University of Nottingham
Biodiscovery Institute

Identification and characterisation of tail components modulating host range of *Clostridium difficile* phages

Joanna Steczynska

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

October 2022

SNIPRBIOME
A CRISPR COMPANY

BBSRC
bioscience for the future

UK Research and Innovation

Declaration

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

Joanna Steczynska, B.Sc. (Hons),
October 2022

Abstract

Clostridium difficile is an anaerobic, spore-forming, pathogenic bacterium that is the leading cause of antibiotic-associated diarrhoea. The infection is treatable with a few antibiotics, however, due to the ever-increasing rates of antibiotic resistance and risk of relapse, novel therapies are in demand. This study focused on advancing the path towards using phages as therapeutic agents against the infection. Genes which are potentially responsible for encoding the receptor binding protein (RBP) found on the phage were investigated. The RBP, encoding the tail fibre, is responsible for garnering the binding specificity of phages. Currently, antibiotics used against the infection are broad spectrum which contributes to the problem, whereas, the phages infecting *C. difficile* present the opposite problem of having a too narrow host range to target all clinically relevant strains with a single phage. This study generated RBP mutants using a CRISPR/Cas9 system in phages ΦCD1801, ΦCD2301, and ΦCD2315 infecting different PCR ribotypes, to confirm the putative RBP gene as responsible for binding specificity. Mutant phages were assayed for their binding ability, with ΦCD1801 imaged under TEM to examine the resulting virion. Further, accessory and structural genes surrounding the RBP were studied in the same manner, through deletion and exchange mutants, swapping various components within the tail module between ΦCD1801 and ΦCD2301 to achieve a host range change. A resulting mutant phage, Ex11, based on ΦCD2301, was able to infect CD1801, the propagating host of ΦCD1801 when carrying its RBP supported by an accessory gene. The tail module was also characterised, establishing the essential and specificity-conferring genes in *Myoviridae*. In addition, novel phages ΦL12S, ΦL12W and ΦL14 were isolated from environmental samples and characterised for their particle morphology, genome content and host range. No identifiable *integrase* gene was found in ΦL14, and the phage was unable to form lysogens *in vitro*, making it the first potentially lytic phage of *C. difficile* to be reported. These findings address the two major problems facing therapeutic application of phages against *C. difficile* and will be instrumental in broadening the host range and overcoming lysogeny.

Acknowledgements

I am very grateful to Prof. Nigel Minton and Dr. Ruth Griffin for their understanding and guidance throughout the whole project. I would like to express particular gratitude to Dr. Michaella Whittle and Dr. Terry Bilverstone for their continued encouragement, and expert knowledge. I want to thank all the past and present members of the SBRC for fostering a great lab environment and continued support. I am eternally grateful to all the lab support staff, especially Magda Lindstrom for always helping me. This work was made possible by UoN's BBSRC Doctoral Training Partnership, so a big thanks to them for the funding and to the team for always being there for their students. This project was also sponsored in part by SNIPR Biome, so a huge thank you for providing the funding, facilitating my internship at the company in Denmark and expert contributions from Dr. Emre Gençay, Dr. Virginia Martinez and (soon to be Dr.) Camilla Kristensen.

Special thanks go to (future Dr.) Sarah Kerr, for just everything and for being my Starbucks/phage partner in crime – I will always do your subcultures for you! I also want to thank (another future Dr.) Temiloluwa Daike for her support, advice and always asking 'why are you here so late?', while staying even later than me. I also want to express gratitude to (also a future Dr.) Jake Yeboah for his emotional support, staying logical in times of stress, all the gym advice and always being ready to debate a new lunchtime topic. I would also like to thank (a future Dr. as-well) Danny McLaughlin for keeping me sane and always getting me through tough times, usually with a takeaway. I could not have done this project without the support of all my friends, and the endless supply of jokes and bucket cocktails they provided, so I am very grateful to all of them.

Finally, I would like to thank my family, particularly my mom and both of my grandmothers – I dedicate this thesis to you. Thank you for always encouraging my pursuit of education and independence. I am grateful you made me stubborn enough to spend four years on this thesis and persist through it.
I hope this makes you proud.

COVID-19 Impact statement

Major parts of the work showcased in this thesis were interrupted by and carried out during the COVID-19 global pandemic. Initially, the laboratory facilities were shut for over 4 months, therefore any bench work was unable to be continued. Following re-opening, the work capacity was severely inhibited by the restrictive rules put in place to prevent any COVID-19 transmission. The laboratory operated on a shift basis, limiting each person to a 7.5-hour shift, thus delaying generation of data for Chapters 5 and 6, as the plaque assay utilised requires at least 8 consecutive hours to run. The shifts switched every week from morning (6.30-14.00) to afternoon (14.30 until 22.00), adding to difficulty. Due to the nature of shift work it was possible to carry out the assays every other week with the help from a colleague, massively slowing down any pace of obtaining data and affecting downstream experiments. In addition, due to social distancing, the use of laboratory facilities was often restricted, extending time-frames for completing experiments during the allocated shift. Also, many staple supplies required for experiments took weeks or months to procure due to global shortages of supplies, further delaying results.

To compensate for loss of laboratory time, the project was re-directed to generation of mutants which would be tested at a later date, as this could be broken down to smaller tasks and did not require prolonged consecutive hours of laboratory access, like the plaque assays. In addition, where possible, additional shifts (which became available later) were utilised to complete the large volume of plaque assays required, particularly for Chapter 6. This resulted in extended lab hours and very long working days, often sacrificing personal comfort.

These circumstances made it difficult to take full advantage of using the laboratory facilities, severely impacting the project progression. In addition, the pandemic took a toll on mental health due to personal circumstances further impacting the work. It was still a great privilege to return to laboratory facilities a lot sooner than many researchers and be able to carry on the work. Nevertheless, the data presented was obtained with much sacrifice and under difficult circumstances, which should be considered when reviewing this thesis.

Abbreviations

%	Percentage
° C	Degrees Celsius
µg	Micrograms
µl	Microlitres
µm	Micrometre
µM	Micromolar
BHI	Brain Heart Infusion
BHIs	Brain Heart Infusion Supplemented
bp	Base Pair
CC	*Clostridium difficile* Selective Supplement
CDI	*Clostridium difficile* Infection
cfu	Colony Forming Units
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Deoxyribonucleic Acid
dH_2O	Distilled water
dsDNA	Double stranded DNA
ELISA	Enzyme Linked Immunosorbent Assay
ESBL	Extended-spectrum beta lactamase
EtOH	Ethanol
FDA	Food and Drug Administration
FMT	Faecal Microbiota Transplantation
g	Grams
g	Gravity
h	Hours
HAs	Homology Arms
ICTV	International Committee on Taxonomy of Viruses
kbp	Kilo Base Pairs
kg	Kilograms
l	Litres
LB	Lysogeny Broth
LHA	Left Homology Arm
mg	Milligrams
min	Minutes
ml	Millilitres
mM	Millimolar
MOI	Multiplicity of Infection
mRNA	Messenger RNA
NEB	New England Biolabs
ng	Nanograms
nm	Nanometre
OD	Optical Density
PaLoc	Pathogenicity Locus
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
pfu	Plaque Forming Units
PMC	Pseudomembranous Colitis
RNA	Ribonucleic Acid
RHA	Right Homology Arm
rpm	Rotations Per Minute
rRNA	Ribosomal RNA
s	Seconds
sgRNA	Single guide RNA
SNP	Single Nucleotide Polymorphism
SOC	Super Optimal Broth
TAE	Tris-acetate-EDTA

TEM	Transmission Electron Microscopy
UK	United Kingdom
US	United States
UV	Ultraviolet
V	Volts
VRE	Vancomycin Resistant Enterococci

Contents

Joanna Steczynska 1
Declaration 2
Abstract 3
Acknowledgements 4
COVID-19 Impact statement 5
Abbreviations 6
Chapter 1: Introduction 13
1. Introduction 14
1.1 Clostridia 14
1.2 *Clostridium difficile* 14
1.2.1 Risk factors and disease epidemiology 15
1.2.2 Hypervirulent strains and outbreaks 17
1.3 Pathogenesis and disease 19
1.3.1 Sporulation 19
1.3.2 Toxins 20
1.3.3 Disease 23
1.4 Diagnosis and treatment 23
1.4.1 Diagnostic methods 23
1.4.2 Treatment 24
1.4.3 Alternative treatments 24
1.5 Bacteriophages 27
1.5.1 Diversity, Morphology and Classification 27
1.5.2 Classification 29
1.6.1 Life cycle 29
1.6.2 Lysogeny 30
1.7 Bacteriophage therapy 31
1.8 Phages of *C. difficile* 32
1.8.1 Phage isolation 32
1.8.2 Temperate phages 32
1.8.3 Characteristics and Morphology of *C. difficile* phages 33
1.8.4 Receptors 33
1.9 Phage therapy in *C. difficile* 34

1.9.1.1 Disadvantages of Lysogeny 34
1.9.1.2 Phage-resistant isolates 35
1.9.2 Advantages of phage therapy in *C. difficile* 36
Chapter 2: Project Aims 38
2. Project Aims 39
Chapter 3: Materials and Methods 40
3. Materials and Methods 41
3.1 Bacterial strains 41
3.2 Microbiology techniques and media formulation 46
3.2.2 Antibiotic supplements 47
3.3 Aerobic strain growth and storage 47
3.3.1 Preparation of Electrocompetent *E. coli* 48
3.3.2 Transformation of *E. coli* through Electroporation 48
3.3.3 Preparation of Chemically competent *E. coli* 48
3.3.4 Chemical transformation of *E. coli* 48
3.4 Anaerobic strains and storage 49
3.4.1 Conjugation of DNA from *E. coli* to *C. difficile* 49
3.5 Phage techniques 50
3.5.1 Plaque assay 50
3.5.2 Phage stock production 51
3.5.3 Phage storage 51
3.5.4 Lysogen isolation 51
3.5.5 Transmission electron microscopy 53
3.5.6 Determination of phage binding to bacterial cells 53
3.5.7 Isolation of phage from environmental samples 53
3.5.8 Screening of environmental filtrate for phage 54
3.6.8 Extraction of Phage Genomic DNA 54
3.6 Molecular techniques 55
3.6.1 Extraction of genomic DNA from *C. difficile* 55
3.6.2 Extraction of plasmid DNA from *E. coli* 55
3.6.3 Restriction digest and dephosphorylation of DNA 55
3.6.4 Plasmid ligation 56
3.6.5 Gel electrophoresis 56

3.6.6 DNA extraction from Agarose gel 56
3.6.7 Polymerase Chain Reaction (PCR) 57
3.6.8 Quantitative Polymerase Chain Reaction (qPCR) 58
3.7 Bioinformatic tools 59
3.7.1 Primer design 59
3.7.2 CRISPR/Cas9 Vector and insert design 59
3.7.3 CRISPR TargetAID (activation-induced cytidine deaminase) 61
3.7.4 Sequencing 61
3.8 Plasmids 62
3.9 Primers 65
Chapter 4: Identification of receptor binding proteins in phages infecting *Clostridium difficile* 71
Chapter 4. Identification of receptor binding proteins in phages infecting *Clostridium difficile* 72
4.1 Introduction 72
4.2 Results 73
4.2.1 Identification of target genes 73
4.2.2 Vector design and construction 76
4.2.3 Addition of a bookmark 78
4.2.4 Optimising DNA transfer and mutagenesis 80
4.2.5 Induction of mutants 89
4.2.6 Testing of mutant phage 90
4.3 Discussion 102
4.3.1 Accuracy of selection of target genes 102
4.3.2 The RBP of ΦCD2315 102
4.3.3 The RBPs of ΦCD1801 and ΦCD2301 105
4.3.4 Summary and importance 106
Chapter 5: Further characterisation of the tail region to achieve a change of host range in phages infecting *Clostridium difficile* 109
Chapter 5. Further characterisation of the tail region to achieve a change of host range in phages infecting *Clostridium difficile* 110
5.1 Introduction 110
5.2 Results 112
5.2.1 Generation and testing of Swap 1-4 mutant phages 112

5.2.2 Vector construction, obtaining and testing the mutant phage for Expansion-1 ... 116
5.2.3 Secondary/ irreversible binding ... 121
5.2.4 Accessory proteins ... 122
5.3 Discussion ... 155
5.3.1 RBP Swapping and Expanding ... 155
5.3.2 Characterising the structural tail gene module ... 157
5.3.3 Host range vs binding range ... 164
5.3.4 Summary and importance ... 166
Chapter 6: Isolation and characterisation of novel phages infecting *Clostridium difficile* from the environment ... 167
Chapter 6. Isolation and characterisation of novel phages infecting *Clostridium difficile* from the environment ... 168
6.1 Introduction ... 168
6.2 Results ... 170
6.2.1 Isolation of phages from the environmental samples against a panel of 027 strains ... 170
6.2.2 Characterisation of isolated phage ... 173
6.2.3 Strain characterisation and Lysogenisation ... 189
6.3 Discussion ... 198
6.3.1 Isolation ... 198
6.3.2 Incomplete infection and host range ... 198
6.3.3 The different host range of ΦL12 ... 201
6.3.4 ΦL14 – A potentially lytic phage ... 202
6.3.5 Summary and importance ... 204
Chapter 7: General discussion and the future of phage therapy in *Clostridium difficile* ... 205
Chapter 7. General discussion and the future of phage therapy in *Clostridium difficile* ... 206
7.1 Aims of the study ... 206
7.2 Future work and potential impact ... 208
Appendix ... 211
8. Appendix ... 212
8.1 Amino Acid alignment of the Hyp14 protein from four *Myoviridae* phages ... 212
8.2 Full list of *C. difficile* strains used in phage isolation – all PCR ribotype 027 ... 213

8.3 Full list of *C. difficile* strains used in host range and EOP testing. 214
Bibliography 217
9. Bibliography 218

Chapter 1:
Introduction

1. Introduction

1.1 Clostridia

Clostridium is a very diverse genus of microorganisms ranging from those which produce industrially useful compounds, to those that cause disease. Its members are Gram-positive anaerobes, regarded for their ability to form endospores. They are commonly found within soil, with some also present in the human and animal intestinal tracts, like *Clostridium butyricum* which is known to play a role in maintaining the gut microbiome (Seki et al., 2003). Industrially relevant members include *Clostridium acetobutylicum*, *Clostridium autoethanogenum, Clostridium beijerinckii* and *Clostridium saccharobutylicum*, which utilize starches to produce acetone, butanol and ethanol in varying ratios (Moon et al., 2016). Other useful species include *Clostridium botulinum* sourced for the production of botulinum toxin, commonly known as Botox, used in aesthetic medicine and treatment of muscle disorders. Despite their commercial usefulness, Clostridia are most recognized for their ability to cause diseases, which can sometimes be fatal. *Clostridium tetani* and *C. botulinum* both produce potent neurotoxins which interfere with actions of neurons, the former leading to tetanus and the latter to botulism.

1.2 *Clostridium difficile*

C. difficile is an anaerobic, Gram-positive, rod-shaped, motile and endospore forming bacterium. First discovered and described by Hall and O'Toole (1935) as *Bacillus difficilis* due to its anaerobic nature, making it difficult to culture and slow growing; it was later renamed to *Clostridium difficile*. Recently, the name sparked debate again, and through genomic analysis it was proposed that the bacterium be renamed to *Peptoclostridium difficile* (Yutin & Galperin, 2013). To avoid public confusion over change from well known "C. diff" to "P. diff", another name was proposed (Lawson et al., 2016) and the bacterium is now known as *Clostridioides difficile*. However, in this document and for the purpose of the study the organisms will be referred to as *Clostridium difficile*.

The bacterium was first isolated from the stool of healthy infants. Asymptomatic carriage of the bacterium in new-borns is abnormally high, compared to the adult population. The number of infants colonized at 1 month of age is 37%, decreasing to 30% at 6 months and finally reaching 0 to 3% at 12 months (Schutze & Willoughby, 2013), which reflects the rate of carriage in adults. Despite the abnormal carriage

rates, disease is rarely reported in infants younger than 12 months. Older children become sensitised to the effects of *C. difficile* toxins at 12 to 24 months of age, as they develop a gut microbiome similar to that of an adult. Disease is not caused even when the pathogen is present at high density, including toxigenic strains, both of which are required for symptoms. Instead, it is thought that infants lack the receptors necessary to mediate the action of toxins (Jangi & Lamont, 2010), and are only affected by disease later in their development.

1.2.1 Risk factors and disease epidemiology

C. difficile is most commonly identified as the cause of antibiotic-associated diarrhoea and responsible for almost all cases of pseudomembranous colitis (PMC). The latter was first described in 1893 (Bartlett, 1994), however, *C. difficile* was pinpointed as its cause over 85 years later by George et al., (1978). The bacterium also causes the less common, but more fatal, toxic megacolon and can lead to colon perforation.

The nosocomial infection is usually prevalent in hospitals and elderly care facilities, where patients are typically undergoing antibiotic treatment for another infection. Their close proximity to each other, increased probability of a weakened immune system due to advanced age and exposure to broad spectrum antibiotics are significant risk factors which contribute to likelihood of disease (Czepiel et al., 2019; Webb et al., 2020).

C. difficile infection (CDI) occurs worldwide, however, in the last decade it has become especially prevalent in developed countries, with the increasing incidence of hypervirulent strains. As a consequence, CDI has replaced the “superbug”, MRSA, as the most common nosocomial associated disease (Miller et al., 2011). However, underdeveloped nations may not be monitoring for incidence of CDI, hence it may be under-reported in such areas. Nearly half a million cases occurred in the US in 2011, with 29,000 deaths. This has also led to an increased burden on the healthcare system with costs increasing from $1.1 billion in a 2002 study (Kyne et al., 2002) compared to a 2015 study with costs of $6.3 billion annually (Levy et al., 2015). The 2011 trend has continued, and largely similar figures for CDI were reported in 2017 (Guh et al., 2020). More recently, *C. difficile* has been named as the most common pathogen to cause healthcare-associated infections in the US, making up 15% of all infections (Magill et al., 2018). Although the rate of incidence of CDI is higher in the

US compared to England, where the frequency has decreased over recent years (A. King et al., 2017), CDI still remains a problem, especially among the geriatric population, which is most at risk. CDI was reported as the 18th leading cause of death in the elderly, in the US with 93% of deaths occurring in people over the age of 65 (Miniño et al., 2010). This can be explained by several factors, such as a weakened immune system at an advanced age, increased amount of exposure to antibiotics due to illness, and recurrent visits with extended stays at healthcare facilities. These factors decrease the patient's defences to *C. difficile* and lead to frequent exposure to the bacterium, leading to a higher chance of infection and mortality.

The overall CDI incidence decrease in the UK, particularly during the recent years, can be partially attributed to the COVID-19 pandemic (Voona et al., 2022) which led to very stringent infection control measures involving frequent hand sanitising and more advanced cleaning protocols and social distancing. This likely halted the spread of many other infectious disease, thus also decreasing CDI incidence. However, according to UKHSA data there has been a notable increase in community associated CDI, responsible for ~40% of cases in 2021/22; a ~10% increase since the 2017/18 period; similar trends have also been observed in the US with incidence shifting towards a higher prevalence in the community (Guh et al., 2020). Therefore, CDI still remains a problem, particularly with the focus shifting to communities outside of the hospital setting, where the prevention measures are not implemented in the same manner.

The main risk factor for CDI is antibiotic therapy, where broad spectrum antibiotics are administered for another infection and in the process, cause gut dysbiosis by disrupting the gut microbiome. Gut dysbiosis is characterised by an imbalance in the commensal microorganisms which usually protect the niche (Toor et al., 2019), – they are largely cleared as a side effect of antibiotic therapy. This allows *C. difficile* to flourish in the niche which is now sparsely populated by commensals as the competition from the other microbes is no longer present in the colon. This event is key in order for CDI to establish itself, leading to growth of vegetative cells from spores, toxin production and disease in the patient. Gut dysbiosis also results in an alteration of bile acid content, leading to germination of spores (Allegretti et al., 2016) and infection. Fluoroquinolones, cephalosporins and carbapenems all increase the risk of CDI, with clindamycin being the main antibiotic implicated as a risk factor

(Luciano & Zuckerbraun, 2014). The level and profile of dysbiosis will vary in each patient mainly depending on which antibiotic therapy thus potentially influencing the level of infection. Overall, a wide range of antibiotics which are broad spectrum can expose the patients to developing CDI. In addition, acid suppression medication used to treat acid reflux, such as proton pump inhibitors and H2 receptor antagonists also increase the risk, although to a lesser extent (Howell et al., 2010). Most CDI occurs in hospitalised patients; however, recently community-acquired CDI has been on the rise and has almost doubled in the last decade. Worryingly, as much as 40% of these cases required hospitalisation due to severity of the infection, likely linked to emerging hypervirulent PCR ribotype 027 and 078 strains. While advanced age remains a key risk, CDI is now more likely to occur in younger patients, especially females and those exposed to infants (Ofori et al., 2018).

C. difficile strains are commonly classed into ribotypes. PCR ribotyping is based on amplification of the intergenic spacer region located between the 16S and 23S rRNA genes. The region varies in length (200-600bp) and paired with different numbers of alleles of the ribosomal operon a particular banding pattern is produced using a single pair of primers, allowing to "type" the strain; recent developments allow to do this directly from samples without isolation of bacteria (Janezic, 2016). The former standard method has been adapted by many institutions in Europe and America especially following outbreaks of the 027 Ribotype strains (Stabler et al., 2009). In the UK, hospital CDI data is collected and analysed by the Clostridioides Difficile Ribotyping Network (CDRN), which has reported an overall decrease in CDI and in prevalence of outbreak ribotype 027 overall (CDRN Report - gov.uk, 2014, 2017-2022); however, it still remains a problem in some regions and in other parts of the world.

1.2.2 Hypervirulent strains and outbreaks

Several outbreaks have taken place around the world and in the UK involving hypervirulent strains that are more infectious and have higher mortality rates. Most notably, strain BI/NAP1/027 first identified in an outbreak in Quebec, was then also implicated in two severe outbreaks at the Stoke Mandeville Hospital in the UK between 2003-2005 (O'Connor et al., 2009). The UK outbreak led to 334 confirmed cases of CDI, including 38 deaths and was one of the most severe to date. The hypervirulent strain is thought to produce higher levels of toxins responsible for

disease due to a mutated regulatory gene, *tcdC*, leading to a non-functional protein (Matamouros et al., 2007). It is believed that the strain acquired resistance to all fluoroquinolones allowing it to spread across the US and Europe due to their wide use in hospitals; their use was the main risk factor during the Quebec outbreak (Pépin et al., 2005). Another PCR ribotype, 078, has emerged as a hypervirulent strain and has a similar toxin profile to the aforementioned strain. Severity and mortality is similar between the two PCR ribotypes, however, 078 affects younger patients and causes community-associated CDI at a much higher rate of 17.5% compared to 6.7% of 027 strains (Goorhuis et al., 2008). Emergence and spread of these hypervirulent strains, partially attributed to use of fluoroquinolones (He et al., 2013), is concerning and has led to mandatory reporting of CDI across the UK in an effort to prevent another outbreak.

Beyond the well-established risk factors of advanced age and increased antibiotic use, the spread of such hypervirulent strains can be explained by a more 'free society'. It is easier and cheaper than ever to travel between different countries, encouraging many people to visit new places, or make more frequent visits to relatives, thus spreading the disease. This reasoning can be applied to many infections; studies support this idea showing travel as a risk factor for CDI (Michal Stevens et al., 2017), particularly community acquired. Additionally, our modern lifestyle including unhealthy diets and consumption of nicotine and alcohol have shown to increase the state of gut dysbiosis (Martinez et al., 2021), the main CDI prerequisite. Collectively our diets consist of foods that have been more processed than ever before (Marino et al., 2021), particularly with the inclusion of unhealthy ingredients and many 'unnatural' additives. Dietary components have been shown to increase CDI-related mortality in a mice model, where the mice were fed a typical 'Western diet' (Hazleton et al., 2022), a high (saturated) fat, low fibre diet representative of the current average US diet composition. Some obscure and less-obvious factors like the inclusion of trehalose as a food additive in the early 2000s have been blamed for global emergence hypervirulent strains of PCR ribotypes 027 and 078 (Collins et al., 2018, 2019). The authors showed that these strains have evolved to metabolise the disaccharide, and thus its increased use in many foods has allow said strains to proliferate; additionally, they used a mouse model to demonstrate enhanced virulence of a 027 strain when trehalose was added. This theory has been largely opposed, as trehalose utilisation is not exclusive to outbreak

strains and even those able to use it, did not confer increased 30-day all-cause mortality in a PCR ribotype 015 model (Eyre et al., 2019). In fact, the authors showed evidence of toxin suppression in 027 strains, further supported by lack of increase in mortality and disease severity in patients infected with strains able to utilise trehalose (Saund et al., 2020). Further findings showed that trehalose remodelled the microbiome preventing CDI in a gut model and when compared to glucose or saline (Buckley et al., 2021). Collectively it was also speculated that the trehalose increase in the gut was too minimal to cause much difference in CDI, particularly as other species present are also able to metabolise it, prompting further investigation as to the real reason for CDI outbreaks and emergence in the community.

1.3 Pathogenesis and disease

1.3.1 Sporulation

C. difficile has the remarkable ability to form endospores, which allows it to survive in tough conditions and persist in the environment. The spores are metabolically inactive with a tough exterior making them resistant to desiccation, heat, radiation, antibiotic treatment and commonly used disinfectants. Due to its anaerobic nature, *C. difficile* isolates cannot exist as vegetative cells in the environment and instead, its spore form is used during transmission between people and objects. Sporulation remains the main vehicle of infection; mutants defective in genes involved in the spore formation pathway were still able to cause disease in animal models, however, transmission cannot take place, confining the infection to its origin (Deakin et al., 2012). Spores from the environment are transmitted to the patient making hospitals the ideal place for the pathogen, where healthcare workers can spread the spores between patients and ineffective disinfectants allow the spores to persist (Dubberke et al., 2014). However, measures have now been implemented in hospitals employing special disinfectants and protocols to limit the spread of the bacterium. Persistent presence of spores contributes to relapse in as many as 25% of hospitalised patients (Ramsay et al., 2018), as they can be re-infected following treatment.

Once the spores enter the gastro-intestinal tract, they travel through the stomach, able to resist the low pH of acid present and germinate in the small intestine. Low levels of oxygen and presence of specific bile acids – cholate derivatives, allows the spores to germinate upon binding to CspC receptors (Sorg & Sonenshein, 2008).

Taurocholate is thought to be the most powerful germinant (Wilson et al., 1982), however, it is not enough to cause germination by itself. Co-germinants in the form of amino acids, mainly glycine or cations, such as calcium and magnesium are required and work most efficiently when combined (Kochan et al., 2018). This is an essential step in the pathogenesis of *C. difficile* as spore formation and germination facilitate disease and spread. This makes *C. difficile* germination an attractive target for therapy.

1.3.2 Toxins

Upon germination, *C. difficile* produces two toxins, TcdA and TcdB, implicated in causing disease and symptoms. Some strains also produce binary toxin, however, its role in disease remains largely unknown. TcdA and TcdB are recognized as major virulence factors in CDI and are some of the largest bacterial toxins studied. The toxins enter the cells via receptor-mediated endocytosis. TcdB is able to enter using chondroitin sulfate proteoglycan 4 (Yuan et al., 2015), Nectin 3 (LaFrance et al., 2015), Frizzled 1/2/7 (Tao et al., 2016) and some glycans (Kordus et al., 2021). Entry mechanism for TcdA is less characterised, although recent developments have shown that sulfated glycosaminoglycans and low-density lipoproteins can act as receptors (Tao et al., 2019), with suggestions that TcdA interacts with multiple receptors at once to allow efficient entry (Kordus et al., 2021). Both toxins are glucosyltransferases and can inactivate the action of Rho, Rac and Cdc42, which are a group of GTPases responsible for assembly and reorganization of the actin cytoskeleton of cells (Tapon & Hall, 1997). Double knock-out of the toxin genes renders strains unable to cause symptoms highlighting their importance in virulence. Single knock-outs still allow either of the toxins to cause disease in animal models through exhibiting their cytotoxic activity (Kuehne et al., 2010).

The toxins are found on PaLoc, a pathogenicity island 19.6kb in size (Figure 1.1), present only in toxigenic strains. The island also contains accessory genes which control the expression of the toxins and allow their release. The gene *tcdC* is highly expressed during the exponential growth phase and as it declines, levels of *tcdA* and *tcdB* gene expression increase, suggesting that it encodes a negative regulator of toxin production. The gene *tcdR* encodes a sigma factor and facilitates toxin production by binding to RNA polymerase and allowing transcription. In addition, it may play a role in modulation of sporulation genes, as mutant spores are more

sensitive to heat and require a higher amount of germinant (Girinathan et al., 2017). Finally, since both toxins lack a signal peptide for secretion, *tcdE* is thought to be responsible for that. It encodes a holin-like protein, similar to that found in phages and allows release of the toxins. *TcdE* mutants were found to have the same level of toxin expression, however, a significantly lower amount of toxins accumulated outside the cells. In addition, when *tcdE* was inserted into a bacteriophage lambda lacking holin, the holin activity was restored in the phage (Govind & Dupuy, 2012). However, despite the similarity to the phage system, TcdE does not compromise the structural integrity of the membrane, as cell contents are not released along with the toxins.

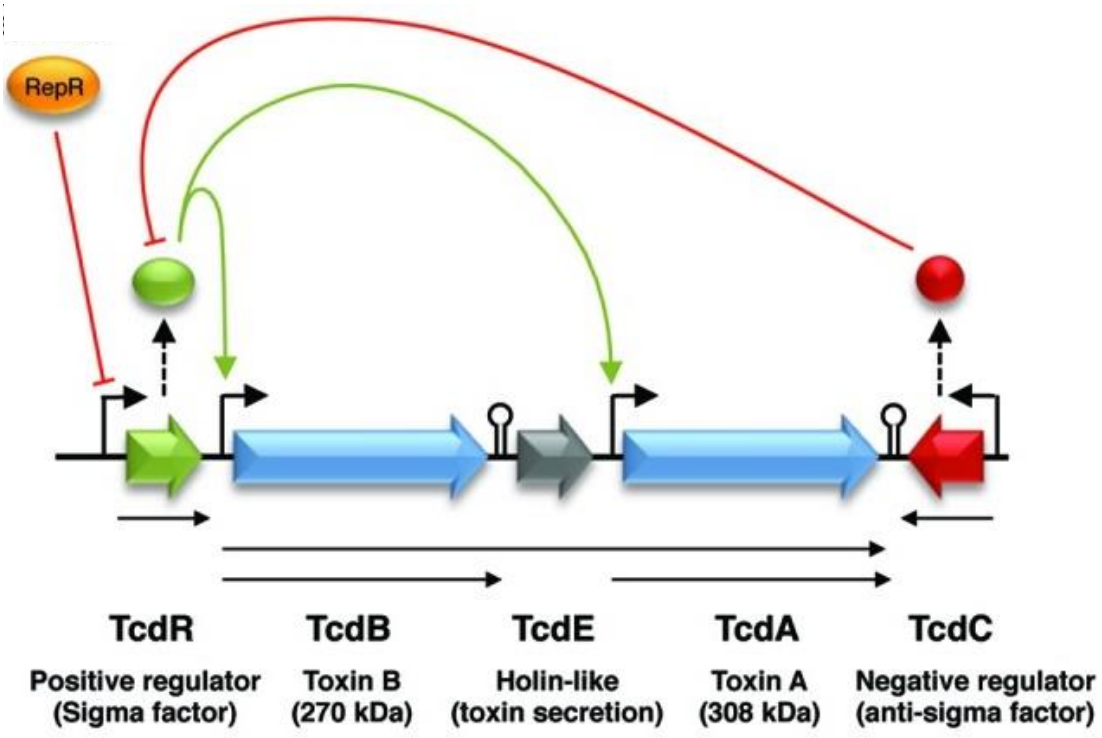


Figure 1.1 PaLoc organization in *C. difficile.*

The diagram shows the order of genes and their regulation, found on the pathogenicity island. *tcdA* and *tcdB* encode toxin gene, *tcdC* is a repressor that prevents transcription of the toxins. *tcdR* upregulates toxin transcription and *tcdE* allows the toxins to be excreted from the cells. RepR is a phage-encoded repressor that downregulates the locus transcription leading to lower toxin levels.

Adapted from Fortier & Sekulovic (2013).

The binary toxin is only found in some strains and was reported in outbreak ribotype 027 strains (O'Connor et al., 2009). Its enzymatic component is encoded by *cdtA* and binding component by *cdtB*, found on the Cdt Locus, outside of PaLoc. This produces an adenosine diphosphate ribosyltranferase protein that, like TcdA and TcdB, causes changes in the actin cytoskeleton leading to its disruption in host cells (Popoff et al., 1988). It is suggested that the binary toxins add to the cumulative effect of the other toxins, however, strains carrying only the binary toxin are unable to cause disease in the hamster model commonly used to study CDI (Geric et al., 2006).

1.3.3 Disease

C. difficile causes a wide spectrum of disease, ranging from diarrhoea to PMC, toxic megacolon and can lead to perforation of the colon. Disease symptoms include fever, abdominal pain and distinctive foul-smelling diarrhoea commonly associated with CDI. PMC is associated with more severe infection and named after the distinctive white plaques that form psuedomembranes on the epithelial surface, caused by the toxins. Following entry into the intestinal epithelial cells, TcdA causes structural disturbances, leading to fluid loss, necrosis and cell death (Hecht et al., 1988). TcdB can exert its cytotoxic effect on a wider range of target cells, also leading to loss of cell integrity and fluids, altering signalling pathways and ending in cell death. Following loss of function of epithelial cells and their apoptosis, white plaques full of cell debris form on the intestinal surface (Voth & Ballard, 2005). In addition, various components of the immune system are directed to site of infection through increased signalling via cytokines and chemokines, leading to massive inflammation and accumulation of neutrophils (Solomon, 2013). Toxic megacolon is a complication of PMC, where the severe inflammation causes neural injury and uncontrolled dilation of the colon. While only up to 3% of patients are affected by this complication (Berman et al., 2008), it is highly lethal, varying from 38% to 80% mortality (Hall & Berger, 2008) and usually requires surgical intervention.

1.4 Diagnosis and treatment

1.4.1 Diagnostic methods

Historically, the most common method for diagnosis of *C. difficile* was the use of cell culture cytotoxicity assays which involve preparing a stool filtrate from a patient sample and applying that filtrate to a monolayer of specific cell lines such as Vero cells or other fibroblasts. Following a 24-48h incubation period, if toxins are present, the cells should exhibit rounding (Planche & Wilcox, 2011). Although cell culture assays were formerly considered as the gold standard, they are labour intensive and time consuming. Through recent developments, they've been surpassed by enzyme-linked immunosorbent assays (ELISAs) with several commercially available testing kits (Shetty et al., 2011). Currently, a mixture of cytotoxic assays due to their accuracy and ELISA detection kits testing for glutamate dehydrogenase due to their quickness are combined for diagnostics.

1.4.2 Treatment

CDI is typically treated with antibiotics, with metronidazole being preferred for mild CDI and vancomycin reserved for more severe cases. Vancomycin is a so-called 'antibiotic of last resort' and the emergence of vancomycin resistant enterococci (VRE) is increasing (Peng et al., 2017), leading to more controlled use. Some have questioned the effectiveness of metronidazole (Gorbach, 1992), while others urge to use it to preserve the effectiveness of vancomycin in more severe cases (Hospital Infection Control Practices Advisory Committee, 1995). For recurrent infections, which take place in up to 25% of patients, reduced vancomycin dosing is preferred and was the only FDA approved treatment for CDI in the US. Both drugs are administered orally for 10 days, with metronidazole dosed at 1-1.5g and vancomycin at 450mg per day.

Another drug has been recently used for treatment of CDI successfully and approved by the FDA. Fidaxomicin is a narrow spectrum antibiotic, leaving the gut microbiome relatively unaffected during administration. Hence, it is thought to cause lower rates of recurrence of CDI when compared to vancomycin, 13.3% and 24%, respectively (Momani et al., 2018), while achieving the same clinical effectiveness. However, it is one of the most expensive antibiotics prescribed with a 10-day course costing upwards of £1,350, compared to much cheaper metronidazole.

1.4.3 Alternative treatments

Gut dysbiosis is a key factor in allowing *C. difficile* to colonize the gut and cause disease. Accordingly, therapies based on preventing it have been employed including, the use of probiotics and faecal microbiota transplants (FMT). Probiotics are microorganisms beneficial to the host and include *Lactobacillus* species and *Saccharomyces boulardii* which contribute to restoring microbial balance. Results from trials have been mixed in terms of preventing CDI, however *S. boulardii* shows some promise in preventing recurrence (Na & Kelly, 2011). FMT shows more potential and was first used in 1958 to treat PMC (Eiseman et al., 1958), long before *C. difficile* was associated with the disease. Success rates with FMT treatment for primary and secondary infections reach cure rates as high as 70 to 90% (Gianotti & Moss, 2017), with the treatment being especially good at preventing recurrence. FMT is potentially likely to be more successful than probiotic treatment as it contains a wider variety of microbes adapted to specifically outcompete *C. difficile*. However,

a major obstacle of FMT is its standing in the public eye and the stigma associated with it. This has led to development of organisations, such as OpenMicrobiome which focus on educating the public through, for example, collaborations with scientific YouTube channels that promote FMT as a treatment and encouraging research into it. Currently the long-term effects of FMT on the patient are unknown and require further study to determine whether FMT is a viable strategy for *C. difficile* treatment. The safety of FMT has also been brought into question following two incidents of transmission of extended-spectrum beta lactamase (ESBL) producing *E. coli* to patients undergoing FMT for a CDI infection (DeFilipp et al., 2019). Tragically one of the patients died due to sepsis, and the other became very sick but recovered. The samples were screened according to FDA mandate, however following this incident, screening protocols should be reviewed and enhanced, particularly as FMT has many benefits and therapeutic potential which should not be dismissed based on this case alone.

Monoclonal antibodies (MABs) – another alternative treatment, which target the toxins themselves thus preventing symptoms have also been employed. Naturally occurring antibodies (post CDI) have exhibited protective effects against disease symptoms by inhibiting the action of toxins and have been implicated in preventing disease recurrence through stabilising the microbiome (Dieterle & Young, 2017). Following development, bezlotoxumab, under the name Zinplava was approved by the FDA in 2017 (Markham, 2016), and then a year later in Europe for prevention of recurrent CDI in adults following phase 3 trials. Initially, two antibodies were developed, including actoxumab and bezlotoxumab targeting TcdA and TcdB respectively; however, development of actoxumab was halted following a higher number of adverse effects and higher mortality observed in the actoxumab only group and lack of increase in efficacy when the MABs were combined. Bezlotoxumab showed lower recurrence (17% compared to 28%) during the MODIFY 1 trial (Wilcox et al., 2017). Analysis following both MODIFY 1 and 2 trials demonstrated that patients only benefitted from decreased recurrence rates if they had at least three risk factors; no recurrence benefit was observed for patients with no predefined risk factors (Gerding et al., 2018). It is an expensive treatment costed at 4,560$ per vial, but only administered as a single 60-minute intravenous infusion at 10mg/kg of body weight (Y. Lee et al., 2017). Cost evaluations (Bainum et al., 2023; Prabhu et al., 2018) have demonstrated justification for its use through showing decreased costs

and burden associated with recurrence of CDI. However, it is not a stand-alone treatment for the infection and is administered alongside antibiotic therapy, not truly addressing the problem of initial CDI, although curbing recurrence thus reducing the burden on the healthcare system.

Another proposed treatment utilizes *Clostridium scindens* as a probiotic to interfere with germination of spores of *C. difficile*. Spore germination is a necessary step for virulence and requires primary bile acids. *C. scindens*, using its enzyme activity, transforms primary bile acids into secondary bile acids which are known inhibitors of *C. difficile* spores (Buffie et al., 2014). Caution and a lot more research is necessary as secondary bile acids are thought to cause some forms of gastrointestinal cancer (Bernstein et al., 2005) necessitating evaluation of the long-term safety of the treatment.

In addition, so-called 'microbiome biotherapeutics', combining bile acid modulation and probiotics have become the latest addition to the market. These commercial products, SER-109 (Seres Therapeutics Inc, USA) and RBX2660 (Rebiotix, USA) are in phase 2 and 3 clinical trials, and show some promise, especially against recurrence (Dubberke et al., 2018; McGovern et al., 2021). The former, SER-109 is a formulation of five species from the Firmicutes phylum, which are spore forming bacteria that are able to compete for nutrients with *C. difficile* and modulate bile acid profiles of patients thus restoring colonisation resistance and supressing spore germination. This is crucial as antibiotics only target live bacteria, leaving the spores unaffected and able to germinate following completion of antibiotic treatment for CDI (Theriot & Young, 2015). In a recent phase 3 clinical trial patients with recurrent CDI (three or more episodes) were given SER-109 or a placebo (four capsules per day, over three days); the trial showed much lower levels of further recurrence at 12% for the treatment group, compared to 40% for the placebo group (Feuerstadt et al., 2022). The species associated with the treatment were detected early on leading to an altered bile acid profile associated with inhibition of spore germination. The treatment did not show any adverse safety concerns beyond those experienced in the placebo group and was effective regardless of age and prior antibiotic treatment (Sims et al., 2023). While SER-109 shows much promise, it still only addresses the recurrence prevention rather than treatment of CDI itself and does not completely eliminate recurrence.

1.5 Bacteriophages

Bacteriophages (phages) are viruses that can infect and kill bacteria. They were discovered at the beginning of the 20th century, first by Frederick Twort in 1915, then again independently reported by Félix d'Hérelle in 1917. D'Hérelle identified the organism as a phage, literally meaning "eater of bacteria" in Greek after he observed clear plaques where bacteria were killed on a plate. Phages require hosts to replicate like any viruses; and in the process they disrupt the normal functions of bacteria through hijacking their molecular machinery to express viral proteins, forming virions and eventually bursting the cell to release phage progeny. Phages are present where their bacterial hosts are found and are distributed throughout the environment, making them the most numerous entities on Earth reaching up to 10^{31-33} (Hatfull & Hendrix, 2011). They outnumber bacteria 10 to 1 and it is estimated that half of the bacterial population is killed every 48h (Rohwer et al., 2009).

1.5.1 Diversity, Morphology and Classification

Phages have existed for over 3 billion years leading to huge diversity between them in genome size, shape, their complexity and structure. Phage tails typically range in length from 20nm to 250nm, however, many fall outside of that spectrum. Phage genomes can be single or double stranded and linear or circular in shape, using RNA or DNA to encode genetic information (Gelderblom, 1996). Their genome sizes vary considerably, from 2.4kb in L5 *Leuconostoc* phage to 316.6kb in *Pseudomonas* 201phi2−1 phage (Hatfull, 2008), even reaching almost 500kb in phage G found in *Bacillus megaterium* (Hatfull & Hendrix, 2011). Their genomic diversity is remarkable and not a single gene is shared across all annotated phage genomes. Typically, they are mosaics of each other, if infecting the same species, due to many gene reshuffling events that have recurred over time. Their genome organisation often shows genes grouped into modules by their function; such that the genes encoding the capsid would be grouped together, as would the genes encoding the tail and the genes encoding replication and integration-related proteins (Belcaid et al., 2010). It has been proposed that double stranded DNA (dsDNA) phages share common ancestry and can exchange their gene modules making many of the phages mosaics of genes from a shared common gene pool (Hendrix et al., 1999).

Structural genes dictate phage morphology. The phage head (capsid) consists of monomers of the same protein and encapsulates the genetic information that is stored inside it, protecting it until it is delivered (Fokine et al., 2005). The head is connected to the tail, sometimes by a neck and collar; the tail is a hollow tube that facilitates the passage of genetic material from the head to the host cell (Hofer, 2016). The tail is connected to a baseplate and has tail fibres attached to it (Figure 1.2) which allow cell receptor recognition and binding (Le et al., 2013). The fibre function can be fulfilled by a tail fibre, a short tail fibre, a long tail fibre or a tail spike, depending on the infected species and phage. This is thought to dictate host specificity of phages, as they can only recognise a range of receptors. The tail length is determined by a tail tape measure protein, with its length being proportional to the length of the tail. Binding of the tail triggers the release of nucleic acid from the phage and its entry into the host (White et al., 2007).

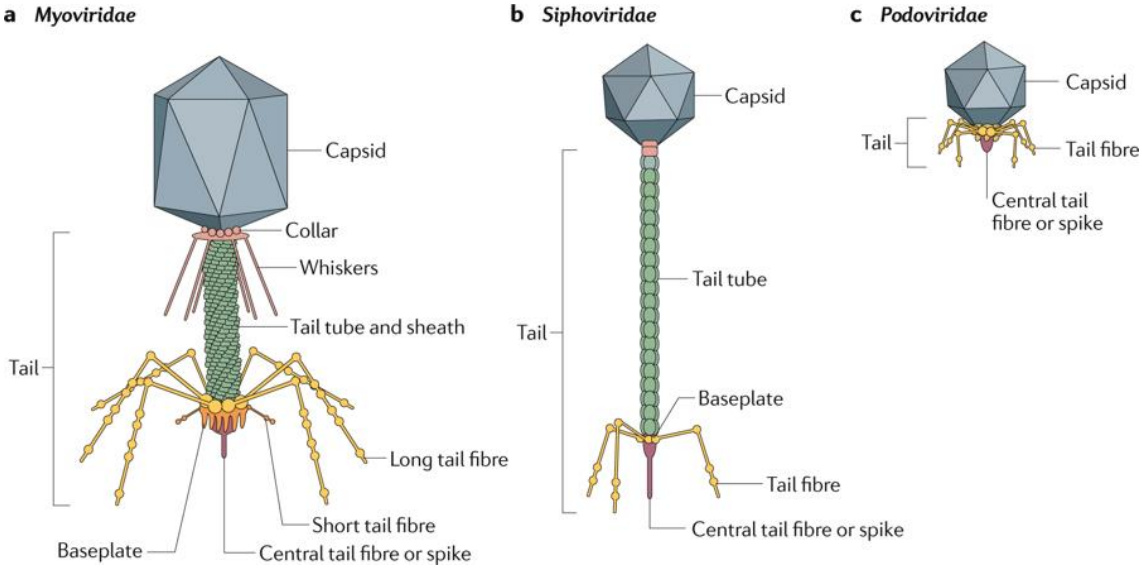


Figure 1.2 Structures of different phage families

The figure shows the three phage families, belonging to the order *Caudovirales*. The figure highlights differences and similarities in structure between the different families. All phages have a capsid and tail with fibres. The *Podoviridae* lack a tail tube and are relatively small in comparison. The *Myoviridae* have longer tails and more components like the collar connecting the capsid and tail. The *Siphoviridae* have the longest tails, and unlike *Myoviridae* lack the collar and tail sheath, which allows their tails to be flexible and narrow.

Adapted from Nobrega et al. (2018)

1.5.2 Classification

Most characterised phages belong into the order *Caudovirales* and fall into three families*: Podoviridae*, *Myoviridae* and *Siphoviridae* (Figure 1.2). Cauda meaning tail in Latin, these phages are tailed with dsDNA and therefore fall under group I of the Baltimore scheme (Ackermann, 1998). The *Podoviridae* are characterised by having very short (about 20nm in length) and non-contractile tails which typically have six subterminal fibres (ICTV, 2012). In contrast, the *Myoviridae* are a larger family characterised by their longer, rigid tails enveloped by contractile sheaths, connected to the head by a neck. Finally, the *Siphoviridae* have very long flexible tails, some over 400nm in length (Fortier & Moineau, 2007), that lack a contractile sheath. Tails are connected directly to the capsid, without a collar protein and end in a terminal spike, with subterminal fibres attached to the tail end (Maniloff & Ackermann, 1998). However, many phages remain undiscovered and out of those observed, a majority cannot be cultured – this is attributed to the great plaque anomaly (Serwer *et al.*, 2007), as the culture conditions remain unknown. This remains an obstacle in phage classification, and the International Committee on Taxonomy of Viruses (ICTV) is always working to classify new phages as they are discovered and characterised.

1.6.1 Life cycle

Phages initiate the infection cycle by binding to the specific host receptors and following conformational changes, inject their genetic material into the bacterium. There are two main life cycles that phages can follow from this point – lytic and lysogenic (Figure 1.3) (Kutter & Sulakvelidze, 2005). Strictly lytic phages hijack bacterial ribosomes to translate viral mRNA to proteins, following transcription. Some phages can induce the bacterial RNA polymerase to preferentially transcribe it over host DNA through the use of promoters and proteins. This ensures expression of viral genes, usually divided into early and late stage genes (Yang et al., 2014), based on function. Once viral proteins are synthesised, virions are assembled, and the genetic material is packaged. The host cells then burst releasing viral progeny, either due to the pressure or phage-encoded endolysins that break down peptidoglycan lysing the cell (Young et al., 2000), allowing the cycle to continue.

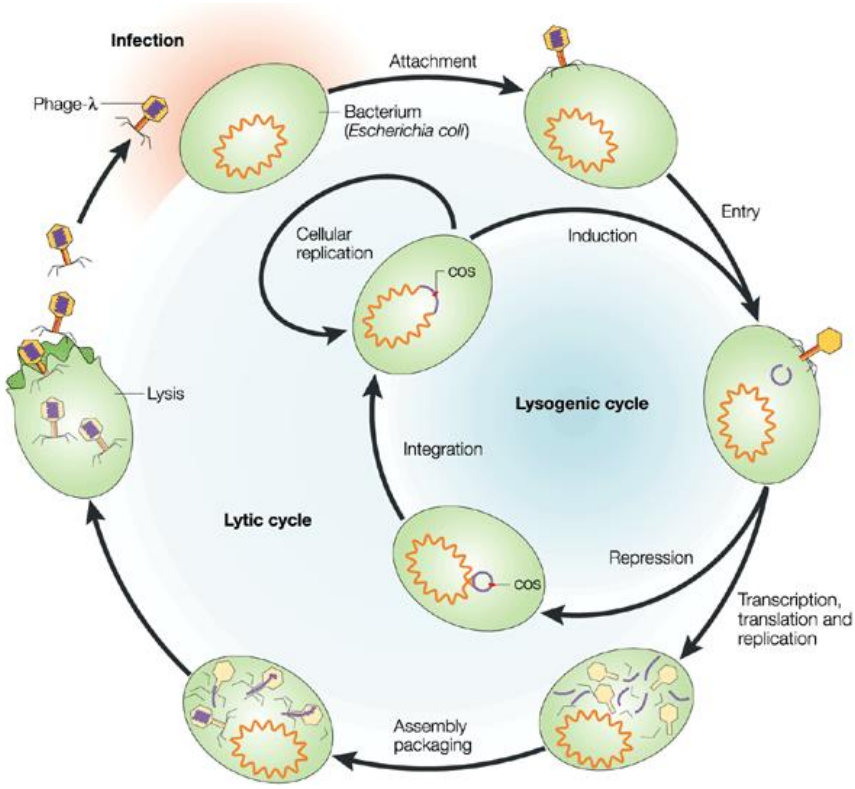


Figure 1.3 Phage Life Cycle

The figure shows the two life cycles of lytic and temperate phages. Lytic phages inject their genome into the bacterium, replicate and synthesise new progeny, then cause cell lysis to release new virions. Temperate phages can follow this cycle, however they can also interrupt it and integrate into the host genome instead, following the lysogenic cycle. During this, the phage genome will be replicated with the bacterial chromosome until conditions are correct.

Taken from Campbell (2003).

1.6.2 Lysogeny

Temperate phages have a choice of following the lytic cycle, or the lysogenic cycle – where their genome is maintained in the host cell post-injection and replicated with the cell, without expressing or synthesising viral proteins (Campbell, 2003). The DNA can be inserted into the host chromosome or maintained in the cytoplasm. The phage exists as a prophage in this state and the bacterium becomes known as a lysogen (lysogenised by the phage). This prevents other phages, from the same immunity group from infecting the lysogen (Ptashne, 2004). These temperate phages can choose which life cycle to follow and this decision is determined by factors like

availability of hosts and nutrient conditions, although specific mechanisms remain largely unknown for many phages.

Recently, it has been discovered that phages can communicate with their progeny and influence their lysis-lysogeny decision through using a signal peptide. The Arbitrium system used by *Bacillus subtilis* phages encodes a communication peptide, which is excreted into the surrounding environment and initiates lysogeny when its concentration is high, as this signals to phage progeny that the number of hosts is low (del Sol et al., 2019). Different versions of this peptide are encoded by different phages, showing this communication to be specific to each phage (Erez et al., 2017). Lately it has also been reported that some phages, like those infecting *Vibrio cholerae*, are able to "listen" to host quorum sensing and detect host peptides, which allows them to modulate their lysis-lysogeny cycle accordingly (Silpe & Bassler, 2019). Many of these systems are only now being discovered and in the future synthetic manipulation of lysis-lysogeny may be possible.

1.7 Bacteriophage therapy

Félix d'Hérelle is considered the father of phage therapy and was the first to use it to cure dysentery in a 12 year old boy in France in 1919. The phage preparation was ingested by d'Hérelle himself and his interns to confirm its safety before being administered to the boy. Following his recovery, more patients were treated successfully (Sulakvelidze et al., 2001). D'Hérelle went on to treat cholera and bubonic plague in India, and commercialised several phage preparations in his laboratory in France, which later became the well-known cosmetics company, L'Oréal (Wilkinson, 2001). Others followed suit and phage preparations were being produced by companies like Eli Lily, against many pathogens to treat conditions like: wound abscesses, respiratory infections and ear infections. However, the war and discovery of antibiotics overshadowed phage therapy, and it was largely abandoned in the West. Its development carried on in the former Soviet Union, where it is still used in conjunction with antibiotics and several institutions for phage therapy exist, like the Eliava Institute in Georgia and Hirszfeld Institute in Poland.

1.8 Phages of *C. difficile*

1.8.1 Phage isolation

Phages infecting *C. difficile* were first isolated by Sell and colleagues through induction with Mitomycin C on ten clinical isolates for the purpose of phage typing (1983). Phages were isolated with some characterisation and used to study the epidemiology of CDI in human and animal models (Bacon et al., 1988; Hawkins et al., 1984). Later, Dei (1989) found that this method was unreliable as a high proportion of isolates tested were not sensitive to the available phage preparations and phage typing was replaced by other molecular techniques. Although, many phages have been discovered in the intervening years they have all proven to be temperate and strictly lytic phages have yet to be isolated. The process of isolating phages has proven to be difficult for *C. difficile* and initially, techniques attempting to isolate free phages from the environment and patient samples were unsuccessful (Goh et al., 2005; Sekulovic et al., 2014). Application of enrichment methods has allowed this hurdle to be overcome and lead to phage isolation from patient stool samples (Meessen-Pinard et al., 2012).

Many phages have been isolated through induction from lysogenised clinical isolates rather than as free phages from the environment. Temperate phages can be induced to excise from the bacterial genome and enter the lytic cycle in several ways. Most commonly, Mitomycin C and norfloxacin can be used to induce lysis (Nagy and Földes, 1991; Ramesh, Fralick and Rolfe, 1999; Horgan, et al., 2010; Govind, Fralick and Rolfe, 2011; Shan et al., 2012), as well as applying UV light (Sekulovic et al., 2014). These are known to induce DNA damage and prompt the phage to make its escape. Interestingly, Nale et al., (2012) found that Mitomycin C and norfloxacin were able to induce different phages from the same isolates in four cases. The isolated phages have been characterized by their various features, using transmission electron microscopy (TEM), whole genome sequencing, strain susceptibility, burst size experiments and phage attachment experiments (M. Clokie & Kropinski, 2009).

1.8.2 Temperate phages

Lack of lytic phages is a major obstacle for traditional phage therapy. It has been suggested that due to the spore-forming nature of *C. difficile* and high frequency of prophage genomes present, temperate phages are more prevalent (Hargreaves et al., 2013). However, lytic phages have been isolated for other spore-forming bacteria

like *Bacillus cereus* (El-Arabi et al., 2013) and others, including clostridial species, for example *Clostridium perfringens* (Volozhantsev et al., 2012). Both studies used large screening libraries containing many strains, and it has been suggested this is a necessity when isolating phage (Hargreaves & Clokie, 2014), especially when screening for lytic phages which may be less prevalent in *C. difficile*. This is because sporulation would prevent lytic phages from surviving as they require metabolically active cells and spores are dormant. Therefore, as temperate phages exist within the genome, they survive until the spore germinates. However, it has been observed that lytic phages infecting *B. subtilis* are able to survive sporulation as their DNA is maintained in the spore as free circular DNA and upon germination, the lytic cycle ensues (Sonenshein, 2006).

1.8.3 Characteristics and Morphology of *C. difficile* phages

All isolated *C. difficile* phages belong to *Caudovirales* and fall under the *Myoviridae* or *Siphoviridae* families; no *Podoviridae* phages have been isolated to date. All have dsDNA, with genome sizes ranging from 31kb to over 131kb. Most, however, fall at the lower end of the spectrum, at <60kb. The GC content is low, at around 26 to 31%. Some gene cassettes are shared between different phages, suggesting recombination and shuffling has occurred (Hargreaves & Clokie, 2014). As is typical of phages, the genome is organized in modules which encode genes for a specified function. For example, the genes encoding tail proteins are often organized together on the genome. The genomes all contain integrase-encoding genes thus confirming them as temperate phages (Hargreaves & Clokie, 2014). Phage DNA can be packaged using headful packaging, as is true for the first *C. difficile* phage sequenced, ΦCD119 (Govind et al., 2006) or by other methods where the genome ends are defined. During headful packaging, concatemers of dsDNA are formed and cut by a terminase once the capsid is full of DNA, leading to different genome ends (Garneau et al., 2017).

