Towards phage therapy against *Clostridioides difficile*

Zuzana Grmanova

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

September 2021

Declaration

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

Zuzana Grmanova

August 2021

Abstract

Clostridioides difficile is a serious public health threat and the rise of CDI epidemics over the last few years has triggered an increased interest in using phages as a form of therapy. Natural phage therapies rely on the lytic activity of phages to eradicate infection. To this date, no strictly lytic phage infecting *C. difficile* has been isolated. To address this issue, we propose to test the concept of using temperate phage to deliver lethal cargo to this pathogen.

We have isolated phages ΦCD3801, ΦCD8601 and ΦCD9301. Several locations in the ΦCD9301 genome have been identified and experimentally proven suitable for cargo sequence knock-in using an antibiotic marker. Initially, the construction of recombinant temperate phage carrying Small Acid Soluble Protein (SASP) as an antimicrobial against *C. difficile* was investigated. In vegetative cells, SASP binds to the DNA and interfere with DNA replication and gene expression leading to cell death. Two SASPs cargos were individually integrated into the ΦCD9301 genome, however no change in the phage activity was observed *in vitro*. Next, we proposed to test the concept of re-directing endogenous CRISPR-Cas systems towards toxic genes specific to this pathogen using recombinant phage ΦCD9301. Although, endogenous CRISPR-Cas was successfully re-directed to autoimmunity with the use of a plasmid, this result was not detected using recombinant phage carrying enCRISPR miniarrays.

Overall, this work shed some light on the generation of recombinant phages infecting *C. difficile*. Moreover, an endogenous CRISPR system of this pathogen was characterised and successfully redirected to autoimmunity. Further work is required to determine the optimal choice of promotor systems to achieve strong enough expression of cargo without the interference with the manufacturing strain.

Acknowledgement

I would like to thank my supervisors, Professor Nigel Minton and Dr. Ruth Griffin, for the opportunity to conduct this research, as well as for their guidance and help. I would also like to thank to the University of Nottingham for funding the project.

This work would not have been possible without the help and support of beautiful people I have met during this course. Thank you Cansu for all the yoga/food/gin sessions, Erik, Bashir, Chris & Jen for weekend breaks, Marcel for his homemade bread and will to listen to my constant complaints and of course thanks to everyone from the former A31 office for the advice, laugh and lunch breaks.

I would also like to thank my family. Mami, oci dakujem ze ste ma vzdy viedli k vytrvalosti. Bez Vas and Vasej podpory by som to nedala. Thanks to Jessie and Paul for giving me home in a foreign country.

Last but not least, I want to thank my partner Skyler for extraordinary patience and support. Thank you for being there for me.

Abbreviations

%	Percentage
°C	Degrees Celsius
µg	Micrograms
µl	Microlitres
µm	Micrometre
µM	Micromolar
5-FC	5-fluorocytosine
5-FOA	5-fluoroorotic acid
Abi	Abortive Infection
ACE	Allelic-Coupled Exchange
aTc	Anhydrous tetracycline
BHIS	Brain Heart Infusion Supplemented
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
CC	*Clostridioides difficile* Selective Supplement
CDI	*Clostridioides difficile* Infection
CFU	Colonies Forming Units
Cm	Chloramphenicol
cm	Centimetres
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
dH20	Deionised Water
DNA	Deoxyribonucleic Acid
ECDC	European Centre for Disease Protection and Control
EDTA	Ethylenediaminetetraacetic Acid
enCISPR	endogenous CRISPR
EOP	Efficiency of Plating
Erm	Erythromycin

EtOH	Ethanol
FDA	Food and Drug Administration
FMT	Faecal Microbiota Transplantation
g	Grams
G	Gravity
GBDP	Genome-BLAST Distance Phylogeny
GC	Guanine-cytosine
GI	Gastrointestinal
GOI	Gene of Interest
HMW	High Molecular Weight
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kbp	Kilo Base Pairs
kg	Kilograms
L	Litres
LB	Lysogeny Broth (Luria-Bertani broth)
M	Mole
mg	Milligrams
min	Minutes
ml	Millilitres
mM	Millimolar
mm	Millimetres
mRNA	Messenger RNA
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs
ng	Nanograms
nm	Nanometre
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
RAST	Rapid Annotation using Subsystem Technology
RBS	Ribosome Binding Site
RNA	Ribonucleic Acid
rpm	Rotations Per Minute
rRNA	Ribosomal RNA
s	Seconds
SASP	Small Acid Soluble Protein
SBRC	Synthetic Biology Research Centre
SDS	Sodium Dodecyl Sulphate
SNV	Single Nucleotide Variant
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

List of Figures

List of Tables

Table of Content

Chapter one:

General Introduction

1. General Introduction

1.1 *Clostridioides difficile*

Clostridioides difficile (formerly *Clostridium difficile*) is Gram-positive, spore-forming bacillus that has long been recognised to be the most common etiologic pathogen of antibiotic-associated diarrhoea worldwide (Crobach *et al*., 2018). It was originally described by Hall and O'Toole in 1935 when it was isolated as part of the stool flora from healthy new born infants.

The ability of *C. difficile* to produce spores contributes to its spread in the community. The spectrum of illness caused by *C. difficile* ranges in severity from asymptomatic carriage to small bowel enteritis, fulminant colitis, life-threatening pseudomembranous colitis and death (Napolitano and Edmiston, 2017). In spite of the fact that risk factors for *C. difficile* infections (CDI) include older people above 65 years of age, previous antibiotic treatment and prolonged hospitalization, recent studies have described the pathogen spreading further into the community causing an increase in the number of cases and severity (Kampouri *et al*., 2021).

Yutin and Galperin (2013) conducted phylogenetic reorganisation of the *Clostridia* and subsequently renamed *C. difficile* to *Peptoclostridium difficile.* More recently, the novel genus *Clostridioides* was proposed for *C. difficile* based on phenotypic, chemotaxonomic and phylogenetic analyses (Lawson *et al*., 2016).

1.1.1 *C. difficile* pathogenesis

Pathogenesis of virulent *C. difficile* strains in the gastro-intestinal tract (GI) is associated with spore germination, colonisation by vegetative cells, the production of toxins, sporulation and interaction with the gut microbiota that contributes to protection from disease (Abt *et al*., 2016; Zhu *et al*., 2018).

1.1.1.1 Toxins

C. difficile produces toxins that primarily target the epithelial cells of the colon leading to cell necrosis and disruption of epithelial membrane integrity, bowel inflammation and potentially life-threatening diarrhoea (Saavedra *et al*., 2018). The pathogen is

capable of producing at least three toxins, namely toxin A (TcdA), toxin B (TcdB) and the binary toxin, however the number of toxins and the amounts produced differ between *C. difficile* strains (Janezic *et al*., 2020).

TcdA and TcdB belong to the family of large clostridial glucosylation toxins which glucosylate Rho GTPases within the host cell cytosol. Toxin A is an enterotoxin because it causes fluid accumulation in the bowel. Toxin B does not cause fluid accumulation but is extremely cytopathic for tissue cells and is considered to be essential for *C. difficile* virulence (Lawson *et al*., 2016). They are both transcribed from a 19.6 kb pathogenicity locus (PaLoc) and are the major virulence factors causing symptoms of CDI (Rodriguez *et al.,* 2016). The PaLoc (Figure 1.1) is variably present among strains and generally contains five genes: the 2 toxin-encoding genes *tcdA* and *tcdB* and three regulatory genes – *tcdR*, that encodes a RNA polymerase sigma factor and plays an important role in upregulating transcription of toxin genes, *tcdC* which encodes a putative negative regulator of *tcdA* and *tcdB* transcription and *tcdE*, encoding a holin-like protein that is thought to be required for toxin secretion (Kodori *et al*., 2020).

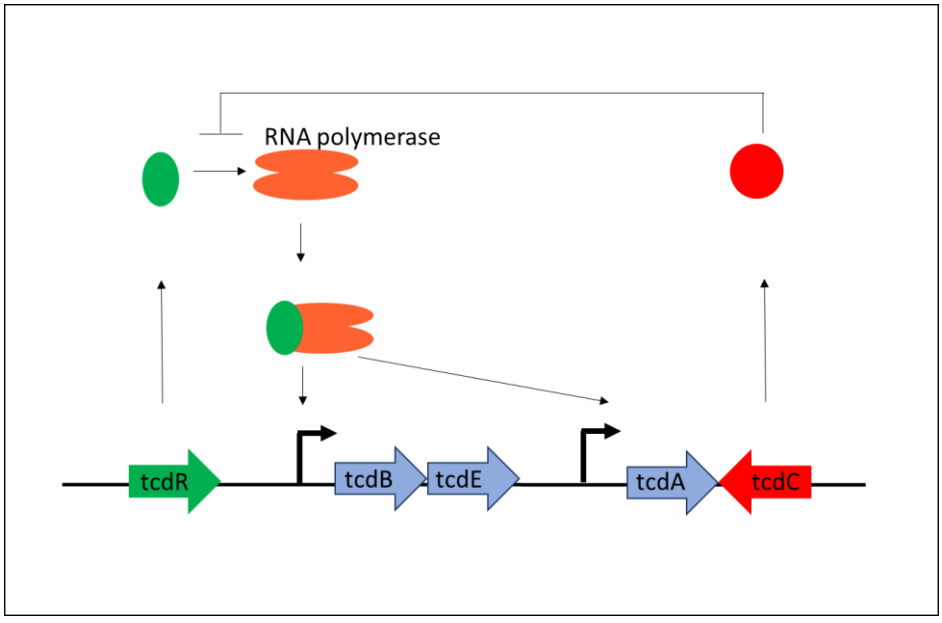

Figure 1. 1: Schematic representation of the pathogenicity locus of *C. difficile*.The PaLoc (19, 6 kb) contains genes *tcdA* and *tcdB* encoding toxins and the regulatory genes *tcdR*, *tcdC* and *tcdE*. *tcdR* encodes a positive regulator of toxin transcription, *tcdC* gene product is a putative negative regulator of transcription and *tcdE* encodes holin-like protein with a putative function in toxin release (Abt *et al*., 2016).

Binary toxin, also known as transferase toxin (CDT), is frequently observed in hypervirulent *C. difficile* strains such as PCR ribotype RT027 strains (Abeyawardhane *et al*., 2021). CDT belongs to the family of binary actin ADP-ribosylating toxins. Although, the precise mechanism remains unknown, it has been shown that CDT has a disruptive effect not only on the actin part of the cytoskeleton but also on the microtubule system, which leads to an increase in *C. difficile* adhesion, colonisation and inflammation of the colon (Abeyawardhane *et al*., 2021). The binary toxin is encoded by a 6.2 kb locus, CdtLoc. CdtLoc consists of toxin genes *cdtA* and *cdtB* and a gene *cdtR* encoding a regulatory protein (Kodori *et al*., 2020). cdtA ADP ribosylates actin, leading to disruption of actin cytoskeleton and the formation of microtubule protrusions. cdtB binds to the host cell and is necessary for ctdA internalisation (Gerding *et al*., 2014).

The orphan response regulator, CdtR, encoded within the CdtLoc, up-regulates CDT production. Moreover, in the case of some clinically relevant strains it has been shown that CdtR has a positive effect on *tcdA* and *tcdB* transcription (Lyon *et al.,* 2016). The number of strains carrying the CdtLoc has increased significantly over the last ten years with 6% of clinical isolates encoding CDT in 2004 compared to 33.5% of isolates now (Lyon *et al*., 2016).

1.1.1.2 Sporulation

The ability of *C. difficile* to sporulate contributes to its transmission, allowing it to disseminate through patient-to-patient contact and environmental contamination (Edwards and McBride, 2014). *C. difficile* sporulation is controlled by sensor histidine kinases that phosphorylate and activate transcription factor stage 0 sporulation protein A (Spo0A) (Martins *et al*., 2021).

Upon entry into the small intestine, spores germinate to become metabolically active vegetative cells. Germination of *C. difficile* spores is an important step for initiating CDI and occurs only in the lower GI because of the low level of oxygen and presence of substances such as bile acids that induce germination of the spore into an actively replicating vegetative cell (Koenigsknecht *et al.,* 2015). Bile acids play a dynamic and crucial role for both pathogenesis and the infectious cycle of *C. difficile*. Importantly, bile acids differ in their ability to promote germination and to influence vegetative

growth. The bile salt, taurocholate, is the primary germinant for *C. difficile* spores with amino acids, including glycine and histidine, serving as co-germinants which enhance germination in the presence of taurocholate (Kochan *et al*., 2018). Notably, the secondary bile acids lithocholic and ursodeoxycholic acids have been shown to have an inhibitory effect on *C. difficile* germination *in vitro* (Winston *et al*., 2020). The bile acid receptor CspC, located in the spore coat was identified to play a crucial role in *C. difficile* germination process, with CspC mutant spores being unresponsive to taurocholate (Francis *et al*., 2013).

1.1.2 *Clostridium difficile* infection

C. difficile is the most frequently reported nosocomial pathogen. In 2017, there were 223,900 cases and around 12 800 deaths associated with CDI in USA (2019 AR Threats Report). The first European study estimated that approximately 124,000 patients developed health-care-associated CDI within the European Union in 2011 and 2012. In England, the infection rate represents 14,139 infections per year (Herbert *et al*., 2019). Nosocomial CDI is a burden on hospitalisation expenditures, increasing annual cost by more than $4 billion in the United States (Lessa *et al*., 2015) In Europe, the economic burden was estimated to be €3000 million per year (Jones *et al*., 2013).

The incidence of CDI among hospitalized patients varies widely from year to year and in different locations but has generally been increasing (Leffler and Lamont, 2015). In the early 2000s, hospitals in North America and Europe began reporting dramatic increases in severity and incidence of CDI, which were associated with PCR ribotype RT027 (He *et al.,* 2013, Couturier *et al*., 2018) and PCR ribotype RT078 (Álvarez-Pérez *et al*., 2017). A similar pattern of increased incidence was also observed in Australia, Asia and Central America (Clements *et al*., 2010). In addition, CDI has been reported increasingly outside of the hospital environment, including in community and nursing homes settings. Increased production of toxins, fluoroquinolone resistance and production of binary toxin are the main factors contributing to outbreaks of CDI caused by hypervirulent PCR ribotype RT027 strains (Brown and Wilson, 2018; Kampouri *et al*., 2021).

Normally, the microbiota of healthy individuals provides a protective barrier against *C. difficile* colonisation by: (1) physically occupying the gut epithelial layer; (2) competing for nutrients, and; (3), converting primary bile salts required for germination of *C. difficile* spores into secondary bile salts that inhibit germination (Vogt and Finlay, 2017; Czepiel *et al*., 2019). Disruption of the gut flora by antibiotic exposure has a negative effect on the physical protective barrier and can result in changes in the composition of bile salts by the elimination of key organisms such as *Clostridium scindens* that convert primary bile salts to secondary, permitting *C. difficile* spore germination, and colonisation (Kachrimanidou and Tsintarakis, 2020). As *C. difficile* spores germinate into metabolically active vegetative cells, the secretion of toxins leads to disruption of the cytoskeletal system of epithelial cells and loss of the tight junctions between cells leading to increased epithelial permeability (Aktories *et al*., 2017).

Symptoms of CDI in patients include mild diarrhoea, abdominal pain, fever and leucocytosis. If symptoms progress, CDI patients can develop life-threatening pseudomembranous colitis and toxic megacolon which can lead to death (Rohana *et al*., 2020).

CDI relapse is characterized by the return of symptoms within eight weeks of primary diagnosis after convalescence or apparent recovery. CDI recurrence after initial treatment occurs in up to 25% of treated patients (Ramsay *et al*., 2018).

Recurrent CDI can be explained by the resistance of *C. difficile* spores towards antibiotic treatment and the negative effect of antibiotics on the host microbiota enabling germination of the spores (Budi *et al.*, 2020). Another possible explanation is a new exogenous infection via spore contamination of the local environment (Castro-Córdova *et al*., 2021). Therefore, relapse and reinfection may be difficult to distinguish. Moreover, external factors such as the age of the patient, use of other antibiotics, and prolonged hospital stay may also increase the risk of recurring infection (Ramsay *et al*., 2018).

1.1.2.1 Current CDI treatment options

Treatment of CDI has become a rising problem due to the increasing occurrence of hypervirulent strains, severity, and high frequency of recurrence. Currently, the standard treatment of CDI involves antibiotic therapy, with first occurrences treated with metronidazole or vancomycin (Jarmo *et al*., 2020).

Despite their effectiveness, treatment of CDI with either metronidazole or vancomycin has been associated with high risk of relapse due to their broad spectrum of activity. Specifically metronidazole targets anaerobic bacteria while vancomycin targets Gram-positive bacteria. These antibiotics have a disruptive effect on the gut microbiota compromising natural protection against *C. difficile* and resulting in high recurrence rates (Jarrad *et al*., 2015).

Fidaxomicin is a newer antimicrobial being used to treat CDI and is similar to vancomycin in treatment efficacy. Despite lowest rate of relapse, the high cost of this antibiotic precludes its routine use in comparison with metronidazole and vancomycin (Cho *et al*., 2020).

Due to the high recurrence level of CDI, novel antibiotics have been studied such as Nitazoxanide, Cadazolid or Ridinilazole. Although, clinical trials have shown promising results, the cost often limits their use for primary CDI (Dieterle *et al*., 2019).

Despite the promising trials of new antibiotics associated with the treatment of CDI, there is a need for further studies to define their clinical utility. Therefore, alternative therapeutic strategies for CDI have been investigated.

Probiotics aim to provide beneficial microbes into the GI tract to recolonise the colon and prevent invasion by pathogens. Common microbes used in probiotic formulations include *Saccharomyces boulardii*, *Lactobacilli*, *Clostridia*, *Streptococci* and *Bifidobacteria* (O'Horo *et al*., 2013).

In the case of recurrent CDI, *S. boulardii* has the strongest evidence supporting its beneficial effect found in two placebo-controlled studies where CDI patients showed 35% less recurrence than the control group. However, due to inadequate randomisation of antibiotics used in this study, the exact effectiveness of *S. boulardii* remains unclear.

Although, probiotics provide potential benefits in the treatment of CDI, to this date, there is not enough data to support their efficacy and there are some drawbacks related to their use. Unlike faecal microbiota transplantation, probiotics typically contain only a limited number of microbial species and since they are dietary supplements, the lack of regulation in their manufacturing process can lead to the production of supplements that are not evaluated for safety (Yip *et al*., 2017).

Several different passive immunisation strategies have been developed and tested over the past two decades with an aim to tackle CDI, such as (I) the development of polyclonal and monoclonal antibodies against toxins A and B, (II) the development of antibody subunits against toxins, and (III) fragment antibodies against surface protein.

To date, bezlotoxumab (ZinplavaTM) is the only antibody-based product used against CDI approved for use in the USA and Europe. Systematically administering the antibody results in a decrease in the recurrence rates of CDI, however it does not provide a locally oriented and long lasting protection (Bruxelle *et al*., 2018).

Faecal microbiota transplantation (FMT) is a treatment method that involves the introduction of a stool from a healthy donor into the colon of the patient to restore a healthy gut microbiome. FMT has recently emerged as an accepted, safe and effective treatment for recurrent CDI and studies have shown a cure rate greater than 90% after 90 days of FMT, following a course of vancomycin treatment (Kassam *et al*., 2013).

Bacteriocins have also been investigated as an alternative or complementary CDI treatment. Nicin, produced by lactic acid bacteria have shown promising potential in inhibition of *C. difficile* vegetative cells and spore germination *in vitro*. More *in vivo* experiments are still required before the possible application of nisin as a therapeutic agent for the treatment of CDI (Le Lay *et al*., 2016).

1.2 Bacteriophages

Bacteriophage (phage) research dates back to 1915 when Frederick Twort first discovered most of the essential features of phages and hypothesised this was either

a manifestation of the bacterial life cycle, an enzyme produced by the bacteria, or some sort of "ultra-microscopic virus" (Twort, 1915). Unable to confirm this hypothesis, he called it a bacteriolytic agent and his discovery was ignored. In 1917, Félix d'Herelle independently discovered ultraviruses that invade bacteria and multiply at their expense. D'Herelle called those ultraviruses phages and introduced the concept of phage therapy for the first time (d'Herelle, 1917).

Phages are the most abundant organisms on Earth with an estimated 10^{31} phage particles on the planet (Mushegian, 2020). Phages have been isolated from every environment colonised by bacteria such as sewage, soil, water and it is believed that at least one type of bacteriophage infects every bacterial strain (Hyman, 2019).

Phage particles vary in size, shape, complexity and genetic information with the size of phage particles ranging from 20 to 200 nm in length (Sinha *et al*., 2018). Phage genomes differ in size from 3.4 kb to 735 kb, and unlike bacteria, there is no single gene conserved in all phage genomes. In fact, phage genomes are highly mosaic representing different combinations of the same modules that have been shuffled and reshuffled over billions of years (Keen, 2015; Al Shayeb *et al*., 2020).

A phage particle consists of a capsid (head), that can be polyhedral (*Microviridae*, *Corticoviridae*, *Tectiviridae*, *Leviviridae* and *Cystoviridae*), filamentous (*Inoviridae*), pleomorphic (*Plasmaviridae*) or connected to a tail (*Caudovirales*). To date, most isolated phages are tailed and have dsDNA genomes (Figure 1.2).

The capsid is a protein coat that encapsulates the genetic information, either DNA (double stranded or single stranded) or RNA (Dion *et al*., 2020). The tail structure contains a hollow tube through which phage nucleic acid passes into the host during infection and tail fibres that allow attachment of phage to the surface of the bacterial cell (Tamma and Suh, 2021).

To date, the most well characterised phages in the literature belong to the order *Caudovirales* and are double stranded DNA tailed phages. Based on their tails the order *Caudovirales* is divided into five families: (1) *Siphoviridae* have a long flexible tail, (2) *Myoviridae* are characterised by a contractile tail consisting of two layers, (3) *Podoviridae* have very short tails, (4) *Ackermannviridae* have a morphology similar to

Myoviridae with tail spikes at the base of the tail and (5) *Herelleviridae* although an official family, share the same morphology as *Myoviridae* (Dion *et al*., 2020).

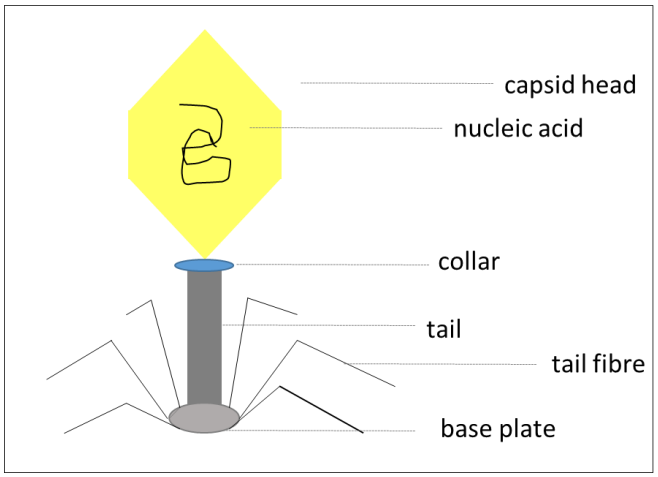

Figure 1.2: Schematic structure of typical tailed phage. Phage particle consists of an icosahedral head made of protein and nucleic acid (DNA or RNA), a collar and the tail that is responsible for the recognition of the host cell and the viral genome delivery. Finally, tailed phages also possess a base plate with tail fibres used for attachment to bacterial cells.

1.2.1 Life cycle of bacteriophages

Most phages display two types of life cycle: the lytic cycle and the lysogenic cycle (Figure 1.3). In the lytic cycle, phage particles attach to the bacterial surface through various cell surface receptors and inject their nucleic acid into the cytosol. The phage nucleic acid hijacks the bacterial machinery and causes it to divert from its normal tasks in bacterial growth to replicate phage nucleic acid. The phage exploits the host machinery to express its genes encoding the outer components such as the capsid and phage particles are assembled. Once a threshold level of mature viral particles are formed, they cause the bacterial cell to burst open. The released phage particles go on to infect more bacterial cells (Kortright *et al*., 2019).

Unlike strictly lytic phages, temperate phages can enter both lytic and lysogenic life cycle. The lysogenic cycle is characterised by the integration of phage nucleic acid into the genome of bacteria or by forming an extrachromosomal episome within the host cell. In this state the phage is referred to as a prophage and is replicated along with bacterial genome and passed on to the daughter cells. Once in a state of lysogeny, phages can remain within their hosts for many generations. Under certain

conditions, such as high stress, the phage genome is excised and temperate phage re-enter a lytic life cycle (Ofir and Sorek, 2018).

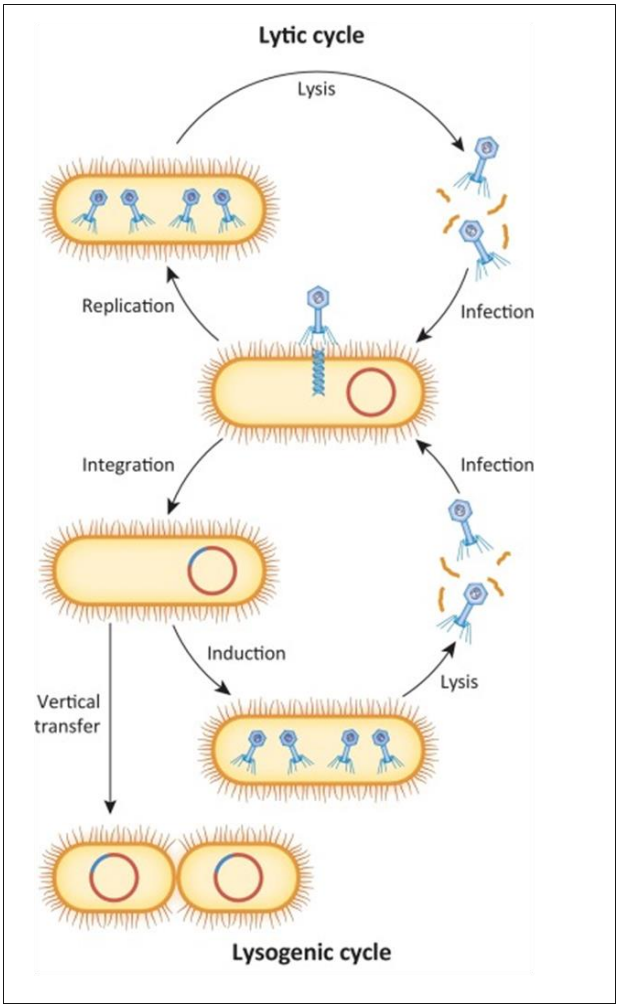

Figure 1.3: Schematic diagram of life cycle of temperate phage λ. During the lysogenic cycle, the phage genome integrates into the host cell genome and replicates passively in the form of a prophage. In the lytic life cycle the phage replicates independently of the host chromosome, and assembles into new viral particles which result in cell lysis (Keen and Dantas, 2018).

Temperate phages can affect bacterial populations in both a positive and negative manner. For example, phage can provide bacteria with protection against infection by other phages or can mediate horizontal gene transfer and subsequent gain of specific features that can be advantageous to bacteria such as antibiotic resistance. Conversely, temperate phages can alter host gene expression or even kill host bacteria upon induction (Howard-Varona *et al*., 2017).

A temperate phage needs to decide which life cycle pathway to choose every time it infects a bacterial cell. This decision was thoroughly investigated in the *E. coli* phage lambda. After the discovery of the Cro protein it was first thought that the lysis/lysogeny decision might simply be explained by a direct competition for operator binding between CI and Cro proteins. We now know that it is a complex process that includes transcriptional repressors and transcriptional activators, as well as RNA degradation, transcription anti-termination, and proteolysis. The decision-making process also integrates information on the metabolic state of the host, host density and the number of infecting phages per cell (Ofir and Sorek, 2018; Benler and Koonin, 2020).

While the lambda lysis-lysogeny switch was studied for decades, the decision-making process in other phages remained almost completely unknown. In contrast to the decision-making network of lambda phage, the mycobacteriophage system seems to be less complex and involves only genes for repressor, Cro, and integrase. The integration site on the phage genome (*attP*) lies within the repressor gene and integration into the bacterial genome removes the proteolytic degradation signal from the C terminus of the repressor, resulting in more stable repressor protein and maintaining lysogeny. In the lytic cycle, both the repressor and the integrase proteins are degraded by cellular proteases, preventing lysogeny (Villanueva *et al*., 2015).

A more recent study that examined the lysogeny dynamics of the *Bacillus* phage phi3T has shown that phages use the arbitrium communication system to decide whether to employ the lytic or lysogenic life cycle. According to Erez and colleagues, phages use a small peptide, six amino acids in length, to communicate with their progeny. Progeny phages measure the concentration of this peptide and choose to infect either through the lytic or the lysogenic cycle based on the concentration. In later stages of the infection decreased population of bacteria is a risk for phage progeny. In this situation, after several cycles of infection, arbitrium peptide is accumulated in the medium and its higher concentration is a signal for progeny phages to enter the lysogenic life cycle with the aim to preserve chances for viable reproduction (Erez *et al.,* 2017).

1.2.2 Phage therapy

1.2.2.1 History of phage therapy

Phage therapy dates back to the discovery of phages in 1917, when d'Hérelle suggested their therapeutic use as antibacterial agents (d'Hérelle, 1917). Shortly after his discovery, d'Hérelle successfully treated dysentery using phages (d'Hérelle and Smith, 1922). Another study using phages to treat infections in humans was published in 1921 when Bruynoghe and Maisin used phage therapy to treat staphylococcal skin disease. Several more studies have followed (Rice, 1930; Schless, 1932). Soon after, d'Herelle started producing phage preparations against various bacterial infections commercially. They were marketed by what later became the large French company L'Oreal (Summers, 1999). In the 1940s, American company Eli Lilly produced several phage products for human use targeting pathogens such as staphylococci, streptococci and *Escherichia coli* (Abedon, 2017).

During World War II and with the discovery of antibiotics, the commercial production of therapeutic phages declined in most of the Western world, especially the USA. However, despite the increasing popularity of antibiotics the use of phages as the treatment for bacterial infections has persisted in various regions of the world and continues to take place to this day in Eastern Europe and in the former Soviet Union (Abedon, 2017). Several institutions in these countries play an important role in phage therapy research such as the Eliava Institute of Bacteriophage, Microbiology, and Virology (EIBMV) in Georgia, and the Hirszfeld Institute of Immunology and Experimental Therapy (HIIET) in Poland (Żaczek *et al*., 2020).

Over the last decade, interest in phage therapy has seen a significant increase with the threat of multi-drug resistant (MDR) infections and concerns that humankind is entering a post-antibiotic era (Luong *et al*., 2020).

1.2.2.2 Benefits of phage therapy

Lytic phages, display remarkable antibacterial activity and have many advantages over antibiotic use. Although, some phages have shown to be able to infect more than one relatively closely related bacterial genus (Hamdi *et al*., 2017), this phenomenon is quite rare with most phages being specific to one species or strain.

The highly specific nature of phages is one of their biggest advantages as a therapy. Unlike antibiotics, which tend to have a broad spectrum of activity, phages have minimal impact on host and health-protecting commensal bacteria of the gut flora, reducing the possibility of opportunistic infections (Abedon, 2017; Hyman *et al*., 2019).

Many *in vivo* animal studies have proven successful use of phages in the treatment of lung infections (Waters *et al*., 2017; Anand *et al*., 2020), GI infections (Xue *et al*., 2020) or localised infections (Kifelew *et al*., 2020), however, reports on the clinical use of phages have been relatively limited. Leszczyński and colleagues used phage to eliminate gastrointestinal colonization by methicillin resistant *S. aureus*. The patient received a cocktail of three phages three times a day after gastric acid neutralisation. After one week of the treatment rectal swabs showed elimination of the pathogen (Leszczyński *et al.* 2006). Another Polish study has demonstrated successful eradication of *Enterococcus faecalis* prostate infections which had previously resisted antibiotic treatment (Letkiewicz *et al*., 2009). More recently, a phage cocktail was administered to a patient with necrotizing pancreatitis complicated by an MDR *Acinetobacter baumannii* infection. Administration of phages intravenously and percutaneously into the abscess cavities was associated with clinical improvement and clearance of the infection (Schooley *et al*., 2017). Another example is a cystic fibrosis (CF) patient, with a multi-drug resistant MDR *Achromobacter xylosoxidans* lung infection who was successfully managed with phage cocktail administered both orally and by inhalation (Lebeaux *et al*., 2021).

It is also believed that phage therapy is significantly safer and better tolerated, compared to antibiotics, as they replicate only in the target bacterium but cannot infect mammalian cells. This conclusion seems supported by studies conducted in Eastern Europe as well as recent experimental use in animals and humans, which have not reported significant adverse effects following phage administration (Principi *et al*., 2019).

Another advantage of phage therapy is biofilm degradation with several studies proving phage ability to penetrate the biofilm structure and killing biofilm associated

cells that are difficult to target by conventional antibiotic (Azeredo and Sutherland, 2008; Alves *et al*., 2016, Bhardwaj *et al*, 2020).

Since phages replicate *in vivo* where hosts are located, very few doses are required in the site of infection after the initial administration due to their ability to increase in density. This could have a positive impact on the cost of treatment as well as the safety of the product (Principi *et al*., 2019).

Furthermore, new phages are usually easy to isolate, often from soil, water, sewage and other waste materials that contain high concentrations of bacteria. Phages are regarded to have a low environmental impact, with discarded phages being easily inactivated by environmental factors such as sunlight or extreme temperatures (Chanishvili and Aminov, R. 2019).

1.2.2.3 Phage therapy drawbacks

Despite the promising research done to this date, there are some limitations that should be addressed before phage therapy can be fully accepted in modern clinical practice. One of the possible drawback is bacterial resistance against the phage. There are several mechanisms bacteria have evolved to protect against a phage infection, such as blocking of phage adsorption by mutation of the bacterial receptor, restriction modification system or by CRISPR-Cas adaptive immunity (Oeschlin, 2018). It is believed that, although bacteria can develop phage resistance, it might be easier to overcome than resistance to antibiotics due to phage-bacteria coevolution. Just as bacteria can develop resistance, phages can develop mechanisms to overcome this resistance by various mechanisms including adaptation to new receptors and masking and modifying restriction sites or production of anti-CRISPR proteins (Orzechowska and Mohammed, 2019; Benler and Koonin, 2020). Also, bacterial resistance to phage is expected to have a less severe effect compared to their resistance to antibiotics, this is mainly due to the enormous amounts and the great diversity of phages (Roach and Debarbieux, 2017).

To conclude, the development of phage-resistance is a possible risk of phage therapy that can be easily overcome by use of a different phage or by applying a preparation containing multiple types of phage. Another option would be to use a combination of antibiotics and phage therapy (Kortright *et al*., 2019).

The minimal impact of phage on the microbiome is an appealing feature of phage therapy. However, the benefits of narrow host range should be balanced with a cost associated with identifying the appropriate phage prior to initiation of treatment. There is often no need to determine susceptibility of target bacteria before broad-spectrum antibiotics are used, which makes them more convenient to prescribe (Abedon *et al*., 2017).

Development of phage preparations that possess a broad spectrum of activity is more difficult because phages are often not effective against all strains, even within a single bacterial species. One of the options to tackle this issue is the use of multiple phages. Formulation of phages into a cocktail can substantially broaden the spectrum of activity and increase their potential (Kortright *et al*., 2019). Moreover, it is possible to develop polyphage treatments targeting multiple bacterial species that makes it possible to treat infections without first identifying the target bacterial species. It has been reported that scientists at the Eliava Institute produced a cocktail containing hundreds of different phages targeting more than 28 species of bacteria (Kutter *et al.*, 2014).

One of the major concerns associated with clinical phage therapy is the interaction of phages with the immune system. Roach and colleagues compared the efficacies of phage application in a variety of immunocompetent and immunocompromised mouse models suffering with respiratory tract infections caused by *Pseudomonas aeruginosa*. According to their results, phage therapy was effective in healthy mice and phages were working in synergy with their immune system to eradicate infection however, phage therapy did not work in mice depleted of neutrophils. Furthermore, they observed that phages were well tolerated by host organisms and were not significantly neutralized by immune defence mechanisms (Roach *et al*., 2017).

To this date, there have been no reports of phages causing disease in humans or animals, however, more investigation is needed to determine how innate and adaptive immune systems respond to phages (Principi *et al*., 2019).

Other concerns associated with phage therapy include the risk of phages carrying genes that encode harmful products, such as toxins and their potential to transduce

DNA between bacteria. Both of those issues can be resolved by employing strictly lytic phages and by characterisation of phages before selection (Principi *et al*., 2019).

1.2.2.4 Synthetic phages

With an advance in sequencing and genetic engineering, design and production of synthetic phages has become possible. Many studies have described the development of recombinant phages with broader host range (Ando *et al*., 2015; Gebhart *et al*., 2017; Heilpern and Waldor, 2003; Lin *et al*., 2012; Scholl *et al*., 2009; Yosef *et al*., 2017). For instance, Li and colleagues generated hybrid T4-like phages by changing the receptor specificity determinant domain of gene 37. Comparing to the wild type, mutant phages exhibited expanded host range and antimicrobial activity against pathogenic *E. coli* (Li *et al*., 2020).

Synthetic biology can also be used to enhance the phage ability to eradicate biofilms. T7 bacteriophage was engineered to express a lactonase enzyme that interferes with bacterial quorum sensing necessary for biofilm formation (Pei *et al*., 2014). More recently, recombinant phage Y2 was engineered to express the depolymerase gene *dpoL1-C* from phage L1 (Born *et al*., 2017).

Although, phages that undergo lysogeny are generally avoided in traditional phage therapy, the advance in synthetic biology made the creation of lytic variants of temperate phages possible. For example, Zhang *et al*., 2013 created a lytic version of the *E. faecalis* phage by deleting the lysogeny module in the genome of phage ΦEf11, and by substituting the *cro* repressor with a nisin-inducible repressor. The mutant phage was incapable of lysogeny and had a significantly extended host range. Temperate phages infecting *Mycobacterium abscessus* were also engineered to produce a lytic derivative, by total or partial removal of the repressor gene required for lysogeny. Moreover, the phage cocktail was administered intravenously to the patient with mycobacterial infection. This is the first case of phage therapy in humans using recombinant phages. The use of this phage cocktail was associated with clinical improvement and no significant side effects (Dedrick *et al*., 2019).

1.2.3 *C. difficile* phages

Phages infecting *C. difficile* were isolated for the first time in 1983 by Sell *et al.* (1983) for the purpose of *C. difficile* typing. Soon after, different research groups focused on *C. difficile* phages either to study transmission and epidemiology of various *C. difficile* strains (Hawkins *et al*., 1984; Bacon *et al*., 1988) or the impact of integrated prophages onto bacterial phenotype and virulence (Nagy and Foldes, 1991).

Several phages infecting *C. difficile* have been isolated either from bacterial lysogens or as free agents from supernatants. Although the level of characterisation between isolated phages vary, all of them carry elements associated with lysogeny control including integrases, repressors and anti-repressors (Hawkins *et al*., 1984; Mahony *et al*., 1985, 1991; Bacon *et al*., 1988; Ramesh *et al*., 1999; Goh *et al*., 2005; Sekulovic *et al*., 2014, Ramirez-Vargas *et al*., 2018).

To this date, no strictly lytic *C. difficile* phage has been isolated. It is thought this could be due to the strictly anaerobic lifestyle of *C. difficile* and its ability to form spores in the environment (Meader *et al.,* 2013; Heuler *et al*., 2021) however, strictly lytic phages have been previously isolated for other spore forming bacteria such as *Bacillus cereus* (El-Arabi *et al*., 2013). Large scale screenings with diverse bacterial strains appear to be important for the detection of lytic phages suggesting that *C. difficile* lytic phages are yet to be discovered (Hargreaves and Clokie, 2014).

Usually, the genome of *C. difficile* carries between 1 and 3 prophages, however, as many as 5-6 prophages were previously identified in a single *C. difficile* genome (Amy *et al*., 2018; Ramírez-Vargas *et al*., 2018). Recent studies pointed at the high prevalence of large phage genomes that reside as extrachromosomal DNA in *C. difficile*. The genome of large phiCD211 and related phages, with a size of ≥131 kb, have been detected in 5% of analysed 2,584 *C. difficile* genomes (Wittmann *et al*., 2015; Garneau *et al*., 2018). Ramírez-Vargas *et al*., 2018 described ten other phage genomes with size ranging between 128–135 kb in *C. difficile* isolates from around the world.

All 29 *C. difficile* phages that have been sequenced at the time of writing (June, 2021) carry dsDNA and belong to either the *Myoviridae* or the *Siphoviridae* family of the order *Caudovirales* (Appendix, Table 7.3). Based on the morphology of phage particle,

C. difficile phages can be divided into 4 groups: the long tailed myoviruses (LTMs), the medium myoviruses (MMs), the small myoviruses (SMVs) and morphologically distinct siphoviruses (SVs) (Hargreaves and Clokie, 2014). More recently, genome analysis enabled classification of *C. difficile* phages into two genera, *phiCD119-like* phages and *phiMMP04-like* phages. Some phages, however, do not fall into either genera (Rashid *et al*., 2016). The size of their genomes ranges from ~31 to ~135 kbp with similar GC content between 26% and 31% (Heuler *et al.*, 2021).

The majority of *C. difficile* phages have relatively narrow host ranges, with most phages infecting few strains (Goh *et al*., 2005b; Sekulovic *et al*., 2011, 2014; Rashid *et al*., 2016) The interpretation of phage host range, however, depends on the number of tested *C. difficile* strains and their diversity (Heuler *et al*., 2021). The narrow host range of phages may be explained by the high prevalence of prophages in *C. difficile* genomes, the availability of a suitable host receptor or the presence of multiple CRISPR arrays in *C. difficile* (Boudry *et al*., 2015). Another possibility is the superinfection exclusion system. Sekulovic and colleagues (2015) have demonstrated that *C. difficile* cell surface protein CwpV blocks replication of phage DNA similarly like superinfection exclusion systems normally encoded on prophages (Sekulovic *et al*., 2015).

The impact of prophages on the physiology and virulence of *C. difficile* is a topic of great interest, considering their high prevalence (Fortier, 2018). Many *C. difficile* prophages encode genes that are suspected to affect their host. For example, some large *phiCD211-like* phages have been shown to carry spore proteases, putative multidrug resistance genes, and multiple regulators that could interfere with host regulation of gene expression (Garneau *et al*., 2018).

The presence of CRISPR arrays has also been identified in several *C. difficile* phages, (Hargreaves *et al*., 2014a; Boudry *et al*., 2015, Sekulovic and Fortier, 2015, Garneau *et al*., 2018). A partial transposase gene has been identified upstream of some CRISPR array regions, suggesting that that these have been transferred via horizontal gene transfer events (Hargreaves *et al*., 2014a). Transcriptomic analyses showed that these CRISPR arrays are transcribed and thus suggest that specific prophages in *C.*

difficile could contribute to *C. difficile* immunity to invading DNA (Boudry *et al*., 2015; Sekulovic and Fortier, 2015).

Moreover, several research groups have reported that *C. difficile* phages can modulate production of toxins when integrated into the bacterial genome, although the impact was variable (Goh *et al*., 2005; Govind *et al*., 2009; Sekulovic *et al*., 2011).

So far, prophages seem to impact *C. difficile* in various ways, depending on the genetic background of the host. A better understanding of the phage-host interaction, as well as the impact on the whole microbiome, will be crucial for future phage therapy applications (Fortier *et al*., 2018). For example, phages naturally occurring in faecal microbiota samples seems to contribute to the successful treatment of recurrent CDI during faecal microbiota transplantation, however, more data is needed to fully understand interplay between the virome and the microbiome (Zuo *et al*., 2018).

1.2.3.1 *C. difficile* phage therapy

The rise of CDI epidemics over the last few years has triggered an increase of interest in using phages as a form of treatment. There are several benefits related with the use of phage therapy against CDI. Unlike antibiotics, phages have a highly specific bacterial host range targeting *C. difficile* while leaving other commensal bacteria intact. This is particularly important when treating CDI because it has been shown that the gut microbiota provides a natural protection against CDI (Cornely *et al*., 2012). Moreover, bacterial biofilms are often resistant to antibiotic treatment and *C. difficile* aggregates have been observed on the surface of the caecum and colon *in vivo* (Buckley *et al*., 2011). Importantly, the nature of CDI makes phage therapy a possible option, as there should not be many problems relating to the delivery of phages to the physical location of CDI (Hargreaves and Clokie, 2014).

The current knowledge regarding the use of phages to treat CDI is limited. In 1999, the first *in vivo* model was developed to test the effect of phage ΦCD140 on ileocecitis in hamsters. In this study, one group of hamsters were administered clindamycin to disrupt the gut microbiota and were subsequently challenged with 10^3 CFU.ml^{-1} *C. difficile* followed by a single dose of phage (10^8 PFU.ml^{-1}). Another two

groups of hamsters were additionally treated with phage immediately after *C. difficile* challenge and every 8 h up to 48 and 72 h, respectively. Promisingly, administration of phages protected hamsters in all three groups, whereas control animals died within 96 hours. Despite the promising results, phage treatment did not fully clear the bacterial infection and the longevity of phage protection was not proved, possibly due to stomach acid interference with the phage preparation (Ramesh *et al*., 1999).

Phage treatment of *C. difficile* was also demonstrated *in vitro* using a batch fermentation system. In this model, the effect of phage ΦCD27 on pathogen viability and toxin production was tested in remedial and prophylactic treatment regimes. While prophylactic phage treatment resulted in a significant decrease of *C. difficile* CFU.ml^{-1} and abolished toxin production, the remedial therapy regime led to a short-term decrease in viable bacteria and reduced production of toxins. As expected, both regimes did not have any significant detrimental effect on any other bacterial commensals colonising the gut. This study also showed integration of ΦCD27 into the host genome after treatment which could be one of the reasons behind incomplete elimination of the pathogen (Meader *et al*., 2010).

Another *in vitro* study used a human colon model system to test the impact of phage ΦCD27 on *C. difficile* with three replicates performed with conclusive results. In two replicates, no vegetative *C. difficile* cells were isolated after the treatment, but the amount of spores had increased significantly, while the results of the third replicate were similar to those of the control group. The authors hypothesised that ΦCD27 lysogenisation occurred at an early stage of treatment in this replicate, resulting in a phage resistant population that was able to colonise the system. Interestingly, toxin production decreased in all three replicates and was probably due to a reduction in the amount of vegetative *C. difficile* cells in the first two replicates and as a result of phage lysogeny in the third replicate. Due to phage selectivity, no detrimental effect on commensal bacteria was observed (Meader *et al*., 2013).

A more recent study tested the activity of seven *C. difficile* phages individually and in combination against clinically relevant strains *in vitro* and *in vivo* using a hamster model. Animals were infected with a combination of *C. difficile* spores instead of the vegetative cells as seen in previous *in vivo* studies and treated with individual phages or their combinations. The results have shown that the phage cocktails consisting of

phiCDHM1, phiCDHM2, phiCDHM5 and phiCDHM6 significantly reduced pathogen colonisation, disease symptoms and extended longevity in the tested hamsters (Nale *et al*., 2016a). In a wax moth model, the same phage cocktail significantly reduced *C. difficile* biofilms, significantly improved the larval remedial regimen and prevented disease in the prophylaxis group (Nale *et al*., 2016b).

More recently, a batch fermentation model was used to examine the impacts of the same optimised 4-phage cocktail on the human microbiota. The experiment determined that the phage mixture can effectively clear a clinically prevalent *C. difficile* isolate in both the prophylactic and remedial treatment vessels. Furthermore, ~2-log increase in the enterobacteria, lactobacilli, and total anaerobe abundance was seen in the phage-only-treated vessel compared to other treatments, where no effect on commensal enterococci, bifidobacteria, lactobacilli, total anaerobes, and enterobacteria was observed. It has been postulated that this proliferation could prevent further colonisation of *C. difficile*, and thus provide protection from CDI (Nale *et al*., 2018). At the time of writing, Selle *et al*., 2020 had developed a synthetic phage ΦCD24-2 expressing bacterial self-targeting CRISPR RNAs. In this study, authors proved that recombinant ΦCD24-2 is significantly more effective than its wild-type parent bacteriophage at killing *C. difficile* both *in vitro* and in a mouse model of CDI showing promising treatment of CDI.

1.3 Hypothesis and aims

To date, no phage therapy has been developed to treat CDI. The reason behind this is mainly because *C. difficile* has not been a target pathogen in the Eastern countries where phage therapy is commonly used. Advances in this area have also been limited due to challenging isolation of phages in the past and the lack of strictly virulent phages infecting *C. difficile*.

Studies have shown that *C. difficile* phages can reduce bacterial numbers both *in vivo* and *in vitro*. Traditionally, phage therapy use strictly lytic phages to eradicate bacterial infection. To this date, no strictly lytic phage infecting *C. difficile* has been isolated. The main aim of this thesis is the development of engineered temperate phages, carrying lethal cargo targeting *C. difficile*.

This project was designed to allow fundamental research in *C. difficile* phage therapy using four main objectives:

- Isolation of novel phages, that can exert lytic activity on *C. difficile* clinically relevant isolates
- Characterisation of novel phages in terms of morphology, genome sequencing and host range
- Identification of suitable location in the phage genome for cargo knock in using genome analysis and genetic tools
- Generation of recombinant phage carrying lethal cargo targeting *C. difficile*

Chapter two:

Materials and Methods

2. Materials and Methods

2.1 Bacterial strains

A list of bacterial strains used in this study, aside from those used directly for phage isolation is provided in Table 2.1. The full list of *C. difficile* strains used for phage isolation from environmental sample is presented in Table 2.2.

Table 2.1: List of bacterial strains used in this study.

Strain	usage	Source
Escherichia coli TOP10	plasmid storage strain	University of Nottingham, N. Minton
Escherichia coli NEB sExpress	Conjugative transfer strain – *E. coli* NEB Express containing R702 R factor	University of Nottingham, C. Woods
Clostridioides difficile CD9301	Genomic DNA sequencing, PCR amplification of homology arms for RiboCas9 and ACE vectors	UK, Nottingham, M. Lister

Table 2.2: List of *C. difficile* strains used for isolation of phages.

C. difficile strain	PCR ribotype	Source
CD1101	002	UK, Nottingham, N. Minton
CD2301	014	UK, Nottingham, N. Minton
CD2801	027	UK, Nottingham, M. Lister
CD5101	027	UK, Nottingham, M. Lister
CD9201	027	UK, Nottingham, M. Lister
CD3801	078	UK, Nottingham, M. Lister
CD4401	078	UK, Nottingham, M. Lister
CD8601	078	UK, Nottingham, M. Lister
CD9001	078	UK, Nottingham, M. Lister
CD9301	078	UK, Nottingham, M. Lister

CD9501	078	UK, Nottingham, M. Lister
CD9801	078	UK, Nottingham, M. Lister
CD0101	015	UK, Nottingham, M. Lister
CD0301	015	UK, Nottingham, M. Lister
CD3601	015	UK, Nottingham, M. Lister
CD4201	015	UK, Nottingham, M. Lister
CD4501	015	UK, Nottingham, M. Lister
CD1902	015	UK, Nottingham, M. Lister
CD4701	015	UK, Nottingham, M. Lister
CD4202	015	UK, Nottingham, M. Lister
CD5901	015	UK, Nottingham, M. Lister
CD6201	015	UK, Nottingham, M. Lister
CD7301	015	UK, Nottingham, M. Lister
CD4502	015	UK, Nottingham, M. Lister
CD8401	015	UK, Nottingham, M. Lister
CD8501	015	UK, Nottingham, M. Lister
CD8502	015	UK, Nottingham, M. Lister
CD81566	027	France, Ed Kuijper, LUMC for EACCAD grant
CD81568	027	France, Ed Kuijper, LUMC for EACCAD grant
CD5108111	027	The Netherlands, Ed Kuijper, LUMC for EACCAD grant
CD32219	027	Luxembourg, Ed Kuijper, LUMC for EACCAD grant
CDM246	027	Ireland, Ed Kuijper, LUMC for EACCAD grant
CD2191	027	Ireland, Ed Kuijper, LUMC for EACCAD grant
CD51556	027	Germany, Ed Kuijper, LUMC for EACCAD grant
CD60902	027	Switzerland, Ed Kuijper, LUMC for EACCAD grant
CD26131	027	Finland, Ed Kuijper, LUMC for EACCAD grant
CD26173	027	Finland, Ed Kuijper, LUMC for EACCAD grant

CDC32	027	USA (historical), Ed Kuijper, LUMC for EACCAD grant
CDC38	027	USA, Ed Kuijper, LUMC for EACCAD grant
CDM13042	027	Canada, Ed Kuijper, LUMC for EACCAD grant
CDDH349	027	UK, Cambridge, Val Hall, Anaerobic Reference Unit
CDDH320	027	UK, Newcastle, Val Hall, Anaerobic Reference Unit
CDDH1916	027	UK, Torbay, Val Hall, Anaerobic Reference Unit
CDDH1396	027	UK, Slough, Val Hall, Anaerobic Reference Unit
CDDH1432	027	UK, London, Val Hall, Anaerobic Reference Unit
CDDH1834	027	UK, Ipswich, Val Hall, Anaerobic Reference Unit
CDDH1466	027	UK, Northampton, Val Hall, Anaerobic Reference Unit
CDDH1858	027	UK, Sunderland, Val Hall, Anaerobic Reference Unit
CDDH1751	027	UK, Bradford, Val Hall, Anaerobic Reference Unit
CD027	027	Austria, Ed Kuijper, LUMC for EACCAD grant
CD51557	027	Germany, Ed Kuijper, LUMC for EACCAD grant
CDR20291	027	UK, Stoke Mandeville
CDDH361	027	UK, London, Val Hall, Anaerobic Reference Unit
CDDH131	027	UK, Manchester, Val Hall, Anaerobic Reference Unit
CDR20352	027	Canada
CDDH478	027	UK, Taunton, Val Hall, Anaerobic Reference Unit
CDR12087	027	Historical EU strain (mid 80s), Michel Popoff
CDT31694	027	Denmark, Ed Kuijper, LUMC for EACCAD grant
CDDH482	027	UK, Oxford, Val Hall, Anaerobic Reference Unit
CDR24694	027	UK, Haverfordwest, Val Hall, Anaerobic Reference Unit
CDDH326	027	UK, Sheffield, Val Hall, Anaerobic Reference Unit
CDDH183	027	UK, Leicester, Val Hall, Anaerobic Reference Unit

CDDH1342	027	UK, Macclesfield, Val Hall, Anaerobic Reference Unit
CDDH1329	027	UK, Coventry, Val Hall, Anaerobic Reference Unit
CDDH85	027	UK, Birmingham, Val Hall, Anaerobic Reference Unit
CDR23970	027	Ireland, Dublin, Val Hall, Anaerobic Reference Unit
CD3188	027	UK, Stoke Mandeville, Jon Brazier/Ed Kuijper
CD3201	027	Ed Kuijper, LUMC for EACCAD grant
CDR24988	027	Northern Ireland, Antrim, Val Hall, Anaerobic Reference Unit
CD05-1223-046	027	Belgium, Ed Kuijper, LUMC for EACCAD grant
CDLibramont 1	027	Belgium, Ed Kuijper, LUMC for EACCAD grant
CD3288	027	Joe Sorg/ Dale Gerding
CD60109015048	027	The Netherlands, Ed Kuijper
CDR20291	027	Trevor Lawley
CD196	027	Trevor Lawley
CD210	027	UK, Leeds

2.2 Plasmids

Table 2.3: List of the plasmids used in this study.

Plasmid	Description	Source
pMTL83151	Modular plasmid containing pCB102, ColE1+TraJ, *catP*	Heap *et al*., 2009
pMTL_SC7315	KO vector containing pCB102, ColE1+TraJ, *catP*, *codA*	Cartman *et al*., 2012
pMTL_SC1A pMTL_SC1B pMTL_SC2A pMTL_SC2B pMTL_SC3 pMTL_SC4A pMTL_SC5B	pMTL_SC7315 ACE vectors for cargo integration into *C. difficile* lysogen targeting locations 1A, 1B, 2A, 2B, 3, 4A ,5B	This study

pMTL_SC1A:ermB pMTL_SC1B:ermB pMTL_SC_HA2A:ermB pMTL_SC_HA2B:ermB pMTL_SC_HA3:ermB	ACE vectors for integrating *ermB* gene into *C. difficile* lysogen locations 1A, 1B, 2A, 2B, 3	This study
pRECas1-IIE-Cdif	RiboCas9 plasmid containing pCB102, ColE1+TraJ, *catP,* RiboCas9	Canadas *et al*., 2019
pMTL_RC4A pMTL_RC4B pMTL_RC5A pMTL_RC5B	RiboCas9 plasmid, vectors for cargo integration into *C. difficile* lysogen targeting locations 4A, 4B, 5A, 5B	This study
pMTL_RC4A_ermB pMTL_RC4B_ermB pMTL_RC5A_ermB pMTL_RC5B_ermB	RiboCas9 vectors for *ermB* integration into *C. difficile* lysogen locations 4A, 4B, 5A, 5B	This study
pMTL83151_Rb3_saspA pMTL83151_Rb3_saspB pMTL83151_RiboLac_saspA pMTL83151_RiboLac_saspB	Plasmids containing *sasp* genes under the control of theophylline inducible system or RiboLac inducible system	This study
pMTL_RC4A_RiboLac_saspA pMTL_RC4A_RiboLac_saspB pMTL_RC4B_RiboLac_saspA pMTL_RC4B_RiboLac_saspB pMTL_RC5A_RiboLac_saspA pMTL_RC5A_RiboLac_saspB pMTL_RC5B_RiboLac_saspA pMTL_RC5B_RiboLac_saspB	RiboCas9 vectors for *RiboLac_sasp* cargo integration into *C. difficile* lysogen locations 4A, 4B, 5A, 5B	This study
pMTL_SC4A_Rb3_saspA pMTL_SC4A_Rb3_saspB	ACE vectors for *Rb3_sasp* cargo integration into *C. difficile* lysogen locations 4A	This study
pMTL_RCas_*tcdA*1 pMTL_RCas_*tcdA*2 pMTL_RCas_*tcdA*3 pMTL_RCas_*tcdA*4	RiboCas9 plasmids, *tcdA* knock out of *C. difficile*	This study
pMTL_RCas_*tcdB*1 pMTL_RCas_*tcdB*2 pMTL_RCas_*tcdB*3 pMTL_RCas_*tcdB*4	RiboCas9 plasmid, *tcdB* knock out of *C. difficile*	This study

pMTL_RCas_*tcdB*A1 pMTL_RCas_*tcdB*A2 pMTL_RCas_*tcdB*A3 pMTL_RCas_*tcdB*A4	RiboCas9 plasmid, *tcdB_tcdA* knock out of *C. difficile*	This study
pMTL_P_{tet}_*pyrE*PS1 pMTL_P_{tet}_*pyrE*PS2 pMTL_P_{tet}_*pyrE*PS3 pMTL_P_{tet}_*pyrE*PS4 pMTL_P_{tet}_*pyrE*PS5	endogenous CRISPR miniarray plasmid, *pyrE* knock out of *C. difficile*	This study
pMTL_P_{tet}_*tcdB*APS1 pMTL_P_{tet}_*tcdB*APS2 pMTL_P_{tet}_*tcdB*APS3 pMTL_P_{tet}_*tcdB*APS4	endogenous CRISPR miniarray plasmids targeting toxin genes of *C. difficile*	This study
pMTL_RC4A_P_{tet}_*tcdBAPS3* pMTL_RC4B_P_{tet}_*tcdBAPS1* pMTL_RC4B_P_{tet}_*tcdBAPS3* pMTL_RC5A_P_{tet}_*tcdBAPS1* pMTL_RC5A_P_{tet}_*tcdBAPS3*	RiboCas9 vectors for endogenous CRISPR miniarray integration into *C. difficile* lysogen locations 4A, 4B and 5A	This study

2.3 Oligonucleotide primers

Table 2.4: List of primers used in this study.

