain containing protein, no known function	*C. difficile*	99	99	2e-69	WP_021418368.1
CD2301_gp15	XkdS related protein, lysozyme like	*C. difficile*	99	99	4e-97	WP_077745821.1
CD2301_gp16	Baseplate protein J	*C. difficile*	99	95	0	WP_074068729.1

CD2301_gp17	DUF2313 domain containing protein, putative tail component	*C. difficile*	99	96	2e-142	WP_021424423.1
CD2301_gp18	Phage tail protein	*Clostridium sp.* HMSC19B11	99	94	6e-169	WP_070441739.1
CD2301_gp19	Hypothetical protein	*C. difficile*	99	99	0	WP_077709177.1
CD2301_gp20	Hypothetical protein	*C. difficile*	99	100	2e-76	WP_066024976.1
CD2301_gp21	Hypothetical protein	Multispecies [Clostridiales]	99	100	3e-135	WP_021367785.1
CD2301_gp22	Hypothetical protein	Multispecies [Clostridiales]	98	100	5e-63	WP_021363500.1
CD2301_gp23	Holin	*C. difficile*	98	99	3e-52	OJT75652.1
CD2301_gp24	AbiF like protein, abortive infection	*C. difficile* CD38	99	100	2e-71	EQE36637.1
CD2301_gp25	AbiF like protein, abortive infection	*C. difficile* DA00141	99	100	1e-78	EQG53028.1
CD2301_gp26	N-acetylmuramoyl-L-alanine amidase, Endolysin	Multispecies [Clostridiales]	99	100	0	WP_009889444.1
CD2301_gp27	Cell wall binding repeat containing protein	*C. difficile*	99	100	0	WP_021367792.1
CD2301_gp28	N-acetylmuramoyl-L-alanine amidase, Endolysin	*C. difficile*	99	99	0	WP_003427607.1
CD2301_gp29	Hypothetical protein	*Clostridium sp.* HMSC19D07	99	91	2e-76	WP_070480287.1

CD2301_gp30	Par homolog, chromosome partitioning protein	*Clostridium phage phiCD6356*	99	92	0	YP_004306133.1
CD2301_gp31	DNA polymerase III epsilon subunit	*C. difficile*	99	98	1e-168	WP_021404582.1
CD2301_gp32	Type III restriction enzyme	*C. difficile*	99	84	0	SJV93868.1
CD2301_gp33	XRE family transcriptional regulator, Putative cI-like repressor protein	*C. difficile*	99	99	3e-83	WP_076639258.1
CD2301_gp34	Hypothetical protein	*C. difficile*	99	98	4e-93	WP_016056413.1
CD2301_gp35	Hypothetical protein	*C. difficile*	99	98	6e-77	WP_074081164.1
CD2301_gp36	Hypothetical protein	*C. difficile*	99	99	4e-102	WP_016056421.1
CD2301_gp37	Integrase	*C. difficile*	99	100	1e-131	WP_077723228.1
CD2301_gp38	Hypothetical protein	*C. difficile*	98	100	2e-59	WP_074081160.1
CD2301_gp39	HNH endonuclease	*C. difficile*	98	99	1e-74	WP_016056424.1
CD2301_gp40	Hypothetical protein	*C. difficile*	99	95	2e-74	WP_077737381.1

8.3.3 ΦCD418

Locus Tag	Putative Product/Function	Organism (Top BLAST Hit)	BLASTp Query Coverage (%)	BLASTp Identity (%)	BLASTp Expectation (E) value	Accession No. of Top BLAST Hit
CD418_gp1	ATPase subunit of terminase	*C. difficile*	96	84	4e-144	WP_021415870.1
CD418_gp2	PBSX subunit of terminase, large subunit	*C. difficile*	99	100	0	WP_077709167.1
CD418_gp3	Phage portal protein	Multispecies [Clostridiales]	99	99	0	WP_021370619.1
CD418_gp4	Hypothetical protein	*C. difficile*	99	99	0	WP_077726792.1
CD418_gp5	Phage minor structural GP20 family protein	*C. difficile*	99	100	2e-147	WP_021415873.1
CD418_gp6	Hypothetical protein	*C. difficile*	99	100	1e-78	WP_021377503.1
CD418_gp7	Phage capsid E family protein	*C. difficile*	99	100	0	WP_074178576.1
CD418_gp8	Hypothetical protein	*C. difficile*	99	100	6e-179	WP_042741648.1
CD418_gp9	Hypothetical protein	*C. difficile*	99	100	2e-84	WP_022621257.1
CD418_gp10	Hypothetical protein	*C. difficile*	99	100	1e-96	WP_022621258.1
CD418_gp11	Hypothetical protein	*C. difficile*	99	100	0	WP_077709165.1
CD418_gp12	Phage tail sheath protein	*C. difficile*	99	100	0	WP_077700975.1
CD418_gp13	XkdM	Multispecies [Clostridiales]	99	100	2e-84	WP_016056325.1
CD418_gp14	XkdN-like family protein	Multispecies [Clostridiales]	99	97	2e-107	WP_021370606.1

CD418_gp15	Phage tail tape measure protein	*C. difficile*	99	99	0	WP_077700973.1
CD418_gp16	Hypothetical protein	*C. difficile*	99	100	0	WP_077700972.1
CD418_gp17	Phage portal protein	*C. difficile*	98	96	3e-152	KP148695.1
CD418_gp18	XkdQ	*C. difficile*	99	100	0	WP_074115680.1
CD418_gp19	Hypothetical protein	*C. difficile*	99	100	3e-83	WP_074115681.1
CD418_gp20	XkdS-like family protein, lysozyme-like	*C. difficile*	99	100	2e-108	WP_077709037.1
CD418_gp21	Baseplate J-like family protein	*C. difficile*	99	99	0	WP_087565799.1
CD418_gp22	DUF2313 domain containing protein, putative tail protein	*C. difficile*	99	100	0	WP_077709176.1
CD418_gp23	Tail-collar fibre family protein	Multispecies [Clostridiales]	99	98	8e-85	WP_021390990.1
CD418_gp24	Tail-collar fibre family protein, partial	*C. difficile*	72	98	3e-59	WP_021424292.1
CD418_gp25	Hypothetical protein	*C. difficile*	98	100	7e-62	WP_021424289.1
CD418_gp26	Cell wall hydrolase (endolysin)	*C. difficile*	99	100	0	WP_021425974.1
CD418_gp27	Hypothetical protein	*C. difficile*	99	100	2e-114	WP_066024941.1
CD418_gp28	Hypothetical protein	*C. difficile*	99	100	1e-129	WP_066024940.1
CD418_gp29	Hypothetical protein	*C. difficile*	98	100	6e-63	WP_066024939.1
CD418_gp30	Hypothetical protein	*C. difficile*	99	100	7e-102	WP_066024938.1
CD418_gp31	Hypothetical protein	*C. difficile*	99	100	3e-152	WP_077709076.1
CD418_gp32	Hypothetical protein	*C. difficile*	99	100	4e-104	WP_066024965.1
CD418_gp33	Hypothetical protein	*C. difficile*	92	100	4e-65	WP_077709075.1

CD418_gp34	Bla/MecI/CopY family transcriptional regulator	*C. difficile*	99	100	8e-89	WP_074089040.1
CD418_gp35	Penicillinase repressor family protein	*C. difficile*	99	100	4e-83	WP_021362721.1
CD418_gp36	Putative integrase	*C. difficile*	99	100	0	WP_015984868.1
CD418_gp37	Cro/cI-type transcriptional regulator	*Clostridium phage CDKM15*	99	98	6e-85	ANT45190.1
CD418_gp38	Phage regulatory, Rha family protein	*C. difficile*	99	100	0	WP_021362427.1
CD418_gp39	Hypothetical protein, putative excisionase	*C. difficile* Y270	99	100	8e-100	EQ151544.1
CD418_gp40	Putative bacteriophage resistance protein	*C. difficile*	99	99	3e-111	WP_022616289.1
CD418_gp41	Putative essential recombination protein	*C. difficile*	99	100	4e-143	WP_015984885.1
CD418_gp42	Putative DnaD	*C. difficile*	99	99	0	WP_074089066.1
CD418_gp43	Putative phage resolvase	*C. difficile* F314	98	100	3e-32	EQ100254.1
CD418_gp44	Single stranded DNA binding protein	*C. difficile*	99	100	4e-95	WP_07772353.1
CD418_gp45	Hypothetical protein	*C. difficile*	99	100	7e-85	WP_074099290.1
CD418_gp46	YopX family protein	*C. difficile*	99	97	2e-102	WP_021394538.1
CD418_gp47	Hypothetical protein	*C. difficile*	99	100	5e-68	WP_021394543.1
CD418_gp48	Hypothetical protein	*C. difficile*	99	100	6e-65	WP_021394544.1
CD418_gp49	DNA (cytosine-5-)-methyltransferase	*C. difficile*	99	99	0	WP_021382014.1
CD418_gp50	Nucleoid associated protein	*C. difficile*	99	100	0	WP_021361548.1
CD418_gp51	Hypothetical protein	Multispecies [Clostridiales]	99	100	2e-72	WP_021360983.1
CD418_gp52	Hypothetical protein	*C. difficile* CD8	99	100	1e-67	EQE93161.1