1.8.4 Receptors

To initiate the phage life cycle, phages need to bind to the host cells to inject their DNA. They use receptor binding proteins (RBPs) which are thought to be the tail fibre proteins (Rothenberg et al., 2011). The fibres can present in many and sometimes multiple forms, as short or long tail fibres and even tail spikes. Phage particles diffuse around until they encounter a bacterial cell, if their receptor is specific to the host,

they will bind, sometimes using conformational changes and inject their genome into the host (Le et al., 2013). The bacterial receptor bound by phage is thought to be the slpA, encoded S-layer protein A (Whittle et al., 2022). *C. difficile* phages are known to be extremely specific, as many infect as little as single strains, while some may infect most members of a particular PCR ribotype (Sekulovic et al., 2014). It is a significant problem when attempting to isolate phages, as large libraries of hosts are still required and even then, phages may be missed if their specific host is not present.

1.9 Phage therapy in *C. difficile*

1.9.1.1 Disadvantages of Lysogeny

Lysogeny is a potential problem in using phage therapy as it leads to several issues that could impact efficacy and safety. As discussed previously (1.3.2), it has been suggested that PaLoc is of phage origin due to close sequence similarity of *tcdE* and phage *holin* gene (Govind & Dupuy, 2012). This raises concerns about phages being a mode of horizontal gene transfer between *C. difficile* isolates and increasing their virulence through advantageous genes delivered during therapy. It has been thoroughly documented that phages are able to transduce genes between bacteria, with antimicrobial resistance genes being of particular concern (Colavecchio et al., 2017; Leclerc et al., 2022). Although some speculate that this effect is largely overemphasized and virome analysis (Enault et al., 2016) as well animal studies (Allen et al., 2011) have revealed that this may be quite a rare occurrence. Regardless, this problem will need to be addressed in both lytic and lysogenic phages alike and level of horizontal gene transfer will need to be evaluated for each phage to ensure safety.

Another problem is that lysogenized bacteria are not killed and the prophage prevents secondary infection or superinfection by related phage (Kutter & Sulakvelidze, 2005). As a consequence, lysogenized cells remain and continue to produce toxins. Bacteria that also already contain prophages before the therapy is applied would potentially be immune if the phage preparation contains related phage. To overcome this, a phage preparation containing a cocktail of phages with enough diversity to counter the lysogeny effect is required.

Finally, the integration of the phage itself could prove to be problematic. The insertion site can differ depending on isolate and can therefore interrupt genes or

their regulation, including toxin genes; this is comparable to insertion of virus genomes in human cells in certain genes where their dysregulation can lead to cancer. This effect has been reported in ΦCD38.2, where the lysogenized bacteria showed increased transcription of the PaLoc genes and a 1.6- and 2.1-fold increase in TcdA and TcdB production, respectively (Sekulovic et al., 2011). Interestingly, other studies have reported the opposite effect. Transcription of PaLoc is downregulated in strains lysogenized by ΦCD119; the prophage expresses a RepR repressor protein which can bind to *tcdR* within PaLoc and decrease the amount of toxin produced (Figure 1.1) (Govind et al., 2009). Another study found toxin production to be reduced in bacteria lysogenized by ΦCD27, possibly by another repressor used to maintain lysogeny with the ability to modulate toxins as a side effect (Williams et al., 2013). This effect seems largely phage specific and may not be true for all phages, therefore this concern may be misguided, however phages used in therapy should be examined for this phenomenon. As it currently stands, the FDA guidance only allows strictly lytic phages to be used in clinical trials, therefore this needs to be considered in regards to this therapy.

1.9.1.2 Phage-resistant isolates

One of the main concerns in phage therapy is the development of escape mutants, which are bacteria that become resistant to further phage infection rendering the therapy ineffective. Some experimental data has suggested that up to 80% of studies and some human trials examining phage treatments in the gut showed phage-resistant variants emerging (Oechslin, 2018). This is a result of the constant arms race between phage and bacteria, driving microbial evolution (Koskella & Brockhurst, 2014). Often, whether through spontaneous or targeted mutations, bacteria will be able to modify their cell surface to remove the receptor used by the phages for entry; for example, LPS modifications (Bertozzi Silva et al., 2016). This however often comes with fitness costs to the bacterium, often reducing their virulence, particularly when it comes to components like the LPS (León & Bastías, 2015). Maintenance of other common anti-phage defenses like RM systems (Vasu & Nagaraja, 2013) and CRISPR-mediated adaptive immunity (Vale et al., 2015) also add burden to the bacterial host. The bacteria often lose their virulence factors, become more susceptible to antibiotic treatment (Chan et al., 2016) or clearance by the immune system (Foster, 2005; Smith & Huggins, 1982).

Many animal studies have shown that *in vivo* treatment with phage led to a selection for phage-resistant isolates, however in these cases the treatment of animals was still somewhat to mostly successful depending on specific phage application (Oechslin, 2018); typically the faster the therapy was applied, like prior to onset of symptoms, the more successful the treatment and less potential for mutants (Smith & Huggins, 1983). Following this study, the authors addressed the concerns of escape mutants by a series of *in vitro* experiments which generated escape mutants; those were then screened for susceptibility to other phages – this was led to a dual-phage approach (Kortright et al., 2019). Two phages were applied as a treatment due to anticipation of resistance to one phage emerging creating a more effective therapy. This approach is now seen in modern phage treatment of infections, including CDI where multi-phage cocktails have been shown to be more effective at clearing infection compared to single phage preparations (Nale, Spencer, et al., 2016). Prophylactic approaches proved a lot more effective at prevention, rather than treatment proved at clearance and survival (Nale, Chutia, et al., 2016), however this may not always be possible in clinic and more studies with optimization are needed to better treat developed infections.

1.9.2 Advantages of phage therapy in *C. difficile*

Phage therapy research has returned to the spotlight in recent years due to the ever-rising emergence of antibiotic resistance (Ventola, 2015) leading to their decreased effectiveness and emergence of VRE (Cetinkaya et al., 2000) as well as many other multidrug-resistant bacteria. The threat of antimicrobial resistance has never been greater and phage therapy is one of the ways to overcome it, as the search for new antibiotics has proven to be slow. Phages are in an evolutionary arms race with bacteria (Stern & Sorek, 2011) and while bacteria evolve resistance mechanisms against them, the phages evolve to overcome those mechanisms, thus avoiding problems with resistance.

Phages also have the ability to effectively clear biofilms, which *C. difficile* can form and antibiotics will struggle to penetrate. It is advantageous that phages usually encode enzymes for biofilm clearing, however, little is known about *C. difficile* biofilms and their contribution to pathogenesis (Vuotto et al., 2018). Phages are self-limiting and self-sustaining, therefore will only be present while bacteria are there, with a small initial dosage required.

Phages often have narrow host ranges. This is true for *C. difficile*, where specificity is observed at the sub-species level. This will therefore allow the microbiome to recover whilst the phage specifically targets the pathogen. The opposite is true for broad-spectrum antibiotics and adverse effects to the microbiome have been reported, especially when Metronidazole has been administered (Saginur et al., 1980). When compared to Metronidazole in a CDI infection model, phages had no negative effect on gut commensals and a reduction in *C. difficile* numbers was detected (Meader et al., 2010). Phages, therefore, have great potential for prevention of relapse in patients by maintaining gut microbial complexity. However, the very narrow host range is also a double-edged sword, since the specificity of phages means many more need to be discovered to target all the clinical strains of *C. difficile*, and production of such complex phage cocktails is not desirable.

To date, no phage therapy against CDI has been developed, due to phages being largely ignored in the West and CDI being less of a problem where phages are used. *In vivo* studies have demonstrated potential effectiveness in fermentation models (Nale et al., 2018) and hamster models of CDI, where hamsters treated with phage were able to survive a challenge with *C. difficile* (Ramesh et al., 1999). Advancement in the field has been limited due to the challenging nature of the work, especially involving phage isolation and lack of strictly lytic phages. This leaves room for development of novel therapies involving the use of semi-synthetic phages which have broader host range, are able to deliver lethal cargo and eradicate the infection.

Chapter 2:
Project Aims

2. Project Aims

The aim of this project is to contribute to creating a therapeutic phage preparation that can be used to treat infections caused by *C. difficile* in the future. This study is undertaken to elucidate the receptor binding proteins (RBPs), phages use for binding to their corresponding receptors on bacterial cells. This basal knowledge will hopefully allow engineering of semi-synthetic phage, with broad host ranges that can be used in therapeutic applications. The study is also carried out to expand the existing pool of phages available to study and adapt. This will be done through these objectives:

1. Characterise previously isolated phages in terms of their RBPs through the use of a CRISPR/Cas9 knock-out system and establish function by testing mutants in binding assays.
2. Design a semi-synthetic phage with modified RBPs and if necessary, accessory genes, to change its host range and test activity against a panel of *C. difficile* isolates.
3. Isolate new phages from environmental samples using a panel of clinically relevant *C. difficile* strains. Resulting phages will be characterised in terms of their particle morphology, genome sequence and their host range.

Chapter 3:
Materials and Methods

3. Materials and Methods

3.1 Bacterial strains

The colour-coded tables below contain a list of all the existing strains used (Table 3.1), and a list of all the strains generated (Table 3.2) in this study.

Table 3.1 List of strains used.

Strain	Description	Source
	Escherichia coli	
Top10	Cloning and plasmid storage strain	Invitrogen, UK
CA434	Plasmid storage and conjugation donor strain; *E. coli* HB101 with R702 conjugation factor	University of Nottingham, N. Minton (Purdy et al., 2002)
NEB sExpress (sX)	Plasmid storage and conjugation donor strain; *E. coli* NEB Express with R702 conjugation factor from CA434	University of Nottingham, N. Minton
TopSex (Ts)	Plasmid storage and conjugation donor strain; *E. coli* Top10 with R702 conjugation factor from CA434	University of Nottingham, N. Minton
NEB 10-Beta	Cloning strain used for large vector construction	New England Biolabs
NEB 10-Stable	Cloning strain used for vector construction with toxic elements	New England Biolabs
	Clostridium difficile	
CD1801	Host for ΦCD1801 phage used for plaque assays and phage propagation.	University of Nottingham, N. Minton
CD1801L	Lysogenized strain of CD1801, containing ΦCD1801 phage. Used for phage storage and phage RBP investigation.	University of Nottingham, N. Minton, M. Whittle
CD2315	Host for ΦCD2315 phage used for plaque assays and phage propagation.	Isolated from: Hungary Leiden University Medical Centre (LUMC) E. Kuijper
CD2315L	Lysogenized strain of CD2315, containing ΦCD2315 phage. Used for phage storage and phage RBP investigation.	University of Nottingham, N. Minton, M. Whittle
CD2301	Host for ΦCD2301 phage used for plaque assays and phage propagation.	University of Nottingham, N. Minton

Strain	Description	Source
CD2301L	Lysogenized strain of CD2301, containing ΦCD2301 phage. Used for phage storage and phage RBP investigation.	University of Nottingham, N. Minton, M. Whittle
CD08011	Host for ΦCD08011 phage used for plaque assays and phage propagation.	University of Nottingham, N. Minton
CD08011L	Lysogenized strain of CD1801, containing ΦCD08011 phage. Used for phage storage and phage RBP investigation.	University of Nottingham, N. Minton, M. Whittle
CD522418	Host for ΦCD418 phage used for plaque assays and phage propagation.	University of Nottingham, N. Minton
CD522418L	Lysogenized strain of CD522418, containing ΦCD418 phage. Used for phage storage and phage RBP investigation.	University of Nottingham, N. Minton
L12	Used in isolation of novel phages from environmental samples.	Isolated from: Germany (LUMC) E. Kuijper
L14	Used in isolation of novel phages from environmental samples. Host for ΦL14.	Isolated from: Switzerland (LUMC) E. Kuijper
L15	Used in isolation of novel phages from environmental samples. Host for ΦL15.	Isolated from: Denmark (LUMC) E. Kuijper
JS50	Host for ΦL12W and ΦL12S. Unknown strain mistaken for and assumed to be L12. A strain from the SBRC Culture collection was banked assuming to be L12 – was renamed to be JS50. Used for propagation of phage.	Unknown

Table 3.2 List of strains generated.

Strain	Description	Source
Clostridium difficile		
CD1801L TCFP Ko BM6	Used to obtain the TCFP Ko mutant, with a bookmark scar	This study, Chapter 4
CD1801L TCFP Comp	Used to complement the TCFP mutant	Chapter 4
CD2315L TPPSP Ko	Used to obtain the TPPSP Ko mutant	Chapter 4
CD2315L TPPSP Ko BM6	Used to obtain the TPPSP Ko mutant, with a bookmark scar	Chapter 4
CD2315L TPPSP Comp	Used to complement the TPPSP mutant	Chapter 4
CD2315L Kelch v3 Ko	Used to obtain the Kelch Ko mutant	Chapter 4
CD2315L Kelch TA	Used to obtain the Kelch TA mutant, single bp change made using TargetAID	Chapter 4
CD2315L DKO BM6	Used to obtain the double knock-out of TPPSP and Kelch genes, with a bookmark scar	Chapter 4
CD2301L PTP Ko BM6	Used to obtain the PTP Ko mutant, with a bookmark scar	Chapter 4
CD2301L PTP Comp	Used to complement the TCFP mutant	Chapter 4
CD522418L TCFP Ko BM6	Used to obtain the TCFP Ko mutant, with a bookmark scar	Chapter 4
CD1801L Swap 1	Used to obtain the Swap 1 mutant	This study, Chapter 5
CD2301L Swap 2	Used to obtain the Swap 2 mutant	Chapter 5
CD2301L Swap 3	Used to obtain the Swap 3 mutant	Chapter 5
CD2315L Swap 4	Used to obtain the Swap 4 mutant	Chapter 5

Strain	Description	Source
CD1801L Expand 1	Used to obtain the Expand 1 mutant, with an additional RBP	Chapter 5
CD1801L Del1	Used to obtain the Deletion 1 mutant	Chapter 5
CD1801L Del2	Used to obtain the Deletion 2 mutant	Chapter 5
CD1801L Del3	Used to obtain the Deletion 3 mutant	Chapter 5
CD1801L Del4	Used to obtain the Deletion 4 mutant	Chapter 5
CD1801L Del5	Used to obtain the Deletion 5 mutant	Chapter 5
CD1801L Del6	Used to obtain the Deletion 6 mutant	Chapter 5
CD2301L Del7	Used to obtain the Deletion 7 mutant	Chapter 5
CD2301L Del8	Used to obtain the Deletion 8 mutant	Chapter 5
CD2301L Del9	Used to obtain the Deletion 9 mutant	Chapter 5
CD2301L Del10	Used to obtain the Deletion 10 mutant	Chapter 5
CD2301L Del11	Used to obtain the Deletion 11 mutant	Chapter 5
CD2301L Del12	Used to obtain the Deletion 12 mutant	Chapter 5
CD1801L Ex1	Used to obtain the Exchange 1 mutant	Chapter 5
CD1801L Ex2	Used to obtain the Exchange 2 mutant	Chapter 5
CD1801L Ex3	Used to obtain the Exchange 3 mutant	Chapter 5

Strain	Description	Source
CD1801L Ex4	Used to obtain the Exchange 4 mutant	Chapter 5
CD1801L Ex5	Used to obtain the Exchange 5 mutant	Chapter 5
CD1801L Ex6	Used to obtain the Exchange 6 mutant	Chapter 5
CD2301L Ex7	Used to obtain the Exchange 7 mutant	Chapter 5
CD2301L Ex8	Used to obtain the Exchange 8 mutant	Chapter 5
CD2301L Ex9	Used to obtain the Exchange 9 mutant	Chapter 5
CD2301L Ex10	Used to obtain the Exchange 10 mutant	Chapter 5
CD2301L Ex11	Used to obtain the Exchange 11 mutant	Chapter 5
CD2301L Ex12	Used to obtain the Exchange 12 mutant	Chapter 5
JS50 ΦL12W NP1	Lysogenized strain of JS50, isolate NP1 containing ΦL12W phage. Used for phage storage.	Chapter 6
JS50 ΦL12S.4	Lysogenized strain of JS50, isolate .4 containing ΦL12S phage. Used for phage storage.	Chapter 6
L35 ΦL12W W11	Lysogenized strain of L35, isolate W11 containing ΦL12W phage. Used for phage storage.	Chapter 6
L35 ΦL12S S10	Lysogenized strain of L35, isolate S10 containing ΦL12S phage. Used for phage storage.	Chapter 6

3.2 Microbiology techniques and media formulation

Media and other solutions used in this study are listed in Table 3.3 below. The media components were dissolved in distilled water with amounts adjusted to the volume required, then autoclaved at 121°C for 20 minutes. If required, antibiotics or other supplements were added to the media after autoclaving, but before pouring plates. If the media solidified, it was melted and cooled to 50°C prior to addition of antibiotics or supplements. Unless otherwise stated, media components were purchased from Oxoid and chemicals purchased from Sigma.

Table 3.3 Media

Medium	Components	Quantity (g/L)
Lysogeny Broth (LB)	Tryptone Yeast Extract Sodium Chloride	10 5 10
LB Agar	Tryptone Yeast Extract Sodium Chloride (NaCl) Agar Bacteriological No.1	10 5 10 15
Brain Heart Infusion Supplemented (BHIs) Broth	BHI Yeast Extract L-Cysteine Hydrochloride Monohydrate	37 5 1
BHIs Agar	BHI Agar Yeast Extract L-Cysteine Hydrochloride Monohydrate	47 5 1
BHIs Top Agar (0.5%)	BHI Yeast Extract L-Cysteine Hydrochloride Monohydrate Agar Bacteriological No.1	37 5 1 5
BHIs Top Agar (0.2%)	BHI Yeast Extract L-Cysteine Hydrochloride Monohydrate Agar Bacteriological No.1	37 5 1 2
BHIs Top Agar Plus (0.2%)	BHI Yeast Extract L-Cysteine Hydrochloride Monohydrate Agar Bacteriological No.1 Calcium chloride ($CaCl_2$) Magnesium chloride hexahydrate ($MgCl_2 \bullet 6H_2O$)	37 5 1 2 0.555 1

Medium	Components	Quantity (g/L)
Phosphate Buffered Saline (PBS), pH 7.4	NaCl KCl $NaHPO_4$ $KH2PO_4$	8.00 0.20 1.44 0.24
Environmental phage isolation media	BHI Yeast Extract L-Cysteine Hydrochloride Monohydrate Sodium Taurocholate Magnesium chloride hexahydrate ($MgCl_2 \bullet 6H_2O$)	27 5 1 1 10.2

3.2.2 Antibiotic supplements

Antibiotics were prepared as concentrated stock solutions and stored for no longer than a month at 4°C. During preparation, before use, the stocks were filter sterilised using a 0.22µm filter (Millipore). The working concentrations used are listed in Table 3.4 below.

Table 3.4 Supplemented antibiotics.

Antibiotic	Solvent	Stock concentration (mg/ml)	Working Concentration (µg/ml)
Chloramphenicol	100% EtOH	25	*E. coli*, Broth – 12.5 *E. coli*, Plates – 25
Thiamphenicol	50% EtOH + 50% Water	15	*C. difficile* - 15

3.3 Aerobic strain growth and storage

Required *E. coli* strains were grown in LB broth or cultured on plates at 37°C overnight, for 16-18 hours, with liquid cultures agitated at 200rpm. The growth media was supplemented with antibiotics when required to only select for strains harbouring relevant plasmids, usually conferring chloramphenicol resistance. *E. coli* strains were stored at -80°C on beads using Microbank™ vials (Pro-Lab Diagnostics). When required, a bead from the appropriate strain stock was added to 5ml of LB broth supplemented with chloramphenicol and incubated at 37°C with agitation.

3.3.1 Preparation of Electrocompetent *E. coli*

Appropriate *E. coli* strain was grown overnight in 5ml LB broth, at 37°C with agitation. A 1% culture inoculum was added to a sterile flask containing 200ml of pre-warmed LB broth and incubated at 37°C with agitation. The OD_{600} was checked after three hours and subsequently after until the culture reach OD_{600} of 0.5-0.7. The cells were harvested by centrifugation at 3709 x *g* for 10 minutes at 4°C, then the pellet was washed twice with ice cold water. Finally, the pellet was resuspended in 1ml of 1mM MOPS (4-Morpholinepropanesulfonic acid) solution containing 10% glycerol, and 50µl aliquots were stored at -80°C.

3.3.2 Transformation of *E. coli* through Electroporation

Electrocompetent cells were thawed on ice gently. Plasmid (100ng) or dialysed ligation mixture (15µl) was added to thawed cells and mixed carefully. This mixture was then transferred to an electroporation cuvette, 0.2cm (Biorad) and pulsed with an electroporator (Biorad Gene Pulser) set to 200Ω, 2.5kV and 25µF capacitance. Following the pulse, 400µl of Super Optimal Catabolite (SOC) medium was added to the cuvette with cells and then transferred to a 1.5ml tube. The cells were recovered for two hours at 30°C before plating in several dilutions on LB agar plates supplemented with chloramphenicol. The plates were incubated for 48 hours at 30°C, until colonies appeared and could be screened.

3.3.3 Preparation of Chemically competent *E. coli*

Appropriate *E. coli* strains were grown as described previously until OD_{600} of 0.5 – 0.7 was reached. The cells were harvested and left on ice for 15 minutes, then centrifuged at 1709 *g*. The pellets were washed twice in 20ml of 100mM calcium chloride and left on ice for 30 minutes in between the washes, then resuspended in 2ml of ice cold 100mM calcium chloride with 15% glycerol. The prepared cells were stored in microfuge tubes in 100µl aliquots at -80°C.

3.3.4 Chemical transformation of *E. coli*

Competent cells were transformed with assembled plasmids or ligation mixtures according to New England Biolabs High efficiency transformation protocol. Briefly, the cells were thawed on ice for 10 minutes, then mixed with 100ng of plasmid or 15µl of ligation mixture. The cells were left on ice for 10 minutes, then heat shocked at 42°C for 30 seconds, then further placed on ice for 5 minutes. Finally, 950µl of SOC

was added and the cells were left to recover with agitation for two hours at 30°C, then plates in several dilutions on LB agar plates supplemented with chloramphenicol for 48h at 30°C.

3.4 Anaerobic strains and storage

Required strains of *C. difficile* were cultured overnight in an anaerobic cabinet (Don Whitley Scientific Ltd 80% N_2, 10 CO_2, 10% H_2) at 37°C. The strains were streaked on BHIs agar plates supplemented with D-cycloserine (250µg/ml) and cefoxitin (8µg/ml); (*C. difficile* selective supplement, Oxoid, (Heap et al., 2009)), referred to as BHIsCC plates. Overnight liquid cultures were set up in BHIs broth with no antibiotic selection, from plated bacterial stocks, and grown under anaerobic conditions at 37°C. Strain stocks were created from overnight growth on plates, supplemented with antibiotics, if selection was required. Growth was harvested with a sterile 10µl loop and resuspended in 1ml of BHIs broth containing 20% glycerol. Strains were stored at -80°C and revived on plates as needed. Plates were reduced in anaerobic conditions for at least four hours prior to use.

3.4.1 Conjugation of DNA from *E. coli* to *C. difficile*

Assembled plasmids were shuttled from *E. coli* donors (described in Table 1) into *C. difficile* using a protocol adapted from Purdy et al., (2002). A 5ml of the *E. coli* donor strain, containing the plasmid, was grown overnight and 1ml of it was harvested through centrifugation at 1377 x *g* for 1 minute. The cells were washed twice in phosphate-buffered saline (PBS) and the pellet was transferred into the anaerobic cabinet. The *E. coli* pellet was resuspended gently in 200µl of overnight *C. difficile* cultured in 1ml of BHIs broth. The mixture was plated onto BHIs agar plates, with no selective supplement, in 20µl spots and left for 24h to allow mating. Following that, the growth was collected using a 10µl sterile loop and resuspended in PBS, then 100µl of it was plated onto five BHIsCC plates supplemented with thiamphenicol (BHIsCCTm) to select for *C. difficile* colonies containing the plasmid. The plates were incubated for 72h at 37°C under anaerobic conditions and resulting transconjugants were re-streaked onto selection plates to confirm they were not background colonies. The colonies were then induced to express the vector components by plating on BHIsCC plates supplemented with thiamphenicol and theophylline (48mg/ml) and resulting potential mutants were screened by PCR and saved.

3.4.1.1 Optimised conjugation

An alternative modified conjugation protocol was also used with some strains upon encountering difficulty when conjugating with the standard protocol described in 3.4.1 which resulted in no transconjugant colonies. Both, *E. coli* donor and *C. difficile* target strains were grown under the same conditions. However, the *E. coli* cells containing appropriate plasmids were harvested through centrifugation at 1217 x *g* for 2 minutes and 30 seconds, followed by washing the pellet twice. The pellet was then resuspended in 200µl of *C. difficile* overnight and plated on a BHIs agar plate in 15-20 small mating spots for either 8h or 24h depending on the strains used. As per protocol, following transfer to selection plates, these were incubated for 72h or more until potential transconjugants appeared with the rest of the standard protocol followed as described above.

3.4.1.2 Plasmid loss

Once mutants of relevant *C. difficile* strains were obtained and confirmed via sequencing, the strains were cultured to lose the CRISPR/Cas9 plasmid to prevent uncontrolled mutations and strains re-verting to the original version. Relevant mutant strains were cultured overnight in 10ml of BHIs broth then subcultured (10ul into 10ml of fresh broth) for 4-7 days, typically 5 days was enough to lose the plasmid, however this varied by strain. To screen for mutants, several dilutions were plated on BHIsCC plates from day 4 onwards, then after overnight growth on plates, the colonies were picked using a 1ul loop and patch plated on a grid of 52, first on BHIsCC to save the colony and then on BHIsCCTm to screen for plasmid loss. If the colonies did not grow on BHIsCCTm, but grew on BHIsCC they were deemed as having lost the plasmid and were screened further by repeated BHIsCCTm streak and sequencing to ensure the mutant was intact after repeated subculturing.

3.5 Phage techniques

3.5.1 Plaque assay

Phages were assayed for plaque forming unit (p.f.u.) counts using a double agar overlay plaque assay adapted from Clokie et al., (2009). *C. difficile* strains were subcultured, where a 1% of overnight culture was transferred to 20ml of pre-reduced BHIs broth and incubated for 6h at 37°C under anaerobic conditions. 1ml of this culture was mixed with 200µl of phage (dilution) and 3ml of 0.5% BHIs top agar kept at 50°C. The mixture was poured to form an even layer on BHIsCC base agar plates.

The phage and bacteria were mixed in the cabinet, then the molten top agar was added on the bench, where the plates were left to set for 10 minutes, then transferred to the anaerobic cabinet. Plates were incubated overnight at 37°C and counts were observed the next day, calculating p.f.u./ml based on plates containing 2-100 plaques for accuracy.

3.5.2 Phage stock production

3.5.2.1 Titre increase using double overlay

To produce phage stocks and increase phage titres, assays were used. The procedure is the same as described for a plaque assay for preparing the double overlay agar plates, except a softer 0.2% (or 0.2% Plus, depending on the phage) top agar was used. Plates which showed almost confluent lysis were scraped into 10ml of pre-reduced BHIs broth under anaerobic conditions, shaken well and left for the phages to diffuse out into liquid, for two hours. The phage preparation was centrifuged at 10000 x *g* for 10 minutes to pellet the agar and bacterial cells. The supernatant was sterilised using a 0.22µm filter and stored at 4°C.

3.5.2.2 Induction using Mitomycin C

Required *C. difficile* lysogen strains containing phages were cultured overnight at 37°C under anaerobic conditions. The next day Mitomycin C was added to the culture at 3µg/ml and left to induce for at least 16-24h. The following day the induced cultures were centrifuged at 10000 x *g* for 10 minutes to pellet the bacterial cells and debris. The supernatant was sterilised using a 0.22µm filter and stored at 4°C. Titre was checked as required.

3.5.3 Phage storage

The phage preparations were stored in BHIs at 4°C and routinely checked through plaque assays for viability, with fresh stocks made as required. As a backup, lysogenized strains of *C. difficile* were stored at -80°C.

3.5.4 Lysogen isolation

Lysogen on plate

Lysogenized strains containing the phage are needed for long term storage of phage and to make phage mutants. Lysogens were isolated using an on-plate method described by (Govind et al., 2009). Briefly, plates containing a lawn of *C. difficile* on 0.5% BHIs top agar, (as described in 3.5.1), excluding the addition of phage. In the

cabinet, high titre phage preparations were spotted onto the lawn in three spots and incubated overnight at 37°C. Following day, using a 10µl loop colonies growing in the phage spot were picked and re-streaked onto a BHIsCC plate. Then, 20 colonies were picked and further purified through re-streaking three times. To confirm these as lysogens, colony PCR was used as well as induction of prophages and testing for immunity to phage infection by plaque assay.

Lysogen from plaque assay

As described in 3.5.1, a standard plaque assay was carried out, with the addition of phage, diluting down to 10^6. Next day, plaques were observed and picked with a tip, then re-streaked onto a BHIsCC plate. From the re-streaks, 20 colonies were selected and passaged for three days. Then the screening process was carried out as above.

Lysogen from neat plate

As described above, after an assay was carried out, the undiluted plate was left in the anaerobic cabinet for an additional day, as usually that is the confluent lysis plate where it appears clear of any lawn or plaques. After an additional day colonies developed, and should be lysogens. The screening process was carried out as above.

Cross-stitch lysogen

Required *C. difficile* strains were cultured overnight at 37°C under anaerobic conditions, then a 10µl loop was used to make several horizontal lines with the grown culture across a BHIsCC plate. After the plate dried for 10 minutes, several 10µl aliquots of a phage stock were spotted in pre-marked spaces and left to dry. The next day, slight growth developed where the phage was spotted and was harvested and re-streaked on a BHIsCC plate, with 20 colonies selected forward for further screening following a 3 day passage.

Lysogen in liquid

C. difficile strains were cultured overnight at 37°C under anaerobic conditions and a 1% inoculum was transferred to 5ml of fresh, pre-reduced BHIs broth. After two hours, 5ml of high titre (10^7 p.f.u./ml) phage preparation was added. Samples were taken 2, 4, and 6h after the phage was added and screened for presence of lysogens. Lysogens were further confirmed by colony PCR and saved as stocks for further validation.

3.5.5 Transmission electron microscopy

Before examination of phage particles using transmission electron microscopy (TEM), the particles were prepared using ammonium acetate precipitation described by (Fortier & Moineau, 2007). Like in the protocol, 1.5ml of high titre phage preparation was centrifuged at 21000 x *g* for 75 minutes, then 1.4ml of the supernatant was removed taking care to not touch the bottom of the tube. The pellet was resuspended with 1ml of 0.1M ammonium acetate (Sigma); the wash step was repeated twice and the precipitated phage was stored at 4°C. A 13µl aliquot of precipitated phage was placed on a 200 mesh copper grid with carbon film (EM Resolutions) and left to coat the grid for 5 minutes. Excess liquid was blotted with Whatman paper and a 2% uranyl acetate stain was added to the grid for 30 seconds for negative staining. Again, the excess liquid was removed and the grid was dried briefly, before imaging the grid under TEM.

3.5.6 Determination of phage binding to bacterial cells

Relevant *C. difficile* strains were subcultured, where a 1% of overnight culture was transferred to 20ml of pre-reduced BHIs broth and incubated for 4h at 37°C under anaerobic conditions. The 20ml culture was harvested by centrifugation at 10000 x *g* for 10 minutes at room temperature and the pellet was re-suspended in 500µl of BHIs broth. A 100µl aliquot of the resuspended cells was centrifuged again at 3,637 x *g* for 2 minutes; the supernatant was discarded and the cells were resuspended with 10µl of the relevant phage preparation. The phage-cell mixture, alongside phage only controls were incubated for 15 minutes at 37°C under anaerobic conditions to allow for phage binding to the cells. Following the binding period, the samples were resuspended in 1ml of BHIs broth if necessary, and centrifuged at 14,549 x *g* for 2 minutes to pellet the bound phage and cells. The supernatant was sterilised using a 0.22µm filter membrane and the resulting sample was enumerated using a standard plaque assay to determine the phage titre in the filtrate. The titre is indicative of the number of free phage particles/ml and effectively demonstrates whether phage binding occurs when compared to phage-only control samples.

3.5.7 Isolation of phage from environmental samples

Sewage samples were collected from the Oxford sewage treatment plant, provided by Severn Trent Water in May 2019. 14 samples were provided from 7 different sewage treatment plant locations: Aylesbury, Didcot, Wargrave, Swindon, Banbury,

Oxford and Reading. Each location donated 500ml of primary during-digestion sludge and 500ml of secondary post-digestion sludge. Samples were transported on ice and stored at 4°C until required. A 50ml aliquot of each sample was mixed with phage isolation media components (described in Table 3.3) and incubated overnight in an anaerobic cabinet at 37°C. The following day, 40ml of each sample was centrifuged at 10 000 x g for 10 minutes at 4°C to pellet debris. The remaining supernatant was carefully poured off and sterilised using a 0.22µm filter membrane. Resulting filtrate was stored at 4°C, ready for screening for presence of phage.

3.5.8 Screening of environmental filtrate for phage

The filtrates from each location were screened for presence of phage using a panel of strains. The strains were grown in 20ml cultures in BHIs broth overnight, under anaerobic conditions at 37°C, then subcultured for 6h to be used in a plaque assay described in 3.5.1. In place of 200µl of phage, 200µl of the filtrate was used. When potential phage plaques were observed the next day, three plaques were picked using a sterile pipette tip from each plate, resuspended in 900µl of BHIs for an hour under anaerobic conditions, then filter sterilised. The resulting phage suspension was then passaged two more times through further plaque assays to ensure phage purity by picking individual plaques each time. The pure stock was then enumerated using the protocol described in 3.5.2 for phage stock production.

3.6.8 Extraction of Phage Genomic DNA

Phage nucleic acid was extracted from phage lysate preparations, preferably high titre (>10^8 pfu/ml), if possible. The RBC Bioscience Viral Nucleic Acid Extraction kit (YVN100) was initially used per the manufacturer's instructions. Briefly, 200µl of phage lysate was mixed with 400µl of VB Buffer and incubated for 10 minutes at room temperature. The sample was mixed with ethanol, then passed through a binding column, which was then washed with W1 Buffer twice and dried. Centrifugation was carried out at 10 000 x *g*, for 1 minute each time. The DNA was eluted in 50µl of RNase-free water and stored at 4°C to prevent fragmentation.

Phage genomic DNA was extracted for sequencing with the NorgenBiotek Phage DNA Isolation Kit (Cat. 46800) using a modified protocol. 15-20ml of high titre (>10^9 pfu/ml), phage lysate was used, split into individual 1ml preparations. The lysate was treated with NorgenBiotek's RNase-Free DNase I Kit; 10µl of DNase and 40µl of

corresponding buffer was added into each preparation and incubated for 15minutes at room temperature, then heat inactivated at 75°C for 5 minutes. This step removes the host genomic DNA contamination, as at this point phage DNA is protected inside capsids. 500µl of Lysis Buffer B was added and the samples were vigorously vortex-mixed for at least 10 seconds. Following, 20µl of Proteinase K was added to each preparation and incubated for 25 minutes at 55°C to remove protein contamination and increase DNA yields. Further incubation step was carried out for 15 minutes at 65°C, and each tube was inverted 2-3 times during incubation. The samples were mixed with isopropanol and bound to spin columns through centrifugation at 6 000 x *g* for 1 minute. The columns were washed three times with Wash Buffer A mixed with ethanol, then dried. Each column was eluted from twice, resulting in 50-70µl of phage DNA suspended in Elution Buffer. The DNA was stored at 4°C until required.

3.6 Molecular techniques

3.6.1 Extraction of genomic DNA from *C. difficile*

Genomic DNA was extracted from 1ml overnight cultures of appropriate strains using a GenElute™ Bacterial Genomic DNA Kit (Sigma), according to the manufacturer's instructions. Briefly, the overnight culture was harvested by centrifugation and resuspended in lysozyme (10mg/ml) for 30 minutes. The cells were lysed with Proteinase K (20mg/ml) and lysis solution C for 10 minutes at 55°C. Then, DNA was bound to a column and eluted using dH_2O and stored at -20°C.

3.6.2 Extraction of plasmid DNA from *E. coli*

Plasmid DNA was extracted from a 5ml overnight culture of *E. coli* grown at 37°C with agitation grown in LB with chloramphenicol. Monarch® Plasmid Miniprep Kit (NEB) was used, according to manufacturer instructions. Briefly, 5ml of culture was harvested by centrifugation, the cells were resuspended and lysed. The DNA was bound to a column, washed twice and eluted using dH_2O and stored at -20°C.

3.6.3 Restriction digest and dephosphorylation of DNA

Restriction enzymes and buffers from NEB were used to perform digests according to manufacturer's instructions. DNA (1µg) was digested with 1µl of each enzyme required and added to CutSmart® Buffer (1X final concentration), with final reaction volume at 20µl made up with dH_2O. Digestions were incubated at 37°C for at least four hours, up to overnight for plasmids. Digested plasmid vectors were

dephosphorylated to prevent self-ligation using Antarctic Phosphatase (NEB). 1µl of the enzyme was mixed with 1µl of its corresponding buffer (10x concentrated) and added to the digestion reaction, then incubated for 30 minutes to an hour at 37°C.

3.6.4 Plasmid ligation

Purified and digested vectors and DNA fragments (usually PCR products) were mixed in various ratios, using 50ng of vector, 2µl of T4 DNA Ligase (Promega) and 2µl of its associated buffer (10x), then made up to 20µl with dH_2O. Reactions were incubated overnight on ice. If required (for electroporation) ligation mixtures were dialysed to remove excess salts on 0.025µm discs (Millipore) on dH_2O for 30 minutes at room temperature. Alternatively plasmids were assembled using the NEBuilder® HiFi DNA Assembly Master Mix according to the manufacturer's instructions for sgRNA-Cas9 vectors.

3.6.5 Gel electrophoresis

Tris-acetate-EDTA (TAE) buffer was diluted to 1x and used to create a 1% agarose (Sigma) solution. SYBR™ Safe (Thermo Fisher Scientific) DNA Gel stain was added to the gel casting trays prior to adding the agarose solution. Quick-Load® 1 kb Plus DNA Ladder, 0.1-10kb, (NEB), was used as a marker. DNA fragments were mixed with 6x Loading Dye (NEB) and loaded onto the gel. The gel was run in TAE buffer at 100V for 45-60 minutes, then visualised using Gel Doc™ XR+ Gel Documentation System (BioRad) with images captures using the corresponding software, Image Lab™.

3.6.6 DNA extraction from Agarose gel

After visualization, correct bands were excised from the gel using a scalpel, taking care to only excise the relevant parts. The bands were purified using Monarch® DNA Gel extraction kit (NEB), according to the manufacturer's instructions. Each gel fragment was dissolved in four volumes of Dissolving Buffer at 50°C, until the fragment was fully solubilised (around 15 minutes). The solution was passed through the column to bind the DNA and wash the sample with supplied wash buffers. Following centrifugation (at 13000 x *g*), the samples were eluted in 6-20µl of dH_2O, then quantified using NanoDrop™ Lite Spectrophotometer. The samples were stored at -20°C until required.

3.6.7 Polymerase Chain Reaction (PCR)

Primers for DNA amplification were manually designed and produced by Sigma as 100µM stocks. The primers were stored at -20°C on arrival. PCR was performed using Q5® High-Fidelity 2X Master Mix (NEB) or DreamTaq Green PCR 2X Master Mix (Thermo Fisher Scientific) according to the manufacturer's instructions with volumes listed in Table 3.5. Both are master mixes, therefore only primers and DNA had to be added. Annealing temperatures were adjusted for each primer pair using the NEB online calculator for Q5 and decreased by 5°C for DreamTaq (Table 3.6).

Table 3.5 PCR Reaction mix components.

Component	Q5®/ DreamTaq
DNA Template	1µl
Forward/Reverse Primer	1µl/each
Master Mix	10µl
Nuclease Free Water	Up to 20µl

Table 3.6 PCR Reaction steps.

Step	Temperature (°C)		Duration		No. of cycles
	Q5®	DreamTaq	Q5®	DreamTaq	Q5®/ DreamTaq
Initial Denaturation	98	95	30s	10 min	1
Denaturation	98	95	10s	30s	30
Annealing	50-72	Tm-5	30s		
Extension	72		30s/kb	1 min/kb	
Final Extension	72		2 min	5 min	1
Hold	15		∞		∞

3.6.8 Quantitative Polymerase Chain Reaction (qPCR)

Primers for DNA amplification were manually designed, specifically for qPCR and produced by Sigma as 100µM stocks. The primers were stored at -20°C on arrival. The qPCR was performed using Power SYBR Green PCR Master Mix (Appplied Biosystems) according to the manufacturer's instructions with volumes listed in Table 3.7. Both are master mixes, therefore only primers and template had to be added. Annealing temperatures were set at 60°C based on a two-step reaction (Table 3.8). The reaction was performed on a LightCycler 480 and analysed using corresponding software.

Table 3.7 qPCR Reaction mix components.

Component	Power SYBR Green
DNA Template	variable
Forward/Reverse Primer (50-900nm)	variable
Master Mix	10µl (1x)
Nuclease Free Water	Up to 20µl

Table 3.8 qPCR Reaction steps.

Step	Enzyme Activation	PCR		Melt Curve
	Hold	Cycles (40)		
		Denature	Anneal/extend	
Time	10 min	15 sec	60 sec	Continuous
Temp °C	95	95	60	60-95

3.7 Bioinformatic tools

3.7.1 Primer design

Primers were designed manually with sequence visualization using Benchling online software (https://benchling.com/). Primers were analysed for melting temperatures, primer dimers and potential secondary structures using OligoEvaluator™ online tool (http://www.oligoevaluator.com/). Primers designed for pMTL80000 vector series (http://clostron.com/pMTL80000.php) (Heap et al., 2009) containing a CRISPR/Cas9 gene editing system were designed to have 5' overhangs for insertion of restriction sites that would allow ligation of the insert into the vectors.

3.7.2 CRISPR/Cas9 Vector and insert design

The vectors used in this study are from the pMTL80000 series and contain a Gram+ replicon (pCB102), a selection marker (catP), a Gram- replicon (ColE1+tra) and an application specific module (CRISPR/Cas9 and components) (Figure 3.1). The resulting plasmid is called pMTL83151. These modules allow the plasmid to replicate in *E. coli* and in *C. difficile* and be selected for using an antibiotic marker, in this case, chloramphenicol. The CRISPR/Cas9 module includes gene specific components (Figure 3.2); the homology arms (HAs) and a guide sequence to allow knock-out of a specific gene sequence. Each target gene will therefore have a unique vector containing different HAs and guide sequence. The HAs are usually designed to extend 1kb downstream (Right homology arm, RHA) and 1kb upstream (Left homology arm, LHA) of the gene of interest. Homologous recombination takes place using the homology present on the plasmid, allowing the deletion or cargo insertion to be "flipped" in place on the bacterial chromosome. This effectively removes and inserts sequences, leaving flanking regions unaffected. The guide sequence (single guide RNA, sgRNA) is 20bp in length and binds to the target of interest, allowing CRISPR/Cas9 to make a cut in a specific place and kill isolates which did not undergo the allelic exchange, as they will still have the target present, thereby selecting for correct mutants.

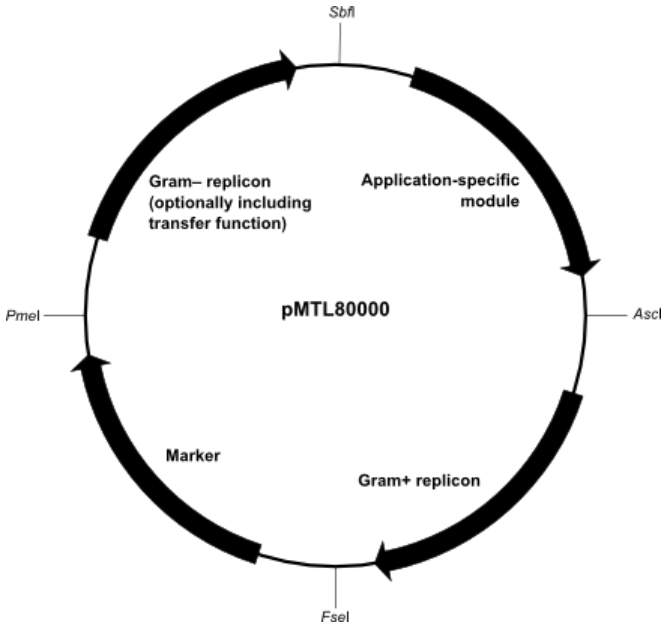


Figure 3.1 Structure of the pMTL80000 vectors

The figure shows a schematic representation of the vector series formulated to allow it to replicate in *E. coli* and in *C. difficile* with an application specific module. In this case, the module contains a CRISPR/Cas9 system capable of knocking out specific genes.

Figure taken from: http://clostron.com/pMTL80000.php

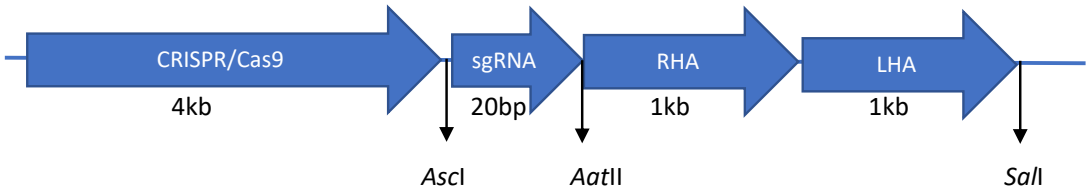


Figure 3.2 Map of application specific module.

The map shows the organization of the locus, its components, and their approximate sizes. As shown on the diagram, the components contain restriction sites inserted through primer overhangs to allow ligation of the fragments in the correct orientation.

The primers for HAs were designed manually using Benchling (https://benchling.com/) to visualize the sequence and create primers for PCR of RHA and LHA individually. The primers were designed to overlap and once the individual HAs were created using PCR, Splicing by Overlap Extension PCR (SOEing PCR) was used to stitch the RHA and LHA together (Horton, 1995). The sgRNAs were also designed using Benchling's CRISPR algorithm set to *C. difficile* 630 genome, with an NGG PAM. Each gene of interest had three different sgRNAs designed to ensure that at least one would have the correct on-target activity and would allow the gene knock-out. The sgRNAs were produced as long primers and annealed with a universal primer. All primers were designed with artificial restriction sites inserted as overhangs to allow digestion of fragments and ligation in the correct orientation. Once the HAs were spliced and extracted and the sgRNA prepared, both along with the vector were digested (described in 3.6.3) and ligated (3.6.4) to form a functioning CRISPR/Cas9 vector, able to remove a specific gene of interest. The vector was confirmed by sequencing.

3.7.3 CRISPR TargetAID (activation-induced cytidine deaminase)

Another method was used to create gene mutants, constructed based on the same pMTL83151 vector (described in Table 3.9 and section 3.7.2). In this case, the application specific module contains a Cas9 fusion to Cytidine Deaminase. The vector only contains the sgRNA as HAs are not necessary as the system mutates a single base, introducing a stop codon rather than relying on homology directed repair like traditional CRISPR/Cas9.

The sgRNA was designed on Microsoft Office Excel Software using an algorithm (Francois Seys, personal communication, PhD, University of Nottingham). The sgRNA was produced as a 20bp primer, with 25bp overhangs to allow assembly, digested and ligated using the NEBuilder® HiFi DNA Assembly Kit (as described in 3.6.4). The vector was confirmed by sequencing.

3.7.4 Sequencing

Purified DNA fragments and vectors were sequenced by Eurofins Genomics to validate expected sequences. The received sequences were aligned with expected sequences using Benchling, with the MAFFT algorithm (Katoh & Standley, 2013) and checked for discrepancies.

3.8 Plasmids

The colour-coded Table (3.9) below contains a list of constructed plasmids and their purpose.

Table 3.9 Plasmids generated.

Plasmid	Description	Source
pMTL83151_Cas9 CRISPR/Cas9 (RiboCas) module	Plasmid containing replicons for *E. coli* (ColE1+traJ) and *C. difficile*, (pCB102) conjugation factors. The application specific module contains CRISPR/Cas9 under the control of a riboswitch controlled by theophylline (RiboCas). Used as a base plasmid for all the plasmids listed below, which contain unique Has and sgRNAs	University of Nottingham, N. Minton (Cañadas et al., 2019)
pMTL83151_Cas9 plasmid base		
TCFP Ko BM6	Used to obtain the TCFP Ko mutant, with a bookmark scar	This study, Chapter 4
TCFP Comp	Used to complement the TCFP mutant	Chapter 4
TPPSP Ko	Used to obtain the TPPSP Ko mutant	Chapter 4
TPPSP Ko BM6	Used to obtain the TPPSP Ko mutant, with a bookmark scar	Chapter 4
TPPSP Comp	Used to complement the TPPSP mutant	Chapter 4
Kelch v3 Ko	Used to obtain the Kelch Ko mutant	Chapter 4
Kelch TA	Used to obtain the Kelch TA mutant, single bp change made using TargetAID, instead of CRISPR/Cas9	Chapter 4
DKO BM6	Used to obtain the double knock-out of TPPSP and Kelch genes, with a bookmark scar	Chapter 4
PTP Ko BM6	Used to obtain the PTP Ko mutant, with a bookmark scar	Chapter 4
PTP Comp	Used to complement the TCFP mutant	Chapter 4
TCFP Ko BM6	Used to obtain the TCFP Ko mutant, with a bookmark scar	Chapter 4
Swap 1	Used to obtain the Swap 1 mutant	This study, Chapter 5
Swap 2	Used to obtain the Swap 2 mutant	Chapter 5

Plasmid	Description	Source
Swap 3	Used to obtain the Swap 3 mutant	Chapter 5
Swap 4	Used to obtain the Swap 4 mutant	Chapter 5
Expand 1	Used to obtain the Expand 1 mutant, with an additional RBP	Chapter 5
Del1	Used to obtain the Deletion 1 mutant	Chapter 5
Del2	Used to obtain the Deletion 2 mutant	Chapter 5
Del3	Used to obtain the Deletion 3 mutant	Chapter 5
Del4	Used to obtain the Deletion 4 mutant	Chapter 5
Del5	Used to obtain the Deletion 5 mutant	Chapter 5
Del6	Used to obtain the Deletion 6 mutant	Chapter 5
Del7	Used to obtain the Deletion 7 mutant	Chapter 5
Del8	Used to obtain the Deletion 8 mutant	Chapter 5
Del9	Used to obtain the Deletion 9 mutant	Chapter 5
Del10	Used to obtain the Deletion 10 mutant	Chapter 5
Del11	Used to obtain the Deletion 11 mutant	Chapter 5
Del12	Used to obtain the Deletion 12 mutant	Chapter 5
Ex1	Used to obtain the Exchange 1 mutant	Chapter 5
Ex2	Used to obtain the Exchange 2 mutant	Chapter 5
Ex3	Used to obtain the Exchange 3 mutant	Chapter 5
Ex4	Used to obtain the Exchange 4 mutant	Chapter 5
Ex5	Used to obtain the Exchange 5 mutant	Chapter 5
Ex6	Used to obtain the Exchange 6 mutant	Chapter 5
Ex7	Used to obtain the Exchange 7 mutant	Chapter 5

Plasmid	Description	Source
Ex8	Used to obtain the Exchange 8 mutant	Chapter 5
Ex9	Used to obtain the Exchange 9 mutant	Chapter 5
Ex10	Used to obtain the Exchange 10 mutant	Chapter 5
Ex11	Used to obtain the Exchange 11 mutant	Chapter 5
Ex12	Used to obtain the Exchange 12 mutant	Chapter 5

3.9 Primers

The table below (Table 3.10) contains a list of all the primers used in this study.

Table 3.10 Primers.