Name	Sequence 5' -> 3'	Description
ZZ17	atagcgatcgcgatagcatgtatatttttt	F, LHA for *tcdA* knockouts
ZZ29	aaaatgtcgactaataaatagtttttatcgagttttagagctagaa atagcaagttaaaataaggctagtccgttatcaacttgaaaaagt ggcaccgagtcggtgc	F, gRNA for *tcdA* knockout
ZZ30	aaaatgtcgacatacttaatagtaacttaccgttttagagctagaa atagcaagttaaaataaggctagtccgttatcaacttgaaaaagt ggcaccgagtcggtgc	F, gRNA for *tcdA* knockout
ZZ31	aaaatgtcgactccttcaagagtattttggtgttttagagctagaa atagcaagttaaaataaggctagtccgttatcaacttgaaaaagt ggcaccgagtcggtgc	F, gRNA for *tcdA* knockout
ZZ32	aaaatgtcgacctttatgcctgatactgctagttttagagctagaa atagcaagttaaaataaggctagtccgttatcaacttgaaaaagt ggcaccgagtcggtgc	F, gRNA for *tcdA* knockout

ZZ33	aaaatgtcgaccgtataagagatcactacaagttttagagctaga aatagcaagttaaaataaggctagtccgttatcaacttgaaaaag tggcaccgagtcggtgc	F, gRNA for *tcdB* knockout
ZZ34	aaaatgtcgacacatcatagatagataagtggttttagagctaga aatagcaagttaaaataaggctagtccgttatcaacttgaaaaag tggcaccgagtcggtgc	F, gRNA for *tcdB* knockout
ZZ35	aaaatgtcgacgaaacctagttcagtaacaggttttagagctaga aatagcaagttaaaataaggctagtccgttatcaacttgaaaaag tggcaccgagtcggtgc	F, gRNA for *tcdB* knockout
ZZ36	aaaatgtcgaccatatgttcctcctgagccagttttagagctagaa atagcaagttaaaataaggctagtccgttatcaacttgaaaaagt ggcaccgagtcggtgc	F, gRNA for *tcdB* knockout
ZZ37	cgcgcgcggcgatcgcataaaaataagaagcctgcaaatgcagg cttcttatttttataaaaaaagcaccgactcggtgccacttttcaa gttg	R, gRNA with AsiSi restriction sites
ZZ40	catatattctagccataagacataaaaacctcctagt	R, LHA for *tcdA* knockout
ZZ41	aggaggtttttatgtcttatggctagaatatatgtttg	F, RHA for *tcdA* knockout
ZZ42	atggcgcgcctaatgaaggaaaaggaatc	R, RHA for *tcdA* and *tcdB*A knockout
ZZ43	atgcgatcgctgaatattttcatataaaactt	F, LHA for *tcdB* knockout
ZZ44	ttttttatctattcactactcataaaattttctccttt	R, LHA for *tcdB* knockout
ZZ45	aggagaaaattttatgagtagtgaatagataaaaatatgt	F, RHA for *tcdB* knockout
ZZ46	atggcgcgcctgatacttattttctctat	R, RHA for *tcdB* knockout
ZZ47	gcgacgtctgaatattttcatataaaacttaatttatttt	F, LHA for *tcdB*A knockout, colony PCR
ZZ51	tgatgatgagacgtcataaaaataagaagcctgcaaatgcaggc ttcttatttttataaaaaaagcaccgactcggtgccacttttcaag ttg	R, gRNA with *Aat*II RS

ZZ52	aaaatgtcgactaaatgtgaaatctggcgaagttttagagctaga aatagcaagttaaaataaggctagtccgttatcaacttgaaaaag tggcaccgagtcggtgc	F, gRNA for *tcdBA* knockout
ZZ53	aaaatgtcgactattctttccatggctcagggttttagagctagaa atagcaagttaaaataaggctagtccgttatcaacttgaaaaagt ggcaccgagtcggtgc	F, gRNA for *tcdBA* knockout
ZZ54	aaaatgtcgacaagctgccgcatatgcttcagttttagagctagaa atagcaagttaaaataaggctagtccgttatcaacttgaaaaagt ggcaccgagtcggtgc	F, gRNA for *tcdBA* knockout
ZZ55	aaaatgtcgacgttttctacttctctgattcgttttagagctagaaat agcaagttaaaataaggctagtccgttatcaacttgaaaaagtgg caccgagtcggtgc	F, gRNA for *tcdBA* knockout
ZZ56	tacgtgatgaaggacaaaatgatagaga	F, sequencing *tcdA* knockouts
ZZ57	ctaaaagtcatgtattacaataacacatgac	R, sequencing *tcdA* *tcdBA* knockouts
ZZ59	gcttaggggaaatatattttaggaaaatatgaatatataatttttagtc	R, colony PCR tcdA and *tcdBA* knockouts, sequencing
ZZ60	gcttaagagcaataaaaacacatcagtttaattcaagc	F, sequencing *tcdA* knockouts
zz61	gtgttctatctaatatgaagatttaccaataaaaaggtgaactatgat	F, sequencing *tcdB* knokcouts
zz63	ctcaattagttaattgactcttctattagataaggagat	R, sequencing *tcdB* knockouts
ZZ62	acgatttgtttctataatactttattaatatgaaatttcttgc	F, colony PCR, sequencing *tcdB*, tcdBA knockouts
ZZ63	ctcaattagttaattgactcttctattagataaggagat	R, sequencing *tcdB* knockouts
ZZ68	gcttttcagaaaaaagtgatgctattatag	F, sequencing *tcdB* , *tcdBA* knockouts
ZZ69	gagtgggacgaagaaaagaataaaaaagttc	F, close genome phage ΦCD3801
ZZ70	gcactaccaaagcaccacttttg	R, close genome phage ΦCD3801

ZZ75	ggagaaaatttttatgagttatggctagaatatatgtttgacaaaaatttattcatg	F, amplify RHA of *tcdB*A
ZZ76	tatattctagccataactcataaaattttctcctttactataa	R, amplify LHA of *tcdB*A, sequencing
ZZ77	gaagtagaaaagatatttaaagccccgcctata	F, close genome phage ΦCD8601
ZZ78	tcttcgaaacctccacttacttttgataaatct	R, close genome phage ΦCD8601
ZZ79	tgaggaactgataagatggttatatcaatctaaagtcaa	F, close genome phage ΦCD9301
ZZ80	atatccaacatcctcgtaacctaatccttttga	R, close genome phage ΦCD9301
ZZ81	atcccggggagtattgttaaagaagttaagagagtagcaacaaaatc	F, amplify LHA for pMTL_SC_HA1B
ZZ82	acatcctagggaatggcgcctttcggcaattagaatcaataac	R, amplify LHA for pMTL_SC_HA1B
ZZ83	cgaaaggcgccattccctaggatgtgccgtttttttaattttagacatgag	F, amplify RHA for pMTL_SC_HA1B
ZZ84	atctcgagtacttaagttcacaattgaatagccactcattaaaatctcctgaatat	R, amplify RHA for pMTL_SC_HA1B
ZZ86	gtcctaggaaaaaaagcaccgactcgg	R, amplify RiboCas9 cargo
ZZ87	atcccggggctcccaaaatataatagtatatgcta	F, amplify LHA for pMTL_SC_HA2A
ZZ88	aagttaatcctagggaatggcgccctattctgtaggtttaaaact	R, amplify LHA for pMTL_SC_HA2A
ZZ89	gaatagggcgccattccctaggattaacttgtgcagaattttacattaac	F, amplify RHA for pMTL_SC_HA2A
ZZ90	atctcgagactttatacaagagtagctatttccaatcagattaaaaataacttatttt	R, amplify RHA for pMTL_SC_HA2A
ZZ91	atcccggggctactcttgtataaagtaaattatttccttgctt	F, amplify LHA for pMTL_SC_HA3
ZZ92	ctcctagggaatggcgccctatcacctcacttaaggaattattaa	R, amplify LHA for pMTL_SC_HA3

ZZ93	gatagggcgccattccctaggagttgcagaaagtaagaataatac	F, amplify RHA for pMTL_SC_HA3
ZZ94	atctcgaggttgtatgatacttcgctttgagtaagtaaaaaac	R, amplify RHA for pMTL_SC_HA3
ZZ103	ccttttgcgtttctcagtcacctcctagctgactc	F, amplify RiboCas9 cargo
ZZ104	aggaggtgactgagaaacgcaaaaaggccatccg	R, amplify T1T2 terminator with shared homology to RiboCas9 cargo
ZZ105	atggcgcctccaggcatcaaataaaacgaaagg	F, amplify T1T2 terminator
ZZ107	cggatggccttttgcgtttctttatttattaatttaatattgagag	F, amplify *ermB* cargo
ZZ108	gtcctaggttatttcctcccgttaaataatagataactattaaaaatagac	R, amplify *ermB* cargo
ZZ109	atcccggggttacctatagtgtttgaggaagaaataagacttttaaatagtcaattatcaata	F, amplify LHA for pMTL_SC_HA1A
ZZ110	cctagggaatggcgccttaccaagccccacccgattttta	R, amplify LHA for pMTL_SC_HA1A
ZZ111	ggcgccattccctaggcctcactcaccatgcaagta	F, amplify RHA for pMTL_SC_HA1A
ZZ112	atctcgagcccgattataacaggcttgcctttataagaagttaatactaaatgtgta	R, amplify RHA for pMTL_SC_HA1A
ZZ113	atcccggggtatatgctataattgtagtgtatattgttatattttaggagg	F, amplify LHA for pMTL_SC_HA2B
ZZ114	agcctagggaatggcgccttaatgtaaaattctgcacaag	R, amplify LHA for pMTL_SC_HA2B
ZZ115	taaggcgccattccctaggctatgtaggttcaaaattaaatag	F, amplify RHA for pMTL_SC_HA2B
ZZ116	atctcgagctaatttgtggtataataaaagcaaggaaataatttactttatac	R, amplify RHA for pMTL_SC_HA2B
ZZ117	ctcaatattaaattaataaataaagaaacgcaaaaaggccatccgtcaggatggcc	R, amplify T1T2 terminator with shared homology to *ermB* cargo

ZZ118	gtttaaaaacatgatacaataagttatggttgg	F, sequence pMTL_SC_HA vectors
ZZ119	gttgaatttgcaggattacggcat	F, sequence pMTL_SC_HA vectors
ZZ120	cggttggggatttaatgaaagggataagtt	F, sequence pMTL_SC_HA vectors
ZZ121	gtttaggtatattataatgtgcatacttaaaattaaatac	F, sequence pMTL_SC_HA vectors
ZZ122	gattaagttgggtaacgccaggg	R, sequence pMTL_SC_HA vectors
ZZ123	gaacatagttagaattgaactaacaac	R, sequence pMTL_SC_HA vectors
ZZ124	gtaatacctaataatttcaagggcaaagc	F, sequence pMTL_SC_HA vectors
ZZ125	cagataagagaggttttagattaactatc	F, sequence pMTL_SC_HA vectors
ZZ126	ctataaagggattatgctaaaaagcaa	R, sequence pMTL_SC_HA vectors
ZZ127	ggatgaatgaaggctttatattaaggtttac	F, sequence pMTL_SC_HA vectors
ZZ128	gagaaatttacaagtaacagtaatatccttttc	F, sequence pMTL_SC_HA vectors
ZZ129	cgtttatttgaatatgttttgccatacaatataagtc	R, sequence pMTL_SC_HA vectors
ZZ130	ctttcctctattttatttaaacaatgaaatttgag	R, sequence pMTL_SC_HA vectors
ZZ137	ccttagcaaccactaggactgaataagc	F, sequence *cas9* cargo
ZZ138	ggaacgttatccgatttaccacgatttttatc	F, sequence *cas9* cargo
ZZ139	gagcatatgttttaagtctttcctcaatcatct	F, sequence *cas9* cargo
ZZ140	gatcaaaaaagatttctttatacttttctggaagttg	F, sequence *cas9* cargo

ZZ141	gatgatagtttctttcatcgacttgaagagtc	R, sequence *cas9* cargo
ZZ142	cttgacttttcgaattccttattatgttggt	R, sequence *cas9* cargo
ZZ143	cgagagcgtatgaaacgaatcgaagaagg	R, sequence *cas9* cargo
ZZ144	agttagagaactttatggatgatgcaccattattc	F, colony PCR ΦCD9301 with integrated cargo in position 1B
ZZ145	agaatctttttaacttcttcttcctccagctc	R, colony PCR ΦCD9301 with integrated cargo in position 1B
ZZ146	gatgttagttaatgaaataaagttaaatttagcaggtggtg	F, colony PCR ΦCD9301 with integrated cargo in position 2A, 2B
ZZ147	aaaaatattatatgcaatatttttttctaatttgtggtataataaaagcaagg	R, colony PCR ΦCD9301 with integrated cargo in positions 2A, 2B
ZZ148	gccaataagagtggtttaaaaaaataagttatttttaatctgattgg	F, colony PCR ΦCD9301 with integrated cargo in position 3
ZZ149	tttattgtcattaaatacttgttttgctatatcttcaaccaacttatc	R, colony PCR ΦCD9301 with integrated cargo in position 3
ZZ150	gctataaagcgttgtcaattaaataaagtattagagattgtaaac	F, colony PCR ΦCD9301 with integrated cargo in position 1A
ZZ151	ctaaagagtttgataatgtttttctttttatcgtaaaatacttattagctaaag	R, colony PCR ΦCD9301 with integrated cargo in position 1A

ZZ152	ctaaaataagtaaacaggtaacgtctattgaattagacag	F, sequence *ermB* cargo
ZZ153	cctgtttacttattttagccagtttcgtcg	R, sequence *ermB* cargo
ZZ154	gcactaagaactttatctaaattggcatct	F, sequence *cas9* cargo
ZZ155	gtcttcttgtcttctcaaaatagcatgc	F, sequence *cas9* cargo
ZZ156	gcccctataagatttttttgatactgtgg	F, sequence *cas9* cargo
ZZ164	aaaatgtcgacgcagagaggattttgaataagttttagagctaga aatagcaagttaaaataaggctagtccgttatcaacttgaaaaag tggcaccgagtcggtgc	F, gRNA for ΦCD9301 knocking in position HA4A
ZZ165	aaaatgtcgacgtttacaaacagttgaaggagttttagagctaga aatagcaagttaaaataaggctagtccgttatcaacttgaaaaag tggcaccgagtcggtgc	F, gRNA for ΦCD9301 knocking in position HA4B
ZZ166	aaaatgtcgacaatttaactggtgaagagctgttttagagctaga aatagcaagttaaaataaggctagtccgttatcaacttgaaaaag tggcaccgagtcggtgc	F, gRNA for ΦCD9301 knocking in position HA5A
ZZ167	aaaatgtcgacgaatataaggtcgataaaccgttttagagctaga aatagcaagttaaaataaggctagtccgttatcaacttgaaaaag tggcaccgagtcggtgc	F, gRNA for ΦCD9301 knocking in position HA5B
ZZ197	gcgacgtcgcaagtttgaataatacactagttg	F, amplify RiboCas9 LHA4A
ZZ200	atggcgcgccctctataagtaagtcatatcctcc	R, amplify RiboCas9 RHA4A
ZZ201	gcgacgtcgatacaataacagaattacattatattgatgttg	F, amplify RiboCas9 LHA4B
ZZ204	atggcgcgccgtggtaactgatgaaatagttacc	F, amplify RiboCas9 RHA4B
ZZ205	gcgacgtctccccatcttaaattgcttcttaattaattatataataac	F, amplify RiboCas9 LHA5A
ZZ208	atggcgcgcctctctactctcaactaagtatttattg	R, amplify RiboCas9 RHA5A

ZZ209	gcgacgtccgctagttacgtaacacttattttaac	F, amplify RiboCas9 LHA5B
ZZ212	atggcgcgccaggaaataattttgaaaacagttattgc	F, amplify RiboCas9 RHA5B
ZZ231	cctagggaatggcgccaaaatcctctctgctaatattcttac	R, amplify RiboCas9 LHA4A
ZZ232	ttttggcgccattccctagggaataaaggattaaaccctctaaaaaatc	F, amplify RiboCas9 RHA4A
ZZ233	cctagggaatggcgccaactgtttgtaaacttgatttgc	R, amplify RiboCas9 LHA4B
ZZ234	ttggcgccattccctagggaaggatggtgataaataatgaaaatatg	F, amplify RiboCas9 RHA4B
ZZ235	acctagggaatggcgccactggtgaagagcttggtaaaatactaaatg	R, amplify RiboCas9 LHA5A
ZZ236	tggcgccattccctaggtaaatttttttgaactcctatttcttttaatc	F, amplify RiboCas9 RHA5A
ZZ237	cctagggaatggcgccaaggtcgataaaccgggtctaattattaattc	R, amplify RiboCas9 LHA5B
ZZ238	ttggcgccattccctaggatattctttgcgaattttatctttaagttc	F, amplify RiboCas9 RHA5B
ZZ275	atggcgccctgccaatagataaaataaagtctgccacaattgtggcagactttattttatctattgatttatttcatttttttaatctagcctcgagatatttgccacccaagttttgttc	F, amplify saspA cargo (for knock in)
ZZ276	gtcctagggtgtagtagcctgtgaaataagtaaggaaaaaaaagaagtaagtgttatatatgatgattattttgtagatg	R, amplify Pfdx_Rb3 (for knock in)
ZZ277	atggcgccctgccaatagataaaataaagtctgccacaattgtggcagactttattttatctattgatttatttcatttttttaatctagcctcgagttgttgaccatgaaattgtccacc	F, amplify *saspB* cargo (for knock in)
ZZ280	gcaaaacaagtattataaatactttagcattaaagaatatag	F, colony PCR ΦCD9301 with integrated cargo in position 4A
ZZ281	ctatagtaaaatacttctgttccaaaagctccta	R, colony PCR ΦCD9301 with

		integrated cargo in position 4A
ZZ282	caaactggagaagcaacaggagatatactagaac	F, colony PCR ΦCD9301 with integrated cargo in position 4B
ZZ283	cgacatgggtttttaaaaaaacagcacttgaaagagaatataac	R, colony PCR ΦCD9301 with integrated cargo in position 4B
ZZ284	caaacactattctaccctcctaatttagttatgtaaaatacac	F, colony PCR ΦCD9301 with integrated cargo in position 5A
ZZ285	gcactctgctctaatacacttatatatcctcttatacttcttaataatt g	R, colony PCR ΦCD9301 with integrated cargo in position 5A
ZZ294	gaagagaggggaggtgttacaatgttatataaaatattac	F, colony PCR ΦCD9301 with integrated cargo in position 5B
ZZ295	gatttttttactatgcagaacaaatgtaaaacaagattttattcatga taaatgtaatc	R, colony PCR ΦCD9301 with integrated cargo in position 5B
ZZ300	atcccggggcaagtttgaataatacactagttgagttatcagag	F, amplify HA4A for ACE vector
ZZ301	atctcgagctctataagtaagtcatatcctccgctattaactctagg	R, amplify HA4A for ACE vector
ZZ302	atcccggggatacaataacagaattacattatattgatgttgagta tggaag	F, amplify HA4B for ACE vector
ZZ303	atctcgaggtggtaactgatgaaatagttaccactctctacaaact tatttc	R, amplify HA4B for ACE vector
ZZ304	gcactctgctctaatacacttatatatcctcttatacttcttaataatt g	R, amplify *saspA* cargo with homology to Pfdx_Rb3

ZZ305	gttagccatcttgttgttacctccttagcagggtgctgccaa	F, amplify Pfdx_Rb3 with homology to *saspA* cargo
ZZ306	ctgctaaggaggtaacaacaagatggcaaagttactaattccacaggctgcg	R, amplify *saspB* cargo with homology to Pfdx_Rb3
ZZ307	ctttgccatcttgttgttacctccttagcagggtgctgcc	F, amplify Pfdx_Rb3 with homology to *saspB* cargo
ZZ312	ctgtggaattagtaactttgccatcatatgaccctcccaatacatttaaaataattatg	R, amplify RiboLac system with homology to *saspB* (for testing)
ZZ313	gggtcatatgatggcaaagttactaattccacaggctgcgtct	F, amplify *saspB* with homology to RiboLac (for testing)
ZZ315	cattttgatagttagccatcatatgaccctcccaatacatttaaaataattatg	R, amplify RiboLac system with homology to *saspA* (for testing)
ZZ316	gggtcatatgatggctaactatcaaaatgcctcaaacagaaattctagtaataag	F, amplify *saspA* with shared homology to RiboLac (for testing)
ZZ317	gtcctaggattaaaaaatgaaataaatcaatagataaaataaagtctgccac	R, amplify *RiboLac_sasp* for knock in
ZZ318	atggcgccgctaatttagatattaattctaaattaagtg	F, amplify *RiboLac_sasp* for knock in
ZZ321	atcccgggtccccatcttaaattgcttcttaattaattatataataacatatattattttaaac	F, amplify HA5A for ACE vectors
ZZ322	atctcgagtctctactctcaactaagtatttattgttttgtttgataattg	R, amplify HA5A for ACE vectors
ZZ323	gaattatgtttcttaaatatacaatcatgtttcatgaatac	F, sequence sasp vectors
ZZ324	gactactataaaattggctaattttataggtaccaatac	F, sequence sasp vectors

ZZ325	cttttagaatagttaaaaagttctaaaatatgttataatg	F, sequence sasp vectors
ZZ326	gacccggggctaatttagatattaattctaaattaagtg	F, amplify riboLAC system with *Xma*I (For testing)
ZZ327a	gtcgacattaaaaaatgaaataaatcaatagataaaataaagtctgccacaattgtggcagactttattttatctattggcagctagcctcgagatatttgccacccaagttttg	R, amplify *saspA* with SalI (for RiboLac testing)
ZZ327b	gtcgacattaaaaaatgaaataaatcaatagataaaataaagtctgccacaattgtggcagactttattttatctattggcagctagcctcgagttgttgaccatgaaattg	R, amplify *saspB* with SalI (for RiboLac testing)
ZZ328b	gacccgggctgccaatagataaaataaagtctgccacaattgtggcagactttattttatctattgatttatttcatttttttaatctagcctcgagatatttgccacccaagttttg	F, amplify *Xma_Rb3saspA_SalI*
ZZ328a	gacccgggctgccaatagataaaataaagtctgccacaattgtggcagactttattttatctattgatttatttcatttttttaatctagcctcgagttgttgaccatgaaattgtc	F, amplify *Xma_Rb3saspB_SalI*
ZZ329	gagtcgacgtgtagtagcctgtgaaataagtaaggaaaaaaaagaagtaagtg	R, amplify *Xma_Rb3saspAB_SalI*
ZZ330	gctttgtacattctaaggtttattaaatatctttgag	F, sequence *sasp* vectors
ZZ331	gaagttctatttaaatactcattagcatttgaaccactaaatc	F, sequence *sasp* vectors
ZZ332	tttttgcggccgcttaagacccactttcacatttaagttgtttttc	F, amplify tet system
ZZ333	tttggatccatgtatttcctcctcttcaatatatttaaggtctctcgcg	R, amplify tet system
ZZ334	tgacactctatcattgatagag	F, sequence enCRISPR miniarrays
ZZ335	ctggtcttctatttttttatatcaattgccattgc	R, sequence *pyrE* plasmid and knockout
ZZ336	caccaaatgcagaagacattgtag	R, sequence *pyrE* plasmid and knockout
ZZ337	gctcctacagctcctactccaaaaaatacaacctttttc	R, sequence *pyrE* plasmid and knockout
ZZ338	ggagtaggagctgtaggagctacttttgcag	F, sequence *pyrE* plasmid

ZZ340	ttttgtcgacataaaaataagaagcctgcaaatgcaggcttcttat ttttatctttacataccacttagttaatataaaacaaatcatcatga gattaattttt	R, amplify CRISPR miniarray PS1 for tcd testing
ZZ342	ttttgtcgacataaaaataagaagcctgcaaatgcaggcttcttat ttttatctttacataccacttagttaatataaaactgagaaagagat aatgcatatgt	R, amplify CRISPR miniarray PS2 for tcd testing
ZZ344	ttttgtcgacataaaaataagaagcctgcaaatgcaggcttcttat ttttatctttacataccacttagttaatataaaacaatcatccaattc attttgttgt	R, amplify CRISPR miniarray PS3 for tcd testing
ZZ353	tttgtcgacataaaaataagaagcctgcaaatgcaggcttcttatt tttatatttaaatacatctcatgttactgtttaactaatgcatatgttc ctcctgag	R, amplify CRISPR miniarray PS4 for tcd testing
ZZ354	tttggatccgttttatattaactaagtggtatgtaaagtttacatctc cagaaaaattaatctcatgatgatttgttttatattaactaagtg	F, amplify CRISPR miniarray PS1 for tcd testing
ZZ355	tttggatccgttttatattaactaagtggtatgtaaagtggctcagg aggaacatatgcattatctctttctcagttttatattaactaagtg	F, amplify CRISPR miniarray PS2 for tcd testing
ZZ356	tttggatccgttttatattaactaagtggtatgtaaagtatacgattt aaaacaacaaaatgaattggatgattgttttatattaactaagtg	F, amplify CRISPR miniarray PS3 for tcd testing
ZZ358	tttggatccgttaaacagtaacatgagatgtatttaaattattctttc catggctcaggaggaacatatgcattagttaaacagtaacatgag a	F, amplify CRISPR miniarray PS4 for tcd testing
ZZ393	ttggcgccctgccaatagataaaataaagtctgccacaattgtggc agactttattttatctattgatttatttcatttttttaatttaagaccca ctttcacatttaagttgtttttcta	F, amplify CRISPR miniarray PS1 or PS3 for phage knockin
ZZ394	aaacctaggataaaaataagaagcctgcaaatgcaggcttcttat ttttatctttacataccacttagttaatataaaac	R, amplify CRISPR miniarray PS1 or PS3 for phage knockin

P1	tttttGTCGACctaatgttggaggtgggaca	F, amplify LHA for *pyrE* knockout
P2	catggaggaaaaataatatgagtaaagtagatcaaagataata	R, amplify LHA for *pyrE* knockout
P3	taaagtagatcaaagataataattgctgagatgttg	F, amplify RHA for *pyrE* knockout
P4	Tttttggcgcgccgtcagcttcttgtttatattcatctc	R, amplify RHA for *pyrE* knockout
P5	tttttggatccgttttatattaactaagtggtatgtaaaggaatatg ctgctgaggtattaagtacagttgttgaagttttatattaactaagt g	F, amplify *pyrE* miniarray PS1
P6	Tttttgtcgacataaaaataagaagcctgcaaatgcaggcttctta ttttatctttacataccacttagttaatataaaacttcaacaactgt acttaatacct	R, amplify *pyrE* miniarray PS1
P7	tttttggatccgttttatattaactaagtggtatgtaaaggctatgg gtggagtaatagtttcttatgaattgggagttttatattaactaagt g	F, amplify *pyrE* miniarray PS2
P8	tttttgtcgacataaaaataagaagcctgcaaatgcaggcttctta ttttatctttacataccacttagttaatataaaactcccaattcata agaaactatta	R, amplify *pyrE* miniarray PS2
P10	tttttggatccgttttatattaactaagtggtatgtaaagactactaa gtctatatctaagtcttttatttgttcagttttatattaactaagtg	F, amplify *pyrE* miniarray PS3
P11	tttttgtcgacataaaaataagaagcctgcaaatgcaggcttctta ttttatctttacataccacttagttaatataaaactgaacaaataa aagacttagata	R, amplify *pyrE* miniarray PS3
P12	tttttggatccgttttatattaactaagtggtatgtaaagcttcttaa ctccattgtattgtcttttctctcagtagttttatattaactaagtg	F, amplify *pyrE* miniarray PS4
P13	tttttgtcgacataaaaataagaagcctgcaaatgcaggcttctta ttttatctttacataccacttagttaatataaaactactgagagaa aagacaatacaa	R, amplify *pyrE* miniarray PS4
P14	tttttggatccgttaaacagtaacatgagatgtatttaaatggaaa acacagtaatagatatgtacaatgtgcaaaagttaaacagtaaca tgag	F, amplify *pyrE* miniarray PS5
P15	tttttgtcgacataaaaataagaagcctgcaaatgcaggcttctta ttttatatttaaatacatctcatgttactgtttaacttttgcacattg tacatatcta	R, amplify *pyrE* miniarray PS5

M120_diag_F	ggtaacaatggtatgagag	F, colony PCR *pyrE* knockouts
M120_diag_R	gcagagacattatgactg	R, colony PCR *pyrE* knockouts

2.4 Bioinformatics tools

2.4.1 Plasmid design

All plasmids were designed using the Benchling online software available at www.benchling.com.

2.4.2 Oligonucleotide design

All primers were manually designed using the Benchling software. Secondary structures and primer dimer formations were predicted together with melting temperature using the Benchling or OligoAnalyser tool available online at https://www.idtdna.com/calc/analyzer.

2.4.3 Sanger Sequencing

Plasmid DNA or PCR products were sequenced by Source Bioscience or Eurofins Genomics. Data was aligned using Benchling or analysed using the Basic Local Alignment Search Tool (BLAST) available at www.ncbi.com/blast.

2.4.4 Genome sequencing

Genome sequencing was provided either by MicrobesNG (Birmingham) or DeepSeq (University of Nottingham). Genomic DNA samples were sequenced by Illumina MiSeq 500bp V2 SBS sequencing and sequencing reads were assembled using CLC Genomics Workbench 11 (Qiagen). For bacterial DNA, trimmed sequencing reads were assembled by their alignment to the reference genome of *C. difficile* CD9301. Identification of single nucleotide variants (SNVs) was carried out using the CLC Genomics Basic Variant Detection tool following the program's default parameters. For phage genomes, sequencing reads were assembled into a single contig *de novo*

using the CLC Genomics *De Novo* Assembly function with default parameters. Open reading frames (ORFs) with a minimum of 100 amino acid residues were predicted using the Artemis software. ORFs were 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). Each ORF was then manually annotated using the online software BLASTp available at www.ncbi.com/blast or UniProt www.uniprot.org.

2.4.5 Phylogenomic analysis

Virus Classification and Tree Building Online Resource (VICTOR) was used for phylogenomic analysis. Nucleotide sequences of isolated phages were compared pairwise with 29 published *C. difficile* phage genomes using the Genome-BLAST Distance Phylogeny (GBDP) method (Meier-Kolthoff *et al*., 2013) under the settings recommended for prokaryotic viruses (Meier-Kolthoff and Göker, 2017). A balanced minimum evolution tree with branch support was created using resulting intergenomic distances via FASTME including SPR postprocessing (Lefort *et al*., 2015), for each of the formulas D0, D4 and D6, respectively. Branch support was concluded from 100 pseudo-bootstrap replicates each. Each tree was rooted at the mindpoint (Farris 1972) and visualized with FigTree (Rambaut, 2006).

The OPTSIL programme (Göker *et al*., 2009) was used to estimate taxon boundaries at the family, genus and specie level with the recommended clustering thresholds (Meier-Kolthoff and Göker, 2017) and with an F value (fraction of links required for cluster fusion) of 0.5 (Meier-Kolthoff *et al*., 2014) being used.

2.4.6 Prediction of terminator sequences

Rho-independent terminators were detected in both strands of the phage genome using the ARNOLD terminator prediction tool, which combines two algorithms, Erpin (Gautheret & Lambert, 2001) and RNAmotif (Macke *et al*. 2001). The software is available online at http://rssf.i2bc.paris-saclay.fr/toolbox/arnold/index.php.

2.4.7 Predictions of promoter sequences

PePPER promoter prediction tool (de Jong *et al*., 2012) and Genome Compiler web software have been used to identify the presence of promoters and RBS sites of integrase genes. Both tools are available online at http://www.cbs.dtu.dk/services/Promoter/ and http://www.genomecompiler.com/ respectively.

2.4.8 Guide RNA design

For the *Streptococcus pyogenes* Cas9 used in this study, the guide RNA consists of a 20 variable nucleotide seed region used for targeting, and an approximately 150 nucleotide sgRNA handle. Seed sequences for CRISPR gRNA were designed using the free online tool CRISPy-web available on https://crispy.secondarymetabolites.org/#/input or using the Benchling CRISPR Guide Design tool, accessible at https://benchling.com. At least two seed sequences were identified for each target region and analysed for off-target and on-target scores using the Benchling CRISPR guide design tool based on Doench *et al*. (2016) algorithm or CRISP-web algorithm based on Ronda *et al*. (2014). Only seed sequences with the highest off-target and on-target scores were used.

2.4.9 Characterisation of the endogenous CRISPR system in *C. difficile* RT078

Endogenous CRISPR arrays and *cas* clusters were identified using the CRISPRcas finder online tool (Grissa *et al*., 2007) available at https://crisprcas.i2bc.paris-saclay.fr/. Identification of *C. difficile* M120 CRISPR target sequences was carried out using the CRISPRTarget software (Biswas *et al*, 2013) following the program's default parameters, available at http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html. PHASTER phage prediction tool (Arndt *et al*., 2016) available at: https://phaster.ca/ was used to determine whether any of the CRISPR arrays were localised in the prophage regions of *C. difficile* M120 genome.

2.5 General microbiology tools

2.5.1 Growth media and antibiotics

Growth media were prepared by dissolving adequate amount of components into distilled water and sterilised by autoclaving at 121°C for 20 minutes (Table 2.5). If required, antibiotics and supplements were added to the media after autoclaving and cooling media to 50°C (Table 2.6).

Table 2.5: List and composition of growth media used in this study.

Media	Composition	Quantity (g x L^{-1})
Luria-Bertani (LB)	Sodium chloride	10
	Tryptone	10
	Yeast extract	5
LB agar	Sodium chloride	10
	Tryptone	10
	Yeast extract	5
	Agar	15
Brain Heart Infusion Supplemented (BHIS)	Brain Heart Infusion	37
	Yeast extract	5
	L- cysteine	1
BHIS agar	Brain Heart Infusion agar	47
	Yeast extract	5
	L- cysteine	1
BHIS top agar 0,5%	Brain Heart Infusion	37
	Yeast extract	5
	L- cysteine	1
	Agar	5
BHIS top agar 0,2%	Brain Heart Infusion	37
	Yeast extract	5
	L- cysteine	1
	Agar	2
C. difficile minimal media (CDMM)	Casamino acids	10
	D-glucose	5
	L-tryptophan	0.5
	L-cysteine	0.5
	Na_2HPO_4	5
	$NaHCO_3$	5
	KH_2PO_4	0.9
	NaCl	0.9
	$(NH_4)_2SO_4$	0.04
	$CaCl_2.2H_2O$	0.026
	$MgCl_2.6H_2O$	0.02

	$MnCl_2.4H_2O$	0.01
	$CoCl_2.6H_2O$	0.001
	$FeSO_4.7H_20$	0.0013
	Ca-D-pantothenate	0.001
	D-biotin	0.001

Table 2.6: List of antibiotics and supplements used in this study.

Antibiotic	**Solvent**	**Stock concentration (mg x ml^{-1})**	**Working concentration for *E. coli* (μg x ml^{-1})**	**Working concentration *Clostridium* (μg x ml^{-1})**
Chloramphenicol (Cm)	100% ethanol	25	25 (agar) 12.5 (broth)	
Thiamphenicol (Tm)	50% ethanol	15	-	15
D-cycloserine	dH_2O	50	-	250
Cefoxitin	dH_2O	50	-	8
Theophylline	100% DMSO	48	-	900
Erythromycin (Em)	100% ethanol	50	50	500
5-Fluorocytosine (5-FC)	dH_2O	500		50
5-Fluoroorotic acid (5-FOA)	100% DMSO	100	-	4000
Uracil	dH_2O	1	-	20
Anhydrous tetracycline	dH_2O	2	-	0.1

2.5.2 Aerobic bacteria

2.5.2.1 Growth condition

All *E. coli* strains were grown in aerobic conditions on Luria-Bertani (LB) agar plates at 30/37°C or in liquid broth at 30/37°C with shaking at 200 revolutions per minute (rpm) (Table 2.5).

2.5.2.2 Strain storage

E. coli strains were stored at -80°C with single colonies picked using a 10 µl plastic loop and re-suspended in Microbank Long Term Storage tubes (Pro-Lab Diagnostics).

2.5.2.3 Preparation of chemically competent *E. coli*

E. coli TOP10 or NEB sExpress LB broth cultures (5 ml) were inoculated directly from the storage tubes using a single bead and incubated overnight at 37°C with shaking at 200 rpm. 2 ml of bacterial overnight culture was used to inoculate 200 ml LB broth and incubated at 37°C with 200 rpm shaking until an optical cell density (OD) at 600 nm 0.3-0.5 was obtained. Bacterial culture and 100 mM $CaCl_2$ was chilled on ice for 30 minutes. Cells were split into 6 falcon tubes and harvested by centrifugation (3500 x g, 10 min, 4°C). The pellet in each tube was re-suspended in 10 ml 100 mM $CaCl_2$ and centrifugation was repeated under the same conditions. Cells were re-suspended again in 10 ml 100 mM $CaCl_2$ and incubated on ice for 30 minutes. Centrifugation was repeated under the same conditions and the final pellet in each tube was re-suspended in 1.5 ml 100 mM $CaCl_2$ and individual tubes were combined. Cells were incubated on ice for an hour before 2 ml of sterilised 50% glycerol was added into the culture. Finally, cells were dispensed into 50 µl aliquots and stored at -80°C.

2.5.2.4 *E. coli* heat shock transformation

Chemically competent *E. coli* cells (50 µl) were thawed on ice for approximately 3 minutes before 1 µl of plasmid DNA or 10 µl of ligation product was added and gently mixed by pipetting. The cells were incubated on ice for 30 minutes before heat shock at 42°C in a water bath for 45 seconds. Immediately, cells were incubated on ice for 2 minutes prior to adding 200 µl SOC medium and incubation for 1 hour at 37°C with 200 rpm shaking. Following incubation, recovered cells (200 µl) were plated onto LB

agar medium containing the appropriate antibiotic and incubated at 37°C or 30°C overnight (Table 2.5 and 2.6).

2.5.2.5 Sampling and optical density measurement of *E. coli*

For measurement of optical density at 600 nm (OD_{600nm}) at specific time points, the CLARIOstar 96 well-plate reader (BMG Labtech, Germany) was used. During growth experiments, *E. coli* cultures were grown overnight (37°C, 200 rpm) in 10 ml LB with appropriate selection to antibiotic. The next day, cultures were diluted 1:5 using LB and 200 µl of sample was applied to a 96 well-plate before the OD_{600nm} was measured. Next, the OD_{600nm} value for each sample was multiplied by 5 and all samples were diluted to OD_{600nm} = 0.01 in 30 ml LB with appropriate selection or/and inducer. Samples were grown for 24 hours (37°C, 200 rpm) and OD_{600nm} was measured every 2, 4, 6, 8 and 24 hours. Samples entering exponential phase were diluted 1:5 with LB to allow OD_{600nm} measurements to be read within the spectrophotometric range. LB was also used as a reference blank.

2.5.3 Anaerobic bacteria

2.5.3.1 Growth condition

C. difficile strains were grown under anaerobic conditions at 37°C inside the MG1000 Mark II Anaerobic Workstation (Don Whitley Scientific Ltd, UK), with an internal atmosphere of Nitrogen (80%), Carbon dioxide (10%) and Hydrogen (10%). BHIS agar or BHIS broth was used to culture *C. difficile* strains and pre-reduced before use in an anaerobic workstation for at least 3 to 12 hours respectively. BHIS agar plates were supplemented with D-cycloserine and Cefoxitin (*C. difficile* selective supplement, Oxoid) according to manufacturer instructions (Table 2.5 and 2.6).

2.5.3.2 Strain storage

C. difficile cells were grown overnight on BHIS agar plate. On the following day, a 10 µl plastic loop was used to re-suspend the cells in 1 ml BHIS broth with 10% (v/v) sterilised glycerol under anaerobic conditions. Strains were stored at -80°C using screw cap tubes.

2.5.3.3 Conjugation of plasmid DNA into *C. difficile*

Prior to conjugation, overnight cultures of donor *E. coli* NEB sExpress cells carrying the required plasmid DNA were prepared in 5 ml LB broth with appropriate antibiotics and incubated at 37°C or 30°C with 200 rpm shaking. Overnight cultures of recipient *C. difficile* cells were made in 1 ml BHIS liquid and incubated at 37°C under anaerobic conditions. 1 ml of *E. coli* overnight culture was harvested by centrifugation (1800 x g, 1 min, 25°C). The pellet was washed twice in 400 µl PBS using the same centrifugation conditions. The pellet was then transferred into the anaerobic cabinet and re-suspended in 200 µl of recipient *C. difficile* strain. The resulting mating mixture was plated onto BHIS agar plates containing no antibiotic using 10 spots of 20 µl volume and incubated anaerobically for 24 hours. On the following day, cells were scraped using a 10 µl plastic loop and re-suspended in 500 µl of sterile PBS. The cell suspension was plated onto five BHIS agar plates containing *C. difficile* supplement and appropriate antibiotic in 100 µl aliquots. Plates were incubated under anaerobic conditions for 48-72 hours.

2.5.3.4 Sampling and optical density measurement of *C. difficile*

For measurement of optical density at 600 nm at specific time points, the CLARIOstar 96 well-plate reader (BMG Labtech, Germany) was used. During growth experiments, *C. difficile* cultures were grown anaerobically overnight at 37°C in 10 ml BHIS with appropriate selection. The next day, cultures were diluted 1:5 using BHIS and 200 µl of sample was applied to a 96 well-plate before the OD_{600nm} was measured. Next, the OD_{600nm} value for each sample was multiplied by 5 and all samples were diluted to OD_{600nm} = 0.01 in 30 ml BHIS with appropriate selection or/and inducer. Samples were grown anaerobically for 24 hours at 37°C and OD_{600nm} was measured every 2, 4, 6, 8 and 24 hours. Samples entering exponential phase were diluted 1:5 using BHIS to allow OD_{600nm} measurements to be read within the spectrophotometric range. BHIS was also used as a reference blank.

2.5.4 Bacteriophages

2.5.4.1 Isolation of bacteriophages from environmental samples

Post-anaerobic digester samples were obtained from Stoke Bardolph Sewage Treatment Plant in Nottinghamshire, U.K., on 17/1/2017 and 6/11/2017 and stored at 4°C. Components of BHIS broth were added to 50 ml of sewage sample together with 0.1% taurocholate and 1% $MgCl_2$ and sample was incubated anaerobically overnight at 37°C. The following day, 40 ml of sample was split into 2 falcon tubes, centrifuged (10 000 x g, 10 min, 4°C). The resulting supernatant was filter sterilised using a 0.22 µm membrane and stored at 4°C, until a plaque assay (section 2.5.4.2) was used to detect phages in the supernatant. Following a plaque assay, a single phage plaque was picked from BHIS agar plate using a sterile pipette, re-suspended in 900 µl BHIS broth and incubated for 1 hour. Individual plaque suspension was then filter sterilised using a 0.22 µm membrane. This procedure was repeated 3 times to purify the phage.

2.5.4.2 Plaque assay

Double agar overlay plaque assay was described by Clokie and Kropinski (2009) and used with some modification in this study. Overnight culture of a *C. difficile* indicator strain was made in 20 ml BHIS broth and incubated at 37°C anaerobically. A falcon tube with 20 ml BHIS broth was inoculated with 1% of overnight culture and incubated at 37°C for 6 hours anaerobically. After 6 hours, 1 ml of indicator *C. difficile* strain was mixed either with 200 µl of supernatant from environmental sample or 200 µl of individual phage plaque (section 2.5.4.1). Outside the anaerobic cabinet, 3 ml of 0.5% BHIS top agar (cooled down to 50°C) was added to the sample and the resulting mixture was plated onto 1% BHIS agar plates containing *C. difficile* supplement and left to dry for no longer than 10 minutes before plates were transferred back to the anaerobic cabinet and incubated at 37°C overnight. On the following day, phages were detected by the presence of plaques and a titre was set by their counting. To increase the titre of the purified phage, plaque assays were repeated using 0.2% BHIS top agar. Following incubation, the whole top layer of the BHIS plate was harvested using a plastic spreader, transferred into 10 ml BHIS broth and incubated for an hour. Post incubation samples were centrifuged (10 000 x g, 10

min, 25°C) and supernatant was filter sterilised using a 0.22 µM membrane. The filtered sample was stored at 4°C or used for proceeding plaque assays until the titre reached a minimum value of 10^8 PFU.ml^{-1}.

2.5.4.3 Phage storage

Phages were stored at 4°C as filtered supernatant of BHIS broth harvested from enriched environmental sample.

2.5.4.4 Preparation of *C. difficile* lysogens

To modify the phage genome, lysogens of *C. difficile* were isolated using the spot of the lawn method (Govind *et al*., 2009). *C. difficile* strains were grown anaerobically as a lawn on 0.5 % BHIS agar plate overnight. The following morning, 20 µl of phage lysate (>10^8 PFU.ml^{-1}) and its 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6} dilutions were spotted onto a BHIS plate containing *C. difficile* lawn and incubated anaerobically at 37°C overnight. Next, a few colonies growing within a zone of clearing were re-streaked 3 times on 1% BHIS agar plates containing *C. difficile* supplement. Lysogens were confirmed by mitomycin C induction (section 2.5.4.5) and by plaque assay screening for immunity of potential lysogens against infection with the same phage.

2.5.4.5 Mitomycin C induction of lysogens

Phages were induced from *C. difficile* lysogen strains using Mitomycin C (Sigma Aldrich) at a final concentration of 3 µg.ml^{-1} using the method described by Sell *et al*., (1983) and Nale *et al*., (2012). Lysogens were grown overnight at 37°C in 20 ml BHIS under anaerobic conditions. The next morning, mitomycin C was added to the culture to a final concentration of 3 µg.ml^{-1} and samples were incubated for 24 hours at 37°C anaerobically. After incubation, samples were centrifuged (3 400 x g, 10 min, 4°C), the supernatant was sterilised using a 0.22 µm membrane filter, used in a plaque assay to test for the presence of phages and stored at 4°C.

2.5.4.6 Transmission electron microscopy

An ammonium acetate method (Fortier and Moineau, 2007) was used to precipitate phage lysates before transmission electron microscopy (TEM). Briefly, 1.5 ml of phage lysate (>10^8 PFU.ml^{-1}) was spun down (21 000 x g, 75 min, 25°C) and 1.4 ml of supernatant was carefully discarded. The remaining 100 µl of phage lysate was mixed

with 1 ml of ammonium acetate (0.1 M, Sigma) and the mixture was centrifuged (21 000 x g, 75 min, 25°C). Precipitation was repeated twice. Phage particles were then stored at 4°C. In the next step, 5 µl of the precipitated phage was added to a 200 mesh Formvar carbon coated copper grid (Agar Scientific). After 30 seconds of incubation, the excess liquid was removed using Whatman paper. To negatively stain the phage, 10 µl of 2 % uranyl acetate (Sigma) was added to the grid. After 30 seconds of incubation, liquid was wiped off with Whatman paper and the grid was left to dry for a few minutes before TEM analysis.

2.5.4.7 Phage host range analysis

A spot on the lawn assay (section 2.5.4.4) was used to determine the sensitivity of tested *C. difficile* strains to phage. For each sensitive strain, a double agar overlay plaque assay (Chapter 2.5.4.2) was used to establish the phage titre. Efficiency of plating (EOP) was calculated by dividing the phage titre obtained when the indicator strain tested for sensitivity was used in the plaque assay with the phage titre obtained using the *C. difficile* propagating strain.

2.5.4.8 *In vitro* phage activity assay

The activity of phages was measured *in vitro* using the protocol described by Selle *et al*., 2020. Briefly, an overnight culture of the *C. difficile* indicator strain was made in 20 ml BHIS broth and incubated at 37°C anaerobically. On the following day, *C. difficile* culture was subcultured 1:100 into BHIS and incubated anaerobically at 37°C to an OD_{600nm} of 0.2. Phage stocks were diluted to 2 x 10^8 $PFU.ml^{-1}$ in BHIS and mixed 1:1 with *C. difficile*. As a negative control, the bacterial culture was mixed 1:1 with BHIS. Samples were induced either with anhydrous tetracycline (0.25 µM) for mutant phages carrying endogenous CRISPR miniarray or with theophylline (2 mM) and lactose (10 mM) for mutant phages carrying *sasp* cargo. Following the induction, all cultures were grown at 37°C anaerobically for 22 hours. Every 2, 4, 6 and 22 hours samples were serially diluted up to 10^{-6} and 5 µl of each dilution was spotted onto 1% BHIS agar plates containing *C. difficile* supplement and appropriate inducer. The plates were incubated anaerobically at 37°C overnight. On the next day, the number of colonies in the densest countable spot was counted and $CFU.ml^{-1}$ was calculated for each sample.

2.6 General molecular tools

2.6.1 Isolation of genomic DNA from *C. difficile*

Genomic DNA was isolated from *C. difficile* using GenElute™ Bacterial Genomic DNA Kit (Sigma). Overnight culture of *C. difficile* (1 ml) was centrifuged (16 000 x g, 5 min, 25°C), the pellet was re-suspended in 200 µl lysis buffer (10 $mg.ml^{-1}$ lysozyme in PBS) and incubated for 30 minutes at 37°C. After the incubation, 20 µl of RNase A was added to the mixture and incubated for 2 minutes at room temperature. A 20 µl of Proteinase K (20 $mg.ml^{-1}$) followed by 200 µl of lysis solution C was added to the mixture and incubated for 10 minutes at 55°C. Following this, 200 µl of absolute ethanol was added and sample was mixed thoroughly until homogenous. The lysate was transferred to the column and centrifuged (7000 x g, 1 min, 25°C). The column was washed two times by centrifugation (13 000 x g, 3 min, 25°C) prior to elution of DNA in 35 µl dH_20.

2.6.2 Phenol-Chloroform isolation of genomic DNA from *C. difficile*

The Phenol-Chloroform method of DNA extraction (Sambrook and Russel, 2001) was used before sequencing of genomic DNA due to higher yield and volume of samples. An overnight culture of *C. difficile* (5 ml) was centrifuged (16 000 x g, 3 minutes, 25°C), the pellet was re-suspended in 180 µl lysis solution (10 $mg.ml^{-1}$ lysozyme in PBS) and incubated for 30 minutes at 37°C. After the incubation, 4 µl of RNase A was added to the mixture and incubated for 15 minutes at room temperature. A 20 µl of Proteinase K (20 $mg.ml^{-1}$) followed by distilled water (85 µl) and 10% SDS was added to the mixture and incubated for 10 minutes at 65°C. After the incubation, samples were transferred to the fume hood before adding 400 µl of phenol:chloroform:isoamylalcohol (25:24:1) to the samples, mixing thoroughly by inversion and transferring to a Phase lock gel tube (Eppendorf) preceding centrifugation (13 000 x g, 3 min, 25°C). After separation, the top phase of the mixture was transferred to another phase-lock tube and the extraction was repeated two more times. Following this, 40 µl of sodium acetate (3 M) and 800 µl of ice cold ethanol (100%) were added to the final supernatant. The sample was gently mixed and incubated for 30 minutes at -80°C. Following the incubation, the sample was

centrifuged (16 000 x g, 15 minutes, 4°C) and the pellet was re-suspended in 1 ml ethanol (70%) before another centrifugation step was performed (16 000 x g, 15 minutes, 4°C). Next, the supernatant was quickly removed and the pellet was left to dry at room temperature for 45 minutes, before being re-suspended in 50 µl sterile, distilled water and stored at -20°C.

2.6.3 Isolation of genomic DNA from the phage

Genomic DNA was isolated from the phage using a RBC Bioscience Viral Nucleic Acid Extraction kit (YVN100) according the manufacturer's instructions. Pure phage lysate of high titre (>10^8 PFU.ml^{-1}) was used. Phage lysate (200 µl) was mixed together with VB buffer (400 µl) and the mixture was incubated for 10 minutes at room temperature. Ethanol (95 %, 500 µl) was added to the lysate and mixed thoroughly by vortexing. The mixture was then passed through a column twice by centrifugation (10 000 x g, 1 min, 25°C). Column was washed two times by centrifugation (10 000 x g, 30 s, 25°C) using W1 buffer (400 µl) and R-Wash buffer (600 µl). Finally, DNA was eluted in 50 µl dH_2O and stored at 4°C.

2.6.4 Phenol-Chloroform isolation of genomic DNA from the phage

Isolation of viral genomic DNA was performed using a modified Phenol-Chloroform method (Sambrook and Russell, 2001; Nale *et al.* 2016). Phage lysate (2 ml) of minimal titre 10^8 PFU.ml^{-1} was transferred into the 15 ml tube. A volume of 25 µl of $MgCl_2$ (1 M) was added to the tube and mixed gently before adding 0.8 µl DNAseI (2000 U.ml^{-1} , Thermo Fisher Scientific) and 20 µl RNAseA (10 mg.ml^{-1}, Thermo Fisher Scientific) to the content of the tube. The sample was mixed briefly using a vortex and incubated for 30 minutes at room temperature. Following this, 80 µl EDTA (0.5 M), 5 µl proteinase K (20 mg.ml^{-1}) and 100 µl SDS (10%) were added and the sample was mixed again using a vortex and incubated for an hour at 55°C with agitation after every 20 minutes. Mixture was then split into 4 phase lock gel tubes (Eppendorf) before being transferred into the fume hood. An equal volume of phenol:chloroform:isoamylalcohol (25:24:1) was added to the sample, mixed thoroughly by inversion and centrifuged (10 000 x g, 10 min, 25°C). After separation, the top layer of samples were transferred to another set of phase-lock tubes and the

extraction was repeated two more times. Following this, 40 µl of sodium acetate (3 M) and 800 µl of ice cold ethanol (100%) were added to final supernatants. The sample was gently mixed and incubated for 5 minutes on ice to aid DNA precipitation. Following incubation, the sample was centrifuged (10 000 x g, 10 minutes, 25°C) and the pellet was washed in 1 ml ethanol (70%) before another centrifugation step was performed (10 000 x g, 10 minutes, 25°C). Supernatant was quickly removed and the pellet was left to dry at room temperature for 45 minutes, before being re-suspended in 50 µl sterile, distilled water and stored at 4°C.

2.6.5 Plasmid DNA extraction from *E. coli*

Plasmid DNA was isolated from *E. coli* harbouring the plasmid using the Monarch Plasmid DNA Miniprep Kit (NEB), in accordance with the manufacturer's instructions. Briefly, 5 ml of *E. coli* overnight culture was centrifuged (16 000 x g, 5 min, 25°C), the pellet was re-suspended in lysis solution and incubated for 5 minutes at room temperature. After incubation, the sample was neutralised and passed through a column by centrifugation (16 000 x g, 1 min, 25°C). The plasmid DNA was then washed and precipitated using ethanol and eluted in 40 µl of sterile dH20 under the same centrifugation conditions. Plasmid DNA was stored at -20°C until required.

2.6.6 Measurement of DNA concentration

Concentration of plasmid DNA or PCR samples were measured using a nanodrop ND-1000 spectrophotometer (Thermo Scientific) according to manufacturer's instructions. The concentration of genomic DNA samples were determined using a Qubit 3.0 Fluorometer (Thermo Scientific), in accordance with the manufacturer's instructions.

2.6.7 Polymerase Chain reaction (PCR)

Polymerase chain reaction (PCR) was performed using either OneTaq Quick-Load 2X Master Mix (NEB) or Phusion Green Hot Start II High-Fidelity PCR Master Mix (Thermo Fisher). Both polymerases were used according to manufacturer's instructions and reactions were performed in a Biometra TRIO Thermocycler. The composition of a typical 20 µl reaction and thermocycling conditions are outlined in table 2.7 and table

2.8, 2.9 respectively. Annealing temperature of primers were determined using the online calculator (NEB) available at https://tmcalculator.neb.com. The duration of the elongation step was dependent on the size of the amplified fragment, in general 1 minute per 1 kb of product in the case of One Taq polymerase and 30 seconds per 1 kb when the Phusion polymerase was used.

Table 2.7: PCR components for OneTaq Quick-Load 2X Master Mix (NEB) or Phusion Green High-Fidelity DNA Polymerase Master Mix (Thermo Fisher).

PCR component	Volume
Master Mix	10 µl
Forward primer (10 uM)	1 µl
Reverse primer (10 uM)	1 µl
template	10-50 ng
nuclease free H2O	to 20 µl

Table 2.8: Thermocycling conditions for PCR performed using OneTaq Quick-Load 2X Master Mix.

Reaction step	Temperature	Duration	Number of cycles
Initial denaturation	94°C	30 sec	1
Denaturation	94°C	30 sec	25
Annealing	X	1 min	
Elongation	72°C	1 min/ 1 kb	
Final elongation	72°C	5 min	1
Hold	15°C	-	

Table 2.9: Thermocycling conditions for PCR performed using Phusion Green Hot Start II High-Fidelity PCR Master Mix.

Reaction step	Temperature	Duration	Number of cycles
Initial denaturation	98°C	30 sec	1
Denaturation	98°C	30 sec	25
Annealing	X	1 min	
Elongation	72°C	30 s/ 1 kb	
Final elongation	72°C	5 min	1
Hold	15°C	-	

2.6.8 Colony PCR

Colony PCR was performed for the purpose of strain screening using OneTaq Quick-Load 2X Master Mix Polymerase (see chapter 2.6.7). Single colonies of *E. coli* were

picked from the agar plate using a sterile micropipette tip, re-suspended in 50 µl of LB (*E. coli*) or BHIS (*C. difficile*) and 1 µl of suspension was used as a template for PCR reaction. The typical PCR protocol was followed, as listed in Table 2.8.

2.6.9 Splicing by overhang extension PCR

Splicing by overhang extension (SOE) PCR was used (Horton *et al*., 1990) for cloning purposes in order to splice two DNA sequences. Phusion Green Hot Start II High-Fidelity PCR Master Mix (Thermo Fisher Scientific) was used together with specific primers with shared homology that function like bridges between the two molecules of DNA. The typical PCR protocol was followed, as listed in Table 2.9.

2.6.10 Restriction digest

DNA samples were digested using restriction endonucleases and buffers (NEB) following the instructions of the manufacturer. Briefly, 1 µl of enzyme (10 000 U), together with 5 µl CutSmart® Buffer was used to cut 1 µg of DNA. Reactions were made up to a total 45 µl volume with dH2O and incubated at 37°C overnight to ensure complete digestion.