CD418_gp53	Hypothetical protein	Multispecies [Clostridiales]	99	100	2e-80	WP_032509864.1
CD418_gp54	Hypothetical protein, RinA transcriptional activator	*C. difficile*	99	100	5e-112	WP_021424396.1
CD418_gp55	Hypothetical protein	*C. difficile*	99	99	0	WP_021360975.1
CD418_gp56	Putative cI-like repressor protein, transcriptional regulator	*C. difficile*	99	99	0	PBF34916.1
CD418_gp57	GNAT family acetyltransferase	*C. difficile*	99	99	3e-142	WP_021401921.1
CD418_gp58	Nucleoside 2-deoxyribosyl transferase like domain protein, Toll-like receptor	*C. difficile*	99	100	9e-95	WP_077709168.1

8.3.4 ΦCD1801

Locus Tag	Putative Product/Function	Organism (Top BLAST Hit)	BLASTp Query Coverage (%)	BLASTp Identity (%)	BLASTp Expectation (E) value	Accession No. of Top BLAST Hit
CD1801_gp1	Terminase	*C. difficile*	99	100	0	WP_074088970.1
CD1801_gp2	Phage portal protein	*C. difficile*	100	99	0	SJQ73496.1
CD1801_gp3	Putative portal protein	*C. difficile*	99	100	5e-76	WP_077715741.1
CD1801_gp4	Phage head morphogenesis protein	*C. difficile*	99	99	0	WP_074088968.1
CD1801_gp5	Hypothetical protein, DnaD domain	*C. difficile*	99	100	5e-131	SJV41264.1
CD1801_gp6	Hypothetical protein, putative prefoldin	*C. difficile*	99	99	1e-135	WP_074037471.1
CD1801_gp7	Phage capsid protein	*C. difficile*	99	100	0	WP_015981494.1
CD1801_gp8	Putative head connector protein	*C. difficile*	99	91	6e-65	EQK57073.1
CD1801_gp9	Hypothetical protein	*C. difficile*	99	100	2e-79	WP_015981496.1
CD1801_gp10	Hypothetical protein, putative tail component	*C. difficile*	99	100	9e-96	WP_074074702.1
CD1801_gp11	Hypothetical protein	*C. difficile*	99	99	3e-103	WP_015981498.1
CD1801_gp12	Phage portal protein	*C. difficile*	99	100	0	WP_042741634.1
CD1801_gp13	XkdM protein	*C. difficile*	99	100	4e-99	WP_015981500.1
CD1801_gp14	XkdN protein	*C. difficile*	99	100	7e-106	WP_015981501.1
CD1801_gp15	Phage tail tape measure protein	*C. difficile*	99	100	0	WP_042741633.1
CD1801_gp16	Putative tail tape measure protein, partial	Multispecies [Clostridiales]	99	100	3e-175	WP_021391027.1

CD1801_gp17	LysM	*C. difficile*	99	99	3e-153	CDS82755.1
CD1801_gp18	Hypothetical protein	Multispecies [Clostridiales]	99	99	2e-70	WP_015981506.1
CD1801_gp19	Hypothetical protein	Multispecies [Clostridiales]	98	100	6e-51	WP_015981507.1
CD1801_gp20	Cell wall hydrolase, NlpC/P60 family protein	*C. difficile*	99	100	0	WP_074074700.1
CD1801_gp21	DUF2577 domain containing protein, unknown function	*C. difficile*	99	99	5e-70	WP_015981510.1
CD1801_gp22	XkdS protein, lysozyme-like	*C. difficile*	99	100	2e-97	WP_021435020.1
CD1801_gp23	Baseplate protein J	*C. difficile*	99	99	0	WP_059027009.1
CD1801_gp24	DUF2313 domain containing protein, putative tail protein	*C. difficile*	99	99	2e-148	WP_059027007.1
CD1801_gp25	Tail-collar fibre protein	*C. difficile*	99	87	2e-150	WP_021400374.1
CD1801_gp26	Hypothetical protein	*C. difficile*	99	99	0	WP_077723327.1
CD1801_gp27	Hypothetical protein	*C. difficile*	98	100	9e-63	WP_077723326.1
CD1801_gp28	N-acetylmuramoyl-L-alanine amidase, endolysin	*C. difficile*	99	96	0	WP_074183847.1
CD1801_gp29	Hypothetical protein	*C. difficile*	99	100	2e-60	WP_021361301.1
CD1801_gp30	AP2 domain protein	*C. difficile*	99	99	3e-149	SJP46746.1
CD1801_gp31	Hypothetical protein	*C. difficile*	99	100	3e-107	WP_077709030.1
CD1801_gp32	Hypothetical protein	*C. difficile*	99	100	2e-64	WP_074082925.1

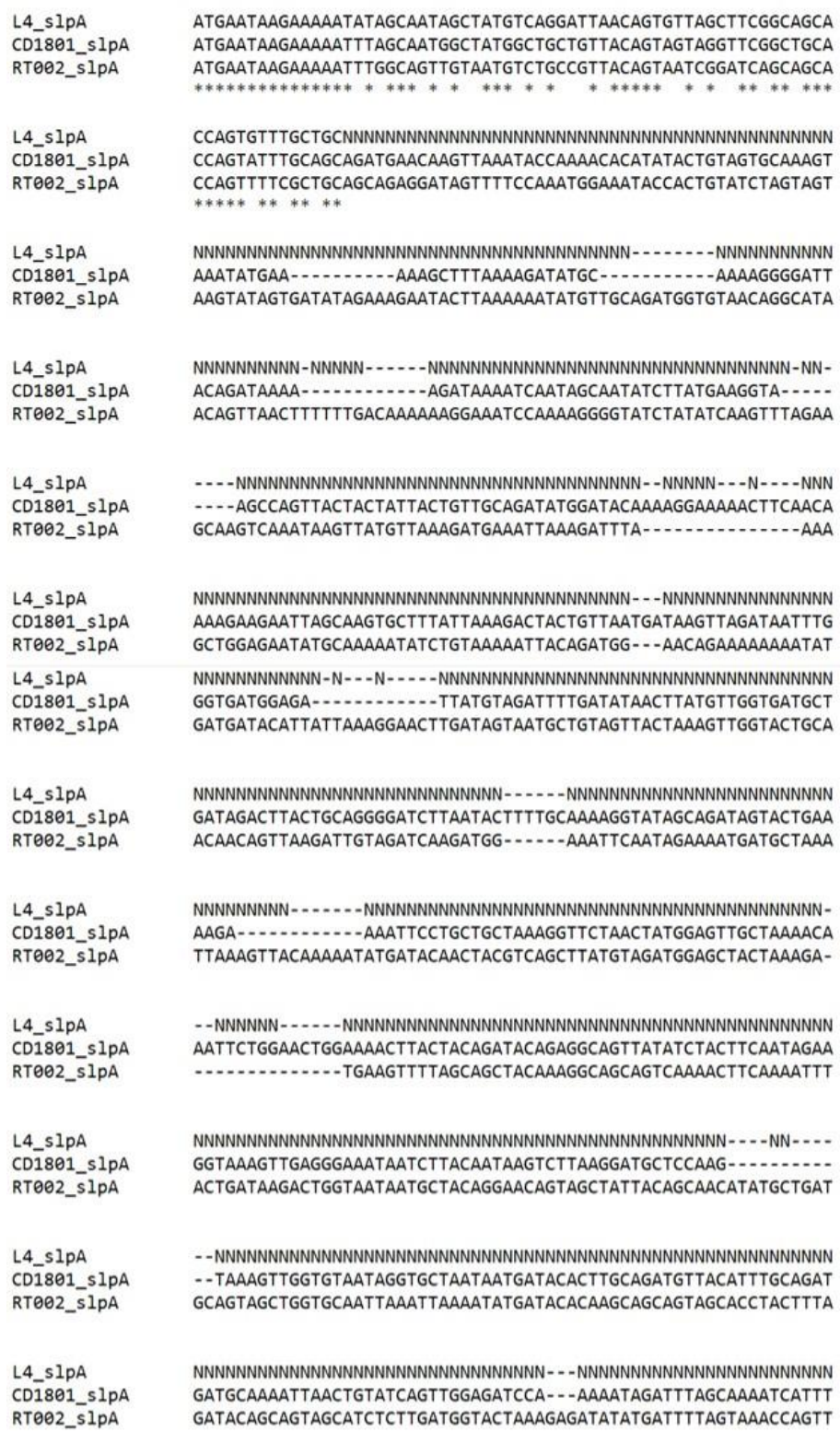
CD1801_gp33	Integrase	Multispecies [Clostridiales]	99	100	0	WP_011861760.1
CD1801_gp34	ImmA/IrrE family metallo-endopeptidase	Multispecies [Clostridiales]	99	100	1e-109	WP_011861759.1
CD1801_gp35	Putative cI-like repressor protein	*C. difficile* Y266	99	100	3e-105	EQ154821.1
CD1801_gp36	Putative anti-repressor	*C. difficile*	98	97	0	WP_103927985.1
CD1801_gp37	Hypothetical protein	Multispecies [Clostridiales]	99	99	1e-65	WP_011861753.1
CD1801_gp38	Putative recombination protein Bet	*C. difficile*	99	98	0	WP_074091101.1
CD1801_gp39	Hypothetical protein	*C. difficile*	99	92	9e-176	WP_074081574.1
CD1801_gp40	YopX family protein	*C. difficile*	98	99	1e-86	WP_021379065.1
CD1801_gp41	Hypothetical protein	*C. difficile*	99	100	8e-80	WP_021401493.1
CD1801_gp42	Hypothetical protein	*C. difficile*	99	99	2e-77	WP_021423237.1
CD1801_gp43	Putative peptide chain release factor protein	*C. difficile*	99	99	1e-73	WP_021435118.1
CD1801_gp44	RusA family crossover junction endodeoxyribonuclease	*C. difficile*	99	100	6e-103	WP_021423240.1
CD1801_gp45	Putative anti-repressor	*C. difficile*	99	99	0	WP_042741824.1
CD1801_gp46	Hypothetical protein	*C. difficile*	99	100	4e-112	WP_021364199.1
CD1801_gp47	Hypothetical protein	*C. difficile*	99	100	2e-70	WP_074033762.1
CD1801_gp48	DUF4868 domain containing protein, bacteriocin synthesis	*C. difficile*	99	99	0	WP_015981549.1