Primer	Primer name	Sequence (5' - 3')
JS0001	CollarKO_LHA_F1	TATATAGACGTCGGAGAAAATAATCAAGCTGTTG
JS0002	CollarKO_LHA_R1	GAGGTGGTATAGGTTGGCACCAATGTAGA
JS0003	CollarKO_RHA_F2	TAGGTTGGCACCAATGTAGAAAGAGAGGTGAG
JS0004	CollarKO_RHA_R2	CACACTTGTTGAGTTATCAGAGGGCGCGCCTATATA
JS0005	phiCD1801recep_sgRNA1_F	TTTTCGTCGACGTTAAATTACAGCTTGGTGAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0006	phiCD1801recep_sgRNA2_F	TTTTCGTCGACAGGGGCAGAATATAATCCAAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0007	phiCD1801recep_sgRNA3_F	TTTTCGTCGACTTTGAAAGTGGCAATGTAGAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0008	sgRNAuniversal_R	GTATCAGACGTCATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTATAA AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCC
JS0009	CD2315_TPPSP_KO_LHA_F1	TATATAGACGTCCGAAGAATTACAAATTATTAAAAG
JS0010	CD2315_TPPSP_KO_LHA_R1	GGATGGTGAAATAATGGAAATGATTTAAA
JS0011	CD2315_TPPSP_KO_RHA_F2	AATAATGGAAATGATTTAAAAAGGAGCG
JS0012	CD2315_TPPSP_KO_RHA_R2	GTTTAAATTTATATGTATCAGATTATGGGGCGCGCCTATATA
JS0013	phiCD2315_sgRNAF1	TTTTCGTCGACATAACAGTAAAAAAGGGCAAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0014	phiCD2315_sgRNAF2	TTTTCGTCGACAGTAATTAATGTAGACGCAGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0015	phiCD2315_sgRNAF3	TTTTCGTCGACCGATAGGAGATACAAGGCAGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0016	CD2315_Kelch_KO_LHA_F1	TATATAGACGTCGGAAAAAAGTTTTGTATTCAAC
JS0017	CD2315_Kelch_KO_LHA_R1	GGAGCGTGAAAGATGACAGAATTATAAA
JS0018	CD2315_Kelch_KO_RHA_F2	AAAGATGACAGAATTATAAAAAAAGGAGAAAATAAA
JS0019	CD2315_Kelch_KO_RHA_R2	GCAGATGATACTAAAATTGCGGCGCGCCTATATA
JS0020	phiCD2315__Kelch_sgRNAF1	TTTTCGTCGACGCAGACTTCGAAAAAGAACTGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0021	phiCD2315__Kelch_sgRNAF2	TTTTCGTCGACGCAAAAAATGACTTGCAAGTGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0022	phiCD2315__Kelch_sgRNAF3	TTTTCGTCGACGCAAATATGCCTACCGCAAGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0023	pMTLCas9_F	CCATATCTACATCTACAAAATAATCATC
JS0024	pCB102-R1	CTGTTATGCCTTTTGACTATC
JS0025	phiCD2315_LC_F1	GGAAAATTATGGCGATG
JS0026	phiCD2315_LC_R1	GGTCGATATAGACAACATGG
JS0027	phiCD1801_MC_F1	GCTACAAGTGGAAATAAAGC
JS0028	phiCD1801_MC_R1	CCAGTGGTACTATTGATGTTG
JS0029	phiCD1801_MC_TCFP_F1	GGAGAATTGACTATAACTACAGG
JS0030	pCB102-F1	GCAAAATACATTCGTTGATG
JS0031	pCB102-F2	GTCAAGTATGAAATCATAAATAAAG
JS0032	catP-F1	GATAAATAGTTAACTTCAGGTTTGTC
JS0033	ColE1+tra-F1	CTGTGGATAACCGTATTACC
JS0034	MCS_R1	GGTCATAGCTGTTTCCTG
JS0035	Cas9scr1	ATGGATAAGAAATACTCAATAGGCTTAG
JS0036	Cas9scr2	GCTTTGTCATTGGGTTTGAC
JS0037	Cas9scr3	GTCGATAAAGGTGCTTCAGC
JS0038	Cas9scr4	GAACATATTGCAAATTTAGCTGG
JS0039	Cas9scr5	CTGACTTCCGAAAAGATTTCC
JS0040	Cas9scr6	GAGTTAGAAAACGGTCGTAAACG
JS0041	Cas9scr1R	GAAACTTAATCATATGCGCTAAGG
JS0042	sgRNA3.0_F	CTAGATTTATATTTAGTCCCTTGCCTTGC
JS0043	CD2315_Kelch_RHA_R2v2	GCTAAGTACAAGGTTGCAACGGCGCGCCTATATA
JS0044	CD2315_Kelch_RHA_R2v3	GATAAAAGCACATGAGGAACGGCGCGCCTATATA
JS0045	phiCD2315_Kelch_TAsgRNA	ATCTTAAGGAGGAGTTTTCGTCGACAATCAAATACATGTTTTGTAGTTTTAGAGCTA GAAATAGCAAGTT

JS0046	pCB102-R1	CTGTTATGCCTTTTGACTATC
JS0047	phiCD2315_Kelch_TA_RevComp	AACTTGCTATTTCTAGCTCTAAAACTACAAAACATGTATTTGATTGTCGACGAAAACTCCTCCTTAAGAT
JS0048	phiCD2315_KelchTA_MC_F1	GGCTATTCAGCTCCTACAAC
JS0049	phiCD2315_KelchTA_MC_R1	CTATAATAGATTCAAAACAATCGC
JS0050	phiCD2315_Kelchv3_MC_F1	GTGGAATGAAATTAAAAATTACG
JS0051	phiCD2315_Kelchv3_MC_R1	CATTAAGAGAAGAAACAGCAAC
JS0052	phiCD2315_TPPSP_MC_F1	GGAATTGATAATTTAGAAATGTC
JS0053	phiCD2315_TPPSP_MC_R1	TTTTAATAAAGGAGCTGTTCC
JS0054	phiCD2315_TPPSP_MC_F2	GGTTTTGAGAGAAACGAAG
JS0055	1801 TCFP LHA+ R1	GAAACGTACAAACCAACTGCTGACTCCGGGAGTGTAAAA
JS0056	1801 TCFP RHA+ F2	CTGACTCCGGGAGTGTAAAAGACTTAGTTATAAAAATTATTTTAGTAGTTTC
JS0057	1801 TCFP LHA+ F1	TATATAGCGGCCGCGGAGAAAATAATCAAGCTGTTG
JS0058	1801 TCFP RHA+ R2	CACACTTGTTGAGTTATCAGAGCTCGAGTATATA
JS0059	027_slpA_F2	SOURCE = M. WHITTLE
JS0060	027_slpA_R2	
JS0061	fur_F1	SOURCE = A. COCKAYNE
JS0062	fur_R1	
JS0063	CD1801_pyrEScreen_F1	SOURCE = M. WHITTLE
JS0064	CD1801_pyrEScreen_R1	
JS0065	1801L_TCFP__LHA_R1bv2	GAGGTGGTATAGGTTGGCAACCGATACCGTTTACGAAATAGG
JS0066	1801L_TCFP__RHA_F2bv2	ACCGATACCGTTTACGAAATAGGACCAATGTAGAAAGAGAGGTGAG
JS0067	CD2301_gp18KO_LHA_F1JS	TATATAGACGTCGATGTAGAGAACATATTGAGGC
JS0068	CD2301_gp18KO_LHA_R1JS	GGAGGTGATATAGTTGGCAACCGATACCGTTTACGAAATAGG
JS0069	CD2301_gp18KO_RHA_F1JS	ACCGATACCGTTTACGAAATAGG ACCTATTTAAGAAAGGAGAGTG
JS0070	CD2301_gp18KO_RHA_R1JS	GCTACAGCAGAAAAAGTTGTAAGGGCGCGCCTATATA
JS0071	phiCD2315_KelchTA_MC_F1v2	CCTAAAAATGTATTTATTGAAGC
JS0072	phiCD2315_KelchTA_MC_R1v2	GGAGTTATTATTTTTTCTAGTAGTGG
JS0073	ColE1+tra-F1	SOURCE = SBRC
JS0074	pCB102_F1	
JS0075	catP_R1(2)	
JS0076	ColE1_F1	
JS0077	MCS_F1(2)	
JS0078	cas9_scr2	
JS0079	cas9_scr4	
JS0080	cas9_scr6	
JS0081	p15a+tra_F2	
JS0082	phiCD08011TCFP_LHA_F_JS	TATATAGACGTCGGAGTAAATAATCAAGCTGTTG
JS0083	phiCD08011TCFP_LHA_R1_JSb	GAGGTGATAAAATTGGCAACCGATACCGTTTACGAAATAGG
JS0084	phiCD08011TCFP_RHA_F1_JSb	ACCGATACCGTTTACGAAATAGG ACCAATGTAAGAAGGTGATAAAG
JS0085	phiCD08011TCFP_RHA_R1_JS	CAAGTTTTGATAGTTCAGATGCGGCGCGCCTATATA
JS0086	phiCD08011_MC_F1	GAATGGGCTTTGGAAGTAG
JS0087	phiCD08011_MC_R1	GTTGCATCAGGACGTACC
JS0088	phiCD08011_MC_TCFP_F1	GCTATTTGTAGTGGAAATGATAG
JS0089	phiCD2301_MC_F1	GTCAAAGTATTAATCTTTGGAGAG
JS0090	phiCD2301_MC_R1	GGTTCTACAACAGATGAAGTTAC
JS0091	phiCD2301_MC_PTP_F1	CAGGTGAAACAGTTACTATATATCC
JS0092	1801 TCFP LHA+ R1_Wat	GAAACGTACAAACCAACTGCTGACTCCGGGAGTGTAAAA
JS0093	1801 TCFP RHA+ F2_Wat	CTGACTCCGGGAGTGTAAAAGACTTAGTTATAAAAATTATTTTAGTAGTTTC
JS0094	phiCD1801TCFP_Complement_BM6_F	TTTTCGTCGACACCGATACCGTTTACGAAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0095	CD418_TCFP_LHA_F1JS	TATATAGACGTCGATAAGATTGATTTAGGTGGAAC
JS0096	CD418_TCFP_LHA_R1JS	GAGGAAGTGTTATAAGTGGCAACCGATACCGTTTACGAAATAGG
JS0097	CD418_TCFP_RHA_F1JS	ACCGATACCGTTTACGAAATAGGACCTATTTAAGATAGGAGAGTGAAA
JS0098	CD418_TCFP_RHA_R1JS	GATAGTTATAAGCCAGCAAATTATAGGGCGCGCCTATATA

JS0099	phiCD418_MC_F1	GGAAACTGTACTAAACAATCTAAAAAC
JS0100	phiCD418_MC_R1	GGTGGAAATACTACGAAACAAG
JS0101	phiCD418_MC_TCFP_F1	GAATGGAATAGAAATGTTAAATGG
JS0102	CD2301_gp18Wat2_R1	GAAAATTTTAACGAGGTTAATATATCTATACAAAATGATATGTTAC
JS0103	CD2301_gp18Wat2_F1	TATACAAAATGATATGTTACCTATTTAAGAAAGGAGAGTGATAAAAAT
JS0104	CD2315_TPPSPbk_LHA_R2	GGATGGTGAAATAATGGAAACCGATACCGTTTACGAAATAGG
JS0105	CD2315_TPPSPbk_RHA_F2	ACCGATACCGTTTACGAAATAGGTATGATTTAAAAAGGAGCG
JS0106	CD2315_TPPSPWat_compR1	GGATGAAGAAATAAGTAATTTCAACGTTTGGCGAGAAAATCAA
JS0107	CD2315_TPPSPWat_compF2	ACGTTTGGCGAGAAAATCAAGAGAATAAACAAGAAGATTGGTATAAC
JS0108	phiCD418_WT_TCFP_F2	CTTGCAGATATTCAACCACTTTA
JS0109	phiCD418_WT_TCFP_F3	GGTGACTTAATAGCAATAAAAAC
JS0110	phiCD418_WT_TCFP_R2	GCAACAAGTAATCTTATTGCTACG
JS0111	phiCD418_MC_TCFP_F2	CTTTATCATGAGCCAATAAGC
JS0112	1801_LHA_Swap1_R1	CCTAGACAGGAGGTGGTATAGGCTCGAGATATATATCCCGGG
JS0113	1801_RHA_Swap1_F2	CTCGAGATATATATCCCGGGAAAGAGAGGTGAGGATATGAG
JS0114	2301_PTP_XhoI_F1	TATATACTCGAGTTGGCAGAACAAAAATATTATAC
JS0115	2301_PTP_XmaI_R1	CAAAATGACATGTTACCTATTTAACCCGGGTATATA
JS0116	1801_LHA_Swap2PTPCD2301_R1_2	CTAGACAGGAGGTGGTATAGGTTGGCAGAACAAAAATATTATAC
JS0117	2301_PTP_1801LHAov_F1_3	CTAGACAGGAGGTGGTATAGGTTGGCAGAACAAAAATATTATAC
JS0118	2301_PTP_1801RHAov_R1_4	CAAAATGACATGTTACCTATTTAAAAAGAGAGGTGAGGATATGAG
JS0119	1801_RHA_Swap2PTPCD2301_F1_5	CAAAATGACATGTTACCTATTTAAAAAGAGAGGTGAGGATATGAG
JS0120	2301_LHA_Swap2TCFPCD1801_R1_2	CTTTATACCCTAAGGAGGTGATATAGTTGGCAGAACAAAAATATTATAC
JS0121	1801_TCFP_2301LHAov_F1_3	CTTTATACCCTAAGGAGGTGATATAGTTGGCAGAACAAAAATATTATAC
JS0122	1801_TCFP_2301RHAov_R1_4	GTAAATAATAATTTGATACCAATGTAGGAAAGGAGAGTGATAAAAATGC
JS0123	1801_RHA_Swap2PTP2301_F1_5	GTAAATAATAATTTGATACCAATGTAGGAAAGGAGAGTGATAAAAATGC
JS0124	2301_LHA_Swap3TPPSP2315_R1_2	CTTTATACCCTAAGGAGGTGATATAGATGGAAAAAAGTTTTGTATTCAAC
JS0125	2315_TPPSP_2301LHAov_F1_3	CTTTATACCCTAAGGAGGTGATATAGATGGAAAAAAGTTTTGTATTCAAC
JS0126	2315_TPPSP_2301RHAov_R1_4	CTATAGTAAACAACTTAATAGATATGATTTAAGAAAGGAGAGTGATAAAAATGC
JS0127	2315_RHA_Swap3PTP2301_F1_5	CTATAGTAAACAACTTAATAGATATGATTTAAGAAAGGAGAGTGATAAAAATGC
JS0128	2315_LHA_Swap4PTP2301_R1_2	GAAGATAAAAAGGATGGTGAAATATTGGCAGAACAAAAATATTATAC
JS0129	2301_PTP_2315LHAov_F1_3	GAAGATAAAAAGGATGGTGAAATATTGGCAGAACAAAAATATTATAC
JS0130	2315_TPPSP_2301RHAov_R1_4	CAAAATGACATGTTACCTATTTAAAAAGGAGCGTGAAAGATG
JS0131	2315_RHA_Swap4PTP2301_F1_5	CAAAATGACATGTTACCTATTTAAAAAGGAGCGTGAAAGATG
JS0132	1801_LHA_Expand1_PTPCD2301_R1_2	CTAGACAGGAGGTGGTATAGGTTGGCAGAAC
JS0133	2301_PTP_1801LHAov_Expand1_F1_3	GTGGTATAGGTTGGCAGAACAAAAATATTATAC
JS0134	2301PTP_1801TCFP__Expand1_R1_4	CAAAATGACATGTTACCTATTTAATTGGCAGAACAAAAATATTATAC
JS0135	2301PTP_1801TCFP__Expand1F1_5	CAAAATGACATGTTACCTATTTAATTGGCAGAACAAAAATATTATAC
JS0136	1801TCFP_RHA_Expand1_R1_6	GTAAATAATAATTTGATACCAATGTAGAAAGAGAGGTGAGGATATGAG
JS0137	1801TCFP_RHA_Expand1_F1_7	GTAAATAATAATTTGATACCAATGTAGAAAGAGAGGTGAGGATATGAG
JS0138	2315TPPSP_KO_LHA_R_2	GGATGGTGAAATAATGGAAACCGATACCGTTTACGAAATAGG

JS0139	2315Kelch_KO_RHA_F_3v2	ACCGATACCGTTTACGAAATAGGGAATTATAAAAAAAGGAGAAAATAAA
JS0140	1801_del1hyp12_LF1_1	TATATAGACGTCGATTTAGATATCTATAAGGGTGAGG
JS0141	1801_del1hyp12_LR1_2	GTGAGGTGGTTTAATTGAAAACCGATACCGTTTACGAAATAGG
JS0142	1801_del1hyp12_RF2_3	ACCGATACCGTTTACGAAATAGGGGTGGTATAGGTTGGCAG
JS0143	1801_del1hyp12_RR2_4	CAAAACACAAATATATTCTTAACCACGGCGCGCCTATATA
JS0144	1801_del1_MC_F1	GATTTATGGGGAGGTAGAAG
JS0145	1801_del1_MC_R1	CAGATTTAGAAATGAAACAAATAC
JS0146	1801_del1_MCHyp12_F2	GTACAGAAGGTTTACATGACTTTAG
JS0147	phiCD1801_MCDel2_F2	CAGAAGGTTTACATGACTTTAGTA
JS0148	1801_del3_LF1_1	TATATAGACGTCGCCAAGGAAAAGCAGG
JS0149	1801_del3_LR1_2	GGAGGTAGAAGTTAATGTATAGTAGTCAACCGATACCGTTTACGAAATAGG
JS0150	1801_del3_MC_F1	CTTTTGATTGTAGTGGATTCAC
JS0151	1801_del3_F2	CAAAATTTACAGATGATTTATTAAGTG
JS0152	phiCD1801Del1_sgRNA1_F	TTTTCGTCGACTTGTTGACACAGCAACGTGGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0153	phiCD1801Del1_sgRNA2_F	TTTTCGTCGACAACATTATATCCTAGACAGGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0154	phiCD1801Del1_sgRNA3_F	TTTTCGTCGACTTTTTGTTGACACAGCAACGGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0155	1801_del4_LF1_1	TATATAGACGTCCTTGAACACAAATTTGAAATG
JS0156	1801_del4_LR1_2	GAAAGAGAGGTGAGGATATGAGAACCGATACCGTTTACGAAATAGG
JS0157	1801_del4_RF2_3	ACCGATACCGTTTACGAAATAGGATTTATTTAATAAATGAGGTGATAAAATG
JS0158	1801_del4_RR2_4	GGGAGTAGCCATAGGAGCGGCGCGCCTATATA
JS0159	1801_del4_MC_F1	AACAAATGATTGTAGTTATAACACTATT
JS0160	1801_del4_MC_F2	GCAAGAAGTAAAAGGTAATAGAAGTAG
JS0161	1801_del4_MC_R1	GCAATAGTACCAGGACACAC
JS0162	phiCD1801Del4_sgRNA1_F	TTTTCGTCGACACCTTAAGTGATATTAGCAAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0163	phiCD1801Del4_sgRNA2_F	TTTTCGTCGACGAAATTAGTACAGAACAAGGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0164	phiCD1801Del4_sgRNA3_F	TTTTCGTCGACTCAATCTTAAGCCAAGTACCGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0165	2301_del7_LF1_1	TATATAGACGTCCCTAGATATTTACAAAGGTGAAGG
JS0166	2301_del7_LR1_2	GAGGTGGTAGTTCAATGAAAACCGATACCGTTTACGAAATAGG
JS0167	2301_del7_RF2_3	ACCGATACCGTTTACGAAATAGG GGTGATATAGTTGGCAGAAC
JS0168	2301_del7_RR2_4	CTAATCTACAAATTTGTGTCGGGGCGCGCCTATATA
JS0169	2301_del7_MC_F1	GGGGAGGTGGAAGTTAATG
JS0170	2301_del7_MCHyp12_F2	GAGGGCATACATGACTTTAG
JS0171	2301_del7_MC_R1	GTTTTAACGGTGCAGGC
JS0172	phiCD2301Del7_sgRNAF1	TTTTCGTCGACCCAATAATCCAATCCCCAAGGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0173	phiCD2301Del7_sgRNAF2	TTTTCGTCGACGCAGGTTCTACAATGAATACGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0174	phiCD2301Del7_sgRNAF3	TTTTCGTCGACATATCATTGTTATAAAAAGAGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0175	2301_del9_LF1_1	TATATAGACGTCCAAGAGACCAGGGAAAGG
JS0176	2301_del9_LR1_2	GGGAGGTGGAAGTTAATGTATACCGATACCGTTTACGAAATAGG
JS0177	2301_del9_MC_F1	CTTTTGATTGCAGTGGG
JS0178	2301_del9_MCBaseJ_F2	CAAGGATACAAAATTTACAGATG
JS0179	2301_del10_LF1_1	TATATAGACGTCGCACATCTTGAACATAAATTTG
JS0180	2301_del10_LR1_2	GGAGAGTGATAAAAATGCAAACCGATACCGTTTACGAAATAGG
JS0181	2301_del10_RF2_3	ACCGATACCGTTTACGAAATAGGTTATGAATAATATGAAAGGTGGT
JS0182	2301_del10_RR2_4	TGTCTATGTTTACTTTAGAAGAAGTGGGCGCGCCTATATA
JS0183	2301_del10_MC_F1	GTTGATGTAGAAGTATTTAGTAATCTATTTAG
JS0184	2301_del10_MCHyp14_F2	GAAAAATAGAAAATTTTAACGAGG
JS0185	2301_del10_MC_R1	CTGAACGTGCATGGATTAC
JS0186	phiCD2301Del10_sgRNAF1	TTTTCGTCGACGATGTTGGTATATATCCAGGGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0187	phiCD2301Del10_sgRNAF2	TTTTCGTCGACGGTGGAATAGGAATAGGTGGGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
JS0188	phiCD2301Del10_sgRNAF3	TTTTCGTCGACGGTTATGAGAAAGTTCCCGAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC

JS0189	ex1_LHA_R1_2	CTAAATTTTAGTGAGGTGGTTTAAATGAAATTAA
JS0190	Ex1_hyp12_F1_3	GGTGGTTTAAATGAAATTAATTGATAAGCTACC
JS0191	Ex1_hyp12_R1_4	TTATACCCTAAGGAGGTGATATAGGTTGGCAGAA
JS0192	ex1_RHA_F1_5	GGTGATATAGGTTGGCAGAACAAAAATATTATAC
JS0193	Ex2_ptp_R_4	CAAAATGACATGTTACCTATTTAAAAAGAGAGGT
JS0194	Ex2_RHA_F1_5	ACCTATTTAAAAAGAGAGGTGAGGATATGAG
JS0195	Ex3_LHA_R1_2	GATTTATGGGGAGGTAGAAGTTAATGTATAGTA
JS0196	Ex3_BaseJ_F1_3	GTAGAAGTTAATGTATAGTAGTCAAACATATGATGTTG
JS0197	Ex4_LHA_R1_2	TGTAGAAAGAGAGGTGAGGATATGCAAACTG
JS0198	Ex4_hyp14_F1_3	AGGTGAGGATATGCAAACTGAATGGAATTTT
JS0199	Ex4_hyp14_R_4	GTTTTACATGGTATGCTTATGAATAATAAATGAGGT
JS0200	Ex4_RHA_F_5	TTATGAATAATAAATGAGGTGATAAAATGAATAGAG
JS0201	Ex5_LHA_R_2	CTAGACAGGAGGTGGTATAGGTTGGCAGAAC
JS0202	Ex5_Ptp_F_3	GTGGTATAGGTTGGCAGAACAAAAATATTATAC
JS0203	Ex7_LHA_R_2	GAATTTAGTGAGGTGGTAGTTCATTGAAATTAA
JS0204	Ex7_Hyp12_F_3	TGGTAGTTCATTGAAATTAATTGATAAACTACCTTC
JS0205	Ex7_Hyp12_R_4	CCTAGACAGGAGGTGGTATAGTTGGCAGAAC
JS0206	Ex7_RHA_F_5	GGTGGTATAGTTGGCAGAACAAAAATATTATAC
JS0207	Ex8_TCFP_R_4	GTAAATAATAATTTGATACCAATGTAGGAAAGGAGAG
JS0208	Ex8_RHA_F_5	ACCAATGTAGGAAAGGAGAGTGATAAAAATGC
JS0209	Ex9_LHA_R_2	TGGGGAGGTGGAAGTTAATGTATAGTA
JS0210	Ex9_BaseJ_F_3	GTGGAAGTTAATGTATAGTAGTCAAACTTATGATGTC
JS0211	Ex10_Hyp14_F_3	AGTGATAAAAATGAGACATTATGAAAGAACAAC
JS0212	Ex10_LHA_R_2	ATTTAAGAAAGGAGAGTGATAAAAATGAGACATT
JS0213	Ex10_Hyp14_R_4	GGCATGCTATAAAATTTATTTAATATGAAAGGT
JS0214	Ex10_RHA_F_5	ATTTATTTAATATGAAAGGTGGTAATTTAAATGG
JS0215	Ex11_LHA_R_2	CTTTATACCCTAAGGAGGTGATATAGTTGGCAGAAC
JS0216	Ex11_TCFP_F_3	GGTGATATAGTTGGCAGAACAAAAATATTATAC
JS0217	1801_seq1_F	GTGTGCAAGATTTGATGAATC
JS0218	2301_seq1_F	CCTGGATATATACCAACATCTG
JS0219	2301_seq2_F	CAACTTGCTTCTGTTGG
JS0220	L12S_TTMP_Lys_F	GTGGAAGAAAATTAATTTAAAATAAG
JS0221	L12S_TTMP_Lys_R	GGAATCAGCTGGCAAC
JS0222	L12W_TTMP_Lys_F	CTTAATAATTTTGAAATAAACCAAG
JS0223	L12W_TTMP_Lys_R	GATATATGGCTTTGGCTG
JS0224	1801_del1hyp12v2_LF2_1	TATATAGACGTCAATAATCAAGCTGTTGACACAG
JS0225	2301_del10short_RR2_4	ATAGAGATTCTGAGATAGTAAGCACGGCGCGCCTATATA
JS0226	2301_del10_MC_R1_short	AATTAAATAACAATTTATAGGAGGA
JS0227	L12W_Lys2_F	GGAGGAATCGAGGATG
JS0228	L12W_Lys2_R	CTAGGAGAAATTAAAGTAGAGCAG
JS0229	L12W_Lys2_R2	CCGGAATTATATAAAGATAACTGG
JS0230	L12W_Lys3_F	GAGACGCCTAAAAAAGCTC
JS0231	L12W_Lys3_R	GTTTGGATGGTAATTCGTG
JS0232	L14WSeqchk2_F1	SOURCE = S. KERR
JS0233	L14WSeqchk2_R1	
JS0234	rpsJF	
JS0235	rpsJR	
JS0236	1801_PS_LR1_2	CCAACAGTAATTTTAGCAACTTTAAAAGATATACAAGAT
JS0237	2301_PS_IR1_F3	TTTTAGCAACTTTAAAAGATATACAAGATTTAGACTCTAAAATTG
JS0238	2301_PS_IR2_R4	ACAAAATGACATGTTACCTATTTAAAAAGAGAGGTGAGGA
JS0239	1801_PS_RF1_5	ATGTTACCTATTTAAAAAGAGAGGTGAGGATATGAG
JS0240	TCFP_PTP_sqR1	ACAACTAATTGCAGTATCTAAATATCC
JS0241	2301_PS_LR1_2	GTCTTGTATGTCTTTAAGTGTTGCTAAAATAACAGTTG
JS0242	1801_PS2_IR_F3	ATTTTAGCAACACTTAAAGACATACAAGACTTAGACACTAAAAT
JS0243	1801_PS2_IR_R4	TATCACTCTCCTTTCCTACATTGGTATCAAATTATTATTTACTG

JS0244	2301_PS2_RF1_5	GATACCAATGTAGGAAAGGAGAGTGATAAAAATGC
JS0245	BA_1F	CAGGTGAAACAGTTACTATATATCCT
JS0246	BA_1R	CATTTGCAATAGATGCCTT
JS0247	BA_2F	TCCAAACTGGATAGTAATAGAAACT
JS0248	BA_2R	GAGTCAGCAGTTGGTTTGTA
JS0249	BA_3F	AAACAGAATTAACAAGCAACATAG
JS0250	BA_3R	CTTGTTTTCTCTGCTGTTGTT
JS0251	BA_3.2F	GATACAACTAAAACAGAATTAACAAGCAAC
JS0252	BA_3.2R	CTTACTACAACAGATAAAACAAATATAGTTAGTGC
JS0253	BA_2.2F	CGAAGAAAATCCAAACTGGATAGTAATAG
JS0254	BA_2.2R	CAACTAATTGCAGTATCTAAATATCCAGAAAC

Chapter 4:
Identification of receptor binding proteins in phages infecting *Clostridium difficile*

Chapter 4. Identification of receptor binding proteins in phages infecting *Clostridium difficile*

4.1 Introduction

Recent emergence of antibiotic resistance and concerns over relapse in patients with cases of CDI have led to renewed interest in phage therapy (Ventola, 2015). One of the advantages of using phages as therapy is their specificity, which allows them to only target their host organisms. In this case, the depleted microbiome of a patient would have the opportunity to recover, while the phages kill the *C. difficile* and remove the infection (Meader et al., 2010).

This advantage is also a severe limitation, as the phages infecting *C. difficile* have a very narrow host range and often only infect a single PCR Ribotype (Sekulovic et al., 2014). Using the many required phages presents pharmaceutical companies with many challenges surrounding their propagation and maintenance. Before any phage-based treatment can be developed, more fundamental research needs to be carried out to elucidate phage binding mechanics. Factors controlling the host range can be clearly established, and hopefully later modified to expand the phage infection profile. First, the receptor binding protein (RBP) on the phage has to be identified as this binds to the receptor on bacteria and allows attachment and injection of phage genetic material (Le et al., 2013).

The RBP is what likely determines the specificity and contributes to host range (Rothenberg et al., 2011), as this is the initial point of contact between the bacteria and phages. Therefore, this study sought to investigate receptor-binding determinants in *C. difficile* phages. In this Chapter, a selection of previously isolated phages within the research group (Whittle et al., 2022) were used to investigate RBPs. Five phages: ΦCD1801 (Ribotype 078), ΦCD2315 (Ribotype 078), ΦCD2301 (Ribotype 014), ΦCD418 (Ribotype 023) and ΦCD08011 (Ribotype 002), were investigated using a CRISPR/Cas9 system to delete the suspected RBP genes and assess their presence and impact on virion assembly and host strain infection. Drawn conclusions are confirmed by genetic complementation of the deleted gene.

4.2 Results

4.2.1 Identification of target genes

Firstly, putative RBP candidate genes were identified. A selection of phages infecting different *C. difficile* PCR Ribotypes were chosen to study how the potential RBP genes differ between the isolates and whether their architectures are conserved, creating a blueprint to identify the genes in other phages for future therapies. Additionally, increasing the number of RBPs studied will increase the understanding of RBPs for phage infection. These findings will have the potential to inform the future development of phages with broadened host range activity.

Phage genes are typically organised into modules according to their function (Belcaid et al., 2010). The RBP is likely found attached to the tail structures and therefore the suspected RBP encoding gene likely falls into this module (Bielmann et al., 2015). In a previous study (A. Dowah, 2018), an anti-gp29 protein was able to block ΦCDHM1 infection, suggesting that gp29 is the RBP allowing the phage to bind to its host, CD105HE1 (PCR Ribotype 076). BLASTp of gp29 protein sequence identified it as "*putative tail fibre protein*". This is highly similar (76% positive amino acid sequence, 58% complete identity; NCBI alignment) to gp25, identified as "*phage tail-collar fibre protein*" in ΦCD1801, the first phage candidate. Identification of this potential RBP set the benchmark for what criteria were used to narrow down potential RBPs in the other phages.

The *tail-collar fibre protein* (*TCFP*) encoding gene is 783bp in size, located downstream of the *tail tape measure protein* and upstream of the *holin* encoding gene. Suspected RBPs in other phage were labelled under the same name (ΦCD418 and ΦCD08011) and/or were typically 700-1000bp in size and found in the assumed position near the structural genes (Figure 4.1). In ΦCD2301 the identified gene was labelled as "*phage tail protein*" (*PTP*) and was highly similar to *TCFP* in ΦCD1801.

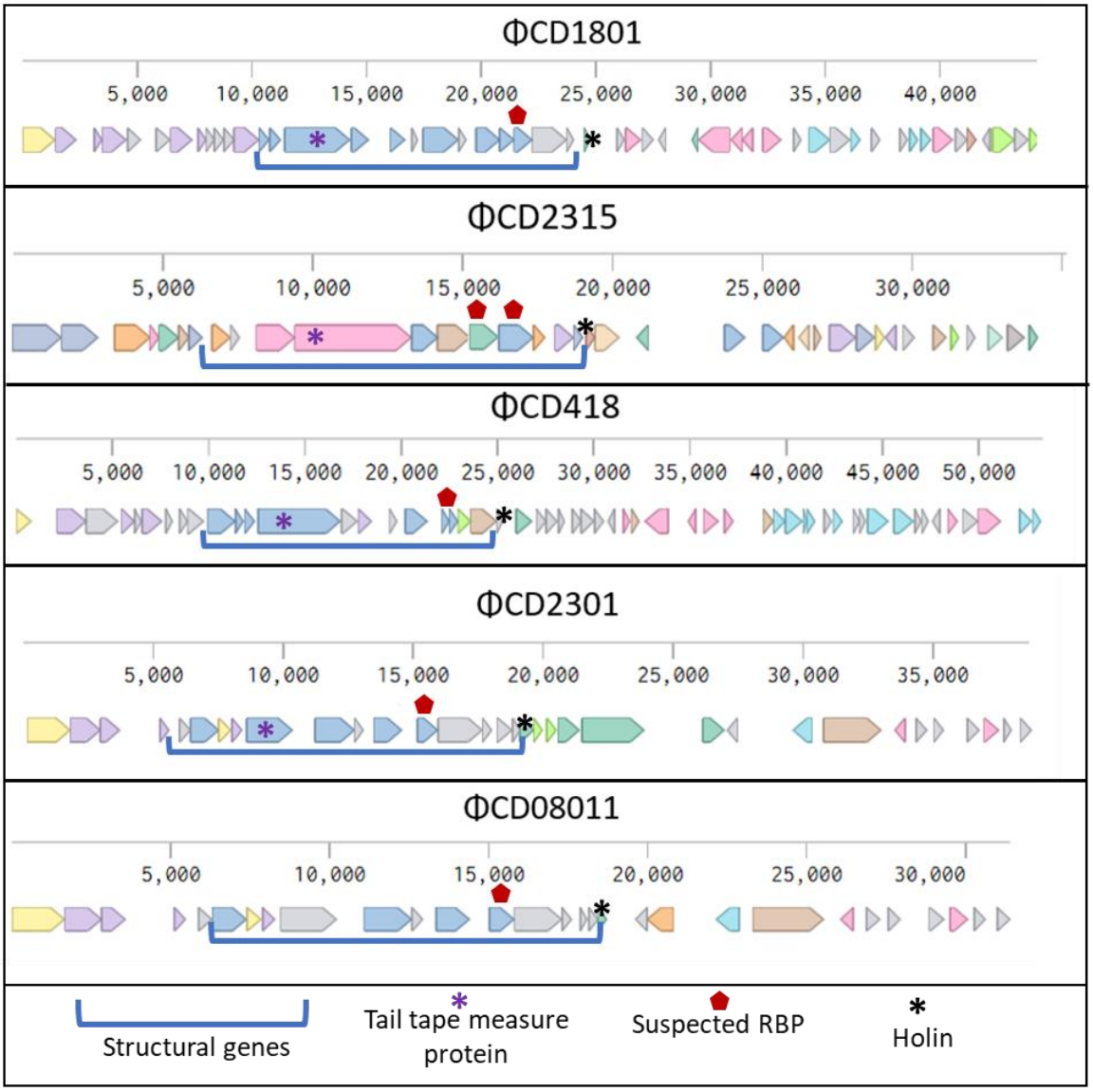


Figure 4.1 Selection of potential RBP encoding genes in candidate phages.

The figure shows a schematic representation of candidate phage genomes. Each arrow block is a potential gene annotated through blast searches. As phage genomes are organised in modules, the structural gene module in each phage is underlined in blue, as it is highly likely that the RBP gene is found in this region. Further annotations show the relative positioning of the suspected RBP gene in relation to the *tail tape measure protein* and *holin* genes, which were used as markers for narrowing down which gene is most likely to encode the RBP.

In phage ΦCD2315 there were several gene candidates that had the potential to be an RBP. The first candidate, *truncated putative phage structural protein* (*TPPSP*) gene is 942bp in size, located appropriately within the tail gene cluster and identified as a 97% match to a structural tail protein in another *C. difficile* phage, ΦCD6356. The second candidate, *Kelch repeat-containing protein* (*Kelch*) encoding gene, while larger in size (1146bp) than the other target genes, has the correct location and a Kelch motif. Kelch domains are known to mediat(Prag & Adams, 2003)teractions (Prag & Adams, 2003), making this a valid candidate for the RBP. Additionally, the Kelch gene was shown to have a 94% similarity to a gene encoding *tail fiber protein H* in ΦCD6356. This gene has been implicated in conferring specificity in phages infecting other organisms (Garcia et al., 2008), therefore showing potential as an RBP in ΦCD2315. The selected genes were compared to each other and found to be highly similar. A summary of the selected candidate genes for each phage is listed in the table below (Table 4.1).

Table 4.1 The table lists candidate genes for each phage, their sizes and abbreviations used in the reminder of the report.

Phage	Gene name	Acronym	Size
ΦCD1801	*Tail collar fibre protein*	*TCFP*	783bp
ΦCD2315	*Truncated putative phage structural protein*	*TPPSP*	942bp
	Kelch repeat-containing protein	*Kelch*	1146bp
ΦCD2301	*Phage tail protein*	*PTP*	780bp
ΦCD418	*Tail collar fibre protein*	*TCFP*	821bp
ΦCD08011	*Tail collar fibre protein*	*TCFP*	783bp

4.2.2 Vector design and construction

Following the selection of candidate genes, the appropriate vectors were designed and constructed with components specific for each gene, ligated into the pMTL83151 empty base vector containing a CRISPR/Cas9 module (Methods: 3.7.2; 3.8, Table 6 – Plasmids). Initially, this was carried out for ΦCD1801 targeting the *TCFP* gene. Individual vector components were prepared as described and ligated (3.6.4) into *E. coli* Top10. The correct construction of the first vector took several attempts and had to be optimised to obtain a SNP-free accurately assembled vector, with attempted conditions listed below (Table 4.2).

Table 4.2 List of attempted conditions used to construct a CRISPR/Cas9 Ko vector.

Vector: HA Ratio	sgRNA amount	sgRNA used	Antarctic Phosphatase (AP) treatment	Transform-ation	Change	Result
1:3	20ng	F1	No	Electroporation	-	Colonies with self-ligated vectors
1:3	20ng	F1	Yes	Electroporation	AP treatment	No colonies
1:3 1:5	40ng/ 20ng	F1	Yes	Electroporation	Different amount of sgRNA	No colonies
1:3 1:5 1:7	20ng	F1	Yes	Electroporation	Different Vector:HA ratios	No colonies
1:3 1:5 1:7 1:1	20ng	F1, F2, F3	Yes	Electroporation	Different Vector:HA ratios with different sgRNAs	No colonies
1:3	20ng	F1, F2, F3	Yes	Electroporation (into CA434 and Top10)	*E. coli* competent cells	No colonies
1:3	**20ng**	**F1, F2, F3**	**Yes**	**Chemical transformation**	**Chemical transformation instead of electroporation**	**Colonies with correctly constructed vectors**

The correct vector was constructed using a vector:HA ratio of 1:3 and 20ng of sgRNA, with the ligation mixture transformed into chemically competent cells. The use of chemically competent cells resulted in transformants, therefore all future experiments used chemically, not electrocompetent cells. Following screening and sequencing for verification, vectors containing sgRNAF1 and F3 were obtained for the ΦCD1801 *TCFP* gene knock-out.

Next, vectors targeting the *TPPSP* and *Kelch* genes in ΦCD2315 were constructed. The vector targeting the *Kelch* gene contained many SNPs in the right homology arm (RHA). This was potentially due to toxicity of CRISPR/Cas9 vectors in *E. coli* in general, and potential specific toxicity of the homology arms (HAs), which sometimes contain entire gene products which cells struggle to tolerate (Cui et al., 2018). In the *Kelch* gene, this was potentially due to presence of gene products encoded in the HA, including a whole *hypothetical protein* gene and approximately 50% of another in the RHA (Figure 4.2). To mitigate this, a shorter RHA was designed (283bp) to avoid having complete genes within the RHA; the vector was constructed correctly without SNPs (Kelchv3). While the vector contains terminators preventing transcription, these can be leaky and resulting products can still be toxic, therefore it was important to minimise any genes present while still having a functional RHA. Shorter HAs may lead to reduced efficiency of the recombination event, however they have been reported to be functional (K. Li et al., 2014) and should still lead to a knock-out of target gene.

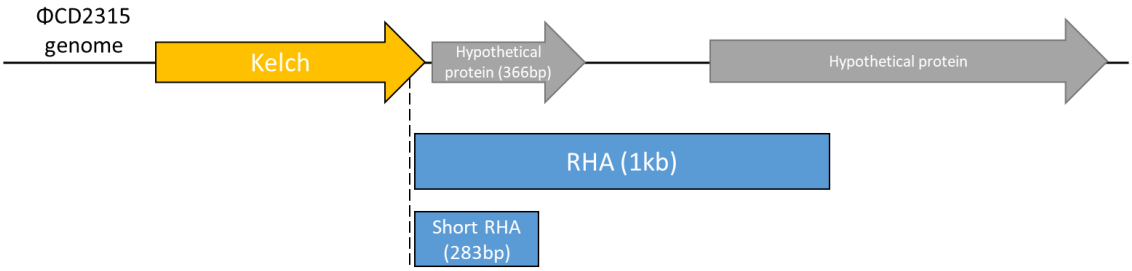


Figure 4.2. Schematic representation of the RHA constructed for the *Kelch* gene knock-out.

The figure shows the *Kelch* gene locus and downstream *hypothetical protein* genes fully or partially contained within the original 1kb long RHA. As this RHA had SNPs, a shorter version was designed to omit the HA containing any complete genes.

4.2.2.1 Target AID

In addition to fully knocking out the *Kelch* gene, a CRISPR TargetAID (Q. Li et al., 2019) vector (TA Kelch) was constructed. In contrast to traditional CRISPR/Cas9 mediated knock-outs, this vector only contains the sgRNA target and does not rely on HA, which was ideal as the RHA is what initially proved problematic. The method relies on an introduction of a single strand cut and substitution of a single base,

changing the amino acid into a stop codon (Nishida et al., 2016). A single place in the gene was selected by the algorithm (3.7.3; François Seys, University of Nottingham PhD thesis) as an appropriate target, placing the introduced stop codon in the middle of the *Kelch* gene, at 663bp. Ideally, the closer the stop codon is placed to the 5' end of the gene, the better as less of the encoded protein is transcribed and translated. This process is less lethal to the cell, as no double strand breaks are introduced like with traditional CRSPR. The vector was simpler to construct as no HAs are required, and it has potential for multiplexing different sgRNAs if required downstream. However, a potential drawback is that the stop codon is placed approximately half-way through the *Kelch* gene, meaning that it can still be partially transcribed. The partial product could have polar effects and limited functions, especially if certain N-terminal domains are present and functional.

4.2.3 Addition of a bookmark

The initial vectors were designed to have flanking regions, HAs, either side of the target gene (suspected RBP), 1kb in size each and in between a "scar" was left remaining to show the remnants of the knocked-out gene. At a later stage in the project, it was discovered that this type of knock-out made it very complicated to complement the target gene back in, which was necessary for further validation. This led to the inclusion of the bookmark; BM6 (Seys et al., 2020), in the scar. The vectors were re-constructed by re-digesting out the HAs with *Aat*II and *Sal*I and inserting the updated HAs which now contained a scar with the bookmark (Figure 4.3). The bookmark is a very strong CRISPR target, therefore a complementation vector can be designed with sgRNA to target BM6, and the HA can carry cargo, in this case the previously knocked-out RBP gene, thus allowing for complementation.

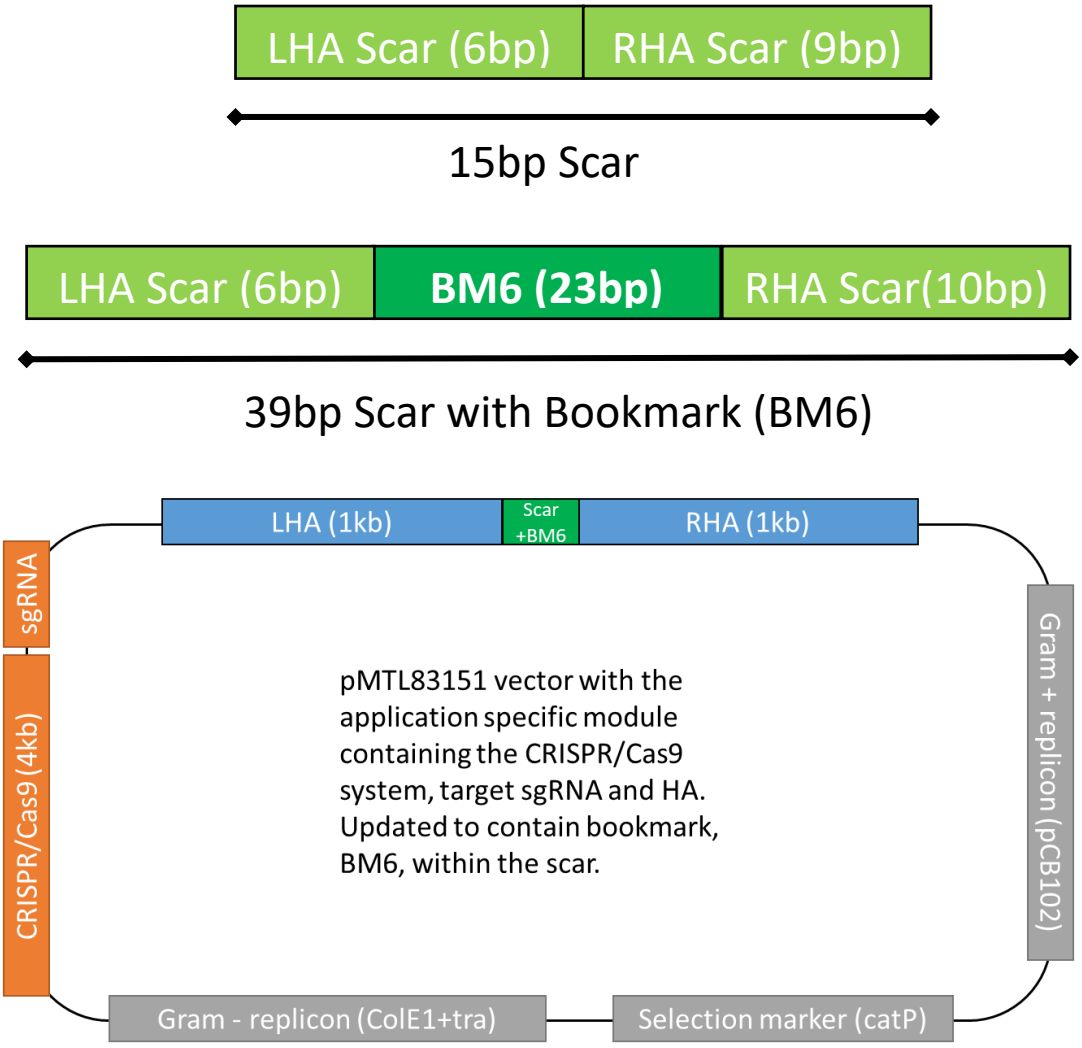


Figure 4.3 BM6 and vector overview.

The figure shows a representation of the remnant in place of the knocked-out candidate gene, the "scar". The small fragment, initially 15bp allowed to locate where the gene was deleted from, while maintaining the reading frame. The updated version includes the bookmark, BM6, which now allows for easy complementation of the gene back in, in downstream experiments. The RHA scar was extended to be 10bp in size, to maintain the reading frame. The vector map shows a schematic representation of the plasmid, also updated to represent the addition of the bookmark. The sgRNA and HA fragments are unique for each candidate gene.

Three more vectors targeting ΦCD2301, ΦCD08011 and ΦCD418 without bookmarks were designed and constructed by another student as part of his PhD rotation project (Stanley Ho, project 2018/19). These were then later used as templates for the reconstruction of these plasmids to include modified HAs and crucially a bookmark for retargeting of CRIPSR/Cas9.

4.2.4 Optimising DNA transfer and mutagenesis

Following construction and verification of the vectors, they were transferred to the appropriate *E. coli* donor strains. Then they were conjugated (3.4.1.; 3.4.2) into each *C. difficile* strain containing the lysogenised phage genome. Before conjugation of the actual knock-out plasmid, the best donor strain was chosen through test conjugations with the appropriate strains. For CD1801L, *E. coli* CA4343 and sX strains were tested, with tests showing dramatically improved efficiency when CA434 was used. However, in practice, with the actual knock-out plasmid, sX proved to be the more efficient strain. For CD2315L *E. coli* Ts was also tested, as cloning and conjugation can take place in the same strain due to presence of the R702 conjugation factor plasmid. However, none of the strains were able to conjugate the knock-out plasmid into CD2315L, therefore the protocol was adapted (3.4.2.). The main change included a lower centrifugation speed with extended time and an *E. coli* culture temperature at 30°C. This prevents the mechanical shearing of pili on the cells, which are required for conjugation and resulted in a much improved conjugation efficiency. A modified heat shock protocol was also tested as, a study showed drastically improved conjugation efficiency when the *C. difficile* cells are heat shocked for 5 minutes at 52°C and grown O/N in 2xYT Broth (normally BHIs Broth), then mated on BHI Agar plates (normally BHIs Agar) (Kirk & Fagan, 2016). However, no improvement in conjugal efficiency was observed in this experiment.

Another change introduced to improve the efficiency, was the modification of mating time on plates. Some strains require more time on the plate with donor cells (CD1801L, CD2301L, CD522418L) and were left for 24h to allow for optimal transfer of the plasmid. Others, (CD2315L and CD08011L) require a much shorter mating time, 8h, as this is enough to transfer the plasmid and past that time background colonies become a problem and take over, preventing transconjugants from developing on selection plates. The optimal conditions for each strain were found through testing and conjugating and are summarised below (Table 4.3).

Table 4.3 A summary of conjugation conditions optimised to work best for each lysogenised strain.

Strain	Protocol	Mating time	Success
CD1801L	Standard	8h/24h	Yes
CD2315L	Modified	8h	Yes
CD2301L	Standard	24h	Yes
CD522418L	Modified	24h	Yes (limited)
CD80011L	Standard/Modified	?	No, all background colonies

Following optimisation and successful conjugation, the resulting transconjugants – which are *C. difficile* strains containing the relevant plasmid, were plated on induction plates containing theophylline. This induces the CRISPR/Cas9 system to select for colonies which have undergone homologous recombination with the plasmid (Figure 4.4a). The colonies that have not, will still contain the target gene (suspected RBP) on the genome and will be targeted by the CRISPR/Cas9 system leading to a double strand break which is lethal to the cells. This method only allows colonies which have undergone homologous recombination to survive, as the target gene is no longer in the genome, but has now swapped over onto the knock-out plasmid (Figure 4.4b). The sgRNA is used as a guide which allows the Cas9 to cleave in the correct place, hence it is important to design guides with high on target (chances of targeting the right place) and low off target (chances of a cut elsewhere in the genome) scores. Multiple guides (F1/2/3) are usually designed as backups, due to software predictions of guide efficiency not always being accurate as they are based on the *C. difficile* 630 genome, not the strain they will actually be used in.

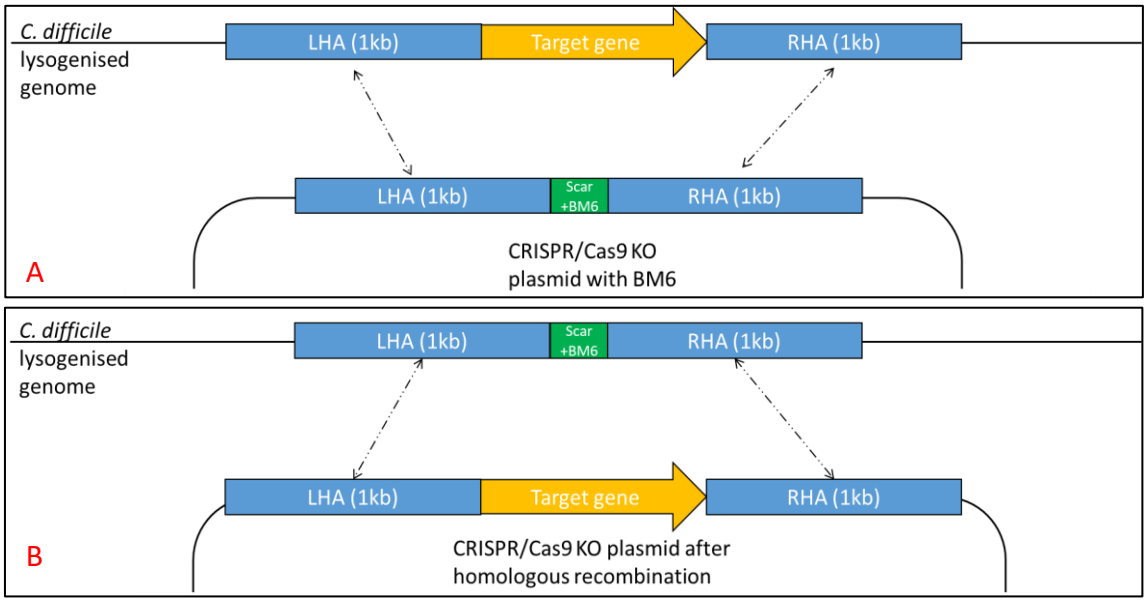


Figure 4.4 Homologous recombination between knock-out plasmids and lysogenised genome.

The figure shows the process that takes place once CRISPR/Cas9 is induced through plating on theophylline. Following conjugation, some colonies will undergo homologous recombination with the plasmid, where the LHA and RHA with the scar and BM6 are flipped into the lysogenised genome, and the target gene is flipped out onto the plasmid; essentially leading to the gene knock-out. Plating on theophylline removes the repression and allows the CRISPR/Cas9 to select for colonies containing no target gene. Colonies that still have the target gene have not undergone homologous recombination and are targeted by CRISPR/Cas9 which causes a double strand break in the genome which is lethal.

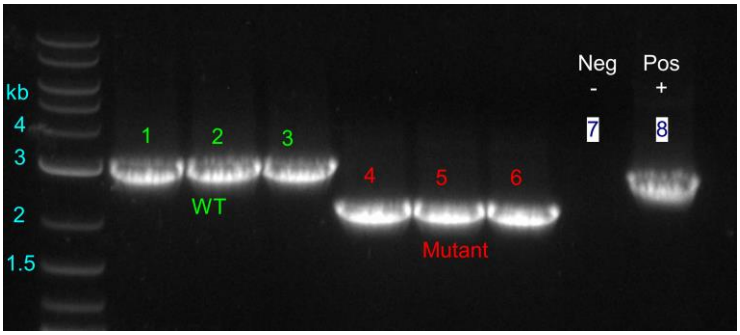
4.2.4.1 Generation of putative RBP mutants

RBP deletion mutants containing bookmark sequences were obtained within lysogenised strains: CD1801L, CD2301L, CD2315L and CD522418L (Table 4.4, Figure 4.5). While a plasmid for CD08011L was constructed, despite repeated conjugations using a standard, modified and heat shock protocols a mutant was not generated. The conjugation proved unsuccessful as selection plates contained an extremely high amount of background colonies instead of transconjugants. Until advances in DNA transfer are made, it may not be possible to obtain a mutant for this strain. All mutants went through the process of plasmid loss (3.4.1.2) and were verified through sequencing.

Table 4.4 Summary of the obtained mutant for each strain, with the knocked-out gene highlighted. These mutants all include BM6 as a bookmark, unless specified and were complemented (with a watermark) to generate corresponding complement mutants.

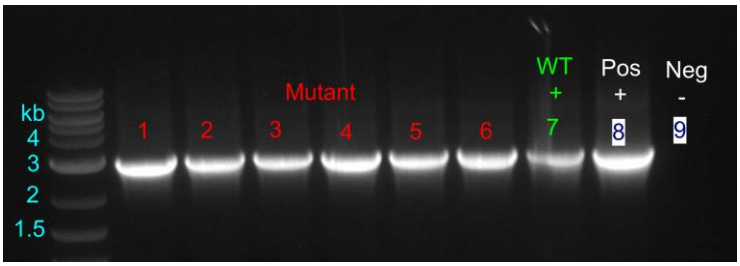
Strain	Gene	Name of mutant	Complemented?
CD1801L	*TCFP*	CD1801L TCFP Ko BM6	Yes
CD2315L	*TPPSP*	CD2315L TPPSP Ko BM6	Yes
CD2301L	*PTP*	CD2301L PTP Ko BM6	Yes
CD522418L	*TCFP*	CD522418L TCFP Ko BM6	No
CD2315L	*Kelch* (No bookmark)	CD2315L Kelch v3 Ko	No
CD2315L	*Kelch* (TargetAID)	CD2315L Kelch TA	No
CD2315L	*TPPSP* and *Kelch* (Double knock-out)	CD2315L DKO BM6	No

A CD1801 TCFP Ko



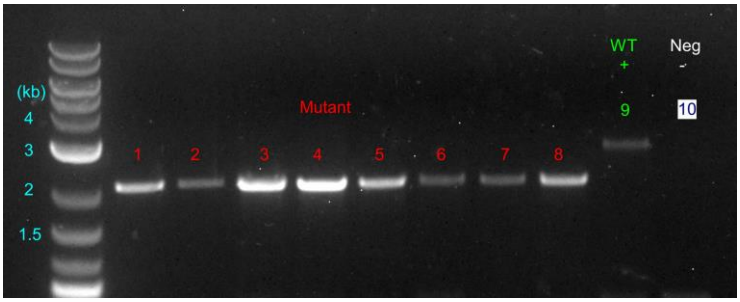
The gel image shows WT control colonies (CD1801L, ~3kb, green, 1-3) and Mutant obtained colonies (CD1801L TCFP Ko, ~2.2kb, red, 4-6), alongside a negative (7) and positive (CD1801L gDNA, 8) controls.

B CD1801 TCFP Comp



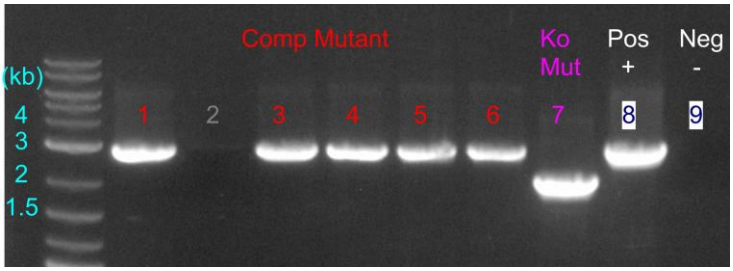
The gel image shows a WT control colony (CD1801L, ~3kb, green, 7) and Mutant obtained colonies (CD1801L TCFP Ko Comp, ~3kb, red, 1-6), alongside a negative (9) and positive (CD1801L gDNA, 8) controls.

C CD2301 PTP Ko



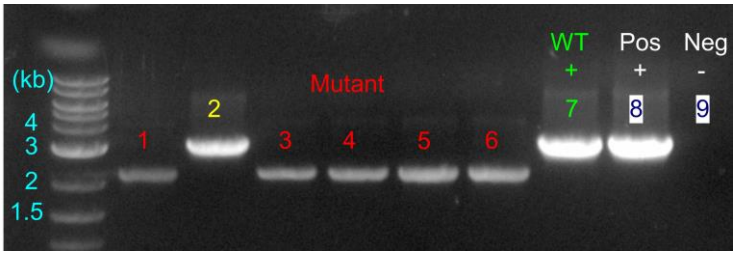
The gel image shows a WT control colony (CD2301L, ~3kb, green, 9) and Mutant obtained colonies (CD2310L PTP Ko, ~2.2kb, red, 1-8), alongside a negative (10) control.

D CD2301 PTP Comp



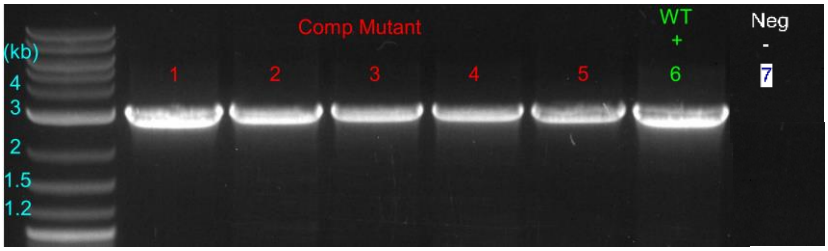
The gel image shows a mutant control colony (CD2310L PTP Ko, ~2.2kb, purple, 7) and Mutant obtained colonies (CD2301L PTP Comp, ~3kb, red, 1, 3-6), alongside a negative (9) and positive (CD2301L gDNA, 8) controls.

E CD2315 TPPSP Ko



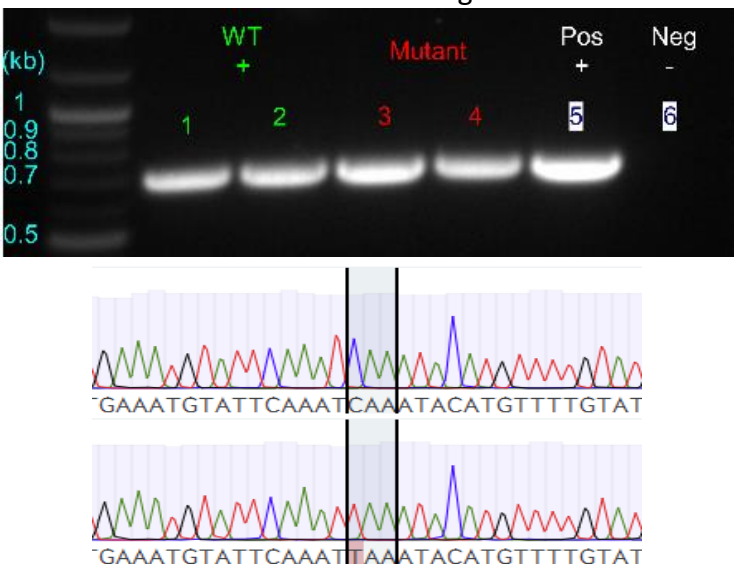
The gel image shows a WT control colony (CD2315L, ~3kb, green, 7) and Mutant obtained colonies (CD2315L TPPSP Ko, ~2.2kb, red, 1, 3-6), alongside a negative (9) and positive (CD2315L gDNA, 8) controls. The non-mutated colony (2, yellow) was discarded.

F CD2315 TPPSP Comp



The gel image shows a WT control colony (CD2315L, ~3kb, green, 6) and Mutant obtained colonies (CD2315L TPPSP Comp, ~3kb, red, 1-5), alongside a negative (7) control.

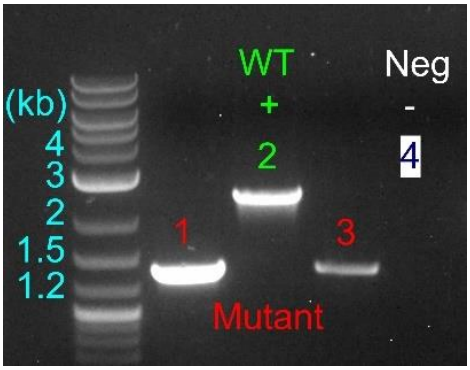
G CD2315L Kelch TargetAID



The gel image shows WT control colonies (CD2315L, ~0.7kb, green, 1-2) and Mutant obtained colonies (CD2315L TA, ~0.7kb, red, 3-4), alongside positive (CD2315L gDNA, ~0.7kb, 5) and a negative (6) control.