2.6.11 Agarose gel electrophoresis

Visualisation and separation of DNA molecules was performed using electrophoresis (400 A, 100 V, 60 min) through 1% agarose gel in 1 x TAE buffer. Agarose gel was made with 1 x TAE buffer and 0.01% (v/v) SYBR Safe DNA Gel Stain (Thermo Scientific). If needed, purple gel loading dye (6X, NEB) was added to the sample before separation and Quick-Load® Purple 2-Log DNA ladder (NEB, 0.1-10.0 kb) was used as a reference. DNA molecules were visualised under blue light using the Gel Doc™ XR+ Gel Documentation System and software (BioRad).

2.6.12 Extraction of DNA from agarose

After separation, if required, DNA molecules were cut from the agarose gel under blue light using a scalpel. The Monarch® DNA Gel Extraction Kit (NEB) was used to purify samples following the manufacturer's instructions. Briefly, excised piece of gel was dissolved in Dissolving Buffer at 55°C for 10 minutes before it was passed through

a column by centrifugation (16 000 x g, 1 min, 25°C). The DNA sample was then washed and precipitated using ethanol and eluted in 16 µl sterile dH_20 under the same centrifugation conditions. DNA samples were stored at -20°C, until required.

2.6.13 DNA ligation

Digested and purified plasmid DNA and inserts with compatible ends were ligated using T4 DNA ligase (Promega) according to manufacturer's instructions in a 10 µl reaction volume. Samples were incubated at 4°C overnight before transformation into *E. coli* cells.

2.6.14 Generation of mutant strains using RiboCas9

The inducible RiboCas9 genome editing tool used in this study is based on work of I. Canadas (Canadas *et al*., 2019). The RiboCas9 vectors designed and constructed in this study are derived from the pRECas1-IIE-Cdif plasmid created by Canadas *et al*., 2019 and contain the Cas9 enzyme under the inducible riboswitch system controlled by theophylline as an inducer.

RiboCas9 plasmids with the specific sgRNA and editing templates were transferred into *C. difficile* CD9301 or its lysogens using conjugation with the *E. coli* sExpress donor strain harbouring the plasmid (see section 2.5.3.3). Conjugal cells were selected on a BHIS agar plate supplemented with *C. difficile* supplement and thiamphenicol and incubated anaerobically at 37°C for 48 hours. Individual colonies were re-streaked onto BHIS agar containing *C. difficile* supplement, thiamphenicol and theophylline as a Cas9 inducer. Resulting single colonies were screened via colony PCR with primers binding in the chromosome upstream and downstream of the homology arms. Following this, single colonies were subcultured 3 x every 24 hours, serially diluted and 100 µl of 10^{-5} dilutions were plated onto BHIS agar containing *C. difficile* supplement. The following morning, single colonies were re-streaked onto BHIS agar plates with *C. difficile* supplement and the same media with additional thiamphenicol, to screen for thiamphenicol sensitive colonies that have lost the plasmid. Mutant strains were further confirmed by Sanger sequencing and stored in BHIS broth supplemented with 10% glycerol at -80°C.

2.6.15 Generation of mutant strains using Allele Coupled Exchange (ACE)

ACE was used to insert the gene of interest into *C. difficile* CD9301 or *C. difficile* lysogens according to the protocol described by Cartman *et al*. (2012). The ACE plasmids designed and constructed in this study are derived from the pMTL_SC7315 plasmid created by Stephen Cartman (Cartman *et. al.*, 2012).

ACE plasmids were transferred into the host strain using conjugation with the *E. coli* sExpress donor strain harbouring the plasmids (see section 2.5.3.3). Conjugal cells were selected on BHIS agar plate supplemented with *C. difficile* supplement and thiamphenicol and incubated anaerobically at 37°C for 48 hours. Faster growing single crossover colonies were re-streaked to purity onto the same medium. To allow second recombination event to occur, colonies were re-streaked 2-3 times onto nonselective BHIS medium and incubated anaerobically at 37°C for 24 hours. In the following step, single colonies were grown overnight in liquid BHIS, serially diluted to 10^{-5} and 100 µl of each dilution was plated on CDMM supplemented with 5-FC. After 24 hours, clones resistant to 5-FC were patch plated onto BHIS agar supplemented with *C. difficile* supplement and the same media containing thiamphenicol to screen for plasmid loss. 5-FC resistant clones sensitive to thiamphenicol were then analysed by PCR using flanking primers binding within the genome upstream and downstream of homology arms. Mutant strains were analysed by sequencing the PCR amplified region around the inserted gene and stored in BHIS broth supplemented with 10% glycerol at -80°C.

2.6.16 Construction of *ΔpyrE* mutants using the endogenous CRISPR

Plasmids used to redirect the *C. difficile* endogenous CRISPR are derivates of pMTL83151 and contain a CRISPR miniarray under the control of a tetracycline inducible system and editing template. Plasmids were conjugated into the *C. difficile* host using the *E. coli* sExpress donor strain carrying the plasmids (see chapter 2.5.3.3). Transconjugants were selected on BHIS agar plate supplemented with *C. difficile* supplement and thiamphenicol and incubated anaerobically at 37°C for 48 hours. To induce the expression of CRISPR miniarray, single colonies were subsequently restreaked onto BHIS media supplemented with *C. difficile* supplement,

thiamphenicol and anhydrous tetracycline. Following this, single colonies were grown overnight in BHIS. Next day, serial dilutions were made and 100 µl of 10^{-5} dilution for each colony were plated onto CDMM supplemented with 5-FOA and uracil. Single colonies were screened via colony PCR with primers binding in the genome upstream and downstream of the homology arms. Mutant colonies were then grown overnight, subcultured 3x every 24 hours, serially diluted and 100 µl of 10^{-5} dilutions were plated on BHIS agar containing *C. difficile* supplement. Following morning, single colonies were re-streaked onto BHIS agar plates with *C. difficile* supplement and the same media with additional thiamphenicol to screen for thiamphenicol sensitive colonies that have lost the plasmid. Mutant strains were further confirmed by Sanger sequencing and stored in BHIS broth supplemented with 10% glycerol at -80°C.

2.6.17 Re-directing the endogenous CRISPR to autoimmunity

Plasmids used to redirect the *C. difficile* endogenous CRISPR to autoimmunity are derivates of pMTL83151 and contain the CRISPR miniarray under the control of a tetracycline inducible system. Plasmids were conjugated into the *C. difficile* host strain using the *E. coli* sExpress donor strain carrying the plasmids (see chapter 2.5.3.3). Transconjugants were plated either on BHIS agar plate supplemented with *C. difficile* supplement and thiamphenicol or BHIS agar plate supplemented with *C. difficile* supplement, thiamphenicol and anhydrous tetracycline and incubated anaerobically at 37°C for 48 hours before counting the number of colonies.

2.6.18 Determination of the conjugation efficiency

The effect of the re-directing the endogenous CRISPR of *C. difficile* to autoimmunity was assessed by determining the conjugation efficiency. Before the conjugation, additional 1 mL of donor strain was pelleted and washed in PBS. The pellet was resuspended in 1 mL of PBS, serially diluted and plated (100 µl) onto LB agar plates with appropriate selection. This allowed the enumeration of the donor strain by Colony Forming Units ($CFU.ml^{-1}$) the following day. After the conjugation, $CFU.ml^{-1}$ was established for each transconjugant by counting the number of colonies on the selective plate 48 hours after incubation. Conjugation efficiency was calculated per donor strain by dividing $CFU.ml^{-1}$ of transconjugants obtained following the transfer

of plasmids containing endogenous CRISPR miniarrays or control plasmid by CFU.ml^{-1} of donor cells.

$$\text{Conjugation efficiency per donor} = \frac{CFU/ml\ transconjugants}{CFU/ml\ donor}$$

2.7 Plasmid cloning

2.7.1 Construction of plasmids used in Chapter 4

pMTL_SC plasmids

ACE plasmids were constructed by cloning homology arms and cargo into pMTL_SC7315 backbone separately. Homology arms were generated in a two-step SOE PCR. In first step of reaction, homology arms were amplified from ΦCD9301 genomic DNA using flanking primers with overhangs containing *XmaI* and *XhoI* restriction sites and the internal primers carrying overhangs complementary to the other homology arm and *Kas*I and *Avr*II restriction sites. Left homology arms for locations 1A, 1B, 2A, 2B and 3 were amplified using primer sets ZZ109 and ZZ110, ZZ81 and ZZ82, ZZ87 and ZZ88, ZZ113 and ZZ114, ZZ91 and ZZ92 respectively. Right homology arms were amplified using primers ZZ111 and ZZ112, ZZ83 and ZZ84, ZZ89 and ZZ90, ZZ115 and ZZ116, ZZ93 and ZZ94. In second reaction, homology arms were spliced together through shared homology using flanking primers ZZ109/ZZ112, ZZ81/ZZ84, ZZ89/90, ZZ113/ZZ116 and ZZ91/94 for the locations 1A, 1B, 2A, 2B and 3. In the following step, homology arms and pMTLSC7315 backbone vector were digested using *Xma*I and *Xho*I enzymes, ligated and transformed into the competent *E. coli.* Resulting plasmids were named pMTL_SC1A, pMTL_SC1B, pMTL_SC2A, pMTL_SC2B and pMTL_SC3.

pMTL_SC:*ermB* plasmids

The T1T2 terminator was amplified from plasmid pJ202-WOK016 using forward primer ZZ105 with *Kas*I restriction site and reverse primer ZZ117 with overhangs complementary to *ermB* sequence. *ermB* gene was amplified together with the promoter from modular vector pMTL83251 using forward primer ZZ107 with

overhangs complementary to T1T2 terminator and reverse primer ZZ108 with *Avr*II restriction site. T1T2 terminator was then spliced together with *ermB* using flanking primers ZZ105 and ZZ108 and cargo sequence was cloned into 5 pMTL_SC_HA plasmids using *Kas*I and *Avr*II restriction sites. The resulting plasmids were sequenced and named pMTL_SC1A:ermB, pMTL_SC1B:ermB, pMTL_SC2A:ermB, pMTL_SC2B:ermB and pMTL_SC3:ermB.

pMTL_RC plasmids

pMTL_RC plasmids were constructed by cloning homology arms and sgRNA into the backbone vector pRECas1-IIE-Cdif (Canadas *et al*., 2019) in one step ligation reaction. Homology arms were generated in a two-step SOE PCR. In first step of reaction, homology arms were amplified from ΦCD9301 genomic DNA using flanking primers with overhangs containing *AatII* and *AscI* restriction sites and the internal primers carrying overhangs complementary to the other homology arm. Left homology arms for locations 4A, 4B, 5A, 5B were amplified using primer sets ZZ197/ZZ231, ZZ201/ZZ233, ZZ205/ZZ235, and ZZ209/ZZ237 respectively. Right homology arms were amplified using primers ZZ232/ZZ200, ZZ234/ZZ204, ZZ236/ZZ208, and ZZ238/ZZ212. In second reaction, homology arms were spliced together through shared homology using flanking primers ZZ197/ZZ200, ZZ201/ZZ204, ZZ205/ZZ208, and ZZ209/ZZ212. sgRNAs were generated in PCR reaction using primer dimers with shared homology. Forward primers ZZ164, ZZ165, ZZ166, and ZZ167 contained *Sal*I restriction site and variable 20 nt seed region for locations 4A, 4B, 5A, 5B. Reverse primer ZZ51 contained complementary sequence to forward primer and *Aat*II restriction site. Finally, individual sgRNA and homology arms were ligated directly into the pMTL83151_RCas9 vector using *Sal*I, *Aat*II and *Asc*I restriction sites in 1:5:3 ratio (vector: homology arm: sgRNA). Plasmids were confirmed by sequencing and named pMTL_RC4A, pMTL_RC4B, pMTL_RC5A and pMTL_RC5B.

Table 2.10: Seed sgRNA sequences used to target locations 4A, 4B, 5A and 5B within ΦCD9301.

Name	Sequence	PAM	Position	Strand
HA4A	gcagagaggattttgaataa	AGG	3893	+
HA4B	gtttacaaacagttgaagga	TGG	4556	+
HA5A	aatttaactggtgaagagct	TGG	13395	-
HA5B	gaatataaggtcgataaacc	GGG	12194	-

pMTL_RC:*ermB* plasmids

T1T2 terminator was amplified from plasmid pJ202-WOK016 using forward primer ZZ105 containing *Kas*I restriction site and reverse primer ZZ117 with overhangs complementary to *ermB* cargo. *ermB* cargo was amplified together with the promoter from modular vector pMTL83251 using forward primer ZZ107 with overhangs complementary to T1T2 terminator and reverse primer ZZ108 containing *Avr*II restriction site. T1T2 terminator was then spliced together with *ermB* using primers ZZ105 and ZZ108 and cloned into pMTL_RC plasmids using *Kas*I and *Avr*II restriction sites. The resulting plasmids were sequenced and named pMTL_RC4A:ermB, pMTL_RC4B:ermB, pMTL_RC5A:ermB, pMTL_RC5B:ermB.

pMTL83151_Rb3_*sasp* plasmids

saspA or *saspB* genes from *B. subtilis* had been synthesized by DNA 2.0. *saspA* gene was amplified with forward primer ZZ328b containing *Xma*I restriction site and T1 terminator and reverse primer ZZ304 with shared homology to Pfdx_Rb3 promoter. Pfdx_Rb3 promoter was amplified using forward primer ZZ305 containing shared homology to *saspA* gene and ZZ329 with *Sal*I restriction site. Pfdx promoter was then spliced together with *saspA* in SOE PCR reaction using flanking primers ZZ328b/ZZ329. Similarly *saspB* gene was amplified with forward primer ZZ328a containing *Xma*I restriction site and T1 terminator and reverse primer ZZ306. Pfdx_Rb3 promoter was amplified using forward primer ZZ307 containing shared homology to *saspB* gene and ZZ329 with *Sal*I restriction site. Pfdx promoter was then spliced together with *saspB* in SOE PCR reaction using flanking primers ZZ328a/ZZ329.

Finally, genes were cloned into the pMTL83151 backbone via *Xma*I/*Sal*I restriction sites, sequenced and named pMTL83151_PfdxRb3_saspA and pMTL83151_PfdxRb3_saspB.

pMTL83151_RiboLac_*sasp* plasmids

RiboLac system was amplified from pMTL-ICG1 plasmid using forward primer ZZ326 with *Xma*I restriction site and reverse primer ZZ315. *saspA* gene was amplified using forward primer ZZ316 with shared homology to the RiboLac system and reverse primer ZZ327a containing *Sal*I restriction site. Both were spliced together using primer set ZZ326/ZZ327a and cloned into pMTL83151 plasmids via *Xma*I and *Sal*I restriction sites. Plasmid was sequenced and named pMTL83151_RiboLac_saspA.

Similarly RiboLac system was amplified from pMTL_ICG1 using forward primer ZZ326 with *Xma*I restriction site and reverse primer ZZ312 with homology to *saspB*. *saspB* was amplified using forward primer ZZ313 and reverse primer ZZ327b containing *Sal*I restriction site. Both were spliced together using primer set ZZ326/ZZ327b and cloned into pMTL83151 plasmids via *Xma*I and *Sal*I restriction sites. Plasmid was sequenced and named pMTL83151_RiboLac_saspB.

pMTL_RC_RiboLac_*sasp* plasmids

RiboCas9 plasmids were used as a backbone to knock in RiboLac_*sasp* cargo sequences into the lysogen. Both *saspA* and *saspB* genes were amplified together with RiboLac promotor system and terminators from pMTL83151_RiboLac_saspA or pMTL83151_RiboLac_saspB using primer sets ZZ318/ZZ317 containing *Kas*I and *Avr*II sites. Cargos were individually cloned into the pMTL_RC4A or pMTL_RC4B plasmids, sequenced and named pMTL_RC4A_RiboLac_saspA, pMTL_RC4A_RiboLac_saspB, pMTL_RC4B_RiboLac_saspA and pMTL_RC4B__RiboLac_saspB.

pMTL_SC_Rb3_*sasp* plasmids

Homology arms targeting location 4A were amplified from pMTL_RC4A plasmid using primer set ZZ300/ ZZ301 and cloned into the pMTL7315 backbone via *Xma*I and *Xho*I restriction sites. Following this, Rb3_saspA or Rb3_saspB cargo sequences were amplified from pMTL83151_PfdxRb3_saspA or pMTL83151_PfdxRb3_saspB plasmids

using primer sets ZZ275/ZZ276 or ZZ277/ZZ276 respectively and cloned into the pMTL_SC4A backbone using *Kas*I and *Avr*II restriction enzymes. Plasmids were verified by sequencing and named pMTL_SC4A_Rb3_sapA and pMTL_SC4A_Rb3_saspB.

2.7.2 Construction of plasmids used in Chapter 5

pMTL_RCas_*tcdA* plasmids

Left homology arm was amplified from *C. difficile* genomic DNA using forward primer containing *AsiS*I restriction site and reverse primer ZZ40 with homology region to right homology arms. Right homology arm was amplified with forward primer ZZ41 with shared homoloy to left arm and reverse primer ZZ42 containing *Asc*I restriction site. A knock-out cassette was spliced together using primers ZZ17 and ZZ42 and double digested using *AsiS*I and *Asc*I restriction endonucleases.

Four seed sgRNA sequences within the *tcdA* gene of *C. difficile* genome were identified using Benchling or CRISPy web. Four sgRNAs were generated in PCR reactions using forward primers ZZ29, ZZ30, ZZ31 or ZZ32 and reverse primer ZZ37.

Table 2.11: Seed sgRNA sequences used to target tcdA gene within *C. difficile* CD9301.

Name	Sequence	PAM	Position	Strand
tcdA_1	taataaatagtttttatcga	GGG	793852	-
tcdA_2	atacttaatagtaacttacc	AGG	794328	-
tcdA_3	tccttcaagagtattttggt	GGG	795880	+
tcdA_4	ctttatgcctgatactgcta	TGG	800224	+

sgRNAs was digested with *Sal*I and *AsiS*I restriction endonucleases and cloned into pRECas1-IIE-Cdif backbone together with the knock out cassette via restriction digest at *Sal*I *and Asc*I sites, generating vectors pMTL_RCas_tcdA1, pMTL_RCas_tcdA2, pMTL_RCas_tcdA3 and pMTL_RCas_tcdA4.

pMTL_RCas_*tcdB* plasmids

The knock-out cassette was amplified from *C. difficile* genomic DNA using SOE PCR. LHA was PCR amplified using primers ZZ43 and ZZ44 containing flanking *Asi*SI restriction site and short region of homology to RHA. RHA was amplified using primers ZZ45 carrying overlapping homology sequence to LHA and ZZ46 with *Asc*I restriction site. Homology arms were spliced using primer set ZZ43/ZZ46.

Four individual sgRNA targeting *tcdB* gene were assembled using a primer dimer PCR reaction with forward primers ZZ33, ZZ34, ZZ35, ZZ36 and reverse primer ZZ37.

Table 2.12: Seed sgRNA sequences used to target tcdB gene within *C. difficile* CD9301.

Name	Sequence	PAM	Position	Strand
tcdB_1	cgtataagagatcactacaa	AGG	787506	+
tcdB_2	acatcatagatagataagtg	TGG	787330	-
tcdB_3	gaaacctagttcagtaacag	TGG	784655	+
tcdB_4	catatgttcctcctgagcca	TGG	787686	-

Individual sgRNAs were digested using using *AsiS*I and *Sal*I restriction enzymes and cloned together with homology arms into pRECas1-IIE-Cdif backbone together via *Sal*I *and Asc*I sites, generating vectors pMTL_RCas_tcdB1, pMTL_RCas_tcdB2, pMTL_RCas_tcdB3 and pMTL_RCas_tcdB4.

pMTL_RCas_*tcdBA* plasmids

Homology arms were amplified from *C. difficile* genomic DNA. LHA was PCR amplified using primer set ZZ47/ZZ76 containing *Aat*II restriction site and homology to the right homology arm and RHA was amplified using primers ZZ75 and ZZ42 containing homology to left arm and *Asc*I site. Homology arms were in second step spliced together in SOE PCR and digested using *Aat*II and *Asc*I restriction enzymes.

Table 2.13: Seed sgRNA sequences used to target tcdB and tcdA gene within *C. difficile* CD9301.

Name	Sequence	PAM	Position	Strand
tcdBA_1	taaatgtgaaatctggcgaa	TGG	787244	+
tcdBA_2	tattctttccatggctcagg	AGG	787689	+
tcdBA_3	aagctgccgcatatgcttca	GGG	785148	+
tcdBA_4	gttttctacttctctgattc	TGG	785344	+

Four sgRNAs were generated in PCR reactions using forward primers ZZ52, ZZ53, ZZ54, ZZ55 and reverse primer ZZ51. Each sgRNA digested using *Sal*I and *Aat*II restriction enzymes. Next step, homology arms and sgRNA were cloned together into pRECas1-IIE-Cdif backbone using *Sal*I and *Asc*I restriction sites, resulting in vectors pMTL_RCas_tcdBA1, pMTL_RCas_tcdBA2, pMTL_RCas_tcdBA3 and pMTL_RCas_tcdBA4.

pMTL_P_{tet}_*pyrE*PS plasmids

Tetracycline inducible system was amplified from pMTL_CW21 with primers ZZ332 and ZZ333 and cloned into the pMTL83151 backbone via *Not*I and *BamH*I restriction enzymes. Resulting plasmid was sequenced and named pMTL83151_ P_{tet}. Following this, left homology arm flanking *pyrE* gene of *C. difficile* CD9301 was amplified using forward primer P1 containing *Sal*I site and P2 containing homology region to right arm. Right homology arm was amplified with primer P3 containing homology region to left arm and P4 primer with *Asc*I restriction site. Both arms were spliced together in SOE PCR reaction using flanking primers. 5 CRISPR miniarrays were synthetized in primer-dimer PCR reaction using forward primers P5, P7, P9, P11 and P13 containing *BamH*I site and respective reverse primers P6, P8, P10, P12 and P14 containing *Sal*I restriction site. Finally, CRISPR miniarrays were cloned into the pMTL83151_P_{tet} plasmid together with homology arms using *BamH*I, *Sal*I and *Asc*I restriction sites in 1:5:3 ratio (vector: homology arm: miniarray). Plasmids were confirmed by colony PCR, sequenced and named pMTL_P_{tet}_*pyrE*SP1, pMTL_P_{tet}_*pyrE*SP2, pMTL_P_{tet}_*pyrE*SP3, pMTL_P_{tet}_*pyrE*SP4 and pMTL_P_{tet}_*pyrE*SP5.

pMTL_P_{tet}_*tcdBASP* plasmids

CRISPR miniarrays were synthetized in primer-dimer PCR reaction using combinations of primers ZZ354/ZZ340, ZZ355/ZZ342, ZZ356/ZZ344 and ZZ358/ZZ353. Following this, CRISPR miniarrays were digested and cloned into the pMTL83151_P_{tet} backbone using *BamH*I and *Sal*I restriction enzymes. Plasmids were confirmed by sequencing and named pMTL_P_{tet}_tcdBASP1, pMTL_P_{tet}_tcdBASP2, pMT_P_{tet}_tcdBASP3 and pMTL_P_{tet}_tcdBASP4.

pMTL_RC_P_{tet}_*tcdBASP* plasmids

CRISPR miniarrays tcdBASP1 and tcdBASP3 targeting toxin genes of *C. difficile* were amplified together with their tetracycline inducible system from pMTL83151_P_{tet}_tcdBASP1 and pMTL83151_P_{tet}_tcdBASP3 plasmids using forward primer ZZ393 containing *KasI* restriction site and T1 terminator and reverse primer ZZ394 containing *AvrII* restriction site. CRISPR miniarrays were digested and cloned into the backbone vectors pMTL_RC4A, pMTL_RC4B, pMTL_RC5A. Plasmids were verified by Sanger sequencing and named pMTL_RC4A_P_{tet}_tcdBAPS3, pMTL_RC4B_P_{tet}_tcdBAPS1, pMTL_RC4B_P_{tet}_tcdBAPS3, pMTL_RC5A_P_{tet}_tcdBAPS1 and pMTL_RC5A_P_{tet}_tcdBAPS3.

Chapter three:

Isolation and Characterisation of Novel Bacteriophages Infecting *C. difficile*

3. Isolation and Characterisation of Novel Bacteriophages Infecting *C. difficile*

3.1 Introduction

Antibiotic resistance, high frequency of relapse and the negative impact of antibiotics on the gut microbiota has led to increased interest in the use of bacteriophages against CDI. Despite phage therapy becoming more popular, research on bacteriophages active against *C. difficile* is relatively limited, partly due to challenges in their technically demanding isolation and propagation (Phothichaisri *et al*., 2018).

Phages specific to *C. difficile* were isolated for the first time in 1983 and used for strain typing purposes (Sell *et al*., 1983). Thereafter, bacteriophages have been isolated from *C. difficile* lysogens following mitomycin C induction (Mahony *et al*., 1985; Mayer *et al*., 2008; Nale *et al*., 2012; Phothichaisri *et al.*, 2018; Ramirez-Vargas *et al*., 2018). Attempts to isolate free bacteriophages from the environment initially failed (Goh *et al.*, 2005b, Nale *et al*., 2012), however, more recent attempts to isolate free phages from the faeces of CDI patients (Meessen-Pinard *et al.,* 2012) and soil samples in Iraq (Rashid *et al*., 2016) have proven to be successful. At the time of writing (June, 2021) all isolated phages infecting *C. difficile* have been temperate and able to enter both the lysogenic and lytic life cycle (Goh *et al*., 2005b; Hargreaves *et al*., 2014b; Sekulovic *et al*., 2011; Sekulovic *et al*., 2014; Heuler *et al*., 2021).

Despite the extensive screening of *C. difficile* phages from environmental samples, there has been no strictly lytic phage isolated with activity towards *C. difficile*. The ability of *C. difficile* phages to integrate into their host genome and their relatively narrow host spectrum, may require the use of a cocktail of several phages to successfully eradicate CDI in animal models (Nale *et al*., 2018). As such, isolation of novel phages and construction of phage libraries is essential for the advancement of phage therapy against CDI.

In terms of characterisation of the phage, most studies use transmission electron microscopy (TEM) to establish phage particle morphology. Other common analysis techniques include host range studies to determine phage specificity, whole genome sequencing and growth dynamics.

All *C. difficile* bacteriophages isolated to date (June, 2021) belong to the order *Caudovirales* and can be grouped by particle morphology into myoviruses (*Myoviridae*) and siphoviruses (*Siphoviridae*). Myoviruses have a contractile tail which can vary in length and can be divided into: (i) long tailed myoviruses, with tail lengths of 150-260 nm and capsid diameters between 60-70 nm; (ii) medium myoviruses, which have 110-130 nm long tails and capsid diameters of 60-70 nm, and; (iii) small myoviruses that have tail lengths of 105-110 nm and capsid diameters of 40-60 nm. The tail of siphoviruses are non-contractile and long (Hargreaves and Clokie, 2015).

The decrease in the cost of whole genome sequencing has led to improved phage characterisation. At the time of writing 29 genomes of *C. difficile* phages have been published on NCBI. Genomes of all phages isolated so far consist of double stranded DNA of variable length in either linear or circular forms. Putative integrase genes have been identified in all sequenced genomes, which suggests they can enter the lysogenic life cycle (Hargreaves and Clokie, 2014).

The analysis of phage genomes is essential in order to determine their suitability for therapeutic purposes as phages can contain toxin genes or genes that can alter the physiology of their host. For example, genomes of large phiCD211-like phages contain genes encoding putative multidrug resistance genes, spore proteases and CRISPR arrays (Garneau *et al*., 2018). CRISPR arrays have been reported in several other *C. difficile* phages (Hargreaves *et al*., 2014; Boudry *et al*., 2015; Sekulovic and Fortier, 2016) suggesting *C. difficile* phages possibly contribute to *C. difficile* resistance to invading DNA.

It has also been shown that some *C. difficile* bacteriophages can interfere with bacterial toxin production, although the impact of phages on toxin production seems to vary. Phage ΦCD119 express the RepR repressor that downregulates toxin genes located in the PaLoc (Govind *et al*., 2009). In contrast, phage ΦCD38-2 was shown to have a positive impact on transcription of all PaLoc genes leading to increased toxin production *in vitro*, although the impact of phage on toxin production was dependent on the host strain (Sekulovic *et*. al, 2011). Moreover, phage phiSemix9P1 has been shown to carry a complete binary toxin locus (CdtLoc) in its genome. CdtLoc is normally present on the chromosome of a fraction of *C. difficile* isolates, including

the hypervirulent PCR ribotype RT027 and RT078 and consists of three genes, all of which were shown to be transcribed by prophage phiSemix9P1 (Riedel *et al*., 2017).

Genome data has also led to improved phage classification and division of some myoviruses into the genus *phicd119likevirus* and *phiMMP04likevirus*, however, some *C. difficile* phages do not fall into either (Rashid *et al*., 2016). The genome of phages belonging to the *phicd119likevirus* genus contain a gene encoding the cytosine-C5 specific metylase and a cassette containing three genes encoding DnaD, a unannotated hypothetical protein and single-stranded DNA (ssDNA) binding protein. Phages within the genus *phiMMP04likevirus* possess the gene encoding a homologue to chromosomal partitioning protein ParA, usually located downstream of the lysis gene cassette and the gene encoding Clp protease. Notably, members of *C. difficile* siphoviruses also contain the gene for the ParA homologue, as well as common endolysin and holin genes (Hargreaves and Clokie, 2015).

Characterisation of phages on the basis of their host range specificity is very important for therapeutic purposes. Phage host spectrum refers to the genera, species and strains of bacteria that can be lysed by an individual phage. Although it is generally believed that *C. difficile*-specific bacteriophages infect one or few strains, some studies have shown that phages are able to infect not only several strains but also different PCR ribotypes of *C. difficile*. For example, it has been shown that the phage ΦCD6356 was able to infect 35% of 37 tested strains belonging to 5 ribotypes (Horgan *et al*., 2010) and phage ΦCD38-2 infected 47% of 207 *C. difficile* isolates that belonged to 11 ribotypes (Sekulovic *et al*., 2014). A more recent survey tested the host range of bacteriophage against 80 *C. difficile* strains that belonged to 21 ribotypes from three countries. Phages ΦCDHM3 and ΦCDHS1 used in this study were active against 39% and 38% of strains belonging to 12 and 11 ribotypes, respectively (Nale *et al*., 2016a).

Even though the host range seems to vary between isolated phages it is essential to note that the host range analysis is dependent upon the panel of strains selected for use in the array and the degree to which the strains in the panel are related.

3.1.1 Aim of this study

Natural phage therapies rely on the lytic activity of phages to eradicate infection. To this date, no strictly lytic phage infecting *C. difficile* has been isolated. To address this issue, we propose to test the concept of using temperate phage to deliver a lethal cargo to eradicate the infection. The analysis of phage is essential in order to determine their suitability for therapeutic purposes. This study aimed at the isolation of novel phages infecting *C. difficile* and their characterisation through phage morphology, analysis of their genome and characterisation of host range specificity.

3.2 Results

3.2.1 Isolation of Novel Phages from Environmental Samples

C. difficile phages were isolated using a standard double layer agar plaque assay (Clokie and Kropinski, 2009) modified by M. Whittle (PhD thesis, 2017). Samples were collected from post-anaerobic digester material (Stoke Bardolph Sewage Treatment Plant in Nottinghamshire, U.K) that previously proved successful for isolating phages in the SBRC. The *C. difficile* germinant, taurocholate was added to the sample as an enrichment agent. Various *C. difficile* indicator strains were used to test for the presence of bacteriophages in the sample. Strains were grown for 6 hours, from a 1% inoculum to obtain a confluent growth within a top agar layer, necessary to detect phage plaques. As a negative control, a plaque assay was performed for each indicator strain without the addition of sample to ensure that strains unable to produce confluent growth were not selected for phage isolation.

Post-anaerobic digester samples used in this study were obtained in January 2017 (sample A) and November 2017 (sample B). In total, 30 *C. difficile* strains were tested as indicator strains using sample A (Table 7.1 in Appendix). Strains were isolated in the U.K. and were of PCR ribotype RT002, RT014, RT015, RT027 and RT078. These PCR ribotypes were the most commonly occurring PCR ribotypes in the SBRC culture collection. Attempts to isolate phages from sewage sample A were unsuccessful with no plaques being observed during phage isolation assay. It was assumed that sample A was stored for a prolonged period of time and consequently a new sewage sample (B) was obtained and tested.

In total, 74 potential host *C. difficile* strains were used in the plaque assays with sample B (Table 7.2 in Appendix) belonging to PCR ribotype RT015, RT027 and RT078 strains. Strains belonging to PCR ribotype RT027 are clinically relevant, 'so-called' hypervirulent strains and were chosen in this assay due to their association with severe CDI and outbreaks worldwide. All *C. difficile* PCR ribotype RT027 strains in possession of the research group were tested. Using PCR ribotype RT015 and RT027 strains as potential hosts, no phage was isolated from sample B. On the contrary, plaques were observed using hypervirulent PCR ribotype RT078 indicator strains

CD3801, CD8601, CD9001, CD9301 and CD9501. In each case, plaques were picked and purified three times. Isolated bacteriophages were named ΦCD3801, ΦCD8601, ΦCD9001, ΦCD9301 and ΦCD9501 after their isolation host strains. In all cases plaques formed by these phages were approximately 1 mm diameter with small and hazy morphology (Figure 3.1).

Phage titres were increased by harvesting the top agar layer from the plate, followed by a plaque assay. The process was repeated until the high titres of $\geq 10^8$ PFU x ml^{-1} were obtained. Attempts to increase the titre of phages using the plaque assay proved to be problematic, with no direct linear correlation between number of plaque assay repetitions and the titre of individual phage being observed. The plaque assay was repeated 10 times for each phage and varying titres were obtained (Table 3.1). Only 3 out of 5 phages (ΦCD3801, ΦCD8601 and ΦCD9301) reached the required concentrations and were therefore chosen for further analysis. Bacteriophages were stored as a filter sterilised supernatant in BHIS broth at 4°C created using a 0.2% BHIS top agar plaque assay. The stability of this storage method was investigated by M. Whittle (PhD thesis, 2017) and shown to provide reasonable long-term storage for up to 6 months.

Table 3.1 List of isolated *C. difficile* phages.

Phage	Propagating strain	Ribotype	Titre (PFU x ml^{-1})
ΦCD3801	*C. difficile* CD3801	RT078	6.5×10^8
ΦCD8601	*C. difficile* CD8601	RT078	3.2×10^8
ΦCD9001	*C. difficile* CD9001	RT078	1.02×10^3
ΦCD9301	*C. difficile* CD9301	RT078	2.15×10^9
ΦCD9501	*C. difficile* CD9501	RT078	3.6×10^3

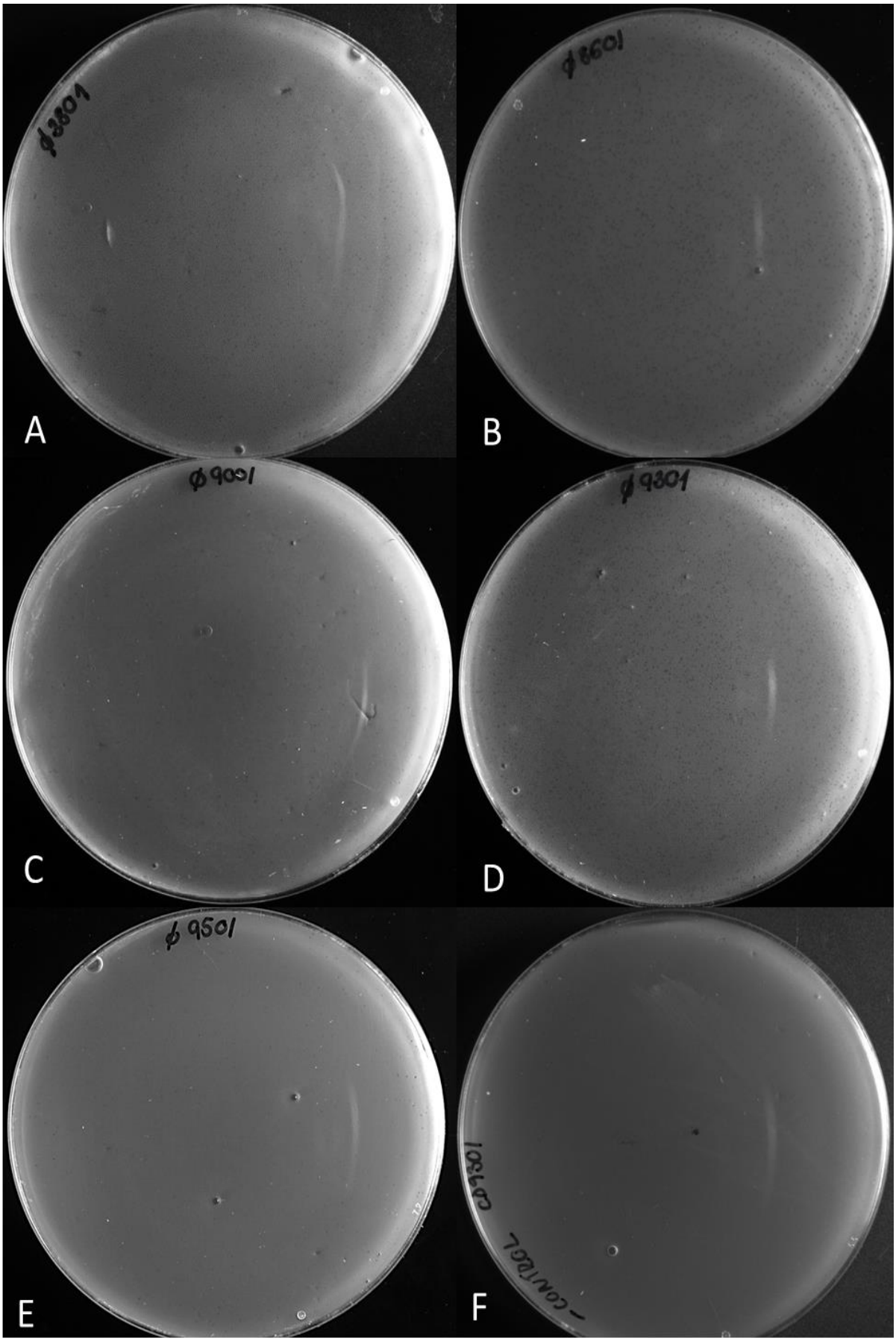
ϕ3801
A
ϕ8601
B
ϕ9001
C
ϕ9301
D
ϕ9501
E
F

Figure 3.1: Plaques of isolated bacteriophages. A. ΦCD3801 phage stock infecting *C. difficile* CD3801, **B**. ΦCD8601 phage stock infecting *C. difficile* CD8601 **C**. ΦCD9001 phage stock infecting *C. difficile* CD9001, **D**. ΦCD9301 phage stock infecting *C. difficile* CD9301, **E**. phage stock ΦCD9501 infecting *C. difficile* CD9501, **F.** Negative control, *C. difficile* CD9301 lawn grown without the presence of the phage.

3.2.2 Isolation of *C. difficile* lysogens

Lysogens of isolated phages ΦCD3801, ΦCD8601 and ΦCD9301 were made using their *C. difficile* propagating strains for long-term storage as glycerol stocks at -80°C. Each phage lysate was spotted on a BHIS plate with a lawn of an appropriate *C. difficile* propagating strain and incubated overnight to form zones of clearing. The following day, a few colonies of potential lysogens growing within a zone of clearing were re-streaked 3 times to increase purity. Phage particles were induced from each lysogen using mitomycin C and tested in a plaque assay using their corresponding propagating *C. difficile* strain.

Induction titres were determined with phage ΦCD9301 showing the highest titre at 1.1×10^3 PFU x ml^{-1}. Both phages ΦCD3801 and ΦCD8601 induced to a similar titre ~ 10^2 PFU x ml^{-1} (Table 3.2). Furthermore, a plaque assay was performed with phages ΦCD3801, ΦCD8601 and ΦCD9301 against the lawn of their respective *C. difficile* lysogen to confirm the immunity of each lysogen against further phage infection. Lysogens were named after their *C. difficile* propagating strains CD3801L, CD8601L and CD9301L and stored as a glycerol stock at -80°C.

Table 3.2. Isolation of *C. difficile* lysogens.

C. difficile **lysogen**	**Titre of induced phage (PFU x ml^{-1})**	**Immunity to phage infection**	
		Phage	**Immunity**
CD3801L	3.6×10^2	ΦCD3801	Yes
CD8601L	1.9×10^2	ΦCD8601	Yes
CD9301L	1.1×10^3	ΦCD9301	Yes

3.2.3 Establishing Phage Morphology using Transmission Electron Microscopy (TEM)

TEM analysis was conducted to access the particle morphology of the isolated phages ΦCD3801, ΦCD8601 and ΦCD9301. All three bacteriophage have tails confirming that they belong to the *Caudovirales* order.

Phages ΦCD3801 and ΦCD8601 appear to have long tails of 300.8 ± 16.9 nm and 285.3 ± 5.5 nm, respectively, with respective icosahedral capsids of diameters, 54.2 ± 5.5 nm and 66.6 ± 5.7 nm (Figure 3.2, Figure 3.3). Tails of both phages ΦCD3801 and ΦCD8601 appear to be flexible and longer than 260 nm, which suggest that they belong to the *Siphoviridae* family. Interestingly, phage ΦCD8601 appears to have a neck connecting tail and the head as well as an enveloped tail, which is typical for phages belonging to the *Myoviridae* family.

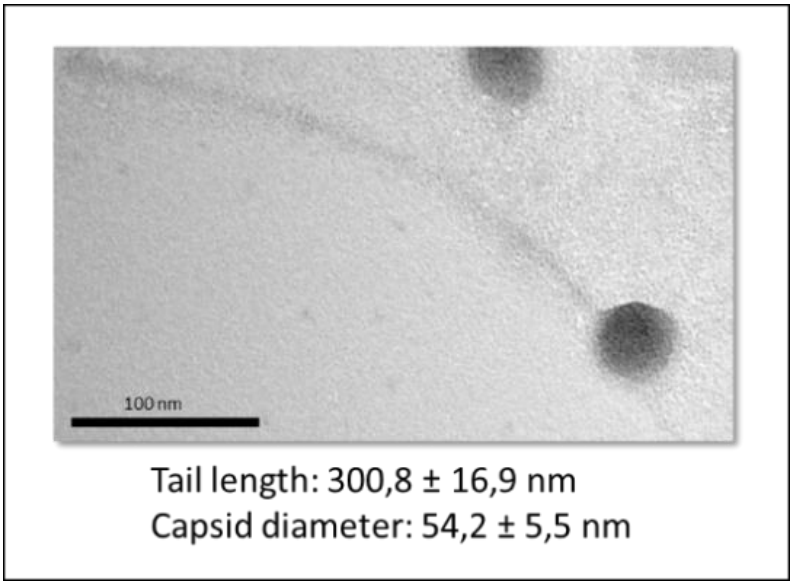

Figure 3.2: ΦCD3801 phage particle under TEM. The phage has been classified as *Siphoviridae* due to long non-enveloped tail that appears to be flexible. The capsid seems to be icosahedral and the neck is not clearly visible. Measurements are the average values of five phage particles and the standard deviations are shown.

According to Hargreaves and Clokie (2015), long tail myoviruses possess tails of 150 – 260nm in length with capsid diameters of 60 – 70 nm. Phage ΦCD9301 has an enveloped tail of 173 ± 13.3 nm and icosahedral capsid of 58 ± 59 nm in diameter

(Figure 3.4). Unlike the other two phages, the tail seems to be too short for *Siphoviridae* but can be grouped into long myoviruses. The presence of a sheath and visible neck connecting the head and tail further confirms this phage belongs to the *Myoviridae* family. TEM also reveals the possible presence of phage tail-like particles (PTLPs).

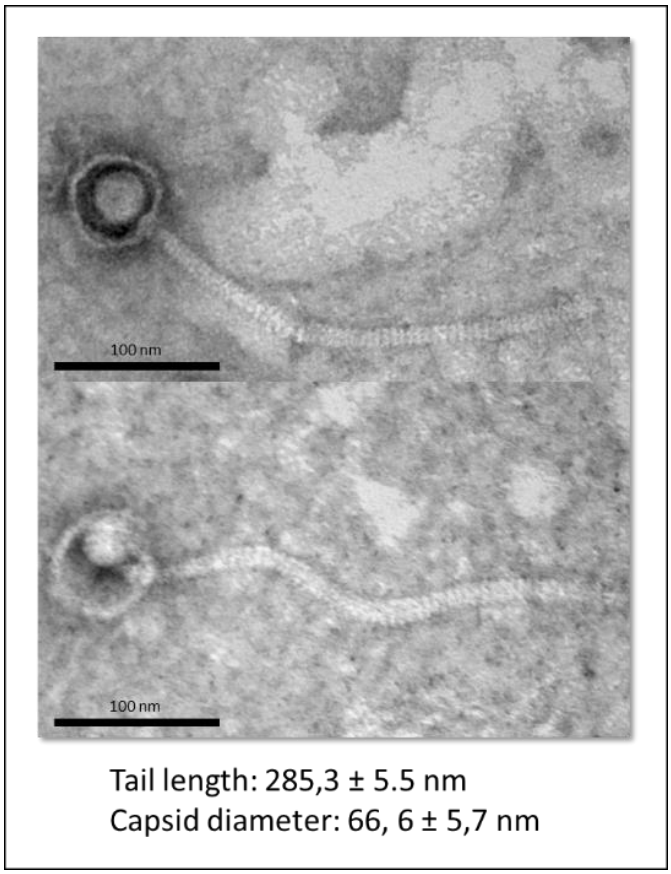

Figure 3.3: ΦCD8601 phage particle under TEM. The tail appears to be flexible and long which suggest that ΦCD8601 is a member of *Siphoviridae* family, however, the presence of sheath, typical for *Myoviridae* family makes a classification of this phage more difficult. Capsid looks icosahedral and the presence of neck is not clearly visible. Measurements are the average values of five phage particles and the standard deviations are shown.

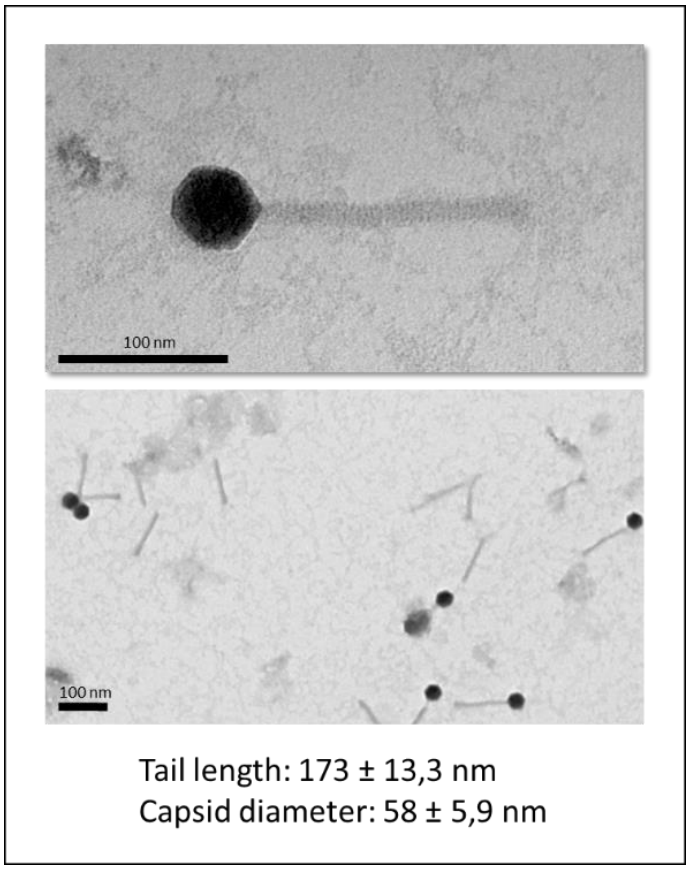

Figure 3.4: ΦCD9301 phage particle under TEM. The phage belongs to *Myoviridae* family. The capsid is icosahedral, connected to the tail with neck. The tail is short, non-flexible with visible sheath. Tail resembling particles called diffocins were also observed in the sample. Measurements are the average values of five phage particles and the standard deviations are shown.

3.2.4 Genotypic analysis

Phage genomes were sequenced using the Illumina MiSeq 500bp V2 SBS sequencing service and sequencing reads were assembled using the CLC Genomics Workbench 11. Open reading frames (ORFs) of a minimum of 100 amino acid residues were predicted using Artemis software (Sanger Institute, Cambridge, UK). ORFs were manually trimmed to the most likely start codon based on the presence of a potential ribosome binding site (RBS). Each ORF was then manually annotated using the online software BLASTp or UniProt.

PCR was used to determine the linear or circular form of phage genomes using primers binding to the end regions of the contig obtained from the sequencing data. For phage ΦCD3801, primers ZZ69 and ZZ70 were used to amplify the end of contig,

while primer sets ZZ77/ZZ78 and ZZ79/ZZ80 were used to determine the genome of phages ΦCD8601 and ΦCD9301, respectively. The PCR products were subsequently purified and sequenced using Sanger sequencing.

The genomes of isolated phages consist of circular, double stranded DNA (dsDNA) with low GC content (~ 30%), which is similar to other published *C. difficile* phages (Table 7.3 in Appendix). Although almost half of the manually annotated ORFs encode genes with unknown function, it appears that the genomes exhibit a modular organisation with modules of gene products predicted to be involved in bacterial cell lysis, lysogeny control, DNA replication or phage assembly. Analysis of the genomes also revealed no presence of the known toxin genes or CRISPR arrays and no genes coding for abortive infection proteins.

3.2.4.1 Phage ΦCD3801 and ΦCD8601

Phages ΦCD3801 and ΦCD8601 were isolated using environmental sample B and the *C. difficile* indicator strains CD3801 and CD8601, both of PCR ribotype RT078. The analysis of both phage genomes revealed that they were almost identical (query cover 96%, identity 99.9%). Due to the high genomic identity, only the genome of the ΦCD3801 was analysed to represent both phages.

The genome of phage ΦCD3801 is composed of 35 159 bp dsDNA with low GC content (28.4%) in keeping with that of the *C. difficile* genome. A single contig was assembled, using CLC Genomics Workbench 11, from 550 894 reads with an average read length of 227 bp. Artemis predicted 52 ORFs, out of which 35 were manually annotated and assigned a putative function (Table 7.4 in Appendix). PCR was used to determine the circular form of the genome using primers flanking the ends of the contig. Sequencing of PCR product confirmed no missing bases from the genome sequencing data.

The lysogeny module is often situated between the lysis module and the DNA replication and regulation cassette. Usually, this cassette consists of genes for transcription regulators, repressors, antirepressors and integrases (Hargreaves and Clokie, 2014). The lysogeny segment of phage ΦCD3801 contains the integrase gene (phiCD3801_GP33) and two XRE family transcriptional regulators that likely play a role in the maintenance of lysogeny (Hargreaves and Clokie, 2014). The presence of

these genes indicate that phage ΦCD3801 is temperate. Production of lysogens further confirmed the temperate nature of the phage. In case of ΦCD3801, lysogeny genes are not clearly organised in a module (Figure 3.5) but are scattered together with DNA replication genes such as genes encoding ssDNA binding protein (phiCD3801_GP30) and HNN motif endonuclease (phiCD3801_GP34). This was previously seen in the case of two other siphoviruses, phiCD38-2 and phiCD6356, and a myovirus, phiMMP04 (Sekulovic *et al*., 2011; Horgan *et al*., 2010; Meessen-Pinard *et al*., 2012).

In the phage assembly module, gene phiCD3801_GP1 encodes a terminase protein that is required for packing the phage DNA into the capsid. Another gene from the module, phiCD3801_GP2, encodes a portal protein that forms a passage for phage DNA during packaging and ejection. It most probably also participates in procapsid assembly (Govind *et al*., 2006).

Head packaging genes are followed by structural genes responsible for head morphogenesis and tail structure (Figure 3.5). The ORF encoding the putative head-tail joining protein (phiCD3801_GP7), shares homology with phage ΦCD6356. Tail structure genes encode for the major tail protein (phiCD3801_GP9), the tale protein (phiCD3801_GP12) and the tail-tape measure protein, which is predicted to determine the length of tail (Govind *et al*., 2006) and is encoded by the largest ORF in the genome (phiCD3801_GP11).

According to Govind *et al*. (2006) most phage require the presence of both holin and endolysin proteins for successful host lysis. The gene phiCD3801_GP6 codes for an endolysin that contains an N-acetylmuramoyl-L-alanine amidase domain that enables the digestion of peptidoglycan. Even though no holin encoding gene homologue was detected in the ΦCD3801 genome, plaques were formed when ΦCD3801 was used in a plaque assay with several *C. difficile* strains proving successful bacterial lysis.

The product of the gene phiCD3801_GP21 shares homology with a putative plasmid partition protein ParA. The presence of the gene encoding a ParA family protein and the absence of the Clp protease gene suggests that the phage belongs to the *Siphoviridae* family.

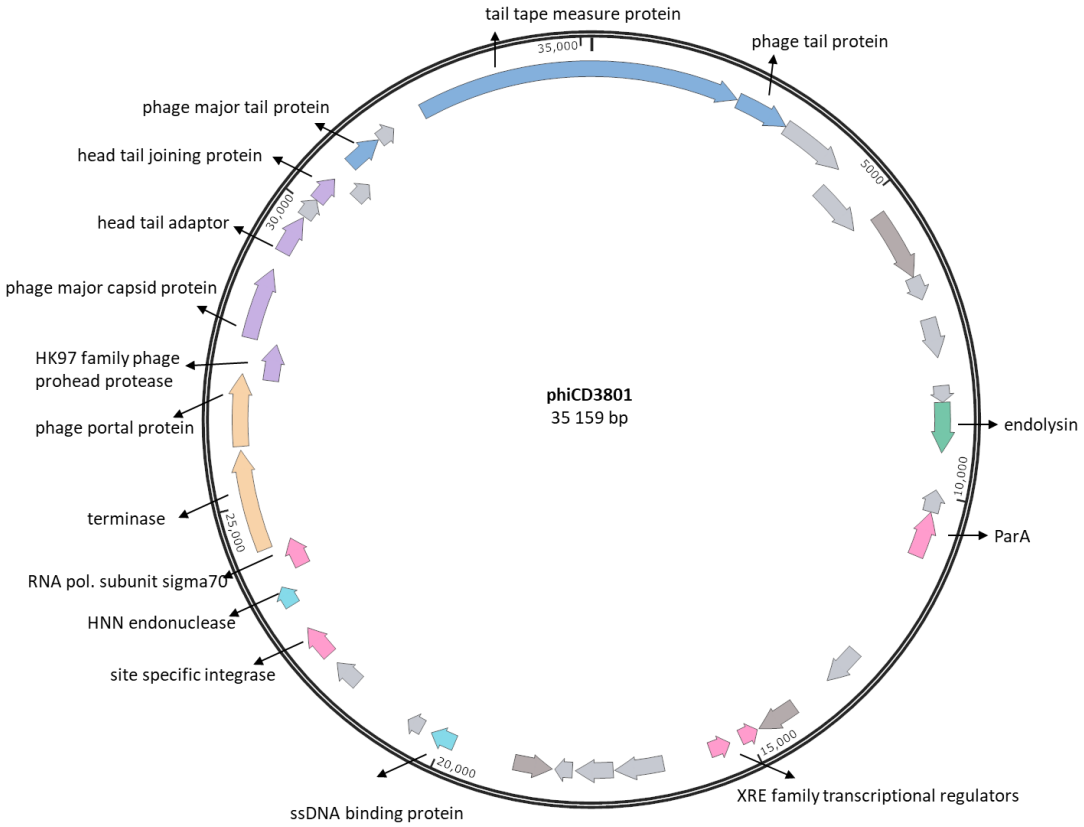

Figure 3.5: Schematic organisation of ΦCD3801 genome. The genome is 35 159 bp long, circular and contains 35 manually annotated ORFs that were predicted by Artemis (Sanger Institute, Cambridge, UK). ORFs are marked with arrows and colours indicate functional modules: DNA replication (turquoise), lysogeny control (pink), head packaging (light orange), head structural components (purple), tail structural components (blue) and cell lysis (green). Proteins with no assigned functionality are shown in grey.

3.2.4.2 Phage ΦCD9301

Phage ΦCD9301 was isolated using environmental sample B and *C. difficile* CD9301 indicator strain PCR ribotype RT078. The genome of the phage is 50 502 bp in length, dsDNA with a low GC content of 30%.

A single contig was assembled, using CLC Genomics Workbench 11, from 674 754 reads with an average read length of 234.58 bp. Primers annealing to the ends of the contig were used to amplify a PCR product that was sequenced to confirm the circular form of the phage genome. Artemis predicted 79 ORFs, out of which, 56 were manually annotated and assigned a putative product (Table 7.6 in Appendix).

The genome of ΦCD9301 is also organised in modules (Figure 3.6). In contrast to phage ΦCD3801, the DNA replication cassette is distinct from the lysogeny module

and contains more genes. The lysogeny section is located between phage assembly proteins and DNA replication proteins and contains several transcriptional regulators and the integrase (phiCD9301_GP34), as well as two penicillinase repressor proteins (phiCD9301_GP32, phiCD9301_GP33).

In common with other phages, genes encoding structural components of the head follow assembly cassette genes encoding terminase and phage portal proteins. The gene cluster positioned in the tail structure area encode component proteins, such as tail protein, tail-sheat, tail-tape measure protein, baseplate protein, LysM peptidoglycan binding protein and, similar to phage ΦCD119, the XkdN protein (Figure 3.6).

Like ΦCD3801, the lysis module of ΦCD9301 lacks a holin protein but contains the endolysin encoded by phiCD9301_GP27 and the cell wall hydrolase (phiCD9301_GP19) that is located within the tail structural genes.

The genome also contains a three gene cassette present in phages of the *phiCD119likevirus* genus. The three gene cassette contains genes encoding DnaD domain protein (phiCD9301_GP42), hypothetical protein (phiCD9301_GP44) and a ssDNA binding protein (phiCD9301_GP43). The presence of these genes in its genome suggests that phage ΦCD9301 belongs to the *phiCD119likevirus* genus of myoviruses. The presence of a tail sheath protein encoded by phiCD9301_GP13 coupled with the absence of a ParA homologue, further confirms this phage as *Myoviridae*. Interestingly, the genome also contains two genes encoding a penicillinase repressor proteins. These genes are located on the antisense strand in the lysogenic conversion module and homologues have been found in several *C. difficile* phages, such as ΦCD119, ΦCDHM1 and phiMMP02. The impact of these genes is not yet known, but their presence could support the phage mediated bacterial gene transfer hypothesis (Hargreaves and Clokie, 2014).