CD1801_gp49	Hypothetical protein	*C. difficile*	99	100	8e-131	CDS82795.1
CD1801_gp50	Transposase	*C. difficile*	99	97	1e-99	WP_074136174.1

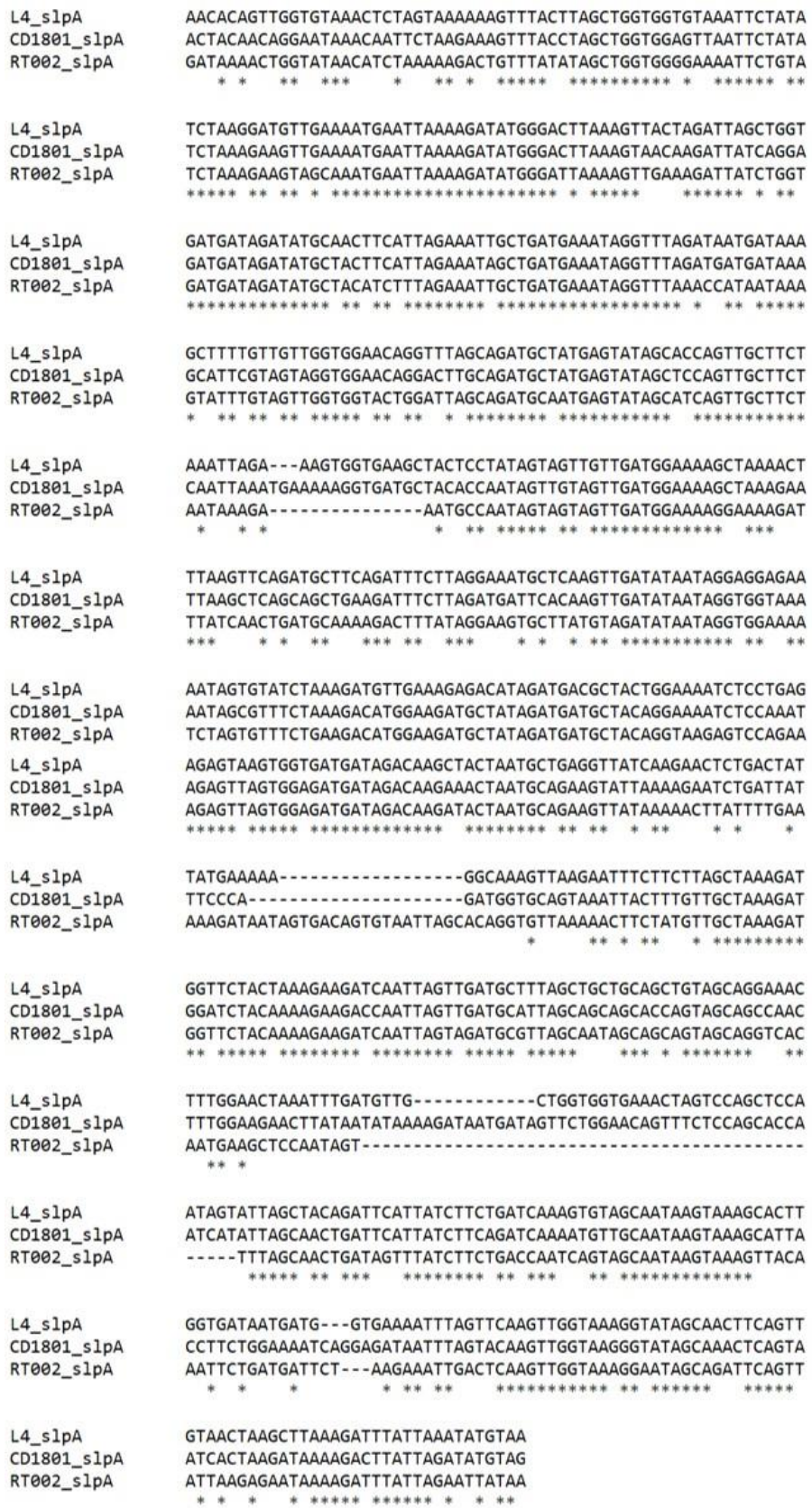
8.4 *C. difficile* S Layer Cassettes and Associated PCR Ribotypes

S layer cassette	PCR ribotype
1	003, 012, 053, 305, 454
2	029, 085, 316
3	033, 284
4	001/072, 011, 015, 027, 106
5	015, 026, 194
6	014. 018
7	012, 019, 017, 038, 137
8	011, 012, 045, 056, 066, 262
9	002, 015
10	013, 014, 020, 069, 106, 129, 153, 321
11	005, 011, 015, 022, 023, 050, 054, 058, 075, 138, 375
12	010, 014
H2/6	078, 193
14	046

8.5 Nucleotide alignment of *slpA* gene from CD051223046 with *slpA* from S layer cassettes 9 and H2/6



```
L4_slpA         NNNNNNNNN---NNNNNNNNNNNNNNN------NNNNNNNNNNNNNNNNNNNNNNNNNN
CD1801_slpA     ATTTTTGAT---ACTAAAACTGGTAAA------TTAGGTGGAATAGTAGAAAAAGAAAAT
RT002_slpA      ATAACGGTTAATACTACTGATGGTACATCAAAATTAAGTTTTGAAAAAACTGGAGAAAAA

L4_slpA         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-NNNNNNNNNNNNNNN-NNNNNNNNN--N
CD1801_slpA     GATGCAACTGAACATGCATATGTAAGAGTAATAAATGCTAAAGAACAAACAATAGATTTA
RT002_slpA      GTTGGAGCAGTTGAATCAAATGTAACAGTTATAGCAGCATCAGATGAAACAGTTACAATA

L4_slpA         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
CD1801_slpA     GATGCTAGTTCATATAAATCAGCTGAAGATTTAGCTAAAGCTTATGCATTTGATGTAAAT
RT002_slpA      TCTGGAGATGCAAAAGAAAAAGCAGAAGCTTTAGCTAAAAAATATGTATTTAAAGATACT

L4_slpA         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
CD1801_slpA     GAGCTTAAGACTCTTTACACTGAAATAGAAGCATATCAAAAAGATAGTAATAATAAAACT
RT002_slpA      GAATTAGAAGATGCATATAAGACAGTAACAGCAAGTGATTT-------------------

L4_slpA         NNNNNNNNNNNNNNNAGTTAATGGTAAGTATCAAGTTGAAATATTTGCAGATGGAAAGAGA
CD1801_slpA     GATAAAGTTCAAATAGTTGATGGTAAATATCAAACAATACTTTATGCTGAAGGAAAGAGA
RT002_slpA      -----TGAAAAAACTGACAATGATTATTATGAGGTAGTTCTTTATCCAACAGGTAAAAGA
                             *   *** * * *** *        * * * *    ** ** ***

L4_slpA         TTAAATACTCTATCAGCTGTA-----------------------AATTCAACAATAGCA
CD1801_slpA     TTAACTACTAAATCAGCAACTCAAGCTTCTAAATTAGCAGAT------------------
RT002_slpA      TTAAATACAGCTTCAACATATGCATCATCTAACTATAAGCCAGAATTACCTACAAGTGAT
                **** ***    *** *

L4_slpA         GATCCAGATTCTACAGCTAAGGTAATAATTAAAGCTGACAAATTAAAAGATTTAAAAGAT
CD1801_slpA     ---GAAAACTCACCACTTAAGTTAGTTATAAAAGCAGATAAATTAAAAGATTTAAAGGAT
RT002_slpA      AGAGTAGATACTCCTGCTATAATAACATTAAGATCTACTAATAAGAATAACTTAAAATCA

L4_slpA         GATCCAGATTCTACAGCTAAGGTAATAATTAAAGCTGACAAATTAAAAGATTTAAAAGAT
CD1801_slpA     ---GAAAACTCACCACTTAAGTTAGTTATAAAAGCAGATAAATTAAAAGATTTAAAGGAT
RT002_slpA      AGAGTAGATACTCCTGCTATAATAACATTAAGATCTACTAATAAGAATAACTTAAAATCA
                     * *  *  *   **   **    * * * *    **    **  * *****

L4_slpA         TATGTAGATGATTTAAAAACTTATAATAATGGATACTCAAATTCAACAAATGTTTCAGGT
CD1801_slpA     TATGTTGAAGATTTAAGAAATGCTAATAATGGTTACTCAAATACTGTTACTGTAGCAGGT
RT002_slpA      GCTTTAGATGAATTAAGAACTTATAATAATTCTTATTCAAATAATTCTGTATTAGCTGGT
                  * * ** ** **** ** *  *******   ** ******         *  * ***

L4_slpA         GATGATAGAATAGAAACAGCAATAGCATTAAGTGAAAAATAC---TACAATTCAGATGAT
CD1801_slpA     GATGATAGAATAGAAACTGCAATAGAATTAAGTAGTAAATACTATAATAATTCAGATGAA
RT002_slpA      GATGATAGAATAGAGACTGCAATAGAAATAAGTAAAGACTCT---TATAATGCAGATGGA
                ************** ** ******* * *****    * *      * *** ******

L4_slpA         GATAATGCATTATTTGAAAAAGAAGTAAATAGTGTAGTTTTAGTTGGTTCAAATGCTATA
CD1801_slpA     GACAATGCTATAACAGAAGATGCTGTTAATAATGTTGTATTAGTTGGATCTCAAGCAATA
RT002_slpA      GGAATTAAAGGAGATTATGTAGAAGCTAATGAGGTAGTATTAGTAGGATCACAATCAATA
                *  * *     *    *    *  *  ***   ** ** ***** ** **  *  * ***

L4_slpA         GTTGATGGTCTTGTAGCATCTCCTTTAGCAGCAGAAAAGAAAGCTCCTTTATTATTAACA
CD1801_slpA     GTTGATGGACTTGTTGCATCACCTTTAGCTTCAGAGAAGAAAGCTCCTTTATTATTAACT
RT002_slpA      GTAGATGGTCTAGTAGCTTCACCTTTAGCTGCTGTTAAAGATGCTCCATTATTATTAACT
                ** ***** ** ** ** ** ********  * *  **  * ***** ***********

L4_slpA         TCAAAAGATAAATTAGATAGTAATGTTAAGTCTGAGATAAAAAGAGTTATGAACTTGAAA
CD1801_slpA     TCAAAAGATAAATTAGATTCAAATGTAAAATCTGAAATTAAAAGAGTTATGAATTTAAAA
RT002_slpA      TCAAAAGATAAGTTAGATTCATCAGTAAAATCTGAAATAAAGAGAGTAATGGGATTAGAT
                *********** ******      ** ** ***** ** ** ***** ***   **  *

L4_slpA         AACACAGTTGGTGTAAACTCTAGTAAAAAAGTTTACTTAGCTGGTGGTGTAAATTCTATA
CD1801_slpA     ACTACAACAGGAATAAACAATTCTAAGAAAGTTTACCTAGCTGGTGGAGTTAATTCTATA
RT002_slpA      GATAAAACTGGTATAACATCTAAAAAGACTGTTTATATAGCTGGTGGGGAAAATTCTGTA
```