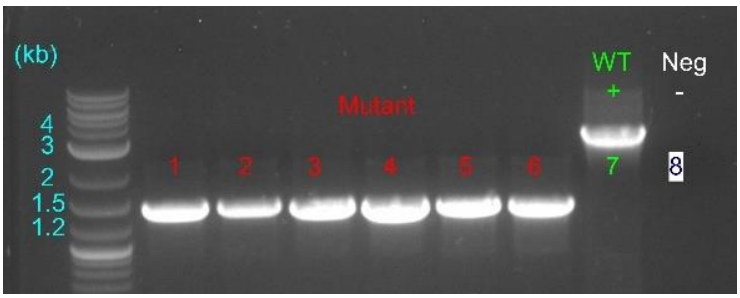
The sequencing result shows the single bp change from C to T, inducing a stop codon.

H CD2315L Kelchv3 Ko



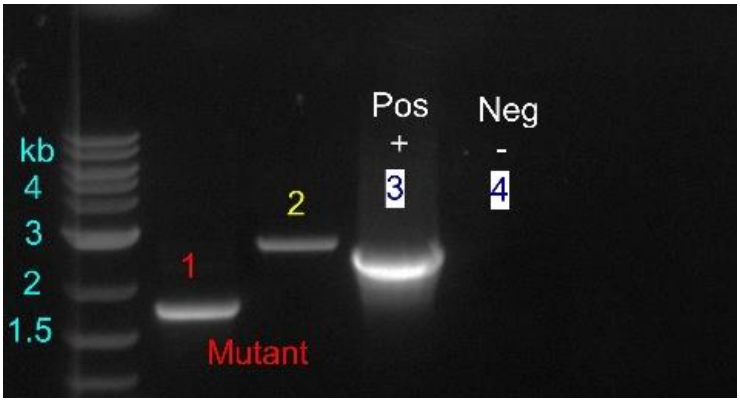
The gel image shows a WT control colony (CD2315L, ~2.5kb, green, 2) and Mutant obtained colonies (CD2315L Kelch v3, ~1.4kb, red, 1,3), alongside a negative (4) control.

I CD2315L DKO



The gel image shows a WT control colony (CD2315L, ~3.4kb, green, 7) and Mutant obtained colonies (CD2315L DKO, ~1.3kb, red, 1-6), alongside a negative (8) control.

J CD522418L TCFP Ko



The gel image shows a Mutant obtained colony (CD522418L TCFP Ko, ~1.8kb, red, 1), alongside negative (4) and positive (CD522418L gDNA, ~2.6kb, 3) controls. The non-mutated colony (2, yellow) was discarded.

Figure 4.5 Gel images confirming desired mutations via PCR.

The figure shows gel images of PCR products resulting from colony DNA amplification to demonstrate the mutants obtained. Each gel has its own mini-legend. The gels all have the Quick-Load® 1 kb Plus DNA Ladder (NEB) as a size indicator and were carried out as described in 3.6.5.

4.2.4.2 Generation of putative RBP complementation mutants

To show that any effect seen in the RBP Ko mutants was not due to a polar effect of the knock-out, the generated RBP mutants were complemented (Table 4.4, Figure 4.5) by restoring the RBP gene in place with a watermark. The complementation was carried out once the RBP Ko mutants were tested (4.2.6), resulting in three complement mutants for: ΦCD1801 *TCFP*, ΦCD2315 *TPPSP* and ΦCD2301 *PTP* genes. The complemented mutants were generated in the same method as above, with the HA cargo containing the deleted gene (with a watermark mutation) and a sgRNA targeting bookmark BM6.

In ΦCD1801 the *TCFP* gene was constructed to contain a single base pair change, resulting in a silent mutation where the encoded amino acid remains the same; this inserts a watermark into the strain allowing it to be distinguished from the WT and any WT contamination. In addition to sequencing, the watermark removes a restriction site, so the PCR products can be digested and compared while keeping the aa sequence the same (Figure 4.6).

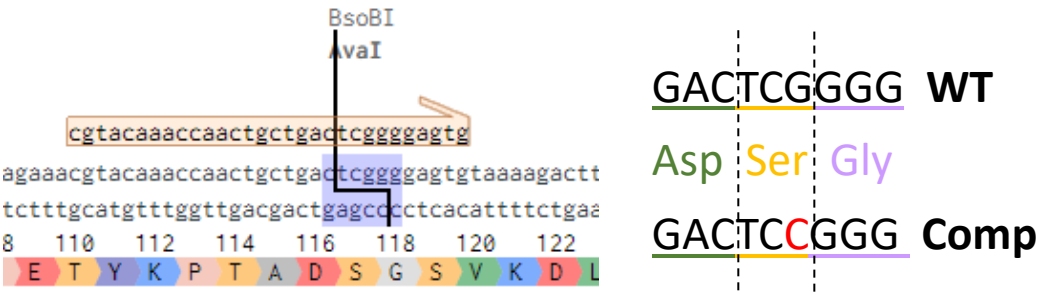


Figure 4.6 Engineering of the watermark.

The figure shows the DNA sequence of the *TCFP* gene, with the recognition sequence for *AvaI* restriction site highlighted (left). Using primers and SOEing PCR the *TCFP* gene contained between the HA was designed to have a single base pair change, from Guanine to a Cytosine (Right, in red), removing the restriction site in the completed version of the strain.

The total PCR product was 3018bp in size; when the WT was digested with *Aval*, it was cut into two fragments, 1461bp and 1557bp. The complemented mutant, although the same in total size, was not cut by the enzyme as the restriction site was removed through the change of base pair and the enzyme can no longer recognise it, thus the fragment remained 3018bp (Figure 4.7).

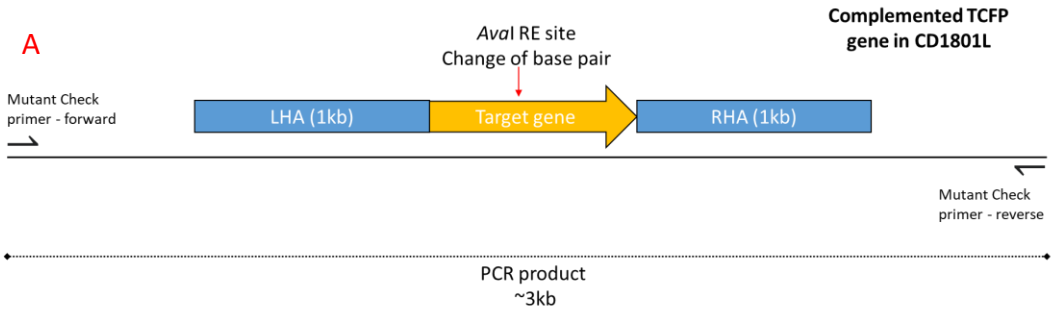


Figure 4.7a Schematic representation of the complemented TCFP locus in ΦCD1801.

The figure shows the expected layout of the TCFP locus once it is complemented. The watermark is in the middle of the TCFP gene and is highlighted on the diagram. The primers used for diagnostic PCR are placed and the undigested product is expected to be approximately 3kb in size.

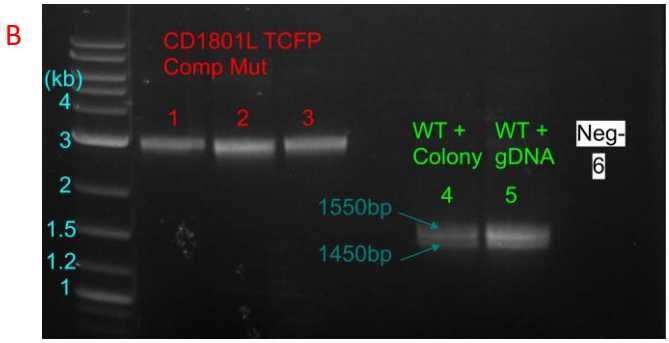


Figure 4.7b Gel image of the digested complemented and WT DNA of CD1801L strains by *Aval*.

The gel image shows the digestion of PCR amplified DNA from WT CD1801L colonies and extracted gDNA (4 and 5, green). Due to presence of the restriction site, the DNA has been cut into two fragments (~1.45kb and ~1.55kb, dark blue). The image also shows DNA of 3 complemented isolates (1-3, red) which has been digested, but not cut as the restriction site is no longer present, thus it remained the same size (~3kb).

4.2.5 Induction of mutants

Once mutants were obtained, verified, and lost the CRISPR/Cas9 plasmids, the next stage was the induction of mutant phage from the lysogenised bacterial hosts. The mutant Ko phage: ΦCD1801 TCFP, ΦCD2301 PTP, ΦCD2315 TPPSP, ΦCD2315 Kelch v3 and ΦCD2315 Kelch TA were induced out through the addition of Mitomycin C at 3µg/ml (Clokie and Kropinski, 2009; 3.5.2.2), alongside the WT lysogenised parent strains. The induction allowed the testing of mutant phage which is required to evaluate whether the selected gene is the RBP. This was repeated once the complementation mutants were generated, to show restoration of phenotype.

Mutant phage originating from CD1801L, CD2315L and CD2301L induced out successfully, verified by plaque assay (in cases where plaques were observed) in the propagating strain (Table 4.5) and presence of phage gDNA in the supernatant confirmed by PCR (whether plaques were observed or not). However, CD522418L WT and mutant were unable to induce using the standard protocol. Modifications, including doubling Mitomycin C (to 6µg/µl) and wrapping tubes in tinfoil (Mitomycin C is light sensitive) were applied without success. Further attempts using UV light may be attempted in the future.

Table 4.5 List of all the induced mutant phage alongside WT controls. The table shows the detected p.f.u./ml at the time of induction, indicating the amount of phage present.

Phage	Mutant (p.f.u./ml)	WT (p.f.u./ml)
ΦCD1801 TCFP Ko	0	$4.8x10^8$
ΦCD1801 TCFP Ko with BM6	0	$1.45x10^7$
ΦCD1801 TCFP Complement	$4.8x10^8$	$6.5x10^8$
ΦCD2315 Kelch v3 Ko	$1.1x10^2$	$3x10^2$
ΦCD2315 Kelch TA Ko	$1.25x10^3$	$1.35x10^3$
ΦCD2315 TPPSP Ko	0	$3x10^2$
ΦCD2315 TPPSP Ko with BM6	0	$9x10^2$
ΦCD2315 TPPSP Complement	$8.5x10^2$	$9x10^2$
ΦCD2301 PTP Ko with BM6	0	$1.4x10^3$
ΦCD2301 PTP Complement	$6x10^2$	$1.75x10^3$
ΦCD522418 TCFP Ko with BM6	0	0

4.2.6 Testing of mutant phage

The phenotypes of the mutant phages were tested by means of a binding assay (3.5.6). Briefly, the assay involves incubation of phage with its propagating host cells (1:10 ratio, unless specified) to allow binding to the host cell surface, then following resuspension, the supernatant is removed and filter sterilised. All the unbound, "free", phage are in the supernatant which is then used in a standard plaque assay (3.5.1). This determines the amount of phage that did not bind, and when compared to the WT, the degree of binding can be established in the mutant. WT phage and phage only (no bacterial cells) controls are used as a baseline so the percentage of bound phage can be calculated. In WT samples, it was expected that most of the phage will bind to the cells, therefore the number of free phage will be lower when compared to the no cell control. The mutant phages were expected to exhibit a much lower, if not a complete lack of binding as the suspected RBP responsible for attachment has been removed. It was expected that the complementation mutants would behave like WT phage, as the RBP gene would be restored, allowing the phages to bind again.

4.2.6.1 ΦCD1801

4.2.6.1.1 Binding assay

The initial assay showed that as expected, with the ΦCD1801 WT phage, almost all, 99.1%, of the phage particles were bound to cells (Figure 4.8). However, with the mutant phage, no phage plaques were detected. Initially, this may suggest that the mutant phage was unable to induce or form complete infectious virion particles. However, the presence of the mutant phage gene was confirmed by PCR in the filtrate, hence another more likely explanation was that, as the RBP has been removed, the phage is unable to bind, therefore it cannot infect the cells. This was expected, as the initial induction also showed no plaques, and the binding assay relies on plaque formation which can only happen when virions that are able to bind. The assay showed complete restoration of function (98.6% binding) upon complementation of the RBP gene, as expected.

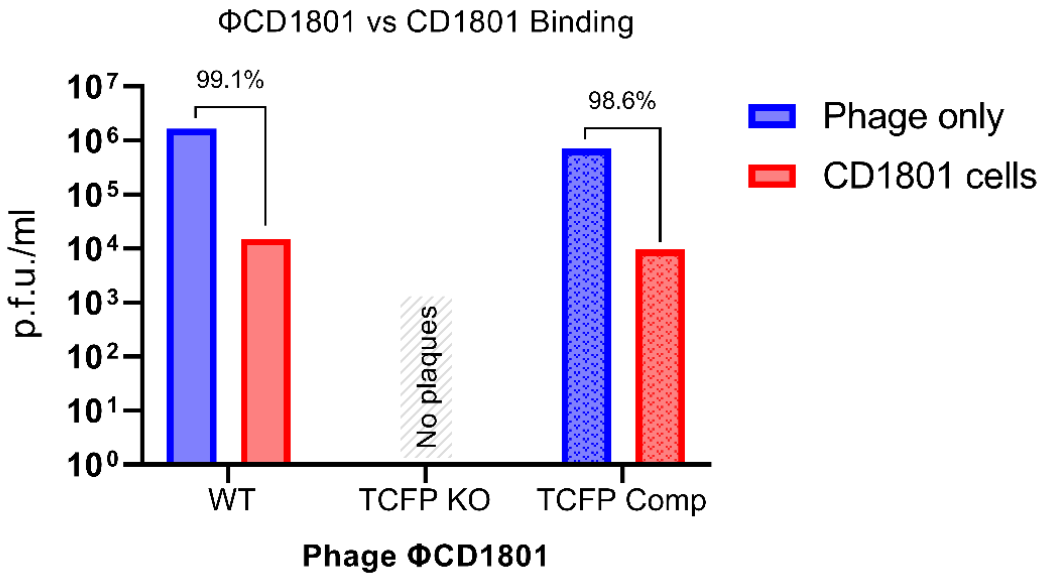


Figure 4.8 Binding assay of ΦCD1801 phages with CD1801 host cells.

The figure shows the binding profiles of the mutant phages in comparison to the WT. Almost all WT phages were able to bind to the cells, hence the sharp decrease in the free phages (indicated by column height). The same was observed for the complemented mutant (TCFP Comp). However, no plaques were seen for the RBP mutant (TCFP Ko). There was no significant difference detected between the WT and the TCFP Comp infection titres (One-way ANOVA, Dunnett's multiple comparisons p=0.68), showing that the complemented mutant behaves like the WT phage. The data presented here was generated during a single experiment but is representative of an overall observed trend.

This is further proof that the *TCFP* is the likely RBP encoding gene in ΦCD1801 as when the gene is removed the binding is completely unable to take place, however when the gene is complemented back in, the binding is restored to WT. The complementation shows that this behaviour is a likely result of the removal of the *TCFP* gene and not any downstream effects or unpredicted knock-outs as the gene was complemented into the mutant *TCFP* Ko BM6 strain. If any other unpredicted effects happened in the mutant, they were not accounted for as only the *TCFP* gene was restored and all indications so far show that it is responsible for the predicted RBP function.

4.2.6.1.2 TEM

The mutant showed complete lack of binding, however other tests were necessary to show that the mutant phage still forms a complete virion particle. The presence of phage was verified through phage gDNA PCR showing presence of mutant gene, indicating induction. Although, the induction levels could not be verified, owing to the inability for mutant phages to form plaques, therefore it was assumed that rate of induction was comparable to the WT phage. However, it was unknown whether the phage particle assembled correctly as the deleted *TCFP* gene could be responsible for other functions and prevent the virion assembly due to a missing part or an incomplete mechanism. To verify the assembly and appearance of the phage, TEM was carried out (3.5.5).

The TEM showed that the mutant phage virion particles were present in the sample and that they assemble correctly with all the structures intact (Figure 4.9). It also indicated similar levels of induction based on the number of virions observed during microscopy, which typically requires at least 10^{6-7} p.f.u./ml, for virions to be seen. At this resolution, the WT and mutant particles could not be distinguished from each other and appear to look the same. The knocked-out gene, *TCFP*, encodes the RBP which is most likely to be the tail fibres, normally located at the terminal end of the phage tail; these are not visible under TEM as they are extremely small, therefore were unable to be detected. However, if seen, the tail fibres, likely, would only be present on WT and not the mutant particle. This verified the complete assembly of the virions and demonstrated that the lack of binding is down to the removal of the *TCFP* gene which is likely to be the RBP in ΦCD1801.

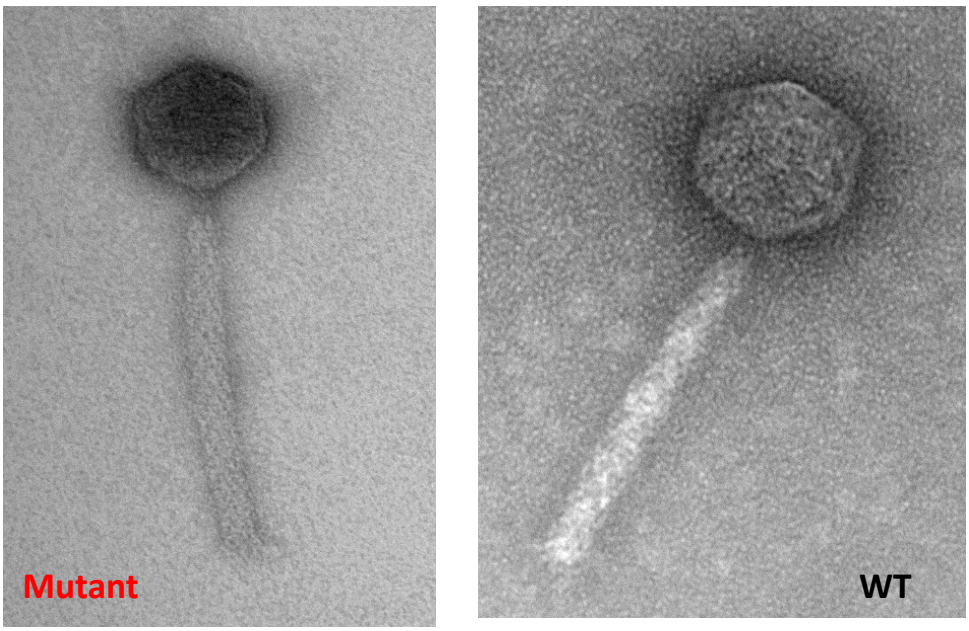


Figure 4.9 TEM of WT and TCFP Ko BM6 mutant of ΦCD1801 phage particles. The figure shows particles of ΦCD1801 phage in its WT and mutant state. The WT particle has a capsid attached to a sheathed tail structure as expected of a Myovirus. The mutant particle cannot be distinguished from the WT and shows that the mutation does not prevent particle assembly. The knocked-out gene (TCFP) likely encodes tail fibres which are hair-like structures expected at the terminal end of the tail; they cannot be seen on either particle as the resolution of TEM is too low to determine their presence.

4.4.6.2 ΦCD2301

Parallel to ΦCD1801, WT and mutants of ΦCD2301 were tested in a binding assay, and the supernatant quantified for the presence of unbound phage particles. The WT phage bound normally as expected, with 97.5% attached to cells (Figure 4.10) and no plaques were detected for the mutant phage. As previously, the complemented mutant showed restoration of the binding function, with 97.1% of phage bound to cells. The same conclusions can be drawn for ΦCD2301 as in ΦCD1801. These results indicate high potential for the *PTP* gene to encode the RBP as binding is completely absent when the gene is removed, and restored when the gene is complemented.

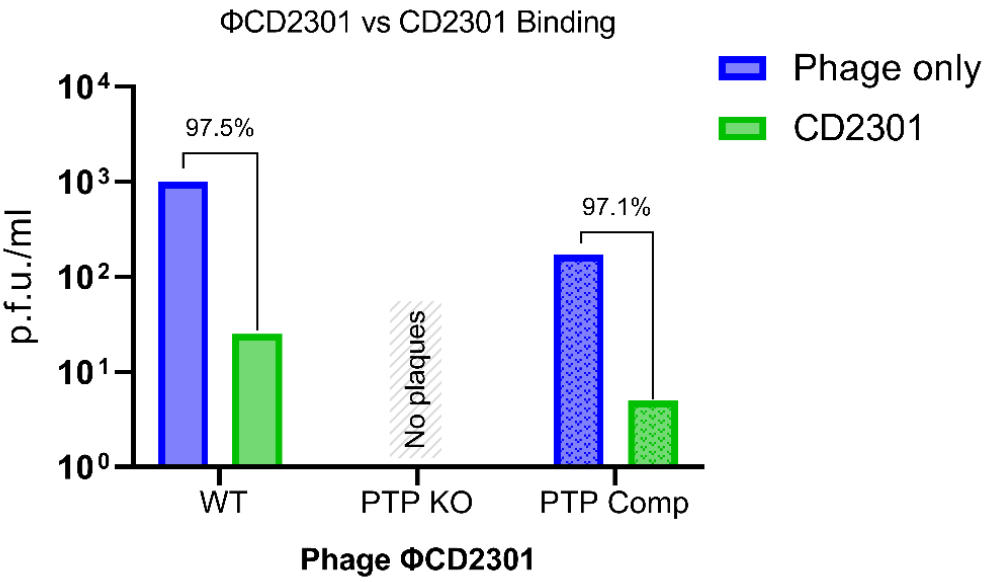


Figure 4.10 Binding assay of ΦCD2301 phages to CD2301 host cells.

The figure shows the binding profiles of the mutant phages in comparison to the WT. Majority of the WT and complemented mutant (PTP Comp) phages were able to bind to cells conferring a high decline in the amount of free phages present. The RBP mutant (PTP Ko) was unable to plaque therefore there was no binding observed. There was no significant difference between the WT and the PTP Comp titres (One-way ANOVA, Dunnett's multiple comparisons p=0.53), showing that the complemented mutant restores the binding to WT levels. The data presented here was generated during a single experiment but is representative of an overall observed trend.

The induction of these phages was verified through PCR; however the formation of virion particles could not be confirmed through TEM. This is due to the ΦCD2301 phage inducing at 10^3, which is too low to use for TEM. The WT virion has been imaged in the past (Whittle et al., 2022), however that was following several rounds of titre increase assays. It was not possible in this case, as the mutant virion is unable to bind and form plaques, therefore its titre could not be increased. The only way to obtain the required >10^7 titre, would be to induce from an extremely high volume of overnight cultures (1,000-10,000ml), which was not logistically possible, due to anaerobic-condition space constraints and time. It can only be inferred, based on ΦCD1801 that this phage and its mutants are likely to behave similarly, due to high genome and structural homologies.

4.2.6.3 ΦCD2315

The same binding assay was repeated for ΦCD2315 and its respective mutants (Kelchv3 Ko, Kelch TargetAID, TPPSP Ko BM6, TPPSP Comp). The binding assay had to be modified as the phage induced at very low p.f.u./ml values (Table 4.5) and since so few phage particles were present, the data was initially skewed. In the original assay, a ratio of 10:1 (cells:phage) was used, and following optimisation testing, in the modified assay a ratio of 1:1 proved most consistent. The ratio had to be adapted as, originally the phage sample was diluted, due to a low 10^{2-3} p.f.u./ml of ΦCD2315 compared to 10^{7-8} of ΦCD1801. To account for low induction titre, the volume was increased leading to a higher number of particles in the binding assay, and more reliable data. The titre was not increased to account for the low induction, as that would skew the data further, as titre does not always increase evenly and the assay included different mutant phage, which had unpredictable binding.

In the case of ΦCD2315, two candidate genes were screened for their ability to be a RBP gene; *Kelch* (Kelch TA, Kelchv3 Ko) and *TPPSP* (TPPSP Ko, BM6). For the Kelch gene, two versions of the mutant were obtained; a traditional CRISPR/Cas9 knock-out where the Kelch gene was removed and replaced by a scar and a CRISPR TargetAID version where a single base pair was changed (4.2.2.1) leading to a stop codon in the middle of the gene and thus incomplete gene products generated during transcription and translation. It was hypothesised that since only roughly half of the gene would be transcribed, it would not be enough for the resulting protein to carry out its function and this mutant would behave in a similar manner to a traditional CRISPR/Cas9 knock-out.

In the initial binding assay (Figure 4.11), only the TPPSP mutant exhibited no binding, therefore only TPPSP was complemented (TPPSP Comp) and the binding assay was repeated to include it. The assay showed the WT phage binding as expected, with 97.4% particles bound in the initial assay. The Kelch mutants were able to plaque, and bound at rates comparable to the WT phage, 100% and 98% respectively for the Kelch TA Ko and Kelch v3 Ko. The Kelch v3 Ko mutant exhibited a slightly unusual binding rate of a 100%, where all the available phage particles were bound, this was likely because the initial induction titre was low compared to the WT and other mutants. Thus, less phage particles were available in the first place and there was a higher chance for all of them to bind. This was not seen in the TA mutant as it

induced at a higher rate, but still behaved as the WT. The induction titres were observed to be more variable than expected, than those observed in other phages; some statistical significance (Wilcoxon test, p=0.032) was found between the samples. This could be due to natural biological variation, an effect of mutations or some other unknown factor at play. This assay showed that the *Kelch* gene was unrelated to binding and infection of CD2315 as no significant difference was found between the binding rates of WT and both Kelch mutants (Wilcoxon test, p=0.25), and they all bound at 97.4-100%. It is unknown whether there is a difference between the two types of mutations as the phenotype was the same as in the WT, regardless of the presence of the gene.

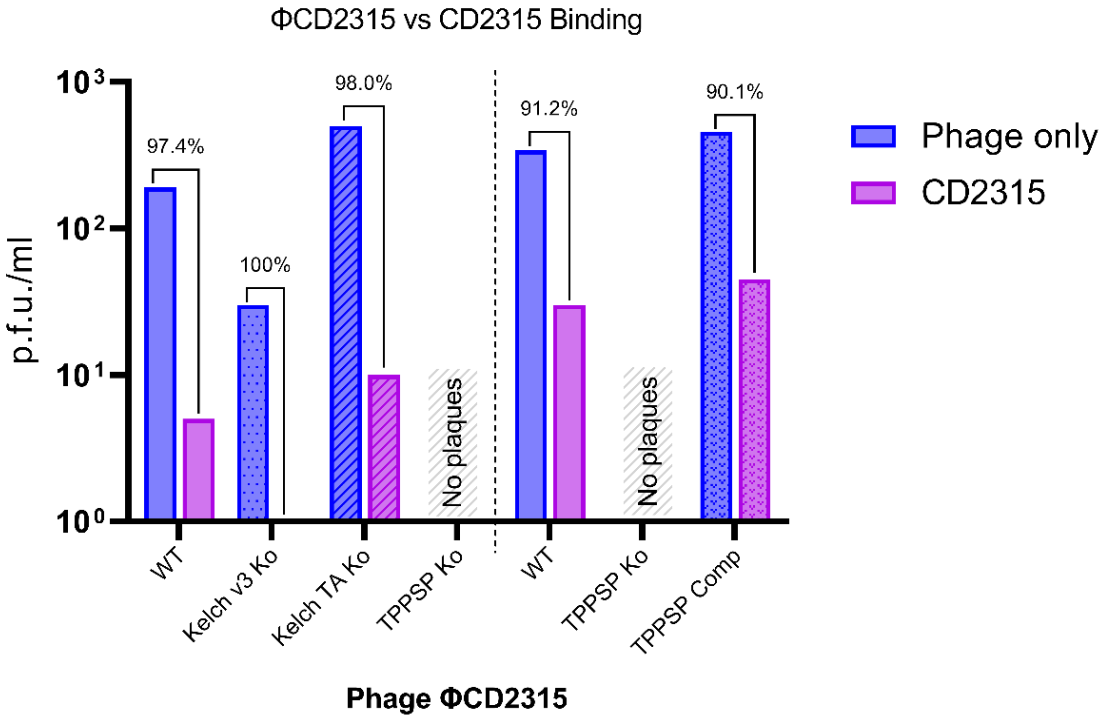


Figure 4.11 Binding assay of ΦCD2315 phages to CD2315 host cells.

The figure shows the binding profiles of several mutant phages obtained through two plaque assays in comparison to the WT. Most phages were able to bind to the WT in both assays, as well both of the Kelch mutants and the TPPSP Comp mutant. The TPPSP Ko exhibited no plaquing in both assays, hence no binding was observed pinpointing it as the RBP. The data presented here was generated during a single experiment but is representative of an overall observed trend.

The TPPSP mutant was unable to plaque implicating it as the RBP gene, thus it was complemented and re-tested. Upon complementation, the binding rate was restored to that of the WT, at slightly lower than usual rates of 90.1% and 91.2% respectively. Although these rates were not a concern as their differences were found not to be statistically significant (One-way ANOVA, Dunnett's multiple comparisons p=0.63) and lower rates of binding have been observed before (~84% or higher) in WT phages. For the same reasons as in ΦCD2301, the mutants were unable to be imaged under TEM, and their titres were even lower than those of the aforementioned phage. Thus, the complete loss of function upon the deletion of the gene, and complete restoration of function upon the complementation are the only evidence indicating this gene is involved in binding, however there could be other reasons for lack of infection.

4.2.6.3.1 Investigation of alternative host range of ΦCD2315 mutants

The lack of change in phenotype in Kelch mutants appeared inconsistent with the bioinformatic data, labelling it as a *tail fibre protein H,* previously known to confer specificity (Garcia et al., 2008). As Siphoviridae, are known to have a broader host range than Myoviridae, it could have been possible that the *Kelch* gene is used to encode an alternative RBP which is used to infect other strains, not CD2315. It was unexpected that a gene in the tail module appeared to play no role in binding and infection, therefore its potential role in other strains was investigated.

The induced mutants were assayed against a panel of 25 other strains previously known to be infected by ΦCD2315, alongside the propagating host, CD2315 (Michaella Whittle, unpublished data). Compared to the WT, the rates of infection by different mutants seemed quite erratic (Figure 4.12) and differed between the mutants, within the same strain. Induction rates were expected to be similar as the level of infection by the Kelch mutants and the WT in CD2315 was highly similar. Further testing showed that the TPPSP mutant (Figure 4.12, TPPSP R2) was able to plaque on all strains, not including CD2315. This result was extremely unexpected.

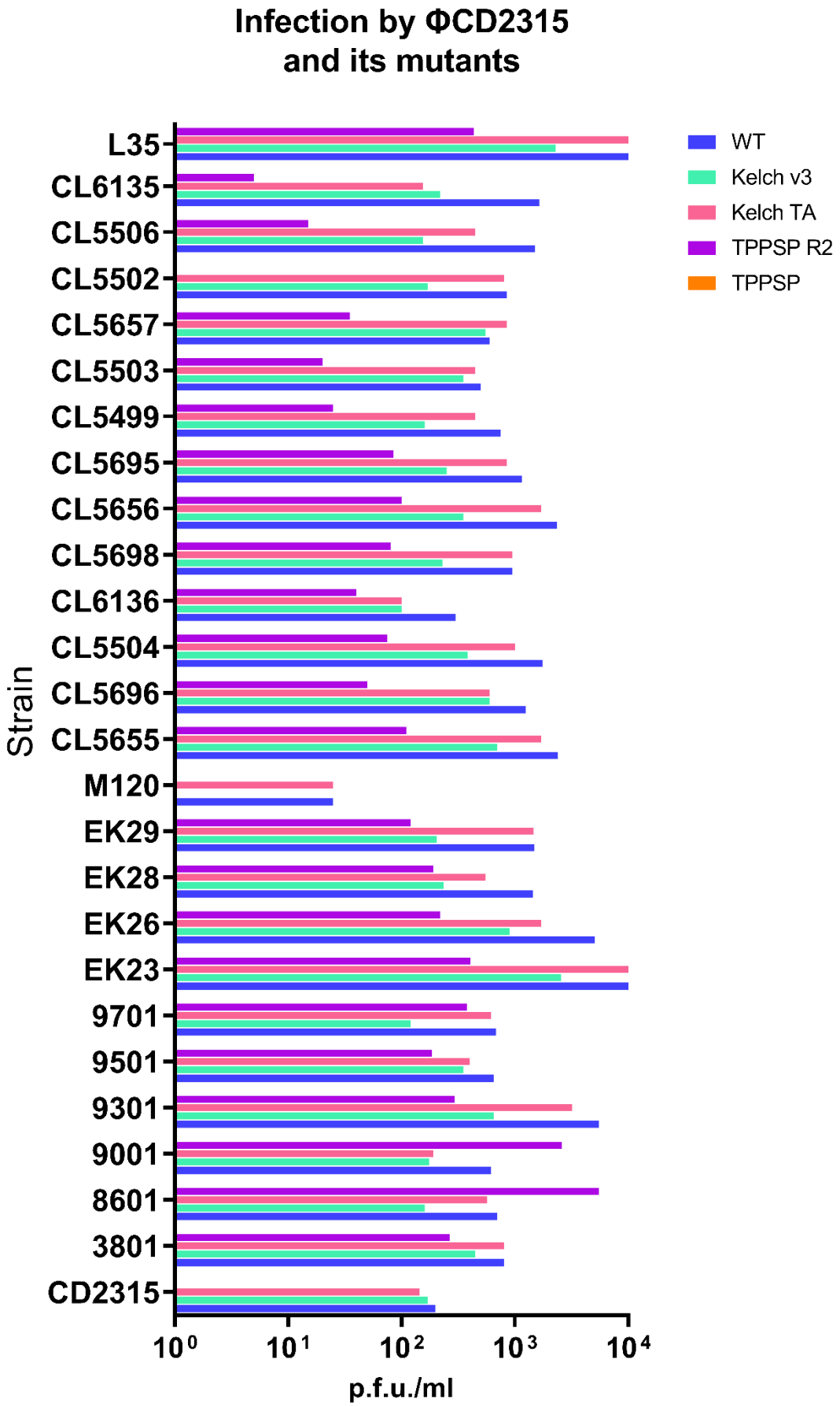

Infection by ΦCD2315
and its mutants
WT
Kelch v3
Kelch TA
TPPSP R2
TPPSP
Strain
L35
CL6135
CL5506
CL5502
CL5657
CL5503
CL5499
CL5695
CL5656
CL5698
CL6136
CL5504
CL5696
CL5655
M120
EK29
EK28
EK26
EK23
9701
9501
9301
9001
8601
3801
CD2315
10⁰
10¹
10²
10³
10⁴
p.f.u./ml

Figure 4.12 Infection of multiples hosts by ΦCD2315 and mutants.

The figure shows an extensive panel of strains, previously used to test the EOP of ΦCD2315. The alternative hosts were used to test whether the Kelch mutants had the ability to infect them. The data showed that all mutants, including TPPSP were surprisingly able to infect. The initial test revealed no infection (TPPSP, in orange), however a further round of plaque assays revealed the opposite (TPPSP R2, in purple). The p.f.u./ml varied widely between and within the strains, between different mutants, showing widespread inconsistency.

To further investigate this, a double knock-out mutant (CD2315L DKO), removing the *Kelch* and *TPPSP* genes was obtained using the same methods as earlier in this Chapter. The mutant was then induced alongside the WT and the other previously generated mutants and used to infect CD2315 and three other strains from the previous panel, selected at random. The assay showed, as previously that the WT, and Kelch mutants were able to infect CD2315, whilst TPPSP Ko and the new DKO mutants were not able to do so. This was expected as the DKO did not contain the *TPPSP* gene, shown to be necessary in infection (4.2.6.3). However, the other strains were infected by the WT and all mutants to varying degrees, again at slightly more varied rates than expected (Figure 4.13).

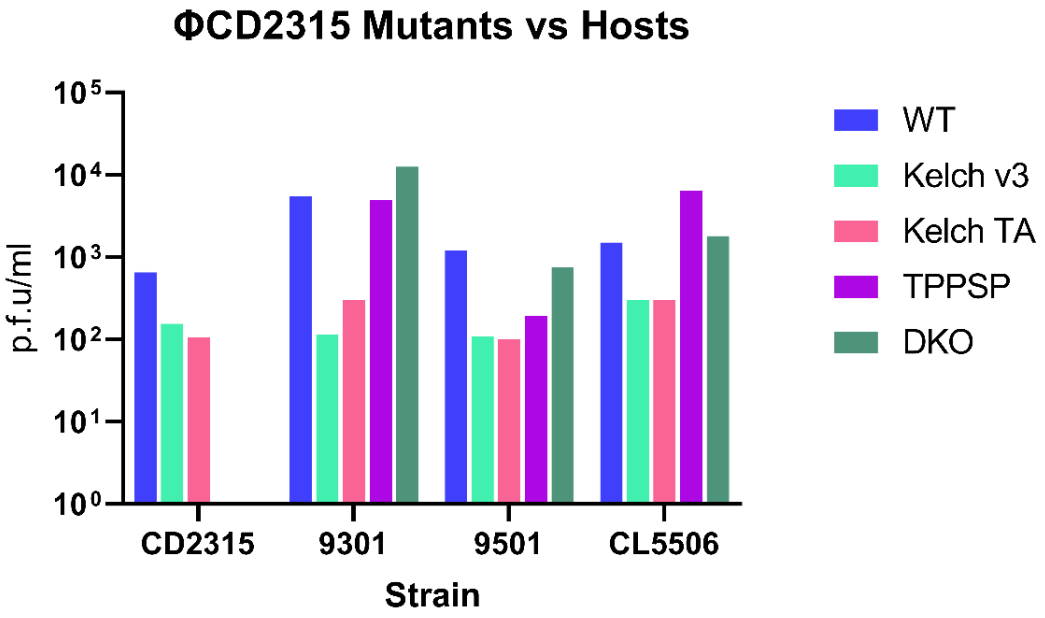


Figure 4.13 Infection of hosts by ΦCD2315 DKO mutant.

The figure shows an additional mutant, DKO, and its infection rates alongside a comparison panel comprising of the WT ΦCD2315 and previous mutants. All mutants were able to infect the three additional strains, whilst only the WT and Kelch mutants were able to infect CD2315.

This erratic data could be explained by the presence of another phage, therefore the propagating host, CD2315 was assayed for a prophage. This was the only logical explanation as the TPPSP mutant should not be able to plaque, and neither should DKO, while the infection data obtained should be more consistent. If present, during routine protocols, the prophage would be induced out alongside the ΦCD2315 phage unknowingly, but it would usually not interfere as long as the induction filtrate was targeted against the CD2315 host. This is because if the prophage is already contained within the host it cannot re-enter. However, other strains may be susceptible to infection by the prophage as well as ΦCD2315, depending on phage RBPs and host factors. To test this, the propagating strain CD2315 and the lysogenised host CD2315L (containing ΦCD2315) were both induced using Mitomycin C (3.5.2.2) and assayed against another panel of different strains.

The data confirmed the presence of a prophage (ΦCD2315 Pro), as the filtrate from CD2315 which did not contain ΦCD2315, was able to produce plaques on alternative hosts, but not CD2315 (Figure 4.14). The prophage cannot infect CD2315 as is contained within that strain, however it can infect other strains which are not lysogenised by it or other phages. The true titre cannot be calculated for the prophage as it is unknown which phage induces more, and the degree of induction seems to vary with each induction event, and even within events. The degree of difference in titres is highlighted in Figure 4.14, however it cannot be reliably used to predict which phage induces more, as the amount of ΦCD2315 Pro compared to just WT ΦCD2315 present in the sample cannot be calculated.

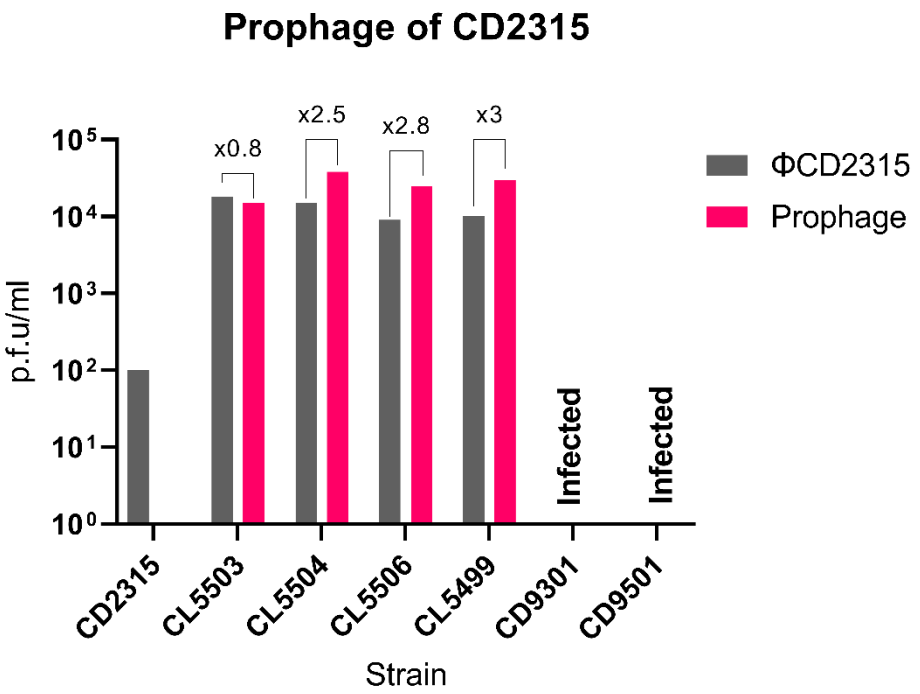


Figure 4.14 Prophage, ΦCD2315 Pro, detection.

The figure shows titres of assayed filtrates from CD2315 and CD2315L used to screen for presence of a prophage. The prophage was able to plaque on all strains, although its titre nor the titre of ΦCD2315 were able to be determined for CD9301 and CD9501. The presence of plaques in the induction filtrate from CD2315 confirms the presence of another phage. For CL strains, the titre difference is indicated, although the difference cannot be attributed to the prophage alone, nor used to calculate the two titres as other factors like induction levels and infection mechanics of individual hosts cannot be controlled for.

The presence of the prophage in CD2315 explains the unreliable results obtained when infecting other hosts with the mutants. This data cannot be used to infer anything about the function of the *Kelch* gene (or the *TPPSP* gene) in the phage as it is skewed by the presence of the prophage. The function of this gene remains to be elucidated.

4.3 Discussion

4.3.1 Accuracy of selection of target genes

The target genes selected for this study were based on a previous study (A. Dowah, 2018) and previously described criteria. It was only a single study where an anti-suspected RBP protein was used to block binding of phage. Based on the sequence of the antagonised protein, a similar gene was found in the initial phage candidate, ΦCD1801. While this evidence may seem sparse due to the very vague link, other logical criteria were also applied. The gene location was key, as phage tail genes in other Gram-positive infecting phage were also found between the *tail tape measure protein* and the *endolysin/holin* genes; for example, as is the case in *Staphylococcus aureus* (X. Li et al., 2016). The tail structural gene module was identified, and the genes within it were excluded one by one, due to either size or already having an already attributed function. The allocated functions were however based only on annotations (Joshi & Xu, 2007) due to sequence similarity and could be wrong; unfortunately, there is no other way to account for that, other than through experimentation.

Through exclusion, the most logical target was identified for each phage. Most genes were excluded based on their high similarity to genes with already assigned functions such as tail structure genes, or *tail tape measure protein* genes. However, for ΦCD2315, there was no clear single candidate gene, and two most likely options emerged. Both genes had merit as a potential RBPs, as *TPPSP* was highly similar to other identified genes, however the *Kelch* gene, while an outlier in similarity to other candidate genes, was highly similar to *tail fibre protein H* which was previously implicated in conferring specificity in phage binding. Therefore, both had potential to be RBPs and were worth testing. If anything, whichever gene did not emerge as the RBP through the binding assay, would only serve to provide further evidence for the RBP gene as data emerged.

4.3.2 The RBP of ΦCD2315

Through experimental exploration it was shown that *TPPSP*, and not *Kelch* was more likely to be the RBP as the binding diminished upon its knock-out, whereas it remained the same in Kelch mutants as in WT phage. The difficulty in predicting which gene was the RBP correctly in regards to the three phages studied, was due to ΦCD2315 being a Siphovirus, whereas all the other phages mentioned in the study

are from the *Myoviridae* family. This reflects the population of currently isolated phages infecting *C. difficile*, as majority are *Myoviridae*, with fewer *Siphoviridae*. The currently isolated *Siphoviridae* tend to be a lot more genetically diverse and infect a wider range of hosts (Sekulovic et al., 2014), hence it was included in the study. This is promising for phage therapy as phages infecting a wider range of *C. difficile* hosts are always desired to reduce the number of phages required in a therapeutic cocktail, therefore ΦCD2315 and its RBP were worth investigating.

It is interesting that regardless of the type of mutation; whether complete knock-out of the *Kelch* gene or expected loss of function through introduction of a stop codon, either mutant does not seem to behave differently. The mutant was tested in binding and plaque assays and its behaviour is that of the WT, therefore it is unknown whether the TargetAID method worked as desired, or if the product was still partially expressed since it does not seem to have an obvious detectable function in the conditions studied. This indicated that the gene may not have a necessary function, and only confers advantage in specific conditions, which remain unknown and may not be related to binding.

A further investigation of this gene was carried out as the gene could still encode an RBP, just not for its propagating host – CD2315. The gene could encode a RBP that is necessary for it to bind to other host strains that were not tested in the binding assay. The ΦCD2315 phage has the widest host range out of all the phages used in this study and can infect a majority of the PCR Ribotype 078 strains tested within the group (Michaella Whittle, unpublished data). Therefore, due to such wide host range it may be necessary for an additional RBP gene to be present to allow infection. This theory was not without merit, as there have been previous reports (Hayes et al., 2018; Takeuchi et al., 2016) of phages using multiple RBPs, which grants them wider host range, therefore this may also be possible with *C. difficile* infecting phages.

It is unlikely that the gene has no function, as viruses, particularly phage, that can only package a very limited amount of genetic material, would evolve to maintain a gene that does not confer an advantage. Phage genomes are extremely compact; genes are often co-transcribed and have overlapping ORFs (Kang et al., 2017), with cases of genes being encoded within each other also reported (Cahill et al., 2017; Summer et al., 2007).

To elucidate whether the *Kelch* gene was a host range determining factor in alternative hosts, all the mutants were tested against a panel of strains previously infected by WT ΦCD2315. The results were extremely unexpected as all the mutants, including the TPPSP Ko, and later DKO were able to infect these strains as plaquing was observed in various degrees. Further investigation revealed that the WT CD2315 contained a prophage which was responsible for the plaques seen, rather than the mutants. The presence of the prophage was undetected as all assays related to ΦCD2315 were carried out in CD2315 which the ΦCD2315 Pro cannot re-infect as it is already in the genome (Labrie et al., 2010). However due to the prophage contamination of the induction filtrates, the function of the *Kelch* gene could not be investigated and its impact in other strains remains unknown. There is no reliable method to distinguish which phage is responsible for which plaques observed, therefore any effects of *Kelch* on alternative strains were masked by the prophage contamination. Although plaques were observed for TPPPSP Ko, they can also be disregarded as prophage contamination, and if the prophage was not present the mutant would likely not be able to infect the alternative strains, although this cannot be confirmed. The evidence for *TPPSP* being the encoded RBP can only be demonstrated in CD2315 which is not impacted by the phage. The exact levels of prophage likely vary with each induction event and the phage does not behave in a replicable manner, therefore the data is not usable. As both phages are present in induction filtrates of CD2315L they can't be distinguished from each other even when compared to the filtrate of CD2315. This is because the phages may have different binding mechanics in each strain tested, and the composition of the filtrate is unknown, hence any calculations based on infection of CD2315 are inaccurate.

The mutant phages also cannot be imaged under TEM due to the low induction titre, also a likely by-product of the prophage. The *TPPSP* gene is still the most likely candidate as the RBP, however it could be responsible for another function as it is unknown whether a virion particle (without the RBP) is able to assemble. Lack of infection by the mutant only provides limited evidence for it being the RBP. However together with the gene location, size, some homology and a mutant unable to infect, it is likely that the gene is involved in interaction with the host surface.

Advances have been made through crystallisation of an RBP of a Siphovirus infecting *C. difficile*, identified as a candidate using a similar bioinformatic approach taken

here, but in phage CDHS-1 (Dowah et al., 2021*). The data is in preprint (study deposited on bioRxiv*), therefore it should be utilized with caution. The study showed that gp22 of CDHS-1 has high homology to putative tail fibres of ΦCD111, ΦCD146 and ΦCD38.2 which are the same type of phage. Crystallisation revealed it to establish a new class of RBPs, quite large in size (~600aa). Although the genome of the studied phage has not been published, based on the aforementioned phages which have genomes available, there was no homologous protein found in ΦCD2315. The three phage genomes had a different type of architecture in the tail gene module, therefore they cannot be compared to ΦCD2315, and thus providing no contribution towards the RBP of this phage, although advancing the field further.

4.3.3 The RBPs of ΦCD1801 and ΦCD2301

The predicted RBPs of *Myoviridae* phages in this study, the *TCFP* of ΦCD1801 and *PTP* of ΦCD2301, have been shown to be involved in binding. This is evident in both phages, where the knock-out of the predicted RBP genes completely removed their ability to bind to host and the phage could not be detected in assays. Once complemented, the phage binding capability was restored to WT, where majority of the phage virions were attached to cells.

As the mutant ΦCD1801 phage was unable to bind and its numbers could not be detected in plaque assays, TEM was necessary to show that the phage virions were assembled correctly. The TEM showed that the virions looked identical to WT as the resolution is too low to show the suspected RBP. In other phage, for example *E. coli* T4, it has been shown (Hyman & van Raaij, 2018) that the tail fibres are responsible for binding, however, are too thin to be observed under TEM. The TEM also showed the presence of phage in the sample in the first place, demonstrating that there are no problems with phage induction. A phage gDNA PCR was not enough by itself to confirm that complete virions induced to a high enough degree. Based on observation of frequency of particles in the mutant and WT sample, the phage induced to the same degree as both TEM samples contained a similar amount of virions on the grid. The TEM for ΦCD2301 was not carried out as its induction titre is too low for imaging.

There is a small possibility that the *TCFP* and *PTP* genes are only accessories in binding and may not actually determine specificity, as viruses often use multiple proteins to bind and infect their hosts (Bertozzi Silva et al., 2016). For example, a

study (Bebeacua et al., 2013) showed that p2, a *Lactococcus lactis* phage attachment is a multiple step process involving initial reversible binding, which if correct will then proceed further. Similar mechanisms involving secondary binding have been reported in *C. difficile* phages (A. Dowah, 2018), although it is not known whether multiple proteins are involved, or whether this is carried out by a single RBP gene. Therefore, the knocked-out *TCFP* and *PTP* genes could be an accessory protein in the binding chain of events. In this case, this seems unlikely as there are no other candidates that could potentially be the RBP within reason as others have previously attributed functions, or do not fit the criteria for an RBP. In addition, in has been shown that the tail fibres often act as the RBP and have been shown to determine the specificity in many phages (Dowah & Clokie, 2018).

4.3.4 Summary and importance

In order to establish a path towards using phages as a *C. difficile* therapy it is necessary to characterise the binding mechanisms, so they can be strategically engineered. Alternative approaches, such as isolating new phages from the environment is costly and time consuming (Nilsson, 2014; Schmidt, 2019), and so far has not yielded a phage capable of infecting all, or even most clinically relevant PCR Ribotypes. Another approach is to combine the phages with the broadest host ranges into cocktails (Gordillo Altamirano & Barr, 2019), however this would require each preparation to be meticulously quality controlled, which maybe prove problematic and again costly. This prevents companies from pursuing this approach, and in the case of *C. difficile* it would require many (13-16) phages in a cocktail also differing by location, showcasing it as a complex approach which may not be feasible.

To overcome these obstacles in the road to phage therapy, the phage can be engineered to infect strains of interest, however an in-depth understanding of the RBP interactions and mechanisms is a prerequisite. Through a logical approach this project chapter aimed to first identify the said RBP which confers the specificity and contributes to the narrow host range. To confirm the identified targets, the suspected RBPs were knocked-out, and the complemented back in while being assayed in several tests at all stages to observe the changes in phenotype and binding behaviour.

The RBPs of ΦCD1801 and ΦCD2301 were identified to be the *TCFP* and *PTP* genes respectively, with sufficient supporting experimental data, particularly for the former. The TEM confirmed virion assembly in the TCFP Ko in ΦCD1801, and although ΦCD2301 could not be imaged its structure and genome is similar enough to also exhibit the same function as its counterpart phage. The RBP of ΦCD2315 proved more complex, and although *TPPSP* was established as the likely RBP candidate, there is less evidence for it being so, due to lack of TEM imaging and complications with prophage contamination. The other aforementioned phages cannot be used as examples, as they belong to a different virus family – the *Myoviridae*, therefore they have different architectures which are unlikely to be comparable, including different tail mechanics to *Siphoviridae*.

Once more RBPs have been identified, the data can be used in multiple ways. An informed engineering approach can be used to either modify the RBP to exhibit a broader host range or, exchange the RBP for one that confers broader specificity, although only a limited pool of those is available. Alternatively multiple broad receptors can be introduced, however this could prove difficult as phage have limited genetic material packaging capacity and having too many RBPs may lead to an incorrectly assembled virion, unable to infect its hosts.

Few existing projects have used CRISPR/Cas9 systems to engineer phages infecting *E. coli* (Tao et al., 2017), *Streptococcus thermophilus* (Martel & Moineau, 2014), *L. lactis* (Lemay et al., 2017) and *B. subtilis* (Schilling et al., 2018), however this is the first project to focus on *C. difficile*. Thus far, engineering of phage RBPs has been established in Gram-negative bacteria, mainly *E. coli* (Ando et al., 2015; Yehl et al., 2019; Yosef et al., 2017). For Gram-positive species, a lot less data and not many approaches are available. Although, in this study (Dunne et al., 2019), a similar approach as in this project was applied, where the authors first identified the potential RBP and verified it using a structural biology approach. Next, using targeted mutations they generated mutant RBPs which resulted in shifting of the host range. Then, the authors produced a phage carrying two RBPs making it polyvalent (Nobrega et al., 2018), then exchanged the domains within them to develop "chimeras" leading to an expanded host range.

A similar approach might be applied in the future of *C. difficile* here where identification and more detailed understanding of RBPs within this project will

provide the necessary data to allow logical engineering, whether its polyvalent phage, targeted mutations or even a combination of both. The next step in RBPs of *C. difficile* phages is to establish whether the identified genes are able to confer specificity on their own. It also needs to be established whether RBPs are harmonious parts able to be shuffled between phages and form hybrid virions while maintaining particle integrity through compatibility to other tail structural components.

Chapter 5:
Further characterisation of the tail region to achieve a change of host range in phages infecting *Clostridium difficile*

Chapter 5. Further characterisation of the tail region to achieve a change of host range in phages infecting *Clostridium difficile*

5.1 Introduction

Before the therapeutic potential of phage therapy can be realised for the treatment of CDI, a number of challenges must first be circumvented. Chief among these is the provision of broad host range phage, either through their isolation or the engineering of phage RBPs. While phage therapy has previously been used to effectively treat a number of infections, they have all relied on lytic phages to infect appropriate hosts without the added complication of lysogeny. However, all isolated phages of *C. difficile* described in the literature have been lysogenic. Meanwhile, attempts to engineer lytic phages have fallen short (Selle et al., 2020).The lack of lytic phages infecting *C. difficile* is addressed in Chapter 6.

The other main problem restricting the potential of applying phage therapy in the treatment of CDI, is the limited host range observed for the phages isolated to date. The specificity of phages is usually to their advantage, where they only infect selected hosts and leave other bacteria unharmed. From a therapeutic standpoint, this preserves the microbial diversity of the niche, the gut, in the case of CDI, allowing a speedy recovery. This would be one of the main advantages of phage therapy against CDI as the current antibiotics employed to treat the infection are broad spectrum and therefore disrupt the microbiome, causing gut dysbiosis. This allows *C. difficile* spores to germinate and colonise the niche once again following treatment, leading to relapse or re-infection which happens in 20-25% of cases (Ramsay et al., 2018).

Currently, the isolated phages have very narrow host ranges where they can only infect select isolates of one or few ribotypes. To overcome this obstacle, this study has previously identified and characterised the genes responsible for encoding the phage RBP in Chapter 4 in three phages. To take full advantage of these novel findings, this chapter attempts to alter phage host range through genetic modification. Through this process, nearby structural tail genes were also

characterised to establish their function, specificity and contribution as a receptor-binding determinant. Many of the genes encode hypothetical proteins and lack homology to any specific functions beyond being tail genes.

The aim of this chapter was to use CRISPR/Cas9-based, genome editing to determine the exact gene set required for alteration of the host range of two *Myoviridae* phages. Once established, this can be repeated in other phages, as they are often homologous, but lack annotation of the genes involved.

5.2 Results

5.2.1 Generation and testing of Swap 1-4 mutant phages

In the previous Chapter, the genes responsible for binding were identified as the RBP genes for phages: ΦCD1801 (Ribotype 078), ΦCD2315 (Ribotype 078) and ΦCD2301 (Ribotype 014). The phages all exhibit a unique host range, therefore in theory, if the RBP genes were swapped between them (Table 5.1), the phages should be able to now infect the new host range.

Table 5.1. The table lists the RBP gene changes made in the planned 4 swaps, indicated by the change in colour. For example, in Swap 1, phage ΦCD1801 originally encoding the *TCFP* gene, would now carry the *PTP* gene from ΦCD2301, hopefully expressing its RBP. The change in RBP should lead to a change in host range, which is determined by the RBP.

Swap	Phage	Original gene, acronym	New Gene
1	ΦCD1801	*Tail collar fibre protein, TCFP*	*PTP*
2	ΦCD2301	*Phage tail protein, PTP*	*TCFP*
3	ΦCD2301	*Phage tail protein, PTP*	*TPPSP*
4	ΦCD2315	*Truncated putative phage structural protein, TPPSP*	*PTP*

5.2.1.1 Vector construction

Following the selection of Swaps, the vectors were constructed as previously described in Chapter 4 (4.2.2 and 4.2.3) and made to contain HAs targeting the genetic sequence of the original phage. The new gene was placed in between the HAs as cargo; essentially swapping the current gene for a homologous gene from a different phage. The vectors all used the same sgRNA targeting bookmark BM6, as all the deletion mutants (generated in Chapter 4) contained this scar in place of the deleted RBP gene to allow complementation and, now swapping (Figure 5.1).