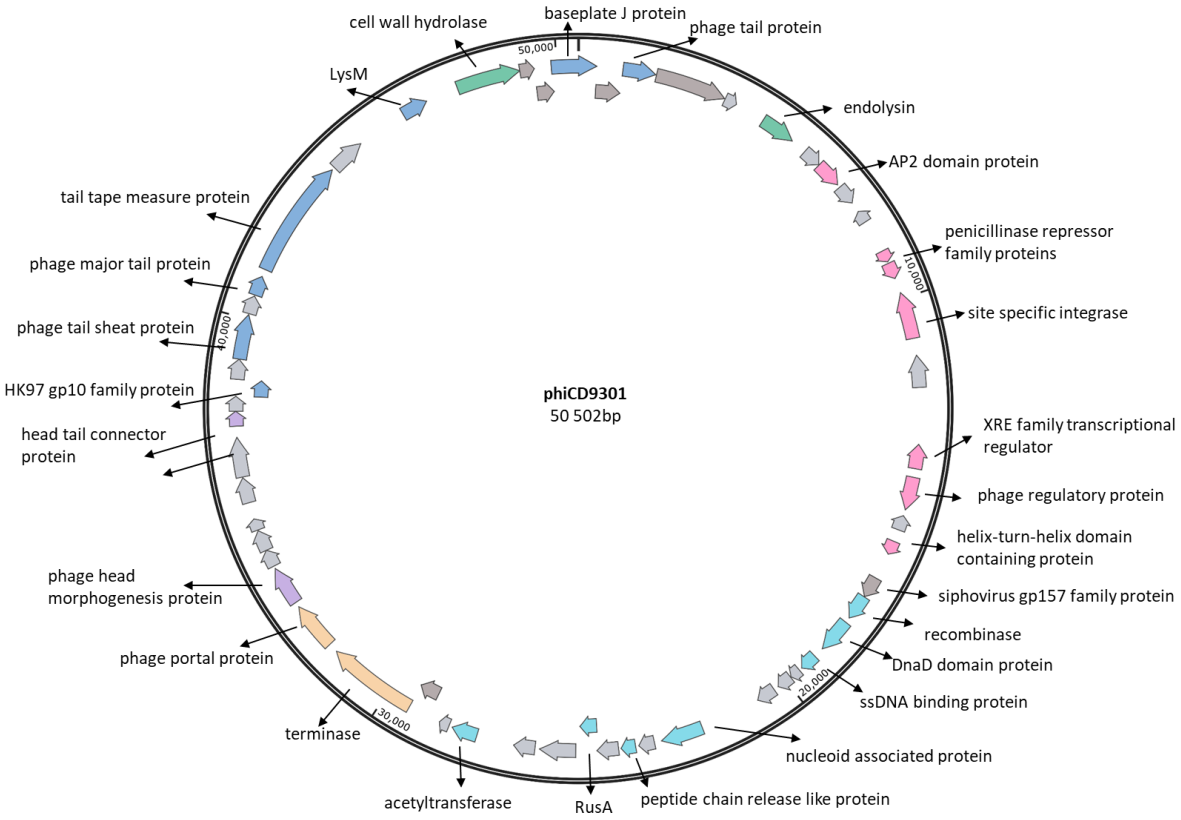

Figure 3.6: Schematic organisation of ΦCD9301 genome. The genome is circular, 50 502 bp in length and contains 56 manually annotated ORFs that were predicted by Artemis (Sanger Institute, Cambridge, UK). ORFs are marked with arrows and colours indicate functional modules: DNA replication (turquoise), lysogeny control (pink), head structural components (purple), DNA packaging (light orange), tail structural components (blue) and cell lysis (green). Hypothetical proteins with no assigned functionality are shown in grey. The presence of three gene cassette encoding ssDNA binding protein, DnaD and hypothetical protein suggests the phage belongs to *Myoviridae* family.

3.2.3 Characterisation of the integrase gene

Bacteriophage integrases are site-specific recombinases necessary for integration and excision of the phage genome during lysogeny and the transition from the lysogenic to the lytic cycle (Groth and Calos, 2014). The integration is completed solely by the activity of the integrase mediating recombination between the *attP* site in the phage genome and the *attB* site of the host chromosome, while the excision process involves additional proteins. Based on their catalytical domain, integrases belong to either the tyrosine or the serine recombinase family (Marintcheva, 2018).

Genome sequencing revealed that both ΦCD3801 and ΦCD9301 carry genes encoding integrases. NCBI protein BLAST was used to characterise the conserved domains of individual integrases and to determine their family. CLUSTAL omega alignment tool was used to compare the genes. It has been shown that the ΦCD3801 integrase is encoded on the sense strand while the ΦCD9301 integrase is located on the antisense strand and both proteins differ significantly in length and amino acid composition. The ΦCD9301 integrase is comprised of 372 amino acids in contrast to 182 residues contained within the integrase of phage ΦCD3801. Despite the difference, analysis showed that both integrases are part of the tyrosine recombinase family (Figure 3.7).

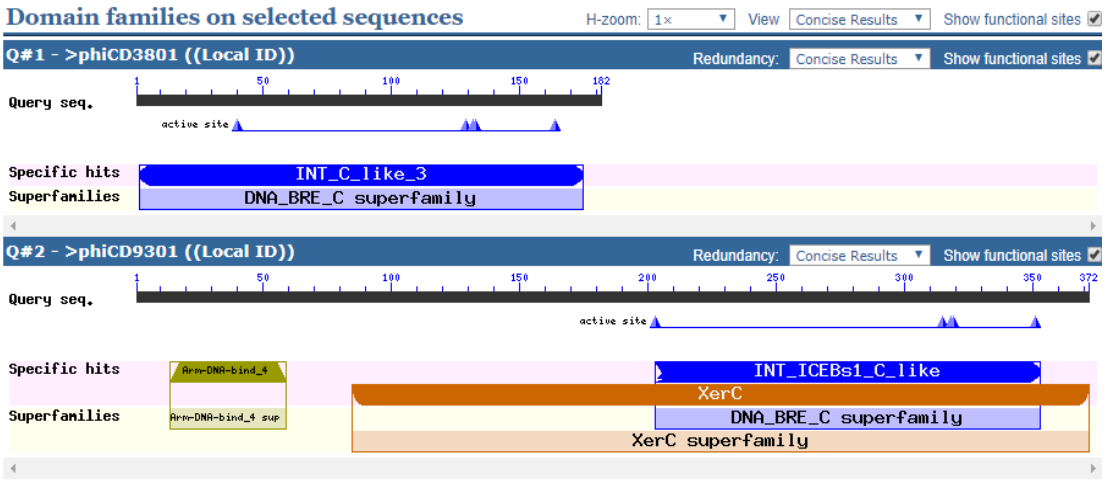

Figure 3.7: NCBI Blast prediction of conserved domains in integrases of isolated phages. Both integrases contain a conserved C terminal catalytic domain characteristic of tyrosine-based site-specific recombinases, which suggest they belong to tyrosine family. The ΦCD9301 integrase also contain Arm DNA-binding domain that has been found in several phage integrases, including the Int protein of phage lambda (Cho *et al.*, 2002).

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. PePPER promoter prediction tool and Genome Compiler software have been used to identify the presence of promoters and RBS sites of integrase genes. This analysis confirmed the presence of promoter and RBS sequences upstream of both the ΦCD3801 and ΦCD9301 integrase genes suggesting that they are transcribed and translated. This finding was further supported experimentally by the isolation of lysogens.

3.2.4 Host Range Analysis

To determine the host range, phage ΦCD3801 and ΦCD9301 (>10^8 PFU x ml^{-1} stock) were tested against several *C. difficile* strains using the standard double layer plaque assay (Chapter 2.5.4.2). In the case of infection, plaque assays were repeated using multiple dilutions of the phage to establish the level of *C. difficile* sensitivity to the phage. Various *C. difficile* isolates in the SBRC strain collection were used, including environmental and clinical strains belonging to different PCR ribotypes.

The host range analysis revealed that each phage formed different numbers of plaques with certain isolates compared to others. To establish the level of *C. difficile* susceptibility to phage infection, the efficiency of plating (EOP) was examined. For each sensitive strain, the EOP was calculated by dividing the phage titre obtained when the indicator strain tested for sensitivity was used in the plaque assay with the phage titre obtained using the *C. difficile* propagating strain. The efficiency of infection was determined as high if the EOP value was between 0.1 and 1, very high if the value was higher than 1 and low if the EOP was < 0.1. The EOP values (Figure 3.8) suggest that the level of infection differs for each phage among different isolates.

A full list of the tested isolates and their ribotype can be found in Figure 3.8. In total, phage host ranges were tested on 100 *C. difficile* strains across 37 PCR ribotypes. Of the 100 isolates tested, 28 isolates were sensitive to phage infection.

	C. difficile strain	RT	ΦCD9301 EOP	ΦCD3801 EOP
1	CD3001	001	X	X
2	CD3401	001	X	X
3	CD3901	001	X	X
4	CD6601_1	001	X	1.6
5	CD7501	001	X	X
6	CD0801	002	1.251	3.36
7	CD1101	002	X	X
8	CD1501	002	X	X
9	CD2001	002	X	X
10	CD2101	002	0.544	X
11	CD4101	002	X	X
12	CD5201	002	X	X
13	CD6001	002	0.0079	0.109
14	CD3301	003	X	X
15	CD0501	005	X	X
16	CD0901	005	X	X
17	CD4801	005	X	X
18	CD7001	005	X	X
19	CD6801	007	X	X
20	CD7101	011	X	X
21	CD6301_1	012	X	X
22	CD7601	012	X	X
23	CD1301	013	X	X
24	CD2701	013	X	X
25	CD2301	014	0.869	1.6
26	CD2601	014	X	1.81
27	CD4001	014	0.813	1
28	CD5601	014	0.93	1.63
29	CD6701	014	X	X
30	CD7701_1	014	X	X
31	CD0101	015	X	X
32	CD0301	015	X	X
33	CD1901	015	X	X
34	CD3601	015	X	X
35	CD4201	015	X	0.001
36	CD4501	015	X	X
37	CD4701	015	X	X
38	CD42_2	015	X	X
39	CD19_2	015	0.511	0.345
40	CD5901	015	X	X
41	CD6201	015	X	X
42	CD7301	015	X	X
43	CD1601	017	X	X
44	CD5001	018	X	X
45	CD1401	020	X	X
46	CD6301_2	020	X	X
47	CD4901	021	0.0981	0.709
48	CD0601	023	X	X
49	CD0701	023	X	X
50	CD1201	023	X	X
51	CD1201_2	023	X	X
52	CD2401	023	X	X
53	CD3701	023	X	X
54	CD5301	023	X	X
55	CD5401	023	0.00511	0.045
56	CD6601_2	026	X	X
57	CD2801	027	X	X
58	CD5101	027	X	X
59	CD4301	029	X	X
60	CD0401	031	X	X
61	CD2201	032	X	1.19
62	CD4002	035	X	X
63	CD2201_2	050	X	X
64	CD6501	050	X	X
65	CD5501	054	X	X
66	CD6901	054	X	X
67	CD5502	054	X	X
68	CD7401	054	X	X
69	CD2501	056	X	X
70	CD0801_2	062	X	X
71	CD1001	062	X	X
72	CD0201	070	X	X
73	CD3501	070	X	X
74	CD5701	076	X	X
75	CD5801	076	X	X
76	CD6101	076	X	X
77	CD9301	078	1	0.59
78	CD3801	078	0.819	1
79	CD8601	078	0.883	1.27
80	CD9001	078	0.595	1.054
81	CD9501	078	0.948	1.054
82	CD9701	078	0.623	0.581
83	CD1801	078	1.079	1.436
84	CD2917	078	1.1	1.772
85	CD2916	078	0.93	1.827
86	CD2918	078	1.125	1.472
87	CD2906	078	1.26	2.9
88	CD3205	078	1.08	1.54
89	CD3176	078	0.883	1.45
90	CD3669	078	X	X
91	CD3177	078	0.53	2.045
92	CD1801	078	0.79	1.7
93	CD3201	181	X	X
94	CD1701	220	X	X
95	CD7201	220	X	X
96	CD0401_2	224	X	X
97	CD6401	255	X	X
98	CD3101	294	X	X
99	CD2901	365	X	X
100	CD4601	479	X	X

Figure 3.8: Sensitivity of various *C. difficile* strains to isolated phages ΦCD3801 and ΦCD9301. EOP was calculated by dividing the phage titre obtained when the propagating strain tested for sensitivity was used in a plaque assay with the phage titre obtained using *C. difficile* indicator strain. The efficiency of infection was determined as high if EOP value was between 0.1 and 1 (orange colour), very high (red colour) if the value was higher than 1 and low if (yellow colour) EOP <0.1. Strains belonging to the same ribotype are represented with the same colour. Single ribotypes are represented in black.

The analysis showed that phage ΦCD3801 has a broader host range compared to ΦCD9301. Phage ΦCD3801 infected 26/100 isolates belonging to 8 different ribotypes including RT078, RT002, RT032, RT014, RT015, RT021, RT023, RT001. Similarly, ΦCD9301 infected 24/100 isolates across 6 different ribotypes RT078, RT002, RT014, RT021, RT015, RT023. Moreover, an overlap was observed as 22 *C. difficile* strains belonging to 6 ribotypes were susceptible to infection with both ΦCD3801 and ΦCD9301. In contrast, the phages could not infect strains belonging to 29 out of 37 tested PCR ribotypes.

Testing isolates from the same PCR ribotype as the propagating strain revealed that both phages have a broad host range activity infecting all tested PCR ribotype RT078 isolates apart from CD3669. Broad range activity against hypervirulent PCR ribotype RT078 demonstrates that phages ΦCD3801 and ΦCD9301 have exciting potential in the therapeutic field.

Overall, phages ΦCD3801 and ΦCD9301 infected multiple *C. difficile* strains across different PCR ribotypes and both phages were shown to infect clinically relevant strains, such as epidemic isolates of ribotype RT001 or RT078, which enhances their therapeutic potential.

3.2.5 Phylogenomic analysis

Virus Classification and Tree Building Online Resource (VICTOR) was used for phage phylogenomic analysis. Nucleotide sequences of isolated phages ΦCD3801 and ΦCD9301 were compared pairwise with 29 published *C. difficile* phage genomes (Table 7.3 in Appendix) using the Genome-BLAST Distance Phylogeny (GBDP) method (Meier-Kolthoff *et al*., 2013) under settings recommended for prokaryotic viruses (Meier-Kolthoff and Göker, 2017).

The phylogenomic trees were inferred using the three different formulas D0 (figure 3.9), D4 (figure 3.10) and D6 (figure3.11) that explored distinct aspects of genome evolution and yielded an average support of 55%, 38% and 59%, respectively. Formula D0 calculates the relatedness of genomes based on the length of all high scoring segment pairs (HSPs) divided by the total genome length. The D4 formula uses the sum of all identities found in HSPs divided by overall HSP length, whereas formula D6 uses the sum of all identities found in HSPs divided by the total genome length. Recommended, formula D4, is independent of genome length and is thus robust against the use of incomplete genomes.

The OPTSIL programme (Göker *et al*., 2009) was used to estimate taxon boundaries at the family, genus and species level with the recommended clustering thresholds. The OPTSIL clustering yielded 30 (D0), 29 (D4) and 27 (D6) species clusters, respectively. At the genus level, five (D0), thirteen (D4) and four (D6) clusters resulted, respectively. The number of respective clusters determined at the family level were one (D0), seven (D4) and one (D6).

All three phylogenetic trees show that the *C. difficile* phages visibly cluster into the family of siphoviruses and myoviruses. Phages belonging to the myoviruses are further clustered into the group of small myoviruses, medium myoviruses and long myoviruses. Phage ΦCD9301 clustered most closely to ΦCD119 and ΦCDHM119 using formulas D0 and D6. All three phages belong to the *phiCD119like* medium sized myoviruses. Interestingly, using formula D04, ΦCD9301 clustered closely to ΦJD032 belonging to *phiMMP04likevirus* myoviruses. On the other hand, phage ΦCD3801 appears to be evolutionary distant to phage ΦCD9301. Phylogenomic analysis revealed that ΦCD3801 is more closely clustered with other *Siphoviridae* phages ΦCD6356, ΦCD146, ΦCD111, ΦCD38_2, ΦCDSH1, which confirms the previous results obtained from TEM and genomic analysis.

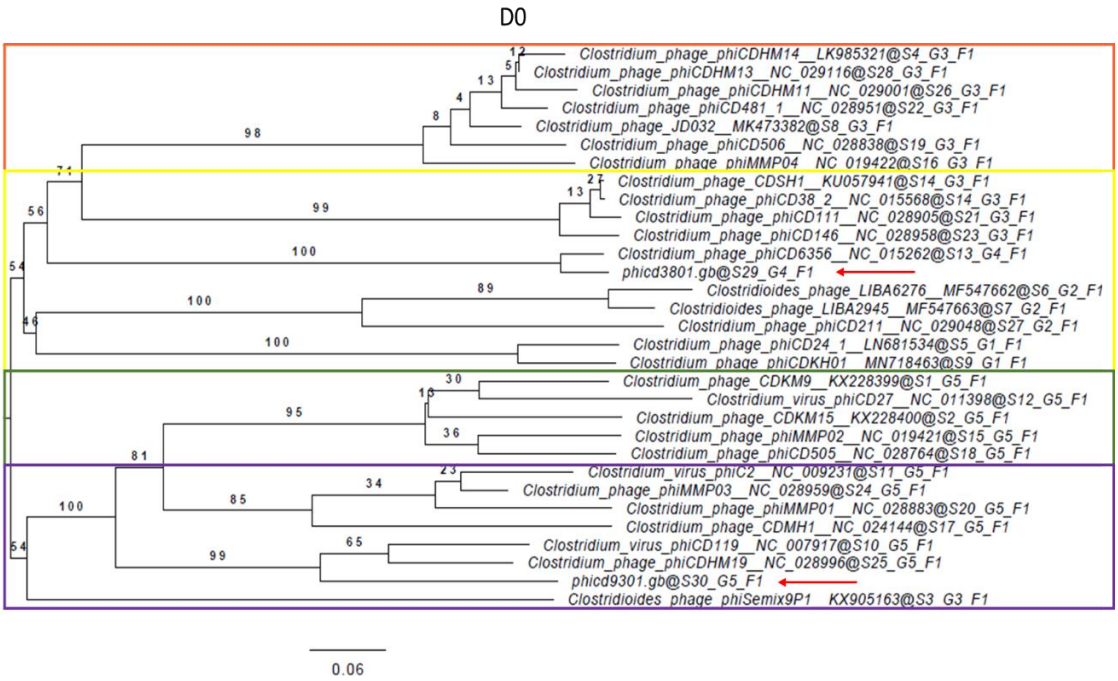

Figure 3.9: Phylogenomic tree based on whole genome comparison of *C. difficile* phage genomes using VICTOR D0 formula. The analysis compare the two novel bacteriophages ΦCD9301 and ΦCD3801 to the genomes of 29 published *C. difficile* phages. Red arrows point to the position of isolated phages. Yellow square marks group of phages belonging to the family of siphoviruses. Red, purple and green colouration marks the groups of phage belonging to small, medium and long myoviruses, respectively.

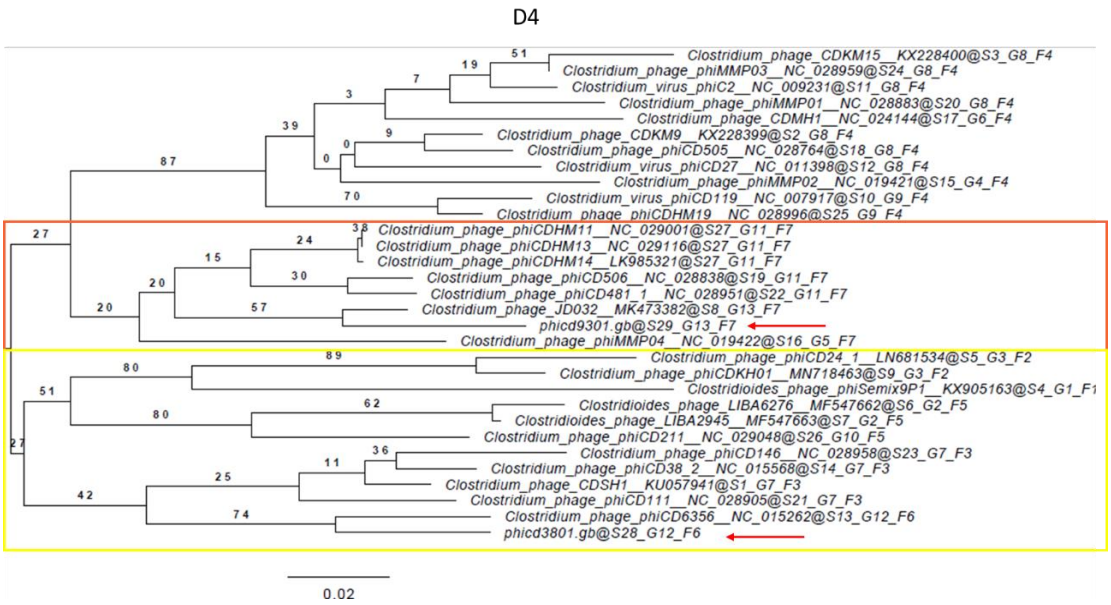

Figure 3.10: Phylogenomic tree based on whole genome comparison of *C. difficile* phage genomes using VICTOR D4 formula. The analysis compare the two novel phage ΦCD9301 and ΦCD3801 to the genomes of 29 published *C. difficile* phages. Red arrows point to the position of isolated phages. Yellow square mark the group of phage belonging to the family of siphoviruses. Red colour marks the groups of phage belonging to small myoviruses, respectively.

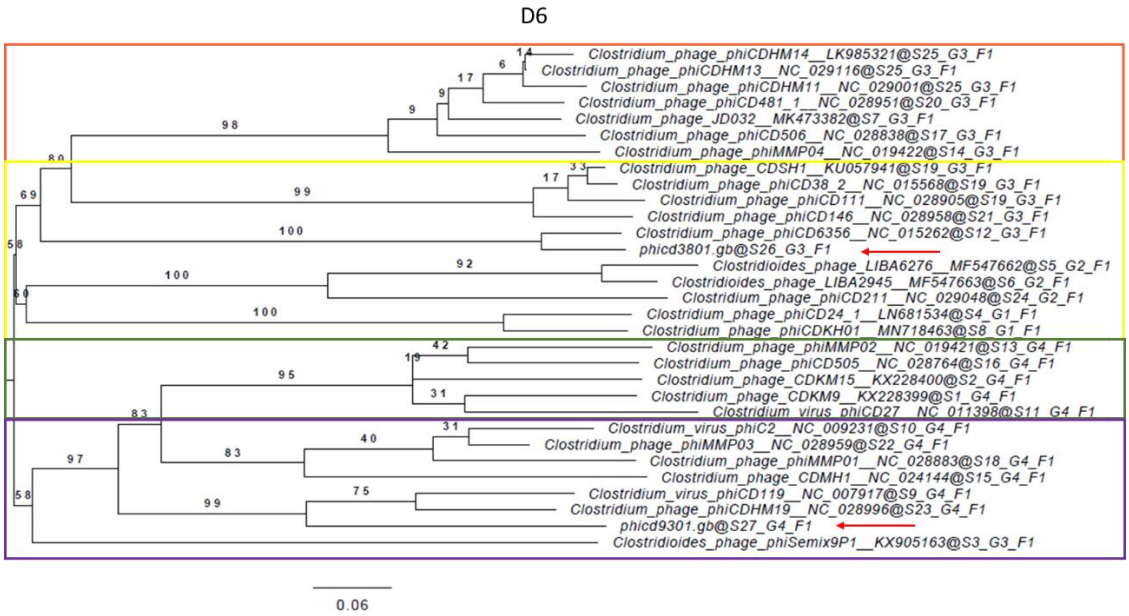

Figure 3.11: Phylogenomic tree based on whole genome comparison of *C. difficile* phage genomes using VICTOR D6 formula. The analysis compare the two novel phage ΦCD9301 and ΦCD3801 to the genomes of 29 published *C. difficile* phages. Red arrows point to the position of isolated phage. Yellow square marks the group of phage belonging to the family of siphoviruses. Red, purple and green colour mark the group of phage belonging to small, medium and long myoviruses, respectively.

3.3 Discussion and summary

Previous work at the SBRC revealed that *C. difficile* phages can be isolated from post-anaerobic digester sewage samples enriched with the germinant taurocholate. In total, five novel phages infecting *C. difficile* ribotype RT078 have been isolated in this study. Phages were named ΦCD3801, ΦCD8601, ΦCD9001, ΦCD9301 and ΦCD9501 after their respective *C. difficile* propagating strains. High phage titre is required for the manipulation of the genome and phage characterisation. Despite several attempts, it only proved possible to obtain the requisite high titres for three of the five phages. Accordingly, only phage ΦCD3801, ΦCD8601 and ΦCD9301 were further characterised using TEM, genome sequencing, host range analysis and phylogenomic analysis.

Despite the extensive screening of *C. difficile* phages from environmental samples, phage with strictly lytic activity have yet to be isolated. Stable lysogens of newly isolated phages were obtained using their propagating strains in order to enable long term phage storage at -80°C. These lysogens were shown to release phage particles after mitomycin C induction and to be immune to further phage infection. The isolation of stable lysogens provides an evidence that the isolated phages are temperate.

Based on their phage particle morphology, all three phages belong to the tailed phage of the *Caudovirales* order. Phage ΦCD3801 seems to belong to the *Siphoviridae* family. Characterisation of ΦCD8601 was more problematic. Phage ΦCD8601 has a long flexible tail suggesting it is the *Siphoviridae* family, however, the presence of a tail sheath is a characteristic feature of myoviruses (Hargreaves and Clokie, 2015). In contrast, based on TEM results, phage ΦCD9301 could be easily classified as a long tail myovirus with a tail length of 173 nm and a capsid diameter of 58 nm. Moreover, the presence of tail sheath and neck further supports the grouping of this phage into the *Myoviridae* family.

TEM also revealed the possible presence of phage tail-like particles (PTLPs). Many bacteria produce PTLPs, often referred to as “defective prophages.” PTLPs cause cell death by creating a channel in the wall of their target cell resulting in a rapid

dissipation of membrane proton potential. They can be divided into contractile (R-type) and non-contractile (F-type) forms. The R-type is structurally similar to tails of myoviruses while the morphology of the latter resembles the siphoviruses (Hegarty *et al.,* 2016). It has been shown that during SOS response, many *C. difficile* strains produce R-type PTLPs, called diffocins that inhibit the growth of other *C. difficile* isolates (Gebhart *et al*., 2015).

The analysis of phage genomes is an essential step in the determination of its suitability for therapeutic purposes. The genomes of isolated phages consist of circular dsDNA with low GC content, which is in common with other *C. difficile* phage published to date. The genomes were annotated manually as the use of computerised systems for automated annotation can be restricted by published annotations. Although almost half of manually annotated ORFs encode hypothetical proteins, genomes appear to be organised in modules containing gene products predicted to be involved in bacterial cell lysis, lysogeny control, DNA replication or phage assembly. Highly mosaic organisation of the genomes is in line with other *C. difficile* phages (Govind *et al*., 2006) and represents different combinations of the same modules that have been shuffled and reshuffled over billions of years (Keen, 2015; Al Shayeb et al., 2020).

The alignment analysis of ΦCD3801 and ΦCD8601 genomes show that they are almost identical. The presence of the gene encoding a ParA family protein and the absence of the Clp protease gene further proves that phages ΦCD3801 and ΦCD8601 belong to the *Siphoviridae* family. Unlike the other two phages, the genome of ΦCD9301 includes the three gene cassette containing genes encoding a DnaD domain protein, hypothetical protein and a ssDNA binding protein. According to Hargreaves and Clokie, 2015 the presence of these genes places phage ΦCD9301 into the *phicd119likevirus* genus of *Myoviridae*.

The presence of a putative integrase gene within the genome of each of the novel phages further confirms that each phage is temperate. To this date, every published *C. difficile* phage carries an integrase gene, however, it is not clear whether they are all active (Heuler *et al.*, 2021). The occurrence of putative RBS and promoter

sequences upstream of both ΦCD9301 and ΦCD3801 phage integrase genes suggests that the genes are expressed.

Phylogenomic analysis of the isolated phages was conducted by by comparing their full nucleotide sequences to the genomes of 29 published *C. difficile* phages. This method has been previously used by Rashid *et al*. (2016) to classify *C. difficile* phages, however, comparison between the endolysin gene, terminase gene and the gene encoding the major capsid protein have also been applied before (Hargreaves and Clokie, 2015).

Because of the high genome identity, phage ΦCD3801 was chosen to also represent ΦCD8601. Phylogenomic trees demonstrated that *C. difficile* phages visibly cluster into families of siphoviruses and myoviruses. VICTOR analysis revealed that phage ΦCD3801 belongs to the *Siphoviridae* family, clustering most closely with *C. difficile* phages ΦCD6356, ΦCD146, ΦCD111, ΦCD38_2 and ΦCDCDSH1. This result is in agreement with the TEM and the genome analysis.

Phage ΦCD9301 appears to be evolutionary distant to the phage ΦCD3801, as phylogenomic analysis revealed it belongs to the *Myoviridae* family. This conclusion was supported by the observed morphology of ΦCD9301. The presence of a gene encoding tail sheath protein and absence of ParA homologue in the phage genome further supports this conclusion. Moreover, ΦCD9301 clustered most closely to ΦCD119 and ΦCDHM119, which both belong to the *phiCD119like* myoviruses. Interestingly, phylogenomic analysis clustered ΦCD9301 together with medium sized myoviruses. According to Hargreaves and Clokie, 2015 medium size myoviruses have 110-130 nm long tails and capsid diameters between 60-70 nm. TEM results show that ΦCD9301 has a long tail of 173 ± 13.3 nm which suggest that it should be grouped together with long tail myoviruses. The length of tail was established as an average value of tail length of five phage particles. A standard deviation of 13.3 indicates that the data points are spread far from the mean. It is, therefore, possible that the length of the tail is shorter than 173 nm and ΦCD9301 indeed belongs to the medium size myoviruses.

Characterisation of phage by their host range is important when considering their application in therapeutic strategies. Although it is generally believed that *C. difficile*

specific phage infect one or few strains, Nale *et al*. (2016) isolated phage ΦCDHM3 and ΦCDHS1 that have proven to be respectively active against 39% and 38% of tested *C. difficile* strains belonging to 12 and 11 ribotypes. It is clear that host range seems to vary between isolated phages and the analysis is dependent on the panel of strains used in the testing.

In total, 100 *C. difficile* strains belonging to 37 PCR ribotypes were selected to test the host range of isolated phages ΦCD3801 and ΦCD9301. Out of the tested strains, 28% were sensitive to phage infection. Phage ΦCD3801 infected isolates belonging to 8 different ribotypes whereas ΦCD9301 infected strains belonging 6 different ribotypes. Testing isolates from the PCR ribotype RT078, which is the same as the propagating strains of isolated phage, revealed that both phage have high host range activity infecting all tested PCR RT078 isolates apart from CD3669 (93.75%). In comparison to the other published phages, Nale *et al*. (2016) isolated 7 phages with relatively broad range infecting *C. difficile* strains across 18 PCR ribotypes, however, only one of the these phage was able to infect *C. difficile* ribotype RT078. Moreover, only 38% of tested strains belonging to RT078 ribotype have shown to be sensitive to the infection with this phage. Similarly, when Rashid and colleagues tested host range of two novel myoviruses CDKM15 and CDKM9, both phage appeared to be unable to infect all 11 tested *C. difficile* RT078 strains (Rashid *et al*., 2016). Broad range activity against hypervirulent ribotype RT078 demonstrates that phages ΦCD3801 and ΦCD9301 have an exciting potential in the therapeutics field. This phenomenon could relate to the sample origin for phage isolation. In our study, we used sewage samples from the Nottinghampshire, whereas other studies utilised purely environmental samples for phage isolation such as soil (Rashid *et al*., 2016).

High prophage carriage has been shown in many of *C. difficile* ribotypes (Garneau *et al*., 2018) and it can explain the resistance observed to the two tested phages. Resistance to infection could also be explained by the CRISPR-Cas system. A single Type IB CRISPR-Cas system was identified in more than 200 *C. difficile* genomes (Andersen *et al*., 2016) with several CRISPR arrays located in prophages (Hargreaves *et al*., 2014; Boudry *et al*., 2015). Moreover, abortive infection (Abi) systems may also be encoded in the genomes and contribute to their resistance against phage.

Bacterial Abi systems represent cell death systems that are activated by phage infection to contain phage infection in bacterial populations (Dy *et al*., 2014). Abi genes are typically encoded by their bacterial hosts although they have been found on ΦC2, which indicates it could inhibit secondary phage infection (Hargreaves and Clokie, 2014).

In conclusion, newly isolated phages ΦCD3801 and ΦCD9301 belonging to the *Siphoviridae* and *Myoviridae* family infect multiple *C. difficile* strains across different PCR ribotypes. Both phage are temperate and genome analysis revealed no presence of toxin genes or CRISPR arrays. While temperate phages are not considered suitable for therapeutic purposes, they can be engineered to their lytic variants or be used as a vehicle to deliver lethal cargo. By doing so we can considerably expand the therapeutic options for the CDI treatment. Moreover, phages isolated in this study are active against clinically relevant strains of ribotype RT001 or RT078. According to Fairley *et al*. (2015) ribotype RT078 strains are more virulent than other common ribotypes in terms of toxin production. Moreover, RT078 is also found to be an important pathogen in pigs. Phages isolated in this study have shown to have a broad spectrum of activity against *C. difficile* RT078 isolates and are suitable for clinical use with high therapeutic potential for CDI treatment.

3.4 Chapter summary

- Three novel phages ΦCD3801, ΦCD8601 and ΦCD9301 infecting *C. difficile* have been isolated, confirmed and characterised.
- TEM analysis revealed that the phages ΦCD3801, ΦCD8601 belong to the *Siphoviridae* family while ΦCD9301 is grouped into the *Myoviridae* family.
- The host range profile of each phage was determined showing that they infect various *C. difficile* PCR ribotypes including clinically relevant strains.
- The genome sequences of novel phages have been established and annotated
- Phylogenomic analysis further confirmed phages ΦCD3801, ΦCD8601 as siphoviruses and ΦCD9301 as myoviruses.

Chapter four:

Generation of a Recombinant Phage Carrying SASP as a Lethal Cargo

4. Generation of a Recombinant Phage Carrying SASP as a Lethal Cargo

4.1 Introduction

Traditional phage therapy encounters multiple barriers that could prevent phages from being developed into functional therapeutics. Fortunately, rational design, modification and construction of recombinant phages have facilitated the expansion of their innate phenotypes.

Many studies have described the development of modified phages with altered host range (Ando *et al*., 2015; Chen *et al*., 2017; Gebhart *et al*., 2017; Heilpern and Waldor, 2003; Lin *et al*., 2012; Nguyen *et al*., 2012; Scholl *et al*., 2009; Yoichi *et al*., 2005; Yosef *et al*., 2017). For instance, using high-throughput mutagenesis Yehl *et al*. (2019) created a library of phages by mutating host range determining regions (HRDRs) in the tail fiber protein of the phage T3. This collection of phages exhibits expanded host range and ability to supress bacterial growth long term by preventing the development of resistance.

Recombinant phage also function as vehicles for antimicrobials that are either incorporated into the phage or attached to its surface. One example is the use of recombinant phage to deliver light-activated antimicrobial agent tin(IV) chlorin e6 (SnCe6), which showed promising results as an alternative to antibiotics for the treatment of common skin infections (Embleton *et al*., 2005). Phage can also be used to deliver immunogenic peptides or proteins to target cells or tissues. Filamentous recombinant phage f88 was engineered to display an immunizing epitope from influenza A virus. In this study, the use of live oral vaccine, comprised of an *E. coli* host expressing invasin in combination with engineered phage, led to the induction of an immune response against an antigen of interest displayed on the phage surface (Deng *et al*., 2017).

In this chapter, the generation of a recombinant phage carrying a lethal cargo in the form of small, acid-soluble protein (SASP) has been investigated. SASPs are produced by endospore-forming bacteria and play an important role in DNA protection within the spore. To date, most of our understanding of SASPs comes from *Bacillus* studies (Setlow and Kornberg, 1970; Setlow, 1988; Setlow, 1992; 2006) and little is known

about SASPs of other spore-forming bacteria. Among the genus *Clostridium*, only SASPs of *C. bifermentas* (Cabrera-Martinez *et al*., 1989), *C. perfringens* (Raju *et al*., 2006, 2007; Li & McClane, 2008; Li *et al*., 2009), *C. acetobutylicum* (Wetzel and Fisher, 2015), *C. botulinum* (Meaney *et al*., 2016) and *C. sporogenes* (Butler *et al*., 2017) have been characterised so far. Very recently, Nerber and Sorg (2021) characterised four SASPs of *C. difficile* R20291. These authors found that SASPs play an important role in UV light protection. Surprisingly, SSPA and SSPB have also been shown to regulate spore formation, which has not been described for any other SASP to date.

SASPs are only synthesized during the late sporulation phase in the spore core and degrade to amino acids rapidly during the early stage of germination. They can be divided into two groups: α/β type SASP and γ type SASP, although the γ type has only been seen in *Bacillus* species and is not able to bind DNA (Francesconi *et al.*, 1988; Vyas *et al.*, 2011).

SASPs consist of 59-75 amino acids resulting in small, 7-10 kDa proteins that have medium to high affinity for bacterial DNA. Binding of α/β type SASP is not dependent on the DNA sequence and results in the change in spore DNA structure from B conformation to an A-like conformation and alternate the supercoiling of DNA (Setlow *et al*., 1992). SASPs form dimers when complexed with DNA and cover approximately 6 bp of DNA sequence. Binding contributes to the protection of spores against UV, heat, desiccation and genotoxic chemicals (Setlow, 2006).

Although SASPs are only produced by spore-forming bacteria, these proteins should be able to bind to any bacterial DNA. In vegetative cells, SASP-DNA binding interferes with DNA replication and gene expression, which leads to irreversible shutdown of cellular activity and cell death. Moreover, SASPs binding to the DNA is not sequence specific and is therefore not compromised by DNA mutations, which suggests that resistance to SASPs resulting from random mutations is improbable. These characteristics make SASPs the promising basis of a new class of antibacterial agent (Setlow, 2006; Fairhead, 2009).

Accordingly, the British company Phico Therapeutics has patented the SASPjet system that use phage vectors for delivering SASPs to target bacteria. Phico's SASPjet phage vectors have deleted lytic genes to ensure they do not finish the normal phage

life cycle in order to increase the time for SASP production as well as minimise the inflammatory response that can occur during conventional lytic phage therapy. SASPjet, therefore, use phage vectors with the sole purpose of delivering SASPs, not as a therapeutic. The SASPjet system has been developed to tackle *Staphyloccocus* infections with proven antibacterial activity against several *S. aureus* isolates, including methicillin-resistant *S. aureus* (Fairhead, 2009).

The choice of an appropriate promotor system appears to be crucial in order to avoid killing the host strain used to harvest modified phage carrying a lethal cargo. To achieve this, tightly regulated inducible systems need to be implemented. Based on previous work in our lab, two inducible systems developed by I. Canadas (Canadas *et al*., 2019) appeared to be appropriate candidates for use against *C. difficile*: theophylline responsive riboswitch Rb3 and the RiboLac inducible system.

A riboswitch is a cis-acting regulatory element located in the 5'UTR region of mRNA that undergoes conformational switching in response to binding of a small ligand. They contain highly conserved ligand binding (aptamer) domains along with an expression platform that undergoes a conformational change leading to modulation of downstream gene expression (Mellin and Cossart, 2015).

Increasing interest in riboswitches led to the development of synthetic aptamer-metabolite combinations that bind non-native ligands without affecting the host's vital functions (Findeiß *et al*., 2017). An example includes theophylline-responsive riboswitches that have shown to be functional in several Gram-positive and Gram-negative bacteria (Topp *et al*., 2010). Theophylline-dependant riboswitches consist of an aptamer and a synthetic Shine-Dalgarno sequence (SD) that is sequestered via pairing with the stem of the riboswitch, resulting in halted translation. Binding of theophylline causes conformational changes leading to release of the SD and its availability to the ribosome, resulting in translation (Lynch *et al*., 2007) (Figure 4.1).

Several novel theophylline-responsive riboswitches have been constructed in the SBRC, out of which, Rb3 outperformed previously published riboswitches, exhibiting the largest dynamic range and a very low basal expression in *C. sporogenes* (Canadas *et al*., 2019).

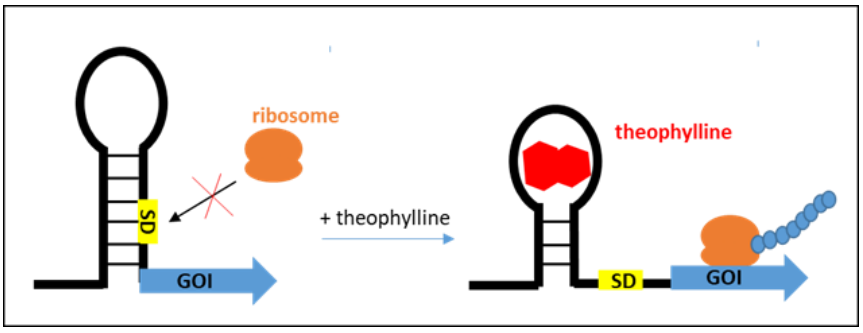

Figure 4.1: Schematic representation of theophylline inducible riboswitch. In the absence of theophylline, Shine Dalgarno sequence (yellow) is sequestered by loop formation of aptamer. Adding theophylline (red) leads to conformational changes of aptamer, making SD available to ribosome, allowing the gene translation (Canadas *et al.*, 2019).

The LAC system from *C. perfringens* strain 13 exhibits a large dynamic range and is functional in several tested *Clostridioides* species. This system contains the P_{bgaL} promotor that is driven by the presence of lactose as an inducer and a divergently orientated *bgaR* gene, whose expression is controlled by the P_{bgaR} promotor. BgaR is a transcriptional regulator that directly interacts with lactose and stimulates P_{bgaL} driven expression (Hartman et. al., 2011).

RiboLac represents a synthetic inducible system that controls the gene expression on both transcriptional and translation levels by combining the theophylline responsive riboswitch with the transcriptionally regulated LAC system from *C. perfringens* strain. In this dual control system, theophylline regulates the translation of BgaR transcriptional factor, which in turn together with lactose regulates the transcription of the gene of interest. In *C. sporogenes*, this system exhibited reduced background expression in relation to the native LAC system with no impact on maximum expression (Canadas, PhD thesis 2019).

Once inserted in the bacterial genome, temperate phages can be engineered using the same strategies developed for their host. Two strategies of *C. difficile* genome editing were investigated in this study. Both involved the integration of the desired cargo gene by standard allelic exchange and the isolation of the desired mutant through counter selection of the original WT allele. In the one case the WT was selected against using the counter selection marker *codA*, while in other counter selection was imposed through targeting of the WT allele with RiboCas9. In the former, the cytosine deaminase encoded by *codA* catalyses the conversion of

cytosine to uracil but also 5-fluorocytosine (5-FC) into the highly toxic 5-fluorouracil (5-FU). 5-FU is a pyrimidine analogue that can be misincorporated into DNA instead of thymine or uracil, leading to cell death. Mutant strains that have lost the plasmid are then easily selected on basal media supplemented with 5-FC (Ehsaan *et al*., 2016). CRISPR/Cas9 genome editing tools have been adapted from the natural immune system of bacteria (Barrangou *et al*., 2007). The technique uses a single DNA endonuclease Cas9 that can be simply programmed to target a desired DNA sequence by altering its guide RNA. The latter is provided as a single guide RNA (sgRNA) which has the dual function of a CRISPR RNA (crRNA) and trans-activated crRNA (tracrRNA). crRNA recognizes the invading DNA through an approximately 20-bp base pairing interaction region and tracrRNA is crucial for crRNA processing and the binding of Cas9 to the target DNA. sgRNA only recognises the target sequence in the presence of protospacer adjacent motif (PAM) next to it. Consequently, any 20 bp sequence with adjacent PAM can be programmed to be cut by the CRISPR/Cas9 system (Wang *et al*., 2016). Once cleaved, DNA sequence can be easily altered via homologous recombination by adding homologous template carrying desired mutation.

RiboCas9 is a CRISPR/Cas9 editing technology that uses a theophylline responsive riboswitch to control the expression of Cas9. Comparing to traditional CRISPR/Cas9, lower basal expression of highly toxic Cas9 facilitates the cloning step and transfer of CRISPR/Cas9 components into the target cells (Canadas *et al*., 2019).

4.1.1 Aims of this study

Temperate phages have often been avoided in traditional phage therapy approaches due to their ability to integrate into the host chromosome and capacity to mediate transfer of genes between bacteria, possibly increasing bacterial virulence. With an advance in sequencing and genetic engineering, design and production of recombinant phages increases the potential use of temperate phages in the treatment of various bacterial infections. The aim of this study was to explore the possibility of using SASPs as an antimicrobial against *C. difficile*. Since no strictly lytic phage have been isolated for *C. difficile*, we aimed to create recombinant temperate

phages carrying SASP under the control of inducible promoter system as a lethal cargo.

4.2 Results

4.2.1 Identification of suitable location in the phage genome for cargo knock-in

Temperate phage can feasibly be engineered when inserted into the bacterial chromosome using the same strategies developed for the bacterial host. It has been predicted that only 10% of additional genetic information can be packed into the viral particle without affecting its functionality (Casjens and Gilcrease, 2009). In the previous chapter, three *C. difficile* phages have been isolated and characterised. Phage ΦCD9301 has the largest genome comparing to the other two isolated phages and was therefore selected as a more suitable candidate for genetic manipulation since it could theoretically package larger DNA constructs into its genome without affecting phage function.

Owing to its small size (1666 bp), cargo containing an *ermB* antibiotic selection marker has been used to test several potential locations in the ΦCD9301 phage genome for gene insertion and assess the downstream effects of DNA insertion on the phage packaging. Two different strategies have been employed to identify a suitable location for successful gene integration into the ΦCD9301 genome. Initially, standard allelic exchange (AE) using counter selection markers was chosen as a tool for cargo insertion. Due to its lower efficiency and long processing time, this strategy was later replaced by the RiboCas9 system.

4.2.1.1 Using AE for *ermB* cargo knocking into the phage ΦCD9301

The AE plasmid used in this study are derivatives of the pMTL_SC7315 vector created by Stephen Cartman (Cartman *et al*., 2012) and contain the *catP* gene as a positive selection marker and *codA* as a counter selection marker. Plasmids also contain the pCB102 Gram-positive origin of replication, the Gram-negative ColE1 origin of replication and TraJ necessary for conjugal DNA transfer, the TCD0164 terminator

from *C. difficile* CD0164 gene and T_{fdx} terminator of the *C. sporogenes* ferredoxin gene, *fdx*.

Low frequency DNA transfer into *C. difficile* via conjugation requires the use of ”pseudo-suicide’’ vectors. Contrary to suicide vectors, these plasmids are able to replicate in the host cell but are segregationally unstable because they replicate at a lower rate in comparison to the host chromosome. Under the selection of the plasmid-borne antibiotic marker, integrant clones grow faster than cells carrying an antibiotic marker on a defective plasmid and are easily identified as larger colonies (Cartman *et al*., 2012).

In silico analysis of *C. difficile* M120, showed lack of a *codA* gene homologue and the presence of *CDIFFM120_03544c* gene encoding uracil phosphoribosyltransferase. To test if AE employing *codA* can be used for gene knock-in, sensitivity to 5-FU and 5-FC was examined by minimal inhibitory concentration testing. *C. difficile* CD9301 was re-streaked on minimal media with increasing concentrations of 5-FU and 5-FC. No growth was observed on medium with the lowest tested level of 5-FU (1 $ug.ml^{-1}$) whereas resistance to high concentrations of 5-FC (750 $ug.ml^{-1}$) was detected. Furthermore, pMTL_SC7315 plasmid carrying the *codA* gene and *catP* gene were transferred into *C. difficile*. The presence of *codA* in *C. difficile* restored the 5-FC sensitivity from 750 $ug.ml^{-1}$ to 10 $ug.ml^{-1}$ in contrast with the WT strain and the strain harbouring the control plasmid pMTL83151.

Five different positions 1A, 1B, 2A, 2B and 3 for cargo insertions into the genome of ΦCD9301 were selected based on the predicted presence of transcriptional terminators. These were predicted using the ARNold prediction software and cargo was designed to be positioned upstream of the predicted terminators. AE vectors for the *ermB* knock-in were generated in two steps. In the first step, 5 homology arms targeting 5 different locations for *ermB* knock-in were amplified from ΦCD9301 genomic DNA and cloned into the pMTL_SC7315 backbone using *Xma*I and *Xho*I restriction sites as outlined in section 2.7.1. Plasmids targeting locations 1A, 1B, 2A, 2B and 3 were sequenced and named pMTL_SC1A, pMTL_SC1B, pMTL_SC2A, pMTL_SC2B and pMTL_SC3 respectively.

In the second cloning step, the T1T2 terminator was spliced together with the *ermB* cargo and its promoter in a SOE PCR reaction and cloned into the individual pMTL_SC plasmids via the *Kas*I and *Avr*II restriction sites as outlined in section 2.7.1 (Figure 4.2). The resulting plasmids were sequenced and named pMTL_SC1A:ermB, pMTL_SC1B:ermB, pMTL_SC2A:ermB, pMTL_SC2B:ermB and pMTL_SC3:ermB.

AE vectors were separately transferred into the CD9301L lysogen of *C. difficile*. Transconjugants were selected with thiamphenicol and re-streaked twice on the same media to allow for plasmid integration into the genome of the ΦCD9301 prophage. Single crossover cells were confirmed with two colony PCR reactions. One PCR was performed using the forward primer binding within a cargo sequence and the reverse primer annealing to the chromosomal sequence (Chapter 2, Table 2.4). A second PCR was performed with the forward primer binding to the chromosome upstream from the cargo sequence and the reverse primer binding within an *ermB* gene. Mutants were subsequently plated onto minimal media supplemented with 5-FC as cells that maintain the *codA* gene (either on autonomously replicating plasmid, or on integrated plasmid) are unable to grow. Individual colonies were then patch plated onto the same media with or without thiamphenicol to confirm plasmid loss. Thiamphenicol sensitive mutants were confirmed by colony PCR using the forward primers (ZZ144, ZZ146, ZZ148 or ZZ150) and the reverse primers (ZZ145, ZZ147, ZZ149 or ZZ151), binding upstream and downstream of regions with shared homology to the knock-in cassette. Individual mutants were further confirmed by Sanger sequencing and named CD9301L:1A_ermB, CD9301L:1B_ermB, CD9301L:2A_ermB, CD9301L:2B_ermB, CD9301L:3_ermB.

Mutants were successfully generated for each tested location in the phage genome (Figure 4.2), however, the efficiency was low in each case. On average, less than 20% of the screened colonies were mutant after plasmid loss. The rest are assumed to be WT cells or mutants that reverted back to the WT.

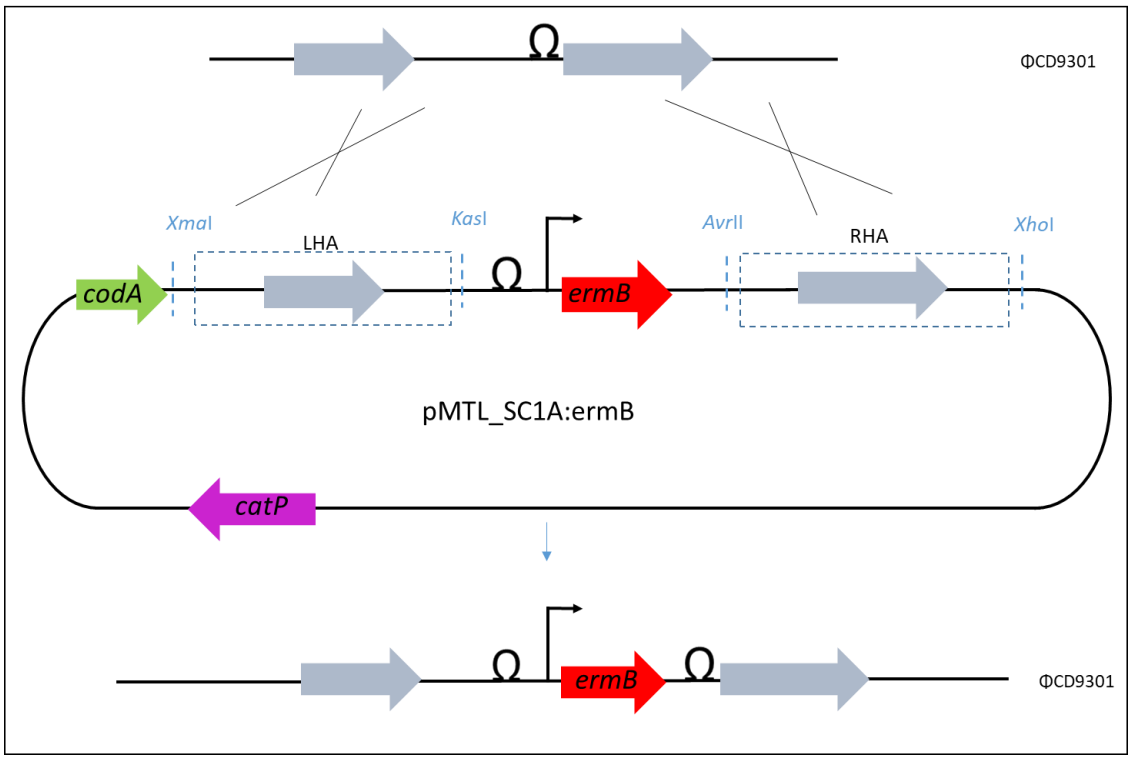

Figure 4.2: Schematic representation of *ermB* cargo knock-in into the CD9301L lysogen using AE. In the first step, transconjugants are re-streaked twice on media containing thiamphenicol to allow for the first cross-over and the integration of plasmid into the genome to occur. Subsequently, cells are plated on the minimal media supplemented with 5-FC to select for the double-crossover cells that have lost the plasmid.

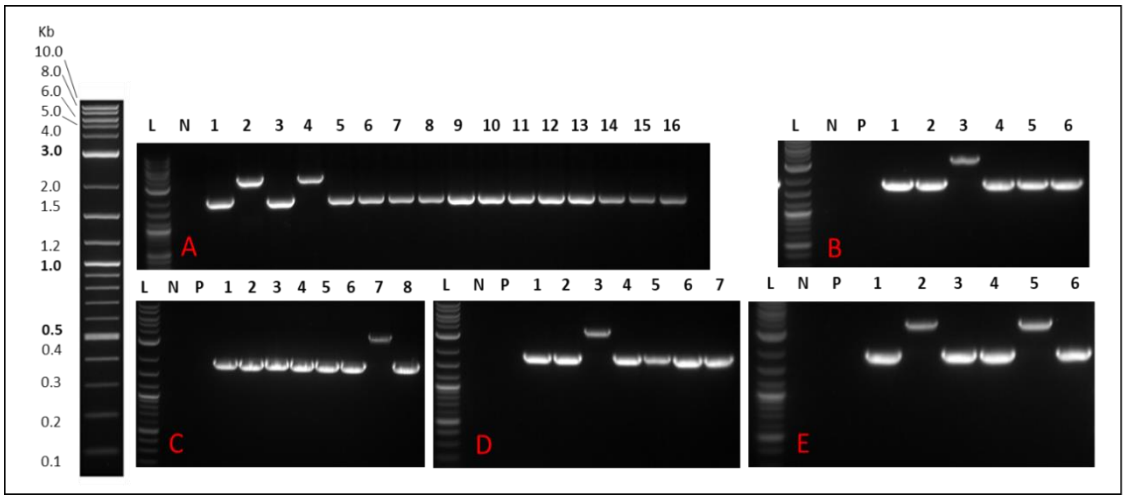

Figure 4.3: Colony PCR screening for *ermB* knock-in into the genome of ΦCD9301 prophage. Colony PCR was performed using Green Taq DNA polymerase with standard thermocycling conditions. **A,** colonies following AE using pMTL_SC1A:ermB plasmid**.** L: 2 log ladder, N: distilled H_2O, 1: ΦCD9301 genomic DNA, 2 -16 individual colonies. **B**, colonies following AE using pMTL_SC1B:ermB plasmid**.** L: 2 log ladder, N: distilled H_2O, P: plasmid pMTL_SC1B:ermB, 1: ΦCD9301 genomic DNA, 2-6 individual colonies. **C**, colonies following AE using pMTL_SC2A:ermB plasmid**.** L: 2 log ladder, N: distilled H_2O, P: plasmid pMTL_SC2A:ermB, 1: ΦCD9301 genomic DNA, 2-8 individual colonies. **D**, colonies following AE using pMTL_SC2B:ermB plasmid**.** L: 2 log ladder, N: distilled H_2O, P: plasmid pMTL_SC:ermB, 1: ΦCD9301 genomic DNA, 2-7 individual colonies. **E**, colonies following AE using pMTL_SC3:ermB plasmid**.** L: 2 log ladder, N: distilled H_2O, P: plasmid pMTL_SC3:ermB, 1: ΦCD9301 genomic DNA, 2-6 individual colonies. Mutant strains are expected to give a band size ~3.56 kbp. Non-mutant strains are expected to give a band size ~1.89 kbp.

Following the AE procedure, mitomycin C was used to induce phage from CD9301L lysogens with the integrated *ermB* cargo. After 24 hours incubation, a plaque assay was performed with induced phages and their replicating strain *C. difficile* CD9301 to check if additional genetic information could be packed into the functional viral particle of modified phage. Unfortunately, no plaque was observed using any of the induced mutant lysogens, which suggest that selected locations (Figure 4.5) in ΦCD9301 genome are not suitable for gene knock-in.

Positions 2A, 2B and 3 are located within the tail structural module and the failure of CD9301L:2A_ermB, CD9301L:2B_ermB and CD9301L:3_ermB to release phage particles after the induction may be explained by cargo interference with structural operon. Positions 1A and 1B are located at the end of replication module and the presence of cargo in this module possibly negatively influenced the phage replication and led to inability of the phage to replicate the DNA and the phage death.

4.2.1.2 Using RiboCas9 for *ermB* cargo knock-in into the phage ΦCD9301

Due to the low mutation efficiency observed and the time-consuming process of counterselection using *codA*, RiboCas9 was chosen as an approach to test several other locations in ΦCD9301 genome for cargo insertion (Canadas *et al*., 2019). Plasmids generated in this study were derived from the RiboCas9 plasmid created by I. Canadas (Canadas *et al*., 2019) containing the components for RiboCas9 tool cloned between the *Not*I and *Asc*I restriction sites within the pMTL83151 modular vector. The CRISPR/Cas9 machinery consists of the *cas9* gene and a single guide RNA (sgRNA), each under the control of individual promoters. The expression of the *cas9* gene used in the vector is regulated on a translational level by theophylline responsive riboswitch located in the P_{fdx} promotor (Figure 4.4). The single guide RNA consists of a 20 nt long protospacer sequence directing the Cas9 endonuclease to the target region within the genome and also contains an 83 nt handle sequence.