C. difficile 051223046 is labelled as L4, CD1801 *slpA* is from S layer cassette H2/6 and RT002 *slpA* is from S layer cassette 9 and is present in *C. difficile* 08011.

Chapter Nine:

References

9 References

Abrini, J., Naveau, H. and Nyns, E.-J. 1994. *Clostridium autoethanogenum*, sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide. *Archives of Microbiology*, 161(4), pp.345–351.

Abt, M.C., McKenney, P.T. and Pamer, E.G. 2016. *Clostridium difficile* colitis: pathogenesis and host defence. *Nature Publishing Group*, 14(10), pp.609–620.

Aksyuk, A.A. and Rossmann, M.G. 2011. Bacteriophage assembly. *Viruses*, 3(3), pp.172–203.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y. and Wishart, D.S. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, 44, pp. 16-21.

Baban, S.T., Kuehne, S.A., Barketi-Klai, A., Cartman, S.T., Kelly, M.L., Hardie, K.R., Kansau, I., Collignon, A. and Minton, M.P. 2013. The Role of Flagella in *Clostridium difficile* Pathogenesis: Comparison between a Non-Epidemic and an Epidemic Strain. *PLoS ONE*, 8(9), pp. 1-9.

Bacon, A.E., Fekety, R., Schaberg, D.R. and Faix, R.G. 1988. Epidemiology of *Clostridium difficile* colonization in newborns: results using a bacteriophage and bacteriocin typing system. *The Journal of Infectious Diseases*, 158(2), pp.349–354.

Bakken, J.S., Borody, T., Brandt, L.J., Brill, J.V., Demarco, D.C., Franzos, M.A., Kelly, C., Khoruts, A., Louie, T., Martinelli, L.P., Moore, T.A., Russell, G. and Surawicz, C. 2011. Treating *Clostridium difficile* Infection with Fecal Microbiota Transplantation. *Clinical Gastroenterology and*

Hepatology, 9(12), pp.1044–1049.

Baptista, C., Santos, M.A. and São-José, C. 2008. Phage SPP1 reversible adsorption to *Bacillus subtilis* cell wall teichoic acids accelerates virus recognition of membrane receptor YueB. *Journal of Bacteriology*, 190(14), pp.4989–4996.

Barbut, F., Mastrantonio, O, Delmée, M., Brazier, J., Kuijper, E. and Poxton, I. 2007. Prospective study of *Clostridium difficile* infections in Europe with phenotypic and genotypic characterisation of the isolates. *Clinical Microbiology and Infection*, 13(11), pp.1048–1057.

Bartlett, J.G. 2008. Historical Perspectives on Studies of *Clostridium difficile* and *C. difficile* Infection. *Clinical Infectious Diseases*, 46(s1), pp.S4–S11.

Bartlett, J.G. and Gerding, D.N. 2008. Clinical Recognition and Diagnosis of *Clostridium difficile* Infection. *Clinical Infectious Diseases*, 46(s1), pp.S12–S18.

Bertozzi Silva, J., Storms, Z. and Sauvageau, D. 2016. Host receptors for bacteriophage adsorption. *FEMS Microbiology Letters*, 363(4), pp. 1-11.

Bertram, R. and Hillen, W. 2008. The application of Tet repressor in prokaryotic gene regulation and expression. *Microbial Biotechnology*, 1(1), pp.2–16.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, L. and Schwede, T. 2014. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42(W1), pp.252–258.

Bidet, P., Barbut, F., Lalande, V., Burghoffer, B. and Petit, J-C. 1999. Development of a new PCR-ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. *FEMS Microbiology Letters*,

175, pp.261–266.

Bliss, D.Z., Johnson, S., Clabots, C.R., Savik, K. and Gerding, D.N. 1997. Comparison of cycloserine-cefoxitin fructose agar (CCFA) and taurocholate-CCFA for recovery of *Clostridium difficile* during surveillance of hospitalized patients. *Diagnostic Microbiology and Infectious Disease*, 29(1), pp.1–4.

Borriello, S.P. 1998. Pathogenesis of *Clostridium difficile* infection. *The Journal of Antimicrobial Chemotherapy*, 41(Suppl C), pp.13–19.

Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M., von Eichel-Streiber. 1996. Defintion of the single integration site of the pathogenicity locus in *Clostridium difficile. Gene,* 181(1), pp. 29-38.

Brazier, J.S. 2001. Typing of *Clostridium difficile*. *Clinical Microbiology and Infection*, 7(8), pp.428–431.

Burnham, C.A.D. and Carroll, K.C. 2013. Diagnosis of *Clostridium difficile* infection: An ongoing conundrum for clinicians and for clinical laboratories. *Clinical Microbiology Reviews*, 26(3), pp.604–630.

Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G. and Fairweather, N. 2001. Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Molecular Microbiology*, 40(5), pp.1187–1199.

Calabi, E. and Fairweather, N. 2002. Patterns of sequence conservation in the S-layer proteins and related sequences in *Clostridium difficile*. *Journal of Bacteriology*, 184(14), pp.3886–3897.

Campbell, A. 2003. The future of bacteriophage biology. *Nature Reviews Genetics*, 4(6), pp.471–477.

Cartman, S.T. and Minton, N.P. 2010. A mariner-Based transposon system for in vivo random mutagenesis of *Clostridium difficile*. *Applied and Environmental Microbiology*, 76(4), pp.1103–1109.

Cartman, S.T., Kelly, M.L., Heeg, D., Heap, J.T. and Minton, N.P. 2012. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype and toxin production. *Applied and Environmental Microbiology*, 78(13), pp.4683–4690.

Cenens, W., Mebrhatu, M.T., Makumi, A., Ceyssens, P-J., Lavigne, R., Van Houdt, R., Taddei, F. and Aertsen, A. 2013. Expression of a Novel P22 ORFan Gene Reveals the Phage Carrier State in *Salmonella typhimurium*. *PLoS Genetics*, 9(2), pp. 1-12.

Cetinkaya, Y., Falk, P. and Mayhall, C.G. 2000. Vancomycin-resistant enterococci. *Clinical Microbiology Reviews*, 13(4), pp.686–707.

Chambers, S.P., Prior, S.E., Barstow, D.A. and Minton, N.P. 1988. The pMTL nic- cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene*, 68(1), pp.139–149.

Chang, A.C.Y. and Cohen, S.N. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *Journal of Bacteriology*, 134(3), pp.1141–1156.

Chaturongakul, S. and Ounjai, P. 2014. Phage-host interplay: Examples from tailed phages and Gram-negative bacterial pathogens. *Frontiers in Microbiology*, 5(442), pp.1–8.

Chopin, M.C., Chopin, A. and Bidnenko, E. 2005. Phage abortive infection in lactococci: Variations on a theme. *Current Opinion in Microbiology*, 8(4), pp.473–479.

Clokie, M.R.J. and Kropinski, A.M. (eds), 2009. *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterisation and Interactions.* Humana Press.

Clokie, M.R.J., Millard, A.D., Letarov, A.V. and Heaphy, S. 2011. Phages in Nature. *Bacteriophage*, 1(1), pp.31–45.

Cohen, S.H., Tang, Y.J. and Silva, J. 2000. Analysis of the pathogenicity locus in *Clostridium difficile* strains. *The Journal of Infectious Diseases*, 181(2), pp.659–663.

Collery, M.M., Kuehne, S.A., McBride, S.M., Kelly, M.L., Monot, M., Cockayne, A., Dupuy, B. and Minton, N.P. 2016. What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of *Clostridium difficile* strain 630 and derivatives. *Virulence*, 0(0), pp.1–15.