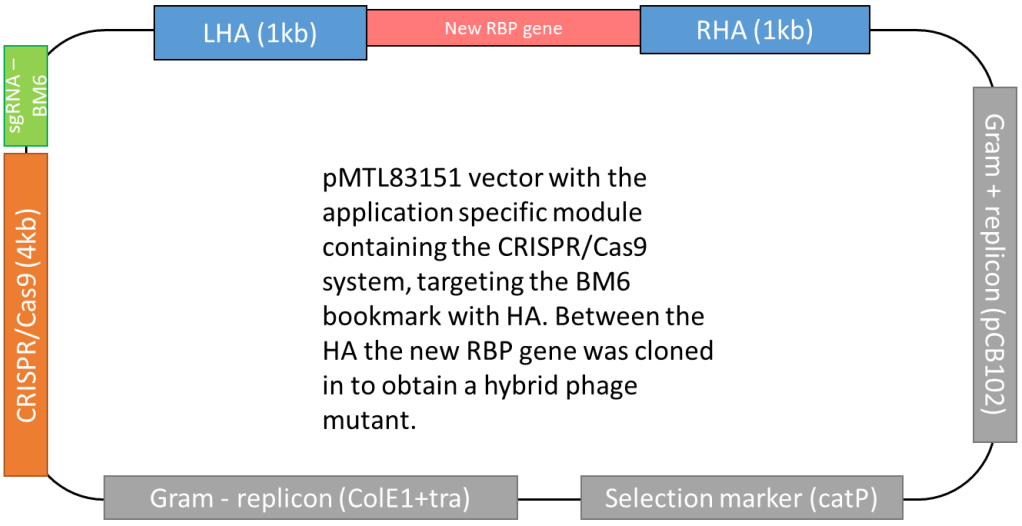


Figure 5.1 Swap plasmid structure

The figure shows a schematic representation of the vectors constructed in order to complete Swaps 1-4 and generate phages containing a RBP from related phage. This is possible as the deletion mutants contain BM6, which is targeted by this vector and used to insert the foreign RBP which is the cargo between the HAs.

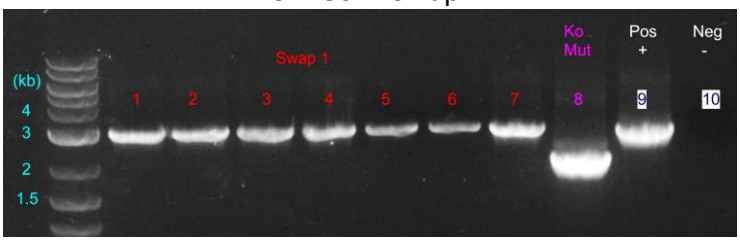
Although phage ΦCD2315 is able to infect CD1801, the reverse is not true and ΦCD1801 cannot infect CD2315. This is likely due to CD2315 containing an unknown prophage discovered in Chapter 4, which has previously interfered with results (4.2.6.3.1.). Accordingly, in view of these complications, no vectors were constructed for swapping the RBP genes between ΦCD1801 and ΦCD2315.

Four vectors were constructed as described previously, where a vector containing BM6 was digested and ligated with a HA+cargo insert constructed by SOEing PCR. The ligation was transformed into chemically competent *E. coli* Top 10 cells. Following screening via PCR and Sanger sequencing, correctly constructed vectors were transformed into *E. coli* sX and conjugated into the appropriate *C. difficile* mutants which contain the lysogenised phage genome, with a deletion of the RBP gene. The conjugation protocol optimised in Chapter 4 was used to successfully obtain the four relevant Swaps, using the modified protocol, and varying the mating time depending on the strain (3.4.1.1).

Once conjugated, the strains were induced using theophylline to obtain mutants. The mutants were verified by PCR (Figure 5.3) and subsequently confirmed for plasmid

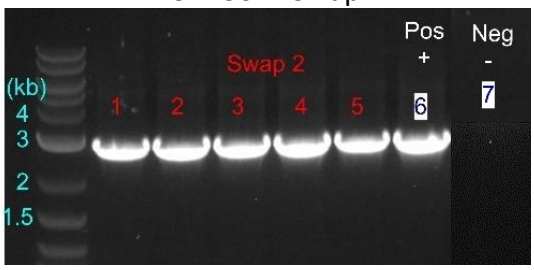
loss (3.4.1.2). The mutant strains, alongside WT controls were then induced (Table 5.2) using Mitomycin C to produce the mutant phages, as described in 3.5.2.2, and the phage preparations were stored at 4°C.

A CD1801L Swap 1



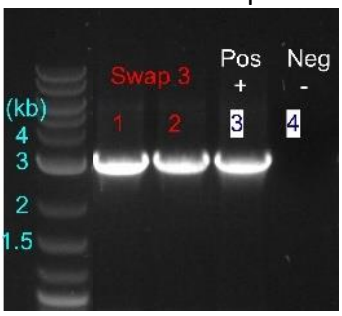
The gel image shows a Ko mutant control colony (CD1801L TCFP Ko, ~2.2kb, pink, 8) and Swap 1 mutant obtained colonies (CD1801L Swap 1, ~2.9kb, red, 1-7), alongside a negative (10) and positive (CD1801L gDNA, 9) controls.

B CD2301L Swap 2



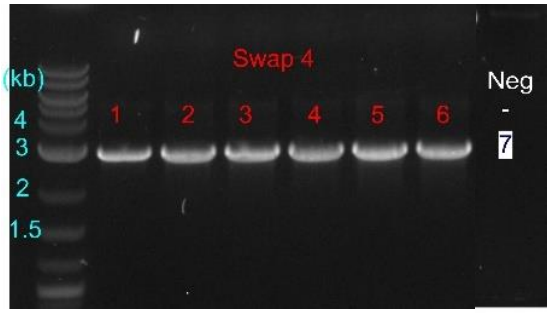
The gel image shows Swap 2 mutant obtained colonies (CD1801L Swap 2, ~2.9kb, red, 1-6), alongside a negative (7) and positive (CD2301L gDNA, 6) controls. Negative lane cropped.

C CD2301L Swap 3



The gel image shows Swap 3 mutant obtained colonies (CD2301L Swap 3, ~3.1kb, red, 1-2), alongside a negative (4) and positive (CD2301L gDNA, 3) controls.

D CD2315L Swap 4



The gel image shows Swap 4 mutant obtained colonies (CD2315L Swap 4, ~3kb, red, 1-6), alongside a negative (7) control.

Figure 5.3 Gel images confirming Swap 1-4 mutants.

The figure shows gel images of PCR products resulting from colony DNA amplification to demonstrate the mutants obtained. Each gel has its own mini-legend. The gels all have the Quick-Load® 1 kb Plus DNA Ladder (NEB) as a size indicator and were carried out as described in 3.6.5.

Table 5.2 Table showing induction titres of tested phages.

Phage	Original host strain (p.f.u./ml)	New host strain (p.f.u./ml)
WT ΦCD1801	$1.2x10^6$	
WT ΦCD2301	$1.25x10^3$	
WT ΦCD2315	$7x10^2$	
Swap 1	0	0
Swap 2	0	0
Swap 3	0	0
Swap 4	0	0

5.2.2. Testing of mutant phage – Swap 1-4

The mutant phages were tested in a basic plaque assay to see whether they were functional. The preparations were assayed against the propagating strain of each phage and against the propagating strain of the theorised new host. For example, in Swap 1, the WT phage infects strain CD1801, and the new hybrid phage should be able to infect CD2301 (as it is carrying the RBP from ΦCD2301) – therefore it was assayed against both strains.

The results indicate that the phage induction proceeded correctly as all WT phages induced at expected levels. However, all Swap mutants showed no signs of infection of either the original or the new theorised host strain. This is likely to be due to the

new RBP not functioning correctly for unknown reasons, and less likely to be due to experimental issues, since the controls induced correctly. Multiple mutant *C. difficile* lysogen clones were induced and subsequent phage preparations were tested to ensure the effect wasn't due to random error and all showed lack of infection.

This indicated that while the RBP is the determining factor in binding, and in determining host range, changing the RBP even between highly similar phages, is not sufficient to modify host range.

5.2.2 Vector construction, obtaining and testing the mutant phage for Expansion-1.

Alongside the swapping of the RBP genes, an attempt was made to expand the host range by inserting an additional RBP gene into a WT phage, without removing the existing RBP gene. The *PTP* gene from ΦCD2301 was inserted in front of the *TCFP* gene in ΦCD1801, as shown in Figure 5.4.

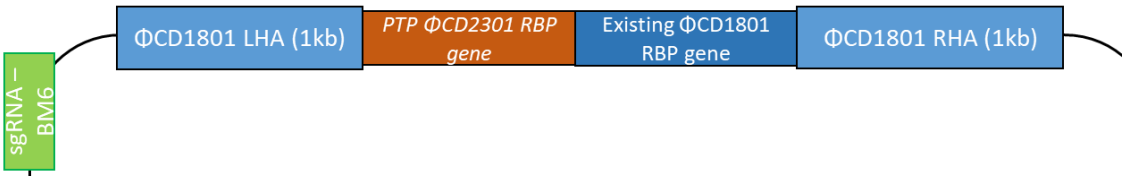


Figure 5.4 Expansion-1 vector map.

The figure shows a schematic representation of the vector used to make the Expansion-1 mutant with HA and two RBP cargo genes in between. The vector is targeting BM6, as previously.

The vector was constructed in a similar fashion to the Swap vectors described above, targeting a BM6 scar in a TCFP Ko mutant in CD1801L generated in Chapter 4 which was also used in Swap 1. The HA insert was constructed by SOEing PCR, including the cargo RBP genes and both HA elements which were first amplified individually to create overlaps using primers and then stitched together in a single reaction. The *PTP* gene was inserted in front of the existing RBP to ensure that it was transcribed and translated correctly as any transcription control elements are found upstream of the gene and would be intact.

The ligation was transformed into competent cells to obtain a correct vector. Following screening, the vector was transformed into *E. coli* sX and conjugated into

the lysogenised *C. difficile* mutant. The resulting colonies were selected, induced and screened (Figure 5.5). Following plasmid loss, and confirmation of the correct insertion of the RBP genes, the modified phage (termed ΦCD1801 Expansion-1 or ΦCD1801E-1 for short) was induced out of the lysogenised bacterium, alongside WT controls of ΦCD1801 and ΦCD2301.

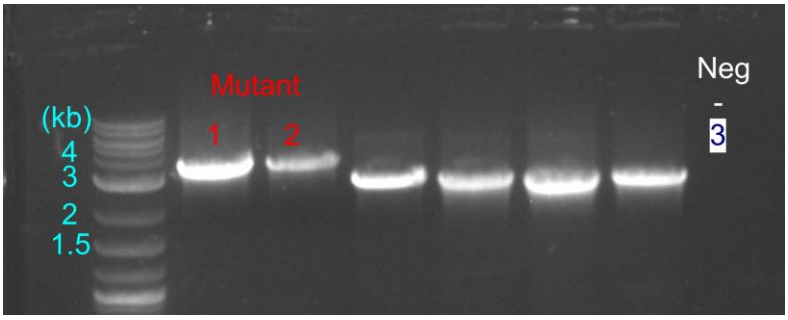


Figure 5.5 Gel image confirming the Expand-1 mutant.
The figure shows a gel image of PCR products resulting from colony DNA amplification to demonstrate the mutants obtained. Lanes 1 and 2 show the E-1 mutant, ~3.8kb in size, with a negative control in lane 3; other bands are unrelated.

All three phages (ΦCD2301, ΦCD1801 and ΦCD1801E-1) were tested against CD1801 and CD2301 (Table 5.3). The controls induced correctly at expected p.f.u./ml values. E-1 mutant phage was able to infect its original host, CD1801, showing the modified particle is able to function as an infective virion. Although the mutant phage seemed to be infecting the original host at a much higher p.f.u./ml than the control WT phage in the first experiment, this is likely due to biological variation that occasionally occurs during induction. Further experiments validated that the phages induce and infect at similar levels. However, ΦCD1801E-1 was unable to infect the new host, CD2301, and the insertion of the *PTP* gene made no difference.

Table 5.3 Infection of CD1801 and CD2301 by E-1 and WT controls.

Phage	CD1801 (p.f.u./ml)	CD2301 (p.f.u./ml)
WT ΦCD1801	$4.2x10^5$	0
WT ΦCD2301	0	$1.2x10^3$
Expansion-1	$1.65x10^8$	0

The above outcome could be for several reasons; the new RBP may not fit in the existing structural domains that hold the RBP in place, the space for the RBP may be occupied by the *TCFP* encoded protein preferentially or the *PTP* gene is not expressed correctly, or its product is not folded in the right configuration making it non-functional. The p.f.u./ml of ΦCD1801E-1 did not seem to be diminished, meaning that all the particles had to contain the *TCFP* product. It's unclear whether the PTP was also inserted into the baseplate, as it tends to contain several tail fibres, typically between 6-8, therefore some insertions could have been possible but were non-functional.

To elucidate whether the fibres were present in the mutant, several binding assays (3.5.6) were carried out with ΦCD1801E-1 and the ΦCD1801 WT control. The mutant was assayed against both hosts. The results (Figure 5.6) show that although ΦCD1801E-1 was able to bind to CD1801, like its WT counterpart, there was no evidence of it binding to CD2301. To check for binding, the ΦCD1801E-1 was incubated with CD2301 as per protocol (3.5.6), however, the resulting supernatant containing free (unbound) phage was then assayed with CD1801, as the mutant phage is able to enumerate on this strain, to determine the p.f.u./ml.

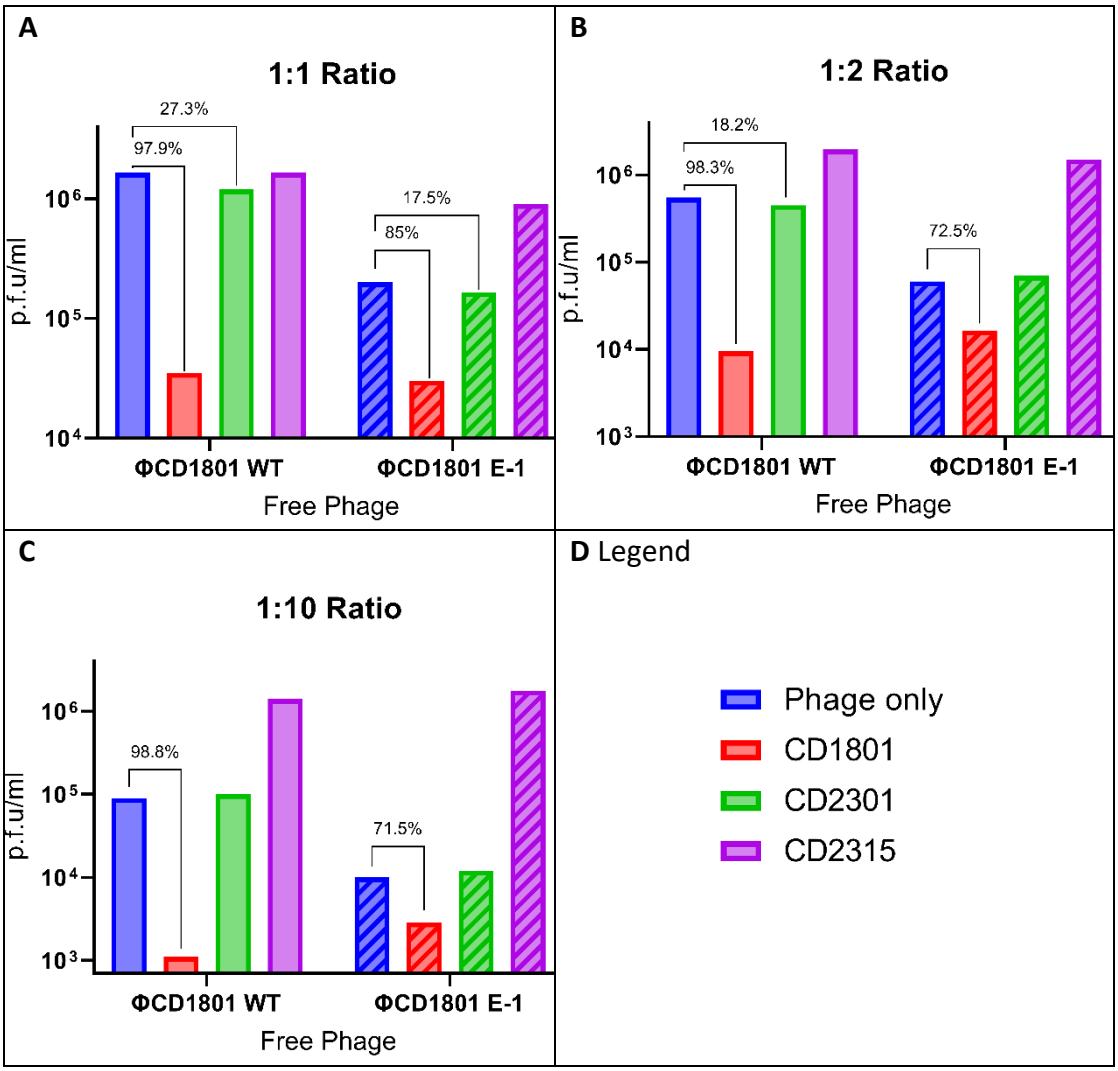


Figure 5.6 Binding assays of ΦCD1801E-1 alongside WT controls with three strains.

The figure shows a binding assay carried out with two phages, the WT and ΦCD1801E-1 mutant, against three cell types: CD1801, CD2301 and CD2315; with three various ratios of phage used. Panels A-C show the ratios used, and each ratio was achieved by varying the amount of phage added while keeping the amount of cells constant. Where there was a decrease in the amount of free phage, i.e., the phages were bound to cells and removed from the supernatant, it is indicated with a percentage decrease when compared to the phage only control. Where not indicated, the amount of free phage was relatively similar or exceeded the number of phages in the phage only control. The data presented here was generated during a single experiment but is representative of an overall observed trend.

At the lowest ratio 1:10 (phage:cells), only binding to the original strain, CD1801 was observed, for both the WT and the ΦCD1801E-1 mutant phage at 98.8% and 71.5%, respectively. As the ratio increased to 1:2, in addition to binding CD1801, the WT ΦCD1801 phage was able to partially bind to CD2301 at 18.2%. At the highest ratio of 1:1, both phages were able to bind both strains, although the ΦCD1801E-1 mutant phage exhibited diminished binding when compared to WT ΦCD1801 phage in all cases. As expected, neither phage was able to bind CD2315. The amount, however, of free phages observed in some of the CD2315 incubated samples seemed to greatly exceed the p.f.u./ml of the phage only control. It is possible that the prophage discovered in Chapter 4 in CD2315 (4.2.6.3.1.) may have spontaneously induced during growth, and skewed the data. In which case, it is difficult to draw insightful conclusions from the CD2315 data, although it is of lesser relevance.

This experiment was conducted to show whether or not the mutant phage particle was able to bind to CD2301, since it was unable to complete a full infection cycle. The mutant phage could have been able to bind to the strain showing the RBP gene is enough to change its binding range, but unable to fully infect for other reasons, – in this case, likely to do with the genetic content of the phage, targeted by host defence systems. If the mutant phage was able to bind, a >50% or more reduction in free phage would be expected when incubated with cells. Typically, binding proficient phage are usually able to achieve a 90-99% reduction in free phage, however, this phage would contain multiple RBPs, potentially leading to less efficient binding overall. The assay showed that the phage was not able to bind, meaning that the particle did not contain the functioning fibre, despite being included in the genome in the best position to be expressed. Despite data showing a 17.5% decrease in free ΦCD1801E-1 mutant phage when incubated with CD2301, it does not indicate binding, as the WT phage showed a 27.3% decrease in phage particles. Therefore, the binding to CD2301 was not improved with the addition of the extra fibre forming the ΦCD1801E-1 mutant phage.

Prior to and following this experiment, for other parts of this study, binding assays were also carried out which showed that the WT ΦCD1801 and to a lesser degree the ΦCD1801E-1 mutant were inconsistently able to bind to CD2301. This finding was inconsistent as often lack of binding to CD2301 was also observed. These observations could be due to experimental variation, as plaquing the phage is not

often the most accurate, yet the only available, method. It may be possible that the phages are minimally binding to CD2301 without proceeding further and would not be surprising as the tail fibres are remarkably similar; discussed later in this chapter.

Surprisingly ΦCD1801E-1 seemed to have diminished overall binding to its original host, CD1801, at 76% on average, compared to a consistent 98% binding rate of the ΦCD1801 WT. This result proved somewhat statistically significant when comparing overall binding rates (paired two-tailed t test, p=0.041*) This may indicate that the presence of the additional fibre interfered with the phage binding mechanism in some way, or reduced expression of the native RBP, effectively making a "worse" phage. The additional fibre was either not expressed, folded incorrectly or did not fit into the baseplate.

5.2.3 Secondary/ irreversible binding

The phage binding mechanism can be broken down into stages. The initial interaction is followed by reversible binding and then finally by irreversible attachment (Quiberoni et al., 2004). The phages randomly collide with the bacteria until they "stumble" upon a receptor (Moldovan et al., 2007), where they bind reversibly (Uchiyama et al., 2011), and then bind permanently to either the same receptor or to a secondary receptor (Dowah & Clokie, 2018). This process involves conformational changes (Bertozzi Silva et al., 2016) which then allow the phages to inject their DNA into the bacteria. A recent study (Thanki et al., 2018) showed that phages infecting *C. difficile* were able to bind reversibly and irreversibly to various hosts, even those they cannot fully infect. Although a high level of irreversible binding was required for an infection.

If the secondary binding was occurring during phage infection, in phages studied here, it would validate further investigation of genes which may be responsible for it as they would likely play a role in the host range. In this study the binding of the three phages ΦCD1801, ΦCD2301 and ΦCD2315 to their hosts was investigated to see whether they were able to bind to their hosts in a reversible/irreversible manner. The protocol based on the above study was a modified version of the binding assay (3.5.6), where the samples were also vortexed to check for the irreversible binding.

However, the results did not indicate any irreversible binding in these phages as the phages either bound to a high degree (>90%) or not at all, showing a lot more binary

results than those observed by Thanki et al., (2018) in their study. Several attempts were made with various ratios, however, none proved fruitful. The phages could simply not exhibit this phenomenon, or the assay protocol used was not sensitive enough to capture this interaction. Consequently, data is not shown. Regardless, accessory proteins were assumed to play a role, although it was not clear whether there was a possibility of secondary binding; the course of the study remained unchanged.

5.2.4 Accessory proteins

As the initial idea of swapping the RBPs did not prove to be successful for changing host range, the possibility that accessory proteins or neighbouring genes may be involved was considered. Inserting additional fibres (like in the ΦCD1801E-1 mutant phage) also did not prove fruitful, indicating more than one gene was likely necessary to alter the host range. This led to a mass deletion and exchange of RBP genes either singly, or in-tandem with upstream or downstream genes, between the two phages ΦCD1801 and ΦCD2301. These two phages were selected as their lysogenised hosts are easy to conjugate into, particularly CD2301L, meaning mutants can be obtained with relative ease. They are also both *Myoviridae* and are therefore more likely to carry compatible elements. Previous criteria also applies where the phages do not infect each other's hosts and have homologous genes in the tail region.

Phage ΦCD2315 was not considered, as being a Siphovirus, it is more distantly related to the ΦCD1801 and ΦCD2301 phages. When analysed, the genome architecture of ΦCD2315 also showed a decreased homology in the tail gene module, indicating possible structural incompatibilities with the *Myoviridae*. Additionally, as shown in Chapter 4, CD2315 also contains a prophage which is functional and induces out which creates a complicated relationship with. These factors make it a less than ideal candidate and since ΦCD1801 and ΦCD2301 are closely related, but infect different hosts made them the obvious choice for the proposed experiments.

5.2.4.1 Bioinformatic analysis of surrounding genes

To determine which genes were going to be deleted in each phage and then respectively exchanged between the two phages, bioinformatic analysis of genes neighbouring the RBP was carried out using several databases. Each gene was analysed for potential function, from the genes upstream the RBP up to and

including the baseplate (*BaseJ*), and downstream the RBP up to the *holin* gene, including the upstream gene encoding a hypothetical protein (Figure 5.7).

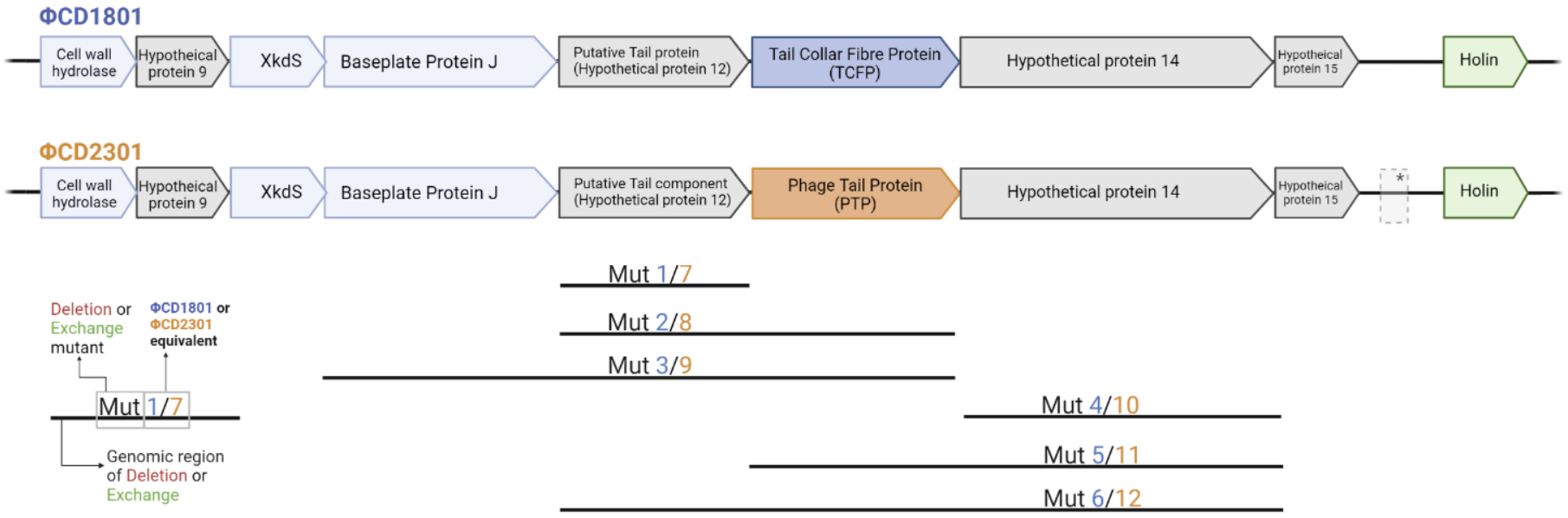

ΦCD1801
Cell wall hydrolase
Hypotheical protein 9
XkdS
Baseplate Protein J
Putative Tail protein (Hypothetical protein 12)
Tail Collar Fibre Protein (TCFP)
Hypothetical protein 14
Hypotheical protein 15
Holin
ΦCD2301
Cell wall hydrolase
Hypotheical protein 9
XkdS
Baseplate Protein J
Putative Tail component (Hypothetical protein 12)
Phage Tail Protein (PTP)
Hypothetical protein 14
Hypotheical protein 15
*
Holin
Deletion or Exchange mutant
ΦCD1801 or ΦCD2301 equivalent
Mut 1/7
Genomic region of Deletion or Exchange
Mut 1/7
Mut 2/8
Mut 3/9
Mut 4/10
Mut 5/11
Mut 6/12

Figure 5.7 Schematic representation of the tail gene module in ΦCD1801 and ΦCD2301.

The figure shows a representation of the tail architecture of both phages, analysed in preparation for the deletions and exchanges. The ΦCD1801 genome, RBP and corresponding mutations are colour-code in blue; the ΦCD2301 is colour-coded in orange. Mutant (line indicator) represents the region that was deleted in the Del 1-12 mutants and exchanged in the Ex1-12 mutants. Mutants 1-6 were made in ΦCD1801, and the corresponding matching mutants 7-12 were made in ΦCD2301. The *holin* gene (green) indicates the border of the structural tail gene module.

There were two likely structural scenarios, outside of genomic contents of the phage, as for why genes other than the RBP were required. The RBP may not fit into the tail structure if it originates from a different phage, therefore, the genes upstream, including *Hypothetical protein 12* (*Hyp12*) and *BaseplateJ* were included in mutant phage generation as they are likely to anchor the RBP gene. In the other scenario, there were accessory genes required to help the RBP function, whether it is in folding, or binding of the receptor. The chaperone genes were likely to be downstream the RBP, hence those were also included, like *Hypothetical protein 14* (*Hyp14*).

Enzymatic activity could have been a third reason as phages often encode such activities as an aid to entering the bacterial cell. However, both phages were found to contain a *cell wall hydrolase* gene, likely involved in this process. As the genes were labelled as the same, with a 96% positive identity, they were deemed to have the same function, thus, swapping them would not likely be beneficial, nor specific to the host range. Further downstream a *hypothetical protein* gene was found, without any annotated function in any database; it was excluded due to high similarity (94%) and relatively distal location to the RBP. The next gene downstream was labelled as an XkdS encoding gene, containing a DUF2634; it seemed to have a connection to gp25 in *E. coli* T4 phage which is responsible for forming a part of the baseplate wedge (Yap et al., 2010) where the tail fibres attach. Gp25 is also involved in a protein complex which causes a sheath contraction, following a conformational change in Long Tail Fibres, allowing injection of DNA (Arisaka et al., 2016). Despite the lack of clear indication as to function, the genes were highly similar at 98%

between the two phages, thus it was determined that an exchange would not likely be beneficial.

Next, the *BaseplateJ* gene, encoding the baseplate was found. Its function as the baseplate is well established where it allows the RBPs to fit into the rest of the phage, connecting it to the tail sheath and tube. As discussed earlier in the chapter, in Figure 5.6, the extra RBP included in the ΦCD1801E-1 mutant did not seem to function, as it may not have been integrated, possibly due to incompatibility with the baseplate. Despite high similarity (95% positive), it was chosen to be included in the mutant generation as any structural irregularities, may prevent the RBPs from fitting in the phage.

The *Hyp12* gene, sandwiched between the baseplate and RBP gene, had no clear indication of function, beyond being a tail component. It had similarity to portal proteins which likely allow the DNA to pass through the tail and be injected, as well as to some tail fibres in distantly related phage. However, the assessment of this gene as a tail fibre was likely incorrect as it was based on an annotation from the first *C. difficile* phage genome sequenced, ΦCD119, where ORF28 was labelled as such (Govind et al., 2006).This was based on similarity to *Chromobacterium violaceum.* An in-depth analysis, however, showed the genome to contain 4 incomplete prophages, where one of the genes encoded, a 600aa product (which had homology to the 132aa ORF28), indicating an unlikely similar function. When the genome was processed by a gene calling pipeline, multiPhATE2 (Zhou et al., 2019; 2021), which uses multiple databases for homology, the *Hyp12* gene was shown to have similarity to the Short Tail Fibres found in phage T4 (Sarah Kerr, University of Nottingham, collaboration). On the other hand, phammer, a domain prediction tool, indicated this gene to be similar to *gp48* in Mu phage. *Gp48* plays a role in formation of the baseplate wedge, making it structurally important. Due to its close proximity and related tail function, it was chosen to be included in the mutations.

Based on the experimental evidence, as well as prior bioinformatic evidence, the *TCFP* and *PTP* genes, located downstream from *Hyp12*, were confirmed to be the RBP and the tail fibre (Chapter 4). Further characterisation was also carried out to show the universal domain, DUF3751 found in this gene. The domain is found at the N-terminus of the protein and is universally known to indicate tail fibres, and simultaneously the RBP (Emre Gencay, SNIPR Biome LLC, personal communication),

even though it is listed as a “domain of unknown function” in literature. The DUF3751 domain is conserved in many phages with this structure, and is approximately 152aa in length in *C. difficile* phages comprising over half of the length of the entire protein (approx. 250-275aa). The domain is a part of the tail fibre that “docks” to the baseplate, whereas the C-terminal domain determines the binding specificity and exhibits a greater degree of divergence between even closely related *C. difficile* phages (Figure 5.8).



Figure 5.8 Amino acid sequence of the RBP proteins highlighting the C-terminal domain differences.

The aa sequence of RBP genes of ΦCD1801 (top sequence) and ΦCD2301 (bottom sequence). The sequences have a large consensus at the N-terminal end of the protein; divergence emerges, particularly towards the C-terminal end, which is responsible for binding. There is a large discrepancy, compared to the rest of the sequence at the C-terminal the end of the protein, likely to be partially responsible for the different host ranges.

The proteins are highly homologous, up until the 200aa mark, even if there are single amino acid discrepancies, they are usually equivalents, whereby they are similar enough to be considered a “positive” and unlikely to have a major impact on function. Towards the C-terminal end, in the last 60aa is where the dissimilarity occurs and is likely correlated with different host ranges. The proteins share a 92% positive identity, and although swapping them alone is not enough to change host range, they are a crucial component of the binding interaction, hence they were included in the mutant generation, in various configurations with the other components.

Following the identification of the domains and homology, vectors were designed to generate PartialSwap-1 and PartialSwap-2 mutants, effectively swapping the C-terminal end of the protein between ΦCD1801 and ΦCD2301, and vice versa. Unfortunately, the constructs proved difficult to clone and so the project was re-directed to focus on the exchange mutants, especially since they had significant potential to elucidate the interactions of the entire tail module.

Downstream of the RBP, another hypothetical protein gene, *Hyp14*, was found; 1.5-1.7kb in size and without any obvious function indications. This is the first gene in the tail module where the proteins differ in size and do not share a high degree of similarity. The Hyp14 protein in ΦCD1801 is made up of 496 residues, compared to 564 in ΦCD2301; the largest tail-related protein outside of the tail tape measure. The two only share a 45% similarity, compared to the usual ~90% homology observed for the proteins discussed above. However, it was included as the difference between the two could be related to the host specificity, and accessory genes are commonly found downstream of the actual RBP, therefore the protein may be required in the binding process.

Due to the low homology between the two, the Hyp14 proteins from ΦCD1801 and ΦCD2301 were also aligned against two further Hyp14 protein sequences (Appendix 8.1) sourced from other in-house *Myoviridae* phages isolated by SBRC researchers, ΦCD9301 and ΦCD08011 using the Clustal Omega (v1.2.4). The alignment showed that in the first part, the sequence lacked homology and had large gaps where the longer sequences contained additional residues. The middle part showed a consensus with many identical residues, suggesting a conserved domain. The third and final part showed a region with some consensus, but less conserved residues than in the middle. The cladogram showed that ΦCD1801 and ΦCD08011 were the most similar to each other, followed by ΦCD9301. The Hyp14 protein of ΦCD2301 was the most distantly related, also explaining why initially there was a lack of homology observed between it and the equivalent protein originating from ΦCD1801. This suggested that the lack of homology in certain areas may be related to the divergent host range between the two phages. Additionally, the multiPhATE2 pipeline found similarities to tail fibre parts of *E. coli* and *Salmonella* spp. phages making its involvement more probable.

The following gene, also a hypothetical protein, *Hyp15*, was excluded from the mutant generation due to lack of any relation to tail genes and its small size (less than 100aa). It had some homology to capsid scaffold proteins which are not involved in binding process, therefore should be excluded from determining the host range specificity. Further downstream genes were also excluded as they were determined to be involved in lysis/lysogeny signalling the end of the tail gene module.

5.2.4.2 Determination of configuration for selected mutants

Based on the bioinformatic evidence presented above and the experimental proof demonstrated in Chapter 4, it was clear that the *TCFP*/*PTP* encode the tail fibre responsible for binding. However, as it is not enough to just swap the fibres, if there is no other interfering factors, the neighbouring accessory proteins were likely at play. One or more of them is likely to be involved as an accessory gene, required for binding. Therefore, they were deleted, and then exchanged on their own or in-tandem, to elucidate their exact involvement and requirements for changing the host range.

Other factors may interfere with a full infection cycle, such as the interaction of the phage genomic content coming in contact with anti-phage systems, for example CRISPR defences. Another factor may be the prophage content of bacterial hosts, however that is largely strain-dependent. Currently these factors cannot be controlled for or easily determined, therefore the project focused on elucidating the function of the neighbouring tail proteins as they are likely to be involved to some degree, and can be manipulated through genetic engineering. In order to successfully engineer a semi-synthetic phage with any potential to be used as a therapy, the basic functions of neighbouring proteins should be known, especially how their removal or modification will affect the infection cycle. Basic characterisation is needed regardless of whether the desired host range change is achieved.

5.2.4.2.1 Single gene deletions and exchanges

Mutants were named based on the type of Mutation, deletion (Del) or exchange (Ex); and the corresponding combination number. For example, a mutant with a deleted gene combination number 2 would be named Del2. Mutants 1-6 were made in the ΦCD1801 phage, and the reminder 7-12 were made in the ΦCD2301, with their corresponding components swapped in the exchange mutants. For example, Ex2

would be a ΦCD1801 phage, containing the *Hyp12* and *PTP* genes from ΦCD2301 in place of the original genes (Figure 5.7).

Mutants 1/7 and 4/10 were deletions, followed by exchange of single genes *Hyp12* and *Hyp14* respectively. The genes were predicted to play a part in the binding process and their individual role had to be elucidated based on the impact of said mutations. This way, their involvement can be separated from the other mutations and it can be easily established which gene is responsible for which effect. For example, in Del2 the *Hyp12* and *TCFP* genes are deleted; the effect of the deletion of *TCFP* has been studied previously (Chapter 4), and since it leads to a complete lack of binding, any effect *Hyp12* may be having would be masked. Therefore, individual deletions were necessary, for a more detailed and well-informed characterisation of the region.

5.2.4.2.2 Combined gene deletions and exchanges

The remaining mutants were deletions and exchanges of multiple genes. Mutants 2/8 were generated to show whether the *Hyp12* gene was assisting the RBP gene and/or was required to allow the correct integration of the RBP structurally. Mutants 3/9 were generated to ensure correct docking of the RBP fibre to the baseplate, as this interaction may be specific, and lack of structural compatibility could result in lack of fibre integration. Mutants 5/11 were obtained to show whether the downstream gene, *Hyp14*, was involved as an accessory to the RBP as it is often common in some phage; and whether it was specific to each RBP to function. Mutants 6/12 were deemed important as both *Hyp12* and *Hyp14* could be involved as accessories and both support the function of the RBP. If that was the case, the *Hyp12* gene would likely have structural involvement, whereas *Hyp14* would act as a chaperone or binding accessory.

Prior deletion was required to ensure that whatever effect was achieved by exchanging the genes, was due to that gene (combination) specifically, rather than any knock-on effects. Additionally, to generate the exchange mutants correctly, the deletions had to take place, as the region had too much homology to achieve the specific homologous recombination required to exchange the genes in the exact fashion. All planned mutants are summarised in Table 5.4.

Table 5.4 Summary of planned deletion mutants and subsequent exchange

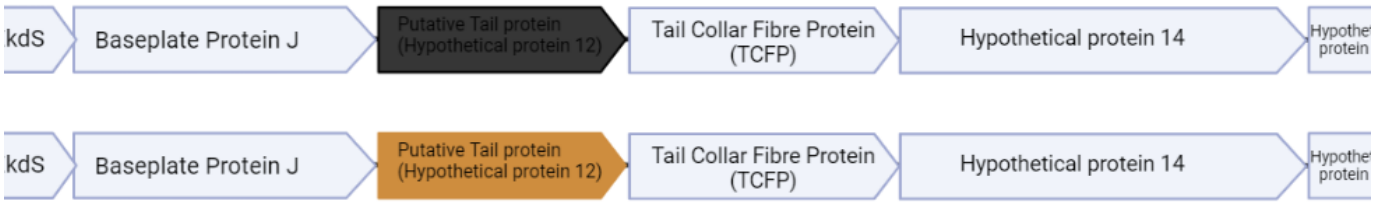
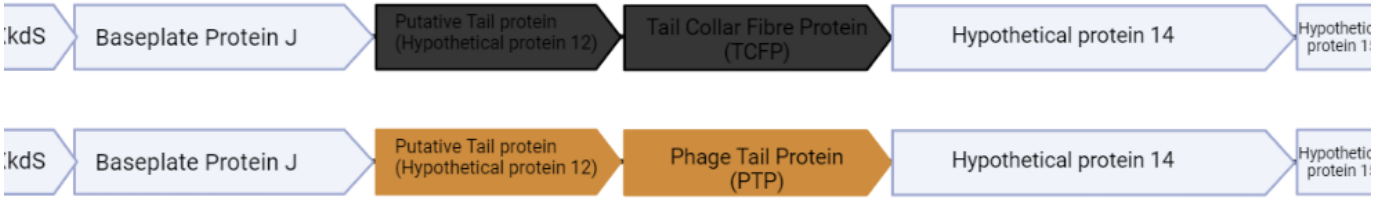
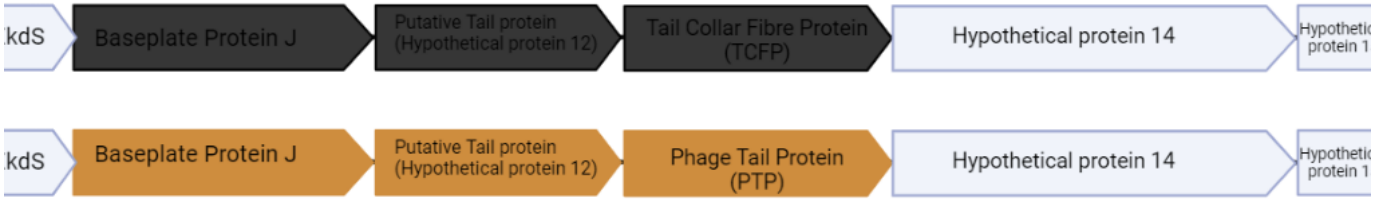
#	Deletion (in WT Lysogen)	Exchange (In Deletion mutant)	Top – Deletion mutant Bottom – Exchange mutant
1	Deletion of Hyp12 in ΦCD1801	Insertion of Hyp12 from ΦCD2301, in ΦCD1801	
2	Deletion of Hyp12 + TCFP in ΦCD1801	Insertion of Hyp12 + PTP from ΦCD2301, in ΦCD1801	
3	Deletion of BaseJ + Hyp12 + TCFP in ΦCD1801	Insertion of BaseJ + Hyp12 + PTP from ΦCD2301 in, ΦCD1801	

Table 5.4 Summary of planned deletion mutants and subsequent exchange

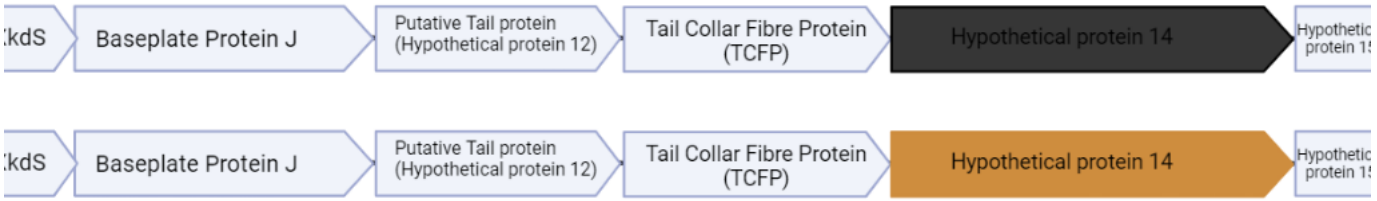
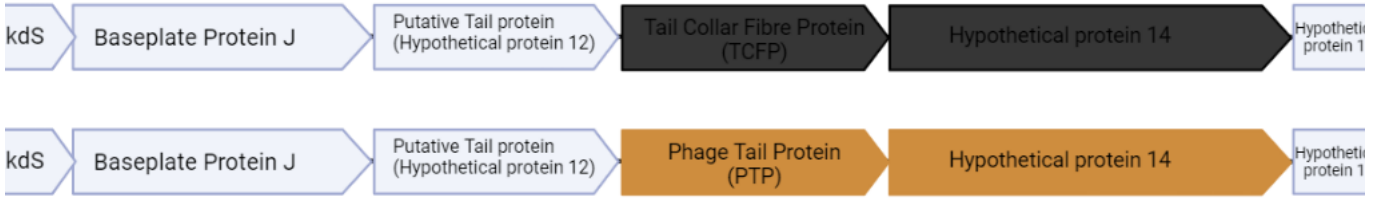
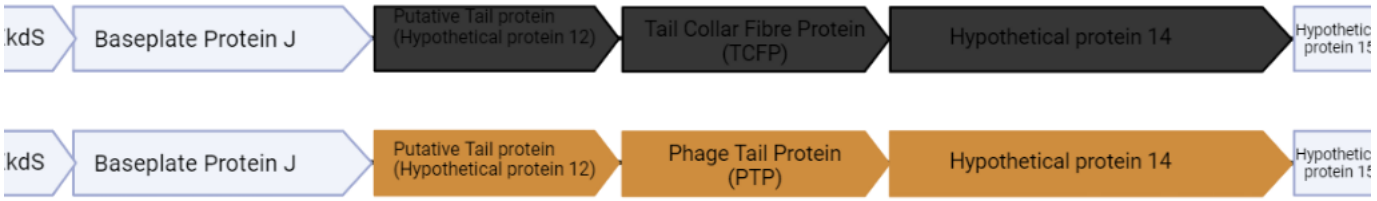
#	Deletion (in WT Lysogen)	Exchange (In Deletion mutant)	Top – Deletion mutant Bottom – Exchange mutant
4	Deletion of Hyp14 in ΦCD1801	Insertion of Hyp14 from ΦCD2301, in ΦCD1801	
5	Deletion of TCFP + Hyp14 in ΦCD1801	Insertion of PTP + Hyp14 from ΦCD2301, in ΦCD1801	
6	Deletion of Hyp12 + TCFP + Hyp14 in ΦCD1801	Insertion of Hyp12 + PTP + Hyp14 from ΦCD2301, in ΦCD1801	

Table 5.4 Summary of planned deletion mutants and subsequent exchange

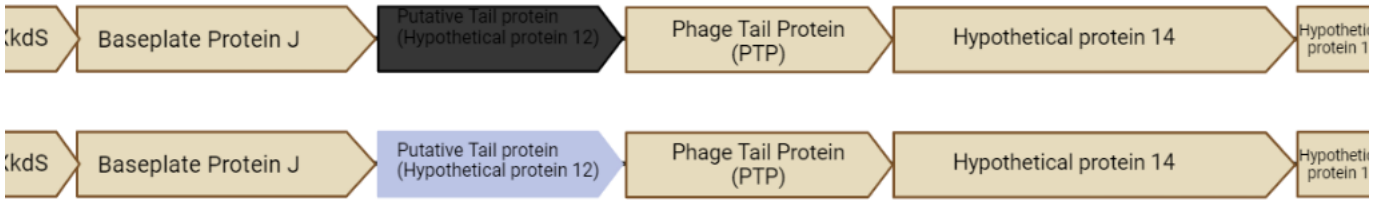
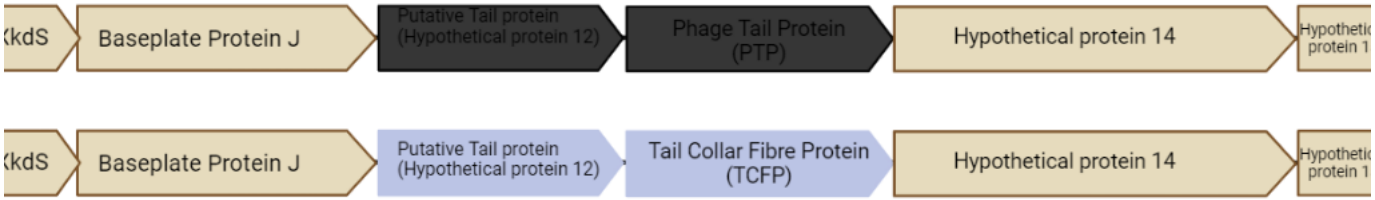
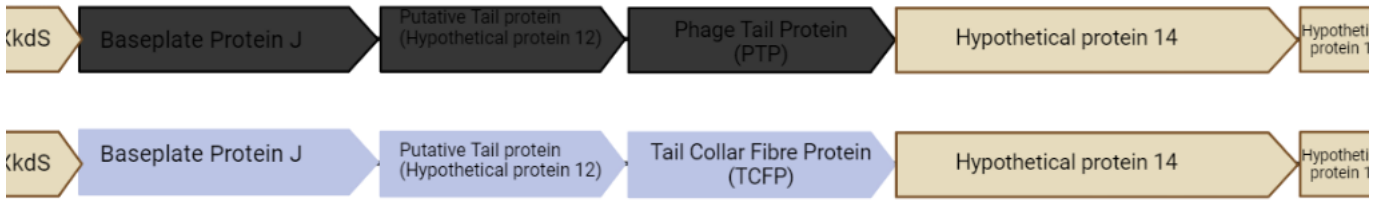
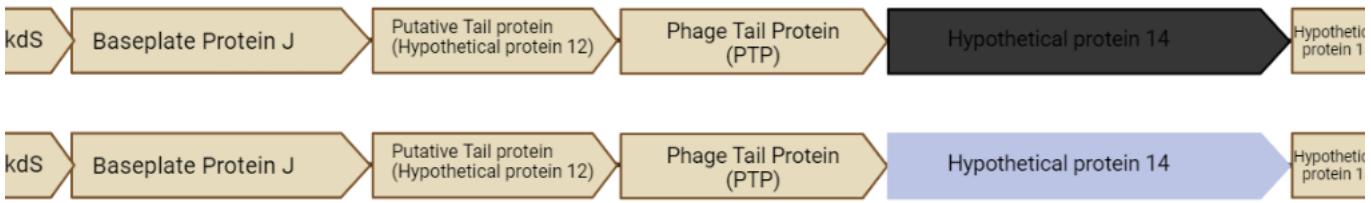
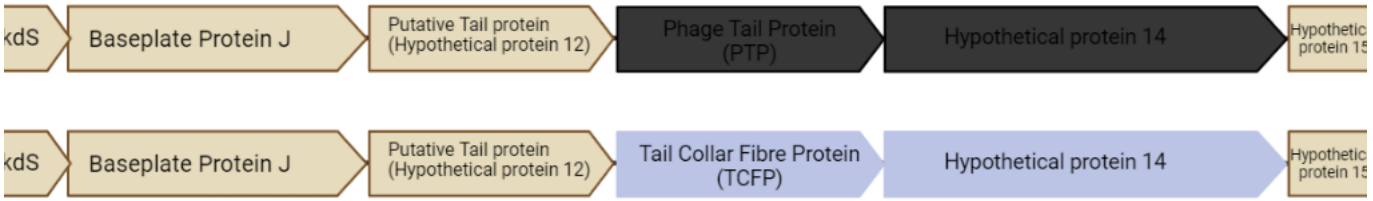
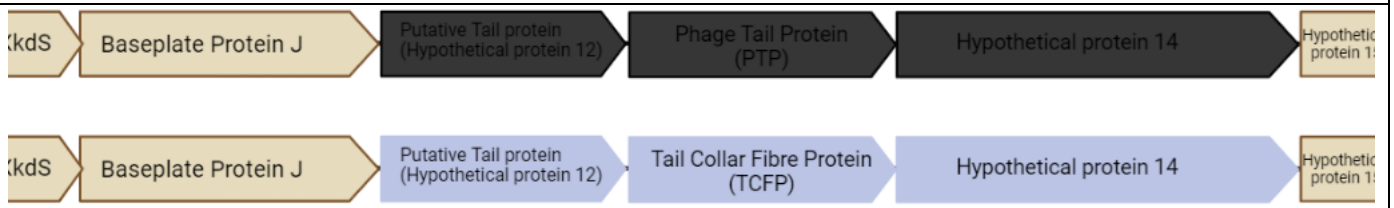
#	Deletion (in WT Lysogen)	Exchange (In Deletion mutant)	Top – Deletion mutant Bottom – Exchange mutant
7	Deletion of Hyp12 in ΦCD2301	Insertion of Hyp12 from ΦCD1801, in ΦCD2301	
8	Deletion of Hyp12 + PTP in ΦCD2301	Insertion of Hyp12 + TCFP from ΦCD1801, in ΦCD2301	
9	Deletion of BaseJ + Hyp12 + PTP in ΦCD2301	Insertion of BaseJ + Hyp12 + TCFP from ΦCD1801, in ΦCD2301	

Table 5.4 Summary of planned deletion mutants and subsequent exchange

#	Deletion (in WT Lysogen)	Exchange (In Deletion mutant)	Top – Deletion mutant Bottom – Exchange mutant
10	Deletion of Hyp14 in ΦCD2301	Insertion of Hyp14 from ΦCD1801, in ΦCD2301	
11	Deletion of PTP + Hyp14 in ΦCD2301	Insertion of TCFP + Hyp14 from ΦCD1801, in ΦCD2301	
12	Deletion of Hyp12 + PTP + Hyp14 in ΦCD2301	Insertion of Hyp12 + TCFP + Hyp14 from ΦCD1801, in ΦCD2301	

5.2.4.3 Vector construction and obtaining mutants

The vectors were constructed as described previously. Initially, after establishing targets, the combinations of genes would be deleted and exchanged. A total of 12 deletion and 12 exchange vectors were to be constructed. For vectors 1 and 4, and the corresponding 7 and 10, sgRNAs had to be designed as the targets were single genes, not previously investigated. Three sgRNAs were designed per target, therefore each new target had three vectors constructed rather than just one. For all the other vectors, as they included the previously targeted RBP gene for each phage, those sgRNAs were used, and a single vector was constructed for each as they have shown to work well before.

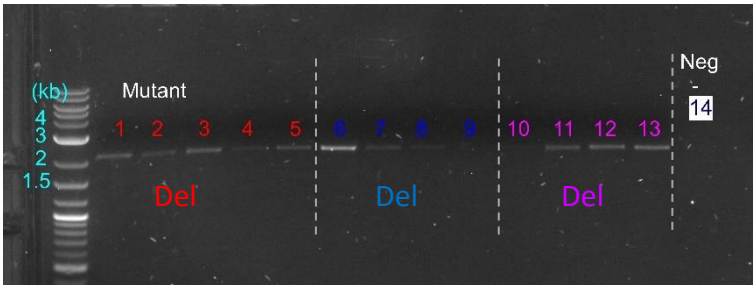
The vectors were constructed either through three way or two way ligations, depending on whether a new sgRNA was required. The HAs were constructed based on the same principle as in the Expand-1 vector. The vectors were constructed in NEB 10-Beta for the larger vectors or NEB 10-Stable. The NEB 10-Beta cells are specifically designed for large vectors, and with large HA inserts of up to 6kb, their deployment was necessary, as this is beyond the typical product size used in CRISPR/Cas9 vectors in-house. The use of more specialised cells, over previously used Top10 improved efficiency, especially needed in such a large-scale mutant construction project. Some vectors required multiple rounds of cloning and screening as regions proved particularly prone to yielding SNPs. Ultimately, all the vectors were generated to contain the expected nucleotide sequences.

Deletion and exchange vectors 10-12 had significant cloning issues, specific to the RHA. The SNPs were highly similar in each event. Further investigation led to re-sequencing of the phage (Sarah Kerr, PhD thesis), showing incorrect initial assembly, where previously there were hypothetical protein genes present between *Hyp15* and the *holin* genes, indicated by the dash-bordered grey box in Figure 5.7. To prevent problems, the HA was shortened to ~300bp (compared to the usual 1kb). Longer HAs typically allow for better conjugation efficiency, however, as CD2301L (target strain) is a very efficient conjugator, no problems were foreseen. The cloning then proceeded without problems.

The verified vectors were transformed into *E. coli* sX and conjugated (3.4.1.1) into appropriate strains. Initially, the deletion mutants had to be made first, before any exchanges could take place, due to high homology (>90%) between the loci. After the

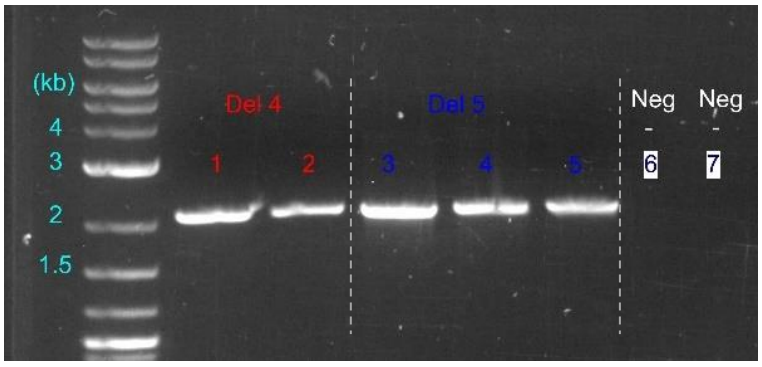
deletion mutants were obtained (Figure 5.8A-E), verified and went through plasmid loss (3.4.1.2), they were stored for future testing and further conjugations. The same protocol was followed for generating exchange mutants, once the vectors were obtained and conjugated. The conjugations took place into the corresponding deletion mutants, allowing to exchange the appropriate gene cassettes between the two phages, ΦCD1801 and ΦCD2301, whilst lysogenised within their host cells. Again, once verified and confirmed through sequencing after plasmid loss, the mutants (Figure 5.8F-L) were stored for further testing.

A CD1801L Del1-3



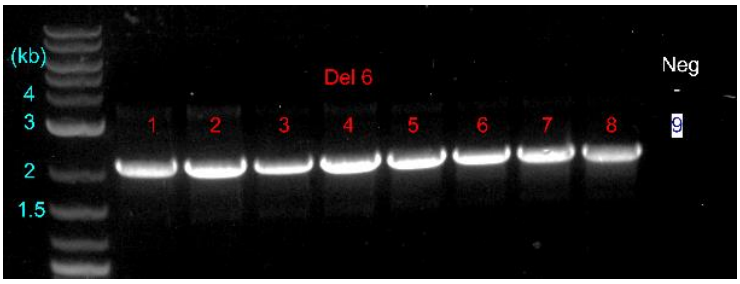
The gel image shows deletion mutants Del1-3 generated in CD1801L colonies, all ~2.2kb as expected. Del1 mutants are in lanes 1-5 (red), non-mutants would be ~2.8kb. Del2 mutants are in lanes 6-9 (blue), non-mutants would be ~3.6kb. Del3 mutants are in lanes 10-13 (purple), non-mutants would be ~4.6kb. Negative control is in lane 14.