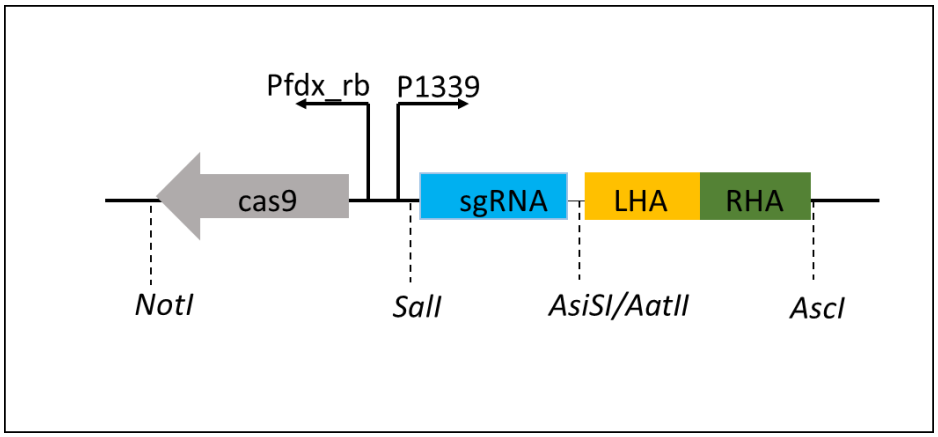

Figure 4.4: Schematic representation of the RiboCas9 components in cloning vector. The RiboCas9 genome editing machinery consists of *cas9* gene under the control of P_{fdx} promotor containing theophylline riboswitch, sgRNA under the control of P1335 arabinose promotor from *C. acetobutylicum* and homology cassette. Homology cassette contains left homology arm (LHA) and right homology arm (RHA) amplified from DNA fragments flanking the chromosomal target region. sgRNA consists of universal handle region and variable seed sequence specific to the target sequence.

In the previous section (4.2.1.1), locations for cargo knock-in were selected based on the presence of phage terminators. Since this strategy led to a limited amount of options, four new locations 4A, 4B, 5A and 5B were selected to be positioned independently of the native terminators. Since locations 1A, 1B, 2A, 2B and 3 were not proven successful for cargo insertion, new locations were also designed outside

of the replication or structural module (Figure 4.5). Benchling was used to select gRNA based on their on and off target score and the homology arms were designed based on the gRNA position. To avoid plasmid self-targeting by Cas9, part of the gRNA was on the LHA, the other part on the RHA, separated by cargo.

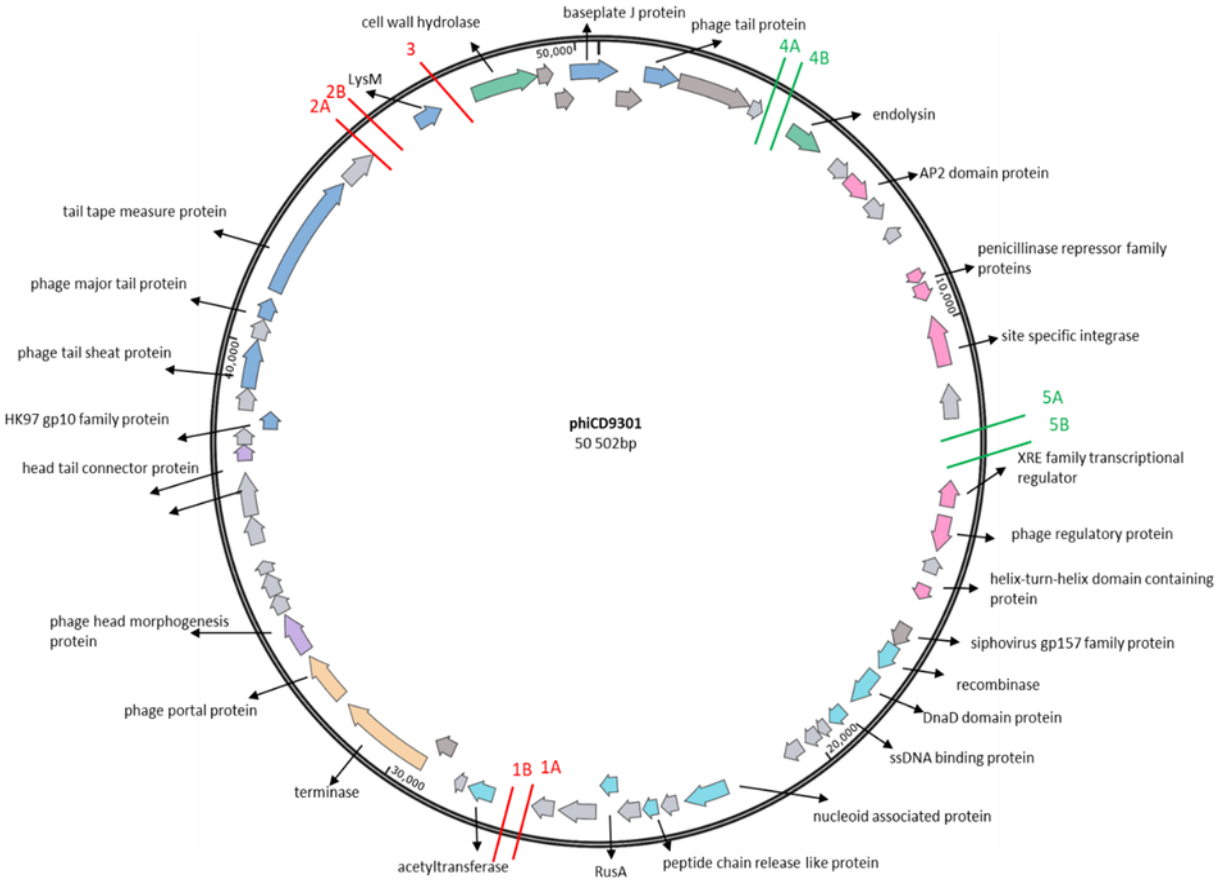

Figure 4.5: Schematic representation of eight locations tested for cargo knock-in into the ΦCD9301 genome. ORFs are marked with arrows and colours indicate functional modules: DNA replication (turquoise), lysogeny control (pink), head structural components (purple), DNA packaging (light orange), tail structural components (blue) and cell lysis (green). Hypothetical proteins with no assigned functionality are shown in grey. Red lines represent positions that were selected for *codA*-based AE cargo insertions. Positions 1A, 1B, 2A, 2B and 3 were designed upstream of the native phage terminators predicted by ARNold web prediction tool. Green lines symbolise locations selected for *ermB* insertion by AE using RiboCas9 as the counter selection system. Due to limited amount of predicted terminators, only location 5B was designed upstream of the native phage terminator.

Similar to the *codA* vectors, RiboCas9 plasmids for *ermB* knock-in were generated in two steps. In the first step, homology arms targeting different locations for cargo insertion were amplified from ΦCD9301 genomic DNA and cloned into the pRECas1-IIE-Cdif backbone together with gRNA using the restriction sites *Asc*I, *Aat*II and *Sal*I as outlined in section 2.7.1. Plasmids were sequenced and named pMTL_RC4A, pMTL_RC4B, pMTL_RC5A and pMTL_RC5B.

In the second step, the T1T2 terminator was spliced together with the *ermB* cargo in an SOE PCR reaction and cloned into the individual pMTL_RC plasmids via the *Kas*I and *Avr*II restriction sites (Figure 4.6). The resulting plasmids were sequenced and named pMTL_RC4A:ermB, pMTL_RC4B:ermB, pMTL_RC5A:ermB, and pMTL_RC5B:ermB.

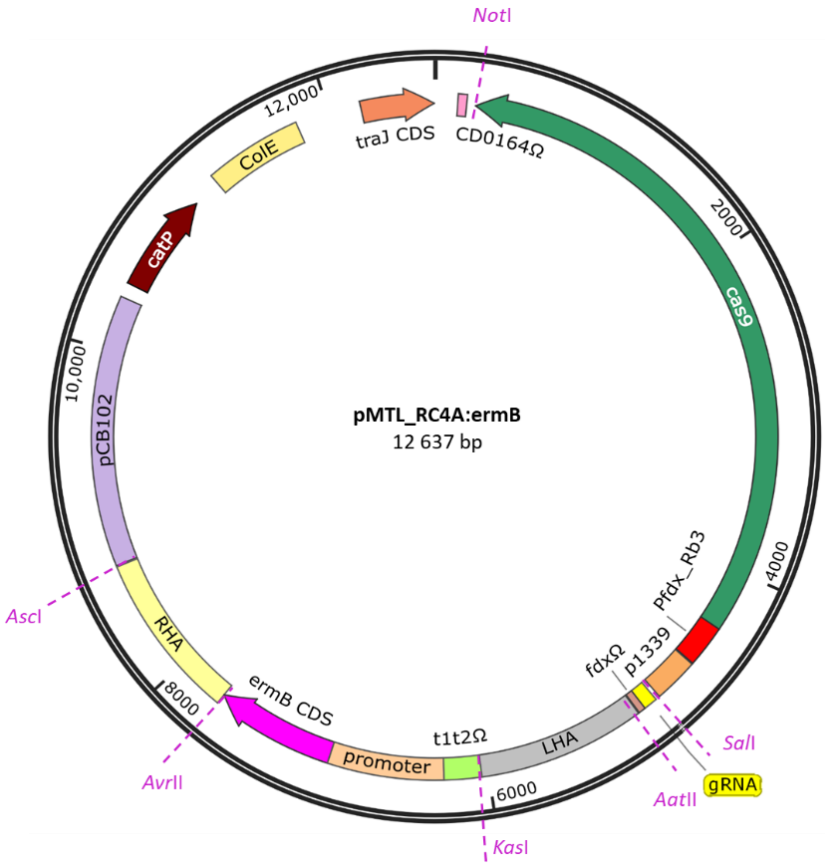

Figure 4.6: Schematic diagram of the RiboCas9 plasmid for *ermB* knock-in. The plasmid was constructed in two steps. In the first step, homology arms and seed gRNA sequence were cloned into the pRECas1-IIE-Cdif backbone using *Sal*I, *Aat*II and *Asc*I restriction enzymes. In the second step *ermB* cargo was cloned into the vector via *Kas*I and *Avr*II sites.

RiboCas9 plasmids were conjugated into the CD9301L lysogen. Transconjugants were selected with thiamphenicol and induced with theophylline as outlined in section 2.6.14. No colonies were observed after the pMTL_RC5A:ermB plasmid transfer, possibly due to the failed conjugation or off-target gRNA binding. Mutant cells were confirmed by colony PCR using the forward primers ZZ280, ZZ282, ZZ294 and the reverse primers ZZ281, ZZ283 and ZZ295 for knock-in mutants generated with plasmids pMTL_RC4A:ermB, pMTL_RC4B:ermB, pMTL_RC5B:ermB respectively.

The RiboCas9 method of generating mutants proved to be faster and more efficient than the use of *codA*, with an average of >80% mutants after colony PCR screening (Figure 4.6). Mutants were sub-cultured to lose the plasmid and thiamphenicol sensitive colonies were further confirmed by Sanger sequencing. Mutant lysogenic strains were named CD9301L:4A_ermB, CD9301L:4B_ermB and CD9301L:5B_ermB and stored as glycerol stocks at - 80°C.

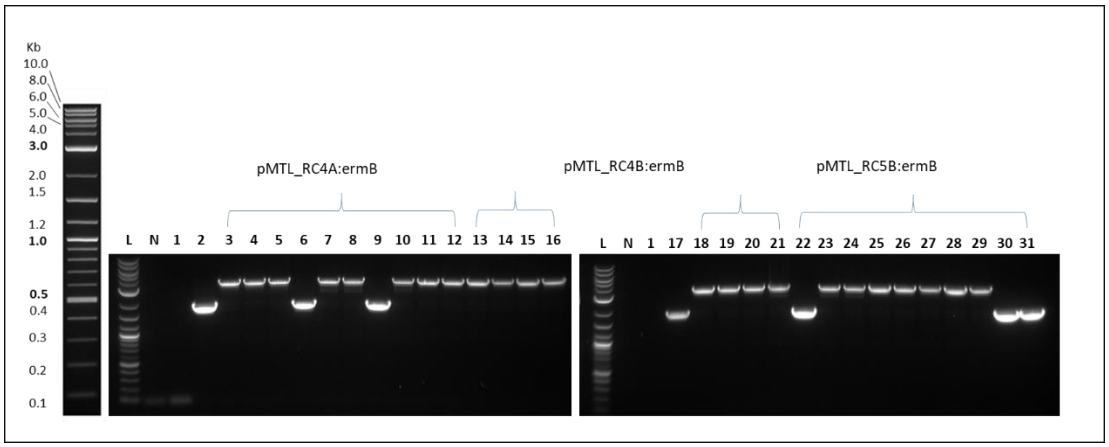

Figure 4.7: Colony PCR screening for ermB knock-in into the genome of ΦCD9301 prophage. Colony PCR was performed using Green Taq DNA polymerase with standard thermocycling conditions. L: 2 log ladder, N: distilled H_2O, 1: plasmid pMTL_RC4A:ermB, 2 and 17: ΦCD9301 genomic DNA, 3-12: mutants obtained following pMTL_RC4A:ermB conjugation 13-21: mutants obtained following pMTL_RC4B:ermB conjugation, 22-31: mutants obtained following pMTL_RC5B:ermB conjugation Mutant strains are expected to give a band size ~3.6 kbp. Non-mutant strains are expected to give a band size ~1.94 kbp.

Mutant lysogens were induced to produce phage particles, using mitomycin C. After 24 hours incubation, a plaque assay was performed with induced phages and their replicating strain *C. difficile* CD9301 to see if additional genetic information could be packed into the phage particle. Plaques were present in the case of all induced

lysogenic strains, however the induction of mutant lysogens produced phage at lower titres comparing to the CD9301L control (Table 4.1). This is probably due to the packaging of additional genetic sequence into the phage virion. The presence of plaques demonstrated that additional genetic information of the size of 1.67 kb can be successfully packed together with the phage genome into the functional phage. These results suggest that positions 4A, 4B and 5B in the phage genome (Figure 4.5) are suitable for inserting cargo sequences into the phage ΦCD9301.

Table 4.1. Titre of the phages induced from mutant lysogens after mitomycin C treatment.

Lysogen	**Titre of induced phage (PFU x ml^{-1})**	**Immunity to ΦCD9301 infection**
CD9301L	8.5 x 10^2	Yes
CD9301L:4A_ermB	1.9 x 10^2	Yes
CD9301L:4B_ermB	2.2 x 10^2	Yes
CD9301L:5B_ermB	3.7 x 10^2	Yes

Following the induction, plaques were picked and purified three times. Similar to plaques formed by WT phage ΦCD9301, plaques formed by mutant phages were small and hazy with approximately 1 mm diameter in size. To further confirm the successful cargo integration into the phage particle, the titre of modified phages was increased by harvesting the top agar layer from the plate followed by performing a plaque assay. The process was repeated until the high phage titre of ≥10^8 PFU x ml^{-1} was obtained. Once, the phage titre reached sufficient values, DNA of modified phages were isolated and the presence of the *ermB* cargo in the phage genome positions 4A, 4B and 5B was once again confirmed by colony PCR using the forward primers ZZ280, ZZ282, ZZ294 and the reverse primers ZZ281, ZZ283 and ZZ295 as well as by Sanger sequencing. Phages were named ΦCD9301:4A_ermB, ΦCD9301:4B_ermB and ΦCD9301:5B_ermB and stored as filtered sterilised supernatant in BHIS broth at 4°C created using 0.2% BHIs top agar plaque assay.

4.2.2 Establishing the effect of SASPs on *E.coli* and *C. difficile*

Initial studies focused on determining the effect of SASP proteins on *E. coli* cell growth. DNA fragments encompassing *saspA* or *saspB* genes from *B. subtilis*, provided by M. Whittle (University of Nottingham, SBRC), had previously been synthesized by DNA 2.0. Genes were individually cloned into the pMTL83151 backbone via the *Xma*I/*Sal*I restriction sites together with the promoter and terminator. Due to the toxic effect of SASPs on vegetative cells, an inducible promotor system was introduced into the vector upstream of *sasp* genes. Based on previous research in our group performed by I. Canadas, synthetic inducible promotor systems such as theophylline induced riboswitch Rb3 or dual lactose-theophylline inducible system RiboLac (figure 4.8) were chosen because of the tight regulation of the gene of interest proven in *C. sporogenes* (Canadas *et al.,* 2019). The plasmid also contains the *catP* selection marker, the pCB102 Gram-positive origin of replication, the Gram-negative ColE1 origin of replication and TraJ necessary for conjugal DNA transfer, the TCD0164 terminator from *C. difficile* CD0164 gene and T_{fdx} terminator of the *C. sporogenes* ferredoxin gene. Plasmid constructs were confirmed by colony PCR, sequenced and named pMTL83151_Rb3_saspA, pMTL83151_Rb3_saspB, pMTL83151_RiboLac_saspA and pMTL83151_RiboLac_saspB.

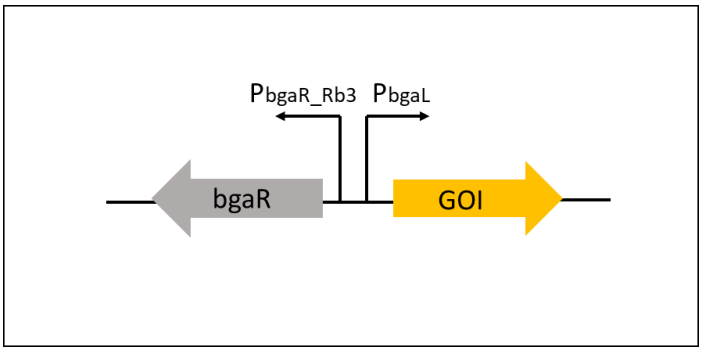

Figure 4.8. Schematic representation of the RiboLac system. The system contains a *bgaR* gene encoding the BgaR transcription activator under the control of the P_{bgaR_Rb3} promoter with theophylline riboswitch. Divergently orientated is gene of interest controlled by lactose inducible promotor P_{gaL}. Theophylline regulates the translation of the BgaR transcriptional factor, which in turn together with lactose regulates the transcription of the gene of interest.

To establish the impact of SASPs on *E. coli,* cells harbouring plasmid-borne *saspA*, *saspB* and appropriate controls were grown in the presence of increasing concentrations of inducers and their optical density (OD_{600nm}) was monitored over time by means of a manual growth curve. The Rb3 inducible system was induced with either 2 mM or 5 mM theophylline. The RiboLac inducible system was induced with both lactose (10 mM) and theophylline (0.5 mM or 2 mM).

As shown in Figure 4.9A, a decrease in OD_{600nm} was observed for both WT and empty plasmid control strains, when comparing the impact of increasing concentrations of the inducer compound on *E. coli* growth (Figure 4.9). The OD_{600nm} of WT cells declined from 2 to 1.3 after 8 hours of incubation when assessed alongside the Rb3 constructs, however, the difference in the growth was far less apparent at the 24h time point (Figure 4.9A). These findings demonstrate an inhibitory effect of theophylline on *E. coli*. Compared to the Rb3 experiments, the toxic effect of inducers was not as noticeable in the RiboLac control groups (Figure 4.9B) exposed to various levels of inducer compounds. This may be explained by the lower doses of theophylline applied compared with the Rb3 experiments.

As presented in Figure 4.9A, both uninduced bacteria harbouring *sasp* genes displayed decreased growth relative to the control, which may indicate that the system is leaky. In which case, even a very low amount of SASP protein can have a lethal effect on *E.coli*. Nonetheless, the toxic effect was more significant with increasing amounts of inducer as cells harbouring *saspA* or *saspB* reduced the OD_{600nm} value by more than 20% and 50% respectively, compared to the induced empty vector control groups.

The inhibitory impact of the SASPA protein was not as evident when the RiboLac system was used, with cells showing similar OD_{600nm} values as the empty vector control. In contrast, *E. coli* cells expressing RiboLac_*saspB* were severely inhibited for growth, with endpoint OD_{600nm} values dropping from 2 as observed in induced control to 0.9. These data sets suggest that SASPB has a greater lethal effect on *E. coli* cells than SASPA.

When considering the two inducible promoter systems, in no case did Rb3 or RiboLac systems achieve maximum control of expression, as uninduced cells carrying *saspB* showing approximately 40-50% reduction in growth after 24 hours in both systems. This finding further demonstrates the potency of the SASPB protein.

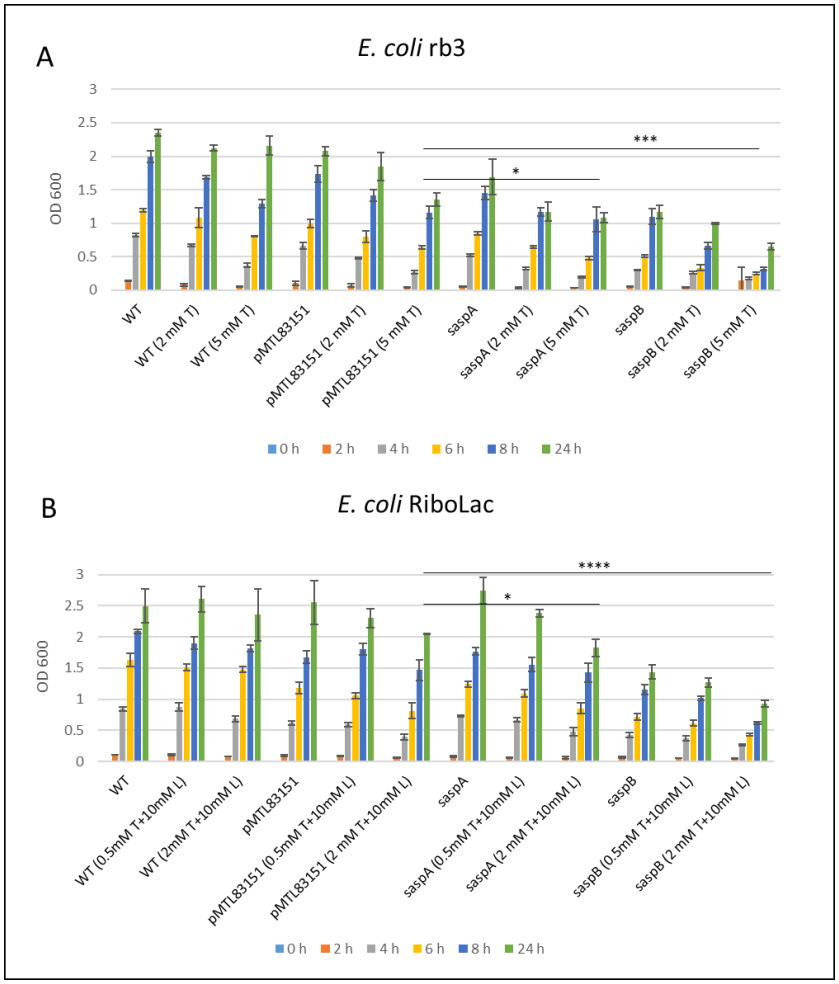

Figure 4.9. Establishing the effect of SASP on the growth of *E. coli*. Cells harboured plasmids with *sasp* genes under the control of **A**, Rb3 inducible system or **B**, RiboLac system and were grown in the presence of increasing dose of inducers theophylline (T) or both theophylline and lactose (L) respectively. In both cases WT cells were cultivated in LB medium and cells carrying plasmids were cultivated in LB medium supplemented with chloramphenicol. Error bars represent standard deviations calculated from 3 biological replicates. Asterisks indicate statistically significant p values for *p<0.05, **p<0.01 ***p≤0.0002, ****p≤0.0001 calculated with unpaired two-tailed Students's *t*-test. A significant decrease in cell growth was observed in Rb3 groups when comparing the cells with empty vector with cells carrying *saspA* (p=0.0174) as well as *saspB* (p=0.0004). Similarly, in RiboLac groups the decline of OD_{600nm} was observed when comparing control to cells carrying *saspA* (p=0.0494) as well as *saspB* (p<0.0001).

To examine the lethal effect of SASP on *C. difficile*, the designated plasmids containing *sasp* genes were transferred into *C. difficile* from their respective *E. coli* donor via conjugation. Transconjugant cells were then exposed to inducer and the OD_{600nm} was measured.

Due to the inhibitory effect of theophylline on *E. coli*, optimal concentration of the inducer on *C. difficile* was established first. *C. difficile* was grown in media containing various concentrations of theophylline ranging from 1 mM to 10 mM and the OD_{600nm} was monitored. In contrast to *E. coli*, theophylline does not appear to have negative impact on *C. difficile* growth when used in concentrations up to 5 mM as no significant difference in OD_{600nm} was observed among control samples in either Rb3 or RiboLac systems. Contrary to this, concentrations higher than 5 mM resulted in a significant reduction of *C. difficile* growth.

No difference in growth between uninduced cells carrying *saspA* and empty vector was detected in either of the two inducible systems (Figure 4.10) which may indicate that both systems exhibit tighter gene expression control in *C. difficile* than in *E. coli*. In contrast, the tighter gene expression was not observed with cells harbouring *saspB*.

Induced cells containing the *saspA* gene showed a decrease in OD_{600nm} by approximately 30% after 8 hours in both systems, however this effect was observed to last only in Rb3 group (Figure 4.10A). Remarkably, a very strong inhibitory effect was observed among *C. difficile* cells carrying *saspB*. In both systems, cells carrying the plasmid with *saspB* displayed delayed growth with OD_{600nm} reaching values more than six times lower than then empty plasmid control after 8 hours of incubation. The difference was not as substantial at the endpoint with reduction in growth by approximately 50% comparing to the control (Figure 4.10). This outcome may be explained by the very lethal effect of SASPB protein on *C. difficile* cells. The difference in the growth curve in *C. difficile saspB* groups may be explained by the delay in growth. It is possible that the toxic effect of SASPB on cells led to selection, favouring cells that lost the plasmid or mutated the *saspB* gene.

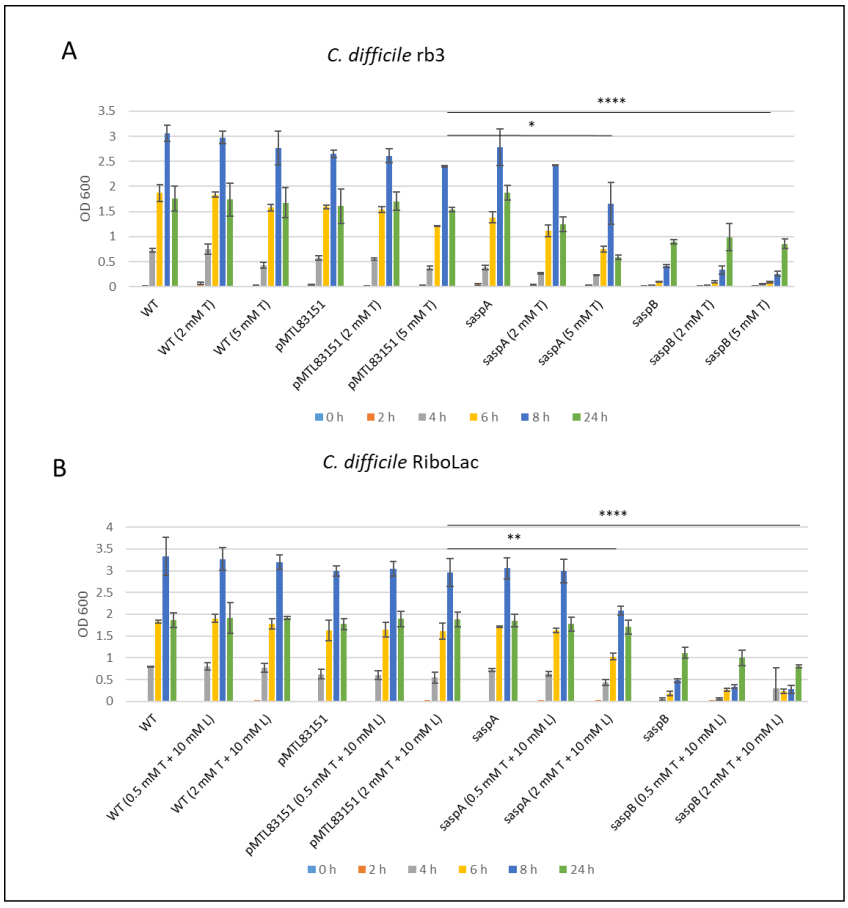

Figure 4.10. Establishing the effect of SASP on the growth of *C. difficile*. Cells were harbouring plasmids with *sasp* genes under the control of **A**, Rb3 inducible system or **B**, RiboLac system and were grown in the presence of increasing dose of inducers theophylline (T) or both theophylline and lactose (L), respectively. In both cases WT cells were cultivated in BHIS medium and cells carrying plasmids were cultivated in BHIS medium supplemented with thiamphenicol. Error bars represent standard deviations calculated from 3 biological replicates. Asterisks indicate statistically significant p values for *p<0.05, **p<0.01 ***p≤0.0002, ****p≤0.0001 calculated with unpaired two-tailed Students's *t*-test. Significant decrease in cell growth was observed in Rb3 groups when comparing the cells with empty vector with cells carrying *saspA* (p=0.0395) as well as *saspB* (p<0.0001). Similarly, in RiboLac groups the decline of OD_{600nm} was observed when comparing control to cells carrying *saspA* (p=0.0037) as well as *saspB* (p<0.0001).

4.2.3 Construction of recombinant phages carrying SASP as a phage lethal cargo

In the previous section, the toxic effect of SASP was proven in *C. difficile*, therefore, the use of ΦCD9301 for the delivery of SASP cargo was proposed. In section 4.2.1.2, locations 4A, 4B and 5B in the genome of ΦCD9301 were identified for a cargo knock-

in as the additional genetic information of the size of 1.6 kB, successfully inserted into the phage genome and packed into the functional viral particle.

Because of its smaller size (549 bp) and similar effectivity, the theophylline inducible system was chosen to be inserted together with *sasp* genes into the genome of ΦCD9301 lysogen location 4A. Because of the use of the Rb3 system in the RiboCas9 plasmids, the *codA*-based AE method was chosen to deliver the cargo.

AE vectors based on *codA* were cloned in two steps. Firstly, homology arms targeting location 4A were amplified from ΦCD9301 genomic DNA and cloned into the pMTL_SC7315 backbone using the *Xma*I and *Xho*I restriction sites as outlined in section 2.7.1. Plasmids were sequenced and named pMTL_SC4A.

In the next step, the inducible promoter Rb3 was spliced together with *sasp* genes and cloned into AE plasmids via the *Kas*I/*Avr*II restriction sites. Plasmids were confirmed by colony PCR, sequenced and called pMTL_SC4A:Rb3saspA and pMTL_SC4A:Rb3saspB.

The *codA* based AE vectors were conjugated into the ΦCD9301 lysogen of *C. difficile* and transconjugants were selected with thiamphenicol and re-streaked twice on the same media. Single crossover cells were confirmed with two colony PCR reactions. One PCR was performed using the forward primer binding within the cargo sequence and the reverse primer annealing to the chromosomal sequence. The second PCR was performed with the forward primer binding the genome upstream from the cargo sequence and the reverse primer binding within the *sasp* cargo. Integrated mutants that lost plasmid were selected using minimal media supplemented with 5-FC and confirmed by colony PCR (Figure 4.11) using primer set ZZ280/ZZ281 binding in the genome outside of the knock-in cassette. Sanger sequencing further confirmed positive colonies.

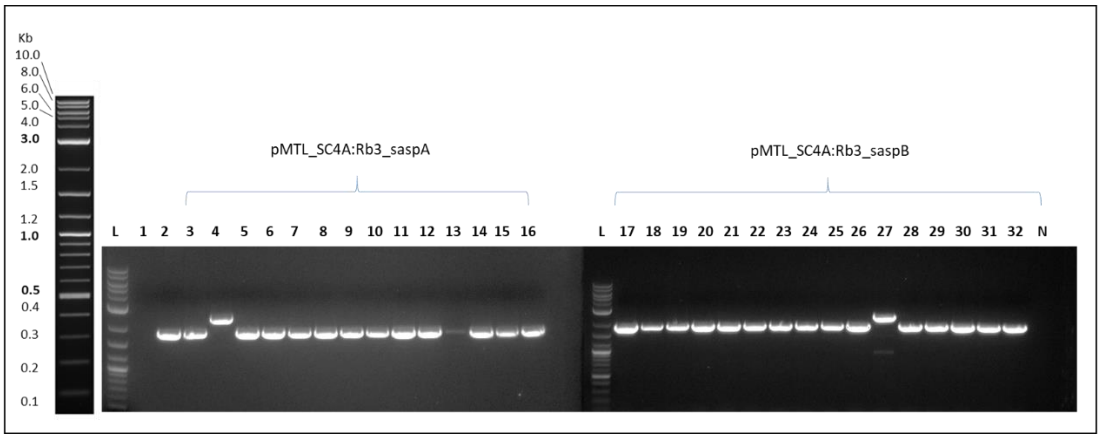

Figure 4.11: Colony PCR screening for *Rb3sasp* knock-in into the genome of ΦCD9301 prophage. Colony PCR was performed using Green Taq DNA polymerase with standard thermocycling conditions. L: 2 log ladder, N: distilled H_2O, 1: plasmid pMTL_SC4A:Rb3_saspA, 2: ΦCD9301 genomic DNA, 3-16: colonies obtained following pMTL_SC4A:Rb3_saspA conjugation, 17-32: colonies obtained following pMTL_SC4A:Rb3_saspB conjugation. Mutant strains are expected to give a band size of ~2.5. Non-mutant strains are expected to give a band size of ~1.94 kbp.

Despite the low efficiency, one *Rb3_saspA* mutant and one *Rb3_saspB* mutant was obtained following AE. The mutant lysogenic strains were named CD9301L:4A_Rb3_saspA and CD9301L:4A_Rb3_saspB. Surprisingly, despite the smaller cargo size compared to the *ermB* cargo, no phages were induced from mutant lysogens following mitomycin C induction.

Due to the unsuccessful cargo integration into the ΦCD9301 genome as well as very low frequency of obtaining mutant lysogens following *codA*-based AE, the RiboCas9 system was chosen to deliver *sasp* genes under the control of the RiboLac into ΦCD9301. RiboCas9 plasmids targeting locations 4A, 4B, 5A and 5B in the ΦCD9301 genome were cloned in section 4.2.1.2 and named pMTL_RC4A, pMTL_RC4B, pMTL_RC5A and pMTL_RC5B.

Genes *saspA* or *saspB* were spliced together with the RiboLac inducible promoter resulting in a cargo of the size 1780 bp. The cargo was cloned into the pMTL_RC4A, pMTL_RC4B, pMTL_RC5A and pMTL_RC5B backbones using the *Kas*I/*Avr*II restriction sites (Figure 4.12). Plasmids were confirmed by colony PCR, sequenced and named pMTL_RC4A_RiboLac_saspA, pMTL_RC4A_RiboLac_saspB, pMTL_RC4B_RiboLac_saspA, pMTL_RC4B__RiboLac_saspB,

pMTL_RC5A_RiboLac_saspA, pMTL_RC5A_RiboLac_saspB, pMTL_RC5B_RiboLac_saspA and pMTL_RC5B_RiboLac_saspB.

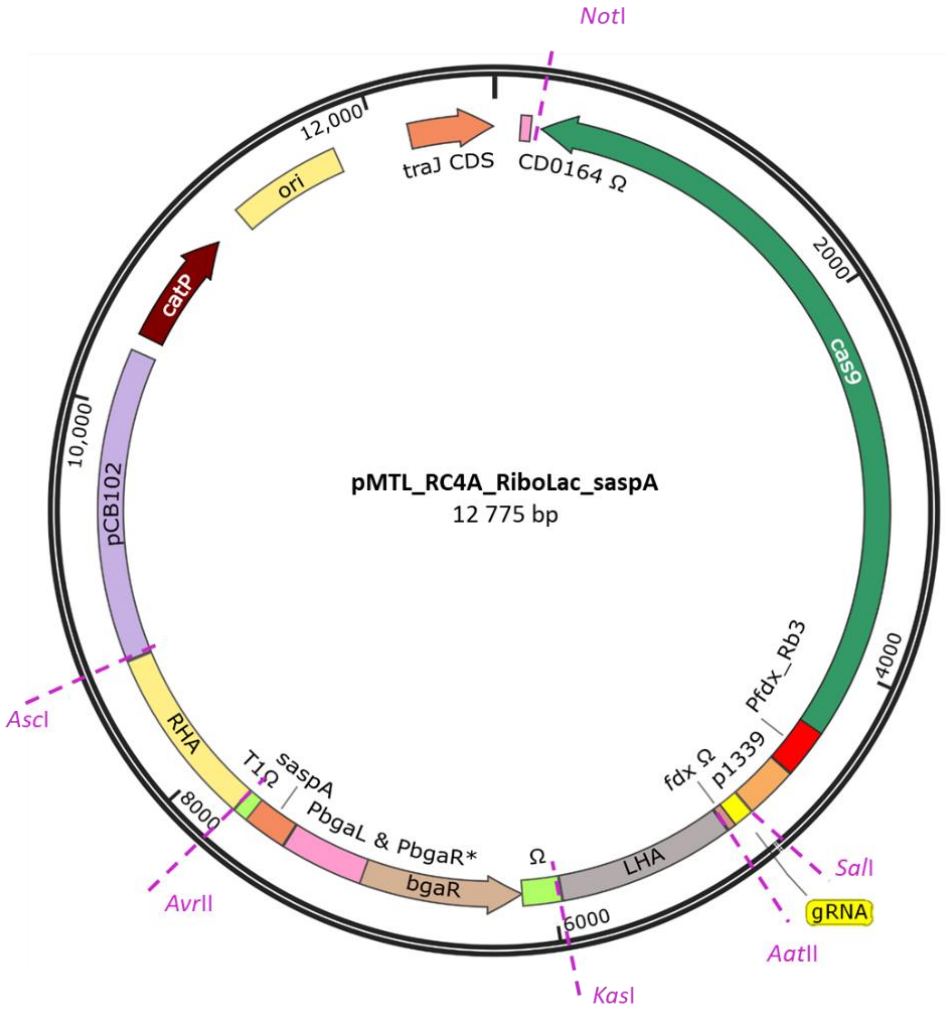

Figure 4.12: Schematic diagram of the RiboCas9 plasmid for *sasp* knockin. The plasmid was constructed in two steps. In first step, homology arms and seed gRNA sequence were cloned into the pRECas1-IIE-Cdif backbone using *Sal*I, *Aat*II and *Asc*I restriction enzymes. In second step RiboLac inducible system was spliced together with *sasp* gene and cloned into the vector via *Kas*I and *Avr*II sites.

All eight plasmids were transferred into the *C. difficile* lysogen ΦCD9301L from *E. coli* via conjugative transfer and transconjugants were selected on media supplemented with thiamphenicol. Following this, Cas9 was induced using theophylline. Similarly, as in the case of the *ermB* cargo, no colonies were observed after conjugation with plasmids targeting location 5A possibly due to failed conjugation or off-target gRNA binding. Mutant colonies obtained using the remaining six plasmids were confirmed

by colony PCR (Figure 4.13) using primer set ZZ280/ZZ281 for mutants generated with plasmids targeting 4A location, primer set ZZ282/ZZ283 for mutant cells made with plasmids pMTL_RC4B_RiboLac_saspA and pMTL_RC4B_RiboLac_saspB and primers ZZ294/ZZ295 for mutants obtained with plasmids targeting location 5B.

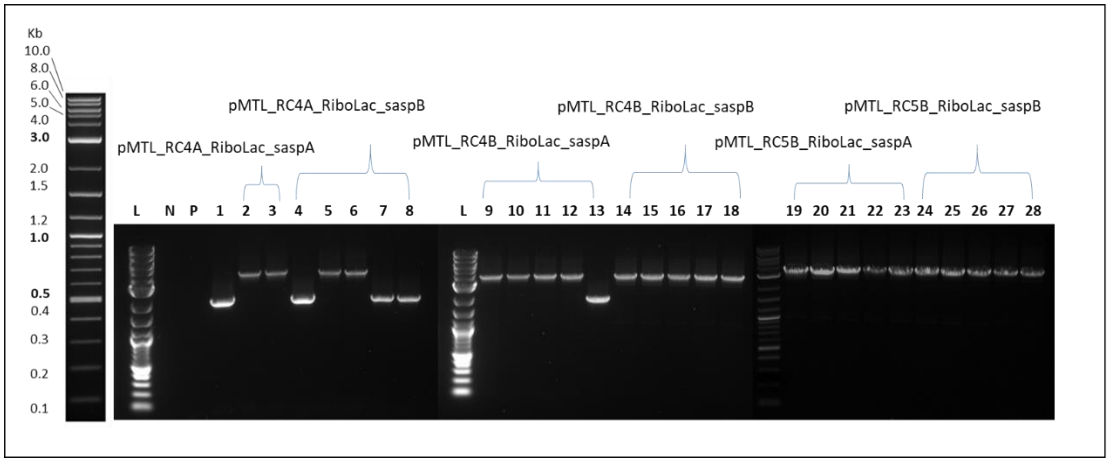

Figure 4.13: Colony PCR screening for *RiboLac_sasp* knock-in into the genome of ΦCD9301 prophage. Colony PCR was performed using Green Taq DNA polymerase with standard thermocycling conditions. L: 2 log ladder, N: distilled H_2O, P: plasmid pMTL_RC4A_RiboLac_saspA, 1: ΦCD9301 genomic DNA, 2-3: mutants obtained following pMTL_RC4A_RiboLac_saspA conjugation 4-8: mutants obtained following pMTL_RC4A_RiboLac_saspB conjugation, 9-13: mutants obtained following pMTL_RC4B_RiboLac_saspA conjugation, 14-18: mutants obtained following pMTL_RC4B_RiboLac_saspB conjugation, 19-23: mutants obtained following pMTL_RC5B_RiboLac_saspA conjugation, 24-28: mutants obtained following pMTL_RC5B_RiboLac_saspB conjugation. Mutant strains are expected to give a band size ~3.7 kbp. Non-mutant strains are expected to give a band size ~1.94 kbp.

Knock-in mutants were sub-cultured to lose the plasmid and colonies sensitive to thiamphenicol were also confirmed by sequencing. Recombinant lysogens were named CD9301L:4A_RiboLac_saspA, CD9301L:4A_RiboLac_saspB, CD9301L:4B_RiboLac_saspA, CD9301L:4B_RiboLac_saspB, CD9301L:5B_RiboLac_saspA and CD9301L:5B_RiboLac_saspB and stored as glycerol stocks at -80°C.

Mitomycin C was used to induce phage from ΦCD9301 lysogens with integrated the *sasp* cargo. After 24 hours incubation, a plaque assay was performed to check if mutant lysogens were able to release functional phage particles. Interestingly,

plaques were only present in case of induced lysogens carrying the *sasp* cargo in position 5B. No plaques were present after the induction of lysogens CD9301L:4A_RiboLac_saspA, CD9301L:4A_RiboLac_saspB, CD9301L:4B_RiboLac_saspA and CD9301L:4B_RiboLac_saspB. The induction was repeated using new mitomycin C stocks which gave the same result. Similarly, CD9301L:4A_Rb3_saspA and CD9301L:4A_Rb3_saspB mutants previously generated with *codA*-based AE carried cargo in position 4A and no plaques were isolated after the induction either. This finding was unexpected because *ermB* cargo sequence of the size of 1.67 kb have been previously shown to be successfully packed into the functional phage when inserted in positions 4A, 4B and 5B of the ΦCD9301 genome. The titre of both recombinant phages after mitomycin C induction from lysogens was approximately 10 times lower compared to the WT control (Table 4.2), which may indicate that basal expression of the *sasp* gene has toxic effects on the lysogen and the phage itself.

Table 4.2. Titre of the phages induced from mutant lysogens after mitomycin C treatment.

Lysogen	Titre of induced phage (PFU x ml^{-1})	Immunity to ΦCD9301 infection
CD9301L	9.6 x 10^2	Yes
CD9301L:5B_RiboLac_saspA	1 x 10^2	Yes
CD9301L:5B_RiboLac_saspB	0.8 x 10^2	Yes

Following the induction of CD9301L:5B_RiboLac_saspA and CD9301L:5B_RiboLac_saspB lysogens, plaques were picked and purified three times. The phage titre was increased by harvesting the top agar layer from plates followed by a plaque assay. The process was repeated until the high phage titre of ≥10^8 PFU x ml^{-1} was obtained. Once the recombinant phages reached the required concentration, genomic DNA was isolated. Mutant phages were once again confirmed via colony PCR using primer set ZZ294/ZZ295 and Sanger sequencing. Phages named ΦCD9301:5B_RiboLac_saspA and ΦCD9301:5B_RiboLac_saspB were

then stored as filtered sterilised supernatant in BHIS broth at 4°C created using 0.2% BHIS top agar plaque assay as well as lysogens stored at -80°C.

4.2.4 Characterisation of recombinant phage carrying SASP as a phage lethal cargo

Similar to the WT ΦCD9301 phage, plaques formed by the phage mutants were small and hazy and of approximately 1 mm in diameter (Figure 4.14).

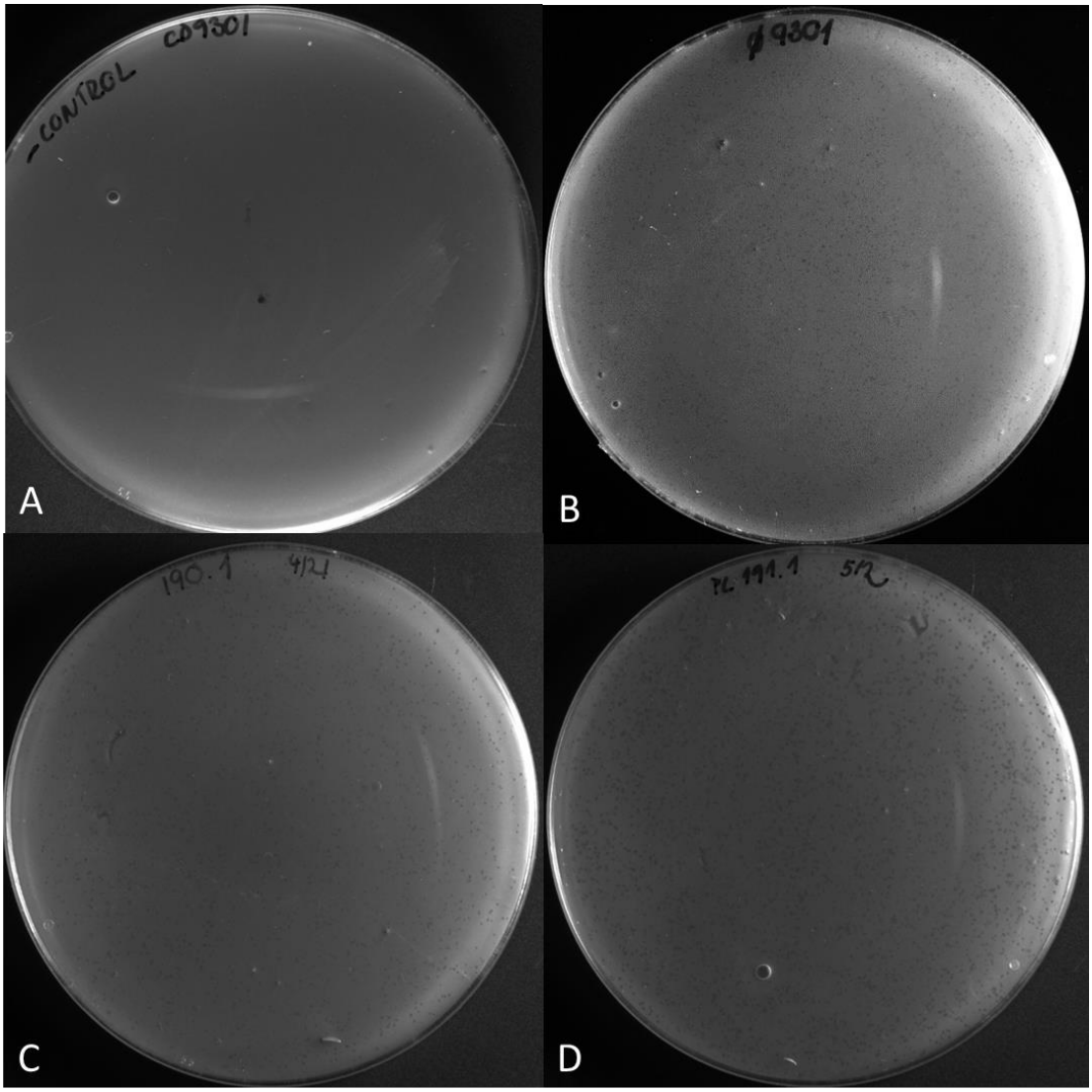

Figure 4.14: Plaques of isolated phage. A. Negative control, *C. difficile* CD9301 lawn grown without the presence of the phage, **B.** ΦCD9301 phage stock infecting *C. difficile* CD9301, **C**. Modified phage ΦCD9301:5B_RiboLac_saspA stock infecting *C. difficile* CD9301, **D**. Modified phage ΦCD9301:5B_RiboLac_saspB stock infecting *C. difficile* CD9301.

Since it was unclear whether genetic modification of the phages would affect their host range, plaque assays were performed with modified phage against several *C. difficile* strains. It was previously demonstrated (section 3.2.4) that the WT phage ΦCD9301 infects 24/100 *C. difficile* isolates across 6 different ribotypes. Both modified phages ΦCD9301:5B_RiboLac_saspA and ΦCD9301:5B_RiboLac_saspB have shown to be able to infect the same *C. difficile* isolates CD0801, CD2101, CD6001, CD2301, CD4001, CD5601, CD19_2, CD4901, CD5401, CD9301, CD3801, CD8601, CD9001, CD9501, CD9701, CD1801, CD2917, CD2916, CD2918, CD2906, CD3205, CD3176, CD3177, CD1801 as the ΦCD9301 phage, indicating that the insertion of the SASP cargo into the phage genome does not affect their host range.

To confirm that the genomes of the modified phages do not carry any undesirable mutations possibly affecting their phenotypes, all genomes were sequenced using Illumina MiSeq 500bp V2 SBS sequencing service. Sequencing reads of both mutant phages with an average read length of 300 bp were mapped to the ΦCD9301 reference genome using CLC Genomics Workbench 11. Firstly, the analysis confirmed that *sasp* cargos were inserted in the location 5B. Secondly no SNV was detected in the genome of recombinant phage when compared to the ΦCD9301 phage. This result confirms that the phages are suitable for further work in this study.
To further understand the inability of the remaining *sasp* lysogens to form phage particles, their genomes were sequenced and mapped to the reference genome of lysogen ΦCD9301L. The analysis showed no presence of SNV or genome alterations that could explain the failure to induce functional phage particles. This data suggest that the induction failure is most probably caused by the position of cargo and the toxic nature of SASPS.

4.2.5 *In vitro* activity of *sasp* recombinant phages against *C. difficile*

To test whether the phage-mediated delivery of a *sasp* cargo could enhance WT phage activity to more efficiently kill a target population of *C. difficile*, the activity of engineered phages was measured together with the appropriate controls using the *in vitro* CFU reduction approach described by Selle *et al*., 2020.

Briefly, an overnight culture *of C. difficile* indicator strain was sub-cultured 1:100 into BHIS and incubated to an OD_{600nm} of 0.2. Phage stocks were mixed 1:1 with *C. difficile* and induced with lactose (10 mM) and theophylline (2 mM). Cultures were grown for 22 hours and their OD_{600nm} as well as number of *C. difficile* colonies grown on the BHIS plate containing the inducer were monitored over time by means of a manual growth curve.

Data from these experiments have shown that WT ΦCD9301 has an inhibitory effect on *C. difficile* growth *in vitro*, as ΦCD9301 infected cultures reached almost 3 times lower OD_{600nm} values up to 6 hours of infection (Figure 4.15A). Cultures treated with WT phage also exhibited an almost 2-log reduction in CFU counts after up to 6 hours of incubation (Figure 4.15B). The inhibitory effect of the phage seemed to lessen with time as no significant difference in OD_{600nm} was observed at the 22 hours endpoint although 1 log reduction was still observed in cells treated with ΦCD9301 comparing to the control following 22 hours incubation in liquid. Both cultures treated with modified phages carrying the *saspA* or *saspB* genes displayed slightly lower OD_{600nm} and CFU/ml counts after 2 and 4 hours incubation comparing to the WT. This reduction in OD was not apparent at the subsequent timepoints of 6 or 22 hours (Figure 4.15). As a consequence, the overall conclusion is that no substantial difference was observed in the activity of the mutants phage compared to that of the WT phage. Several external conditions of the experiment were investigated in an attempt to optimise the growth conditions and thus the expression of lethal genes. Firstly, different concentrations of the inducers were tested in both liquid cultures as well as plates such as 0 mM T (theophylline)/0mM L (lactose); 0.5 mM T/10 mM L; 2 mM T/ 10 mM L and 5 mM T/15 mM L. Secondly, different ratios of *C. difficile*:phage were tested such as 1:2 and 1:3. Time of induction was also investigated. In all cases, no outstanding difference in activity was observed between WT phage ΦCD9301 and ΦCD9301:5B_RiboLac_saspA and ΦCD9301:5B_RiboLac_saspB.

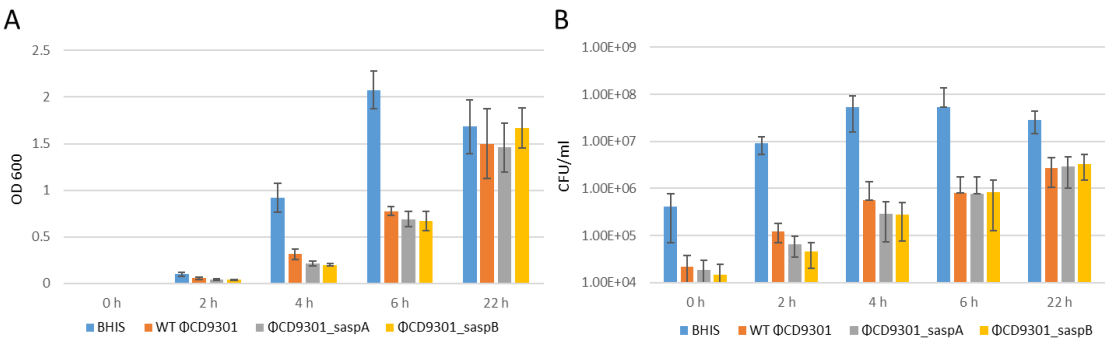

Figure 4.15: Comparison of the phage activity *in vitro*. *C. difficile* cultures were infected with WT or mutant phage carrying the *sasp* gene under the control of the RiboLac inducible system. As a negative control *C. difficile* was mixed with BHIS broth. All four cultures were induced with lactose (10 mM) and theophylline (2 mM) and the activity was monitored over the time by OD_{600nm} measurements as well as by the number of *C. difficile* colonies obatined on the BHIS agar plates containing the inducer. The experiment was repeated 5 times to calculate standard deviations, as represented by error bars. An approximately 3-fold reduction of the OD_{600nm}, as well as an almost 2-log reduction of CFU/ml, was observed in cultures treated with WT phage for up to 6 hours. No significant difference was observed in the activity of mutant phages comparing to the WT phage.

4.3 Discussion and summary

Several studies focused on the development of recombinant phages have been published. With temperate phages, engineering is easily achieved using tools developed for its bacterial host (Park et al., 2017). As it has been predicted that only 10% of additional genetic information can be successfully packed into the phage particle (Casjens and Gilcrease, 2009), novel phage ΦCD9301 with the largest genome (50.503 kb) was been selected for the delivery of lethal cargo into the *C. difficile*.

Initial studies focused on the identification of suitable positions in the phage genome for the cargo knock-in. Because of its small size and limited impact on the host cell viability, the gene encoding erythromycin resistance (*ermB*) was selected as a cargo to test several candidates. In the initial experiments a knock-in vector, based on the *codA* gene as the counter selection, was deployed to integrate cargo into 5 different locations, designed to precede phage natural terminators to avoid interference with the gene expression of the phage. Although, the cargo was successfully integrated into all selected positions, no phage particles were obtained following mitomycin C induction. We postulate that this was caused by the cargo interference with the replication and the packaging of the phage DNA.

Despite successful integration of *ermB* cargo into the ΦCD9301 lysogen, the efficiency of obtaining mutants via AE using *codA* counter selection was low. As such, a strategy employing RiboCas9 system was considered. Four new locations in the phage genome were selected for cargo integration. Selection process independent of the presence of native phage terminators allowed broader range of options. Based on the previous results, all four locations were chosen to be outside of the structural module of the genome and replication genes.

RiboCas9 has proven to be a more suitable system for gene integration as the high efficiency of mutant generation and ease of application made the process simpler and less time consuming. The selection of positions outside of structural and assembly genes appear to be crucial in the generation of recombinant phage as induction of phages from lysogens carrying cargo in positions 4A, 4B and 5B was successful. This is in corelation with previously pubslished study that proved

successful lethal cargo integration into the *S. aureus* phage (Park *et al*., 2017). This data has shown that the additional genetic information of the size of 1.67 kb can be packed into the functional ΦCD9301 particles.

Due to its small size and toxic effect on vegetative cells, SASPs were selected as a suitable candidate for antimicrobial cargo delivery in the phage ΦCD9301. SASPs are small proteins, naturally occurring during sporulation in spore-forming bacteria. At the early stage of sporulation, these proteins bind to the DNA in a no-sequence specific manner, change the DNA conformation and protect it against UV, heat and other agents (Setlow, 2006). It has been hypothesised that SASPs have toxic effect on vegetative cells due to its interference with DNA replication and gene expression which leads to cell death. Indeed, the use of SASPs as an antimicrobial agent is being employed by Phico Therapeutics. They have used phage vectors for the SASP delivery targeting several *S. aureus* strains in order to treat skin infections (Fairhead, 2009). Due to the potential high toxic effect on vegetative cells, the choice of an appropriate promotor system is essential in order to achieve expression of toxic levels of SASPs in the target host while avoiding cell death of the propagating strain. Two inducible promotor systems appeared to be the right candidates based on the previous research in the SBRC. Both Rb3 and RiboLac exhibit a high dynamic range of activity and little background noise when used in *C. sporogenes* (Canadas, PhD thesis, 2019).

To investigate the impact of SASPs on vegetative cells of *E. coli* and *C. difficile*, two *sasp* genes from *B. subtilis* were separately cloned into the plasmid together with the inducible promotor system and transferred into the host cells.

The toxic effect of theophylline was observed in *E. coli* cells incubated in LB with 5 mM concentration of theophylline. This is in correlation with the findings of Kamiura *et al*. (2019) who observed no inhibitory effect at a theophylline concentration less or equal to 1 mM, but a decline in growth at 5 mM theophylline. On a contrary, this finding was not observed in *C. difficile*, which indicates that theophylline is safe to use as an inducer in *C. difficile* with concentrations up to 5 mM. Concentrations higher than 5 mM resulted in a reduction of growth, in agreement with other previously published studies (Desai and Gallivan, 2004; Canadas *et al*., 2019).

In *E. coli*, compared to the Rb3 system, the RiboLac system appeared to be less leaky, with no difference being observed when comparing the OD_{600nm} of uninduced cells carrying *saspA* and the empty vector control. However, it is important to note that this was not observed in the *saspB* group, where uninduced cells showed a significant reduction in OD_{600nm}. A similar phenomenon has been observed when the Rb3 system was used to regulate gene expression. It is postulated that SASPB may bind to the DNA with higher affinity than SASPA and even a minimal amount of copies are sufficient to have lethal effects on the cell.

In *C. difficile*, the difference in the level of basal expression was not as apparent when comparing both the Rb3 and RiboLac promoter systems, however, as previously seen in *E. coli*, the uninduced *saspB* group of *C. difficile* demonstrated significantly lower viability with both systems.

In *E. coli* both SASPs have shown to have a different negative impact on cell viability with cells carrying the *saspB* gene displaying more than 40% less viability than cells carrying the *saspA* gene. This may imply that the latter is less lethal. The detrimental impact of SASPB appeared to be even stronger in *C. difficile*, where *saspB* expressing bacteria presented 50% decrease in cell mass comparing to the *saspA* group.