Corrigan, R.M. and Foster, T.J. 2009. An improved tetracycline-inducible expression vector for *Staphylococcus aureus*. *Plasmid*, 61(2), pp.126–129.

Davison, S., Couture-Tosi, E., Candela, T., Mock, M. and Fouet, A. 2005. Identification of the *Bacillus anthracis* (gamma) phage receptor. *Journal of Bacteriology*, 187(19), pp.6742–6749.

Delbruck, M. 1945. Effects of specific antisera on the growth of bacterial viruses (bacteriophages). *Journal of Bacteriology*, 50, pp.137–150.

Dingle, K.E., Didelot, X., Ansari, M.A., Eyre, D.W., Vaughan, A., Griffiths, D., Ip, C.L.C., Batty, E.M., Golubchik, T., Bowden, R., Jolley, K.A., Hood, D.W., Fawley, W.N., Walker, A.S., Peto, T.E., Wilcox, M.H. and Crook, D.W. 2013. Recombinational switching of the *Clostridium difficile* S-layer and a novel glycosylation gene cluster revealed by large-scale whole-

genome sequencing. *The Journal of Infectious Diseases*, 207(4), pp.675–686.

Dong, H., Zhang, Y,. Dai, Z. and Li, Y. 2010. Engineering *Clostridium* strain to accept unmethylated DNA. *PLoS ONE*, 5(2), pp.1–8.

Drekonja, D.M., Butler, M., MacDonald, R., Bliss, D., Filice, G., Rector, T. and Wilt, T. 2011. Comparative effectiveness of *Clostridium difficile* treatments: a systematic review. *Annals of Internal Medicine*, 155(12), pp.839–847.

Drilling, A.J., Ooi, M.L., Milijkovic, D., James, C., Speck, P., Vreugde, S., Clarke, J. and Wormald, P-J. 2017. Long-Term Safety of Topical Bacteriophage Application to the Frontal Sinus Region. *Frontiers in Cellular and Infection Microbiology*, 7(49), pp.1–8.

Duckworth, D.H. 1976. "Who Discovered Bacteriophage?". *Bacteriological Reviews*, 40(4), pp.739–802.

Edwards, P. and Smit, J. 1991. A transducing bacteriophage for *Caulobacter crescentus* uses the paracrystalline surface layer protein as a receptor. *Journal of Bacteriology*, 173(17), pp.5568–5572.

Ehrt, S., Guo, X.V., Hickey, C.M., Ryou, M., Monteleone, M., Riley, L.W. and Schnappinger, D. 2005. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Research*, 33(2), pp.1-11.

Eidhin, D.N., Ryan, A.W., Doyle, R.M., Walsh, J.B. and Kelleher, D. 2006. Sequence and phylogenetic analysis of the gene for surface layer protein, *slpA*, from 14 PCR ribotypes of *Clostridium difficile*. *Journal of Medical Microbiology*, 55(1), pp.69–83.

Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., Melamed, S., Leavitt, A., Savidor, A., Albeck, S., Amitai, G. and Sorek, R. 2017. Communication between viruses guides lysis–lysogeny decisions. *Nature*, 541(7638), pp.488–493.

Fagan, R.P., Albesa-Jové, D., Qazi, O., Svergun, D.I., Brown, K.A. and Fairweather, N.F. 2009. Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Molecular Microbiology*, 71(5), pp.1308–1322.

Fagan, R.P. and Fairweather, N.F. 2011. *Clostridium difficile* Has Two Parallel and Essential Sec Secretion Systems. *The Journal of Biological Chemistry*, 286(31), pp.27483–27493.

Farooq, P.D., Urrunaga, N.H., Tang, D.M. and von Rosenvinge, E.C. 2015. Pseudomembranous Colitis. *Disease-a-Month*, 61(5), pp.181–206.

Farris, J.S. 1972. Estimating Phylogenetic Trees from Distance Matrices. *The American Naturalist*, 106(951), pp.645–668.

Fien, K., Turck, A., Kang, I., Kielty, S., Wulff, D., McKenney, K. and Rosenberg, M. 1984. CII-dependent activation of the pRE promoter of coliphage lambda fused to the *Escherichia coli galK* gene. *Gene*, 32(1–2), pp.141–150.

Fogg, P.C.M., Colloms, S., Rosser, S., Stark, M. and Smith, M.C.M. 2014. New applications for phage integrases. *Journal of Molecular Biology*, 426(15), pp.2703–2716.

Fortier, L.C. and Moineau, S. 2007. Morphological and genetic diversity of temperate phages in *Clostridium difficile*. *Applied and Environmental Microbiology*, 73(22), pp.7358–7366.

Freeman, J., Baller, M.P., Baines, S.D., Corver, J., Fawley, W.N., Goorhuis, B., Kuijper, E.J. and Wilcox, M.H. 2010. The changing epidemiology of *Clostridium difficile* infections. *Clinical Microbiology Reviews*, 23(3), pp.529–549.

Garvey, P., Fitzgerald, G.F. and Hill, C. 1995. Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid pNP40. *Applied and Environmental Microbiology*, 61(12), pp.4321–4328.

George, R.H., Syminds, J.M., Bimoch, F., Brown, J.D., Arabi, Y., Shinagawa, N., Keighley, M.R.B., Alexander-Williams, J. and Burdon, D.W. 1978. Identification of *Clostridium difficile* as a cause of pseudomembranous colitis. *British Medical Journal*, 1(6114), p.695.

Girinathan, B.P., Braun, S.E. and Govind, R. 2014. *Clostridium difficile* glutamate dehydrogenase is a secreted enzyme that confers resistance to H_2O_2. *Microbiology*, 160(1), pp.47–55.

Goh, S., Riley, T. V and Chang, B.J. 2005. Isolation and Characterization of Temperate Bacteriophages of *Clostridium difficile. Applied and Environmental Microbiology*, 71(2), pp.1079–1083.

Goh, S., Ong, P.F., Song, K.P., Riley, T.V. and Chang, B.J. 2007. The complete genome sequence of *Clostridium difficile* phage φC2 and comparisons to φCD119 and inducible prophages of CD630. *Microbiology*, 153(3), pp.676–685.

Göker, M., Garcia-Blázquez, G., Voglmayr, H,. Telleria, M.T. and Martin, M.P. 2009. Molecular Taxonomy of Phytopathogenic Fungi: A Case Study in *Peronospora*, *PLoS One*, 4(7), pp. 1-14.

Golkar, Z., Bagasra, O. and Pace, D.G. 2014. Bacteriophage therapy: A potential solution for the antibiotic resistance crisis. *Journal of Infection in Developing Countries*, 8(2), pp.129–136.

Govind, R., Fralick, J.A. and Rolfe, R.D. 2006. Genomic Organization and Molecular Characterization of *Clostridiuum dificile* Bacteriophage PhiCD119. *Journal of Bacteriology*, 188(7), pp.2568–2577.

Govind, R., Vediyappan, G., Rolfe, R.D., Dupuy, B. and Fralick, J.A. 2009. Bacteriophage-mediated toxin gene regulation in *Clostridium difficile*. *Journal of Virology*, 83(23), pp.12037–45.

Hargreaves, K.R. et al., 2013. Genetically diverse *Clostridium difficile* strains harboring abundant prophages in an estuarine environment. *Applied and Environmental Microbiology*, 79(20), pp.6236–6243.

Hargreaves, K.R. and Clokie, M.R.J. 2014. *Clostridium difficile* phages: Still difficult? *Frontiers in Microbiology*, 5(184), pp.1–14.

Hargreaves, K.R. and Clokie, M.R.J. 2015. A Taxonomic Review of *Clostridium difficile* Phages and Proposal of a Novel Genus, “Phimmp04likevirus”. *Viruses*, 7(5), pp.2534–41.

Hargreaves, K.R., Flores, C.O., Lawley, T.D. and Clokie, M.R.J. 2014. Abundant and Diverse Clustered Regularly Interspaced Short Palindromic Repeat Spacers in *Clostridium difficile* Strains and Prophages Target Multiple Phage Types within This Pathogen. *mBio*, 5(5), pp.1–10.

Hargreaves, K.R., Kropinski, A.M. and Clokie, M.R.J. 2014. What does the talking? Quorum sensing signalling genes discovered in a bacteriophage genome. *PLoS ONE*, 9(1), pp. 1-9.

Hawkins, C., Buggy, B., Fekety, R. and Schaberg, D. 1984. Epidemiology of

colitis induced by *Clostridium difficile* in hamsters: application of a bacteriophage and bacteriocin typing system. *The Journal of Infectious Diseases*, 149(5), pp.775–780.

Healthcare Commission, 2006. *Investigation into outbreaks of Clostridium difficile at Stoke Mandeville Hospital, Buckinghamshire Hospitals NHS Trust*.

Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P. and Minton, N.P. 2007. The ClosTron: A universal gene knock-out system for the genus *Clostridium*. *Journal of Microbiological Methods*, 70(3), pp.452–464.

Heap, J.T., Pennington, O.J., Cartman, S.T. and Minton, N.P. 2009. A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods*, 78(1), pp.79–85.

Heap, J.T., Kuehne, S.A., Ehsaan, M., Cartman, S.T., Cooksley, C.M., Scott, J.C. and Minton, N.P. 2010. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *Journal of Microbiological Methods*, 80(1), pp.49–55.

Heap, J.T., Ehsaan, M., Cooksley, C.M., Ng, Y-K., Cartman, S.T., Winzer, K. and Minton, N.P. 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research*, 40(8), pp. 1-10.