B CD1801L Del4-5



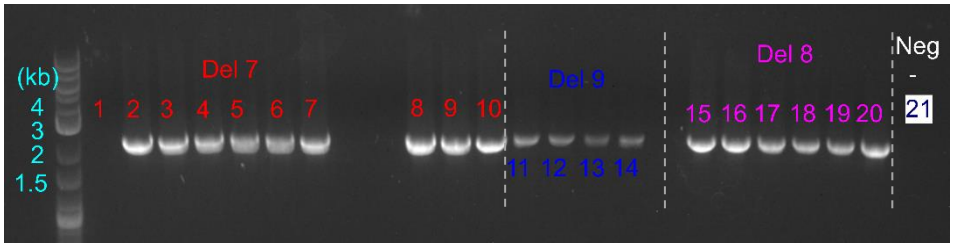
The gel image shows deletion mutants Del4-5 generated in CD1801L colonies, all ~2.2kb as expected. Del4 mutants are in lanes 1-2 (red), non-mutants would be ~3.6kb. Del5 mutants are in lanes 3-5 (blue), non-mutants would be ~4.5kb. Negative controls are in lanes 6 and 7.

C CD1801L Del6



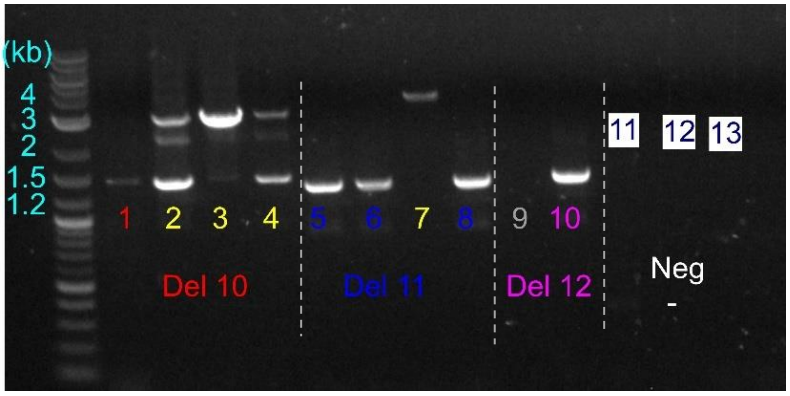
The gel image shows deletion mutants Del6 generated in CD1801L colonies, all ~2.2kb as expected. Del6 mutants are in lanes 1-8 (red), non-mutants would be ~5kb. Negative control is in lane 9.

D CD2301L Del7-9



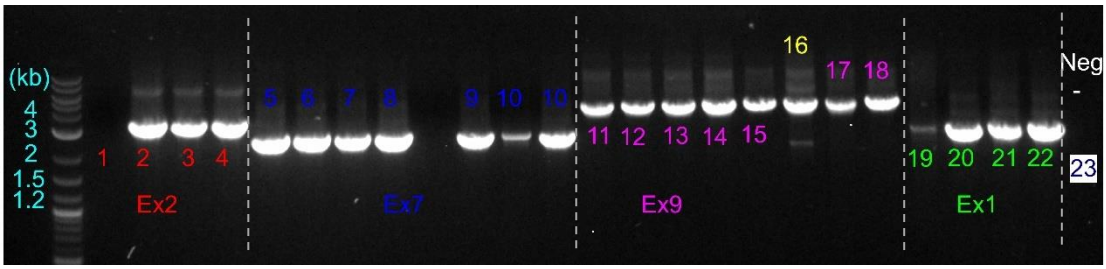
The gel image shows deletion mutants Del7-9 generated in CD2301L colonies, all ~2.2kb as expected. Del7 mutants are in lanes 1-10 (red), non-mutants would be ~2.7kb. Del8 mutants are in lanes 15-20 (purple), non-mutants would be ~3.6kb. Del9 mutants are in lanes 11-14 (blue), non-mutants would be ~4.6kb. Negative control is in lane 21.

E CD2301L Del10-12



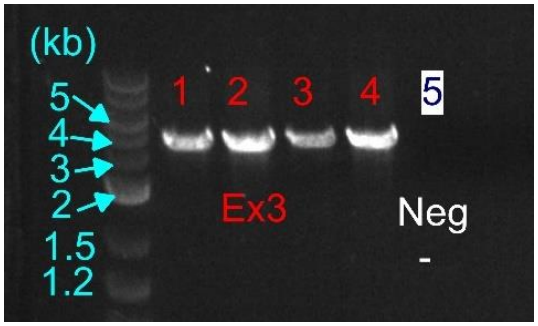
The gel image shows deletion mutants Del10-12 generated in CD2301L colonies, all mutants ~2.2kb as expected. Lanes 2-4 and 7 (yellow) contained the wrong size bands, hence the colonies were discarded as they were non-mutants or mixed. Del10 mutants are in lane 1 (red), non-mutants would be ~3.2kb. Del11 mutants are in lanes 5-6,8 (blue), non-mutants would be ~3.9kb (7). Del12 mutants is in lane 10 (purple), non-mutants would be ~4.5kb. Negative controls are in lanes 11-13.

F CD1801L and CD2301L Ex1-2, Ex7, Ex9



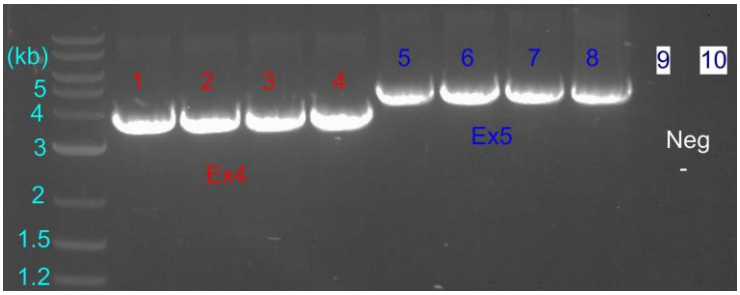
The gel image shows exchange mutants generated in CD1801L/CD2301L colonies, all mutants were various sizes as expected. Lane 16 (yellow) had the two bands, hence the colonies were discarded as they were a mixed population of mutants and non-mutants. Non-mutants would be ~2.2kb in size. Ex1 mutants are in lanes 19-22 (green) and ~2.8kb. Ex2 mutants are in lanes 1-4 (red) and ~3.6kb. Ex7 mutants are in lanes 5-10 (blue) and ~2.8kb. Ex9 mutants are in lanes 11-18 (purple) and ~4.6kb. Negative control is in lane 23.

G CD1801L Ex3



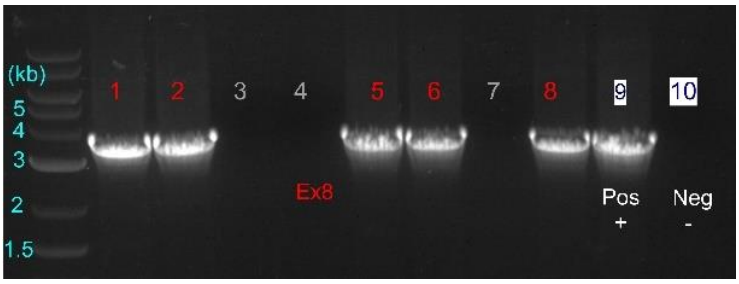
The gel image shows Ex3 mutants in lanes 1-4 (red), at ~4.6kb. Negative control is in lane 5.

H CD1801L Ex4/5



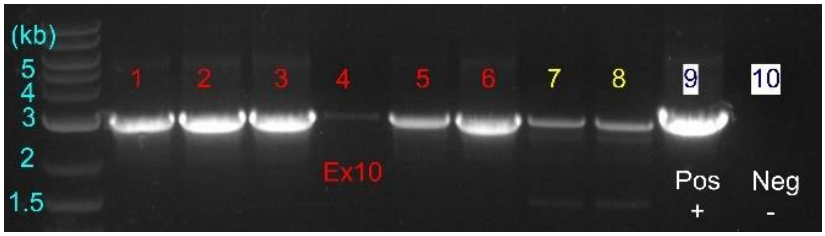
The gel image shows Ex4 mutants in lanes 1-4 (red), at ~3.6kb. Ex5 mutants are in lanes 5-8 (blue), at ~4.5kb. Negative controls are in lanes 9-10.

I CD2301L Ex8



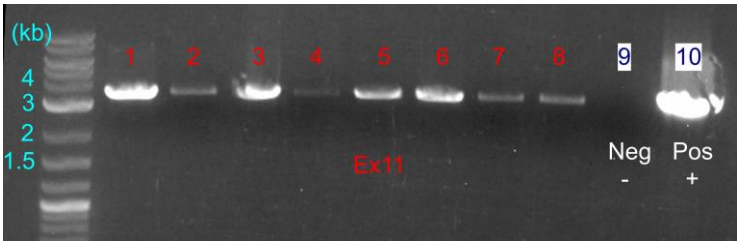
The gel image shows Ex8 mutants in lanes 1-2, 5-6 and 8 (red), at ~3.6kb. Negative control is in lane 10 along with a positive control in lane 9 (CD2301L gDNA).

J CD2301L Ex10



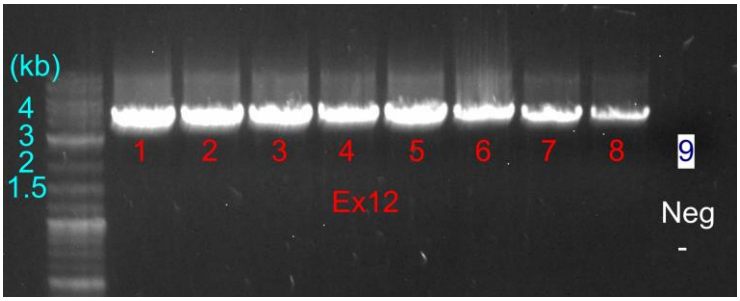
The gel image shows Ex10 mutants in lanes 1-6 (red), at ~3.2kb. Lanes 7 and 8 (yellow) had the two bands, hence the colonies were discarded as they were a mixed population of mutants and non-mutants. Negative control is in lane 10 along with a positive control in lane 9 (CD2301L gDNA).

K CD2301L Ex11



The gel image shows Ex11 mutants in lanes 1-8 (red), at ~3.9kb. Negative control is in lane 9 along with a positive control in lane 10 (CD2301L gDNA).

L CD2301L Ex12



The gel image shows Ex12 mutants in lanes 1-8 (red), at ~4.5kb. Negative control is in lane 9.

Figure 5.8 Gel images of PCR'd mutant colonies for mutants 1-12

The figure shows gel images of PCR products resulting from colony DNA amplification to demonstrate the mutants obtained. Each gel has its own mini-legend. The gels all have the Quick-Load® 1 kb Plus DNA Ladder (NEB) as a size indicator and were carried out as described in 3.6.5.

The vector for Ex6 was obtained with much difficulty, due to problems with SNPs in the LHA, which were overcome by reducing the size to ~400bp. The vector itself contained the largest cargo, at 5kb, compared to the usual 2-3kb. This caused further problems with construction, however, a vector was obtained using the NEB 10-Beta cells. The vector was unable to be conjugated into CD1801L cells, likely due to its large size combined with inefficient conjugation in CD1801L cells. The conjugation was attempted numerous times with different copies of the Ex6 vector, however it proved unsuccessful.

5.2.4.4 Testing Mutants

Once the majority of mutants were ready, they were tested in two batches; due to the number of mutants, it was not possible to effectively process them all at once. All mutants were induced in triplicate, in three separate events per mutant, alongside WT controls to ensure the induction was taking place properly. The Ex11 and Ex12 mutants had an additional biological replicate induced in triplicate each, as they showed a positive result; this was to ensure this was not due to contamination or a random event.

5.2.4.4.1 Deletion mutants

Initially, each deletion mutant phage was tested against two strains, CD1801 and CD2301 (Figure 5.10); that way effects on binding, whether to the original or new strain could be seen. Controls were included in each replicate batch to ensure no contamination and to use as a comparison. This was especially important as phages can induce at different levels each time, therefore a comparison for each replicate event was important. Typically, ΦCD1801 induces at 10^5-10^7, and ΦCD2301 at 10^2-10^4 p.f.u./ml. The detection limit for all plaque assays at absolute minimum is at 5 plaques/ml being equivalent to a single plaque on an undiluted plate (5 p.f.u./ml). However, a single plaque is typically not enough confirmation of infection, therefore only plates containing between 2-100 plaques are counted (per 3.5.1), making the safe minimum detection level $1x10^1$, or 10 p.f.u./ml. This is important to consider

when looking at mutant phages – particularly those with mutations in the tail region, which may have very impaired binding and infection rates and may be below detection limits.

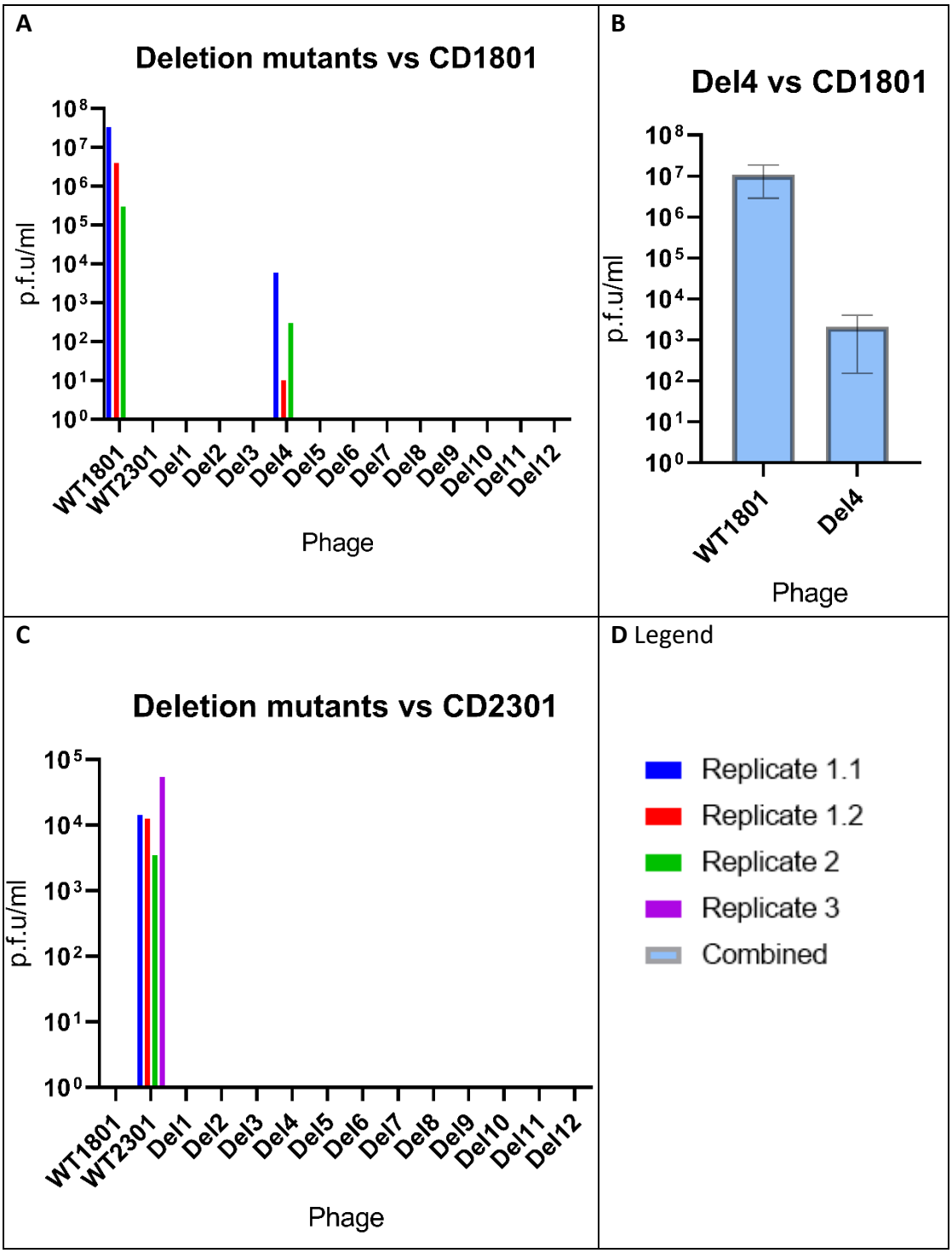


Figure 5.10 Plaque assay showing the Del1-12 mutants against CD1801 and CD2301.

The figure shows the deletion mutants assayed against CD1801 (A) and CD2301 (C), with WT controls. The assay was done in triplicate for each mutant. Panel B shows the combined data for Del4 against CD1801 as it was the only positive result, and the first clue to function. The error bars indicate a standard error of mean (SEM).

When the 12 deletion mutants were incubated against CD1801, only Del4 was able to produce plaques. The WT control induced at $1.07x10^7$ p.f.u./ml on average, compared to the Del4 mutant which was able to infect its original strain at $2.1x10^3$ p.f.u./ml on average (Figure 5.10B). This showed a 99.98% reduction in the phage's ability to bind, which is 5087 times less efficient than the WT. No other deletion mutants were able to plaque on this strain; this was not unexpected as Del2/3/5/6 all included the RBP gene deletion. Previous data has demonstrated that when this gene was not present the phage was unable to complete infection. Del1 mutant was also unable to plaque, suggesting that the *Hyp12* gene is a key tail component necessary for infection. Since Del4 mutants were able to plaque, this gene is not necessary for the phage to function. However, since its deletion meant a severe reduction in the number of plaques, in order for the phage to function efficiently it needs to be present. Del7-12 mutants were unable to infect CD1801 which was not unexpected, as these were ΦCD2301 deletion mutants, and it would be extremely unlikely for them to plaque on a new host without the addition of components.

The same 12 deletion mutants were incubated against CD2301 and none of the phages were able to plaque on this host (Figure 10C). The WT control induced at $1.77x10^4$ p.f.u./ml which is on the higher end of average induction titre, therefore the mutants should also induce as expected. Del1-6 were not expected to plaque as these were ΦCD1801 phages with removed genes, and no additional components to allow plaquing on CD2301. Again, Del8/9/11/12 were not expected to plaque as they were missing the RBP gene. In parallel to Del1, Del7 was also unable to plaque on its original strain, showing the *Hyp12* gene as necessary for infection in both phages. Del10, equivalent to Del4, but in ΦCD2301, was the only mutant with potential to produce plaques. However, an in-depth look at the titres can be used to explain the lack of plaques from the Del10 mutant. Del4 in ΦCD1801, caused a ~5000-fold reduction in the number of plaques, and if the average induction for ΦCD2301 was 17700 particles, the calculated expected number of Del10 particles would be 3.5 p.f.u./ml. However, per protocol, only 200µl of the sample is usually measured, making it 0.7 particles per undiluted plate. Even if a single plaque did appear on the plate, it may be easily missed as typically it is not enough to indicate infection. Therefore, while Del10 may still produce limited plaques, they would be well below the detection limit, and still, as with Del4 – making it a very inefficient and dysfunctional phage.

5.2.4.4.2 Exchange mutants

Next, each exchange mutant phage was tested against both strains, CD1801 and CD2301. The WT ΦCD1801 induced at $8.62x10^6$ and ΦCD2301 at $1.77x10^4$ p.f.u./ml on average. Several of the mutant exchange phages were able to infect CD1801, and one was able to infect CD2301 (Figure 5.11). There was no data for Ex6 as a mutant could not be generated.

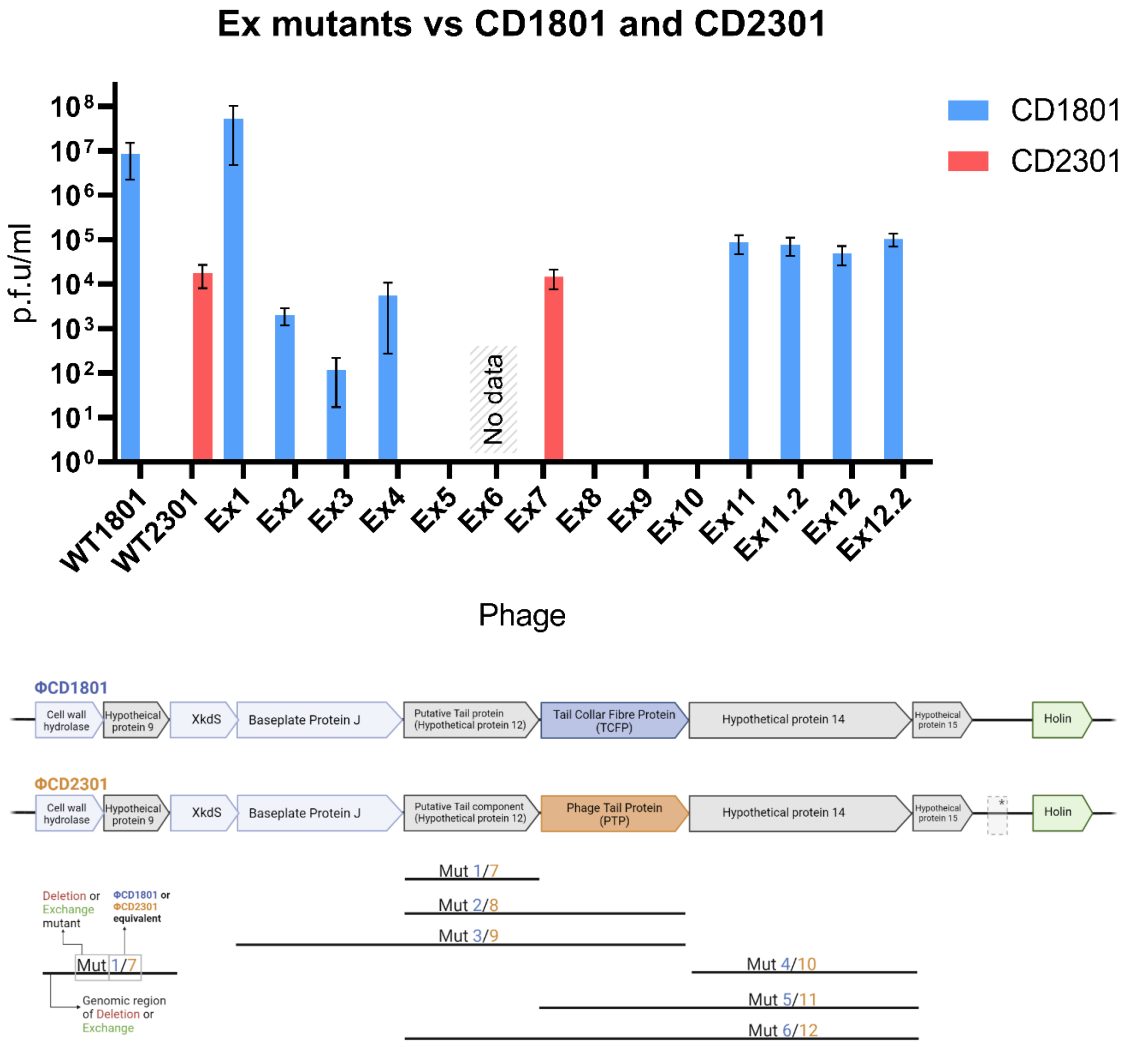


Figure 5.11 Plaque assay showing the Ex mutants against CD1801 and CD2301.

The figure shows screening of mutant Ex phages for their ability to infect CD1801 (blue) and CD2301 (red) strains. Each mutant was tested against each strain, with the results column for each either side of the tick (left for CD1801 and right for CD2301). The mutants were tested in triplicate, with error bars indicating the SEM. Ex11 and Ex12, ΦCD2301 mutants armed with two different tail cassettes, now have the ability to infect CD1801, which was the aim of this study. The reciprocate was not true where Ex5, an Ex11 equivalent in the opposite phage, was not able to infect CD2301. For Ex6, an Ex12 equivalent, there is no data due to unsuccessful conjugation. Map of Mutants included.

The Ex1 mutant was able to infect CD1801, its original strain, once the *Hyp12* gene was replaced, although with that of ΦCD2301 origins. The reciprocal was true for Ex7, which was once again able to infect its original strain, CD2301. Ex1 seemed to be even more efficient, than the WT, infecting at $5.32x10^7$ on average, compared to $8.62\ x10^6$ p.f.u./ml, almost a log increase or 6.17-fold higher. Ex7 infected at a similar rate compared to the WT at $1.45x10^4$ and $1.77x10^4$ p.f.u./ml, respectively. The increased efficiency for Ex1, could be down to a difference in induction rather than a more efficient phage, however that cannot be discounted. There was no statistical significance between the WT and Ex1/7 infection rates (paired two-tailed t test, p=0.40* and p=0.34* respectively). The exchange swapped the *Hyp12* genes between the phages, showing that it is necessary for this gene to be present, however it does not change the host range and it is not specific to each phage. Likely it is a structural gene, unrelated to any specificity and just needs to be present.

The Ex2 mutant was able to only infect CD1801 and at a much lower rate than the WT. The rate of infection decreased by 99.98%, or by 4,300~ times, making it a highly dysfunctional phage. This mutant contained the *Hyp12* and *PTP* genes from ΦCD2301 suggesting some sort of synergy between the two, and showing that the fibre from ΦCD2301 is able to bind to CD1801 in the presence of correct components. The reciprocal was not seen for Ex8, as it was not able to plaque on either strain. This is likely because if any effects were present they were well beyond the detection level, similarly to Del10 discussed earlier. If parallel to Ex2, Ex8 would plaque at 3.4 particles/ml which is not detectable nor functional.

The Ex3 mutant was also able to plaque, but even lower levels than Ex1 and Ex2, and only on its original strain. Out of the three induction events, the Ex3 mutant only induced twice, averaging a 118.3 p.f.u./ml, showing a 99.9986% decrease when compared to the WT. The Ex9 equivalent did not result in any plaques, however, with a 72,865 times decrease in Ex3, only 0.2 particles/ml would be present in theory. The results indicated that inclusion of the baseplate gene (*BaseplateJ*) from ΦCD2301, on top of the *Hyp12* and *PTP* genes decreases the functionality of the phage even further, as the components originate from a different phage and are not adapted to fit well together here. They do provide a restoration of the phage's ability to infect compared to the Del3 mutant, albeit it is very limited and the phage would not be viable in practice.

Ex4 was again, only able to plaque on CD1801, at $5.52x10^4$ p.f.u./ml, which is a 99.94% decrease compared to the WT, although marginally better than the Del4 mutant. The Ex4 mutant had roughly double the titre of the Del4 mutant (Figure 5.12), however, it was not statistically significantly different (paired two-tailed t test, p=0.52), and typically titres so close to each other are considered to be within the same level of infection. This indicates that complementation of the *Hyp14* gene from a different source does not restore the phenotype. Based on this data, the *Hyp14* gene is required for a good level of infection and is specific to each phage as Ex4 did not show significant titre increases. Although, infection still occurred with both mutants, the titre was significantly decreased, and in practice the mutants would make very ineffective phages. This effect could not be confirmed for the ΦCD2301 phage, as, if the impact seen in ΦCD1801 was mirrored, the detection titre would be too low for the Del10/Ex10 mutants to be observed due to a much lower induction titre of ΦCD2301.

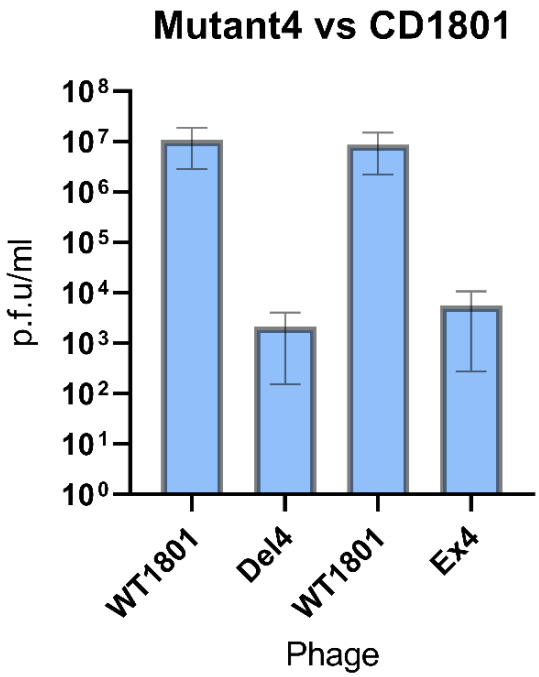


Figure 5.12 A comparison of the Del4 and Ex4 mutants to the WT controls. The graph shows calculated tires based on triplicate repeats, with SEM error bars. There is no significant difference between the two mutants, showing that inclusion of the *Hyp14* gene from different source had little to no effect on the phage titre, and the gene is specific to each phage.

Ex11 was unable to plaque on its original strain, CD2301, however it was able to plaque on CD1801, the new target strain. This was the first ever recorded result documenting a host range change in phages infecting *C. difficile*. The change in host range was achieved by swapping the RBP gene, along with the downstream *Hyp14* gene, making it the host binding range determining module. The mutant phage was able to infect CD1801 at $8.32x10^4$ p.f.u./ml (Figure 5.13). Although it is a ~100 times less effective at infecting CD1801 than its WT phage, its induction titre needs to be taken into account. The WT ΦCD2301 phage infects its host at $1.77x10^4$ p.f.u./ml, making the Ex11 mutant 4.7 times more effective at infecting its new CD1801 host. Therefore, this comparison is not accurate, as it is unknown at what titre the phage induces; the titre is only relative to the strain it induces from (CD2301), however it cannot re-infect the same strain due to change in host range. It cannot be compared as it is the first phage of its kind, to be able to infect a different host through carrying the host specific, binding determining module. It was not induced out of CD1801, therefore, comparing it to the WT ΦCD1801, is also not an accurate measure. Regardless, the hybrid phage is able to infect a new host at a good level, which has never been previously documented.

Ex12 was also able to plaque on CD1801 (Figure 5.13), and its host range changed like in Ex11, where it was no longer able to infect CD2301. This exchange mutant included the *Hyp12* gene in addition to the *Hyp14* and the RBP, in case it was required for structural integrity. The Ex12 mutant infected CD1801 at $7.23x10^4$ p.f.u./ml on average, which was not statistically significant when compared to the Ex11 mutant (Dunn's multiple comparisons test p=0.99, two-tailed t test p=0.69), and their rates were virtually the same. The inclusion of *Hyp12* does not seem to provide the phage with any advantages or disadvantages, particularly relating to host range. The result is consistent with Ex1, where the origin of the *Hyp12* gene did not seem to make a difference, and the phage would be able to infect as long as it was present.

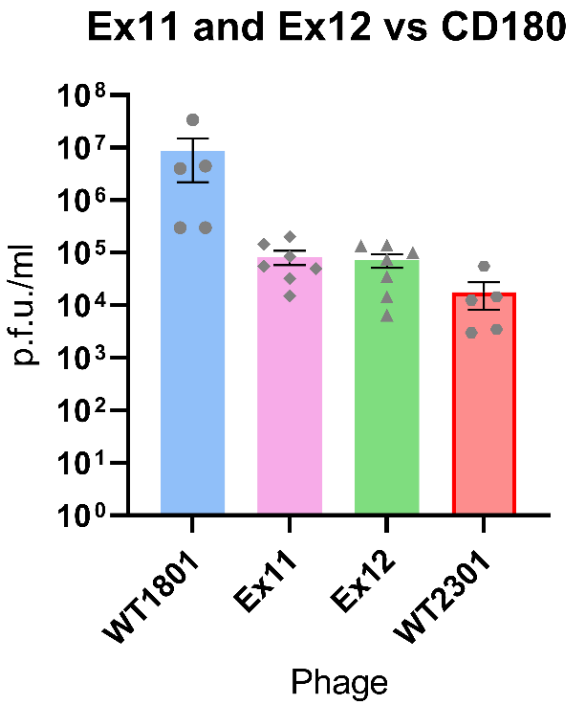


Figure 5.13 Ex11/12 infecting CD1801 alongside WT controls.

The figure shows infection levels of CD1801 by Ex11 and Ex12 phages alongside the WT ΦCD1801 control. The WT ΦCD2301 control is bordered in red as it was plaqued against CD2301 cells as it does not infect CD1801 cells. It was included as comparison since Ex11/12 are hybrid phages containing genomic elements of both phages. Each symbol represents an individual replicate, with the bar heigh indicating the average and the error bars indicating the SEM. There was an overall statistical significance between ΦCD1801 and the Ex11/12 mutants (Kruskal-Wallis test, p=0.013**, Dunn's multiple comparison test, Ex11 compared to WT p=0.02* and Ex12 compared to WT p=0.072**).

Unexpectedly, Ex5 was unable to plaque on either strain. The *Hyp14* and RBP genes were swapped in this case, showing that the *Hyp14* alongside the *TCFP* genes are necessary to infect CD1801. Replacing those genes with those of ΦCD2301 did not allow it to infect CD2301. This was the ΦCD1801 mutant equivalent to Ex11 – which was able to change host range. This is likely because of the genomic contents of the ΦCD1801 phage or because of potential prophages contained within the CD2301 genome. To determine whether Ex5 was able to bind to CD2301 cells even though it was not able to infect it, a binding assay was carried out, followed by qPCR to determine levels of free phage, as this mutant, along with some others were unable to plaque. The mutant phage was also assayed against 7 additional alternative hosts

usually infected by ΦCD2301, using a plaque assay alongside a WT control. Two of the strains, CD88011 and CD8001 exhibited limited infection, whereby the plaques observed were very small and did not dilute out properly – a phenomenon commonly observed later in this chapter (5.2.4.4.3) and in Chapter 6. This phenomenon usually indicates an incomplete infection and interference from the host, making the killing inefficient.

5.2.4.4.3 Ex11 and Ex12 binding to alternative hosts

The Ex11 and Ex12 mutants were assayed against strains infected by ΦCD1801 to determine whether the host range was mimicked exactly, through simple exchange of the host binding module. Strains (n=15) were selected based on previous EOP determined for ΦCD1801 (Whittle et al., 2022). The Ex11 and Ex12 mutants were unable to infect a lot of the strains, especially to the same level as ΦCD1801 (Figure 5.14).

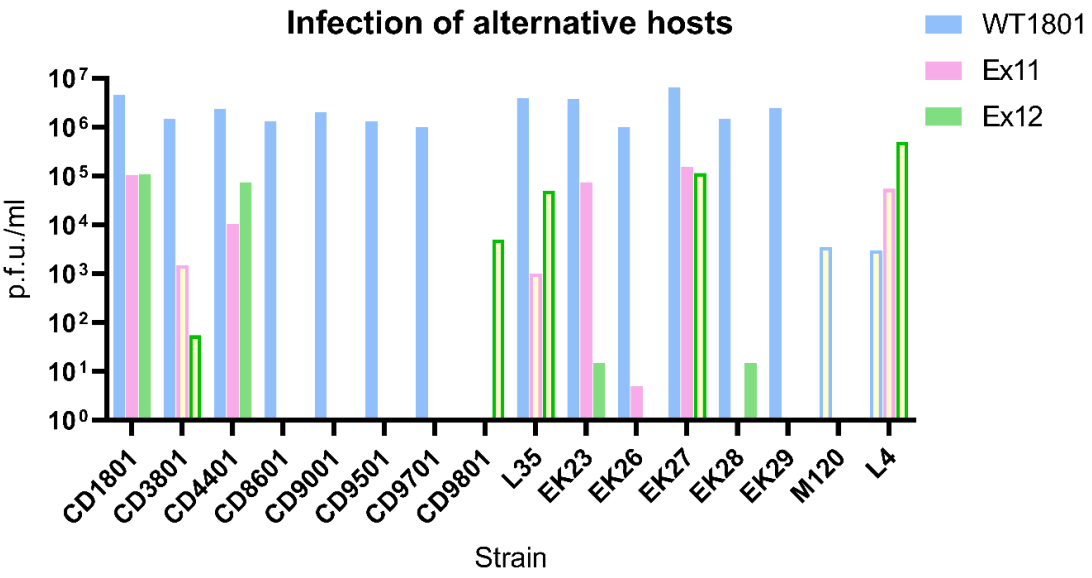


Figure 5.14 Infection of alternative hosts by EX11, Ex12 and WT control.

The figure shows levels of infection of alternative hosts by Ex11 and Ex12 mutants, along with a WT ΦCD1801 control. Where the bar is filled in yellow, with only a coloured border the infection was incomplete, and the plaques observed were very small and on occasion not countable. Overall, the WT phage was able to infect all strains, but CD9801, which may be due to the starting titre and a low EOP. The mutants were not as efficient at infecting (although some of that can be explained by the induction titre) and were only able to infect a portion of the strains.

The exchange mutants were able to infect 9/15 strains combined, with some individual differences. Out of those, five strains showed resistance whereby the plaques observed were very small and sometimes did not dilute properly. This indicates some form of resistance to phage infection, making the infection abortive, incomplete or preventing plaque formation for unclear reasons. It is likely that the reason was founded in genetic contents of the phage, and some genome component of ΦCD2301 was preventing efficient completion of the cycle and was targeted by host resistance mechanisms, such as CRISPR defence systems (further addressed in Chapter 6). The levels of infection between Ex11 and Ex12 also differ in some cases; for example, strain EK23 was infected to a lower degree by Ex12 and strain EK26 was completely uninfected by Ex12. These can be explained by the biological variation and very low titres, where the difference was sometimes seen due to a presence of 1-3 plaques on the undiluted plate, skewing the data in favour of one mutant over the other.

In the instance of strain L4, the exchange mutants outperformed the WT phage, however, this result is not conclusive as all three infections exhibited resistance and the infection rate reflected may not be accurate. Only Ex12 was able to infect CD9801, however, again it was a resisted infection. The resistance events seem to randomly skew the titre and based on previous data are unpredictable and depend on unknown factors that cannot be mitigated. In conclusion, out of the 15 tested strains, only six can be confidently labelled as able to be infected by the hybrid mutant phages.

5.2.4.4.4 Binding assay and qPCR

The induced mutant phages were also tested in a binding assay (3.5.6), with all mutants incubated in a 1:1 ratio with either CD1801, CD2301 or no cells as a control. This served to show, how the exchange mutants were able to bind to cells, even if they were not able to fully infect a given host. The binding assay supernatants from mutants that were able to plaque were then enumerated in a plaque assay to see their binding profile.

The binding data largely showed the expected results, where if a strain was infected, the phage would be able to bind to it at a 97-100% rate, hence it is not presented. There was still some overlap between binding of ΦCD1801 to CD2301 and vice versa, at rate of up to 40%; the data was not clear enough to determine whether this was

simply biological variation in p.f.u./ml or actual binding. Likely, there was some cross-binding between strains, but no cross infection, except in the Ex11 and Ex12 mutants.

As many of the mutant phages were unable to plaque, to determine whether they were still able to bind, a qPCR was planned, with the idea of creating a more complete landscape of genes necessary for binding, and specificity. Lack of plaquing does not necessary mean that the phages are not able to bind, only that they are unable to complete a full infection cycle resulting in a plaque. The phage may still be able to attach and inject its DNA, however as the infection does not result in a release of progeny virions, the binding cannot be detected through a plaque assay. This happens due to host interference with the phage, which tend to be phage-specific, likely due to genome content. For example, a hybrid phage, ΦCD1801 with the correct RBP gene cassette from ΦCD2301 may be able to bind to CD2301, however as the genome of ΦCD1801 is different, a particular region of DNA, not found in ΦCD2301, may be targeted by a CRISPR/Cas defence system, or another anti-phage mechanism. This could be the case with the Ex5 mutant. Another possibility is also exclusion from phage superinfection, whereby any prophages or prophage elements may prevent ΦCD2301 from infecting the host, even if it can bind using the right components.

To determine the degree of phage binding, the amount of free phages remaining in the supernatant is enumerated, following incubation with cells, and a no cell control for comparison. In this assay, instead of enumeration, the number of DNA copies was to be detected by qPCR, rather than actual phage particles. Following the binding assay, a portion of the supernatant was incubated with DNaseI and digested overnight to remove any stray DNA, not encapsulated in a phage particle. After inactivation, the samples were stored at 4°C until ready to be used in a qPCR reaction (3.6.8).

Prior to any qPCR reactions with samples, the protocols and primers have to be very well optimised to ensure the robustness and accuracy of the reaction. The three initial primer pairs designed to target both ΦCD1801 and ΦCD2301 DNA were able to achieve an efficiency of 1.6-1.8 following several optimisations. However, negative controls showed amplification at 30-35 cycles, indicating primer dimer (Figure 5.15).

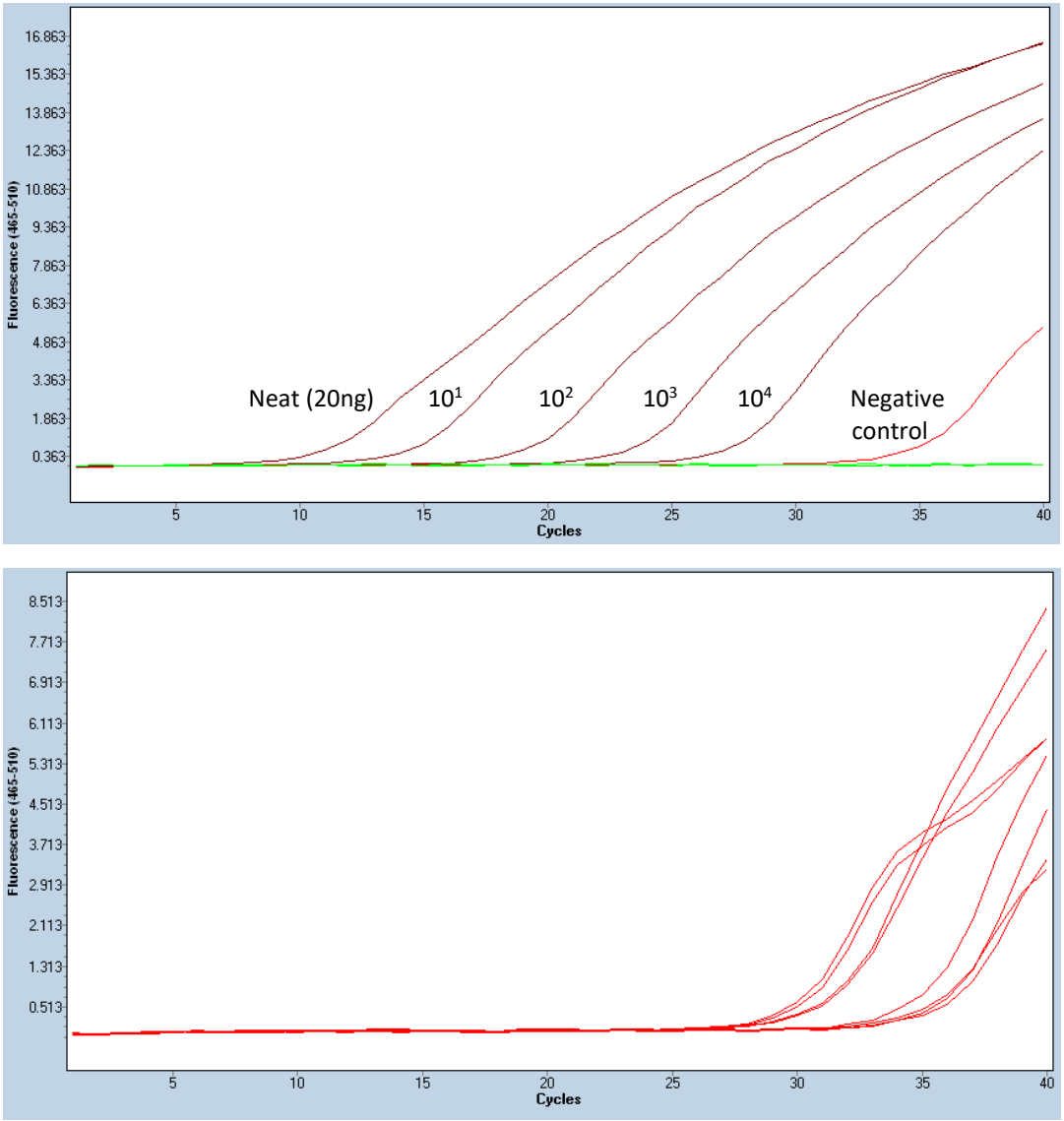


Figure 5.15 qPCR optimisation.

The top panel of the figure shows a good amplification of control gDNA (dark red lines) diluted to 10^4, however in bright red in the below panel (and the smallest bright red line in the top panel) amplification in the negative control (primer only) samples can be seen. This indicates primer dimer and shows the reaction needs further optimisation.

To remove the primer dimer while still amplifying the desired gene product, optimisation reactions were used to determine the annealing temperature (60.2°C) and primer concentration (150nm/300nm), however it was still not satisfactory for the reaction. New primers were re-designed and optimised in the same manner, however no satisfactory conditions were found either, with amplification still seen in negative controls.

Regardless, a run with samples was attempted, where the DNA volume was increased based on preliminary experiments to give the best chance of detection. In theory, the capsids would be able to open during the initial melting step at 98°C. Sample A was the mutant phage incubated with CD1801, Sample B was mutant phage incubated with CD2301, and sample C was a no cell control. Sample C should contain the most phage DNA, as there were no cells present, therefore no potential for adsorption. Either samples A or B would contain less as the phages would be bound in at least some cases, especially with WT phage samples, establishing parameters. However, no DNA was detected in any of the samples on an initial trial run, with 1µl of the digested sample used as template.

With further optimisation the template volume was increased to 7µl, to maximise the chance of the detection, however, it was still unsuccessful. Interestingly, no primer dimer false positive were present, even though they were seen in the negative (no template) control samples (Figure 5.16). The reaction could have failed as there possibly was not enough DNA to be detected in the first place. Typically, in WT controls, the phages are present at 10^5 p.f.u./ml for ΦCD1801, meaning there should be at least 10 particles in 1µl of digested binding assay supernatant. However, if the p.f.u./ml was lower, especially common in ΦCD2301 samples, lack of detection is expected. Another reason for failure could be due to the media used in the binding assay. In the standard protocol, BHIs is used to elute the supernatant as the samples are plaqued and there is no downstream interference. However, as it is an impure broth it contains many salts, sugars and other contaminants which could have possibly interfered with the very sensitive qPCR reaction, especially when used in a higher volume.

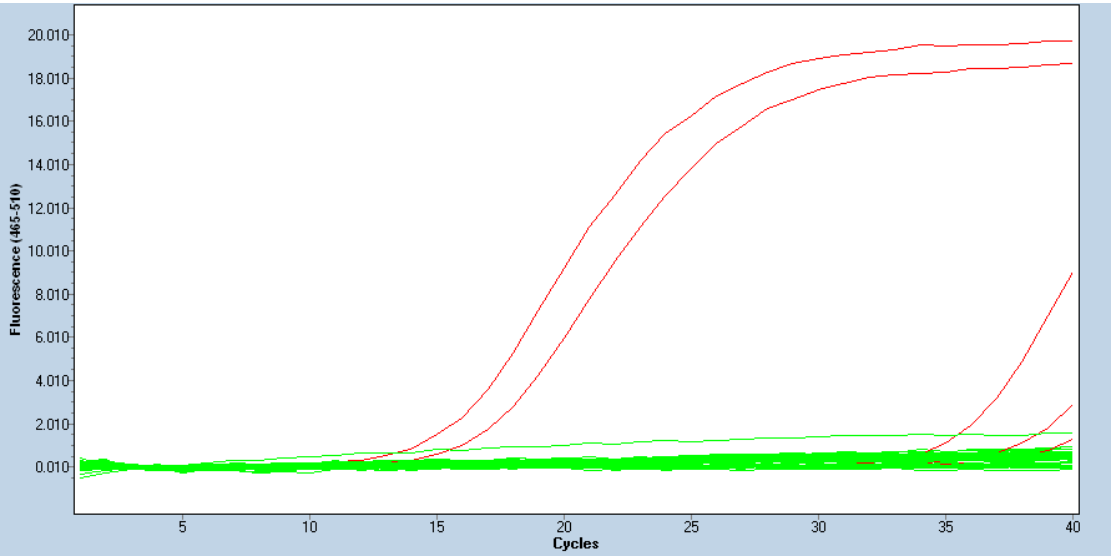


Figure 5.16 qPCR of binding assay samples

Figure shows two amplifications curves for the positive control gDNA. The three lines towards the end, at 35+ cycles, indicate where primer dimer was amplified in negative control samples – not desirable. The binding assay samples were coloured in green and did not amplify, therefore the reaction failed.

The qPCR experiment proved to be unsuccessful in detecting any phage DNA in any of the samples. This idea should be investigated further in the future, however, in this case it is less important as it was not necessary to determine key genes required for change of host range. The Ex11 mutant was able to infect the new target strain, revealing that the RBP gene as well as its downstream accessory gene are required. The other mutants were also key determinants in demonstrating which genes are essential, and which genes pertain to host specificity.

5.3 Discussion

5.3.1 RBP Swapping and Expanding

It is widely accepted that RBPs are host range determinants (de Jonge et al., 2019; Rakhuba et al., 2010) – those being the tail fibres in the case of *C. difficile* phages. Phages use either tail fibres or tail spikes to interact with the bacterial surface, until the receptor is recognised, followed by binding and injection of genetic material. RBPs are a primary determinant of host range, as the first point of contact between the virus and the bacterial cell, hence they were selected for manipulation, following their identification in *C. difficile* phages in Chapter 4.

Several strategies exist for re-directing the host range, many dependent on a high level of characterisation of the tail region. Multiple, high-precision methods all rely on detailed structures of RBPs, which are used as frameworks for engineering of the phages. The structures are generated through X-ray crystallography (Dunne et al., 2018; Garcia-Doval & van Raaij, 2012; North et al., 2019) and cryo-electron microscopy (Ge et al., 2015; Guerrero-Ferreira et al., 2019; Spinelli et al., 2014). None of these structures have been published (Dowah et al., 2021; in pre-print on bioRxiv) for phages infecting *C. difficile*, nor has the structural tail region been characterised. These techniques can be used for strategies like "domain swapping" which allow for production of chimeric RBPs (Ando et al., 2015; Tétart et al., 1998), however, these require a deep understanding of the binding site and mechanism. These methods are not viable in this case as phage genomes in this species are not understood well enough, leading to fundamental characterisation of essential tail genes and their functions in this study.

The initial approach towards changing host range between phages, was attempted through swaps, where the identified tail fibres were replaced with those of phages infecting different strains. In theory, this would have been followed by a host range change, however, mutant phages Swap 1-4 proved unsuccessful. This has been successfully done for several other phages, mainly those infecting *E. coli*, as they are most broadly studied with well characterised genomes. The host range of phage T2 was modified to mirror that of IP008 (Mahichi et al., 2009) and PP01 (Yoichi et al., 2005) through homologous recombination of the RBP genes. In the case of T-even phages, the distal tip of the tail fibre is recognised as the RBP, expressed by *gp37* and with contributions from *gp38*. The approach of this study was more precise, as the

nature of *C. difficile* phages is temperate, and their genomes were edited while lysogenised, then selected for the correct mutation using CRISPR/Cas, allowing for defined editing.

Despite the precise editing of the phage genome to express a different RBP, the host range change did not follow, and the phages were unable to infect the new theorised host or the original host. This was somewhat expected as the tail gene module is not well characterised, and not much is known about how the phages interact with *C. difficile* cells. The binding process is complex involving initial interaction, secondary/irreversible binding, and digestion of the cell surface to inject the DNA in many phages (Quiberoni et al., 2004). The possibility of irreversible binding has been previously demonstrated by (Thanki et al., 2018), however, the phages used in this study did not exhibit this property, or it was not detected with the methods used. It is a well-known phenomenon in T2 phages, requiring involvement from a short tail fibre gene which is encoded by *gp12* and used for secondary irreversible attachment to LPS (Riede, 1987). While the hybrid phage generated by Yoichi et al., (2005), was able to change its host range based on the swap of the RBP, allowing the phage to adsorb to the new host, it was unable to infect it well. The authors generated an "updated" mutant phage, to include *gp12* from PP01 (Hoshiga et al., 2019), which then allowed the mutant to infect the new desired strain.

A similar problem could have been the culprit for the RBP swaps not working in this case. Although the binding mechanisms of T even phages cannot act as a blueprint as they are much bigger (170kb compared to 40-50kb) lending themselves to a more complicated binding process. Lack of homology between mechanisms is evident through its tail structure, where the long tail fibre is encoded by at least 4 genes (*gp34-38*) and is only one of many involved in adsorption. *C. difficile* phages do not have the size capacity for such evolved process, therefore less genes were likely to be involved, giving merit to only swapping the RBP. After confirmation of the tail fibre as the RBP, to achieve a host range change, the next logical step was to swap it, before further investigation. Swaps 1 and 2 were a lot more likely to work as the two phages ΦCD1801 and ΦCD2301 were a lot more closely related. Yet, none proved to be functional suggesting that another gene was involved, or the tail fibres were not compatible to be swapped due to structural issues. Interference from the new hosts could have also been a problem, however, that could not be controlled for.

The attempted expansion of the host range through the ΦCD1801E-1 mutant also did not work, likely for the same, at this point unknown reason as the swaps. Presence of multiple RBPs has been observed in *Salmonella* (Gebhart et al., 2017) and lactococcal phages (Hayes et al., 2018). Such phages have a broader host range than their single-RBP counterparts, although their baseplate architecture differs, allowing for expression of both RBP genes in the tail tip of mature virions. The secondary RBP was found upstream of the "classical" RBP gene, influencing the decision to put a second RBP also upstream in the ΦCD1801E-1 mutant. Another reason for this was to ensure the expression of this gene. The mutant phage was able to infect at a high level, and induced at 10^8, which is on the higher end of the typical induction rates. However, its binding capability seemed slightly diminished, by ~20%; the WT ΦCD1801 binding rate was always high (97-99%). This indicated that there was some interference in binding due to the additional RBP, likely taking up space in the baseplate, but not interacting with the cell surface, probably due to misfolding or structural incompatibility.

In theory, since the N-terminal domain of the RBP is conserved and typically responsible for anchoring the fibre, only swapping the C-terminal end, responsible for binding (Latka et al., 2019) should be enough to achieve the same effect as swapping the entire fibre (Latka et al., 2021). Mutants PartialSwap-1 and 2 were designed for this purpose, however, their generation was never complete due to issues with cloning and time restrictions. As the N-terminal domain (DUF3751) would remain unchanged, the fibre would be anchored and its properties would be examined purely based on binding ability, eliminating structural interference. The mutants would have been used to distinguish whether the ΦCD1801E-1 mutant and the Swap mutants had issues with anchoring or if other factors were at play. Regardless, the expansion mutants resulted in a successful host range change, but further fundamental characterisation is always useful in enlightening the binding landscape.

5.3.2 Characterising the structural tail gene module

As the swaps and expansion did not lead to any functional phages, further investigation of the accessory genes was required. Due to lack of characterisation, any genes involved in the binding process to *C. difficile* cells have not been described, outside of the RBP. As these processes are complex, the involvement of another

gene was likely, although unclear in what capacity. Thus, the bioinformatic analysis of the tail genes was used as the basis for planning which genes were likely to be involved and which combinations they should be swapped in.

5.3.2.1 Upstream structural genes

The primary reason for selecting genes upstream of the RBP – Mut 1-3 and the equivalent 7-10, was to ensure structural integrity of the mutant. Especially if secondary binding (Thanki et al., 2018) is involved, since conformational changes occur (Bertozzi Silva et al., 2016), it is important for conserved domains to be in place to facilitate the process. Although the genes had high homology scores, as little as a few amino acid difference, in key positions, can have an impact on binding and function of the molecular machinery. A single amino acid has the potential to change the host range, as reported for phage JG004, where a lysine to asparagine substitution resulting from a point mutation broadened the host range (Le et al., 2013).

The experimental data showed that the structural genes upstream were not involved in host specificity and just had to be present. *Hyp12* exchange (Ex1/7) between the two phages showed the gene as essential, however, not specific as Ex1/7 could still infect its original hosts at the previous or higher rates. It is still unclear as to what the function of this gene exactly is, as none can be predicted by homology. Software such as AlphaFod2 (Jumper et al., 2021; Senior et al., 2020) may be useful in this case and it has been used previously to elucidate the binding mechanisms (Goulet & Cambillau, 2021, 2022) using artificial intelligence and predict function based on 3D models rather than amino acid homology.

Based on the typical structure of *Myoviridae*, such as phage Mu, the *Hyp12* gene is predicted to be a part of the baseplate wedge due to shared similarity with *gp48*. The wedge is made up of two of each *gp46-48*. Three wedge dimers assemble to form the baseplate along with the central hub encoded by *gp43* and *gp44* (Büttner et al., 2016). The tail fibres are encoded by *gp49* and *gp50*, with a short tail fibre encoded by *gp45*. It is highly possible that *Hyp12* is related to the wedge protein as it is in the correct location and is not a host determinant. Although, the multiPHATE2 pathway suggested the gene as a T4-like short tail fibre homolog (*gp12*), there is little basis for it to present as such. In phage Mu, *gp45* is always found downstream of the baseplate gene (Pfam database), and it is responsible for partially determining host

range as the tail spike used to pierce the cell surface. The *gp12* gene of T4 can be likened to the *gp45* of phage Mu, whereby it is responsible for the secondary binding (Riede, 1987) . Short tail fibres are not encoded in close proximity to the long tail fibres, and are far upstream of the baseplate in T4, providing additional evidence for lack of support. However, as little is known about *C. difficile* phages, neither possibility can be excluded, and the *Hyp12* could very well be a *gp12* like structure, supporting the secondary binding of the phage. It appears not to contribute to specificity, however, the short tail fibre receptor could be the same for phages ΦCD1801 and ΦCD2301 as they are closely related and the gene may not be important for modulating the host range in this case.