Based on previous observations, it has been predicted that both SASPs have different effects on the cell. Therefore, both *sasp* genes were separately introduced into the phage genome. Because of its smaller size (549 bp) and similar effectivity as RiboLac, the theophylline inducible system Rb3 was originally chosen to be integrated together with *sasp* genes into the ΦCD9301 lysogen. Despite the low efficiency, mutant lysogens carrying cargo in position 4A were generated. Unfortunately, assembly of the recombinant phage was not successful despite the fact that selected location had previously shown to be successful for the *ermB* cargo knock-in.

Due to the unsuccessful cargo integration into the ΦCD9301 genome, RiboLac_*sasp* was used to replace it. Although cargo was successfully integrated in all three locations 4A, 4B and 5B, recombinant phage particles were only obtained after induction from lysogens carrying *sasp* cargos in position 5B. This finding was unforeseen as both positions 4A and 4B have previously been shown to be suitable for *ermB* cargo knock-in and phage assembly. Compared to positions 4A and 4B,

recombinant lysogens carrying *ermB* in position 5B were induced to the highest titre following mitomycin C treatment, although the titre was still 2 times lower than the WT phage control. The titre of both recombinant phages ΦCD9301_5B_RiboLac_saspA and ΦCD9301_5B_RiboLac_saspB after the induction from their respective lysogens was approximately 10 times lower compared to the WT control. We postulate that even with the theophylline riboswitch used to control the expression of both inducible systems Rb3 and RiboLac, basal expression of *sasp* genes have toxic effects on the phage itself and interfere with phage replication or assembly. To better understand the impact of *sasp* genes on phage and the manufacturing strain, genome sequences of recombinant lysogens were analysed. The analysis showed no difference in the genome of recombinant lysogens that could explain the failure to induce functional phage particles. It is, therefore, assumed that the failure to induce phage particles is most likely caused by the position of cargo and the toxic nature of SASPS.

Following the induction of CD9301L:5B_RiboLac_saspA and CD9301L:5B_RiboLac_saspB lysogens, genomes as well as the host range of the recombinant phages were analysed. Sequencing data showed that no ancillary mutations had arisen. Similarly like Selle *et al*., 2020 our engineered phages were able to infect the same 24/100 *C. difficile* isolates as the WT phage, proving that the lethal cargo has no negative impact on the host range of the phage.

To investigate the effect of engineered phages on *C. difficile*, their activity was measured together with the WT phage *in vitro.* These experiments demonstrated that despite the presence of the *saspA* or *saspB* lethal cargo, no substantial difference was observed in the activity of recombinant phages comparing to the WT ΦCD9301. All phages displayed similar activity, reducing the numbers of *C. difficile* CFU by approximately 2-log after 6 hours of incubation. The inhibitory effect of the phages seemed to decrease with the time as no significant difference in OD_{600nm} was observed after 22 hours of incubation, although a 1-log reduction in CFU was still detected on plates in *C. difficile* groups treated with phages comparing to the BHIS control. The decrease in the inhibitory effect over time was anticipated as the development of the bacterial resistance towards single phage infection has been

previously shown to be common in many studies (Selle *et al*., 2020). This can be overcome by using a cocktail of different phages.

A similar activity of the mutant phage compared to the WT was unexpected as both *sasp* cargos have been proven to have detrimental effect on *C. difficile*, when delivered by plasmid. High copy number plasmids used in this study deliver multiple copies of the gene of interest to the host, whereas recombinant phage integrated into the host genome delivers only single copy of lethal cargo. This may explain the discrepancy in results observed when sasp was delivered to the host on plasmid comparing to the same cargo delivery by the phage. It is also predicted that SASPs have a negative impact on the phage itself, as unlike phage carrying the *ermB* cargo, phage particles were only induced when the *sasp* cargo was integrated into position 5B and even then, the induction titre was 10 times lower than the control.

Another explanation is that the expression of the *sasp* genes from the RiboLac inducible system was not sufficient to affect the host. Fairhead *et al*., 2009 replaced phage lytic genes with *sasp* and isolated recombinant phage unable to access lytic cycle. In their study, the use of the weak constitutive promotor of the *fbaA* gene foreign to the phage ensured that SASP expression was sufficiently low not to affect the propagating strain but sufficient to kill the target bacteria after phage replication within the bacterial genome and accumulating enough copies of protein. Stable integration of recombinant phage into the host genome accumulated sufficient levels of lethal cargo over time leading to host death. As our recombinant phage has an acces to the lytic life cycle, we postulate that cargo delivery by our phage did not achieve high enough levels of lethal cargo necessary to have detrimental impact on *C. difficile*. Moreover, Fairhead *et al*. (2009) employed the *saspC* gene from *B. megaterium* to tackle *S. aureus* infections using phage vector. Comparing the amino acid sequences of individual proteins, SASPA and SASPB used in this study shared 94% and 69% identity to SASPC. According to Nerber and Sorg (2021), the impact of SASPs on the host differ despite the strong sequence similarity within and between organisms. It is possible that SASPC binds to the DNA with higher affinity comparing to SASPA or SASPB tested in this study. Different effects on the host have been observed in this study when comparing SASPA and SASPB and it can be assumed that

the *saspC* gene together with the weak *fbaA* gene promoter had a detrimental effect on the bacterial host without affecting the phage vector.

In conclusion, we have generated two recombinant *C. difficile* phages carrying lethal *saspA* or *saspB* genes under the control of the RiboLac inducible promoter system. Unfortunately, the integration of *sasp* genes into the phage genome did not lead to the improvement of the phage activity *in vitro*. This might be a consequence of the insufficient expression levels of the lethal cargo or SASP interference with phage itself. In light of these observations, it was decided to explore alternative avenues in the pursuit of a modified phage harbouring lethal cargo.

4.5 Chapter summary

- Three locations in the ΦCD9301 genome have been established to be suitable for cargo knock-in with the phage being able to package additional genetic information of the size of 1.67 kb into the functional particles.
- The effect of SASPA and SASPB on cell viability have been determined in *E. coli* and *C. difficile.*
- Inducible systems Rb3 and RiboLac were compared.
- Both *sasp* genes were individually integrated together with the RiboLac inducible promoter system into the ΦCD9301 genome, however, no change in the phage activity was observed *in vitro* comparing to the WT phage.

Chapter five:

Using CRISPR-Cas Systems as an Antimicrobial against *C. difficile*

5. Using CRISPR-Cas Systems as an Antimicrobial against *C. difficile*

5.1 Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) proteins represent an adaptive, small RNA-based immune system in bacteria and archaea that confer resistance to foreign genetic elements such as bacteriophages or plasmids. The CRISPR defence system consists of three distinct steps: adaptation, expression and interference (Fage *et al*., 2021).

During adaptation, foreign DNA is cut into small fragments and incorporated into the CRISPR locus in the form of short sequences (around 20 bp) referred to as spacers. The nucleotide sequence of the spacer corresponds to the part of exogenous DNA, called protospacer. Bacterial chromosome spacers are incorporated between 24-48 bp long palindromic direct repeat sequences, iterations of which constitute CRISPR arrays. During the expression phase, *cas* genes are expressed and CRISPR loci are transcribed and processed into small CRISPR RNA (crRNA) by Cas enzymes. Finally, in the third stage of CRISPR immunity called interference, crRNA in complex with Cas proteins recognise and cleave the invading DNA based on sequence complementarity. Cleavage of the foreign nucleic acid results in the destruction of the invading genome and immunity (Rath *et al*., 2015).

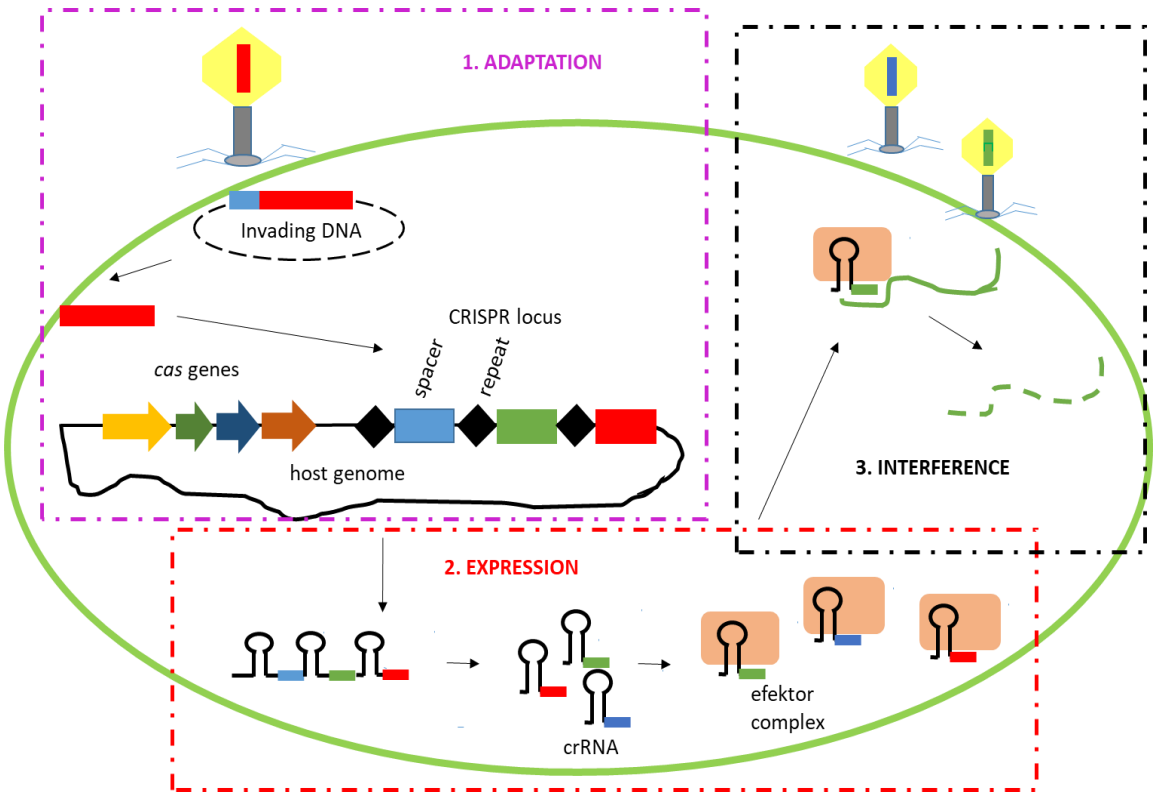

Figure 5.1: Schematic representation of bacterial CRISPR system. During adaptation (purple square) foreign DNA is recognised, cleaved and incorporated into the CRISPR locus as a spacer. Expression step (red square) involves the transcription of CRISPR locus and crRNA processing. Finally, during interference (black square), crRNA in complex with Cas proteins recognise and cleave the invading DNA that match the spacer sequence (adapted from Semenova and Severinov, 2017).

An important component of the CRISPR mediated defence system is the protospacer-adjacent motif, PAM. PAM is a short 3 bp sequence located at the 3' or 5' end of the protospacer, required for protospacer recognition. The absence of PAM in a bacterial CRISPR array prevents the bacteria from self-targeting and autoimmunity (Maikova *et al*., 2019). Cas proteins are involved in all three stages of CRISPR immunity and their gene clusters are often associated with CRISPR arrays. Based on *cas* gene sets variations and operon organisation, prokaryotic CRISPR-Cas systems can be divided into three major types that can be further divided into 12 subtypes (Bourdy *et al*., 2015).

The type I CRISPR/Cas systems represent the most diverse group, currently separated into six subtypes (IA – IF), that have variable numbers of *cas* genes. A common feature for all types are the activity of Cas1 and Cas2 enzymes, involved in spacer acquisition during adaptation. The type I systems are defined by the presence of the Cas3

enzyme responsible for degrading the invading DNA molecule. Apart from Cas1, Cas2 and Cas3, all type I systems encode an enzyme complex called Cascade. In most cases, Cascade processes the crRNA and locates the target. The subtype I-B system has been found in *Clostridium* and is characterised by the presence of the subtype-specific protein Cas8 (Maikova *et al*., 2018; Hille *et al*., 2018).

While types I and III have been identified in both bacteria and archaea, type II is unique to bacteria only. Type II is the most studied and best characterized among the three types. This system encodes Cas1 and Cas2 as well as the signature protein Cas9. Cas9 plays an active role in all three stages of CRISPR defence.

Type III encodes the signature protein Cas10 with unclear function. Most Cas proteins of type III form complexes that are similar to Cascade. Unlike the other two types, type III has been shown to target RNA as well as DNA (Rath *et al*., 2015; Maikova *et al*., 2018).

Recent studies demonstrated the occurrence of an active CRISPR subtype I-B system in *C. difficile*. The presence of CRISPR loci in *C. difficile* was described for the first time in 2006 by Sebaihia *et al*. (2006), but its widespread occurrence and diversity is yet to be explored. The first study demonstrating the expression of multiple CRISPR arrays in *C. difficile* was published in 2013, when Soutourina and co-workers (Soutourina *et al.*, 2013) revealed the presence of diverse crRNA during deep sequencing of regulatory RNAs in *C. difficile*. Subsequently, the type I-B CRISPR-Cas system was identified in more than 2 500 analysed genomes of *C. difficile* (Hargreaves *et al*., 2014; Andersen *et al*., 2016; Maikova *et al*., 2018).

The *C. difficile* CRISPR-Cas system is typified by an unusual high number of CRISPR arrays per genome. For instance, genomes of *C. difficile* 630 and *C. difficile* R20291 have been shown to carry 12 and 9 CRISPR arrays, respectively (Soutourina *et al*., 2013; Boudry *et al*., 2015). On average, *C. difficile* genomes contain 8.5 arrays, which is notably more than the typical one to three CRISPR arrays described in other major pathogenic bacteria (Andersen *et al*., 2016). Interestingly, several *C. difficile* CRISPR arrays have been found to be localised and actively expressed in prophages (Hargreaves *et al*., 2014; Boudry *et al*., 2015). It is assumed that the localization of CRISPR arrays within prophages may protect them and their lysogenic host from the

secondary infection by competing phages (Hargreaves *et al*., 2014). Similarly, like other highly expressed bacterial genes, *C. difficile* CRISPR arrays are positioned in the direction of the chromosome replication (Boudry *et al*., 2015).

The majority of sequenced *C. difficile* genomes also appear to have two or three *cas* gene operons. The complete *cas* operon is usually associated with the longest CRISPR array and encodes all Cas proteins necessary for adaptation (Cas1, Cas2, Cas4) as well as for interference (Cas6, Cas8, Cas7, Cas5, Cas3). The additional *cas* gene sets are missing the spacer acquisition genes. While incomplete *cas* operons are present in almost all sequenced strains, complete *cas* operons were found in approximately 90% of them, suggesting that some *C. difficile* strains have lost their ability to adapt to invading DNA through CRISPR (Boudry *et al*., 2015).

The identification of protospacers in phage genomes has allowed for the prediction of PAM sequences. Multiple PAMs were identified that are recognised by the *C. difficile* CRISPR system. CCA or CCT motifs localised on the 5' end were the most common PAMs, however, alternative PAMs CCC, CCG and TCA were also found (Boudry *et al*. 2015).

Many applications have arisen from *C. difficile* CRISPR systems. Since the order of spacers within an array reflects phage invasion in different populations of the same bacterial species, Andersen *et al*. (2016) proposed the use of *C. difficile* CRISPR for genotyping purposes.

Since less components are required for genomic manipulation, endogenous CRISPR systems are favourable over the most widely adopted exogenous Cas9 system taken from *S. pyogenes*. Endogenous CRISPR type I-B systems have been successfully applied for genome modification in several members of the genus *Clostridium* such as *C. pasteurianum* (Pyne *et al*., 2016), *C. tyrobutyricum* (Zhang *et al*., 2018) or *C. butyricum* (Zhou *et al*., 2021). A recent study successfully redirected the native *C. difficile* CRISPR system. Therein, Maikova *et al*., 2019 used plasmids carrying artificial CRISPR miniarrays with spacers targeting *hfq* in *C. difficile* 630Δerm and R20291. crRNA expressed from the plasmid formed an effector complex with the native Cas machinery, cleaving the protospacer of choice and killing WT cells. Moreover, adding

homologous arms flanking the protospacer resulted in a highly efficient editing plasmid and construction of *Δhfq* recombinant cells (Maikova *et al*., 2019).

Novel therapeutics using both heterologous and endogenous CRISPR systems are rapidly emerging with several groups using phagemids or phage particles as vectors to deliver CRISPR-Cas components to pathogenic bacteria (Bikard and Barrangou, 2017). For instance, M13 phagemid was used to deliver Cas9 together with gRNA targeting various antibiotic resistance genes such as *blaSHV-18*, and *gyrAD87G* into *E. coli* leading to cell death both *in vitro* and *in vivo* using the wax worm model (Bikard *et al*., 2014). In another study, phiNM1 phagemid carrying the CRISPR-Cas9 system eliminated targeted *S. aureus* strains in a mouse skin colonization model (Citorik *et al*., 2014).

Promising results were also obtained with CRISPR/Cas recombinant phage for treating systemic bacterial infections. Park *et al*. (2017) knocked a CRISPR-Cas9 array into the *S. aureus* phage ΦSaBov. The resulting phage displayed significantly enhanced activity against *S. aureus* comparing to the WT phage both *in vitro* and *in vivo* using the murine skin infection model. The efficiency of this phage was also measured in the treatment of osteomyelitis and soft tissue infection in rats infected with *S. aureus*. Notably, recombinant phage was successful in reducing soft tissue infection but not bone infection, suggesting that phages carrying CRISPR/Cas systems may not be efficient against all infections (Cobb *et al*., 2019).

5.1.1 Aims of this study

In the previous chapter, *sasp* genes were integrated into the ΦCD9301 genome as a lethal cargo, however, no change in the phage activity was observed *in vitro* comparing to the WT phage. In this study, we propose to test the concept of re-directing heterologous and endogenous CRISPR/Cas systems towards toxic genes specific to this pathogen using recombinant phage ΦCD9301.

5.2 Results

5.2.1 Construction of the *C. difficile* manufacturing strain

To avoid self-targeting of the *C. difficile* CD9301 manufacturing strain used for CRISPR cargo knock-in into ΦCD9301, the CRISPR targets *tcdA* and *tcdB* must first be deleted. To achieve this, the RiboCas9 genome editing tool (Canadas *et al*., 2019) was applied to generate mutant *C. difficile* CD9301 strains with individual *tcdA* or *tcdB* deletions, or a double *tcdA*/*tcdB* deletion. Plasmids used in this study were derived from the RiboCas9 plasmid created by I. Canadas (Canadas *et al*., 2019). Homology arms flanking *tcdA* or *tcdB* genes or a combination of both *tcdA* and *tcdB* genes were PCR amplified using *C. difficile* CD9301 genomic DNA. In order to select the most efficient gRNA for further work in this study, four gRNA seed sequences were designed using Benchling for each toxin gene. Each gRNA was constructed in a primer dimer PCR reaction and cloned together with the respective homology arms into the pRECas1-IIE-Cdif backbone using the restriction sites *Asc*I, *AsiS*I and *Sal*I or *Asc*I, *Aat*II and *Sal*I as outlined in section 2.7.2. Plasmids were sequenced and named pMTL_RCas_tcdA1-4, pMTL_RCas_tcdB1-4 and pMTL_RCas_tcdBA1-4. The cloning was confirmed by colony PCR and the sequence of constructs was verified by sequencing.

The RiboCas9 plasmids were conjugated into *C. difficile* CD9301. The resultant transconjugants were selected with thiamphenicol and induced with theophylline as outlined in the section 2.6.14. For *C. difficile*Δ*tcdA* knockouts, 28 colonies were screened by PCR using the forward primer ZZ17, binding within the LHA and the reverse primer ZZ59, binding in the genome downstream of the *tcdA* gene (Figure 5.2). The highest CRISPR/Cas9 efficacy (100%) was observed when the tcdA_3 seed sequence sgRNA was used. The lowest CRISPR/Cas9 efficacy (0%) was observed when the tcdA_1 vector was used, since no colony was obtained after conjugation into *C. difficile*. The mutant strains were subcultured three times to achieve plasmid loss and colonies sensitive to thiamphenicol were further confirmed by PCR amplifying and sequencing the deleted region.

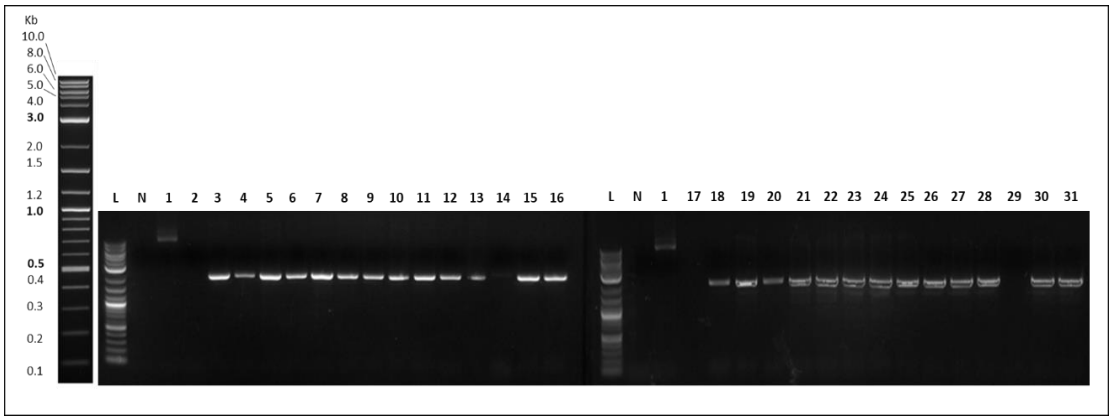

Figure 5.2: Colony PCR screening for *C. difficile* CD9301*ΔtcdA* colonies. Colony PCR was performed using Phusion DNA polymerase with standard thermocycling conditions. L: 2-log DNA ladder (NEB), N: distilled H2O, lane 1: *C. difficile* CD9301 genomic DNA, 2: plasmid pMTL_RCas_tcdA1, lanes 3-31: single colonies following conjugation of pMTL_RCas_tcdA plasmids. Non-mutant strains were expected to give a 10.27 kB band, mutant *ΔtcdA* strains were expected to give a band sized 2.13 kB.

In the case of *C. difficile*Δ*tcdB* mutants, 31 colonies were screened by colony PCR using the forward primer ZZ62, binding in the chromosome upstream of the homology arm and the reverse primer ZZ46, binding in the RHA (Figure 5.3). The highest CRISPR/Cas9 efficacy (100%) was observed when the tcdB_2 seed sequence sgRNA was used. The lowest CRISPR/Cas9 efficacy (16.7%) was observed when the pMTL_RCas9_tcdB_1 vector was used. The mutants were subcultured to lose the plasmid and thiamphenicol sensitive colonies were further confirmed by Sanger sequencing.

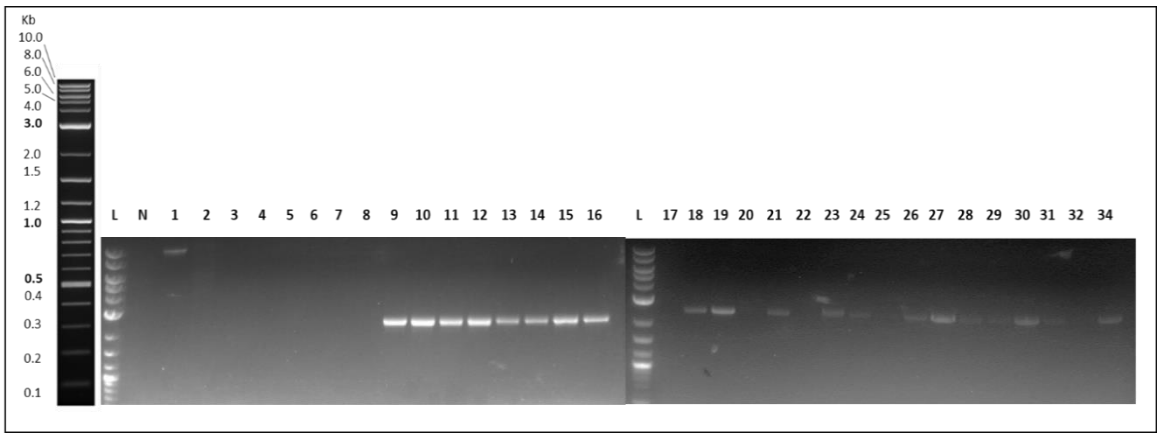

Figure 5.3: Colony PCR screening for *C. difficile* CD9301ΔtcdB mutants. Colony PCR was performed using Phusion DNA polymerase with standard thermocycling conditions. L: 2 log ladder, N: distilled H_2O, 1: *C difficile* CD9301 gDNA, 2: plasmid pMTL_RCas_tcdB1, lanes 3-34: single colonies following conjugation of pMTL_RCas_tcdB plasmids. Primers were designed to bind to the genome outside of the homology arms. Mutant strains were expected to give a band of size of 2.1 kbp. Non-mutant strains are expected to give a band sized of 9.2 kbp.

Four RiboCas9 plasmids targeting both genes *tcdB* and *tcdA* were also delivered into *C. difficile* and colonies were screened via colony PCR (Figure 5.4). Together, 16 colonies were screened using the forward primer ZZ47, binding within the LHA and reverse primer ZZ59 binding in a genome downstream of *tcdA*. No colony was observed following conjugation of plasmids pMTL_RCas_tcdBA1 and pMTL_RCas_tcdBA2. The highest CRISPR/Cas9 efficacy (100%) was observed when tcdBA_3 seed sequence sgRNA was used. Knock out of genes *tcdB* and *tcdA* of the mutants with lost plasmid was further confirmed by PCR amplifying and sequencing the region of deletion. The mutant strains were subcultured to lose the plasmid and thiamphenicol sensitive colonies were further confirmed by Sanger sequencing.

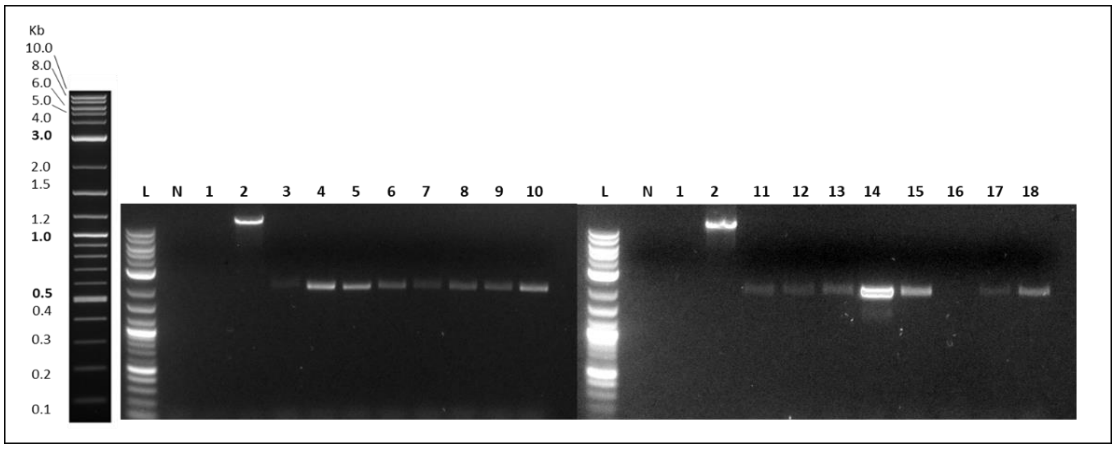

Figure 5.4: Colony PCR screening for *C. difficile* CD9301*ΔtcdB_tcdA* mutants. Colony PCR was performed using Phusion DNA polymerase with standard thermocycling conditions. L: 2 log ladder, N: distilled H_2O, 1: plasmid pMTL_RCas_tcdBA1, 2: *C. difficile* CD9301 gDNA, lanes 3-18: single colonies following conjugation of pMTL_RCas_tcdBA plasmids. Primers were designed to bind in the genome outside of the homology arms. Mutant strains were expected to give a band size of 2.142 kbp. Non-mutant strains were expected to give a band size of 18.716 kbp.

5.2.2 Isolation of *C. difficile* lysogens

Lysogens of isolated phage ΦCD9301 were made using mutant manufacturing strains *C. difficile*Δ*tcdA, C. difficile*Δ*tcdB* and *C. difficile*Δ*tcdB_tcdA* to enable phage genome editing. Each phage lysate was spotted onto a plate with the appropriate lawn of *C. difficile* propagating strain and incubated overnight to form zones of clearing. The following day, a few colonies of potential lysogens growing within a zone of clearing were re-streaked 3 times for purity. Phage particles were induced from each lysogen using mitomycin C and tested in a plaque assay with the propagating *C. difficile* strain CD9301.

Induction titres were determined to be similar in all cases (Table 5.1). Furthermore, a plaque assay was performed with phage ΦCD9301 against the lawn of respective *C. difficile* lysogen to confirm the immunity of each lysogen against further phage infection. Lysogens were named after their *C. difficile* propagating strains CD9301ΔtcdAL, CD9301ΔtcdBL and CD9301ΔtcdA_tcdBL and stored as a glycerol stock at -80°C.

Table 5.1. Isolation of *C. difficile* lysogens.

***C. difficile* lysogen**	**Titre of induced phage (PFU x ml^{-1})**	**Immunity to phage infection**	
		Phage	**Immunity**
CD9301Δ*tcdA*L	8.9 x 10^2	ΦCD9301	Yes
CD9301Δ*tcdB*L	1 x 10^3	ΦCD9301	Yes
CD9301Δ*tcdB_tcdA*L	1.2 x 10^3	ΦCD9301	Yes

5.2.2.1 Genome analysis of CD9301*ΔtcdA_tcdB*L

To confirm the integration of the phage ΦCD9301 into the genome of the manufacturing *C. difficile* strain, genomic DNA of the CD9301*ΔtcdA_tcdB*L lysogen was extracted using phenol-chloroform and sequenced using Illumina MiSeq sequencing service.CLC Genomics Workbench 11 was used to map 2 504 978 reads with an average read length of 300 bp. Sequencing reads were trimmed and assembled by their alignment to the reference genome of *C. difficile* CD9301 and the genome of ΦCD9301.

Firstly, the deletion of *C. difficile* toxin genes was confirmed, with no reads being present in case of both *tcdB* and *tcdA* gene (Figure 5.5).

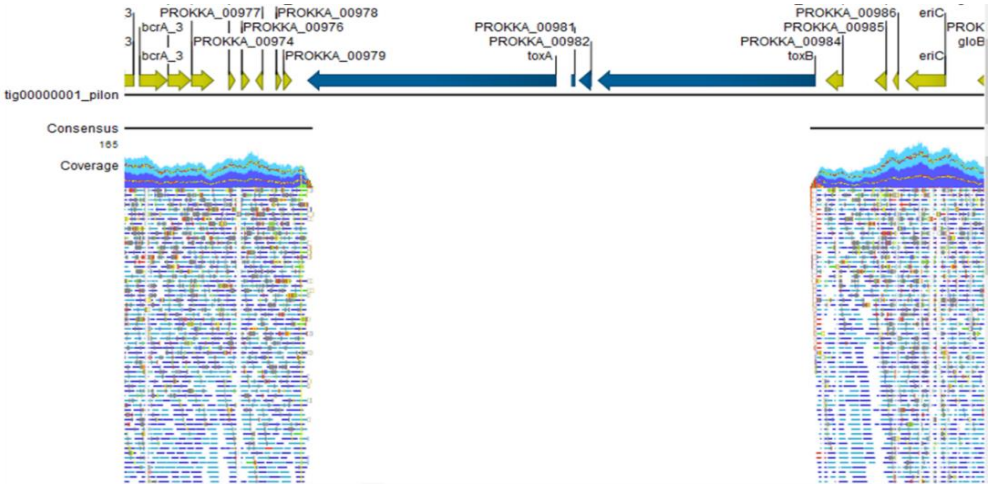

Figure 5.5: Deleted toxin genes of CD9301*ΔtcdA_tcdB*L lysogen. This shows alignment of sequencing reads to the genome of *C. difficile* CD9301. No read coverage is detected in the case of genes encoding for toxin gene B, *tcdR* repressor, hypothetical protein and toxin gene A (blue arrows) as opposed to the high read coverage of the area upstream and downstream of the deletion.

In order to determine the location of ΦCD9301 integration within the bacterial genome, the integration site was manually search for by locating the integrase gene of the phage. Phages usually integrate into the genome of their host by recombination between the phage attachment site (*attP*) and target sequences in bacteria called *attB* (Campbell, 1992). As the *attP* site of most phage is immediately downstream of the integrase (Williams *et al*., 2013), single reads were manually examined to identify a read which contained the phage integrase and bacterial DNA. The analysis revealed that ΦCD9301 integrated in position 2 191 273 of the genome interrupting the *mprA_2* gene of *C. difficile*. This gene codes for a stage 0 sporulation protein A homolog.

The sequencing data was then mapped for the second time using the predicted genome of the CD9301*ΔtcdA_tcdB*L lysogen created in Benchling. This genome had the ΦCD9301 genome inserted into the position 2 191 273 of *C. difficileΔtcdB_tcdA* genome (figure 5.6). As expected sequencing reads aligned across the predicted integration site.

Figure 5.6: The integration site of the phage ΦCD9301 within the genome of *C. difficileΔtcdB_tcdA*L. This shows alignment of the sequencing reads across the predicted integration site of the phage ΦCD9301 (shown in red) within the bacterial DNA (shown in yellow). Phage integrated in the position 2 191 273 of bacterial genome interrupting *mprA_2* gene of *C. difficile,* encoding for the stage 0 sporulation protein A homolog.

To confirm the integration site experimentally, PCR was performed using the forward primer ZZ402, binding within the bacterial genome and the reverse primer ZZ403, binding within the integrase gene of the phage. The resulting PCR product of

expected size of 461 bp was sequenced confirming that ΦCD9301 integrated into the *mprA_2* gene of *C. difficileΔtcdB_tcdA*.

5.2.3 Using heterologous Cas9 as a lethal cargo

As previously mentioned in the introduction, several studies using exogenous Cas9 as an antimicrobial have been published to this date. In Chapter 4, several locations in the phage genome proved to be suitable for the insertion of cargo sequences up to 1666 bp. In this chapter, the use of the inducible RiboCas9 system as a phage lethal cargo was investigated. A RiboCas9 cassette was designed to target toxin genes of *C. difficile* and the lack of a homologous template in the phage should lead to the pathogen's inability to repair DNA double strand breaks (DSBs) caused by Cas9 and cell death (Figure 5.7).

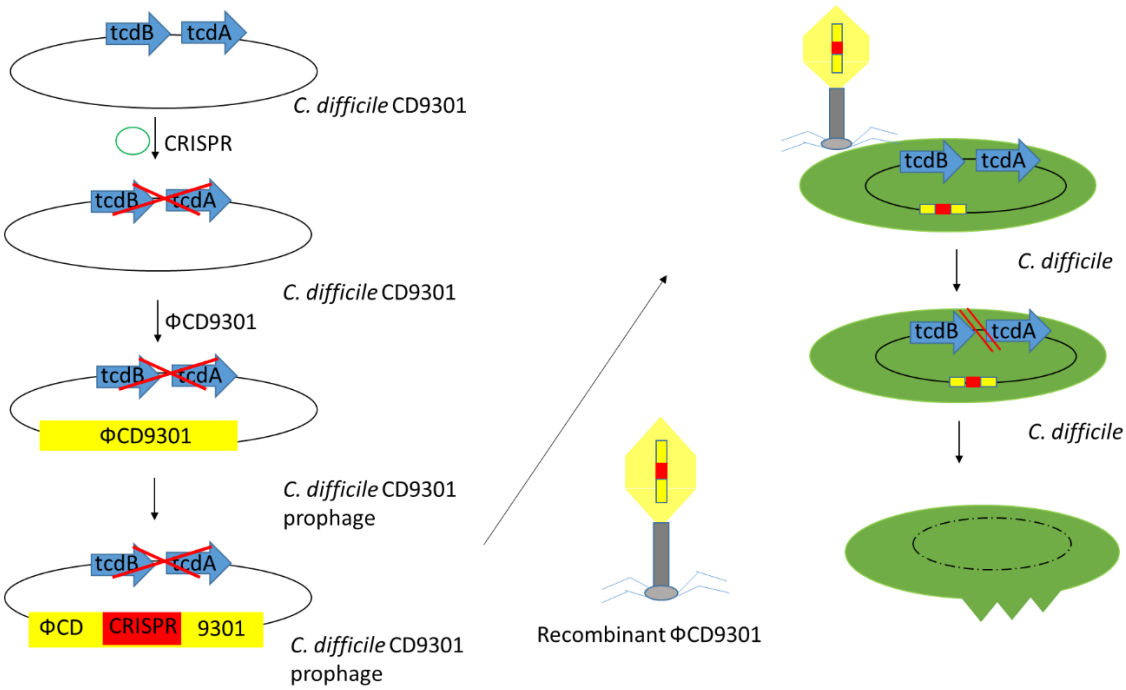

Figure 5.7: Schematic representation of the use of ΦCD9301 for the delivery of heterologous Cas9. In first step, *C. difficile* with deleted toxin genes is generated to serve as a manufacturing strain and ΦCD9301L lysogens are made using this strain. Following this RiboCas9 cargo targeting toxin genes is incorporated into the phage genome and the recombinant phage is induced using mitomycin C. Recombinant phage is then used to infect *C. difficile* WT and expression of Cas9 will cause DSB in toxin genes, leading to cell death due to the lack of homologous template.

Since the RiboCas9 tool was used as a cargo, allelic exchange vectors based on the *codA* counter selection marker were chosen as a method for the cargo knock-in into the phage genome, despite its low efficiency. The constructed plasmids were similar to those described in Chapter 4, which were generated in two steps (Figure 5.8). Firstly, homology arms targeting the locations 4A and 5B were amplified from ΦCD9301 gDNA and cloned into the *codA*-based vector pMTLSC7351 backbone using the *Xma*I and *Xho*I restriction sites resulting in plasmids pMTL_SC_HA4A, and pMTL_SC_HA5B. Plasmids were confirmed by colony PCR and sequence verified.

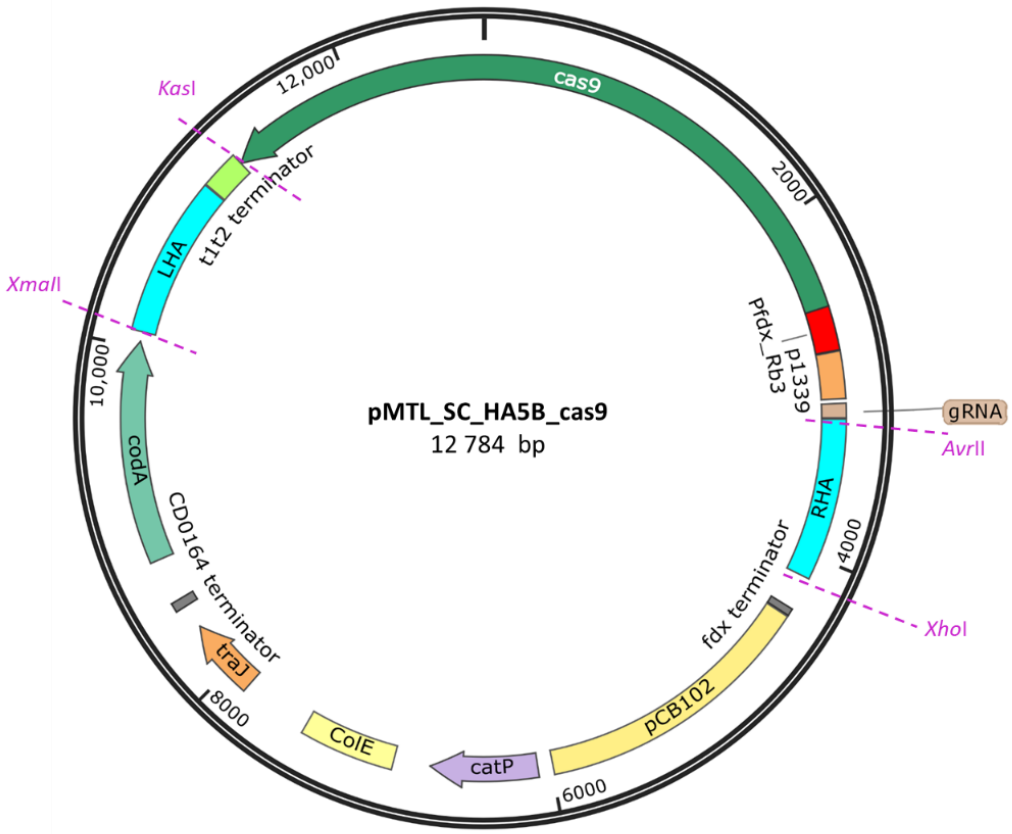

Figure 5.8: Schematic representation of AE plasmid for *cas9* knock-in. In first step, homology arms were cloned into the pMTLSC7315 backbone via *Xma*I and *Xho*I restriction sites. Secondly, cargo was cloned into the vector using *Avr*II and *Kas*I enzymes. Despite several attempts and optimising, all obtained plasmids were mutated in *cas9* gene.

Next, the RiboCas9 cargo was amplified from the plasmid pMTL_RCas_tcdBA3, previously used to knockout both *tcdB* and *tcdA* genes of the manufacturing strain. This plasmid carrying tcdBA3 sgRNA, proved to target *tcdB_tcdA* genes with the highest efficiency (Chapter 5.2.1). The T1T2 terminator was then spliced to the cargo

using SOE PCR and both were cloned into pMTL_SC_HA4A and pMTL_SC_HA5B (Figure 5.8). Unfortunately, the cloning was not successful despite several cloning attempts. As part of the trouble shooting, several *E. coli* cloning strains were used such as DH5α, Top10 or XL1-blue. The incubation temperature during transformation was reduced from 30°C to room temperature. In addition, sgRNA tcdBA3 was replaced with other candidates proven to work in section 5.2.1, whilst the use of low copy replicon backbone vectors with p15a Gram-negative replicon was also investigated. Moreover, homology arms and primers were re-designed and the HiFi cloning kit was used as a different cloning approach. In every case, *cas9* from isolated plasmids carried several mutations and plasmids could not be used for further work in this study.

At the time of cloning, M. Whittle successfully integrated RiboCas9 cargo into the *C. difficile* phage ΦCD1801, however the recombinant phage did not produce functional particles (Whittle, M. unpublished data). RiboCas9 cargo is 4 979 bp in size, which is almost 10% in the case of ΦCD9301 and 10% in ΦCD1801. It was assumed that the RiboCas9 cargo was too big for *C. difficile* phages isolated in our group. In light of the above mentioned difficulties, it was decided to redirect this project towards the use of the endogenous CRISPR system.

5.2.4 Characterisation of *C. difficile* M120 endogenous CRISPR

CRISPRfinder predicted that *C. difficile* CD9301 representing strain, *C. difficile* M120 (PCR ribotype RT078) carries 12 potential CRISPR arrays, 3 of which are questionable either due to a small number of direct repeats or because of their low sequence identity (less than 100%). All arrays appear to be distributed through the chromosome, with confirmed CRISPR arrays 2, 3, 5, 6, 7 being located on the sense strand and CRISPR arrays 8, 9, 10, 12 on the antisense strand (Figure 5.9B). None of the arrays are located in the prophage region. Within 9 confirmed CRISPR arrays of *C. difficile* M120, 168 spacers were identified with an average length of 36 bp. The amount of spacers within an array ranged between 3 in CRISPR array 10 to 44 in CRISPR array 6. The length of direct repeats ranged from 26 bp to 36 bp. The presence of three *cas* clusters type I-B was detected using CRISPRfinder, out of which 2 appear

to be complete (Figure 5.9A). All three *cas* operons are located on the antisense strand.

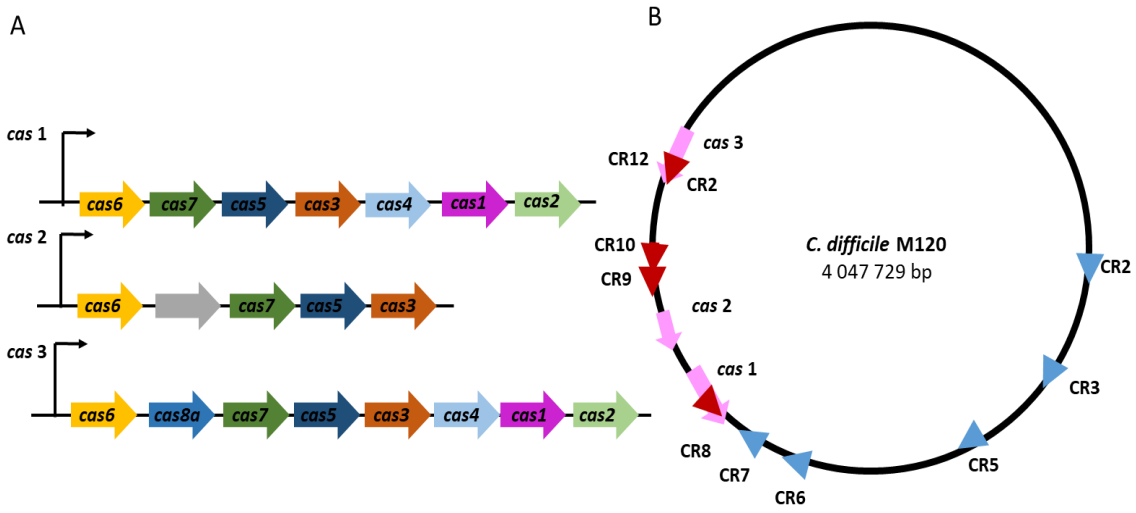

Figure 5.9: Schematic representation of the endogenous CRISPR system in *C. difficile* M120. **A**, Locations of predicted CRISPR arrays. Arrays are numbered according to the CRISPRfinder database and arrows indicate the positions of each array. The positions of each *cas* operon in the genome are indicated. **B**, Organisation of *cas* operons. *cas* 2 operon is partial. *cas* operon 1 and 3 are complete and associated with CRISPR arrays 12 and 8 respectively (edited from Boudry *et al*., 2015).

CRISPRTarget software was used to identify the spacers in CRISPR system of *C. difficile* M120. The analysis revealed that many *C. difficile* spacers correspond to published *Clostridioides* phages. For instance CRISPR array 8 carries spacers that match sequences of *C. difficile* phages ΦCDMH1, ΦCD119, ΦCDIF1296T or ΦCD211 and array 12 targets ΦCD211, ΦCDIF1296T, ΦCD481-1, ΦCDHM14, ΦMMP04, ΦCDHM13, ΦCDHM1, ΦCD111, ΦCD481-1, ΦCD506. Since the analysis is limited by the lack of available published genomes of *C. difficile* phages, it is predicted that the number of spacers correlating to these phage are much higher. Investigation also revealed that numerous spacers within the same array or between arrays target the same phage.

5.2.5 Development of the endogenous CRISPR array for *C. difficile* CD9301 genome editing

The native *pyrE* gene was chosen to evaluate the possibility of using an endogenous *C. difficile* CRISPR system to target bacterial genomes. This gene codes for orotate phosphoribosyltransferase, which is an enzyme, involved in *de novo* pyrimidine biosynthesis. Conversely, it can be used as a positive selection marker as it is essential for cell growth in media lacking exogenous pyrimidines and/or negative selection marker. Cells that lack *pyrE* can be selected on growth medium supplemented with 5-FOA, as this compound is highly toxic only to cells which contain this metabolic pathway (Ehsaan *et al*., 2016).

The CRISPR arrays 8 and 12 were the longest arrays associated with *cas* clusters and were therefore selected for the development of synthetic CRISPR mini-array. CRISPR array 8 contains 38 spacers and is associated with the *cas* operon 1 while CRISPR array 12 is associated with the *cas* gene set 3 and has 32 spacers. The repeat-spacer-repeat motif in synthetic mini-array 1, based on the *C. difficile* M120 CRISPR array 8 contains two 30 bp long repeats separated by a 36 bp spacer sequence targeting the gene of interest. Synthetic mini-array 2 mimics CRISPR array 12 and contains two 29 bp long repeats separated by a 36 bp spacer sequence.

Five endogenous CRISPR plasmids for *pyrE* knockout were designed by Terry Bilverstone (unpublished data) and contain the tetracycline inducible promoter system followed by the CRISPR mini-array. This system was chosen for this study as it was previously proven successful in re-directing endogenous CRISPR of *C. difficile* R20291 and 630Δ*ermB* by Maikova *et al*. (2019).

The tetracycline inducible promoter system contains two divergently orientated promoters, P_{tetR} and P_{tet} each with an overlapping *tet* operator sequence (*tetO*). While P_{tet} drives the expression of the gene of interest, P_{tetR} controls the expression of transcriptional repressor *tetR* that binds to the *tetO* operator and blocks both promoters. Adding tetracycline or anhydrous tetracycline (aTc) causes conformational changes in TetR and prevents its binding to the operator, relieving the repression (Dembek *et al*., 2015).

Three trinucleotide 5' PAMs CCA, CCT and TCA were selected for this study based on previous experimental and *in silico* work in *C. difficile* 630 and R20291 (Boudry *et al*., 2015). Several 36 bp long sequences in the *pyrE* gene associated with 5' PAM were selected by T. Bilverstone to test the re-directing of the endogenous CRISPR system (Table 5.2).

Table 5.2. List of *pyrE* spacers selected in this study.

	PAM	Spacer 5'-3'	CRISPR array
SP1	5'CCA	gaatatgctgctgaggtattaagtacagttgttgaa	CR_12
SP2	5'CCA	gctatgggtggagtaatagtttcttatgaattggga	CR_12
SP3	5'CCT	actactaagtctatatctaagtcttttatttgttca	CR_12
SP4	5'CCT	cttcttaactccattgtattgtcttttctctcagta	CR_12
SP5	5'TCA	ggaaaacacagtaatagatatgtacaatgtgcaaaa	CR_8

Plasmid pMTL83151 was used as a backbone to clone the endogenous CRISPR system as outlined in section 2.7.2. Initially, the tetracycline inducible system was cloned into the backbone using the *Bam*HI and *Not*I restriction enzymes. The resulting plasmid was sequenced and named pMTL83151_P_{tet}. Following this, homology arms flanking the *pyrE* gene of *C. difficile* M120 was amplified using genomic DNA and spliced together in a SOE PCR reaction. Next, 5 CRISPR mini-arrays were synthetized in a primer-dimer PCR reaction and separately cloned into the pMTL83151_P_{tet} plasmid together with the homology arms via the *Bam*HI and *Asc*I restriction sites. Plasmids were confirmed by colony PCR, sequenced and named pMTL_P_{tet}_pyrESP1, pMTL_P_{tet}_pyrESP2, pMTL_P_{tet}_pyrESP3, pMTL_P_{tet}_pyrESP4 and pMTL_P_{tet}_pyrESP5. Sanger sequencing revealed that the plasmid pMTL_P_{tet}_pyrESP3 carries a point mutation G->A in the left homology arm, however this mutation had no impact on the purpose of the experiment as proving the efficiency of the CRISPR mini-array was intended.

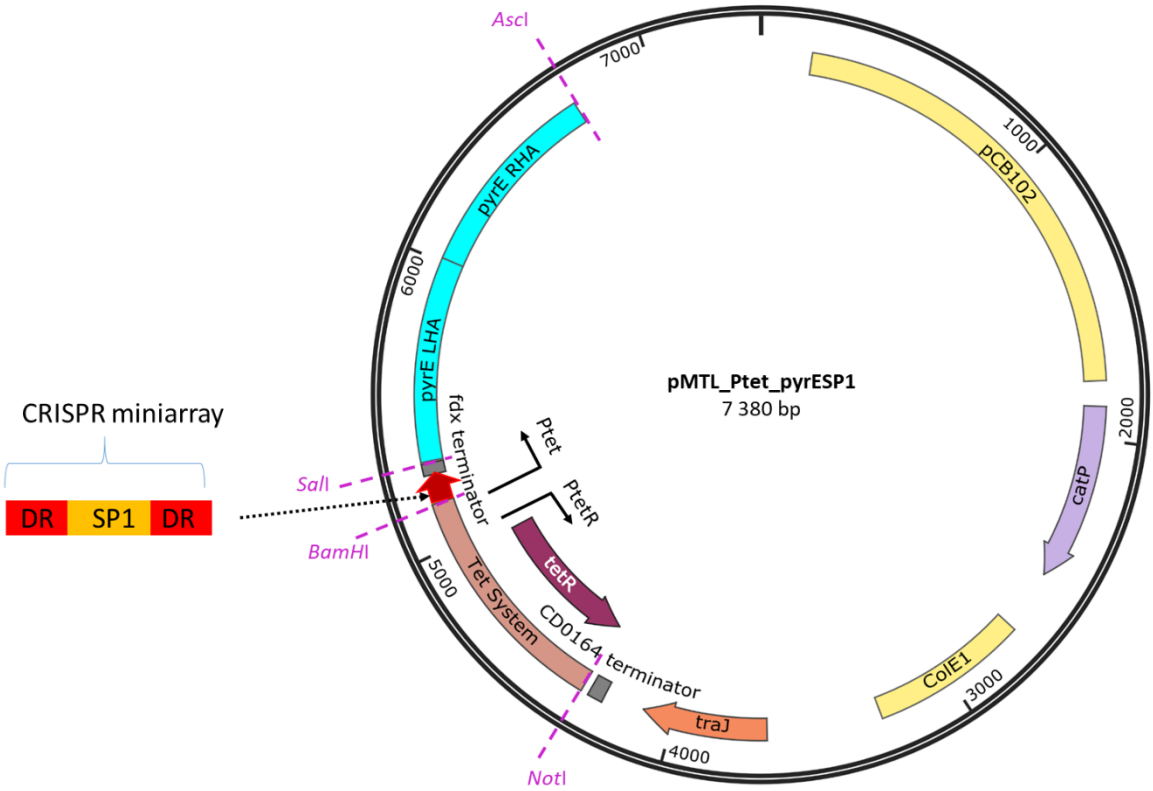

Figure 5.10: Schematic representation of CRISPR mini-array vector for *pyrE* knock out. The CRISPR array consisting of 2 direct repeats separated by the SP1 spacer targeting *pyrE* gene was cloned into the vector together with homology arms via *Bam*HI, *Sal*I and *Asc*I restriction sites. The expression of CRISPR module is under the control of tetracycline inducible system that is composed of P_{tet} promotor driving the expression of the array and P_{tetR} promoter controlling the expression of TetR transcription repressor.

Plasmids were transferred into *C. difficile* CD9301 via conjugal transfer from the *E. coli* donor and transconjugants were selected using thiamphenicol. To achieve the expression of the CRISPR mini-array, several transconjugants for each vector were induced with anhydrous tetracycline as outlined in section 2.6.16 and cells were plated on minimal media supplemented with 5-FOA to select for *ΔpyrE* mutants. Growth was observed for each induced transconjugant, suggesting that homologous recombination between the chromosome and plasmid borne homology arms had occurred. The mutant strains were confirmed by colony PCR using the primer pair M120_diag_F and M120_diag_R, designed to bind in the genome upstream and downstream of homology arms (Figure 5.11). Every screened colony obtained after conjugation with all five plasmids was shown to carry a mutated allele of the *pyrE*

gene. These results advocate that all 5 tested spacers based on CRISPR arrays 8 and 12 as well as the selected PAMs 5'CCA, 5'CCT and 5'TCA allow highly efficient genome editing of *C. difficile* M120. Mutant strains were subcultured three times to lose the plasmid and colonies sensitive to thiamphenicol were confirmed once more by sequencing the PCR amplified region of deletion.

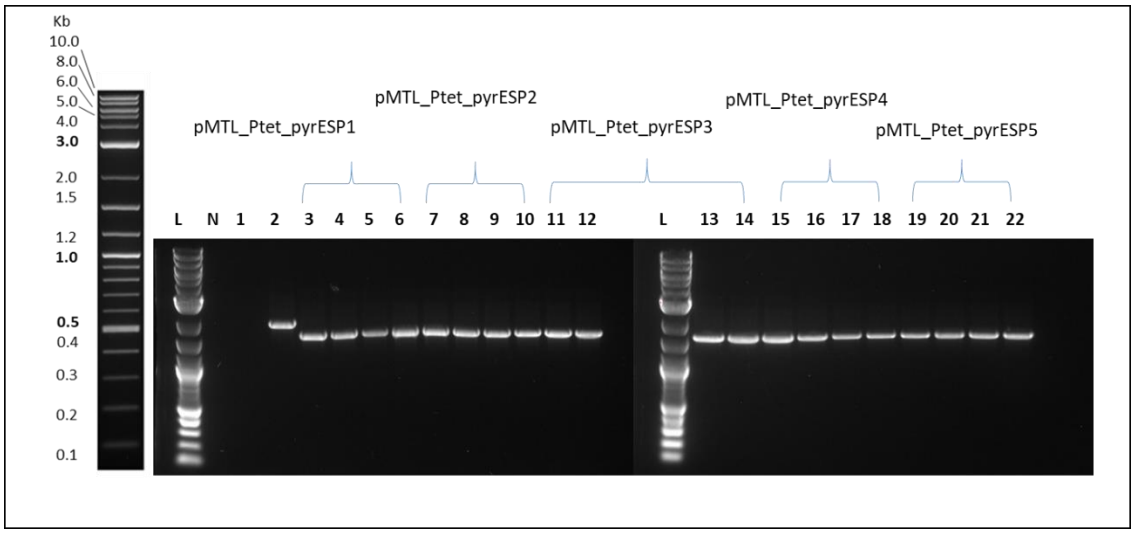

Figure 5.11: Colony PCR confirmation of *C. difficile* M120 *ΔpyrE* mutants. Colony PCR was performed using Green Taq DNA polymerase with standard thermocycling conditions. L: 2 log ladder, N: distilled H_2O, 1: pMTL_P_{tet}_pyrESP1, 2: *C. difficile* CD9301 genomic DNA, 3 -22 individual colonies following conjugation of pMTL_P_{tet}_pyrESP plasmids. Colonies with deleted *pyrE* gene were expected to give a band size of 1.687 kbp. Non-mutant strains were expected to give a band size of 1.994 kbp.

5.2.6 Re-directing *C. difficile* endogenous CRISPR system to autoimmunity

In the previous section, CRISPR arrays 8 and 12 of *C. difficile* M120 proved to be functional and successful for genome editing. This represents the first use of endogenous CRISPR reported for RT078 strains of *C. difficile*. Next, the possibility of redirecting these arrays towards *C. difficile* autoimmunity was investigated.

As in the initial attempt to use heterologous Cas9 as an antimicrobial agent, toxin genes were selected as endogenous CRISPR targets. Four sequences within the *tcdB* gene associated with 5'PAM (Table 5.3) were tested to obtain the most efficient self-targeting response.

Table 5.3. List of selected *C. difficile tcdB* spacers in this study.