Herman, C., Thévenet, D, D'Ari, R. and Bouloc, P. 1997. The HflB protease of *Escherichia coli* degrades its inhibitor λ cIII. *Journal of Bacteriology*, 179(2), pp.358–363.

Horgan, M., O'Sullivan, O., Coffey, A. Fitzgerald, G.F., van Sinderen, D., McAuliffe, O. and Ross, R.P. 2010. Genome analysis of the *Clostridium*

difficile phage φCD6356, a temperate phage of the Siphoviridae family. *Gene*, 462(1–2), pp.34–43.

Howerton, A., Patra, M. and Abel-Santos, E. 2013. A new strategy for the prevention of *Clostridium difficile* infection. *The Journal of Infectious Diseases*, 207(10), pp.1498–504.

Huang, H., Chai, C., Li, N., Rowe, P., Minton, N.P., Yang, S., Jiang, W. and Gu, Y. 2016. CRISPR/Cas9-Based Efficient Genome Editing in *Clostridium ljungdahlii*, an Autotrophic Gas-Fermenting Bacterium. *ACS Synthetic Biology*, 5(12), pp.1355–1361.

Hussain, H.A., Roberts, A.P. and Mullany, P. 2005. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630*Δerm*) and demonstration that the conjugative transposon Tn916ΔE enters the genome of this strain at multiple sites. *Journal of Medical Microbiology*, 54(2), pp.137–141.

Hall, I.C. and O'Toole, E. 1935. Intestinal flora in new-born infants: With a description of a new pathogenic anaerobe, *Bacillus difficilis*. *American Journal of Diseases of Children*, 49(2), pp.390–402.

Johnson, A., Meyer, B.J. and Ptashne, M. 1978. Mechanism of action of the Cro protein of bacteriophage λ. *Biochemistry*, 75(4), pp.1783–1787.

Jones, A.M., Kuijper, E.J. and Wilcox, M.H. 2013. *Clostridium difficile*: A European Perspective. *Journal of Infection*, 66(2), pp.115–128.

Jones, D.T. and Woods, D.R. 1986. Acetone-butanol fermentation revisited. *Microbiological Reviews*, 50(4), pp.484–524.

Karberg, M., Guo, H., Zhong, J., Croon, R., Perutka, J. and Lambowitz, A.M. 2001. Group II introns as controllable gene targeting vectors for genetic

manipulation of bacteria. *Nature Biotechnology*, 19(12), pp.1162–1167.

Karjalainen, T., Waligora-Dupriet, A-J., Cerquetti, M., Spigaglia, P., Maggioni, A., Mauri, P. and Mastrantonio, P. 2001. Molecular and Genomic Analysis of Genes Encoding Surface- Anchored Proteins from *Clostridium difficile*. *Infection and Immunity*, 69(5), pp.3442–3446.

Karjalainen, T., Saumier, N., Barc, M-C., Delmée, M. and Collignon, A. 2002. *Clostridium difficile* genotyping based on *slpA* variable region in S-layer gene sequence: An alternative to serotyping. *Journal of Clinical Microbiology*, 40(7), pp.2452–2458.

Kassam, Z., Lee, C.H., Yuan, Y. and Hunt, R.H. 2013. Fecal microbiota transplantation for *Clostridium difficile* infection: systematic review and meta-analysis. *The American Journal of Gastroenterology*, 108(4), pp.500–508.

Kawata, T., Takeoka, A., Takumi, K. and Masuda, K. 1984. Demonstration and preliminary characterization of a regular array in the cell wall of *Clostridium difficile*. *FEMS Microbiology Letters*, 24(2–3), pp.323–328.

Keessen, E.C., Hopman, N.E.M., van Leengoed, L.A.M.G., van Asten, A.J.A.M., Hermanus, C., Kuijper, E.J. and Lipman, L.J.A. 2011. Evaluation of four different diagnostic tests to detect *Clostridium difficile* in piglets. *Journal of Clinical Microbiology*, 49(5), pp.1816–1821.

Khanna, S., Pardi, D.S., Aronson, S.L., Kammer, P.P., Orenstein, R., Sauver, J.L., Harmsen, W.S. and Zinsmeister, A.K. 2012. The Epidemiology of Community-Acquired *Clostridium difficile* infection: A population-based study. *The American Journal of Gastroenterology*, 2011(1), pp.89–95.

Kim, K.H., Fekety, R., Batts, D.H., Brown, D., Cudmore, M., Silva, J., Waters,

D., 1981. Isolation of *Clostridium difficile* From the Environment and Contacts of Patients With Antibiotic Associated Colitis. *Journal of Infectious Diseases*, 143(1), pp.42–50.

Kirby, J.M., Ahern, H., Roberts, A.K., Kumar, V., Freeman, Z., Acharya, K.R. and Shone, C.C. 2009. Cwp84, a Surface-associated Cysteine Protease, Plays a Role in the Maturation of the Surface Layer of *Clostridium difficile*, 284(50), pp.34666–34673.

Kirk, J.A., Banerji, O. and Fagan, R.P. 2016. Characteristics of the *Clostridium difficile* cell envelope and its importance in therapeutics. *Microbial Biotechnology*, 10(1), pp.76–90.

Kirk, J.A., Gebhart, D., Buckley, A.M., Douce, G.R., Govoni, G.R. and Fagan, R.P. 2017. New class of precision antimicrobials redefines role of *Clostridium difficile* S layer in virulence and viability. *Science Translational Medicine.* 9(406), pp. 1-10.

Köpke, M., Straub, M. and Dürre, P. 2013. *Clostridium difficile* Is an Autotrophic Bacterial Pathogen. *PLoS ONE*, 8(4), pp.1–7.

Koskella, B. and Meaden, S. 2013. Understanding bacteriophage specificity in natural microbial communities. *Viruses*, 5(3), pp.806–823.

Kuehne, S.A. and Minton, N.P. 2012. ClosTron-mediated engineering of *Clostridium*. *Bioengineered*, 3(3), pp.247–254.

Kutter, E. and Sulakvelidze, A. 2005. *Bacteriophages: biology and applications*, CRC Press.

Kwok, C.S., Aruther, A.K., Anibueze, C.I., Singh, S., Cavallazzi, R. and Loke, Y.K. 2012. Risk of *Clostridium difficile* infection with acid suppressing drugs and antibiotics: meta-analysis. *The American Journal of*

Gastroenterology, 107(7), pp.1011–1019.

Lambowitz, A.M. and Zimmerly, S. 2011. Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harbor Perspectives in Biology*, 3(8), pp.1–19.

Lau, C. and Chamberlain, R.S. 2016. Probiotics are effective at preventing *Clostridium difficile*-associated diarrhoea: A systematic review and meta-analysis. *International Journal of General Medicine.* 9(1), pp. 27-37.

Lawson, P.A., Citron, D.M., Tyrrell, K.L. and Finegold, S.M. 2016. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe.* 40(1), pp. 95-99.

Lee, W.J., Billington, C., Hudson, J.A. and Heinemann, J.A. 2011. Isolation and characterization of phages infecting *Bacillus cereus*. *Letters in Applied Microbiology*, 52(5), pp.456–464.

Lefort, V., Desper, R. and Gascuel, O. 2015. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Molecular biology and evolution*, 32(10), pp.2798–2800.

Lessa F.C., Mu, Y., Bamberg, W.M., Beldavs, Z.G., Dumyati, G.K., Dunn, J.R., Farley, M.M., Holzbauer, S.M., Meek, J.I., Phipps, E.C., Wilson, L.E., Winston, L.G., Cohen, J.A., Limbago, B.M., Fridkin, S.K., Gerding, D.N. and McDonald, L.C. 2015. Burden of *Clostridium difficile* infection in the United States. *The New England Journal of Medicine*, 372(9), pp.825–834.

Li, Q., Chen, J, Minton, N.P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng. Z., Ren, X., Yang, J., Gu, Y., Jiang, W., Yiang, Y. and Yang, S. 2016. CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnology Journal*, 11(7),

pp.961–972.

Los, M. and Wegrzyn, G. 2012. Pseudolysogeny. *Advances in Virus Research.* 82(1), pp. 339-349.

Ly-Chatain, M.H., 2014. The factors affecting effectiveness of treatment in phages therapy. *Frontiers in Microbiology*, 5(51), pp.1–7.

MacNair, C.R., Stokes, J.M., Carfrae, L.A., Fiebig-Comyn, A.A., Coombes, B.K., Mulvey, M.R. and Brown, E.D. 2018. Overcoming mcr-1 mediated colistin resistance with colistin in combination with other antibiotics. *Nature Communications.* 9(1), pp. 1-8.

Mahony, D.E., Bell, P.D. and Easterbrook, K.B. 1985. Two Bacteriophages of *Clostridium difficile*. *Journal of Clinical Microbiology*, 21(2), pp.251–254.

Manrique, P., Dills, M. and Young, M.J. 2017. The human gut phage community and its implications for health and disease. *Viruses*, 9(6), pp.9–11.

Marsh, J.W., Arora, R., Schlack, J.L., Shutt, K.A., Curry, S.R. and Harrison, L.H. 2012. Association of relapse of *Clostridium difficile* disease with BI/NAP1/027. *Journal of Clinical Microbiology*, 50(12), pp.4078–4082.

Mayer, M.J., Narbad, A. and Gasson, M.J. 2008. Molecular characterization of a *Clostridium difficile* bacteriophage and its cloned biologically active endolysin. *Journal of Bacteriology*, 190(20), pp.6734–6740.