The inclusion of the RBP and the baseplate in Ex2 and Ex3 respectively, seemed to decrease the effectiveness of the phage at binding to the host. The phage was only able to bind to the original host for ΦCD1801, as the equivalent mutants (Ex8/9) were below the detection level. These exchanges showed that the upstream genes are not important in host range determination, and although plaques were produced, the phages were realistically not viable for any therapy. Viability decreased with the inclusion of more foreign elements. In future design of semi-synthetic phages, based on this data, it would not be beneficial to include genes upstream of the RBP. Their inclusion does not improve binding, likely as the swaps make the molecular machinery less efficient at binding as the parts may not fit quite right. Although it may also be a case by case basis – this data alone cannot be used to predict this effect for every Myovirus infecting *C. difficile*, without further experimental validation.

This data indicates that the structural genes are not a problem and their manipulation is not necessary, likely due to conserved domains allowing the fitting of the RBP seamlessly with other factors at play, downstream from the baseplate structure.

5.3.2.2 Downstream accessory gene – *Hyp14*

Ex11 mutant revealed to be the key answer to completing the host range puzzle and achieving one of the aims of this study. This exchange redirected the host range of ΦCD1801 to reflect that of ΦCD2301 by inserting two of its tail genes – the RBP and the downstream *Hyp14* gene. This phenomenon has not been seen in literature,

where the accessory gene of this size was required to accompany the RBP to create a functional phage with a modified host range.

The gene on its own is not essential. When deleted, the number of functional particles is reduced severely, by almost ~5000-fold, decreasing the titre by 4-logs of magnitude, down to 10^3. Single exchange of the gene had no effect, suggesting that by itself, it has no direct impact on binding and it needs to be paired with other genes. It serves as an accessory for the RBP gene in some capacity and is specific to the RBP of the phage it originated from. The most obvious suggestion for its function is that it is a chaperone-encoding gene, required for correct folding of the RBP structure, based on phenotypes tied to its modification.

Initial assessment of the gene did not show such predictions, as it atypically large in size compared to other identified chaperone genes. Two of the most well recognised tail assembly chaperones, *gp38* in T4 and *Tfa* in phage lambda are a 183aa and 194aa, respectively. The Hyp14 protein is 497aa and 565aa in ΦCD1801 and ΦCD2301, respectively, making it almost 3 times as big. It has been shown that gp38 is essential for correct folding of the *gp34-37* encoded long tail fibre, specifically the chaperone interacts with gp37 which is the distal tip of the fibre, and cannot oligomerise (Hashemolhosseini et al., 1996) in absence of gp38. Although Tfa and gp38 only share 40% homology, Tfa can be used to substitute gp38 leading to formation of infectious T4 virions with fully functioning tail fibres (Montag & Henning, 1987). T4 requires a second chaperone (Herrmann & Wood, 1981), gp57, responsible for assisting in folding of the long and short tail fibres encoded by *gp34*, *gp37* and *gp12* (King & Laemmli 1971). The *gp57* encoded protein is only 79aa in size (Matsui et al., 1997) and is likely used to keep the monomers separate until other tail parts are folded correctly (Bartual et al., 2010).

The role of *gp38* in T4 appears to be different to the role of a comparable *gp38* gene in phages T2 and K3, and Ox2 (Riede et al., 1987). In contrast to T4 not containing the gp38 protein in its mature virions, these T-even like phages have been shown to display this protein at the free ends of the long tail fibres. There, it acts like an adhesin and does not influence the tail folding, rather it serves to recognise the receptor using its C-terminal domain, while bound to the tail fibre with its N-terminal domain (Trojet et al., 2011).

Recently it has been shown that the Tfa of phage Mu actually stays bound to the folded tail fibre and becomes a part of the RBP at the distal tip (North et al., 2019); it has the ability to bind to LPS (the cell surface receptor). This is in contrast to the T4 chaperones, which are seen as true chaperones as they do not form a stable complex with the fibre and are not present in the mature virion. This finding suggests that the Mu chaperone may actually contribute to the host range specificity. In a further study (North & Davidson, 2021), the authors showed that pre-incubating the host cells with purified Tfa from Mu inhibits phage adsorption to its host cells. Their study also showed the same effect for the Tfa from the P2 phage, and the Tfa proteins were specific only to the designated hosts, indicating Tfa's impact in specificity (Hendrix & Duda, 1992).

The Tfa chaperone is made up for the N-terminal RBP binding domain, and a C-terminal RBP folding domain (North et al., 2019) with only a 12aa difference between the Tfa of Mu and P2. A single amino acid substitution in the N-terminal domain shifted the tropism of the chaperone from blocking infection of Mu to P2, indicating that the lone residue is responsible for specificity to the receptor. However, despite having an impact on specificity of the phage host range, simply exchanging the Tfa alone between the two phages did not shift the host range of the phages. The authors showed that the chaperone could only bind in a stable manner to the phage RBP it originated from, thus explaining the attenuation of plaquing of hybrid phages.

To summarise, there are three types of proteins found immediately downstream of the RBP, labelled as gp38 or Tfa. The gp38 protein seems to have a pure chaperone function in T4, where it assists in tail folding and does not relate to the host range. The Tfa proteins of Mu and P2 play a dual role of chaperone, however, are also involved in binding the receptor while bound to the RBP and mediate the virion-host interaction. The final *gp38* encoded proteins of T-even phages like T2 and T6 do not act as tail folding chaperones, instead their role is that of adhesins, mediating the interaction of the phage RBP with the host. The aforementioned Tfa of Mu and P2 are almost a middle group between the two gp38 distinct proteins.

While this cannot be confirmed without further experimentation, the Hyp14 protein is likely an equivalent of the Tfa from Mu and P2. When removed, the phage cannot function well without it, as the RBP probably cannot fold effectively on its own; although some level of function was observed in the Del4 mutant, it was very limited.

When the gene was exchanged (Ex4) between ΦCD1801 and ΦCD2301, the phenotype was the same as that of Del4 showing the protein produced is specific to the RBP it is found downstream of. The ΦCD1801E-1 mutant also provides further support for this as the additional RBP did not prove functional when inserted on its own, and likely required the presence of the accompanying Hyp14, as the native Hyp14 protein was likely incompatible. The protein is only compatible with its designated RBP, and it may also have an adhesin function on top of its chaperone role just like the Tfa of Mu and P2, which bind to the LPS. Future studies can elucidate this role through investigating Hyp14's interaction with the bacterial surface. Advances have been made towards elucidating the bacterial receptor, suspected to be slpA (Royer et al., 2022 - in pre-print on bioRxiv; Whittle et al., 2022) – if the two proteins interact, this will provide further evidence for the binding mechanism. If combined, information from the host and virion perspectives can be able to be used to broaden the host range or at least control and understand it better.

While Hyp14 seems to fulfil the role of chaperone and a binding accessory protein, its role cannot be determined for sure without further experimental evidence. It may be possible that the gene encodes an additional fibre or a protein with enzymatic activities required to inject the genome. Structural biology studies, such as X-ray crystallography or bioinformatic approaches such as AlphaFold2 (Jumper et al., 2021; Senior et al., 2020) may provide further clues as to its exact domains and functions. Based on information present in literature, and the experimental data seen in this study, the Hyp14 possibly mimics the role of a Mu/P2 Tfa protein. Although, its large size does not align with this idea, as well as lack of homology between the two *Hyp14* genes. However, the alignment showed a diverse N-terminal, and slightly less diverse C-terminal end, with a conserved middle, suggesting a conserved function, with specificity to certain proteins. On the other hand, these factors also do not align to any other gene *Hyp14* may be homologous to. It simply may contain additional domains, hence its size, as the Tfa is a very diverse group of proteins, with over 10,000 members (North & Davidson, 2021), thus accounting for limited conservation. Recently, a *C. difficile* Siphovirus tail fibre has been reported (4.2.6.3.; Dowah et al., 2021) and although not homologous, it appears to be ~600aa in size with multiple domains, showing that a large protein may fulfil this function.

The real change of host range occurred in Ex11, where the RBP and the *Hyp14* genes were exchanged together, and the mutant could now infect the new designated strain, CD1801. The Ex12 mutant displayed the same phenotype, except it included the *Hyp12* gene, upstream of the RBP in addition to the other two. As both mutants infected the new host to the same level, the inclusion of *Hyp12* is not necessary and has no benefit to changing host range. Ex12 also further supports the idea that the *Hyp12* gene only needs to be present, however, it is not specific to the phage, likely as long as there is enough homology between the phage of origin and the target phage in the gene sequence or function.

Ex11 can be used as an example in changing the host range of Myoviridae phages infecting *C. difficile*. All other phages of this type isolated in house: ΦCD418, ΦCD08011 (Whittle et al., 2022) and ΦCD9301 (Zuzana Gramanova, PhD Thesis, University of Nottingham), as well as some isolated elsewhere, like ΦC2 (Goh et al., 2007) follow the same tail genome architecture as the *Myoviridae* of this study. The RBP and Hyp14 of these phages can be designated as the host range specificity module and used to change the host range between any of these phages.

This is the first step towards a semi-synthetic phage with a controlled host range which can be changed to infect a designated set of known strains. Although so far it has major limitations which further studies will have to focus on. The generated phage will only be able to infect the strains which are already infected by known *Myoviridae* and its range will not be broader, rather just different. This opens a door to solving the problem of a narrow host range, with further improvement, if a broad host range Myovirus is discovered. Future further characterisation of this locus will allow determination of different domain boundaries, allowing the exchange of domains, leading to chimeric RBPs, as done elsewhere (Ando et al., 2015; Dunne et al., 2019; Yosef et al., 2017). To overcome the problem of limited RBPs available, structure-guided mutagenesis can be used to diversify the RBP binding sites through random or targeted mutations, creating a library of receptor binding domains (Yehl et al., 2019), aiming for those with a broad tropism. These strategies can be applied to the RBP itself as well as the *Hyp14* gene, the latter being of particular interest as its impact has not been reported before, and it has proved to be as important as the RBP in modulating the host range.

5.3.3 Host range vs binding range

When discussed in literature, the tail fibre has often become synonymous with determining host range. On the surface this idea makes sense, as, if the phage cannot bind to a bacterial receptor, it will not be able to infect it. However, often, even if the phage is able to bind to the receptor, it may not be able to infect it. It is very important to distinguish the host range from the binding range (Hyman & Abedon, 2010). In reality, the RBP and accessory elements only determine the binding range, which is one of many factors that influences the host range. This study, especially this chapter, has focused its investigation on the binding range, which is likely a lot broader than the true host range. A host is considered a part of the range when the phage is able to infect it and produce a plaque. Not all binding events results in plaque formation, however, binding is often the only phage-dependant factor.

Other factors can affect the host range – all of them dependant on the host itself. Downstream of the binding, the host can influence a number of phage processes, all of which can prevent a true infection and lead to no formation of the plaques. This is a very important idea to consider, as despite Ex11 (and Ex12) being able to infect CD1801, its new host, through presence of the ΦCD1801 specific RBP and *Hyp14* genes, it cannot infect all alternative hosts of ΦCD1801. The Ex11 phage was only able to infect six out of 15 of ΦCD1801 hosts, and some additional hosts which exhibited resistance to infection through presence of small, often uncountable plaques. That phenomenon, discussed in detail in Chapter 6, indicates an abortive infection where the phage is able to bind, however its infection cycle cannot proceed further due to host interference.

An obvious culprit is the presence of CRISPR/Cas systems (Barrangou et al., 2007) which function like an adaptive immune system and defend the host from previously-encountered phages. Another invading-DNA targeting defence mechanism is type I and II restriction-modification systems which *C. difficile* has been reported to utilise (Purdy et al., 2002). Both of these anti-phage systems are able to interfere as the genomic content of ΦCD2301 is different to that of ΦCD1801, and where ΦCD1801 infection may be permissive, a DNA part of ΦCD2301 could be targeted which is not present in ΦCD1801. Additionally, each phage may have different defence

mechanisms against anti-phage systems, leading to further host range discrepancies, although that is a phage dependant factor.

Another likely factor is the heavy presence of prophages in *C. difficile* genomes (Amy et al., 2018), which exhibit high polylysogeny, where they often contain multiple prophages or prophage elements. Prophages are often implicated in phage superinfection exclusion where an invading phage cannot enter the cell (Bondy-Denomy et al., 2016), as the host is already lysogenised. As much as 25% of all isolated phages exist within bacterial genomes as a prophage (Bondy-Denomy & Davidson, 2014), and this figure is likely a lot higher in *C. difficile*. All *C. difficile* isolated phages are lysogenic, therefore contained within at least one strain of *C. difficile*, presenting a widespread problem, previously implicated in the narrow host range of these phages.

The distinction of host range and binding range is particularly important in regards to *C. difficile* due to the high prevalence of prophages, as often the phages can bind but not infect due to said prophage presence. This problem was likely encountered here, with the even further limited host range of Ex11. Other well-studied organisms, where phage therapy is actively being pursued, do not experience prophage interference to the same levels as *C. difficile*, hence likely the lack of distinction between host range and binding range in studies. Many studies, including review articles, mention engineering the host range, when they are really talking about exchanging or modifying the RBP in some manner, which is only responsible for the binding range. They do not take into account the other factors influencing the host range as much, as the prophage problem is particularly prevalent in *C. difficile,* where it is unavoidable. Although, it is hard to influence the host, which cannot be modified in a clinical setting, to allow infection by the (therapeutic) phage. Instead, ways of allowing the phage to infect despite presence of the prophages should be investigated.

This study has also largely referred to the host range, rather than binding range, to stay consistent with the literature and avoid any confusion. However, it is an oversight to only focus on the RBP and nearby genes as the sole determinants of host range, when they are only responsible for determining the binding range. In other words, we cannot forget the role of the host, when investigating the host range, and

only focus on the binding mechanism, as it is only the first step in a successful infection.

5.3.4 Summary and importance

The attempt to swap the host range between the two phages proved successful, with limitations. The Ex11 mutant showed that a ΦCD2301 phage, armed with the RBP and *Hyp14* gene, from ΦCD1801 is able to infect CD1801, the original host of ΦCD1801. This is a new discovery for *C. difficile* phage research, demonstrating which genes are required to shift the host range, as they confer specificity to binding to host cells. Other mutations revealed that the genes upstream from the RBP are compatible with different RBP genes, and their structures are conserved enough to maintain the integrity of the phage when upstream host range specificity components are exchanged. The Ex11 mutant was unable to follow the exact host range of ΦCD1801, likely due to presence of prophages and/or anti-phage genetic mechanisms. The role of Hyp14 was also revealed to be an unusual chaperone which also confers host specificity. This exciting discovery is the first step in broadening the host range of phages infecting *C. difficile*, which is currently one of the main obstacles restricting the potential of phage therapy against CDI.

Chapter 6:
Isolation and characterisation of novel phages infecting *Clostridium difficile* from the environment

Chapter 6. Isolation and characterisation of novel phages infecting *Clostridium difficile* from the environment

6.1 Introduction

Clostridium difficile infections have risen in the last 20 years and have become recognised as a common cause of hospital-acquired infection following prolonged antibiotic treatment. Many antibiotics have been implicated in facilitating the infection, including those used to treat it – metronidazole and vancomycin. Hence, there is a need for alternative therapies, in particular, narrow-spectrum therapies. As such, research into the potential of phage therapy is rising in popularity. However, the field cannot advance sufficiently without larger available libraries of phages to study the infection and killing mechanisms, and potentially later adapted for use in patients. Subsequently, isolation of novel phages, particularly those infecting relevant clinical ribotypes of *C. difficile* is a pre-requisite.

Early attempts to isolate phages from environmental samples were not successful. This was likely due to the very narrow host range most phages have, partly addressed by this study. Therefore, strain selection is very important, and even if there are phages present in the sample, they may not be able to infect the panel of selected strains and therefore go undetected. Phage isolation requires persistence, a large strain collection and sometimes, some luck. Further attempts were more successful and many phages were isolated since. However, all phages isolated to date are temperate and exhibit narrow host ranges, thus requiring multiple phages in a preparation, making treatment difficult and not profitable for companies. A more extensive library with carefully selected phages will be required for treatments of CDI to be successful.

In most recent times, ribotype 027 and 078 *C. difficile* strains were identified to cause the most infections, and several outbreaks (1.2.1). Based on reports from the CDRN, and overall decrease in infections has been reported, including the 027 ribotype (CDRN Report - gov.uk, 2014). However, it persists to be a problem in certain regions, like the North West and in other parts of the world. Future outbreaks cannot be ruled out as antibiotic resistance continues to increase and treatments may become less effective.

The aim of this chapter was to isolate novel phages infecting ribotype 027 strains and characterise them. This would allow for study of these phages in-house and a potential to find a good phage candidate for clinical studies and engineering for therapy. It will also add to the existing pool of phages available to investigate.

6.2 Results

Previous attempts to isolate phage from environmental samples resulted in successful isolation of phage infecting Ribotypes 002, 014, 023 and 078 (Whittle et al., 2022) yet few phages exist with broad host range activity towards ribotype 027. Therefore, this round of isolation focused on finding phages infecting said ribotype to study.

6.2.1 Isolation of phages from the environmental samples against a panel of 027 strains

6.2.1.1 Phage isolation

Post-anaerobic digester (PD) (n=7) and during-digestion samples (DD) (n=7) were collected from the Oxford Sewage Works plant, supplied from the following locations: Aylesbury, Didcot, Oxford, Swindon, Wargrave, RGD and Banbury. The samples, kindly provided by Severn Trent Water were transported on ice and processed the same day. The processing was previously optimised in-house (Whittle et al., 2022). The sewage samples were enriched with taurocholate, a known *C. difficile* germinant, to encourage outgrowth of spores present in the sample (3.5.7). The samples were centrifuged, then F/S and stored for analysis (3.5.8). PD samples were analysed first as the likelihood of isolating phage was previously higher in these samples, as they were more likely to contain spores after processing. The isolation work was carried out alongside a fellow PhD student and colleague, Sarah Kerr, as both studies were working towards a common goal of isolating novel phages. The workload was shared to speed up the process and increase the likelihood of achieving isolation. Since both studies could utilise any isolated phage and pursue different research goals, there were no conflicts of interest.

The samples were analysed through a previously established and described method (3.5.1) – the double agar layer overlay plaque assay, where 1ml of a 6 hour 1% Subculture of a strain is added to a “phage” sample and mixed with molten Top Agar (0.5%), then poured on top of previously prepared selective agar plates. The plates are examined the next day and all should contain an even lawn, where phage plaques can be observed if phage were present and able to infect the strain.

A panel of 36 in-house ribotype 027 strains was selected for screening the processed sewage samples. The panel included a well-known outbreak strain R20291 from the

Stoke-Mandeville hospital. Strains from England, Europe and USA were present (Appendix 8.2), thus covering a wide range. Initially, strains originating in the UK were used, however, those yielded no results. European strains resulted in isolation of potentially 4 novel phages. From the Wargrave-PD sample, 3 phages were isolated using strains L12 (Germany), L14 (Switzerland) and L15 (Denmark). From the Swindon-PD sample, a single phage was isolated with the L12 strain. The resulting phages were named after their *C. difficile* isolation strain and location: ΦL12W, ΦL14W, ΦL15W and ΦL12S. These phages were furtherer propagated using their isolation strains.

6.2.1.2 Propagating strains

In order to confirm the accuracy of the strain collection, all three isolation strains were sent to the *C. difficile* Ribotyping Network, (CDRN) in Leeds. Strains L14 and L15 came were confirmed as ribotype 027 strains, as expected. However, strain L12 was designated as ribotype 078. It was thought that the initial in-house classification of strain L12 in-house was incorrect, and all following work was carried out under the assumption that the strain being used was the L12 from the culture collection. A personal stock was subsequently used for all experimentation labelled and referred to as JS50. It later transpired that JS50 was different to L12. When a fresh stock of L12 from the collection was used as a propagating strain for a plaque assay, no plaques were detected. Additional tests confirmed L12 as an 027 and distinct from JS50, which was identified as ribotype 078 by CDRN under the initial classification. As the culture collection contains hundreds of strains, it is very difficult to pinpoint the strain identity for JS50. This study reasons, that there is little urgency to do so, since JS50 is only being used for propagation and therefore should not significantly impact phage behaviour. For the purpose of this study and to avoid confusion, the phages will be referred to as ΦL12S and ΦL12W, despite propagation taking place in strain JS50.

Upon initial isolation, the phages were present at very low p.f.u./ml values (Table 6.1). The plaques appeared small (~0.1-0.3cm diameter), usually hazy and uniformly round. Each phage was purified by plaque picking three times subsequently, before increasing titres by using the double agar overlay method with a softer agar (0.2%) where the confluent lysis plates were harvested (3.5.2.1). Despite initial difficulties, a titre of $>10^8$ p.f.u./ml was eventually obtained for each phage.

Table 6.1 Initial isolation values.

Phage	Initial number of plaques (p.f.u./ml)
ΦL12W	13
ΦL14W	3
ΦL15W	7
ΦL12S	8

At this point, following the isolation, the four isolated phages were shared between this study and Sarah Kerr's study. All four phages were found in the samples assayed by this study, however that was just due to coincidence of picking the samples randomly between ourselves. The fact that Sarah Kerr and I equally contributed to the isolation of these phages led to us having equal entitlement to their application. To equalise the workload, this study proceeded with maintaining and preparing the ΦL12 phages, and Sarah Kerr's study focused on ΦL14 and ΦL15 phages. Phage preparations were stored long term at 4^{o}C in BHIs Broth following F/S without significant decrease in titre.

6.2.1.3 Titre increase of isolated phages

Titre of each phage was continually increased using the double overlay agar method (3.5.2), with 0.2% Top agar. On occasion the 0.2% Plus top agar was used which contains Calcium and Magnesium ions (Horgan et al., 2010). Previously, when increasing the titre of another in-house phage ΦCD2315, 0.2% Plus top agar was shown to increase the titre from 1x10^{7} to 5x10^{8}, resulting in a 50-fold increase from a single propagation, whereas the phage remained at 10^{7} when using the standard top agar. ΦL12 phages and ΦCD2315 are *Siphoviridae* with long non-contractile tails indicating that the addition of ions may also be helpful (Yazdi et al., 2018). When tested, this effect was not as pronounced with the ΦL12 phages, and the benefit was only observed some of the time, likely due to the phages having a different ion requirement to ΦCD2315. The variation could also be due to the dilution that was scraped; plates exhibiting confluent lysis are thought to contain the most phage particles. On some occasions, it was challenging to determine which plate contained confluent lysis, especially when the titre reaches above 10^{7}, and confluence appears on 1/10, 1/100 and even 1/1000 dilutions.

It was necessary to increase titre for future experiments; the required titre for successful imaging through TEM is thought to be 10^{7}, as otherwise the likelihood of

visualising phage particles is very low. Similarly for genomic extraction, the titre needs to be equally high, if not higher to obtain a pure preparation that meets the sample submission standards for sequencing.

6.2.2 Characterisation of isolated phage

6.2.2.1 TEM of novel phage

Following isolation, phages were phenotypically characterised through electron microscopy to allow classification, based on virion physiology. All published *C. difficile* phages belong to the order Caudovirales and through specific particle morphology, are members of either the *Myoviridae* or *Siphoviridae* families. Thus far, no *C. difficile* phages have been assigned to the *Podoviridae* family. The phages classified as *Myoviridae* have medium length contractile tails, connected to the head by a neck; the tails are covered by a sheath, keeping them rigid. Those classified as *Siphoviridae* have longer, non-contractile tails which are flexible and lack a sheath.

After a high titre (10^8 to 10^9 p.f.u./ml) preparation of each phage was obtained the phages were imaged under TEM to determine whether they were all different from each other, or if they were the same phage isolated from different samples or by different strains. Before imaging, the samples were prepared by ammonium acetate precipitation (Methods 3.5.5, Fortier and Moineau 2007) and observed following staining.

ΦL12W and ΦL12S were identified as *Siphoviridae*. Both ΦL12W and ΦL12S had very long tails at 311.1±10.4nm and 303.0±11.7nm, respectively (Figure 6.1). Some phage particles in ΦL12S initially appeared to have shorter tails at 215.7±14.6nm, however, after further sampling, most tails appeared to be ≥300nm, therefore it was determined that those measurements were not accurate as they were an artefact or preparation, thus they were excluded from the final result. The sample contained multiple separated tail fragments, suggesting some parts of the tail were detached or ripped off during centrifugation, which is unavoidable, especially with such long tails. Capsids appeared relatively small in size, mostly uniform with 46.5±2.1nm and 49.4±2.1nm in width, respectively. As both L12 phage appeared similar in size, the possibility that they were the same species could not be excluded through imaging alone.

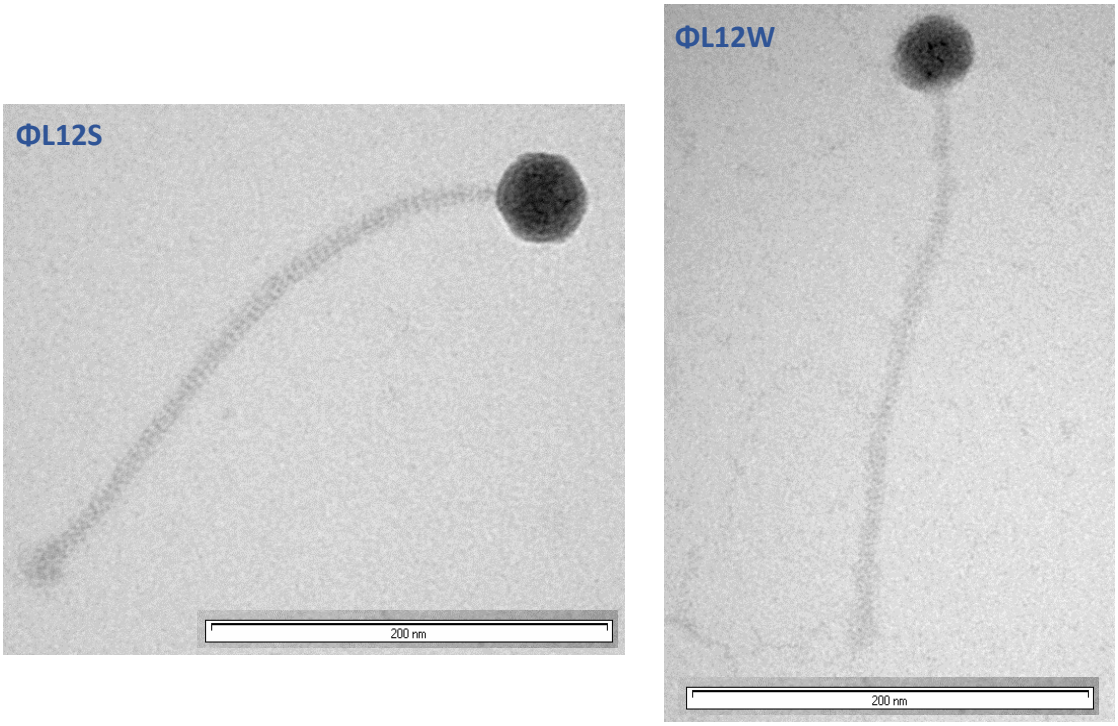


Figure 6.1 Phage TEM of ΦL12S and W

The TEM images obtained, reveal both phages to be *Siphoviridae.* Both appear similar in size, with very long tails at >300nm and relatively small capsids at >50nm. Due to similarity in phenotype, it was possible that they were the same phage, as they cannot be distinguished from each other through size alone.

ΦL14W and ΦL15W were identified as *Myoviridae*. Both had medium length tails, closer to the longer end of the range, at 173.4±9.6nm and 168.2±7.6nm, respectively (Figure 6.2). The tails also appeared much thicker in comparison to the ΦL12 phage as they are sheathed and were roughly double the width of the *Siphoviridae* phage. The phage particles for both samples were mostly intact, although sometimes attached to each other by tails or capsids. The capsids were slightly larger than those of the Siphoviridae at 52.6±3.4nm and 55.6±4.3nm in width, respectively. As with the ΦL12 phages, the ΦL14 and ΦL15 phages were highly similar to each other in size, within the standard deviation expected, therefore they could not be classed as distinct phages without genome sequencing.

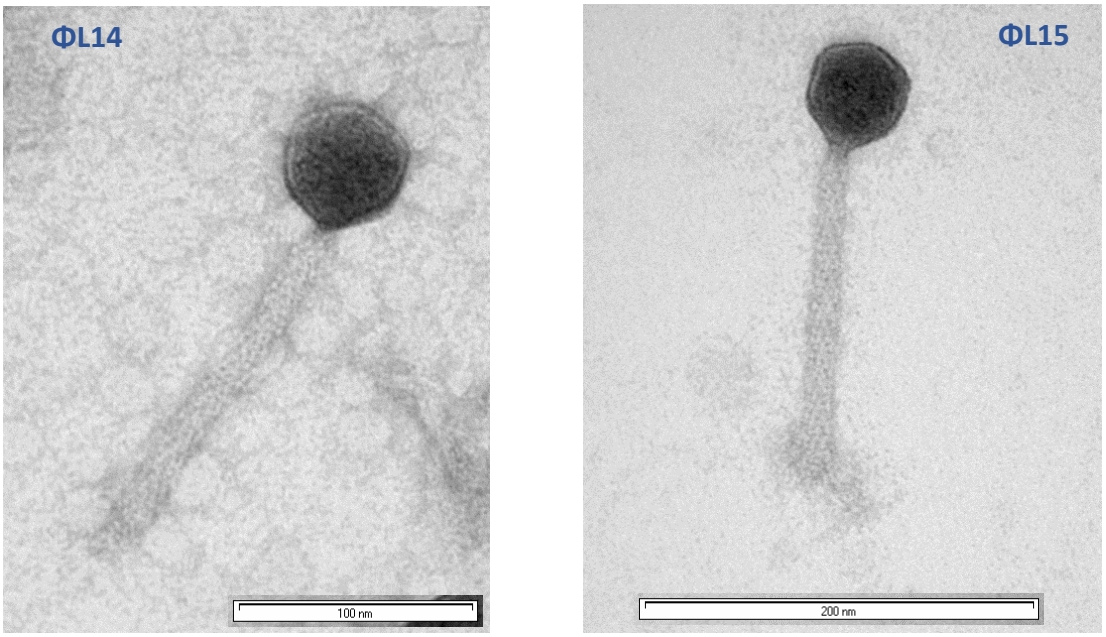


Figure 6.2 Phage TEM of ΦL14 and ΦL15

The TEM images obtained, reveal both phages to be *Myoviridae.* Both appear similar in size, with very long tails at ~170nm and relatively medium capsids at ~55nm. Again, the phages presented as highly similar in phenotype, therefore it was possible that they were the same phage.

6.2.2.2 Genome sequencing and annotation

To determine whether the phages were distinct from each other, the genomes were extracted and sequenced. A NorgenBiotek phage DNA isolation kit was used successfully for the isolation of phage gDNA (3.6.8), with a more rapid column-based protocol without the use of phenol-chloroform or caesium chloride as it common. To achieve sample requirements for sequencing, a large volume of 15-20ml of preparations was used with multiple columns, compared to the 1ml usually required. It was necessary to increase volume as the isolated phage titre could not reach 10^{10} p.f.u./ml, and the yield required for sequencing is much higher than for other downstream applications. This yield is easily achieved for phages infecting other organisms, notably, *E. coli*, however, it is much more challenging to achieve titres beyond 10^8, 10^9 for *C. difficile* phage. This kit has been utilised widely by the phage community (Akhwale et al., 2019; Chaudhary et al., 2021; Luzon-Hidalgo et al., 2021), showing DNA integrity is preserved and of high quality for sequencing.

The phage genome samples were sequenced using Illumina and assembled into single contigs using CLC Genomics Workbench v20, eliminating any bacterial genome background contamination. The genomes were annotated using several pieces of software. Artemis v18.1 was used to predict ORFs, which were manually screened and then confirmed. Some were missed out and annotated later due to alternate start codons which are common in phages. The ΦL12 phage genomes were annotated manually through the use of the NCBI and Uniprot databases. Following the manual annotation, the genomes were also cross-checked through the multiPHATE2 annotation pathway (Zhou et al., 2019; 2021), which utilises Glimmer3, Prodigal and Phannotate for calling gene homology. The corresponding work was carried out for ΦL14/15 by Sarah Kerr (data not shown in this study). Phage ORFs are provided in Tables 6.2 (general sizes) and 6.3 (ΦL12S and W ORFs, with consensus). Alignments of all phages are shown in Figure 6.2.

Table 6.2 Genome sizes and ORFs of novel phages

Phage	Genome Size (kb)	Number of ORFs
ΦL12S	35,682	61
ΦL12W	35,132	59
ΦL14	45,415	70
ΦL15	49,507	75

Table 6.3 ORF annotation and comparison between the ΦL12 phages.

gp	ΦL12S	Consensus	ΦL12W
1		Terminase - small subunit	
2		Terminase - small subunit (part)	
3		Terminase - large subunit	
4		Portal Protein	
5		HK97 family phage prohead protease	
6		Major Capsid protein	
7		Phage gp6-like head-tail connector protein	
8		Head-tail Adaptor protein	
9		Hypothetical protein	
10		Head-tail Joining Protein	
11		M18 family aminopeptidase / Prohead protease	
12		Phage Major Tail protein (phi13 family)	
13		Hypothetical protein	
14		Phage tail domain containing protein	
15		Hypothetical protein	
16		Tail Tape Measure protein	
17		Phage Tail Family protein	
18		Phage Tail Protein	
19		(Truncated) putative phage structural protein	
20		Kelch Repeat Containing protein / Tail Fiber Protein H	
21		Hypothetical protein	
22		XkdX Family protein / (tail sheath?)	
23		Hypothetical protein / Glycine Rich protein	
24		Hypothetical protein	
25		Putative holin	
26		Endolysin / N-acetylmuramoyl-L-alanine amidase	
27		Hypothetical protein	
28		Hypothetical protein	
29		parA family protein	
30		Hypothetical Protein AbrB/MazE/SpoVT family DNA-binding domain-containing protein / PemI	

31		Putative membrane protein / Hypothetical protein	
32		Putative membrane protein / (NSR)	
33		Hypothetical protein / Putative membrane protein	
34		helix-turn-helix domain-containing protein (transcriptional regulator)	
35		Hypothetical protein (NSR)	
36		Hypothetical protein	
37		helix-turn-helix transcriptional regulator / DNA-binding protein	
38	Hypothetical protein	Only present in ΦL12S	Not present
39	DNA-binding helix-turn-helix protein	Different	XRE transcriptional regulator
40	Putative membrane protein	Only present in ΦL12S	Not present
41	Hypothetical protein	Different	Hypothetical protein
42	Hypothetical protein / Putative phage protein	Different	Hypothetical protein / Putative phage protein
43	Phage DNA replication protein	Different size	Phage DNA replication protein
44		Hypothetical protein	
45		Hypothetical protein	
46		Hypothetical DUF2513 protein	
47		Hypothetical protein	
48		Hypothetical protein	
49		Hypothetical protein	
50		Hypothetical protein	
51		Hypothetical protein / DUF3797 domain-containing protein	
52		Single-stranded DNA-binding protein (Helix-destabilizing protein)	
53		Hypothetical protein / Sigma70/sigmaF-like protein / DNA binding protein	
54		Hypothetical protein	
55		Hypothetical protein	
56		Hypothetical protein	
57		Hypothetical protein	

58		DNA-directed RNA polymerase sigma-70 factor	
59		Site-specific Integrase	
60		Hypothetical protein / putative transglycosylase	
61		HNH Endonuclease	

Gene Function	Terminase	Capsid genes	Tail genes
Lysis	Lysogeny	Hypothetical/Unknown	No consensus

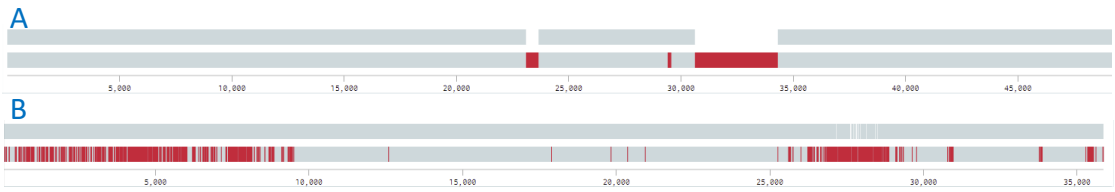


Figure 6.2 Overview of the genome differences between the newly isolated phages.

The figure shows alignments of ΦL14 and ΦL15 (top panel, A), initially indicating discrepancies in three regions (in red). Further work (Sarah Kerr, University of Nottingham, PhD), revealed the sequences to be identical and downstream experiments proceeded for ΦL14. ΦL12S and W were also aligned (bottom panel, B) to show dispersed SNPs (red lines) alongside large regions of dissimilarity (in red). However, sizeable regions were also identical (in grey), indicating that these phages diverged from each other recently.

Phage ΦL14 had discrepancies in the genome (Figure 6.2A), therefore further work was required and carried out by Sarah Kerr's study, which contains the full details. Briefly, the genome was re-sequenced using Oxford Nanopore Technologies MinION system and extracted to be interpreted by several pieces of software for a final version of the genome. Sanger sequencing was also carried out to confirm any absences. The sequencing and downstream processing showed that ΦL14 and ΦL15 were in fact the same phage, therefore only ΦL14 was selected moving forward and its genome is detailed in Sarah Kerr's work.

The annotated genes showed better characterised structural regions, and less characterised regulatory regions, which is often observed for *C. difficile* phages. The genomes were riddled with hypothetical proteins indicating absence of conserved functional domains. Even if the function is hypothesised it can rarely be trusted as it is based on loose and distant annotations.

The ΦL12 phages are distinct, however, they have a 94.1% pairwise identity (Figure 6.2B). ΦL12S contains 61 ORFs compared with ΦL12W which only has 59. However, all the other ORFs are in the same positions and there is a large consensus for 56/59 of them. The three differences lay in a regulatory region which is not well characterised, where two of the proteins are hypothetical and the third seems to be potentially involved in lysogeny (*gp39*). The two extra ORFs (*gp38* and *gp40*) found in ΦL12S are also hypothetical and do not have homology to proteins with established functions. The differences present as individual SNPs (Figure 6.2B) in some areas of the genome. Some genes contain no SNPs, others contain a single SNP or multiple.

The phages are clearly very closely related and likely only just recently diverged from each other with so much similarity. It is worth noting that both ΦL12 phages contain an integrase, marking them as temperate, which all published phages of *C. difficile* are. However, ΦL14 does not contain any genes which are homologous to any integrases previously identified, which indicates that it could be the first lytic phage of *C. difficile* isolated so far. There is no genetic remnants of an integrase, therefore the phage could have never had one in the first place, or lost it completely. ΦL12 and the ΦL14 phages belong to different virus families, thus it is not logical to contrast the two.

ΦL12S and ΦL12W are the most closely related to each other, with the next closest relative being ΦCD6356 at 93.5% and 94.6% similarity, respectively. The other phages that are related are a mixture of Siphoviridae and Myoviridae, and the next closest relatives for both are ~69% or less related, making the ΦL12 phages quite distinct. Phage ΦCDHM19 is the closet relative of ΦL14, sharing 90.1% homology, alongside ΦCD119 sharing 87.4% homology; the other phages in Figure 6.3 are also closely related unlike for the ΦL12 phages. Interestingly, all phages closely related to ΦL14 are Myoviridae, showing further evidence of ΦL14 fitting into this viral family.

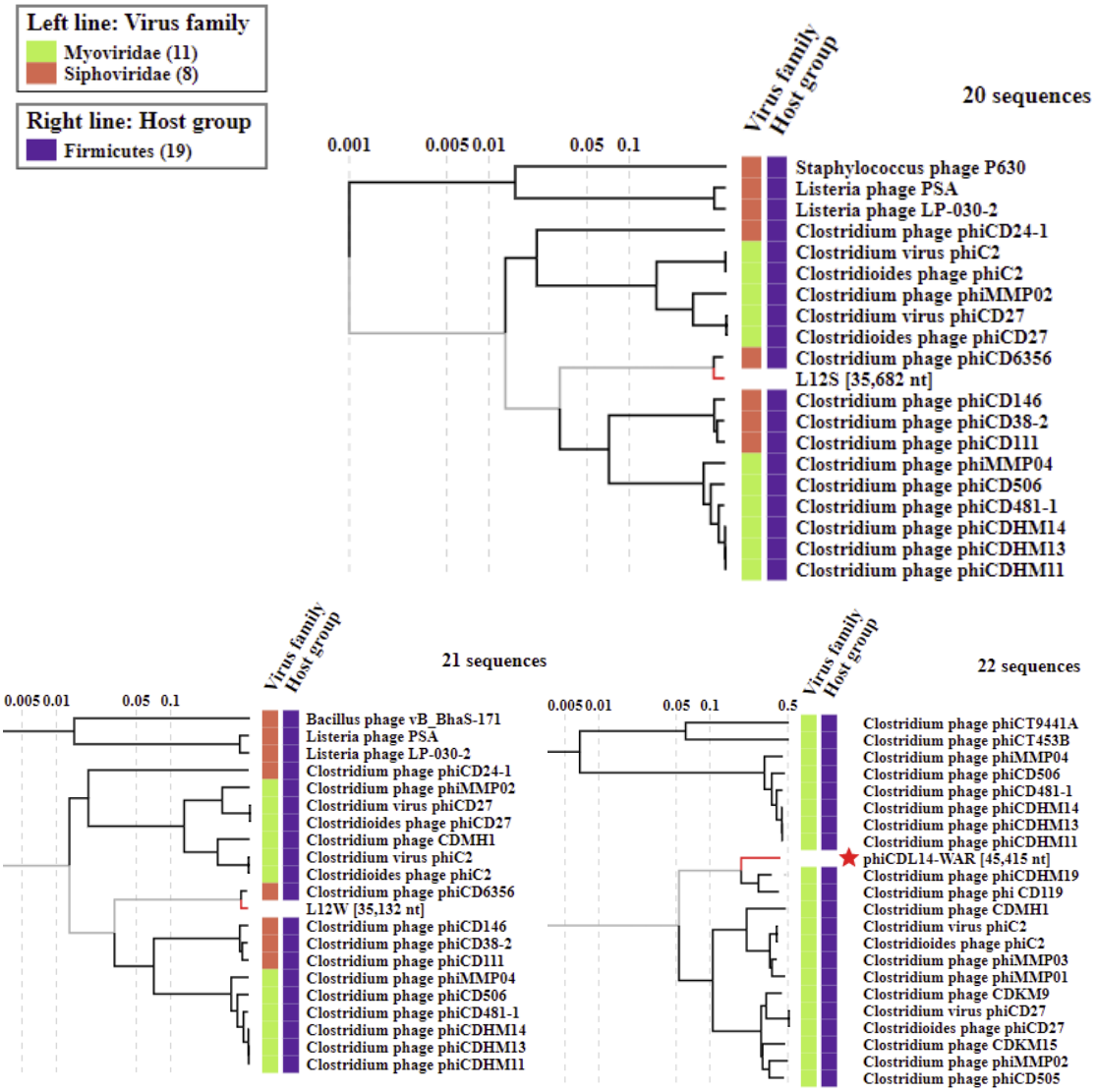


Figure 6.3 Dendrogram of novel phages and their closest genetic relatives.

The three dendrograms (proteomic trees) were generated for each phage using the ViPTree server (Nishimura et al., 2017) which utilises tBLASTx to compute genome-wide sequence homology. These show the closest phage relatives of these viruses. ΦL14 is only related to *Myoviridae*, whereas the ΦL12's are related to both *Sipho*- and *Myoviridae*.

6.2.2.3 Host range analysis

The effectiveness of a phage as a potential treatment is largely dependent on its ability to infect a wide range of hosts. Phages previously isolated in-house (Whittle et al., 2022) had host ranges generally confined to a single ribotype. The newly isolated phages, ΦL12S/W and ΦL14 were assayed against a panel of *C. difficile* strains from the SBRC strain collection (n=190), covering 33 ribotypes. The initial stage of

screening of the phage included all four phages as the genetic content of ΦL14 and ΦL15 was unknown and they had to be treated as separate phages. The bacterial lawns were prepared as according to the double overlay method (3.5.1), without the addition of phage, then transferred to the anaerobic chamber once set. Each plate was divided into quarters, one per phage, and the phage preparation was spotted in 10µl as undiluted and as a 1/1000 dilution (Figure 6.3). The lawns were observed the next day following incubation and strains which were deemed positive were selected forward for a full screening with dilutions to determine the efficiency of plating (EOP). The EOP is a measure of how well a phage infects a test strain, compared to its propagation and/or isolation strain.

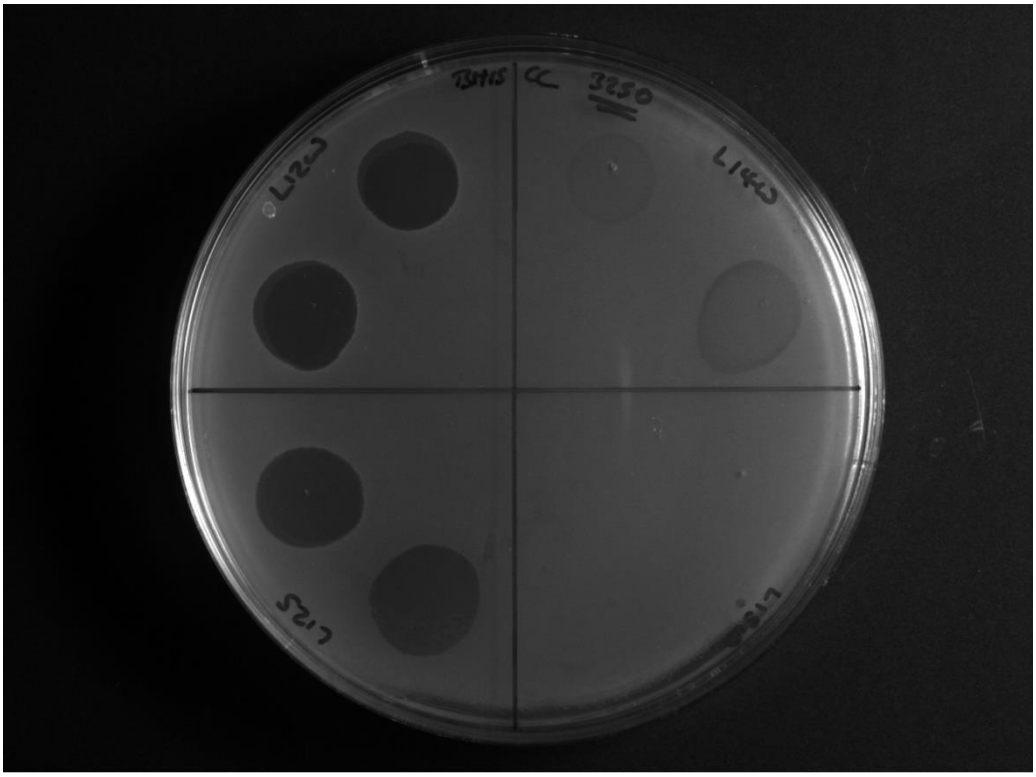


Figure 6.4 Example of initial screening of host range.

The figure shows an imaged agar plate, with a double-overlay, with spotted phage preparations. Each quarter contains one phage and two dilutions (undiluted and 1/1000). The ΦL12 phages (left side) show evident clearing of the lawn, ΦL14 less so (top right), with none observed for ΦL15. These results were noted for each strain, and those which had a single positive spot indicating death of the lawn, were selected forward for a full screening against both ΦL12's and ΦL14.

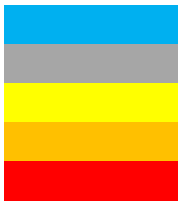
The full screening excluded ΦL15 as it was determined at this point that it was not distinct from ΦL14. A full plaque assay is required for *C. difficile* phage, i.e., each phage dilution is a single plate, as individual spots on a lawn do not give an accurate result and can often be misinterpreted. Due to the extremely high volume of labour and materials required, again, the full host range testing was carried out alongside Sarah Kerr's study with test strains divided equally, where the presented results are an equal contribution from both parties. Roughly half (50.5%, 98/190) of the strains used (Appendix 8.3) showed a positive result for some, if not all phages. Due to a much higher volume of positive strains than expected, the assayed dilutions were: Neat, 1/100, 1/10 000 and 1/1 000 000; meaning the phage stocks were diluted 1/100 rather than the traditional 1/10 employed in all other assays, reducing the number of plates required by half. Despite these modifications, over 2,000 plates were required including any repetitions and for culturing of test strains, therefore these reductions were necessary otherwise this would not be feasible in a timely manner. The results obtained (Table 6.4) were not affected, and full assays were performed in cases with ambiguous findings.

Table 6.4 EOP and host range of newly isolated phages.

CRG Strain name	PCR-ribotype	ΦL12S	ΦL12W	ΦL14
L2	027	x	x	x
L3	027	x	x	x
L6	027	0	0	x
L8	027	x	x	0.56
L9	027	0	0	0.00636364
L10	027	0	0	0.00272727
L11	027	0	0	x
L14	027	0	0	0
L16	027	0	0	x
L17	027	0	0	x
L19	027	0	0	x
L20	027	0	0	0
L21	027	0	0	x
DH349	027	0	0	x
DH32	027	0	0	0.00090909
DH196	027	0	0	x
DH1396	027	0	0	x
DH1432	027	0	0	0.00272727
DH1834	027	0	0	x
DH1466	027	0	0	x
DH1858	027	x	0	0.00272727
DH1751	027	0	0	x
L13	027	0	0	0.00818182
L12	027	0	0	0
R20291	027	5	0.58	0.27272727
DH361	027	0	0	x
DH131	027	0	0	0.00909091
R20352	027	0	0	x
DH478	027	0	0	0.00636364
R12087	027	0	0	1.57142857
L26	001	5.5	0.005	0.002
L25	001	0.07	0.01	0.001
L28	001	0.1	0.20625	0.04285714
L35	001	0.164	4.375	0
L30	001	0.011	0.000513	0
VPI10463	001	0	0	0.001
L34	001	0	0	x
EK26	001	0.213	0.1875	0
EK27	078	0.0005	2.375	0.0002
EK28	078	0.163	2.5	0
M120	078	0	1.25E-06	0
L15	027	0	0	0.35714286

CRG Strain name	PCR-ribotype	ΦL12S	ΦL12W	ΦL14
L24	001	0.025	0.00625	0.000086
L27	001	0.002	3.13E-05	0
L29	001	0.019	x	0
L32	001	x	0.122222	0
DH482	027	0	0	1.43
R24694	027	0	0	0.00285714
DH326	027	0	0	0.00285714
DH1342	027	0.000592	1.22	0.00036667
DH1329	027	0	0	x
DH85	027	0	0	0.00141395
R23970	027	x	x	x
EK15	014	0	x	
EK23	078	0.0005	0.708333	0.000225
EK24	078	0.000317	0.75	0.00015
001	001	0.0675	0.00125	0.005575
O27	027	0	0	x
L1	027	0	0	0.02790698
R24988	027	0	0	x
EK29	078	0.024444	0.230769	
E101		0.036667	0.011538	0.00057143
B14		0	0	0
B32		5.33	1.88	0.0000069
B35		0.0023	0.00045	0
BZ47		0.446667	1.425	0
V23		0.053333	1	0
L5	027	0	0	0.0006
R20298	027	0.0003	0	0.002
R8366	001	0.00155	0.00018	0
NCTC 11205	001	0	3.1	0.04125
NCTC 11207	001	32	6.6	x
NCTC 11209	001	0	0	0
NCTC 11204	001	4.92	3.28	1.27
R20291	027	0	0	x
BI-9	001	0.305	0.3	0
LIV29	001	0.09	0.003	0
CD196	027	0	0	5
TL176	014	0	0	0
M120	078	x	x	
6443 REA type BI 17		0	0	0.025
210	027	0	0	0.02
CD2801	027	x	0	0.00375
CD5101	027	0	0	x
CD9201	027	0	0	x
CD3801	078	0.0325	0.102	0

CRG Strain name	PCR-ribotype	ΦL12S	ΦL12W	ΦL14
CD4401	078		0	
CD8601	078	0.001375	0.289	0.0006
CD9001	078	0.005	0.429	
CD9301	078	0.045	4.238	0
CD9501	078	0.073333	0.762	0
CD3601	015	0	0	0
CD1902	015	0	0	
CD8501	015	0	0	0
CD1801	078	0.000971	1.043	0.000063



The results can be categorised into three sections; the phage does not infect the strain, the phage infects the strain but titre cannot be determined, the phage infects the strain and the titre can be determined. The latter can be divided into three subcategories, by EOP. If the EOP is greater than 1, it indicates a strain that is highly susceptible to phage infection and could be potentially useful in future studies. Where the EOP is between 0.1 – 0.99, the strain is either on par with the propagation strain, or slightly less infected; still useful. If the EOP is below 0.1, the strain is not infected very well, and is therefore less useful for downstream experiments. A total of 85 strains were positive for varying degrees of infection by one or more phages.

Table 6.5 Summary of EOP and number of strains infected in each category.

EOP	ΦL12S	ΦL12W	ΦL14
≥ 1	5	12	4
01. – 0.99	6	12	3
≤ 0.1	23	8	32
Total	34	32	39
Total %	38.2	36	43.8
Average EOP	1.61	1.08	0.27
x – not true infection	8	7	25

Only strains where the EOP can be determined, can be classed as true infections. Strains where the titre could not be determined were designated with an "x" and highlighted in blue (Table 6.4 and 6.5) – for these strains there were obvious signs of infection in the form of lysis on undiluted plates, and then very small, uncountable plaques, particularly towards the edge of the plate down to the 10^{-2} or sometimes 10^{-3} dilutions. This indicated infection which was either abortive, incomplete or could not result in normal plaques for unknown reasons. These types of infections were particularly prevalent in ΦL14, where 25 isolates out of 93 tested exhibited this phenomenon (Table 6.5). It was also observed in ΦL12S/W, however, at a much lower degree, with 8 and 7 isolates, respectively.

Table 6.4 also shows two replicates of R20291 responding differently to phage infection. This was investigated further in ΦL14, where four strain lineages, differing in their origins were assayed. The assay showed that strains CRG0825 and CRG2897 were infected at EOP 0.4 and 0.42, whereas the other two strains, CRG2021 and CRG3661 remained somewhat resistant, exhibiting the incomplete infection characteristic of very small uncountable plaques. The strains have different sources and are genetically divergent which could explain the differences.

There is no obvious "winner" for the most infectious phage out of the three tested. All are able to infect at least three ribotypes: 001, 027 and 078; the latter two being particularly relevant clinical ribotypes. Some strains infected are of unknown ribotypes, therefore the ribotype range could potentially be broader. ΦL14 may seem the most infections on first glance, infecting 43.8% of those tested (n=39), however, only four strains exhibited an EOP ≥ 1, and three exhibited an EOP of 01.–0.99, indicating that effectively only nine strains could be infected efficiently. The other strain infections (n=32) were inefficient with low EOPs, below 0.1, some as low as 0.0000069 for strain B32. A Low EOP does not disqualify a phage completely, as strain B32 still had plaques on the undiluted plate, infecting at $1.85x10^2$ p.f.u./ml; other strains with EOP of 0.1 were also infected at up to 10^6. A lower EOP is simply less ideal, especially in a clinical aspect, and the average EOP for all strains infected by ΦL14 was 0.27, compared with ΦL12S which had the highest average EOP at 1.61.

Based on EOP alone, ΦL12S would be the clear winner, as it infects 5 strains extremely efficiently, with EOPs of over or close to 5 (n=4), and a single EOP of 32 for strain NCTC 11207, which is incredibly high and has not been seen before in-house.

ΦL12W also infected this strain at a high EOP of 6.6, however, ΦL14 was unable to fully infect it and instead uncountable very small plaques were observed. Although ΦL12S infects these few strains very well, the majority of strains (n=23) were again infected at below EOP 0.1, indicating less efficiency. Arguably, ΦL12W can be seen as the most effective phage with the most clinical potential, with an average EOP of 1.08, and able to infect 24 strains at EOP 0.1 or above. Although it had the least total infections at 36%, out of the three phages, the infection of those strains was highly efficient, therefore it can be seen as the most reliable. The overall winner will depend on the specific application of the phage, however, in a clinical setting, relating to host range alone, ΦL12W would likely be most effective.

As ΦL12S and W are both genetically highly similar, their host range does vary slightly. For example, ΦL12W was able to infect strain M120, however, its sister phage ΦL12S was not. The opposite was true for strain R20298, where only ΦL12S could infect it. These results could potentially be explained by the low EOP of infection for both at 0.00000125 and 0.0003, respectively. The test titre (10^7) could have been too low to detect using those strains if the EOP is extremely low for both. However, this does not explain the third instance of difference, where NCTC11205 was only infected by ΦL12W, at a high EOP of 3.1, therefore, infection by ΦL12S was unlikely to be missed. There were also five instances of difference between the two phages in incomplete infections, where titres could not be determined for some assays, for example in strain EK15. Following genetic analysis, there were no differences found in the tail fibre (RBP) gene (*gp19*) responsible for binding to hosts; there was a single SNP in a potential accessory gene (*gp20*), however, it is a synonymous mutation as both proteins are identical at the amino acid level. This indicates that the host range landscape cannot be explained by the binding proteins alone (as also discovered in Chapter 5) and is interesting to observe the differences in host range and EOP between closely related phages. The EOP assay can vary from experiment to experiment and does have limited replicability, however, the differences are too large in some cases to be considered experimental errors or biological variation.

6.2.3 Strain characterisation and Lysogenisation

6.2.3.1 Isolation strain characterisation

The growth profiles of isolation strains, L12 and L14, were characterised to optimise plaque assays and inform future experiments, especially those involving MOI. The strains were grown from subculture in biological triplicate and dilutions were plated in technical triplicate every hour, for the first 8 hours, then at 24 hours. Figure 6.5 shows L12 as a much faster grower, and a more typical propagating strain growth rate, similar to previously mentioned CD1801 or CD2301. It is standard to use those strains in a plaque assay after 6 hours of subculture, which is consistent with these results as the strain reaches its peak after 5 hours.