	PAM	Spacer 5'-3'	CRISPR array
SP1	5'CCA	tttacatctccagaaaaattaatctcatgatgattt	CR_12
SP2	5'CCA	tggctcaggaggaacatatgcattatctctttctca	CR_12
SP3	5'CCT	tatacgatttaaaacaacaaaatgaattggatgatt	CR_12
SP4	5'TCA	tattctttccatggctcaggaggaacatatgcatta	CR_8

CRISPR mini-arrays consisting of a spacer flanked by the direct repeats were synthetized in a primer-dimer PCR reaction and cloned into the pMTL83151_P_{tet} backbone using the *Bam*HI and *Sal*I restriction enzymes. Plasmids were confirmed by colony PCR, sequenced and named pMTL_P_{tet}_tcdBASP1, pMT_P_{tet}_tcdBASP2, pMTL_P_{tet}_tcdBASP3, pMTL_P_{tet}_tcdBASP4. The lack of homologous template in the plasmid results in the cell inability to repair DSBs and ultimately, cell death.

CRISPR mini-array plasmids and pMTL83151_P_{tet} control were conjugated into *C. difficile* CD9301 and transconjugants were plated on media supplemented with thiamphenicol. A control conjugation with the pMTL83151_P_{tet} plasmid resulted in the efficiency of 6.85 CFU.ml^{-1}. The experiment was repeated 3 times with significantly reduced efficiency observed in the conjugation reaction with plasmids carrying CRISPR mini-array comparing to the control. Especially, the number of transconjugants in the conjugations with plasmids pMTL_P_{tet}_tcdBASP1 and pMTL_P_{tet}_tcdBASP3 decreased by more than 60% relative to empty plasmid control (Figure 5.12A). The smaller number of transconjugants observed after conjugation with all four tested plasmids is probably due to leakage of the inducible P_{tet} promoter. To fully induce the expression of the CRISPR mini-array, five transconjugant colonies for each plasmid were re-streaked on the media containing anhydrous tetracycline. While, all five colonies grew on a pMTL_P_{tet} control plate, no growth was observed on plates with tested plasmids, suggesting highly efficient autoimmune response of cells to the CRISPR mini-arrays.

To check the efficiency of self-targeting, all conjugation reactions were repeated and transconjugants were plated on media supplemented with thiamphenicol and inducer. The experiment was repeated 3 times with significant reduction of number of transconjugants being observed in case of all plasmids comparing to the control (Figure 5.12B). No colony was observed when plasmids pMTL_P_{tet}_tcdBASP1 and pMTL_P_{tet}_tcdBASP3 were transferred into *C. difficile*, suggesting these arrays are 100% efficient in self-targeting. Both miniarrays have repeat-spacer-repeat motifs based on the *C. difficile* M120 CRISPR array 12 and target the *tcdB* gene.

Since all CRISPR arrays target toxin genes of *C. difficile*, plasmids were transferred into the *C. difficileΔtcdBA* strain to investigate off-target Cas3 cleavage. No significant difference was observed when comparing the conjugation efficiency between the control groups and tested plasmids (Figure 5.12A). A slight decrease in number of transconjugants was observed when pMTL_P_{tet}_tcdBASP2 and pMTL_P_{tet}_tcdBASP4 groups were plated on media containing inducer (Figure 5.12B). These results may indicate that plasmid borne crRNAs expressed from these two CRISPR arrays may bind elsewhere in the genome with low efficiency and cause DSB.

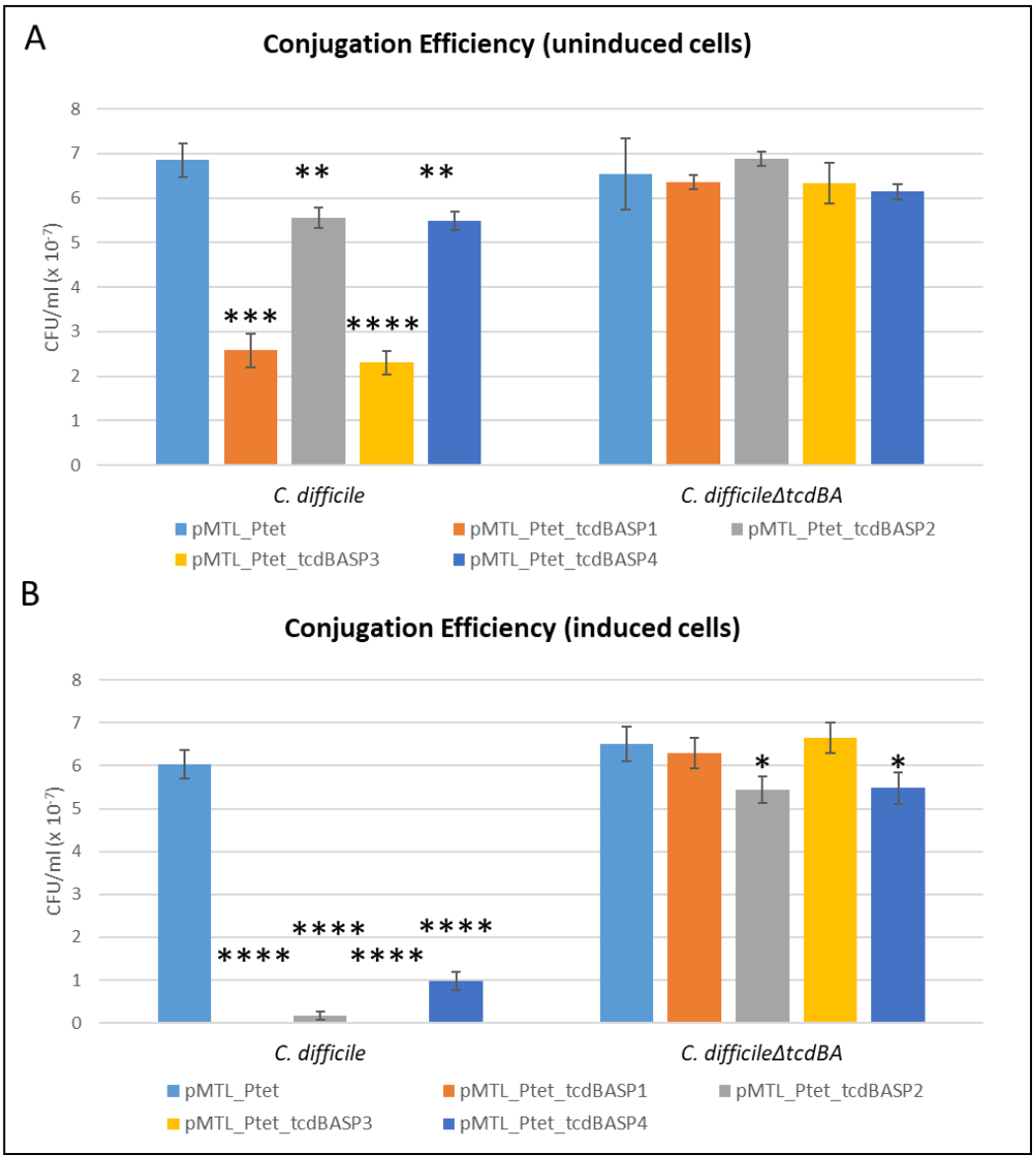

Figure 5.12: Comparison of conjugation efficiency. *C. difficile* CD9301 and *C. difficile* CD9301Δ*tcdBA* were conjugated with plasmids containing a CRISPR mini-array under the control of P_{tet} inducible system. Transconjugants were grown on **A.** BHIS media supplemented with thiamphenicol or **B.** BHIS media supplemented with thiamphenicol and anhydrous tetracycline. Error bars represent standard deviations calculated from 3 conjugations. Asterisks indicate statistically significant p values for *p<0.05, **p<0.01 ***p≤0.0002, ****p≤0.0001 calculated with unpaired two-tailed Students's *t*-test. There was a significant difference in efficiency between control plasmid group and cells carrying a CRISPR mini-array plasmids (p= 0.0002, p= 0.0069, p= <0.0001 and p= 0.0053) in uninduced *C. difficile* WT. The difference was more significant after adding the inducer with p<0.0001 when comparing pMTL83151_P_{tet} to all four tested plasmid groups. Even though efficiency values were observed to be similar among all of uninduced *C.difficileΔtcdBA* groups, a slight decrease in number of colonies was observed in case of pMTL_P_{tet}_tcdBASP2 and pMTL_P_{tet}_tcdBASP4 transconjugants after the induction (p=0.0221 and p=0.0307) suggesting off-target genome binding of crRNA expressed from these plasmids.

5.2.7 Construction of recombinant phages carrying endogenous CRISPR array as a phage lethal cargo

In the previous section, the endogenous CRISPR system of *C. difficile* M120 was characterised and successfully hijacked to autoimmunity. Endogenous CRISPR mini-arrays tcdBASP1 and tcdBASP3 targeting toxin genes of *C. difficile* were therefore selected as antimicrobial candidates for ΦCD9301 phage delivery as they exhibited 100% efficient in their self-targeting response and no off-target cleavage.

Individual CRISPR mini-arrays were delivered into the CD9301*ΔtcdB_tcdA*L lysogen using the RiboCas9 system. The RiboCas9 plasmids targeting locations 4A, 4B, 5A and 5B in ΦCD9301 genome were cloned in section 4.2.1.2 and named pMTL_RC4A, pMTL_RC4B, pMTL_RC5A and pMTL_RC5B.

Cargos tcdBASP1 and tcdBASP3 were amplified together with the aTc inducible promoter system from plasmids pMTL_P_{tet}_tcdBASP1 and pMTL_P_{tet}_tcdBASP3 and cloned into the individual pMTL_RC backbone plasmids via the *Kas*I/*Avr*II restriction sites as described in section 2.7.2. The cloning step proved to be challenging due to the occurrence of recombination events between the two direct repeats within the CRISPR mini-array cassette. Despite several attempts, the cloning of plasmids targeting location 5B was unsuccessful. The remaining plasmids pMTL_RC4A_tcdBASP1, pMTL_RC4A_tcdBASP3, pMTL_RC4B_tcdBASP1, pMTL_RC4B_tcdBASP3, pMTL_RC5A_tcdBASP1 and pMTL_RC5A_tcdBASP3 were confirmed by colony PCR and Sanger sequencing.

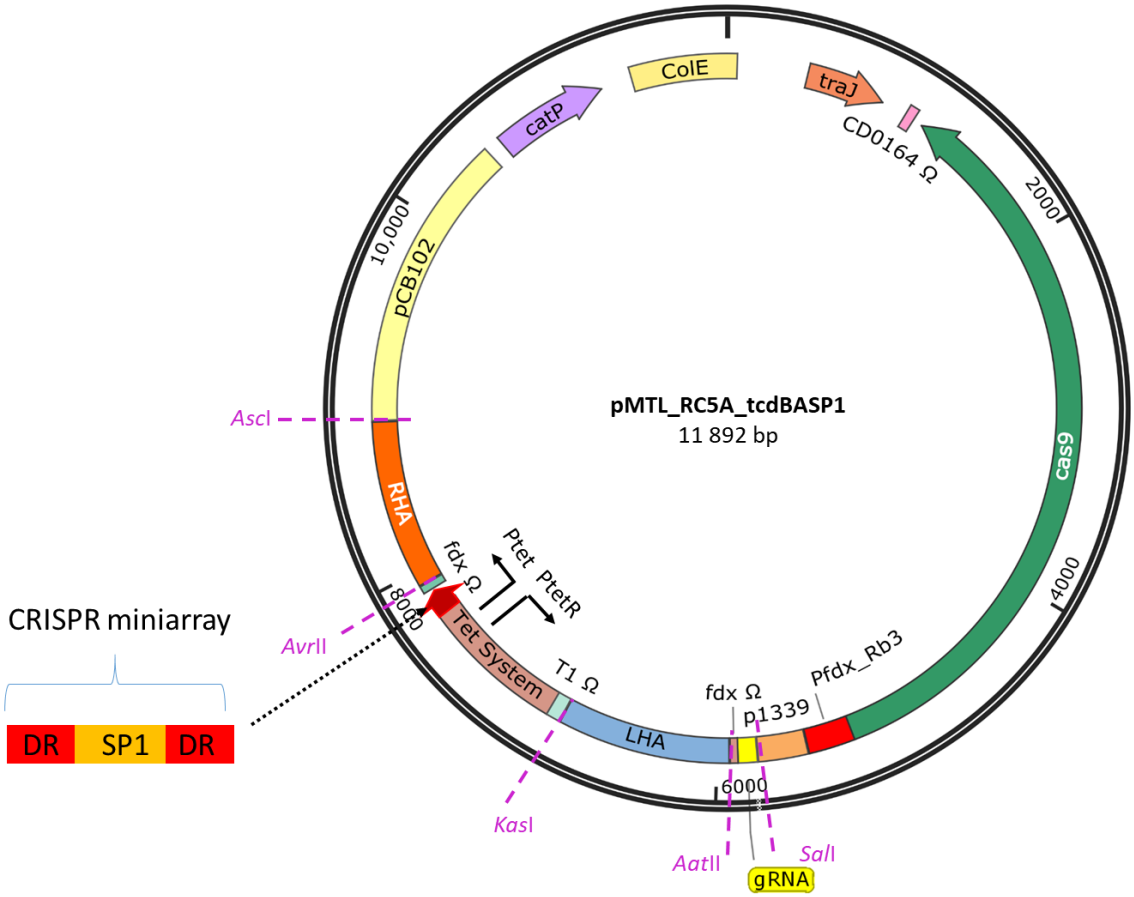

Figure 5.13: Schematic representation of the RiboCas9 plasmid for CRISPR miniarray knock-in. The plasmid was constructed in two steps. In the first step, homology arms and seed gRNA sequence were cloned into the pRECas1-IIE-Cdif backbone using *Sal*I, *Aat*II and *Asc*I restriction enzymes. In the second step the aTc inducible promoter was cloned together with CRISPR mini-array into the vector via *Kas*I and *Avr*II sites.

Plasmids were transferred into the *C. difficile* lysogen CD9301_*ΔtcdB_tcdA*L from their respective *E. coli* NEB express donor cells via conjugative transfer and transconjugants were selected on media supplemented with thiamphenicol. Following this, Cas9 was induced using theophylline and the mutant colonies were confirmed by colony PCR (Figure 5.14). PCR was performed using the forward primers ZZ280, ZZ282, ZZ284 and the reverse primers ZZ281, ZZ283 and ZZ285 for knock-in mutants generated with plasmids targeting positions 4A, 4B and 5A, respectively. Interestingly, despite the previous failed attempts to knock-in *ermB* and *sasp* cargos into the location 5A, conjugation of pMTL_RC5A_tcdBASP1 and pMTL_RC5A_tcdBASP3 plasmids into the lysogen resulted in mutant colonies.

Mutants were sub-cultured to lose the plasmid and Sanger sequencing further confirmed thiamphenicol sensitive colonies. Sequencing analysis revealed that the lysogens CD9301_*ΔtcdB_tcdA*L:4B_tcdBASP1 and CD9301_*ΔtcdB_tcdA*L:4B_tcdBASP3 lacked the spacer sequence, due to the recombination event between the direct repeats. Despite several attempts, the generation of mutant lysogens in position 4B were unsuccessful. The remaining mutant lysogens with the correct sequence were named CD9301*ΔtcdB_tcdA*L:4A_tcdBASP1, CD9301*ΔtcdB_tcdA*L:4A_tcdBASP3, CD9301*ΔtcdB_tcdA*L:5A_tcdBASP1 and CD9301*ΔtcdB_tcdA*L:5A_tcdBASP3 and stored as glycerol stocks at - 80°C.

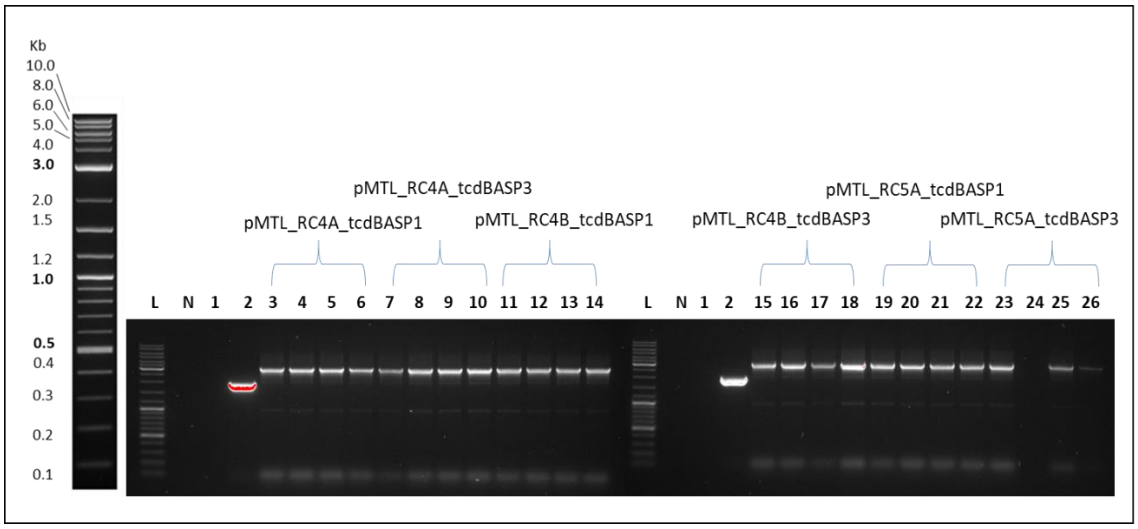

Figure 5.14: Colony PCR screening for CRISPR array knock-in into the genome of ΦCD9301 prophage. Colony PCR was performed using Green Taq DNA polymerase with standard thermocycling conditions. L: 2 log ladder, N: distilled H_2O, 1: pMTL_RC4A_tcdBASP1, 2: ΦCD9301 genomic DNA, 3-26: colonies obtained following conjugation of individual pMTL_RC_tcdBASP plasmids. Colonies with an integrated endogenous CRISPR array were expected to give a band size of 2.997 kbp. Non-mutant strains were expected to give a band size of 1.94 kbp.

All four lysogens with integrated CRISPR arrays were induced with mitomycin C to produce phage particles. The following day, a plaque assay was performed with the induced phages and *C. difficile* CD9301 to see whether they can package additional genetic sequence. Plaques were present only in case of lysogens

CD9301*ΔtcdB_tcdA*L:5A_tcdBASP1 and CD9301*ΔtcdB_tcdA*L:5A_tcdBASP3 (Table 5.2).

Table 5.4. Titre of the phage induced from mutant lysogens after mitomycin C treatment.

Lysogen	Titre of induced phage (PFU x ml^{-1})	Immunity to ΦCD9301 infection
CD9301*ΔtcdB_tcdA*L	1 x 10^3	Yes
CD9301L*ΔtcdB_tcdA*L: 5A_tcdBASP1	8.5 x 10^2	Yes
CD9301L *ΔtcdB_tcdA*L: 5A_tcdBASP3	7.2 x 10^2	Yes

Following the mitomycin C induction of recombinant phage, plaques were picked and purified three times. The phage titre was increased by harvesting the 0.2% top agar layer from plate followed by a plaque assay. The process was repeated until the high phage titre ≥10^8 PFU x ml^{-1} was obtained. Once the recombinant phages reached the required concentration, genomic DNA was isolated. Mutant phage were once again confirmed via colony PCR using the primer set ZZ284/ZZ285 and Sanger sequencing. Phage stocks named ΦCD9301:5A_tcdBASP1 and ΦCD9301:5A_tcdBASP3 were then stored as filtered sterilised supernatant in BHIS broth at 4°C as well as lysogens stored at -80°C.

5.2.8 Characterisation of recombinant phages carrying enCRISPR as a lethal cargo

The plaques formed by ΦCD9301:5A_tcdBASP1 and ΦCD9301:5A_tcdBASP3 phage appeared to be small and hazy with approximately 1 mm diameter in size (Figure 5.15).

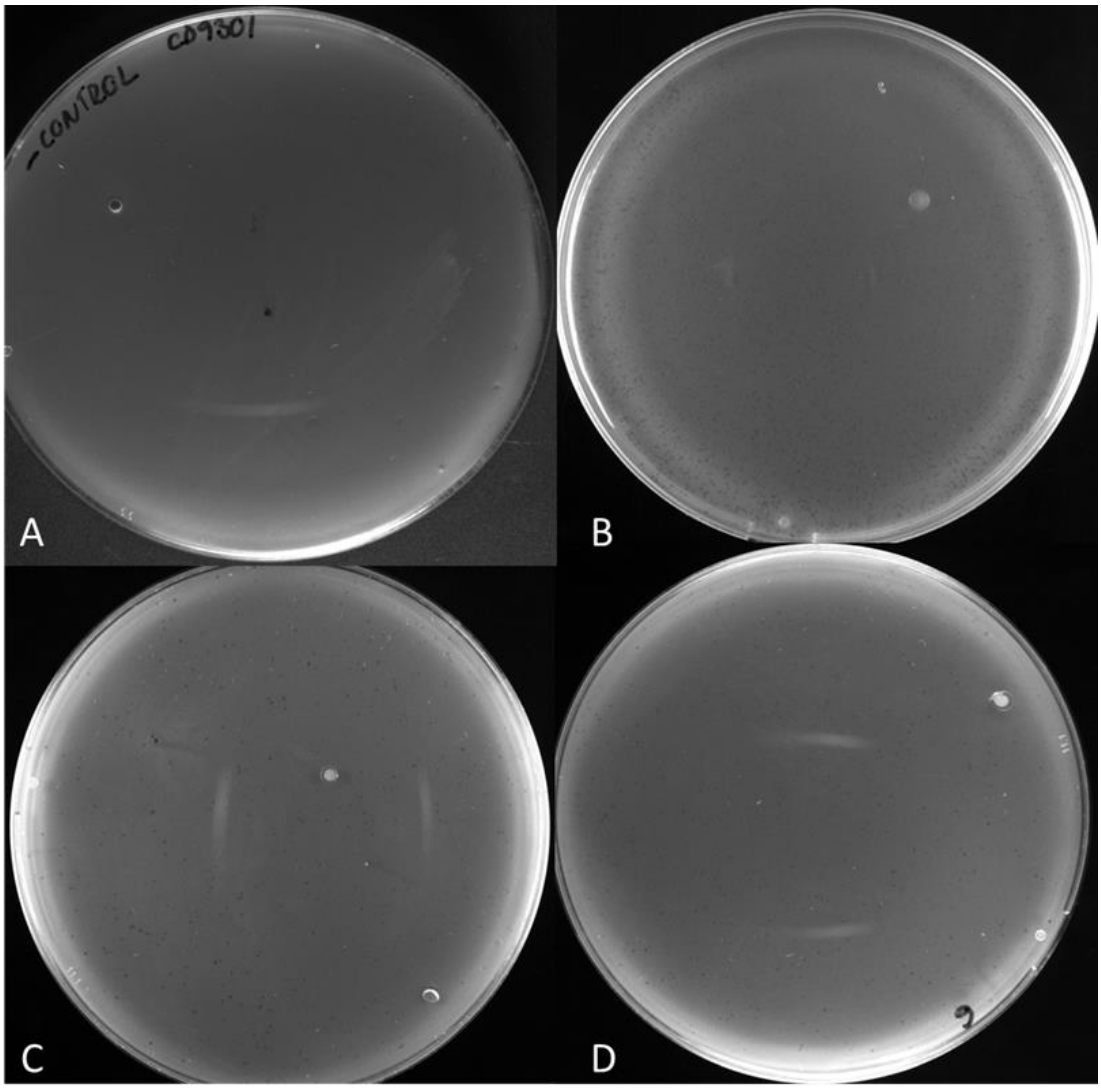

Figure 5.15: Plaques of isolated phages. A. Negative control, *C. difficile* CD9301 lawn grown without the presence of the phage, **B.** ΦCD9301 phage stock infecting *C. difficile* CD9301, **C**. Modified phage ΦCD9301:5A_tcdBASP1 stock infecting *C. difficile* CD9301, **D**. Modified phage ΦCD9301:5A_RtcdBASP3 stock infecting *C. difficile* CD9301.

To confirm that the genetic modification of the recombinant phage did not affect their host range, plaque assays were performed with modified phage against several *C. difficile* strains. The host range analysis revealed that both recombinant phage ΦCD9301:5A_tcdBASP1 and ΦCD9301:5A_tcdBASP3 are, as with the WT phage ΦCD9301, able to infect the same 24/100 *C. difficile* isolates across 6 different ribotypes RT078, RT002, RT014, RT021, RT015, RT023. This data confirm that the insertion of the enCRISPR cargo into the location 5A of ΦCD9301 genome does not have a negative impact on its host range.

Recombinant phages were also analysed by genomic sequencing to prove that their DNA does not carry any undesirable mutations possibly affecting their phenotypes. All genomes were sequenced using the Illumina MiSeq 500bp V2 SBS sequencing service. Sequencing reads of both mutant phage with an average read length of 250 bp were mapped to the ΦCD9301 reference genome using CLC Genomics Workbench 20. Firstly, the analysis confirmed that enCRISPR cargos were inserted in the location 5A. Secondly, no SNV was detected in the genomes of ΦCD9301:5A_tcdBASP1 or ΦCD9301:5A_tcdBASP3 when their sequencing reads were mapped to ΦCD9301, proving their suitability for further work in this study.

To understand the inability of CD9301*ΔtcdB_tcdA*L:4A_tcdBASP1, and CD9301*ΔtcdB_tcdA*L:4A_tcdBASP3 to form functional phage particles, their genome were also analysed. The analysis showed there were no SNV or genome alterations present that could explain the failure to induce functional phage particles. This data suggest that this failure is most probably caused by the choice of cargo location.

5.2.9 *In vitro* activity of enCRISPR recombinant phages against *C. difficile*

The activity of both recombinant phage carrying the enCRISPR array was measured *in vitro* to see whether they outperform the WT ΦCD9301 in the treatment of *C. difficile*. The activity was measured together with the appropriate controls using *in vitro* CFU reduction experiments as described by Selle *et al*. (2020).

Briefly, overnight cultures of the *C. difficile* indicator strain was subcultured 1:100 into BHIS and incubated to an OD_{600nm} of 0.2. Phage stocks were mixed 1:1 with *C. difficile* and induced with anhydrous tetracycline (0.25 µM). Cultures were grown for 22 hours and their OD_{600nm} as well as the number of *C. difficile* colonies grown on the BHIS plate containing the inducer were monitored over time by means of a manual growth curve.

No significant difference was observed when comparing the WT ΦCD9301 to the recombinant phage carrying enCRISPR array (Figure 5.16). All phage displayed similar activity towards *C. difficile*, reducing the OD_{600nm} and the number of CFU approximately 3 times and by 1.5 log respectively for up to 6 hours. The inhibitory effect of phage on *C. difficile* was not as apparent at the endpoint of experiment, with

all phage reducing the number of *C. difficile* CFU by almost 1 log, possibly due to the bacteria developing resistance towards the phage. The induction time as well as different concentrations of the inducer (0 µM, 0.1 µM, 0.25 µM, and 0.5 µM) in liquid cultures as well as plates were investigated in an attempt to optimise the expression of the lethal cargo. Despite the optimisations, no significant difference in the activity of mutant phages was observed comparing to the WT.

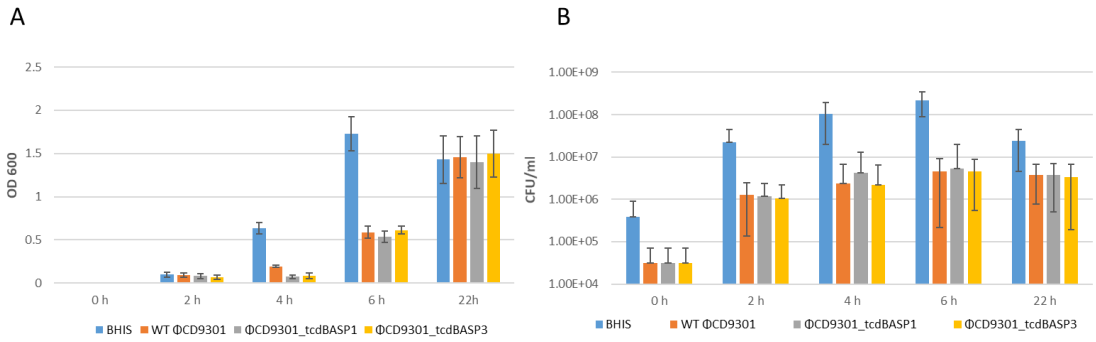

Figure 5.16: Comparison of the phage activity *in vitro*. *C. difficile* cultures were infected with WT phage or mutant phage carrying the enCRISPR mini-array under the control of the aTc inducible promoter system. As a negative control, *C. difficile* was mixed with BHIS broth. All four cultures were induced with anhydrous tetracycline (0.25 µM) and the activity was monitored over the time by OD_{600nm} measurements as well as number of *C. difficile* colonies grown on the BHIS plate containing the inducer. Experiment were repeated five times to calculate standard deviations represented by error bars. No significant difference was observed in the activity of mutant phage comparing to the WT phage. An approximate 3-fold reduction in the OD_{600nm} as well as 1.5 log reduction of CFU/ml was observed in cultures treated with phage for up to 6 hours.

5.3 Discussion and summary

CRISPR/Cas systems are a powerful defence mechanism that naturally evolved to protect bacteria from invasive genetic elements. With the advance in molecular techniques, CRISPR/Cas systems can be reprogrammed to target bacterial chromosomes, allowing genome manipulation or cell death. Although CRISPR/Cas has been shown in principle to be a powerful programmable antimicrobial, the main challenge for its use to eradicate bacterial infection is its delivery, which requires high specificity and efficacy. Thus, several studies have successfully employed phagemids and phage particles as vectors to deliver CRISPR/Cas components to bacterial populations (Bikard *et al*., 2014; Citorik *et al*., 2014; Park *et al*., 2017; Lam *et al*., 2020).

In this chapter, we tested the concept of using the temperate phage ΦCD9301 to deliver lethal CRISPR/Cas cargo directed at toxin genes specific to *C. difficile*. The RiboCas9 genome editing tool was chosen to generate *C. difficile* strains with deleted toxin genes and lysogens of these strains were made to enable genome editing of the phage DNA without the risk of self-targeting of the bacterial manufacturing strain.

Initially, the strategy of using heterologous Cas9 as a lethal phage cargo was investigated as this strategy has been sucesfully employed for phages infecting *S.aureus* (Citorik *et al*., 2014; Park *et al*., 2017). AE vectors based on *codA* were designed to carry RiboCas9 cassette targeting toxin genes and homology arms representing positions in phage genome previously shown to be suitable for cargo knock-in. Despite several attempts and thorough optimisations, cloning of the plasmids has proven to be difficult. As previously shown, it is expected that non-specific binding of Cas9 at off-target positions could have led to alleviated toxicity in the cloning host (Pattanayak *et. al*., 2013; Hsu *et al*., 2013). Moreover, the RiboCas9 cargo is 4 979 bp in size, which is almost 10% in case of ΦCD9301. It is assumed that that the RiboCas9 cargo is too large for the phage presently available in the SBRC. Because of these results, and the difficulties encountered in cloning, the project was redirected to investigate the use of endogenous CRISPR-Cas system of *C. difficile*.

The analysis of the genome of *C. difficile* M120 propagating strain has revealed that it contains an active bacterial CRISPR system of subtype I-B. Similar to other *C. difficile* strains, M120 has been shown to carry unusually high number of CRISPR arrays. Genome analysis revealed the presence of 168 spacers within nine confirmed CRISPR arrays, which is more than *C. difficile* 630 and R20291.

Several spacers within or outside of the same array targets the same phage, possibly reflecting the repeated infections of *C. difficile* by this phage. This mechanism probably increase the efficiency of CRISPR defence and suggest a phage mutational mechanism to escape it. Some spacers have been shown to target highly conserved genes involved in phage replication and assembly. According to Bourdy *et al*. (2015), this strategy ensures economical and efficient defence against several phage.

Even though, some previous studies have found *C. difficile* CRISPR arrays to be localised and actively expressed in prophages (Hargreaves *et al*., 2014), PHAST analysis revealed that none of the CRISPR arrays of M120 are located within prophage regions.

The presence of three *cas* clusters on the antisense strand was identified, of which two appear to be complete. Similarly like most of the analysed *C. difficile* strains to date, the partial *cas2* operon composed of *cas3*, *cas5*, *cas6*, and *cas7* genes has been detected in *C. difficile* M120, suggesting their possible accessory role in CRISPR function as well as less efficient interference driven by the main *cas* operons (Boudry *et al*., 2015).

In the present work, the endogenous CRISPR-Cas system was successfully used for genome editing of *C. difficile* M120. The *pyrE* gene encoding for orotate phosphoribosyltransferase was chosen to evaluate the possibility of using an endogenous *C. difficile* CRISPR-Cas to target the bacterial genome, as cells that lack *pyrE* can be easily selected on growth medium supplemented with 5-FOA (Ehsaan *et al*., 2016).

In *C. difficile* 630 and R20291, the longest CRISPR array is associated with the complete functional *cas* cluster. According to Boudry *et al*. (2015), the longest array associated with the *cas* operon is most probably an active CRISPR functional unit in bacterial genome. In M120, CRISPR array 6 is the array with the highest number of

spacers, however, this array is not associated with the *cas* gene set. Instead, the second largest CRISPR array 8, with 38 spacers, is associated with complete *cas1* cluster as well as CRISPR array 12, which has 32 spacers an has been shown to be associated with the *cas3* cluster. Thus, the possibility of redirecting these arrays towards *C. difficile* was investigated.

Vectors were designed to carry CRISPR mini-arrays consisting of a spacer flanked by two direct repeats under the control of the aTc inducible promoter system and homology arms for *pyrE* knockout. Five 36 bp long spacers associated with a 5'CCA, 5'CCT or 5'TCA PAM inside the *pyrE* gene were tested. This strategy enabled the successfully isolation of *ΔpyrE* mutant strains of *C. difficile* M120. Mutants were generated in case of every mini-array with 100% efficiency after the induction with anhydrous tetracycline, proving that that CRISPR mini-array based on native *C. difficile* M120 array 8 and 12 are actively transcribed and highly efficient for genome editing. This result confirms that plasmid borne crRNA expressed from synthetic CRISPR array can re-direct native CRISPR system of *C. difficile* M120 to recognise and target protospacers in its own genome and that 5'PAMs (CCT, TCA, CCA) predicted by Bourdy *et al*. (2015) are being recognised by M120 *cas* machinery.

Redirecting the endogenous CRISPR-Cas machinery provides considerable benefits over the techniques routinely used for genome editing of *C. difficile*. First of all, mini-array vectors do not carry genes encoding heterologous Cas proteins which may have toxic effects on both cloning host or target bacterium leading to undesirable mutations. Moreover, comparing to other *Clostridiodes* genome manipulation techniques such as ClosTron or the use of counter selection markers such as *codA* or *pyrE*, this method involves less plating steps and is therefore more time efficient.

To repurpose the endogenous CRISPR/Cas system for autoimmunity, vectors carrying targeting mini-array under the aTc inducible promoter without any homologous regions were designed. As in the initial attempt to use heterologous Cas9 as an antimicrobial agent, toxin genes were selected to be targeted by endogenous Cas endonuclease to ensure the elimination of virulent *C. difficile* strains. Four sequences within the *tcdB* gene associated with previously validated PAMs were tested to obtain the most efficient self-targeting response. To check for the off-target cleavage of

designed spacers, CRISPR mini-arrays were transferred into both *C. difficile* WT and *C. difficileΔtcdBA* manufacturing strains for comparison. A 100% efficient self-targeting response was detected with mini-array plasmids pMTL83151_P_{tet}_tcdBASP1 and pMTL83151_P_{tet}_tcdBASP3. Both mini-arrays have repeat-spacer-repeat motif based on *C. difficile* M120 CRISPR array 12 and target the *tcdB* gene. Furthermore, when conjugated in the *C. difficile*Δ*tcdB_tcdA* strain these plasmids exhibited no off-target cleavage activity and were therefore selected as antimicrobial candidates for phage delivery.

Despite the initial problems with recombination between direct repeats during the cloning or conjugation, both enCRISPR mini-arrays were successfully delivered into the ΦCD9301L lysogen locations 4B and 5A. In common with the *sasp* experiments, lysogens carrying enCRISPR in positions 4B did not form phage particles. In the previous chapter, we concluded that the position 4 was not suitable for phage knock-in probably due to the toxic nature of SASP. Similar observations in this study, however, led us to conclude that the positions 4A or 4B are not suitable for cargo insertion, despite the successful generation of *ermB* recombinant phage in these locations.

Following the mitomycin C induction, genome analysis of recombinant phage ΦCD9301:5A_tcdBASP1 and ΦCD9301:5A_tcdBASP3 showed that they carry no undesirable mutations. Host range testing also revealed that the mutant phage were capable of infecting the same 24/100 *C. difficile* as ΦCD9301, demonstrating that the cargo does not have a negative impact on the host range.

The activity of enCRISPR phage was measured *in vitro* to investigate their impact on *C. difficile* compared to the parental ΦCD9301. Unfortunately, no difference was observed in the level of *C. difficile* infection when comparing recombinant phage to the WT phage. All phage reduced the OD and number of *C. difficile* CFU at a similar level. Approximately 3-fold reduction of the OD_{600nm} as well as 1.5 log reduction of CFU/ml was observed in cultures treated with phage for up to 6 hours. As seen in the previous chapter, the phage inhibition of *C. difficile* growth seemed to decrease with the time as no significant difference in OD_{600nm} was observed at the endpoint of

experiment, although 1-log reduction in CFU was still detected on plates in *C. difficile* groups treated with phage compared to the control. This is most probably due to increased amount of lysogens in the culture and because of various mechanism of bacterial resistance towards phage infection such as the receptor-mediated resistance often seen in monophage treatments (Sekulovic *et al*., 2015).

A similar activity of mutant phage comparing to the WT was unexpected, as both enCRISPR mini-arrays had previously been shown to mediate 100% efficient self-targeting with no off-target cleavage activity, when delivered to *C. difficile* on plasmid. This may be due to higher copies of the cargo delivered by the high copy number plasmids versus single copy of gene delivered by the phage integrated into the *C. difficile* genome. At the time of writing, recombinant mycobacteriophage carrying artificial miniarray re-directing endogenous III-A CRISPR-Cas system in *Mycobacterium tuberculosis* H37Ra was generated by Jiang *et al*., 2021. Similarly, like our study, the recombinant phage did not have a higher lethality to *M. tuberculosis* H37Ra than the WT phage. Authors of this study found that endogenous CRISPR-Cas system in *M. tuberculosis* H37Ra could not target the plasmid effectively; however, the mRNA level of the target genes were significantly decreased. Jiang *et al*., (2021) assumed that the low efficiency of the recombinant mycobacteriophage may be caused by the low nucleic acid interference activity of CRISPR-Cas system in *M. tuberculosis* H37Ra. As our enCRISPR mini-arrays have shown 100% efficiency when delivered on plasmid, we postulate that the problem may lay in the amount of lethal cargo copies delivered via the phage rather than the interference activity of *C. difficile* endogeneous CRISPR array.

Selle *et al*. (2020) generated recombinant bacteriophage ΦCD24-2 expressing bacterial self-targeting CRISPR RNAs. In this study, authors proved that recombinant ΦCD24-2 is significantly more effective than its WT parent phage at killing *C. difficile* both *in vitr*o and in a mouse model. Their WT ΦCD24-2 infecting *C. difficile* strain CD19 exhibited an approximately 1-log reduction in the numbers of CFU after 6 h of incubation *in vitro* but recovered rapidly. In contrast, recombinant phage performed 2 log better than WT after 2 hours of infection and 1 log better after 6 hours and at the endpoint of infection, decreasing the recovery of the culture over 22 hours.

Unlike our study, Selle *et al*. (2020) used an endogenous CRISPR miniarray that was preceded by the native CRISPR leader sequence from *C. difficile* 630.

Yosef *et al*., (2015) used T7 phage to deliver CRISPR-Cas system targeting antibiotic resistance-conferring plasmids in *E.coli.* In this study, the engineered phage carrying CRISPR arrays under the control of the strong T7 promoter significantly outperformed WT phage. Due to the lack of time, we used enCRISPR mini-array together with the anhydrous tetracycline inducible promotor system that has been shown to be far from ideal (Muh *et al.*, 2019). It is therefore possible that the aTc inducible promotor system used in our study did not provide sufficient expression levels of the enCRISPR mini-array and should be replaced with a strong constitutive promoter or native leader sequence from *C. difficile* endogenous CRISPR as these strategies were proven successful by Selle *et al*., 2020 and Yosef *et al*., 2015.

To conclude, we have generated two recombinant *C. difficile* phages carrying lethal enCRISPR array redirecting *C. difficile* endogenous CRISPR to autoimmunity. Unfortunately, neither of the recombinant phage exhibit enhanced activity against *C. difficile in vitro* when compared to the WT phage. This is most probably due to the insufficient expression levels of the lethal cargo, as the system has proven to be 100% efficient when delivered by plasmid.

5.4 Chapter summary

- *C. difficile* strains with deleted toxin genes were constructed, to enable genome editing of the phage DNA without the risk of self-targeting of manufacturing bacterial strain.
- An endogenous CRISPR-Cas system of *C. difficile* M120 was characterised and successfully re-directed to autoimmunity.
- Two enCRISPR mini-arrays were individually integrated together with a tetracycline inducible promoter system into the ΦCD9301 genome, however, no change in the phage activity was observed *in vitro* comparing to the WT phage.

Chapter six:

Conclusions and Future Prospects

6. Conclusions and Future Prospects

C. difficile is an important nosocomial pathogen and the rise of CDI epidemics as well as the escalation of antibiotic resistance increased the interest in phage therapy. The work presented in this thesis has provided insight into the generation of synthetic phage infecting *C. difficile*.

Initial work focused on the isolation of novel *C. difficile* bacteriophages. In Chapter three, five phages were successfully isolated, three of which were characterised in terms of therapeutic attributes, including phage particle morphology, genome analysis and their host specificity. The complete genome sequence of phage ΦCD3801, ΦCD8601 and ΦCD9301 has been detailed and analysed, confirming the temperate nature of all three phage as well the absence of toxin gene homologues, CRISPR arrays or genes encoding for abortive infection proteins, confirming their suitability for therapeutic purposes. Moreover, the newly isolated phages were shown to infect multiple *C. difficile* strains across different PCR ribotypes, especially hypervirulent ribotype RT078. This is an interesting finding because in previous studies most of the *C. difficile* PCR ribotype RT078 isolates were resistant to phage infection (Nale *et al*., 2016; Rashid *et al*., 2016). This is important as CDI has been predominantly associated with ribotypes RT027 and RT078. Strains belonging to RT078 are more frequently associated with community-aquired disease, increased toxin production, severe CDI and higher mortality (Walker *et al*., 2013; Li *et al*., 2020). Furthermore, *C. difficile* RT078 strains have been confirmed to be involved in zoonotic transmission. Unlike the *C. difficile* RT027 lineage, RT078 strains have been isolated from both humans and farm animals and bidirectional spread contributes to distribution of antimicrobial resistance genes between humans and animals (Knetsch *et al*., 2018). Broad range activity against hypervirulent RT078 strains demonstrates that newly isolated phage ΦCD3801, ΦCD8601 and ΦCD9301 have an exciting potential in the therapeutics field as well as veterinary medicine.

Since no strictly lytic phage have been isolated for *C. difficile* to the date, we aimed to create recombinant temperate phage carrying lethal cargo. Using phage for the delivery of lethal cargo will kill the target bacteria in two forms, firstly killing solely through lytic phage infection and secondly killing through the delivery and expression

of lethal genes. To aid these studies, following work in Chapter four focused on optimising the DNA transfer into the phage genome and investigating suitable locations for cargo knocking. Using a selectable *ermB* cargo gene, three locations in the ΦCD9301 genome have been established to be suitable for cargo knock-in with the phage being able to package additional genetic information into the functional particles without disturbing its own replication and packaging process. In this chapter, we have shown that *C. difficile* temperate phage genomes can be readily modified using allelic exchange methods based on either the *codA* counter selection marker or RiboCas9.

To investigate the potential lethal cargo candidates, the negative effect of *B. substilis* SASPA and SASPB on *C. difficile* viability was determined using plasmids and both cargos were successfully integrated into the ΦCD9301 genome. Unfortunately, no improvements in the activity of recombinant phage were detected *in vitro* when compared to the wild type phage ΦCD9301. This observation was unexpected as both *sasp* cargos were shown to have detrimental effect on *C. difficile* when plasmid borne. We postulate that the level of expression of the *sasp* genes was not sufficient to kill the host using the RiboLac inducible promoter system. Moreover, RiboLac exhibited significant activity in the absence of the inducers, especially when used together with *saspB*. It is, therefore, possible, that SASPS have a toxic effect on the phage itself and interfere with phage replication or assembly. Fairhead *et al*. (2009) previously constructed recombinant phage carrying SASPC targeting *S. aureus*. In their work, the *sasp* gene was inserted into a lysis module, replacing genes required for host lysis. Lack of access to the phage lytic cycle allowed continued expression and accumulation of SASP and subsequent host death. The use of the weak constitutive promotor of the *fbaA* gene directed the expression of SASPC at a level that was insufficient to kill the propagating strain but sufficient to kill the target bacteria.

Future work should investigate the use of more suitable promotor systems that would accomplish minimal expression of *sasp* to avoid targeting the manufacturing strain and phage but at the same time achieve sufficient expression levels to kill the target bacteria. To achieve this, weak constitutive promotors or tighter inducible

systems that exhibit high levels of gene expression and minimal background expression should be explored and tested in *C. difficile*. For example, a xylose-inducible expression system *xylR* based on the native P_{xyl} promoter of *C. difficile* has been recently proven to drive tightly controlled strong expression of the gene of interest in *C. difficile* (Müh *et al*., 2019). At the time of writing this discussion, four SASPs of *C. difficile* R20291 were characterised by Nerber and Sorg (2021). It would be worthwhile investigating the use of the *C. difficile* native *sspA* promoter coupled with a theophylline riboswitch. Another option worth investigating would be to replace the integrase gene of ΦCD9301 with *sasp* under the control of a weak constitutive promotor from *C. difficile* as a similar tactic was successfully employed for *S. aureus* (Fairhead *et al*., 2009).

It may also be beneficial to employ different *sasp* genes either from *C. difficile* or other members of the *Clostridium* genus. Nerber and Sorg (2021) emphasise that despite the strong sequence similarity, SASPs have varying roles and strength between spore forming organisms. We have demonstrated the different effect of SASPA and SASPB from *B. substillis* on *C. difficile* in this study, it can therefore be useful to explore other SASPs and their impact on the host when delivered by a recombinant phage.

In Chapter five, we tested the concept of re-directing heterologous and endogenous CRISPR-Cas systems towards toxic genes specific to this pathogen using recombinant phage ΦCD9301. Firstly, *C. difficile* strains with deleted toxin genes A and B were successfully constructed to avoid self-targeting of manufacturing strain used to harvest the recombinant phage. The major advance made in this study was the characterisation and re-directing endogenous CRISPR-cas3 system of *C. difficile* M120. This strain belongs to the clinically relevant ribotype RT078. Firstly, we successfully utilized the endogenous CRISPR-Cas system for genome editing of *C. difficile* M120 selecting *pyrE* for gene knockout. Secondly, we effectively repurposed the system towards autoimmunity, achieving a 100% efficient self-targeting response and no off-target cas3 cleavage activity. To my knowledge, this is the first time that an endogenous CRISPR system was re-directed in the *C. difficile* M120 strain.

Finally, two mini-arrays redirecting the endogenous CRISPR of *C. difficile* M120 were successfully integrated into the ΦCD9301 genome. Unfortunately, neither of the recombinant phage generated exhibit enhanced activity against *C. difficile in vitro* when compared to the wild type phage, probably due to the insufficient expression levels of the cargo. As the system was 100% efficient when delivered on plasmid, we conclude that as with the *sasp* recombinant phages, the choice of a more appropriate promoter system appears to be crucial and should be the focus of future experiments. Selle and colleagues (Selle *et al*., 2020) recently generated recombinant *C. difficile* bacteriophage ΦCD24-2 expressing bacterial self-targeting CRISPR RNAs. In this study, recombinant ΦCD24-2 exhibited significantly enhanced bacteriocidal activity compared to the wild type phage, successfully inhibiting the numbers of *C. difficile* CFU both *in vitro* and *in vivo* using animal model. Unlike our study, an endogenous CRISPR mini-array was used that was preceded by the native CRISPR leader sequence from *C. difficile* 630 (Selle *et al*. 2020). Based on these findings, we believe that it will be essential to repeat the experiment with a native CRISPR leader sequence from *C. difficile* M120 as this approach has been proven to be efficient in a different *C. difficile* strain. Another option will be to use a strong constitutive promotor such as Pfdx that has been shown to drive high expression of cargo (Zhang *et al*, 2015).

Once, the expression of lethal cargo in the bacterial host is high enough to inhibit the growth of *C. difficile*, future improvements to this approach can be considered. One such option is the construction of recombinant phage carrying CRISPR mini-arrays targeting multiple genes specific to *C. difficile* to avoid the potential risk of phage resistance. Additionally, albeit beyond the scope of this work, it will be important to evaluate the potential effect that the engineered phage could have in the treatment of CDI in animal model.

Other strategies to make temperate phage more lethal may also be explored in the future. One such option is the creation of a strictly lytic phage infecting *C. difficile*. This can be achieved by removing putative lysogenic genes that promote lysogeny *in vivo*. Genome analysis of the phage ΦCD3801, ΦCD8601 and ΦCD9301 identified regions of the genes involved in lysogeny control. Knock out of the key lysogeny

regulators such as integrase gene and cI repressor is likely to prevent the ability of the phages to lysogenise their bacterial host. This approach can completely remove the limitation of bacterial lysogeny. Another option is identification of genes encoding tail fibre proteins and characterisation host range determining regions involved in receptor binding. Comparison of these genes with other published genomes of *C. difficile* phages could be a step forward in the creation of a synthetic phage which could target a wide spectrum of *C. difficile* strains. Finally, therapeutic cocktails consisting of recombinant phages can be readily developed to expand their host range and reduce the risk of bacteria developing phage resistance.

Overall, the objectives set out at the start of this study were achieved and findings in this study deepened our understanding for the potential of phage therapy in the treatment of *C. difficile* infection. Altogether, we demonstrate the feasibility of construction of a recombinant bacteriophage infecting *C. difficile* that carry lethal cargo either in form of *sasp* genes or CRISPR mini-array targeting the host genome. Once a better suitable promotor system is established, future expansions to this approach can be considered.

Chapter seven:

Appendix

7. Appendix

Table 7.1: List of all *C. difficile* strains tested with sample A. In total, 30 strains were used as indicator strains in phage isolation assay with environmental sample obtained in January 2017.

C. difficile strain	PCR ribotype	Source	Isolation location
CD1101	002	University of Nottingham, N. Minton	UK, Nottingham
CD2301	014	University of Nottingham, N. Minton	UK, Nottingham
CDDH183	027	Val Hall, Anaerobic Reference Unit	UK, Leicester
CDDH85	027	Val Hall, Anaerobic Reference Unit	UK, Birmingham
CD2801	027	M. Lister	UK, Nottingham
CD5101	027	M. Lister	UK, Nottingham
CD9201	027	M. Lister	UK, Nottingham
CD3801	078	M. Lister	UK, Nottingham
CD4401	078	M. Lister	UK, Nottingham
CD8601	078	M. Lister	UK, Nottingham
CD9001	078	M. Lister	UK, Nottingham
CD9301	078	M. Lister	UK, Nottingham
CD9501	078	M. Lister	UK, Nottingham
CD9801	078	M. Lister	UK, Nottingham
CD0101	015	M. Lister	UK, Nottingham
CD0301	015	M. Lister	UK, Nottingham
CD3601	015	M. Lister	UK, Nottingham
CD4201	015	M. Lister	UK, Nottingham
CD4501	015	M. Lister	UK, Nottingham
CD1902	015	M. Lister	UK, Nottingham
CD4701	015	M. Lister	UK, Nottingham
CD4202	015	M. Lister	UK, Nottingham
CD5901	015	M. Lister	UK, Nottingham
CD6201	015	M. Lister	UK, Nottingham
CD7301	015	M. Lister	UK, Nottingham
CD4502	015	M. Lister	UK, Nottingham
CD8401	015	M. Lister	UK, Nottingham
CD8501	015	M. Lister	UK, Nottingham
CD8502	015	M. Lister	UK, Nottingham

Table 7.2: List of all *C. difficile* strains tested with sample B. In total, 74 strains were used as indicator strains in phage isolation assay with environmental sample obtained in November 2017.

C. difficile strain	PCR ribotype	Source	Isolation location
CD2801	027	M. Lister	UK, Nottingham
CD5101	027	M. Lister	UK, Nottingham

CD9201	027	M. Lister	UK, Nottingham
CD3801	078	M. Lister	UK, Nottingham
CD4401	078	M. Lister	UK, Nottingham
CD8601	078	M. Lister	UK, Nottingham
CD9001	078	M. Lister	UK, Nottingham
CD9301	078	M. Lister	UK, Nottingham
CD9501	078	M. Lister	UK, Nottingham
CD9801	078	M. Lister	UK, Nottingham
CD0101	015	M. Lister	UK, Nottingham
CD0301	015	M. Lister	UK, Nottingham
CD3601	015	M. Lister	UK, Nottingham
CD4201	015	M. Lister	UK, Nottingham
CD4501	015	M. Lister	UK, Nottingham
CD1902	015	M. Lister	UK, Nottingham
CD4701	015	M. Lister	UK, Nottingham
CD4202	015	M. Lister	UK, Nottingham
CD5901	015	M. Lister	UK, Nottingham
CD6201	015	M. Lister	UK, Nottingham
CD7301	015	M. Lister	UK, Nottingham
CD4502	015	M. Lister	UK, Nottingham
CD8401	015	M. Lister	UK, Nottingham
CD8501	015	M. Lister	UK, Nottingham
CD8502	015	M. Lister	UK, Nottingham
CD81568	027	Ed Kuijper	France
CD81566	027	Ed Kuijper	France
CD5108111	027	Ed Kuijper	The Netherlands
CD32219	027	Ed Kuijper	Luxembourg
CDM246	027	Ed Kuijper	Ireland
CD2191	027	Ed Kuijper	Ireland
CD51556	027	Ed Kuijper	Germany
CD60902	027	Ed Kuijper	Switzerland
CD26131	027	Ed Kuijper	Finland
CD26173	027	Ed Kuijper	Finland
CDC32	027	Ed Kuijper	USA (historical)
CDC38	027	Ed Kuijper	USA
CDM13042	027	Ed Kuijper	Canada
CDDH349	027	Val Hall, Anaerobic Reference Unit	UK, Cambridge
CDDH320	027	Val Hall, Anaerobic Reference Unit	UK, Newcastle
CDDH1916	027	Val Hall, Anaerobic Reference Unit	UK, Torbay
CDDH1396	027	Val Hall, Anaerobic Reference Unit	UK, Slough
CDDH1432	027	Val Hall, Anaerobic Reference Unit	UK, London
CDDH1834	027	Val Hall, Anaerobic Reference Unit	UK, Ipswich
CDDH1466	027	Val Hall, Anaerobic Reference Unit	UK, Northampton

CDDH1858	027	Val Hall, Anaerobic Reference Unit	UK, Sunderland
CDDH1751	027	Val Hall, Anaerobic Reference Unit	UK, Bradford
CD027 Alexander	027	Ed Kuijper	Austria
CD51557	027	Ed Kuijper	Germany
CDR20291	027	Val Hall, Anaerobic Reference Unit	UK, Stoke Mandeville
CDDH361	027	Val Hall, Anaerobic Reference Unit	UK, London
CDDH131	027	Val Hall, Anaerobic Reference Unit	UK, Manchester
CDR20352	027		Canada
CDDH478	027	Val Hall, Anaerobic Reference Unit	UK, Taunton
CDR12087	027		Historical EU strain
CDT31694	027		Denmark
CDDH482	027	Val Hall, Anaerobic Reference Unit	UK, Oxford
CDR24694	027	Val Hall, Anaerobic Reference Unit	UK, Haverfordwest
CDDH326	027	Val Hall, Anaerobic Reference Unit	UK, Sheffield
CDDH183	027	Val Hall, Anaerobic Reference Unit	UK, Leicester
CDDH1342	027	Val Hall, Anaerobic Reference Unit	UK, Macclesfield
CDDH1329	027	Val Hall, Anaerobic Reference Unit	UK, Coventry
CDDH85	027	Val Hall, Anaerobic Reference Unit	UK, Birmingham
CDR23970	027	Val Hall, Anaerobic Reference Unit	Ireland, Dublin
CD3188	027	Jon Brazier/Ed Kuijper	UK, Stoke Mandeville
CD3201	027	Ed Kuijper	
CDR24988	027	Val Hall, Anaerobic Reference Unit	Northern Ireland, Antrim
CD05-1223-046	027	Ed Kuijper	Belgium
CDLibramont 1	027	Ed Kuijper	Belgium
CD3288	027	Joe Sorg/ Dale Gerding	
CD60109015048	027	Ed Kuijper	The Netherlands
CDR20291	027	Trevor Lawley	
CD196	027	Trevor Lawley	
CD210	027		UK, Leeds

Table 7.3: List of sequenced *C. difficile* phages

Phage ID	Clasification	Genome	Genome size (kbp)	GC (%)	Reference
ΦMMP02	*Myoviridae*	Circular	31.67	29.6	NC_019421.1

ΦMMP04	*Myoviridae*	Circular	31.67	30.0	NC_019422.1
Φ38-2	*Siphoviridae*	Linear	41.09	30.8	NC_015568.1
ΦCD6356	*Siphoviridae*	Linear	37.66	28.4	NC_015262.1
ΦCD27	*Myoviridae*	Linear	50.93	29.4	NC_011398.1
ΦC2	*Myoviridae*	Linear	56.54	28.7	NC_009231.1
ΦCD146	*Siphoviridae*	Linear	41.51	30.7	NC_028958.1
ΦCD119	*Myoviridae*	Linear	53.33	28.7	NC_007917.1
ΦCDHM1	*Myoviridae*	Linear	54.27	28.4	NC_024144.1
ΦCDHM11	*Myoviridae*	Circular	32	30.5	NC_029001.1
ΦCDHM13	*Myoviridae*	Circular	33.6	30	NC_029116.1
ΦCDHM19	*Myoviridae*	Circular	54.3	28.5	NC_028996.1
ΦCD211 = ΦCDIF1296T	*Siphoviridae*	Linear	131.3	26.4	NC_029048.1 CP011970.1

ΦCD505	*Myoviridae*	Linear	49.32	29.4	NC_028764.1
ΦMMP03	*Myoviridae*	Linear	52.26	28.9	NC_028959.1
ΦCD481-1	*Myoviridae*	Linear	32.85	30.3	NC_028951.1
ΦCD506	*Myoviridae*	Linear	33.27	29.6	NC_028838.1
ΦCD111	*Siphoviridae*	Linear	41.56	30.9	NC_028905.1
ΦCDKM15	*Myoviridae*	Linear	50.61	29.0	KX228400.1
ΦCDKM9	*Myoviridae*	Linear	49.82	29.0	KX228399.1
ΦCD24-1	*Siphoviridae*	Linear	44.13	28.5	LN681534.1
ΦCDHM14	*Myoviridae*	Circular	32.65	29.5	LK985321.1
ΦCDSH1	*Siphoviridae*	Linear	41.62	30.8	KU057941.1
ΦMMP01	*Myoviridae*	Linear	44.46	28.9	NC_028883.1
ΦSemix	*Myoviridae*	Circular	56.6	26.7	KX905163.1

ΦJD032	*Myoviridae*	Circular	35.11	29.93	MK473382.1
ΦCDKH01	*Siphoviridae*	Linear	45.09	28.7	MN718463.1
ΦLIBA6276	*Siphoviridae*	Circular	134.25	29.3	MF547662.1
ΦLIBA2945	*Siphoviridae*	Circular	132.5	29.3	MF547663.1

Table 7.4: List of predicted gene products of phage genome ΦCD3801.