Meader, E., Mayer, M.J., Gasson, M.J., Steverding, D., Carding, S.R. and Narbad, A. 2010. Bacteriophage treatment significantly reduces viable *Clostridium difficile* and prevents toxin production in an in vitro model system. *Anaerobe*, 16(6), pp.549–554.

Medscape, 2011. *FDA Approves Fidaxomicin for the Treatment of Clostridium difficile Infection*,

Meessen-Pinard, M., Sekulovic, O. and Fortier, L.C. 2012. Evidence of in vivo prophage induction during *Clostridium difficile* infection. *Applied and Environmental Microbiology*, 78(21), pp.7662–7670.

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P. and Göker, M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, 14(60), pp. 1-14.

Meier-Kolthoff, J.P., Hahnke, R.L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., Rohde, C., Rohde. M., Fartmann, B., Goodwin, L.A., Chertkor, O., Reddy, T., Pati, A., Ivanova, N.N., Markowitz, V., Kyrpides, N.C., Woyke, T., Göker, M. and Klenk, H.P. 2014. Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Standards in Genomic Sciences*, 9(2), pp. 1-19.

Meier-Kolthoff, J.P. and Goeker, M. 2017. VICTOR: Genome-based Phylogeny and Classification of Prokaryotic Viruses. *Bioinformatics, btx440,* Available at: http://doi.org/10.1093/bioinformatics/btx440

Miedzybrodzki, R., Borysowski, J., Weber-Dabrowska, B., Fortuna, W., Letkicwicz, S., Szufnarowski, K., Pawelczyk, Z., Rogoz, P., Klah, M., Wojtasik, E. and Gorski, A. 2012. Clinical aspects of phage therapy. *Advances in Virus Research*, 83, pp.73–121.

Minton, N.P., Ehsaan, M., Humphreys, C.M., Little, G.T., Baker, J., Henstra, A.M., Liew, F., Kelly, M.L., Sheng, L., Schwartz, K. and Zhang, Y. 2016. A roadmap for gene system development in Clostridium. *Anaerobe*, 41, pp.104–112.

Monteville, M.R., Ardestani, B. and Geller, B.L. 1994. Lactococcal

bacteriophages require a host cell wall carbohydrate and a plasma membrane protein for adsorption and ejection of DNA. *Applied and Environmental Microbiology*, 60(9), pp.3204–3211.

Musher, D.M., Manhas, A., Jain, R., Niula, F., Waqar, A., Logan, N., Marino, B. and Graviss, E.A. 2007. Detection of *Clostridium difficile* toxin: Comparison of enzyme immunoassay results with results obtained by cytotoxicity assay. *Journal of Clinical Microbiology*, 45(8), pp.2737–2739.

Muto, C.A., Pokrywka, M., Shutt, K., Mendelsohn, A.B., Nouri, K., Posey, K., Roberts, T., Cryole, K., Krystofiak, S., Patel-Brown, S., Pasculle, A.W., Paterson, D.L., Saul, M. and Harrison, L.H. 2005. A large outbreak of *Clostridium difficile*-associated disease with an unexpected proportion of deaths and colectomies at a teaching hospital following increased fluoroquinolone use. *Infection Control and Hospital Epidemiology*, 26(3), pp.273–280.

Na, X. and Kelly, C. 2011. Probiotics in *Clostridium difficile* Infection. *Journal of Clinical Gastroenterology*, 45(Supplementary), pp.S154-158.

Nagy, E. and Foldes, J. 1991. Electron microscopic investigation of lysogeny of *Clostridium difficile* strains isolated from antibiotic-associated diarrhea cases and from healthy carriers. *APMIS*, 99(4), pp.321–326.

Nale, J.Y., Shan, Y., Hickenbotham, P.T., Fawley, W.N., Wilcox, M.H. and Clokie, M.R.J. 2012. Diverse temperate bacteriophage carriage in *Clostridium difficile* 027 strains. *PLoS ONE*, 7(5), pp.1–9.

Nale, J.Y., Spencer, J., Hargreaves, K.R., Buckley, A.M., Trzepinkski, P., Douce, G.R. and Clokie, M.R.J. 2016. Bacteriophage combinations significantly reduce *Clostridium difficile* growth in vitro and proliferation

in vivo. *Antimicrobial Agents and Chemotherapy*, 60(2), pp.968–981.

Nale, J.Y., Chutia, M., Carr, P., Hickenbotham, P.T. and Clokie, M.R.J. 2016b. "Get in early"; Biofilm and wax moth (Galleria mellonella) models reveal new insights into the therapeutic potential of *Clostridium difficile* bacteriophages. *Frontiers in Microbiology*, 7(1383), pp.1–16.

Ng, Y.K., Ehsaan, M., Philip, S., Collery, M.M., Janoir, C., Collignon, A., Cartman, A., Cartman, S.T. and Minton, N.P. 2013. Expanding the Repertoire of Gene Tools for Precise Manipulation of the *Clostridium difficile* Genome: Allelic Exchange Using pyrE Alleles. *PLoS ONE*, 8(2), pp. 1-13.

Normark, B.H. and Normark, S. 2002. Evolution and spread of antibiotic resistance. *Journal of Internal Medicine*, 252(2), pp.91–106.

Orth, P., Cordes, F., Schnappinger, D., Hillen, W., Saenger, W. and Hinrichs, W. 1998. Conformational changes of the Tet repressor induced by tetracycline trapping. *Journal of Molecular Biology*, 279(2), pp.439–447.

Orth, P., Cordes, F., Schnappinger, D., Hillen, W., Saenger, W. and Hinrichs, W. 2000. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature Structural Biology*, 7(3), pp.215–219.

Owens, R.C., Donskey, C.J., Gaynes, R.P., Loo, V.G. and Muto, C.A. 2008. Antimicrobial-Associated Risk Factors for *Clostridium difficile* Infection. *Clinical Infectious Diseases*, 46(s1), pp.S19–S31.

Pal, A. and Chattopadhyaya, R. 2009. RecA-mediated cleavage of lambda cI repressor accepts repressor dimers: probable role of prolyl cis-trans isomerization and catalytic involvement of H163, K177, and K232 of

RecA. *Journal of Biomolecular Structure & Dynamics*, 27(2), pp.221–233.

Patel, M., Jiang, Q, Woddgate, R., Cox, M.M. and Goodman, M.F., 2010. A New Model for SOS-induced Mutagenesis: How RecA Protein Activates DNA Polymerase V. *Critical Reviews Biochemistry and Molecular Biology*, 45(3), pp.171–184.

Paredes-Sabja, D., Shen, A. and Sorg, J.A. 2014. *Clostridium difficile* spore biology: Sporulation, germination, and spore structural proteins. *Trends in Microbiology*, 22(7), pp.406–416.

Planche, T. and Wilcox, M. 2011. Reference assays for *Clostridium difficile* infection: one or two gold standards? *Journal of Clinical Pathology*, 64(1), pp.1–5.

Ptashne, M. 2004. *A Genetic Switch: Phage Lambda Revisited* 3rd ed., New York: Cold Spring Harbor Laboratory Press.

Ptashne, M. 1967. Isolation of the λ Phage Repressor. *Proceedings of the National Academy of Sciences of the United States of America*, 57(2), pp. 306.

Public Health England, 2016. *Clostridium difficile* Ribotyping Network (CDRN) for England and Northern Ireland 2013-2015. Available at: https://www.gov.uk/government/publications/clostridium-difficile-ribotyping-network-cdrn-report.

Purdy, D., O'Keeffe, T.A.T., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A. and Minton, N.P. 2002. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular Microbiology*, 46(2), pp.439–452.

Rambaut, A. 2006. FigTree 1.4.3 - a graphical viewer of phylogenetic trees and a program for producing publication ready figures. Available at: http://tree.bio.ed.ac.uk/software/figtree/.

Ramesh, V., Fralick, J.A. and Rolfe, R.D. 1999. Prevention of *Clostridium difficile*-induced ileocecitis with Bacteriophage. *Anaerobe*, 5(2), pp.69–78.

Rao, K. and Young, V.B. 2015. Fecal Microbiota Transplantation for the Management of *Clostridium difficile* Infection. *Infectious Disease Clinics of North America*, 29(1), pp.109–122.

Rashid, S.J., Barylski, J., Hargreaves, K.R., Millard, A.A., Vinner, G.K. and Clokie, M.R.J. 2016. Two novel myoviruses from the north of Iraq reveal insights into *Clostridium difficile* phage diversity and biology. *Viruses*, 8(11), pp.1-18.

Reeves, A.E., Theriot, C.M., Bergin, I.L., Huffnagle, G.B., Schloss, P.D. and Young, V.B. 2011. The interplay between microbiome dynamics and pathogen dynamics in a murine model of *Clostridium difficile* infection. *Gut Microbes*, 2(3), pp.145–158.

Rice, P., Longden, I. and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, 16(6), pp.276–277.

Ripp, S. and Miller, R.V. 1998. Dynamics of the pseudolysogenic response in slowly growing cells of *Pseudomonas aeruginosa*. *Microbiology*, 144(8), pp.2225–2232.

Ripp, S. and Miller, R. V. 1997. The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology*, 143(6), pp.2065–2070.

Rokney, A., Kobiler, O., Amir, A., Court, D.L., Stavans, J., Adhya, S. and

Oppenheim, A.B. 2008. Host responses influence on the induction of lambda prophage. *Molecular Microbiology*, 68(1), pp.29–36.

Romano, V., Pasquale, V., Krovacek, K., Mauri, F., Demarta, A. and Dumontet, S. 2012. Toxigenic *Clostridium difficile* PCR ribotypes from wastewater treatment plants in Southern Switzerland. *Applied and Environmental Microbiology*, 78(18), pp.6643–6646.