Gene name	Predicted product	Organism	Query Cover	E value	Identity
phiCD3801_GP1	terminase large subunit	*C. difficile*	99%	0.00E+00	98%
phiCD3801_GP2	phage portal protein	*C. difficile*	98%	0	98%
phiCD3801_GP3	HK97 family phage prohead protease	*C. difficile*	100%	1.00E-148	97%
phiCD3801_GP4	phage major capsid protein	*C. difficile*	100%	0.00E+00	93%
phiCD3801_GP5	head-tail adaptor protein	*C. difficile*	100%	6.00E-143	93%
phiCD3801_GP6	Hypothetical protein	*C. difficile*	100%	3.00E-66	98%
phiCD3801_GP7	Putative head-tail joining protein	*Clostridium* phage phiCD6356	98%	1.10E-95	96.60%

phiCD3801_GP8	Hypothetical protein	*C. difficile*	100%	7.00E-73	98%
phiCD3801_GP9	Phage major tail protein	*C. difficile*	100%	2.00E-141	99%
phiCD3801_GP10	Hypothetical protein	*C. difficile*	100%	4.00E-52	87%
phiCD3801_GP11	phage tail tape measure protein	*C. difficile*	100%	0.00E+00	98%
phiCD3801_GP12	phage tail family protein	*C. difficile*	99%	0.00E+00	99%
phiCD3801_GP13	Hypothetical protein	*C. difficile*	100%	0.00E+00	98%
phiCD3801_GP14	Hypothetical protein	*C. difficile*	100%	0.00E+00	99%
phiCD3801_GP15	kelch repeat-containing protein	*C. difficile*	100%	0.00E+00	100%
phiCD3801_GP16	hypothetical protein	*C. difficile*	100%	2.00E-79	100%
phiCD3801_GP17	hypothetical protein	*C. difficile*	97%	2.00E-148	99%
phiCD3801_GP18	hypothetical protein	*C. difficile*	100%	5.00E-54	100%
phiCD3801_GP19	probable phage N-acetylmuramoyl-L-alanine amidase	*Clostridium* phage phiCD6356	100%	0.00E+00	93%
phiCD3801_GP20	hypothetical protein	*C. difficile*	100%	3.00E-74	97%
phiCD3801_GP21	ParA family protein	*C. difficile*	100%	3.00E-167	93%

phiCD3801_GP22	hypothetical protein	*C. difficile*	100%	3.00E-154	99%
phiCD3801_GP23	Hypothetical protein	*C. difficile*	100%	4.00E-157	99%
phiCD3801_GP24	XRE family transcriptional regulator	*C. difficile*	100%	8.00E-63	100%
phiCD3801_GP25	helix-turn-helix transcriptional regulator	*C. difficile*	99%	8.00E-75	99%
phiCD3801_GP26	Hypothetical protein	*C. difficile*	100%	0.00E+00	99%
phiCD3801_GP27	Hypothetical protein	*C. difficile*	100%	1.00E-141	100%
phiCD3801_GP28	Hypothetical protein	*C. difficile*	100%	5.00E-58	100%
phiCD3801_GP29	DUF2513 domain-containing protein	*C. difficile*	55%	8.00E-75	99%
phiCD3801_GP30	single-stranded DNA-binding protein	*C. difficile*	100%	6.00E-94	100%
phiCD3801_GP31	Hypothetical protein	*C. difficile*	100%	1.00E-53	100%
phiCD3801_GP32	Hypothetical protein	*C. difficile*	100%	2.00E-98	97%
phiCD3801_GP33	site-specific integrase	*C. difficile*	100%	4.00E-131	100%
phiCD3801_GP34	HNH endonuclease	*C. difficile*	100%	6.00E-73	100%
phiCD3801_GP35	RNA polymerase subunit sigma-70	*Clostridium* sp. HMSC19D07	100%	2.00E-117	100%

Table 7.5: List of predicted gene products of phage genome ΦCD9301.

Gene name	predicted product	organism	Query Cover	E value	Identity
phiCD9301_GP1	phage terminase	*C. difficile*	99%	0.00E+00	100%
phiCD9301_GP2	phage portal protein	*C. difficile*	100%	0.00E+00	98%
phiCD9301_GP3	phage head morphogenesis protein	*C. difficile*	100%	0.00E+00	100%
phiCD9301_GP4	hypothetical protein	*C. difficile*	100%	2.00E-92	99%
phiCD9301_GP5	hypothetical protein	*C. difficile*	100%	3.00E-74	74%
phiCD9301_GP6	hypothetical protein	*C. difficile*	100%	3.00E-54	100%
phiCD9301_GP7	hypothetical protein	*C. difficile*	100%	1.00E-136	99%
phiCD9301_GP8	hypothetical protein	*C. difficile*	100%	0.00E+00	99%
phiCD9301_GP9	phage gp6-like head-tail connector family protein	*C. difficile*	100%	2.00E-73	99%
phiCD9301_GP10	hypothetical protein	*C. difficile*	100%	5.00E-80	100%
phiCD9301_GP11	HK97 gp10 family phage protein	*C. difficile*	100%	1.00E-95	100%
phiCD9301_GP12	hypothetical protein	*C. difficile*	94%	6.00E-102	99%
phiCD9301_GP13	phage tail sheath family protein	*C. difficile*	100%	0.00E+00	100%
phiCD9301_GP14	hypothetical protein	*C. difficile*	100%	7.00E-100	100%
phiCD9301_GP15	phage XkdN-like family protein	*C. difficile*	100%	3.00E-105	100%
phiCD9301_GP16	phage tail tape measure protein	*C. difficile*	100%	0.00E+00	91%

phiCD9301_GP17	hypothetical protein	*C. difficile*	98%	6.00E-140	95%
phiCD9301_GP18	LysM peptidoglycan-binding domain-containing protein	*C. difficile*	100%	1.00E-148	99%
phiCD9301_GP19	cell wall hydrolase	*C. difficile*	100%	0.00E+00	97%
phiCD9301_GP20	DUF2577 domain-containing protein	*C. difficile*	95%	5.00E-67	95%
phiCD9301_GP21	DUF2634 domain-containing protein	*C. difficile*	100%	9.00E-94	95%
phiCD9301_GP22	putative baseplate-J protein	*C. difficile*	99%	0.00E+00	96%
phiCD9301_GP23	DUF2313 domain-containing protein	*C. difficile*	100%	2.00E-140	94%
phiCD9301_GP24	phage tail protein	*C. difficile*	100%	4.00E-174	95%
phiCD9301_GP25	glycine rich family protein	*C. difficile*	100%	0.00E+00	72%
phiCD9301_GP26	hypothetical protein	*C. difficile*	100%	6.00E-61	98%
phiCD9301_GP27	N-acetylmuramoyl-L-alanine amidase	*C. difficile*	100%	0.00E+00	92%
phiCD9301_GP28	hypothetical protein	*C. difficile*	100%	3.00E-102	100%
phiCD9301_GP29	AP2 domain-containing protein	*C. difficile*	100%	2.00E-147	99%
phiCD9301_GP30	hypothetical protein	*C. difficile*	100%	7.00E-105	98%
phiCD9301_GP31	hypothetical protein	*C. difficile*	100%	1.00E-59	94%
phiCD9301_GP32	penicillinase repressor family protein	*C. difficile*	100%	4.00E-57	100%
phiCD9301_GP33	penicillinase repressor family protein	*Clostridiales*	100%	3.00E-86	100%

phiCD9301_GP34	site-specific integrase	*C. difficile*	100%	0.00E+00	100%
phiCD9301_GP35	hypothetical protein	*C. difficile*	100%	0.00E+00	100%
phiCD9301_GP36	XRE family transcriptional regulator	*C. difficile*	69%	5.00E-87	98%
phiCD9301_GP37	phage regulatory protein	*C. difficile*	100%	8.00E-178	92%
phiCD9301_GP38	hypothetical protein	*C. difficile*	93%	2.00E-65	95%
phiCD9301_GP39	helix-turn-helix domain-containing protein	*C. difficile*	100%	2.00E-69	97%
phiCD9301_GP40	siphovirus Gp157 family protein	*C. difficile*	100%	6.00E-112	99%
phiCD9301_GP41	recombinase	*C. difficile*	100%	3.00E-144	100%
phiCD9301_GP42	DnaD domain protein	*C. difficile*	100%	0.00E+00	99%
phiCD9301_GP43	single-stranded DNA-binding protein	*C. difficile*	100%	3.00E-94	96%
phiCD9301_GP44	hypothetical protein	*C. difficile*	100%	5.00E-51	99%
phiCD9301_GP45	hypothetical protein	*C. difficile*	100%	5.00E-83	100%
phiCD9301_GP46	hypothetical protein	*C. difficile*	100%	7.00E-97	93%
phiCD9301_GP47	nucleoid-associated protein	*C. difficile*	100%	0.00E+00	94%
phiCD9301_GP48	hypothetical protein	*C. difficile*	100%	4.00E-78	95%
phiCD9301_GP49	peptide chain release factor-like protein	*C. difficile*	100%	3.00E-73	98%
phiCD9301_GP50	hypothetical protein	*C. difficile*	85%	4.00E-91	96%

phiCD9301_GP51	RusA family crossover junction endodeoxyribonuclease	*C. difficile*	97%	1.00E-92	98%
phiCD9301_GP52	hypothetical protein	*C. difficile*	100%	0	99%
phiCD9301_GP53	hypothetical protein	*C. difficile*	100%	1.00E-112	98%
phiCD9301_GP54	GNAT family acetyltransferase	*C. difficile*	100%	2.00E-141	99%
phiCD9301_GP55	hypothetical protein	*C. difficile*	100%	3.00E-52	100%
phiCD9301_GP56	transposase	*C. difficile*	90%	1.00E-103	99%

Chapter eight:

References

8. References

Abedon, S. T. (2017). Bacteriophage Clinical Use as Antibacterial" Drugs": Utility and Precedent. *Microbiology spectrum*, *5*(4).

Abeyawardhane, D. L., Godoy-Ruiz, R., Adipietro, K. A., Varney, K. M., Rustandi, R. R., Pozharski, E., & Weber, D. J. (2021). The Importance of Therapeutically Targeting the Binary Toxin from *Clostridioides difficile*. *International Journal of Molecular Sciences*, *22*(6), 2926.

Abt, M. C., McKenney, P. T., & Pamer, E. G. (2016). Clostridium difficile colitis: pathogenesis and host defence. *Nature Reviews Microbiology*, *14*(10), 609-620.

Aktories, K., Schwan, C., & Jank, T. (2017). *Clostridium difficile* toxin biology. *Annual review of microbiology*, *71*, 281-307.

Al-Shayeb, B., Sachdeva, R., Chen, L. X., Ward, F., Munk, P., Devoto, A., ... & He, C. (2020). Clades of huge phages from across Earth's ecosystems. *Nature*, *578*(7795), 425-431.

Álvarez-Pérez, S., Blanco, J. L., Harmanus, C., Kuijper, E., & García, M. E. (2017). Subtyping and antimicrobial susceptibility of Clostridium difficile PCR ribotype 078/126 isolates of human and animal origin. *Veterinary microbiology*, *199*, 15-22.

Alves, D. R., Perez-Esteban, P., Kot, W., Bean, J. E., Arnot, T., Hansen, L. H., ... & Jenkins, A. T. A. (2016). Bacteriophages to treat P. aeruginosa biofilms. *Microbial Biotechnology*, *9*(1), 61-74.

Amy, J., Bulach, D., Knight, D., Riley, T., Johanesen, P., & Lyras, D. (2018). Identification of large cryptic plasmids in Clostridioides (Clostridium) difficile. *Plasmid*, *96*, 25-38.

Anand, T., Virmani, N., Kumar, S., Mohanty, A. K., Pavulraj, S., Bera, B. C., ... & Tripathi, B. N. (2020). Phage therapy for treatment of virulent Klebsiella pneumoniae infection in a mouse model. *Journal of global antimicrobial resistance*, *21*, 34-41.

Andersen, J. M., Shoup, M., Robinson, C., Britton, R., Olsen, K. E., & Barrangou, R. (2016). CRISPR diversity and microevolution in Clostridium difficile. *Genome biology and evolution*, *8*(9), 2841-2855.

Ando, H., Lemire, S., Pires, D. P., & Lu, T. K. (2015). Engineering modular viral scaffolds for targeted bacterial population editing. *Cell systems*, *1*(3), 187-196.

Arndt, D., Grant, J., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res., 2016 May 3.

Azeredo, J., & Sutherland, I. W. (2008). The use of phages for the removal of infectious biofilms. *Current pharmaceutical biotechnology*, *9*(4), 261-266.

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

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709-1712.

Benler, S., & Koonin, E. V. (2020). Phage lysis-lysogeny switches and programmed cell death: Danse macabre. *BioEssays*, *42*(12), 2000114.

Bhardwaj, S. B., Mehta, M., Sood, S., & Sharma, J. (2020). Isolation of a novel phage and targeting biofilms of drug-resistant oral enterococci. *Journal of Global Infectious Diseases*, *12*(1), 11.

Bikard, D., & Barrangou, R. (2017). Using CRISPR-Cas systems as antimicrobials. *Current opinion in Microbiology*, *37*, 155-160.

Bikard, D., Euler, C. W., Jiang, W., Nussenzweig, P. M., Goldberg, G. W., Duportet, X., ... & Marraffini, L. A. (2014). Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nature biotechnology*, *32*(11), 1146-1150.

Biswas, A., Gagnon, J. N., Brouns, S. J., Fineran, P. C., & Brown, C. M. (2013). CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA biology*, *10*(5), 817-827.

Born, Y., Fieseler, L., Thöny, V., Leimer, N., Duffy, B., & Loessner, M. J. (2017). Engineering of bacteriophages Y2:: dpoL1-C and Y2:: luxAB for efficient control and rapid detection of the fire blight pathogen, Erwinia amylovora. Appl. Environ. Microbiol., 83(12), e00341-17.

Boudry, P., Semenova, E., Monot, M., Datsenko, K. A., Lopatina, A., Sekulovic, O., ... & Soutourina, O. (2015). Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. *MBio*, *6*(5), e01112-15.

Brown, A. W., & Wilson, R. B. (2018*). Clostridium difficile* colitis and zoonotic origins—a narrative review. *Gastroenterology report*, *6*(3), 157-166.

Bruynoghe, R. A. J. M., & Maisin, J. (1921). Essais de thérapeutique au moyen du bacteriophage. *CR Soc. Biol*, *85*, 1120-1121.

Bruxelle, J. F., Péchiné, S., & Collignon, A. (2018). Immunization Strategies Against Clostridium difficile. In Updates on Clostridium difficile in Europe (pp. 197-225). Springer, Cham

Buckley, A. M., Spencer, J., Candlish, D., Irvine, J. J., & Douce, G. R. (2011). Infection of hamsters with the UK Clostridium difficile ribotype 027 outbreak strain R20291. *Journal of medical microbiology*, *60*(8), 1174-1180.

Budi, N., Safdar, N., & Rose, W. E. (2020). Treatment issues in recurrent Clostridioides difficile infections and the possible role of germinants. *FEMS Microbes*, *1*(1), xtaa001.

Butler III, R. R., Schill, K. M., Wang, Y., & Pombert, J. F. (2017). Genetic characterization of the exceptionally high heat resistance of the non-toxic surrogate Clostridium sporogenes PA 3679. *Frontiers in microbiology*, *8*, 545.

Cabrera-Martinez, R. M., Mason, J. M., Setlow, B., Waites, W. M., & Setlow, P. (1989). Purification and amino acid sequence of two small, acid-soluble proteins from Clostridium bifermentans spores. *FEMS microbiology letters*, *61*(1-2), 139-143.

Campbell, A. M. (1992). Chromosomal insertion sites for phages and plasmids. *Journal of bacteriology*, *174*(23), 7495.

Cañadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., & Minton, N. P. (2019). RiboCas: a universal CRISPR-based editing tool for Clostridium. *ACS synthetic biology*, *8*(6), 1379-1390.

Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T., & 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), 4683-4690.

Casjens, S. R., & Gilcrease, E. B. (2009). Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. In *Bacteriophages* (pp. 91-111). Humana Press.

Castro-Córdova, P., Mora-Uribe, P., Reyes-Ramírez, R., Cofré-Araneda, G., Orozco-Aguilar, J., Brito-Silva, C., ... & Paredes-Sabja, D. (2021). Entry of spores into intestinal epithelial cells contributes to recurrence of Clostridioides difficile infection. *Nature communications*, *12*(1), 1-18.

Centres for Disease control and Prevention, 2019 AR threats report. https://www.cdc.gov/drugresistance/pdf/threats-report/clostridioides-difficile-508.pdf

Champredon, D., Shoukat, A., & Moghadas, S. M. (2020). Effectiveness and cost-effectiveness of a Clostridium difficile vaccine candidate in a hospital setting. *Vaccine*, *38*(11), 2585-2591.

Chanishvili, N., & Aminov, R. (2019). Bacteriophage therapy: coping with the growing antibiotic resistance problem. *Microbiology Australia*.

Chen, M., Zhang, L., Xin, S., Yao, H., Lu, C., & Zhang, W. (2017). Inducible prophage mutant of Escherichia coli can lyse new host and the key sites of receptor recognition identification. *Frontiers in microbiology*, *8*, 147.

Cho, E. H., Gumport, R. I., & Gardner, J. F. (2002). Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. *Journal of bacteriology*, *184*(18), 5200-5203.

Cho, J. M., Pardi, D. S., & Khanna, S. (2020, April). Update on treatment of Clostridioides difficile infection. In *Mayo Clinic Proceedings* (Vol. 95, No. 4, pp. 758-769). Elsevier.

Citorik RJ, Mimee M, Lu TK (2014).: Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nat Biotechnol, 32:1141-1145

Clements, A. C., Magalhães, R. J. S., Tatem, A. J., Paterson, D. L., & Riley, T. V. (2010). Clostridium difficile PCR ribotype 027: assessing the risks of further worldwide spread. The Lancet infectious diseases, 10(6), 395-404.

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

Cobb, L. H., Park, J., Swanson, E. A., Beard, M. C., McCabe, E. M., Rourke, A. S., ... & Priddy, L. B. (2019). CRISPR-Cas9 modified bacteriophage for treatment of Staphylococcus aureus induced osteomyelitis and soft tissue infection. *PLoS One*, *14*(11), e0220421.

Cornely, O. A., Miller, M. A., Louie, T. J., Crook, D. W., & Gorbach, S. L. (2012). Treatment of first recurrence of Clostridium difficile infection: fidaxomicin versus vancomycin. Clinical infectious diseases, 55(suppl_2), S154-S161.

Couturier, J., Davies, K., Gateau, C., & Barbut, F. (2018). Ribotypes and new virulent strains across Europe. In Updates on Clostridium difficile in Europe (pp. 45-58). Springer, Cham.

Crobach, M. J., Vernon, J. J., Loo, V. G., Kong, L. Y., Péchiné, S., Wilcox, M. H., & Kuijper, E. J. (2018). Understanding Clostridium difficile colonization. Clinical microbiology reviews, 31(2), e00021-17.

Czepiel, J., Dróżdż, M., Pituch, H., Kuijper, E. J., Perucki, W., Mielimonka, A., ... & Biesiada, G. (2019). Clostridium difficile infection. *European Journal of Clinical Microbiology & Infectious Diseases*, *38*(7), 1211-1221.

d'Herelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. *CR Acad. Sci. Paris*, *165*, 373-375.

d'Hérelle, F., & Smith, G. H. (1922). *The bacteriophage, its role in immunity*. Williams & Wilkins.

de Jong, A., Pietersma, H., Cordes, M., Kuipers, O. P., & Kok, J. (2012). PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC genomics*, *13*(1), 299.

Dedrick, R. M., Guerrero-Bustamante, C. A., Garlena, R. A., Russell, D. A., Ford, K., Harris, K., ... & Spencer, H. (2019). Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. *Nature medicine*, *25*(5), 730-733.

Dembek, M., Barquist, L., Boinett, C. J., Cain, A. K., Mayho, M., Lawley, T. D., ... & Fagan, R. P. (2015). High-throughput analysis of gene essentiality and sporulation in Clostridium difficile. *MBio*, *6*(2), e02383-14.

Deng L, Roose K, Job ER, De Rycke R, Van Hamme E, *et al*. 2017. Oral delivery of Escherichia coli persistently infected with M2e-displaying bacteriophages partially protects against influenza A virus. J. Control. Release 264:55–65

Desai, S. K., & Gallivan, J. P. (2004). Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. Journal of the American Chemical Society, 126(41), 13247-13254.

Dieterle, M. G., Rao, K., & Young, V. B. (2019). Novel therapies and preventative strategies for primary and recurrent Clostridium difficile infections. *Annals of the New York Academy of Sciences*, *1435*(1), 110.

Dion, M. B., Oechslin, F., & Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nature Reviews Microbiology*, 1-14.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., ... & Virgin, H. W. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology*, *34*(2), 184.

Dy, R. L., Przybilski, R., Semeijn, K., Salmond, G. P., & Fineran, P. C. (2014). A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic acids research*, *42*(7), 4590–4605. https://doi.org/10.1093/nar/gkt1419

Edwards, A. N., & McBride, S. M. (2014). Initiation of sporulation in Clostridium difficile: a twist on the classic model. *FEMS microbiology letters*, *358*(2), 110-118.

Ehsaan, M., Kuit, W., Zhang, Y., Cartman, S. T., Heap, J. T., Winzer, K., & Minton, N. P. (2016). Mutant generation by allelic exchange and genome resequencing of the

biobutanol organism Clostridium acetobutylicum ATCC 824. *Biotechnology for biofuels*, *9*(1), 1-20.

El-Arabi, T. F., Griffiths, M. W., She, Y. M., Villegas, A., Lingohr, E. J., & Kropinski, A. M. (2013). Genome sequence and analysis of a broad-host range lytic bacteriophage that infects the Bacillus cereus group. *Virology journal*, *10*(1), 48.

Embleton, M. L., Nair, S. P., Heywood, W., Menon, D. C., Cookson, B. D., & Wilson, M. (2005). Development of a novel targeting system for lethal photosensitization of antibiotic-resistant strains of Staphylococcus aureus. *Antimicrobial agents and chemotherapy*, *49*(9), 3690-3696.

Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., ... & Amitai, G. (2017). Communication between viruses guides lysis–lysogeny decisions. *Nature*, *541*(7638), 488.

European Centre for Disease Prevention and Control. Point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals 2011–2012. ECDU. 2013 online http://www.ecdc.europa.eu/en/publications/Publications/healthcare-associated-infections-antimicrobial-use-PPS.pdf.

Fage, C., Lemire, N., & Moineau, S. (2021). Delivery of CRISPR-Cas systems using phage-based vectors. *Current Opinion in Biotechnology*, *68*, 174-180.

Fairhead, H. (2009). SASP gene delivery: a novel antibacterial approach. *Drug News Perspect*, *22*(4), 197-203.

Fairley, D. J., McKenna, J. P., Stevenson, M., Weaver, J., Gilliland, C., Watt, A., & Coyle, P. V. (2015). Association of Clostridium difficile ribotype 078 with detectable toxin in human stool specimens. Journal of medical microbiology, 64(11), 1341-1345.

Farris JS: Estimating phylogenetic trees from distance matrices. Am Nat 1972, 106:645-667.

Göker M, García-Blázquez G, Voglmayr H, Tellería MT, Martín MP: Molecular taxonomy of phytopathogenic fungi: A case study in Peronospora. PLoS One 2009, 4:8-10.

Lefort V, Desper R, Gascuel O: FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol 2015, 32:2798-2800.

Findeiß, S., Etzel, M., Will, S., Mörl, M., & Stadler, P. F. (2017). Design of artificial riboswitches as biosensors. *Sensors*, *17*(9), 1990.

Fortier, L. C. (2018). Bacteriophages contribute to shaping clostridioides (Clostridium) difficile species. *Frontiers in microbiology*, *9*, 2033.

Fortier, L. C., & Moineau, S. (2007). Morphological and genetic diversity of temperate phages in Clostridium difficile. *Applied and environmental microbiology*, *73*(22), 7358-7366.

Francesconi S.C., MacAlister T.J., Setlow B., Setlow P.. (1988;). Immunoelectron microscopic localization of small, acid-soluble spore proteins in sporulating cells of Bacillus subtilis. J Bacteriol 170: 5963—5967.

Francis, M. B., Allen, C. A., Shrestha, R., & Sorg, J. A. (2013). Bile acid recognition by the Clostridium difficile germinant receptor, CspC, is important for establishing infection. *PLoS pathogens*, *9*(5), e1003356.

Garneau, J. R., Sekulovic, O., Dupuy, B., Soutourina, O., Monot, M., & Fortier, L. C. (2018). High prevalence and genetic diversity of large phiCD211 (phiCDIF1296T)-like prophages in Clostridioides difficile. Applied and environmental microbiology, 84(3), e02164-17.

Gautheret D, Lambert A. (2001) Direct RNA Motif Definition and Identification from Multiple Sequence Alignments using Secondary Structure Profiles. *J Mol Biol.* **313**:1003–11

Gebhart D, Lok S, Clare S, Tomas M, Stares M, Scholl D, Donskey CJ, Lawley TD, Govoni GR (2015) A modifi ed R-type bacteriocin specifi cally targeting Clostridium diffi - cile prevents colonization of mice without affecting gut microbiota diversity. MBio 6(2):e02368-14

Gebhart, D., Williams, S. R., & Scholl, D. (2017). Bacteriophage SP6 encodes a second tailspike protein that recognizes Salmonella enterica serogroups C2 and C3. *Virology*, *507*, 263-266.

Gerding, D. N., Johnson, S., Rupnik, M., & Aktories, K. (2014). Clostridium difficile binary toxin CDT: mechanism, epidemiology, and potential clinical importance. Gut microbes, 5(1), 15-27.

Goh, S., Riley, T. V., & Chang, B. J. (2005). Isolation and characterization of temperate bacteriophages of Clostridium difficile. *Applied and environmental microbiology*, *71*(2), 1079-1083.

Govind, R., Fralick, J. A., & Rolfe, R. D. (2006). Genomic organization and molecular characterization of Clostridium difficile bacteriophage ΦCD119. *Journal of bacteriology*, *188*(7), 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.

Grissa, I., Vergnaud, G., & Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research*, *35*(suppl_2), W52-W57.

Groth, A. C., & Calos, M. P. (2004). Phage integrases: biology and applications. Journal of molecular biology, 335(3), 667-678.

Hall, I. C., & O'TOOLE, E. L. I. Z. A. B. E. T. H. (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), 390-402.

Hamdi, S., Rousseau, G. M., Labrie, S. J., Tremblay, D. M., Kourda, R. S., Slama, K. B., & Moineau, S. (2017). Characterization of two polyvalent phages infecting Enterobacteriaceae. *Scientific reports*, *7*(1), 1-12.

Hargreaves, K. R., Flores, C. O., Lawley, T. D., & Clokie, M. R. (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), e01045-13.

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.

Hartman, A. H., Liu, H., & Melville, S. B. (2011). Construction and characterization of a lactose-inducible promoter system for controlled gene expression in Clostridium perfringens. *Appl. Environ. Microbiol.*, *77*(2), 471-478.

Hawkins, C. C., Buggy, B. P., Fekety, R., & Schaberg, D. R. (1984). Epidemiology of colitis induced by Clostridium difficile in hamsters: application of a bacteriophage and bacteriocin typing system. *Journal of Infectious Diseases*, *149*(5), 775-780.

He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D. J., Martin, M. J., ... & Lawley, T. D. (2013). Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nature genetics*, *45*(1), 109-113.

Heap, J. T., Ehsaan, M., Cooksley, C. M., Ng, Y. K., Cartman, S. T., Winzer, K., & Minton, N. P. (2012). Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic acids research*, *40*(8), e59-e59.

Heap, J. T., Pennington, O. J., Cartman, S. T., & Minton, N. P. (2009). A modular system for Clostridium shuttle plasmids. *Journal of microbiological methods*, *78*(1), 79-85.

Hegarty, J. P., Sangster, W., Ashley, R. E., Myers, R., Hafenstein, S., & Stewart, D. B. (2016). *Induction and Purification of C. difficile Phage Tail-Like Particles. Clostridium Difficile, 167–175.*

Heilpern, A. J., & Waldor, M. K. (2003). pIIICTX, a predicted CTXϕ minor coat protein, can expand the host range of coliphage fd to include Vibrio cholerae. *Journal of bacteriology*, *185*(3), 1037-1044.

Herbert, R., Hatcher, J., Jauneikaite, E., Gharbi, M., d'Arc, S., Obaray, N., ... & Thomas, A. (2019). Two-year analysis of Clostridium difficile ribotypes associated with increased severity. *Journal of Hospital Infection*, *103*(4), 388-394.

Heuler, J., Fortier, L. C., & Sun, X. (2021). Clostridioides difficile phage biology and application. *FEMS Microbiology Reviews*.

Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., & Charpentier, E. (2018). The biology of CRISPR-Cas: backward and forward. *Cell*, *172*(6), 1239-1259.

Horgan, M., O'Sullivan, O., Coffey, A., Fitzgerald, G. F., van Sinderen, D., McAuliffe, O., & Ross, R. P. (2010). Genome analysis of the Clostridium difficile phage ΦCD6356, a temperate phage of the Siphoviridae family. *Gene*, *462*(1-2), 34-43.

Horton, R. M., Cai, Z. L., Ho, S. N., & Pease, L. R. (1990). Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques*, *8*(5), 528-535.

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *The ISME journal*, *11*(7), 1511.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., ... & Cradick, T. J. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, *31*(9), 827.

Hyman, P. (2019). Phages for phage therapy: isolation, characterization, and host range breadth. *Pharmaceuticals*, *12*(1), 35.

Janezic, S., Dingle, K., Alvin, J., Accetto, T., Didelot, X., Crook, D. W., ... & Rupnik, M. (2020). Comparative genomics of Clostridioides difficile toxinotypes identifies module-based toxin gene evolution. *Microbial genomics*, *6*(10).

Jarmo, O., Veli-Jukka, A., & Eero, M. (2020). Treatment of Clostridioides (Clostridium) difficile infection. *Annals of Medicine*, *52*(1-2), 12-20.

Jarrad, A. M., Karoli, T., Blaskovich, M. A., Lyras, D., & Cooper, M. A. (2015). Clostridium difficile drug pipeline: challenges in discovery and development of new agents. *Journal of medicinal chemistry*, *58*(13), 5164-5185.

Jiang, Z., Wei, J., Peng, N., & Li, Y. (2021). Genetic Engineering of a Phage-Based Delivery System for Endogenous III-A CRISPR-Cas System Against Mycobacterium tuberculosis. In CRISPR-Cas Methods (pp. 311-323). Humana, New York, NY.

Jones, A. M., Kuijper, E. J., & Wilcox, M. H. (2013). Clostridium difficile: a European perspective. Journal of Infection, 66(2), 115-128.

Kachrimanidou, M., & Tsintarakis, E. (2020). Insights into the Role of Human Gut Microbiota in Clostridioides difficile Infection. *Microorganisms*, *8*(2), 200.

Kamiura, R., Toya, Y., Matsuda, F., & Shimizu, H. (2019). Theophylline-inducible riboswitch accurately regulates protein expression at low level in Escherichia coli. *Biotechnology letters*, *41*(6-7), 743-751.

Kampouri, E., Croxatto, A., Prod'hom, G., & Guery, B. (2021). Clostridioides difficile Infection, Still a Long Way to Go. *Journal of Clinical Medicine*, *10*(3), 389.

Kassam, Z., Lee, C. H., Yuan, Y., & Hunt, R. H. (2013). Fecal microbiota transplantation for Clostridium difficile infection: systematic review and meta-analysis. The American journal of gastroenterology, 108(4), 500.

Keen, E. C. (2015). A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays*, *37*(1), 6-9.

Keen, E. C., & Dantas, G. (2018). Close encounters of three kinds: bacteriophages, commensal bacteria, and host immunity. *Trends in microbiology*, *26*(11), 943-954.

Kifelew, L. G., Warner, M. S., Morales, S., Vaughan, L., Woodman, R., Fitridge, R., ... & Speck, P. (2020). Efficacy of phage cocktail AB-SA01 therapy in diabetic mouse wound infections caused by multidrug-resistant Staphylococcus aureus. *BMC microbiology*, *20*(1), 1-10.

Knetsch, C. W., Kumar, N., Forster, S. C., Connor, T. R., Browne, H. P., Harmanus, C., ... & Lawley, T. D. (2018). Zoonotic transfer of Clostridium difficile harboring antimicrobial resistance between farm animals and humans. *Journal of clinical microbiology*, *56*(3), e01384-17.

Kochan, T. J., Foley, M. H., Shoshiev, M. S., Somers, M. J., Carlson, P. E., & Hanna, P. C. (2018). Updates to Clostridium difficile spore germination. Journal of bacteriology, 200(16), e00218-18.

Kodori, M., Ghalavand, Z., Yadegar, A., Eslami, G., Azimirad, M., Krutova, M., ... & Zali, M. R. (2020). Molecular characterization of pathogenicity locus (PaLoc) and tcdC genetic diversity among tcdA+ B+ Clostridioides difficile clinical isolates in Tehran, Iran. *Anaerobe*, *66*, 102294.

Koenigsknecht, M. J., Theriot, C. M., Bergin, I. L., Schumacher, C. A., Schloss, P. D., & Young, V. B. (2015). Dynamics and establishment of Clostridium difficile infection in the murine gastrointestinal tract. *Infection and immunity*, *83*(3), 934-941.

Kortright, K. E., Chan, B. K., Koff, J. L., & Turner, P. E. (2019). Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. *Cell host & microbe*, *25*(2), 219-232.

Kutter, E., Borysowski, J., Miedzybrodzki, R., Górski, A., Weber-Dabrowska, B., Kutateladze, M., ... & Adamia, R. (2014). Clinical phage therapy. *Phage Therapy: Current Research and Applications*, 257-288.

Lam, K. N., Spanogiannopoulos, P., Soto-Perez, P., Alexander, M., Nalley, M. J., Bisanz, J. E., ... & Turnbaugh, P. J. (2020). Phage-delivered CRISPR-Cas9 for strain-specific depletion and genomic deletions in the gut microbiota. *bioRxiv*.

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

Lebeaux, D., Merabishvili, M., Caudron, E., Lannoy, D., Van Simaey, L., Duyvejonck, H., ... & Vaneechoutte, M. (2021). A Case of Phage Therapy against Pandrug-Resistant Achromobacter xylosoxidans in a 12-Year-Old Lung-Transplanted Cystic Fibrosis Patient. *Viruses*, *13*(1), 60.

Leffler, D. A., & Lamont, J. T. (2015). Clostridium difficile infection. *New England Journal of Medicine*, *372*(16), 1539-1548.

Le Lay, C., Dridi, L., Bergeron, M. G., & Ouellette, M. (2016). Nisin is an effective inhibitor of Clostridium difficile vegetative cells and spore germination. Journal of medical microbiology, 65(2), 169-175.

Lessa, F. C., Mu, Y., Bamberg, W. M., Beldavs, Z. G., Dumyati, G. K., Dunn, J. R., ... & McDonald, L. C. (2015). Burden of Clostridium difficile infection in the United States. *New England Journal of Medicine*, *372*(9), 825-834.

Leszczyński, P., Weber-Dabrowska, B., Kohutnicka, M., Łuczak, M., Górecki, A., & Górski, A. (2006). Successful eradication of methicillin-resistantStaphylococcus aureus (MRSA) intestinal carrier status in a healthcare worker—case report. *Folia microbiologica*, *51*(3), 236-238.

Letkiewicz, S., Międzybrodzki, R., Fortuna, W., Weber-Dąbrowska, B., & Górski, A. (2009). Eradication of Enterococcus faecalis by phage therapy in chronic bacterial prostatitis—case report. *Folia microbiologica*, *54*(5), 457.

Li, J., & McClane, B. A. (2008). A novel small acid soluble protein variant is important for spore resistance of most Clostridium perfringens food poisoning isolates. *PLoS pathogens*, *4*(5).

Li, J., Paredes-Sabja, D., Sarker, M. R., & McClane, B. A. (2009). Further characterization of Clostridium perfringens small acid soluble protein-4 (Ssp4) properties and expression. *PLoS One*, *4*(7).

Li, M., Shi, D., Li, Y., Xiao, Y., Chen, M., Chen, L., ... & Zhang, W. (2020). Recombination of T4-like Phages and Its Activity against Pathogenic Escherichia coli in Planktonic and Biofilm Forms. *Virologica Sinica*, 1-11.

Li, T., Zhang, Y., Dong, K., Kuo, C. J., Li, C., Zhu, Y. Q., ... & Zhu, Y. (2020). Isolation and characterization of the novel phage JD032 and global transcriptomic response during JD032 infection of Clostridioides difficile ribotype 078. *MSystems*, *5*(3), e00017-20.

Lin, T. Y., Lo, Y. H., Tseng, P. W., Chang, S. F., Lin, Y. T., & Chen, T. S. (2012). A T3 and T7 recombinant phage acquires efficient adsorption and a broader host range. *PloS one*, *7*(2), e30954.

Luong, T., Salabarria, A. C., & Roach, D. R. (2020). Phage therapy in the resistance era: where do we stand and where are we going?. *Clinical Therapeutics*.

Lynch, S. A., Desai, S. K., Sajja, H. K., & Gallivan, J. P. (2007). A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chemistry & biology*, *14*(2), 173-184.

Lyon, S. A., Hutton, M. L., Rood, J. I., Cheung, J. K., & Lyras, D. (2016). CdtR regulates TcdA and TcdB production in Clostridium difficile. PLoS pathogens, 12(7), e1005758.

Macke T, Ecker D, Gutell R, Gautheret D, Case DA and Sampath R. (2001) RNAMotif – A new RNA secondary structure definition and discovery algorithm. *Nucleic Acids Res.* **29**:4724–4735

Magali Naville, Adrien Ghuillot-Gaudeffroy, Antonin Marchais & Daniel Gautheret (2011) ARNold: A web tool for the prediction of Rho-independent transcription terminators, RNA Biology, 8:1, 11-13, DOI: 10.4161/rna.8.1.13346

Mahony, D. E., Bell, P. D., & Easterbrook, K. B. (1985). Two bacteriophages of Clostridium difficile. *Journal of clinical microbiology*, *21*(2), 251-254.

Maikova, A., Severinov, K., & Soutourina, O. (2018). New insights into functions and possible applications of clostridium difficile CRISPR-Cas system. *Frontiers in microbiology*, *9*, 1740.

Maikova, A., Kreis, V., Boutserin, A., Severinov, K., & Soutourina, O. (2019). Using an endogenous CRISPR-Cas system for genome editing in the human pathogen Clostridium difficile. *Applied and environmental microbiology*, *85*(20), e01416-19.

Marintcheva, B. (2018). Lightning Round: Harnessing the Power of Viruses.

Martins, D., DiCandia, M. A., Mendes, A. L., Wetzel, D., McBride, S. M., Henriques, A. O., & Serrano, M. (2021). CD25890, a conserved protein that modulates sporulation initiation in Clostridioides difficile. *Scientific reports*, *11*(1), 1-15.

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

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

Meader, E., Mayer, M. J., Steverding, D., Carding, S. R., & Narbad, A. (2013). Evaluation of bacteriophage therapy to control Clostridium difficile and toxin production in an in vitro human colon model system. *Anaerobe*, *22*, 25-30.

Meaney, C. A., Cartman, S. T., McClure, P. J., & Minton, N. P. (2016). The role of small acid-soluble proteins (SASPs) in protection of spores of Clostridium botulinum against nitrous acid. *International journal of food microbiology*, *216*, 25-30.

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

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M: Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 2013, 14:60.

Meier-Kolthoff JP, Göker M. VICTOR: Genome-based Phylogeny and Classification of Prokaryotic Viruses. Bioinformatics 2017, 33:3396-3404.

Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C, Rohde M, Fartmann B, Goodwin L a., Chertkov O, Reddy TBK, Pati A, Ivanova NN, Markowitz V, Kyrpides NC, Woyke T, Göker M, Klenk H-P: Complete genome sequence of DSM 30083T, the type strain (U5/41T) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci 2014, 9:2.

Mellin, J. R., & Cossart, P. (2015). Unexpected versatility in bacterial riboswitches. *Trends in Genetics*, *31*(3), 150-156.

Moore, J. H., Honrado, C., Stagnaro, V., Kolling, G., Warren, C. A., & Swami, N. S. (2020). Rapid in vitro assessment of Clostridioides difficile inhibition by probiotics using dielectrophoresis to quantify cell structure alterations. *ACS infectious diseases*, *6*(5), 1000-1007.

Müh, U., Pannullo, A. G., Weiss, D. S., & Ellermeier, C. D. (2019). A xylose-inducible expression system and a CRISPR interference plasmid for targeted knockdown of gene expression in Clostridioides difficile. *Journal of Bacteriology*, *201*(14), e00711-18.

Mushegian, A. R. "Are There 1031 Virus Particles on Earth, or More, or Fewer?." *Journal of Bacteriology* 202, no. 9 (2020).

NAGY, E., & Földes, J. (1991). Electron microscopic investigation of lysogeny of Clostridium difficile strains isolated from antibiotic-associated diarrhea cases and from healthy carriers. *Apmis*, *99*(1-6), 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., Trzepiński, P., Douce, G. R., & Clokie, M. R. (2016a). Bacteriophage combinations significantly reduce Clostridium difficile growth in vitro and proliferation in vivo. *Antimicrobial agents and chemotherapy*, *60*(2), 968-981.

Nale, J. Y., Chutia, M., Carr, P., Hickenbotham, P. T., & Clokie, M. R. (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.

Nale, J. Y., Redgwell, T. A., Millard, A., & Clokie, M. R. (2018). Efficacy of an optimised bacteriophage cocktail to clear Clostridium difficile in a batch fermentation model. *Antibiotics*, *7*(1), 13.

Napolitano, L. M., & Edmiston Jr, C. E. (2017). Clostridium difficile disease: diagnosis, pathogenesis, and treatment update. Surgery, 162(2), 325-348.

Nerber, H. N., & Sorg, J. A. (2021). The small acid-soluble proteins of Clostridioides difficile are important for UV resistance and serve as a check point for sporulation. *bioRxiv*.

Nguyen, A. H., Molineux, I. J., Springman, R., & Bull, J. J. (2012). Multiple genetic pathways to similar fitness limits during viral adaptation to a new host. *Evolution: International Journal of Organic Evolution*, *66*(2), 363-374.

Oechslin, F. (2018). Resistance development to bacteriophages occurring during bacteriophage therapy. *Viruses*, *10*(7), 351.

Ofir, G., & Sorek, R. (2018). Contemporary phage biology: from classic models to new insights. *Cell*, *172*(6), 1260-1270.

O'horo, J. C., Jindai, K., Kunzer, B., & Safdar, N. (2014). Treatment of recurrent Clostridium difficile infection: a systematic review. Infection, 42(1), 43-59.

Orzechowska, B., & Mohammed, M. (2019). The War between Bacteria and Bacteriophages. InTech Publishing.

Park, J. Y., Moon, B. Y., Park, J. W., Thornton, J. A., Park, Y. H., & Seo, K. S. (2017). Genetic engineering of a temperate phage-based delivery system for CRISPR/Cas9 antimicrobials against Staphylococcus aureus. *Scientific reports*, *7*(1), 1-13.

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology*, *31*(9), 839.

Pei, R., & Lamas-Samanamud, G. R. (2014). Inhibition of biofilm formation by T7 bacteriophages producing quorum-quenching enzymes. *Applied and environmental microbiology*, *80*(17), 5340-5348.

Phothichaisri, W., Ounjai, P., Phetruen, T., Janvilisri, T., Khunrae, P., Singhakaew, S., ... & Chankhamhaengdecha, S. (2018). Characterization of Bacteriophages Infecting Clinical Isolates of Clostridium difficile. *Frontiers in microbiology*, *9*, 1701.

Principi, N., Silvestri, E., & Esposito, S. (2019). Advantages and limitations of bacteriophages for the treatment of bacterial infections. *Frontiers in pharmacology*, *10*, 513.

Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., & Chou, C. P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Scientific reports*, *6*(1), 1-15.

Raju, D., Waters, M., Setlow, P., & Sarker, M. R. (2006). Investigating the role of small, acid-soluble spore proteins (SASPs) in the resistance of Clostridium perfringens spores to heat. *BMC microbiology*, *6*(1), 50.

Raju, D., Setlow, P., & Sarker, M. R. (2007). Antisense-RNA-mediated decreased synthesis of small, acid-soluble spore proteins leads to decreased resistance of Clostridium perfringens spores to moist heat and UV radiation. *Appl. Environ. Microbiol.*, *73*(7), 2048-2053.

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

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

Ramírez-Vargas, G., Quesada-Gómez, C., Acuña-Amador, L., López-Ureña, D., Murillo, T., del Mar Gamboa-Coronado, M., ... & Rodríguez, C. (2017). A Clostridium difficile lineage endemic to Costa Rican hospitals is multidrug resistant by acquisition of chromosomal mutations and novel mobile genetic elements. *Antimicrobial agents and chemotherapy*, *61*(4), e02054-16.

Ramírez-Vargas, G., Goh, S., & Rodríguez, C. (2018). The novel phages phiCD5763 and phiCD2955 represent two groups of big plasmidial siphoviridae phages of Clostridium difficile. *Frontiers in microbiology*, *9*, 26.

Ramsay, I., Brown, N. M., & Enoch, D. A. (2018). Recent progress for the effective prevention and treatment of recurrent Clostridium difficile infection. *Infectious Diseases: Research and Treatment*, *11*, 1178633718758023.

Rashid, S. J., Barylski, J., Hargreaves, K. R., Millard, A. A., Vinner, G. K., & Clokie, M. R. (2016). Two novel myoviruses from the North of Iraq reveal insights into clostridium difficile phage diversity and biology. *Viruses*, *8*(11), 310.

Rath, D., Amlinger, L., Rath, A., & Lundgren, M. (2015). The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie*, *117*, 119-128..

Rice, T. B. (1930). The use of bacteriophage filtrates in the treatment of suppurative conditions. *The American Journal of the Medical Sciences*, *179*(3), 345-360.

Riedel, T., Wittmann, J., Bunk, B., Schober, I., Spröer, C., Gronow, S., & Overmann, J. (2017). A Clostridioides difficile bacteriophage genome encodes functional binary toxin-associated genes. *Journal of biotechnology*, *250*, 23-28.

Roach, D. R., & Debarbieux, L. (2017). Phage therapy: awakening a sleeping giant. *Emerging Topics in Life Sciences*, *1*(1), 93-103.

Roach, D. R., Leung, C. Y., Henry, M., Morello, E., Singh, D., Di Santo, J. P., ... & Debarbieux, L. (2017). Synergy between the host immune system and bacteriophage is essential for successful phage therapy against an acute respiratory pathogen. *Cell host & microbe*, *22*(1), 38-47.

Rodriguez, C., Van Broeck, J., Taminiau, B., Delmée, M., & Daube, G. (2016). Clostridium difficile infection: Early history, diagnosis and molecular strain typing methods. *Microbial pathogenesis*, *97*, 59-78.

Rohana, H., Azrad, M., Nitzan, O., Adler, A., Binyamin, D., Koren, O., & Peretz, A. (2020). Characterization of Clostridioides difficile Strains, the Disease Severity, and the Microbial Changes They Induce. *Journal of clinical medicine*, *9*(12), 4099.

Ronda, C., Pedersen, L. E., Hansen, H. G., Kallehauge, T. B., Betenbaugh, M. J., Nielsen, A. T., & Kildegaard, H. F. (2014). Accelerating genome editing in CHO cells using

CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnology and bioengineering*, *111*(8), 1604-1616.

Russell, D. W., & Sambrook, J. (2001). *Molecular cloning: a laboratory manual* (Vol. 1, p. 112). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Saavedra, P. H., Huang, L., Ghazavi, F., Kourula, S., Berghe, T. V., Takahashi, N., ... & Lamkanfi, M. (2018). Apoptosis of intestinal epithelial cells restricts Clostridium difficile infection in a model of pseudomembranous colitis. Nature communications, 9(1), 1-10.

SCHLESS, R. A. (1932). Staphylococcus aureus meningitis: treatment with specific bacteriophage. *American Journal of Diseases of Children*, *44*(4), 813-822.

Scholl, D., Cooley, M., Williams, S. R., Gebhart, D., Martin, D., Bates, A., & Mandrell, R. (2009). An engineered R-type pyocin is a highly specific and sensitive bactericidal agent for the food-borne pathogen Escherichia coli O157: H7. *Antimicrobial agents and chemotherapy*, *53*(7), 3074-3080.

Schooley, R. T., Biswas, B., Gill, J. J., Hernandez-Morales, A., Lancaster, J., Lessor, L., ... & Segall, A. M. (2017). Development and use of personalized bacteriophage-based therapeutic cocktails to treat a patient with a disseminated resistant Acinetobacter baumannii infection. *Antimicrobial agents and chemotherapy*, *61*(10), e00954-17.

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., ... & Holden, M. T. (2006). The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature genetics*, *38*(7), 779-786.

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

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

Sekulovic, O., & Fortier, L. C. (2015). Global transcriptional response of Clostridium difficile carrying the φCD38-2 prophage. *Applied and environmental microbiology*, *81*(4), 1364-1374.

Sekulovic, O., Ospina Bedoya, M., Fivian-Hughes, A. S., Fairweather, N. F., & Fortier, L. C. (2015). The C lostridium difficile cell wall protein CwpV confers phase-variable phage resistance. *Molecular microbiology*, *98*(2), 329-342.

Sekulović, O., & Fortier, L. C. (2016). Characterization of functional prophages in Clostridium difficile. In *Clostridium difficile* (pp. 143-165). Humana Press, New York, NY.

Sell, T. L., Schaberg, D. R., & Fekety, F. R. (1983). Bacteriophage and bacteriocin typing scheme for Clostridium difficile. *Journal of clinical microbiology*, *17*(6), 1148-1152.

Selle, K., Fletcher, J. R., Tuson, H., Schmitt, D. S., McMillan, L., Vridhambal, G. S., ... & Theriot, C. M. (2020). In Vivo Targeting of Clostridioides difficile Using Phage-Delivered CRISPR-Cas3 Antimicrobials. *Mbio*, *11*(2).

Semenova, Ekaterina, and Konstantin Severinov. "Interdependencies between the adaptation and interference modules guide efficient crispr-cas immunity." *Evolutionary Biology: Self/Nonself Evolution, Species and Complex Traits Evolution, Methods and Concepts*. Springer, Cham, 2017. 51-62.

Setlow, P. (1988). Small, acid-soluble spore proteins of Bacillus species: structure, synthesis, genetics, function, and degradation. *Annual Reviews in Microbiology*, *42*(1), 319-338.

Setlow, P. (1992). DNA in dormant spores of Bacillus species is in an A-like conformation. *Molecular microbiology*, *6*(5), 563-567.

Setlow, P. (2006). Spores of Bacillus subtilis: their resistance to and killing by radiation, heat and chemicals. *Journal of applied microbiology*, *101*(3), 514-525.

Setlow, P. (2007). I will survive: DNA protection in bacterial spores. *Trends in microbiology*, *15*(4), 172-180.

Setlow, P., & Kornberg, A. (1970). Biochemical Studies of Bacterial Sporulation and Germination XXIII. Nucleotide Metabolism during spore Germination. *Journal of Biological Chemistry*, *245*(14), 3645-3652.

Sinha, S., Grewal, R. K., & Roy, S. (2018). Modeling Bacteria–Phage Interactions and Its Implications for Phage Therapy. In *Advances in applied microbiology* (Vol. 103, pp. 103-141). Academic Press.

Soutourina, O. A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., ... & Dupuy, B. (2013). Genome-wide identification of regulatory RNAs in the human pathogen Clostridium difficile. *PLoS genetics*, *9*(5).

Summers, W. C. (1999). *Felix dHerelle and the Origins of Molecular Biology*. Yale University Press.

Tamma, P. D., & Suh, G. A. (2021). Phage Are All the Rage: Bacteriophage in Clinical Practice. *Journal of the Pediatric Infectious Diseases Society*.

Topp, S., Reynoso, C. M., Seeliger, J. C., Goldlust, I. S., Desai, S. K., Murat, D., ... & Scott, J. R. (2010). Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl. Environ. Microbiol.*, *76*(23), 7881-7884.

Twort, F. W. (1915). An investigation on the nature of ultra-microscopic viruses. *The Lancet*, *186*(4814), 1241-1243.

Villanueva, V. M., Oldfield, L. M., & Hatfull, G. F. (2015). An unusual phage repressor encoded by mycobacteriophage BPs. *PloS one*, *10*(9), e0137187.

Vogt, S. L., & Finlay, B. B. (2017). Gut microbiota-mediated protection against diarrheal infections. *Journal of travel medicine*, *24*(suppl_1), S39-S43.

Vyas J., Cox J., Setlow B., Coleman W.H., Setlow P.. (2011;). Extremely variable conservation of γ-type small, acid-soluble proteins from spores of some species in the bacterial order Bacillales. J Bacteriol 193: 1884—1892 [CrossRef] [PubMed].

Walker, A. S., Eyre, D. W., Wyllie, D. H., Dingle, K. E., Griffiths, D., Shine, B., ... & Infections in Oxfordshire Research Database (IORD). (2013). Relationship between bacterial strain type, host biomarkers, and mortality in Clostridium difficile infection. *Clinical infectious diseases*, *56*(11), 1589-1600.

Wang, H., La Russa, M., & Qi, L. S. (2016). CRISPR/Cas9 in genome editing and beyond. *Annual review of biochemistry*, *85*, 227-264.

Wang, S., Hong, W., Dong, S., Zhang, Z. T., Zhang, J., Wang, L., & Wang, Y. (2018). Genome engineering of Clostridium difficile using the CRISPR-Cas9 system. *Clinical Microbiology and Infection*.

Waters, E. M., Neill, D. R., Kaman, B., Sahota, J. S., Clokie, M. R., Winstanley, C., & Kadioglu, A. (2017). Phage therapy is highly effective against chronic lung infections with Pseudomonas aeruginosa. *Thorax*, *72*(7), 666-667.

Wetzel, D., & Fischer, R. J. (2015). Small acid-soluble spore proteins of Clostridium acetobutylicum are able to protect DNA in vitro and are specifically cleaved by germination protease GPR and spore protease YyaC. *Microbiology*, *161*(11), 2098-2109.

Williams, R., Meader, E., Mayer, M., Narbad, A., Roberts, A. P., & Mullany, P. (2013). Determination of the attP and attB sites of phage φCD27 from Clostridium difficile NCTC 12727. *Journal of medical microbiology*, *62*(9), 1439-1443.

Winston, J. A., Rivera, A. J., Cai, J., Thanissery, R., Montgomery, S. A., Patterson, A. D., & Theriot, C. M. (2020). Ursodeoxycholic acid (UDCA) mitigates the host inflammatory response during Clostridioides difficile infection by altering gut bile acids. *Infection and Immunity*.

Wittmann, J., Riedel, T., Bunk, B., Spröer, C., Gronow, S., & Overmann, J. (2015). Complete genome sequence of the novel temperate Clostridium difficile phage phiCDIF1296T. *Genome announcements*, *3*(4), e00839-15.

Xue, Y., Zhai, S., Wang, Z., Ji, Y., Wang, G., Wang, T., ... & Gu, J. (2020). The Yersinia Phage X1 Administered Orally Efficiently Protects a Murine Chronic Enteritis Model Against Yersinia enterocolitica Infection. *Frontiers in microbiology*, *11*, 351.

Yehl, K., Lemire, S., Yang, A. C., Ando, H., Mimee, M., Torres, M. D. T., ... & Lu, T. K. (2019). Engineering phage host-range and suppressing bacterial resistance through phage tail fiber mutagenesis. *Cell*, *179*(2), 459-469.

Yip, C., Phan, J., & Abel-Santos, E. (2017). Treatment of Clostridium difficile Infections. In Antibiotic Drug Discovery (pp. 1-19).

Yoichi, M., Abe, M., Miyanaga, K., Unno, H., & Tanji, Y. (2005). Alteration of tail fiber protein gp38 enables T2 phage to infect Escherichia coli O157: H7. *Journal of biotechnology*, *115*(1), 101-107.

Yosef, I., Goren, M. G., Globus, R., Molshanski-Mor, S., & Qimron, U. (2017). Extending the host range of bacteriophage particles for DNA transduction. *Molecular cell*, *66*(5), 721-728.

You, J. H., Jiang, X., Lee, W. H., Chan, P. K., & Ng, S. C. (2020). Cost-effectiveness analysis of fecal microbiota transplantation for recurrent Clostridium difficile infection in patients with inflammatory bowel disease. *Journal of gastroenterology and hepatology*, *35*(9), 1515-1523.

Youngster, I., Russell, G. H., Pindar, C., Ziv-Baran, T., Sauk, J., & Hohmann, E. L. (2014). Oral, capsulized, frozen fecal microbiota transplantation for relapsing Clostridium difficile infection. Jama, 312(17), 1772-1778.

Yutin, N., & Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environmental microbiology*, *15*(10), 2631-2641.

Żaczek, M., Weber-Dąbrowska, B., Międzybrodzki, R., Łusiak-Szelachowska, M., & Górski, A. (2020). Phage Therapy in Poland–a Centennial Journey to the First Ethically Approved Treatment Facility in Europe. *Frontiers in Microbiology*, *11*. d'Hérelle, F., & Smith, G. H. (1922). *The bacteriophage, its role in immunity*. Williams & Wilkins.

Zhang, H., Fouts, D. E., DePew, J., & Stevens, R. H. (2013). Genetic modifications to temperate Enterococcus faecalis phage φEf11 that abolish the establishment of lysogeny and sensitivity to repressor, and increase host range and productivity of lytic infection. *Microbiology*, *159*(Pt 6), 1023.

Zhang, J., Zong, W., Hong, W., Zhang, Z. T., & Wang, Y. (2018). Exploiting endogenous CRISPR-Cas system for multiplex genome editing in Clostridium tyrobutyricum and engineer the strain for high-level butanol production. *Metabolic engineering*, *47*, 49-59.

Zhou, X., Wang, X., Luo, H., Wang, Y., Wang, Y., Tu, T., ... & Zhang, J. (2021). Exploiting heterologous and endogenous CRISPR-Cas systems for genome editing in the probiotic Clostridium butyricum. *Biotechnology and Bioengineering*.

Zhu, D., Sorg, J. A., & Sun, X. (2018). Clostridioides difficile biology: sporulation, germination, and corresponding therapies for *C. difficile* infection. *Frontiers in cellular and infection microbiology*, *8*, 29.

Zuo, T., Wong, S. H., Lam, K., Lui, R., Cheung, K., Tang, W., ... & Ng, S. C. (2018). Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome. *Gut*, *67*(4), 634-643.