Rutherford, K., Parkhill, J., Crook, J., Hornsell, T., Rice, P., Rajandream, M.A. and Barrell, B. 2000. Artemis: sequence, visualisation and annotation. *Bioinformatics*, 16(10), pp.944–945.

Saginur, R., Hawley, C.R. and Bartlett, J.G. 1980. Colitis Associated with Metronidazole Therapy. *The Journal of Infectious Diseases*, 141(6), pp.772–774.

Sambrook, J. and Russell, D.W. 2001. *Molecular Cloning: a laboratory manual*, New York: Harbour Cold Spring.

São-José, C., Baptista, C. and Santos, M.A. 2004. *Bacillus subtilis* operon encoding a membrane receptor for bacteriophage SPP1. *Journal of Bacteriology*, 186(24), pp.8337–8346.

Sára, M. and Sleytr, U.B. 2000. S-layer proteins. *Journal of Bacteriology*, 182(4), pp.859–868.

Savariau-Lacomme, M.P., Lebarbier, C., Karjalainen, T., Collignon, A. and Janoir, C. 2003. Transcription and analysis of polymorphism in a cluster of genes encoding surface-associated proteins of *Clostridium difficile*. *Journal of Bacteriology*, 185(15), pp.4461–4470.

Sayedy, L., Kothari, D. and Richards, R.J. 2010. Toxic megacolon associated *Clostridium difficile* colitis. *World Journal of Gastrointestinal Endoscopy*,

2(8), p.293-297.

Schubert, R.A., Dodd, I.B., Egan, J.B. and Shearwin, K.E. 2007. Cro's role in the CI – Cro bistable switch is critical for lambda's transition from lysogeny to lytic development. *Genes and Development*, 21, pp.2461–2472.

Seal, B. 2013. Characterization of bacteriophages virulent for *Clostridium perfringens* and identification of phage lytic enzymes as alternatives to antibiotics for potential control of the bacterium. *Poultry Science*, 92(2), pp.526–533.

Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeño-Tárraga, A.M., Wang, H., Holden, M.T.G., Wright, A., Churcher, C., Quail, M.A., Baker, S., Bason, N., Brooks, K, Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Fetlwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Robbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B. and Parkhill, J. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genetics*, 38(7), pp.779–786.

Sekulovic, O., Meessen-Pinard, M. and Fortier, L.C. 2011. Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. *Journal of Bacteriology*, 193(11), pp.2726–2734.

Sekulovic, O., Garneau, J.R., Néron, A. and Fortier, L-C. 2014. Characterization of temperate phages infecting *Clostridium difficile* isolates of human and animal origins. *Applied and Environmental Microbiology*, 80(8), pp.2555–2563.

Sell, T.L., Schaberg, D.R. and Fekety, F.R. 1983. Bacteriophage and bacteriocin

typing scheme for *Clostridium difficile*. *Journal of Clinical Microbiology*, 17(6), pp.1148–1152.

Setlow, P. 2014. Germination of spores of Bacillus species: What we know and do not know. *Journal of Bacteriology*, 196(7), pp.1297–1305.

Shan, J., Patel, K.V., Hickenbotham, P.T., Nale, J.Y., Hargreaves, K.R. and Clokie, M.R.J. 2012. Prophage carriage and diversity within clinically relevant strains of *Clostridium difficile*. *Applied and Environmental Microbiology*, 78(17), pp.6027–6034.

Shao, Y. and Wang, I.N. 2008. Bacteriophage adsorption rate and optimal lysis time. *Genetics*, 180(1), pp.471–482.

Solomon, K., Fanning, S., McDermott, S., Murray, S., Scott, L., Martin, A., Skally, M., Burns, K, Kuijper, E., Fitzpatrick, F., Fenelon, L. and Kyne, L. 2011. PCR ribotype prevalence and molecular basis of macrolide-lincosamide-streptogramin B (MLSB) and fluoroquinolone resistance in Irish clinical *Clostridium difficile* isolates. *Journal of Antimicrobial Chemotherapy*, 66(9), pp.1976–1982.

Sonenshein, A.L. 2006. Bacteriophages: How bacterial spores capture and protect phage DNA. *Current Biology*, 16(1), pp.14–16.

Stabler, R.A., Dawson, L.F., Phua, L.T.H. and Wren, B.W. 2008. Comparative analysis of BI/NAP1/027 hypervirulent strains reveals novel toxin B-encoding gene (*tcdB*) sequences. *Journal of Medical Microbiology*, 57(6), pp.771–775.

Stern, A. and Sorek, R. 2012. The phage-host arms-race : Shaping the evolution of microbes. *Bioessays*, 33(1), pp.43–51.

Sulakvelidze, A., Alavidze, Z. and Morris, J.G. 2001. Minireview:

Bacteriophage Therapy. *Antimicrobial Agents and Chemotherapy*, 45(3), pp.649–659.

Svenningsen, S.L., Costantino, N., Court, D.L. and Adhya, S. 2005. On the role of Cro in lambda prophage induction. *Proceedings of the National Academy of Sciences of the United States of America*, 102(12), pp.4465–4469.

Tapon, N. and Hall, A. 1997. Rho, Rac and Cdc42 GTPases regulate the organization of the actin cytoskeleton. *Current Opinion in Cell Biology*, 9(1), pp.86–92.

Theriot, C.M., Koenigsknecht, M.J., Carlson, P.E., Hatton, G.E., Nelson, A.M., Li, B., Huffnagle, G.B., Li, J.Z. and Young, V.B. 2014. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nature Communications*, 5(3114), pp.1-10.

Ticehurst, J.R., Aird, D.Z., Dam, L.M., Borek, A.P., Hargreaves, J.T. and Carroll, K.C. 2006. Effective detection of toxigenic *Clostridium difficile* by a two-step algorithm including tests for antigen and cytotoxin. *Journal of Clinical Microbiology*, 44(3), pp.1145–1149.

Tremblay, D.M., Tegoni, M., Spinelli, S., Campanacci, V., Blangy, S., Huyghe, C., Desmyter, A., Labrie, S., Moineau, S. and Cambillau, C. 2006. Receptor-binding protein of *Lactococcus lactis* phages: Identification and characterization of the saccharide receptor-binding site. *Journal of Bacteriology*, 188(7), pp.2400–2410.

Ventura, M., Zomer, A., Canchaya, C., O'Connell-Motherway, M., Kuipers, O., Turroni, F., Ribbera, A., Foroni, E., Buist, G., Wegmann, V., Shearman, C., Gasson, M.J., Fitzgerald, G.F., Gok, J. and van Sinderen, D. 2007.

Comparative analyses of prophage-like elements present in two *Lactococcus lactis* strains. *Applied and Environmental Microbiology*, 73(23), pp.7771–7780.

Venugopal, A.A. and Johnson, S. 2012. Fidaxomicin: A novel macrocyclic antibiotic approved for treatment of *Clostridium difficile* infection. *Clinical Infectious Diseases*, 54(4), pp.568–574.

Viscidi, R., Willey, S. and Bartlett, J.G. 1981. Isolation rates and toxigenic potential of *Clostridium difficile* isolates from various patient populations. *Gastroenterology*, 81(1), pp.5–9.

Voth, D.E. and Ballard, J.D. 2005. *Clostridium difficile* Toxins: Mechanism of Action and Role in Disease. *Clinical Microbiology Reviews*, 18(2), pp.247–263.

Wells, C.L. and Wilkins, T.D. 1996. Clostridia: Sporeforming Anaerobic *Bacilli*. In S. Baron, ed. *Medical Microbiology*. Galveston, Texas: University of Texas Medical Branch at Galveston, p. Chapter 18.

Wilcox, M.H., Hawkey, P.M., Patel, B., Blanche, T. and Stone, S. 2013. Updated guidance on the management and treatment of *Clostridium difficile* infection. *Public Health England*, pp.1–29.

Williams, R., Meader, E., Mayer, M., Narbad, A., Roberts, A.P. and Mullany, P. 2013. Determination of the *attP* and *attB* sites of phage CD27 from *Clostridium difficile* NCTC 12727. *Journal of Medical Microbiology*, 62(9), pp.1439–1443.

Wittebole, X., De Roock, S. and Opal, S.M. 2014. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. *Virulence*, 5(1), pp.226–235.

Wittmann, J., Riedel, T., Bunk, B., Spröer, C., Gronow, S. and Overmann, J. 2015. Complete Genome Sequence of the Novel Temperate *Clostridium difficile* Phage phiCDIF1296T. *Genome Announcements*, 3(4), pp.1–2.

Young, R. 1992. Bacteriophage lysis: mechanism and regulation. *Microbiological Reviews*, 56(3), pp.430–481.

Youngster, I., Russell, G.H., Pinder, C., Ziv-Baran, T., Sauk, J. and Hohmann, E.L. 2014. Oral, capsulized, frozen fecal microbiota transplantation for relapsing *Clostridium difficile* infection. *JAMA*, 312(17), pp.1772–1778.

Yutin, M. and Galperin, N. 2013. A genomic update on clostridial phylogeny: Gram-negative spore-formers and other misplaced clostridia. *Environmental Microbiology*, 15(1), pp. 2631-2641.

Zhang, S., Palazuelos-Munoz, S., Balsells, E.M., Nair, H., Chit, A. and Kyaw, M.H. 2016. Cost of hospital management of *Clostridium difficile* infection in the United States – a meta-analysis and modelling study. *BMC Infectious Disease*, 16(1), pp. 1-18.