L14 is a slower growing strain, which is reflected in the growth curve and plaque assay experiments carried out where the strain typically required 7 hours of growth to produce dense, even lawns. On the other hand, L14 may decrease at a lower rate, as the number of viable cells was almost a log higher after 24 hours, when compared to L12; although this does not seem to have any experimental significance.

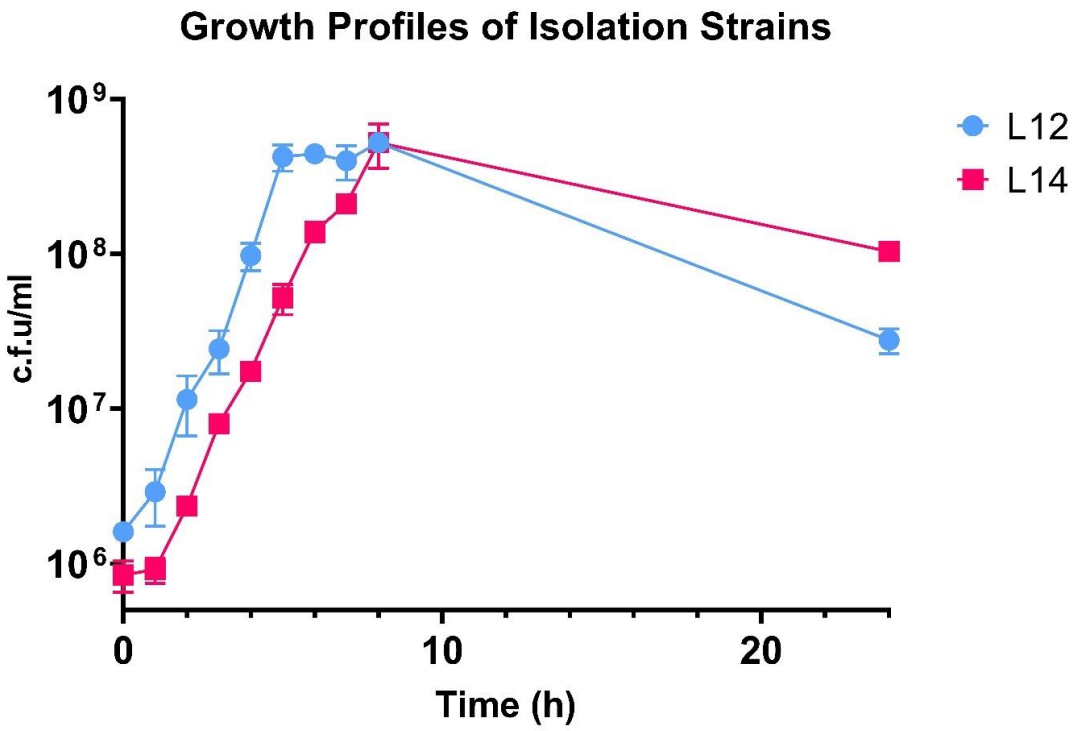


Figure 6.5 Growth profiles of L12 and L14.

The figure shows the growth of the two L12 and L14 isolation strains used to isolate the three phages. Conditions for growth were the same as for a plaque assay where a subculture was inoculated with a 1% overnight growth and measured through spotting dilutions on selective plates. The growth pattern is relatively similar for both, however L12 grows much faster than L14, reaching stationary phase at 5 hours and 8 hours, respectively.

6.2.3.2 Lysogen formation

The final step in phage isolation is to make sure the phages can be stored long term; lysogenic phages are able to integrate into the host genome and be "stored" in the bacterial strain until induced. If a lysogenic phage is able to infect a strain, it can also lysogenise it; the rate of lysogeny is thought to be between 3-10%, although the exact figures are unknown and may vary by strain, with some claiming as high as 15% (Sinha et al., 2017) to below 1% (Weinbauer & Suttle, 1996).

The main in-house method of making lysogens involves a standard plaque assay (3.5.4 – Lysogen from plaque assay), where the next day plaques are "picked" using a pipette tip and re-streaked on BHIsCC plates, then passaged for 3 days. Plaques of lysogenic phages often appear hazy as the colonies that become lysogenised grow inside the plaques. This method relies on picking those colonies and purifying them through passage. The ΦL12S lysogen was successfully obtained through this method,

where colonies 4, 7 and 9 were selected following PCR (Figure 6.6) to confirm they were resistant to further infection and the phage could be induced using Mitomycin C (3.5.2.2, Figure 6.6). This method was unsuccessful for ΦL12W and ΦL14. It was attempted several times, however, this method carries a high risk of contamination with WT *C. difficile* from the lawn around the plaque. To make picking more precise, a fine needle was used, however, this did not improve the outcome. L12S lysogen isolate 4 was selected as the final storage strain as it was deemed most stable.

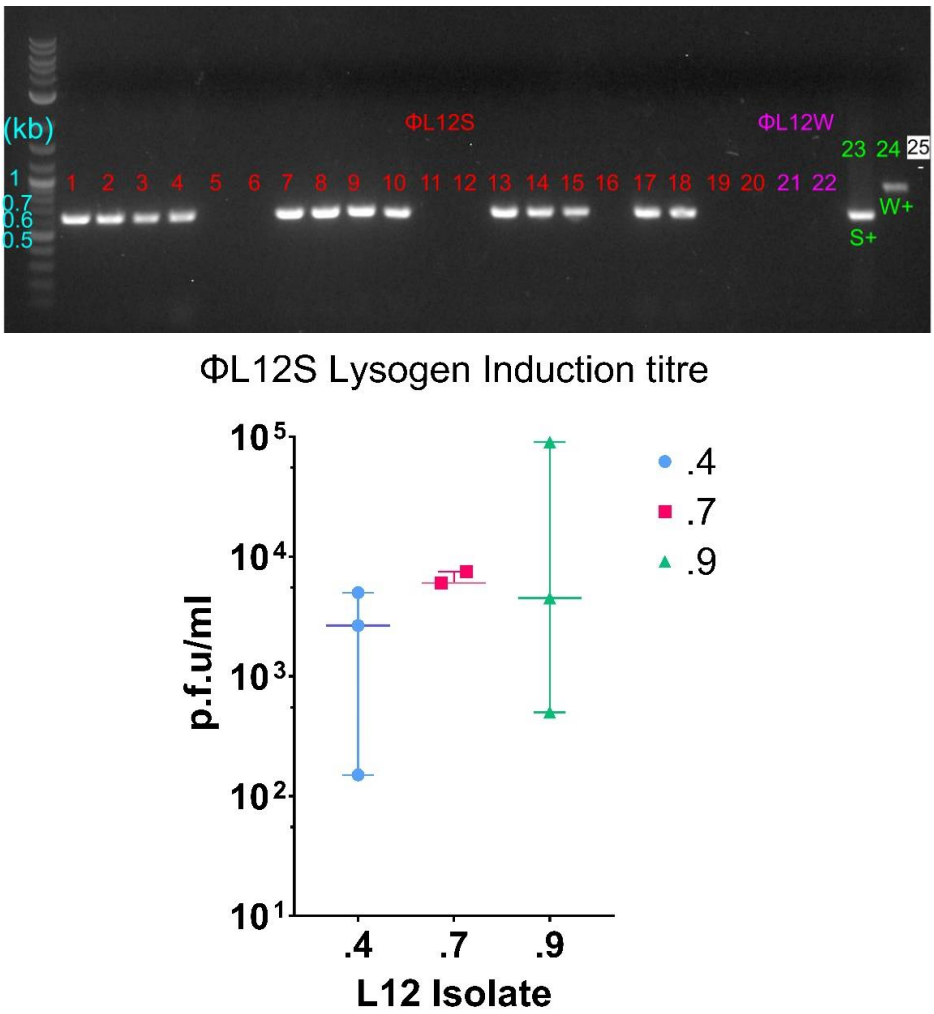


Figure 6.6 Lysogen screening in ΦL12S and titre testing.

The gel image is showing the screening of 20 lysogens for ΦL12S, and two for ΦL12W, with controls. After 3 days, 13 out of 20 colonies were lysogenised by ΦL12S; 0 out of 2 were lysogenised by ΦL12W. Following PCR screening, Isolates 4, 7 and 9 for ΦL12S were randomly selected to be tested further. All isolates were resistant to further ΦL12S infection meaning all colonies were fully lysogenised. All three were tested in triplicate; isolate 7 was excluded as it did not grow consistently well. Isolate 9 induced erratically and was the least consistent in its induction titre. Although isolate 4, induced at lower titres than isolate 9 on average, it was more consistent and reliable, thus selected forward.

6.2.3.2.1 Liquid Lysogen

After unsuccessful attempts with the standard method, ΦL12W and ΦL14 were used to infect liquid subcultures of their respective strains (3.5.4 – Lysogen in liquid). For ΦL12W, initially an MOI of 10 was attempted, calculated based on the L12 growth (Figure 6.5), where a stock of ΦL12W at $1x10^8$ p.f.u./ml, was added at 2 hours where L12 was at $1x10^7$ c.f.u./ml. Samples were taken every 2 hours for 6 hours, then at 24 hours and plated on BHIsCC in dilutions and screened the next day via PCR. In the early hours between 2 and 6 hours, all colonies appeared to have been lysogenised. At 24 hours, 2 out of 5 colonies appeared to contain the phage. Following testing, none of the colonies could be confidently classed as a pure lysogens where the cultures often did not grow well, and lawns were sparse on plaque assays. The PCR result was likely a false positive from phage contamination, rather than the phage genome being truly contained within the bacterium. This also explains the lack of growth of the cultures, as they were likely being killed by the phage as they were growing.

Likely, due to the high MOI all colonies appeared as lysogens early on, however, none of them truly were. Instead, the bacterial culture was killed rather quickly due to the overwhelming number of phage, as indicated by the lack of lysogens at 24 hours. If any were present, they were likely lost amongst the background of other colonies. In the next attempt, a standard MOI of ~1 was achieved by adding an equal amount of ΦL12W at $2.25x10^7$ p.f.u./ml to an equal amount of subculture at 3 hours, where the cell density reached $2.4x10^7$c.f.u./ml. Following sampling and screening, the PCR showed the number of lysogenised isolates increasing as time went by (Figure 6.7); when confirmed in a plaque assay however, the colonies appeared to have no phage or did not grow well. At this point, another method was successful, therefore continuation with this method was not necessary as it was proving unreliable. Despite its success with other phages, the method requires more optimisation, and, in the future, it may be beneficial to purify the colonies by passaging prior to screening.

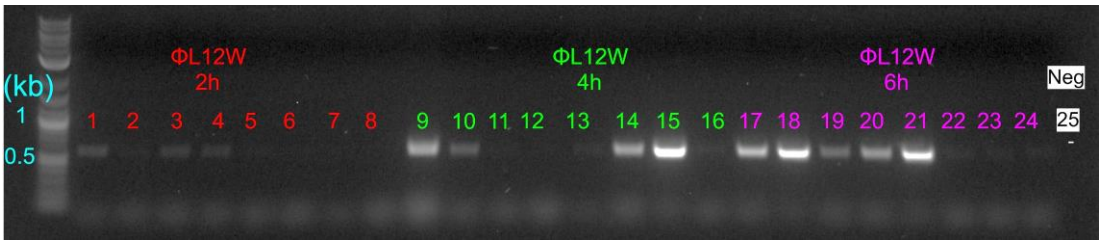


Figure 6.7 PCR screening of potential ΦL12W lysogens.

The figure shows the screening of colonies for being lysogenised by ΦL12W over time. To screen for lysogeny a phage region ~0.6kb was chosen and targeted with a primer pair. As time passed, the number of potential lysogens increased from 3 at 2h (red), to 4 at 4h (green) and 5 at 6h (purple). However, the colonies did not prove to be robust lysogens and were discarded.

6.2.3.2.2 Neat and spot pick

The neat pick (Figure 6.8) and spot pick (Figure 6.9) methods proved fruitful (3.5.4) and lysogens for ΦL12W were easily obtained using both methods. The neat pick method had a 100% efficiency as all colonies picked were lysogens (Figure 6.8), however, only 8 colonies were tested following a 3-day passage. Half of the colonies were screened further; all exhibited resistance to further infection and produced induction titres between 10^2-10^4. Isolate NP1 produced the most consistent, relatively high titre (Figure 6.10), thus it was selected to use in downstream experiments and in long term storage.

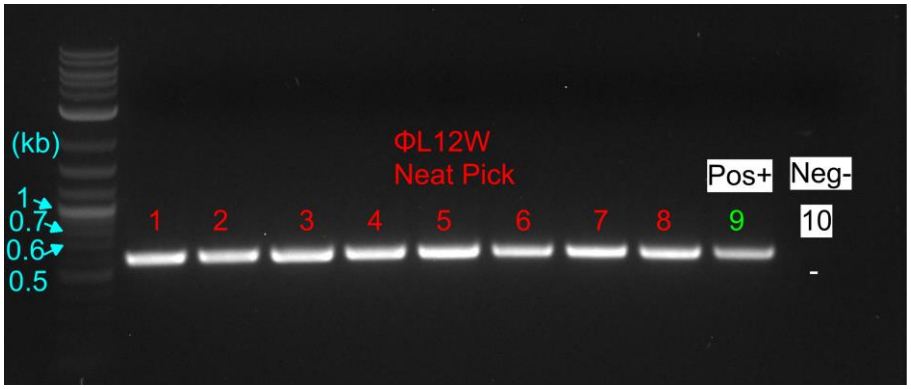


Figure 6.8 Screening potential lysogen colonies generated by the Neat Pick method.

The figure shows potential lysogen colonies (1-8, red) which all appear to be positive for containing the ΦL12W genome. The PCR amplified bands were generated using primers targeting the ΦL12W genome and are ~0.6kb in size, as previous. Alongside is a positive control with gDNA (9, green) and a negative control (10).

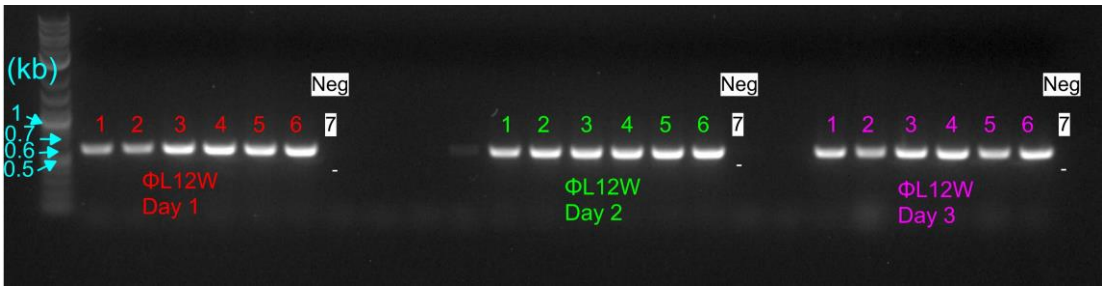


Figure 6.9 Screening potential ΦL12W lysogen colonies generated by the Spot Pick method.

The figure shows potential lysogen colonies (1-6, ~0.6kb) alongside negative controls (7), screened over a 3-day purification period. The colonies are re-streaked to purity using this method and were checked for the presence of the phage genome each day. The results showed that the lysogen remains stable as all colonies retain the genome over the course of the continued culturing.

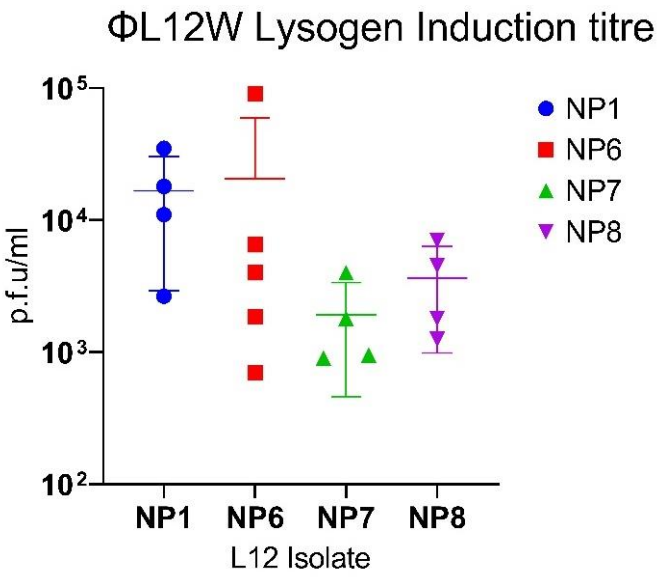


Figure 6.10 Testing the induction titre of ΦL12W lysogens generated using the Neat Pick method.

Isolates 1, 6-8 were selected and induced four times to verify the reproducibility of the induction titre. Isolate NP6 appeared to have the highest induction titre on average at ~$2x10^4$p.f.u./ml, however this was skewed by a single induction event where the titre was much higher than usual and lacked consistency. Isolate NP1 had the second highest average titre at ~$1.7x10^4$p.f.u./ml and remained reliable throughout each induction event, therefore it was selected forward.

These methods were repeated for ΦL14 to show that a lysogen cannot be made as the phage was expected to be potentially lytic, since it lacks the *integrase* gene (analysis by Sarah Kerr) – a hallmark of all lysogenic phages. None of the methods attempted (Figure 6.11 and 6.12), including the standard overlay method resulted in a lysogen. These protocols were repeated for ΦL14 by this study and Sarah Kerr's study multiple times to ensure that the phage was truly not lysogenic. The neat pick and spot pick methods are very reliable in lysogen formation, and alongside the others and multiple repeats, if the phage was lysogenic *in vitro* it should have integrated into the bacterial genome at least some of the time. Sarah Kerr's study also shows that the phage kills the bacteria and the titre decreases over time at varying speed, depending on MOI, further demonstrating lack of lysogen formation. Combined with genome analysis, the phage can be confidently regarded as the first described lytic phage of *C. difficile*.

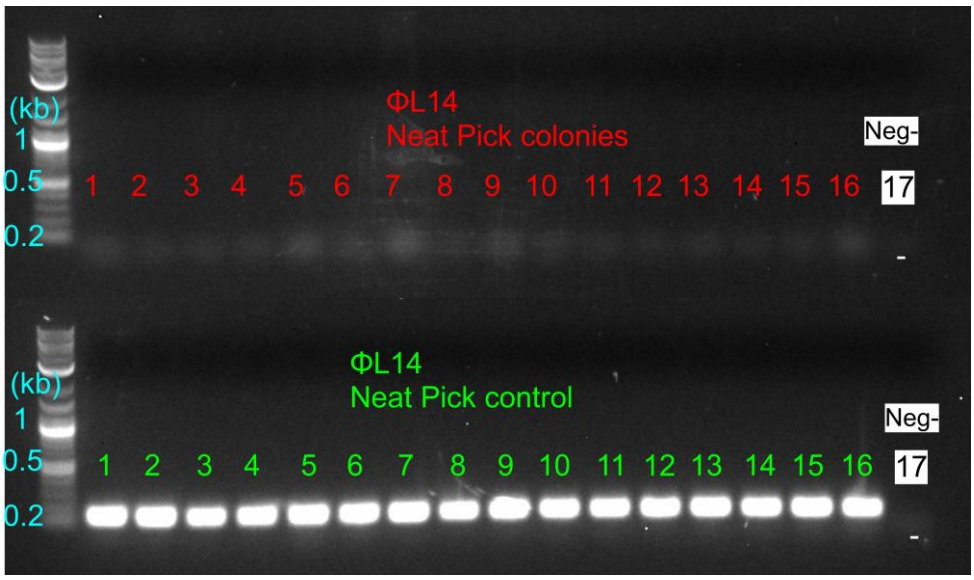


Figure 6.11 Neat Pick method attempted to generate ΦL14 lysogens.

The gel images show a colony PCR of potential ΦL14 lysogenised colonies. The top line shows no positive bands for lysogen formation (1-16, red). The bottom line shows that the colony PCR was successful as the primers were targeting a common *fur* gene to produce a ~200bp product, found in all *C. difficile* isolates. This was to ensure that the colonies were able to be amplified as there was no positive isolates for the lysogen, and that there was no contamination (17).

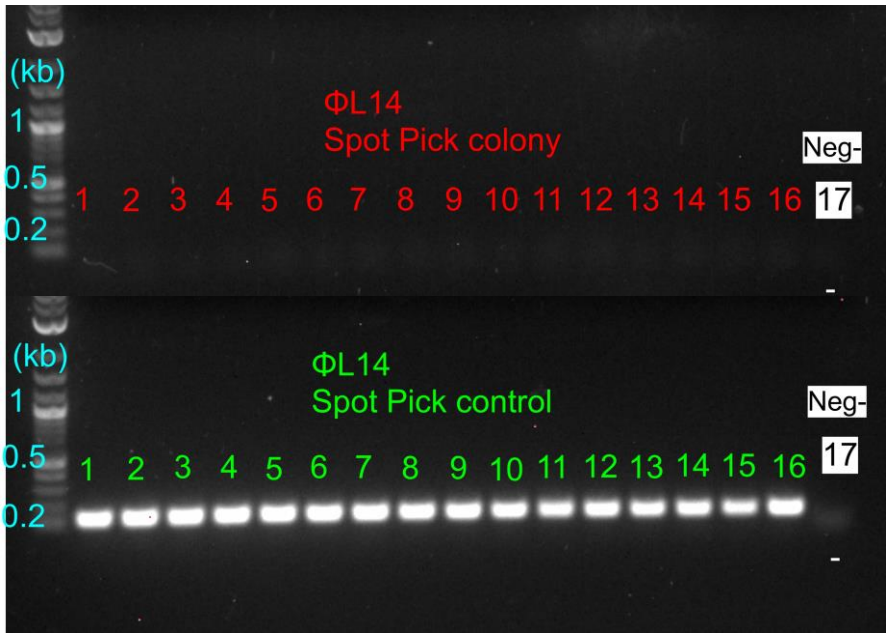


Figure 6.12 Spot Pick method attempted to generate ΦL14 lysogens.

The gel images show a colony PCR of potential ΦL14 lysogenised colonies through the spot pick method. The top line shows no positive bands for lysogen formation (1-16, red). The bottom line shows that the colony PCR was successful as the primers were targeting a common *fur* gene, as above in Figure 6.11. This gel shows no lysogens were obtained, providing further evidence for the lytic nature of ΦL14.

6.2.3.2.3 L35 Lysogenation

As mentioned above (6.2.1.2), the strain believed to be L12 was discovered to be something else at this point, and the strain used the whole time was JS50. Since that strain was unknown and uncharacterised in any way, except for its 078 ribotype, the lysogens for ΦL12S/W were re-made in a more known strain published as "Wilcox 078" and named L35 in-house. The strain was selected due to its previous presence in literature and a reasonable EOP for both phages. Lysogens, L35.S10 and L35.W11 were easily obtained using the spot pick method, with ΦL12S inducing at $4x10^4$p.f.u./ml and ΦL12W at $5x10^5$p.f.u./ml (Figure 6.13), while being resistant to further infection.

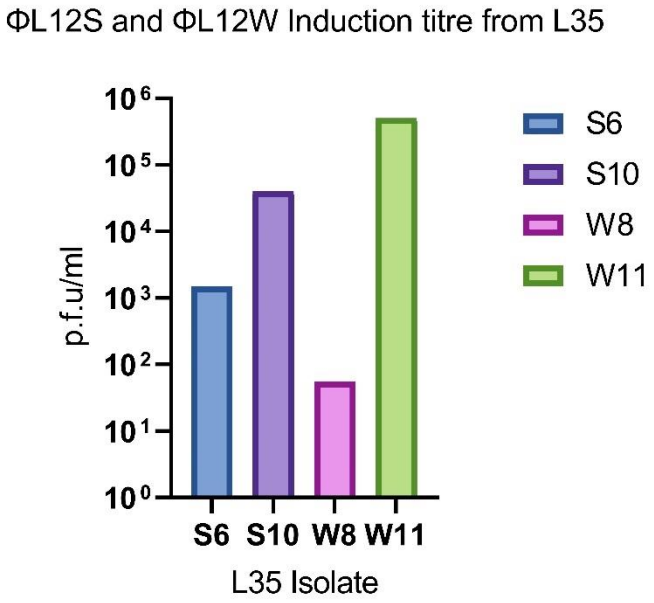


Figure 6.13 Testing induction titre of ΦL12S and ΦL12S lysogens generated in strain L35.

Figure shows the induction titre of four L35 lysogen isolates for ΦL12S and ΦL12W, two each. The isolates selected for storage and further study, S10 and W11 were chosen based on induction titre. Both L35 and JS50 can be used for future propagation, however L35 is a known characterised strain, therefore these lysogens may be priorities depending on application. JS50 is more useful for propagation as the phages are able to infect it at a higher EOP, therefore should be used for titre increase.

6.3 Discussion

6.3.1 Isolation

It was surprising that none of the UK strains resulted in the isolation of phages, however, European strains were successful. The Aylesbury sample(s) were expected to be more likely to produce phage as they contained waste from the Stoke Mandeville hospital, well known for two severe outbreaks between 2003-2005 (O'Connor et al., 2009). The outbreaks were caused by an 027 strain, therefore it was the most likely sample to result in novel phages. However, this can be explained by the high prevalence of prophage carriage within the ribotype (Shan et al., 2012). It can also be a testament of effectiveness of procedures that have been put in place for infection control and prevention, and the samples may no longer contain *C. difficile* or its phage.

The prevalence of PCR Ribotype 027 has decreased sevenfold between 2008-2015 CDRN Report - gov.uk, 2014), increasing the difficulty in isolating relevant phages. Another way of isolating phages is through induction from existing collections. However, even though the phage may be induced, it may be less likely that they would be able to infect a wide range of the 027 ribotype as it is highly lysogenised, and thus immune to further infection a lot of the time.

The isolation resulted in three novel phages, two very genetically closely related *Siphoviridae* - ΦL12S and ΦL12W, and a Myovirus – ΦL14. The ΦL12 phages have genome sizes of ~35kb compared to the larger ΦL14 at ~45kb, consequently with more ORFs encoding potential genes. All three have longer tails than the average members of their groups, however capsids are quite small, at ~50nm. ΦL12S and ΦL12W are genetically similar, although with enough distinctions to be classified as individual phages.

6.3.2 Incomplete infection and host range

The immunity to infection within the 027 ribotype is clearly demonstrated by the host range testing results, showing high resistance by isolates to the newly isolated phage. The testing panel largely comprised of this ribotype (n=49/95, 51.6%), however only three isolates were infected by all three phages. ΦL12S and ΦL12W performed poorly, infecting 3 and 2 isolates respectively. ΦL14 performed much better infecting 23/49 isolates, 46.9%, making it a much better therapeutic phage

against these strains. Additionally, a lot of infection where the titre could not be determined were observed, with 23 additional strains showing this incomplete infection. This phenomenon was also observed in ΦL12S/W, to a lesser degree and it was not specific to just the 027 ribotype. In the other ribotypes infected, 001 and 078, it was observed at a much lesser degree for all phages. As this effect was predominantly observed with ΦL14 against 027 ribotypes, it is likely that existing prophages within the genome are preventing further infection with a similar phage.

In some isolates, 5-6 complete prophage genomes can be found, however, 1-3 is more common (Amy et al., 2018), alongside the presence of genomic islands containing prophage genes it may be more than enough to exclude other phages from entering or completing their infection cycles. Other phage superinfection exclusion mechanisms have been well studied in other species (Labrie et al., 2010), however, examples are limited in *C. difficile*. Given the prevalence of prophages within this ribotype, like Φ027 in R20291, existing phages could be interfering with infection by ΦL14 and others.

Recently, interesting ways of prophages re-programming host expression of genes has been reported (Sekulovic & Fortier, 2015). Out of the 39 bacterial genes, mostly metabolism-related, impacted by the ΦCD38-2 prophage presence, the *cwpV* gene was the most significant relating to prophage interference. The expression of this gene was upregulated to a 20-fold increase in the lysogenised model 027 strain, R20291, responsible for several outbreaks (Stabler et al., 2009). The gene encodes a protein associated with the cell wall, and is expressed under phase variation which is controlled by another bacterial protein, RecV (Fagan & Fairweather, 2014) which changes the on/off switch for the gene, could be influenced by phage genes. In WT cells, only 5% of the cells had the gene switched on, compared to 95% in lysogenised cells. The CwpV protein is thought to play a role in cell adhesion and influence biofilm formation (Reynolds et al., 2011); it can also block the injection of phage DNA preventing superinfection by another phage. WT cells that express this protein are resistant to phage infection, which also carries consequences of its own for the future of phage therapies. Many similar mechanisms to this one may exist and have yet to be reported and could be contributing to this incomplete infection phenomenon.

Another way of preventing phage infection employs CRISPR/Cas systems which are a form of bacterial adaptive immunity and contain a record of past infections (Barrangou et al., 2007) to which the cell is now immune. If the infecting phage or its relative, has previously attempted to infect the cell, it will be recognised by the system as foreign DNA and degraded (Bhaya et al., 2011). CRISPR has been shown to dictate host susceptibility to different groups of phages, by targeting highly conserved genes within those groups.

In addition, phages themselves have been reported to carry CRISPR arrays (Sebaihia et al., 2006), including the model strain CD630 (Soutourina et al., 2013). The prophages likely acquired the spacers targeting other phage during horizontal gene transfer events; these can be used to arm the endogenous bacterial CRISPR against invading phage (Hargreaves et al., 2014), again preventing superinfection. The presence of CRISPR arrays, whether bacterial or prophage carried, cannot determine the host range alone despite some correlation shown, especially as many phages also carry anti-CRISPR proteins (Bondy-Denomy et al., 2013) and can escape the CRISPR attacks through mutations in the protospacer adjacent motifs (Westra et al., 2013). CRISPR interference is a possible explanation for why only half of the assayed 027 ribotype isolates are infected, however, it is unlikely to explain the incomplete infection phenomenon as plaques are still observed, showing a potential partial infection. Similarly, *C. difficile* is known for having at least type I and II restriction-modification systems (Purdy et al., 2002) which work on a similar principle as CRISPR and digest invading DNA and could be targeting incoming phage DNA. However, it also may not explain the phenomenon as partial infection is still required. Although both of these systems could simply be leaky and allow some phage to survive, but not efficient enough for a full infection.

It is interesting to note that four lineages of R20291, divided into three groups of origin responded differently to infection by ΦL14. Strain CRG0825 was obtained for in-house applications from the UK Anaerobe reference unit (ARU), after the Stoke Mandeville outbreak. Strain CRG2021 was obtained from the Brendan Wren laboratory at The London School of Hygiene and Tropical Medicine and was deemed inefficient at plasmid transfer (Monteford et al., 2021). The final strain, CRG3661 was acquired from Novartis, which the company acquired from Brendan Wren's stock. The strains fell into two groups where CRG0825 and its derivative, CRG2897, were

infected by ΦL14, however the other two strains exhibited incomplete infection. One of the differences noted in CRG0825 is its single flagellum (Baban et al., 2013), compared with multiple flagella distributed over the body in isolates of CRG2021 (Martin et al., 2013) and CRG3661. The former also produces more toxins and forms biofilms more readily. These differences result from SNPs and indels present in regulatory regions as a result of continued subculturing of CRG0825 at the ARU prior to in-house acquisition of the strain. This protocol, although no longer practiced for stock production, was deliberately exploited to isolate erythromycin sensitive derivatives of the model strain CD630 (Collery et al., 2017).

As demonstrated with the cwpV protein which is also involved in biofilm formation – upregulation of these host elements can interfere with phage infection. The phenotypic changes in flagella could explain the incomplete infection by phage, as maybe only a small proportion of the isolates are able to be infected due to flagella interfering with the binding process, thus resulting in very small plaques. However, as phage infection is very complex and can be affected by the host in ways which are not well understood, the number of other regulatory mutations could be responsible for this, working behind the scenes of the phenotype to modulate other characteristics of the cell. It is possible that other members of the 027 ribotype in the screening panel had similar mutations acquired naturally or through improper protocols.

The effect may also be phage specific, rather than strain specific as some of the isolated phages were able to fully infect a strain, whilst others exhibited incomplete infection for the same strains; for example L8, NCTC 11205 and CD2801. It is unclear about how this phenomenon would behave in a clinical setting and whether it would have any negative impacts on therapy.

6.3.3 The different host range of ΦL12

Despite the high genetic similarity of ΦL12 phages, their host range is largely similar, but not identical. There are three instances of where only one of the sister phages can infect a strain, and five where the infection was incomplete for one or the other. The phages do contain SNPs, but none which are nonsynonymous in the RBP region responsible for binding described in Chapter 4, therefore both phages should be able to bind to the target strain in an identical manner. However, the discrepancies in the host range and EOP for the same strains, suggest there are other factors in play.

Restriction-modification systems found in the host seem a likely culprit as they target invading DNA and destroy it (Purdy et al., 2002). As the two phages have regions of genetic divergence from each other, it is possible that a specific target that exists in one phage does not exist in the other allowing it to escape the system and infect the given strain.

Another possible explanation is phage defence mechanisms which counteract the bacterial attacks and overcome defence mechanisms to invade the bacterium. Although no such specific proteins have been identified in either phage, it is likely that they would be missed as these are not well characterised in *C. difficile* phages. Any of the numerous hypothetical proteins could serve as a defence mechanism against bacterial attacks especially since phages differ slightly in the number of proteins present which could explain the host range and EOP discrepancies.

6.3.4 ΦL14 – A potentially lytic phage

ΦL14 is the first *C. difficile* phage isolated to date that does not contain an *integrase* gene, all other genomes sequence to date include integrases (Govind, Fralick and Rolfe, 2006; Goh et al., 2007; Horgan, et al., 2010; Hargreaves et al., 2014). The gene is a hallmark of a lysogenic phage, as it allows the phages to integrate in a specified region of the bacterial genome. Often the integrases are also required for excision from the genome (Lauer et al., 2002), when induced by antibiotics or stress.

As *C. difficile* is a spore-forming bacterium, it's believed that this applies an evolutionary pressure that drives its phages to favour the lysogenic lifestyle. This allows the phages to survive when cells are not in a vegetative state and persist within spores until they germinate again, allowing the phages to enter their lytic cycle and continue propagation (Heuler et al., 2021). However, sporulation alone cannot be the single reason for lack of lytic phages as other spore-forming bacteria related bacteria like *C. perfringens* (Ha et al., 2019) or *Bacillus cereus* (J. H. Lee et al., 2013) are infected by strictly lytic phages. *C. difficile* phages are notoriously difficult to isolate due to their host ranges and lytic phages may be extremely rare, therefore it could simply be a "numbers game", and by coincidence this study happened to isolate one.

This could mean significant leaps towards phage therapy as the lysogenic lifestyle comes with many limitations and risks making temperate phages unusable. The

phages can integrate leaving the patient with resistant isolates or can also participate in horizontal gene transfer, where cases of upregulation of toxin production in *C. difficile* have been shown when infected with ΦCD38.2 (Sekulovic et al., 2011).

Attempts at engineering a lytic *C. difficile* phage have been recently reported (Selle et al., 2020). The authors removed the integrase and a cI repressor which is thought to maintain lysogeny. This mutation made the phage ΦCD24-2 lytic *in vitro*, over a 22-hour period; however, when tested *in vivo,* in a mouse model lysogens appeared after 2-4 days post challenge. The mechanism for this is unclear, however, the activity associated with the lysogeny modules could have been compensated for by the prophage genomes present in the tested isolate.

It has been shown thoroughly that ΦL14 is unable to form lysogens *in vitro*, however, this phage would need to be tested *in vivo* to make sure that it behaves like a strictly lytic phage. The study by Selle and colleagues (2020) has severe limitations, as it does not state how the deletion of the lysogeny genes was accomplished or which genes specifically were deleted, especially as there is no official gene identified for the cI repressor protein. Moreover, the genome sequence of ΦCD24-2 has not been published. In the past it has been shown that a gene labelled as the cI-like repressor through database searches, in a *C. difficile* phage (ΦCD2301) was incorrectly labelled as responsible for maintaining lysogeny (Sarah Kerr, PhD, and personal communication, University of Nottingham) and being homologous to the cI gene in lambda phage (Dodd et al., 2005; Ptashne, 2004; Ptashne & Watson, 1967). It may be possible that the true gene responsible for maintaining lysogeny has not been removed and only the integrase gene had to be complemented through presence of prophages in isolates used in the *in vitro* study, thus allowing the mutant phage to lysogenize cells.

A complete list of genes responsible for lysogeny has not been published and the study did not report the phage genome or genes deleted. Therefore, the presence of homologous sequences in ΦL14 cannot be assessed. It is impossible to predict how ΦL14 might behave in a clinical setting, however, as it appears to be naturally lytic, it may have an advantage over the above mentioned engineered phage.

6.3.5 Summary and importance

The attempt to isolate phages from sewage samples led to successful discovery of three new viruses. Two *Siphoviridae* (ΦL12S and ΦL12W) and a Myovirus (ΦL14), all infecting the clinically important ribotype 027 amongst others, exhibiting a relatively wide host range. The genome analysis revealed that the ΦL12 phages are highly similar and only have a slight difference in host range. The ΦL14 genomic analysis revealed a lack of homology to other integrase genes, indicating the phage may not be able to access the temperate lifestyle, making it the first strictly lytic *C. difficile* phage to be reported. The phage is unable to lysogenize several tested strains, further providing evidence for lack of genome integration. This is an extremely important discovery, especially since phages engineered to be lytic were not successful in the past when tested *in vivo*. This promising discovery could potentially change the future of phage therapies used against *C. difficile* infections, overcoming yet another obstacle.

Chapter 7:
General discussion and the future of phage therapy in *Clostridium difficile*

Chapter 7. General discussion and the future of phage therapy in *Clostridium difficile*

7.1 Aims of the study

The objectives of this study aimed to address major obstacles facing *C. difficile* phage therapy, one of which is a narrow host range alongside a limited pool of available phages. The aims were as follows:

- Characterise known phages in terms of their RBPs through the use of a CRISPR/Cas9, a precise gene editing tool, and establish the function of the RBP through molecular genetic techniques and by testing binding.
- Design a semi-synthetic phage with modified RBPs leading to a change in host range and test its efficacy against a panel of relevant *C. difficile* isolates.
- Isolate novel phages from environmental samples through screening against a panel of clinically relevant *C. difficile* strains. Characterise the isolated phages in terms of their particle morphology, genome content and host range.

The work carried out in this study has successfully led to achieving all objectives. Three phages, ΦCD1801, ΦCD2301 and ΦCD2315 were genetically modified using CRISPR/Cas9 vectors to create a deletion of the putative RBP gene for each. Deletion of the gene lead to a loss of the binding ability and resulted in non-infectious particles, showing the RBP involvement in binding to the host. The function was restored upon complementation of the mutation. The phages were analysed for binding ability through assays to demonstrate this impact. TEM of the ΦCD1801 mutant showed that the virion particles retain their structural integrity and can assemble even though they likely lack the tail fibre which functions as the RBP.

After the RBP was established in three phages, in an attempt to change their host range, the genes were swapped between them, however this was unsuccessful and particles were not infectious in all four cases. A mutant phage with two RBP genes was constructed, however its host range remained unaffected by the additional gene, and the phage could only infect its original host. The expanded mutant was only able to bind its original host, although the binding was diminished, likely due to the presence of an unfunctional additional RBP.

This data led to a further investigation of accessory genes surrounding the RBP in *Myoviridae*, ΦCD1801 and ΦCD2301, as the genes were hypothesised to have an impact on the host range and at least one of them was likely required to achieve the desired change. Following a further generation of 12 deletion mutants and corresponding exchange mutants, key genes were characterised for their role in supporting the RBP and relation to host specificity. In particular, the *Hyp12* gene upstream of the RBP, was shown to be required in maintaining structural integrity of the particle, however had no relevance to host range and was not specific to each phage. The downstream *Hyp14* was shown to be specific to the RBP of the phage it originated from and likely plays an accessory role to the RBP. On its own, the Hyp14 encoded protein does not control the host range, however, without it infection of the host is very inefficient. Hyp14 was revealed to contribute to host range and efficiency of infection when combined with the RBP it is encoded with.

The first ever host range change in *C. difficile* infecting phages was documented in the Ex11 mutant. The RBP and *Hyp14* genes of ΦCD1801 were swapped into the ΦCD2301 phage in place of the existing genes creating the semi-synthetic Ex11 hybrid phage. The mutant was able to now infect the CD1801 host, previously infected only by WT ΦCD1801 and not WT ΦCD2301. The mutant was also able mimic the host range of ΦCD1801 and infect some of the other strains previously infected by said phage. Some strains showed immunity to infection, probably due to host resistance mechanisms, such as restriction-modification systems, unrelated to the RBP and the accessory *Hyp14* gene. It is important to make the distinction between the host and binding ranges, as genetically modifying the phage RBP affects its binding range, however host factors which cannot be controlled are also involved in permitting a full infection. The Hyp14 protein likely functions as a chaperone helping the RBP fold correctly, and as an adhesin bridging the binding between the RBP and the host receptor, or in a similar fashion, where it contributes to the host range, but it is not a sole determinant. Therefore, we can now control the binding range of *Myoviridae* phages through exchanging their RBP and *Hyp14* genes, allowing the hybrid phage to infect new hosts, permitting there is no host resistance mechanisms present. This discovery is a huge leap towards broadening the host range and advancing phage therapy against *C. difficile*.

This study also isolated three novel phages, two *Siphoviridae*: ΦL12S and ΦL12W; and a Myovirus: ΦL14. The particles were imaged under TEM allowing their classification and size characterisation, particularly their long tails. The phages displayed a relatively broad host range, all able to infect clinically relevant ribotypes, including the hypervirulent 027 and widespread 078 strains. Genome analysis revealed ΦL12S and W to be very closely related, and lysogenic due to presence of *integrase* genes. No integrase was discovered in the ΦL14 genome, marking it as the first ever potentially lytic phage of *C. difficile* reported.

The ΦL14 phage was investigated for its ability to lysogenise cells *in vitro*, through several methods, two of which are proven to form lysogens almost every single time. After thorough and repeated testing no lysogens were isolated to date, providing further evidence towards the lytic nature of this phage. *In vivo* studies are pending. This discovery addresses the other major obstacle against phage therapy in CDI, as the use of strictly lytic phages is a prerequisite in traditional therapies.

7.2 Future work and potential impact

This study addresses the two major roadblocks holding back phage therapy from being applied in CDI – narrow host range and lack of lytic phages. The tail gene module characterisation revealed the two genes that control the host range of *Myoviridae* and co-function to facilitate the binding. Further advances in characterisation are required to potentially broaden the host range using the same RBP+Hyp14 module with strategic aa substitutions, however, a baseline has now been established, which addresses the problem of narrow host range. Although it has revealed a new roadblock – host resistance mechanisms which largely contribute to the host range. This discovery allows the modulation of the binding range, which is only a factor in host range, and the host itself needs to be considered more in future studies.

Further understanding of phage resistance mechanisms needs to developed, as it was revealed to be a somewhat unexpected obstacle in changing the host range. This was evident when the hybrid phage (Ex11) was unable to fully infect strains that correspond to the encoded binding module. The exact host range was not carried over, as engineering the phage binding is only one of many factors affecting host range, although the only one of significance in other species where phage therapy is

applied. Phage treatments used against other infectious organisms typically have large libraries of easily manipulated phages available as substitutes.

One of the suspected issues is high prophage carriage within *C. difficile*, where superinfection exclusion prevents certain competing phages from entering the bacterium. This mechanism is not well-studied as such widespread prophage competition is uncommon in other organisms. To overcome these, the interaction of prophages should be studied, to find a correlation determining which phages are prevented from entry, and how this is overcome by other phages. Another resistance factor to phage infection is the presence of defences against invading DNA, such as CRISPR or restriction-modification systems. Phages have been reported to encode resistance proteins to these systems, hence they can be circumvented. Further studies are required to discover the exact mechanisms in *C. difficile*, eliminating obstacles to entry and full infection for semi-synthetic phages of the future.

The issue of lysogeny is also being tackled as the newly discovered ΦL14 appears to be the first reported lytic phage of *C. difficile*. To confirm its lack of lysogeny animal studies will need to be utilised in the future. All applications of phages as a therapeutic require them to be strictly lytic, as lysogeny causes many problems including horizontal gene transfer, modulation of bacterial gene expression (like increasing toxin production), and further resistance to phage infection. Discovery of this phage is only the first step; further characterisation will be needed to address its efficacy in a clinical setting, including its ability to tackle the infection. Even if clinically ineffective, ΦL14 is a platform for studying lytic phages. It can be used as an example to elucidate the parameters of a lytic phage, and how these can be applied to existing lysogenic phages, engineering them to remove their lysogenisation activity.

The ΦL14 phage is able to infect a range of clinically relevant strains in the panel tested. Although, owing to the discoveries made in Chapter 6, as it is a Myovirus, its host range can now be changed for that of another phage. These findings can be used to generate a pool of lytic semi-synthetic ΦL14-like phages, with a binding range of ΦCD1801, for example. Combined with the newly published data about the bacterial receptor – slpA, where certain cassettes (12-14 types present) may correspond to certain phage binding modules, the number of required phages in a therapeutic cocktail can be reduced even more.

To further decrease the complexity of the cocktails, exact roles of the RBP, Hyp14 and their interaction needs to be studied. The first logical step in this, is crystallisation of both proteins to reveal the functional domains, and more importantly the amino acids responsible for them, particularly Hyp14, since it is novel and its role is experimentally unconfirmed. This information can be used to establish domain boundaries, and direct focus of mutations towards key amino acids, in order to broaden the specificity of the proteins, while retaining their function. Site-directed mutagenesis can be used to create synthetic versions of the binding domains creating libraries of artificial domains which may confer broader specificity. This eliminates the need for continued isolation of phages as sources of novel RBPs, which is tedious, time consuming and often has limited-success. This can also reveal a correlation between the slpA cassettes and ability of RBPs+Hyp14 to bind them, developing a deeper understanding of the interaction.

This is by no means a complete solution, however the data presented here shows advances made in tackling problems hindering phage therapy applications in CDI. Further advances in the field will be able to use this information to hopefully engineer lytic phages, able to circumvent host defences to entry, with a broad host range able to infect and kill *C. difficile* with high efficiency.

Appendix

8. Appendix

8.1 Amino Acid alignment of the Hyp14 protein from four *Myoviridae* phages



8.2 Full list of *C. difficile* strains used in phage isolation – all PCR ribotype 027

Culture Collection No.	Strain Name	Isolation Location
2021	R20291	Stoke Mandeville, UK
2864	L2	France
2865	L3	France
2866	L6	The Netherlands
2867	L8	Luxembourg
2868	L9	Ireland
2869	L10	Ireland
2870	L11	Germany
2871	L14	Switzerland
2872	L16	Finland
2873	L17	Finland
2874	L19	USA (historical)
2875	L20	USA
2876	L21	Canada
2882	DH1834	Ipswich
2883	DH1466	Northampton
2893	DH1858	Sunderland
2894	DH1751	Bradford
2895	L13	Austria
2896	L12	Germany
2900	R20352	USA
2902	R12087	EU (historical 80s)
3134	L15/L22	Denmark
3150	R24694	Haverfordwest
3153	DH1342	Macclesfield
3154	DH1329	Coventry
3156	R23970	Dublin
3201	L1	Unknown
3202	R24988	Antrim
3271	L4	Belgium
3272	L5	Belgium
3290	R20298	Canada
3331	L7	The Netherlands
None	28-01	Nottinghamshire
	51-01	Nottinghamshire
	92-01	Nottinghamshire

8.3 Full list of *C. difficile* strains used in host range and EOP testing.

CRG Strain name	PCR-ribotype	**CRG Strain name**	PCR-ribotype
L26	001	**DH482**	027
L25	001	**R24694**	027
L28	001	**DH326**	027
L30	001	**DH183**	027
L24	001	**DH1342**	027
L27	001	**DH1329**	027
L29	001	**DH85**	027
001	001	**R23970**	027
R8366	001	**O27**	*027*
NCTC 11205	001	**L1**	*027*
NCTC 11207	001	**R24988**	*027*
NCTC 11209	001	**L4**	027
BI-9	001	**L5**	027
LIV29	001	**R20298**	027
EK9	002	**NCTC 12727**	027
EK2	002	**R20291**	027
EK3	002	**CD196**	027
EK4	002	**210**	027
EK6	002	**CD2801**	027
EK7	002	**CD5101**	027
L33	002	**CD9201**	027
EK8	002	**029**	029
002	002	**7322**	031
EK5	002	**V**	045?
TL178	002	**046**	046
CD1101	002	**053**	053
003	003	**056**	056
3232	010	**1351**	060
7032989	010	**VII**	063
012	012	**070**	070
630WT	012	**075**	075
L31	014	**077**	077
L32	014	**L35**	078
EK14	014	**L34**	078
EK15	014	**EK26**	078
EK16	014	**EK27**	078
014	014	**EK28**	078
TL176	014	**M120**	078
CD2301	014	**EK23**	078
015	015	**EK24**	078
TL174	015	**EK29**	078
CD3601	015	**XVI**	078
CD1902	015	**M120**	078

CD8501	015	**CD1801**	078
CD0301	015	**CD3801**	078
CD4501	015	**CD4401**	078
CD4701	015	**CD8601**	078
CD4202	015	**CD9001**	078
CD5901	015	**CD9301**	078
CD7301	015	**CD9501**	078
CD4502	015	**III A**	080
CD8401	015	**081**	081
EK19	017	**VPI10463**	087
L23	017	**095**	095
L22	017	**EK31**	106
EK20	017	**EK32**	106
EK21	017	**EK34**	106
017	017	**EK36**	106
M68	017	**LIV22**	106
CF5	017	**VI**	127
020	020	**XXIV**	131
023	023	**III C**	251
CD305	023	**1413 REA type M3**	non-toxigenic
L2	027	**1472 REA type M23**	non-toxigenic
L3	027	**1475 REA type T7**	non-toxigenic
L6	027	**NCTC 11204**	not known
L8	027	**P49**	Unknown
L9	027	**P59**	
L10	027	**E4**	
L11	027	**E11**	
L14	027	**E101**	
L16	027	**B9**	
L17	027	**B10**	
L19	027	**B14**	
L20	027	**B32**	
L21	027	**B35**	
DH349	027	**BZ47**	
DH32	027	**CS3**	
DH196	027	**CS8**	
DH1396	027	**V23**	
DH1432	027	**V43 54**	
DH1834	027	**6443 REA type BI 17**	
DH1466	027	**R10432**	106
DH1858	027	**R12801**	106
DH1751	027	**R15347**	106
L13	027	**R22079**	106
L12	027	**R23942**	106

R20291	027	**R25469**	106
DH361	027	**R10459**	106
DH131	027	**R27384**	106
R20352	027	**106**	106
DH478	027	**R108095**	106
R12087	027	**117**	117
L15	027	**126**	126

Bibliography

9. Bibliography

Ackermann, H. W. (1998). Tailed bacteriophages: the order caudovirales. *Advances in Virus Research*, *51*, 135–201.

Akhwale, J. K., Rohde, M., Rohde, C., Bunk, B., Spröer, C., Boga, H. I., Klenk, H. P., & Wittmann, J. (2019). Isolation, characterization and analysis of bacteriophages from the haloalkaline lake Elmenteita, Kenya. *PLoS ONE*, *14*(4). https://doi.org/10.1371/JOURNAL.PONE.0215734

Allegretti, J. R., Kearney, S., Li, N., Bogart, E., Bullock, K., Gerber, G. K., Bry, L., Clish, C. B., Alm, E., & Korzenik, J. R. (2016). Recurrent Clostridium difficile Infection Associates with Distinct Bile Acid and Microbiome Profiles. *Alimentary Pharmacology & Therapeutics*, *43*(11), 1142. https://doi.org/10.1111/APT.13616

Allen, H. K., Looft, T., Bayles, D. O., Humphrey, S., Levine, U. Y., Alt, D., & Stanton, T. B. (2011). Antibiotics in feed induce prophages in swine fecal microbiomes. *MBio*, *2*(6). https://doi.org/10.1128/MBIO.00260-11/SUPPL_FILE/MBO006111199ST4.DOC

Amy, J., Bulach, D., Knight, D., Riley, T., Johanesen, P., & Lyras, D. (2018). Identification of large cryptic plasmids in Clostridioides (Clostridium) difficile. *Plasmid*, *96–97*, 25–38. https://doi.org/10.1016/J.PLASMID.2018.04.001

Ando, H., Lemire, S., Pires, D. P., & Lu, T. K. (2015). Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Systems*. https://doi.org/10.1016/j.cels.2015.08.013

Arisaka, F., Yap, M. L., Kanamaru, S., & Rossmann, M. G. (2016). Molecular assembly and structure of the bacteriophage T4 tail. *Biophysical Reviews*, *8*(4), 385. https://doi.org/10.1007/S12551-016-0230-X

Baban, S. T., Kuehne, S. A., Barketi-Klai, A., Cartman, S. T., Kelly, M. L., Hardie, K. R., Kansau, I., Collignon, A., & Minton, N. P. (2013). The Role of Flagella in Clostridium difficile Pathogenesis: Comparison between a Non-Epidemic and an Epidemic Strain. *PLoS ONE*, *8*(9). https://doi.org/10.1371/JOURNAL.PONE.0073026

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. *The Journal of Infectious Diseases*, *158*(2), 349–354. https://doi.org/10.1093/infdis/158.2.349

Bainum, T. B., Reveles, K. R., Hall, R. G., Cornell, K., & Alvarez, C. A. (2023). Controversies in the Prevention and Treatment of Clostridioides difficile Infection in Adults: A Narrative Review. *Microorganisms 2023, Vol. 11, Page 387*, *11*(2), 387. https://doi.org/10.3390/MICROORGANISMS11020387

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, *315*(5819), 1709–1712. https://doi.org/10.1126/SCIENCE.1138140

Bartlett, J. G. (1994). Clostridium difficile: history of its role as an enteric pathogen and the current state of knowledge about the organism. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *18 Suppl 4*, 265.

Bartual, S. G., Otero, J. M., Garcia-Doval, C., Llamas-Saiz, A. L., Kahn, R., Fox, G. C., & van Raaij, M. J. (2010). Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(47), 20287. https://doi.org/10.1073/PNAS.1011218107

Bebeacua, C., Tremblay, D., Farenc, C., Chapot-Chartier, M.-P., Sadovskaya, I., van Heel, M., Veesler, D., Moineau, S., & Cambillau, C. (2013). Structure, Adsorption to Host, and Infection Mechanism of Virulent Lactococcal Phage p2. *Journal of Virology*, *87*(22), 12302–12312. https://doi.org/10.1128/jvi.02033-13

Belcaid, M., Bergeron, A., & Poisson, G. (2010). Mosaic Graphs and Comparative Genomics in Phage Communities. *Journal of Computational Biology*, *17*(9), 1315–1326. https://doi.org/10.1089/cmb.2010.0108

Berman, L., Carling, T., Fitzgerald, T. N., Bell, R. L., Duffy, A. J., Longo, W. E., & Roberts, K. E. (2008). Defining surgical therapy for pseudomembranous colitis with toxic megacolon. *Journal of Clinical Gastroenterology*, *42*(5), 476–480. https://doi.org/10.1097/MCG.0b013e31804bbe12

Bernstein, H., Bernstein, C., Payne, C. M., Dvorakova, K., & Garewal, H. (2005). Bile acids as carcinogens in human gastrointestinal cancers. *Mutation Research*, *589*(1), 47–65. https://doi.org/10.1016/j.mrrev.2004.08.001

Bertozzi Silva, J., Storms, Z., & Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiology Letters*, *363*(4). https://doi.org/10.1093/FEMSLE/FNW002

Bhaya, D., Davison, M., & Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual Review of Genetics*, *45*, 273–297. https://doi.org/10.1146/ANNUREV-GENET-110410-132430

Bielmann, R., Habann, M., Eugster, M. R., Lurz, R., Calendar, R., Klumpp, J., & Loessner, M. J. (2015). Receptor binding proteins of Listeria monocytogenes bacteriophages A118 and P35 recognize serovar-specific teichoic acids. *Virology*, *477*, 110–118. https://doi.org/10.1016/J.VIROL.2014.12.035

Bondy-Denomy, J., & Davidson, A. R. (2014). When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *Journal of Microbiology 2014 52:3*, *52*(3), 235–242. https://doi.org/10.1007/S12275-014-4083-3

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, *493*(7432), 429–432. https://doi.org/10.1038/NATURE11723

Bondy-Denomy, J., Qian, J., Westra, E. R., Buckling, A., Guttman, D. S., Davidson, A. R., & Maxwell, K. L. (2016). Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal*, *10*(12), 2854. https://doi.org/10.1038/ISMEJ.2016.79

Buckley, A. M., Moura, I. B., Arai, N., Spittal, W., Clark, E., Nishida, Y., Harris, H. C., Bentley, K., Davis, G., Wang, D., Mitra, S., Higashiyama, T., & Wilcox, M. H. (2021). Trehalose-Induced Remodelling of the Human Microbiota Affects Clostridioides difficile Infection Outcome in an In Vitro Colonic Model: A Pilot Study. *Frontiers in Cellular and Infection Microbiology*, *11*, 549. https://doi.org/10.3389/FCIMB.2021.670935/BIBTEX

Buffie, C. G., Bucci, V., Stein, R. R., Mckenney, P. T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., Littmann, E., den Brink, M. R. M., Jenq, R. R., Taur, Y., Sander, C., Cross, J., Toussaint, N. C., Xavier, J. B., & Pamer, E. G. (2014). Precision microbiome restoration of bile acid-mediated resistance to Clostridium difficile. *Nature*, *517*(7533), 205–208. https://doi.org/10.1038/nature13828

Büttner, C. R., Wu, Y., Maxwell, K. L., & Davidson, A. R. (2016). Baseplate assembly of phage Mu: Defining the conserved core components of contractile-tailed phages and related bacterial systems. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(36), 10174–10179. https://doi.org/10.1073/PNAS.1607966113

Cahill, J., Rajaure, M., O'Leary, C., Sloan, J., Marrufo, A., Holt, A., Kulkarni, A., Hernandez, O., & Young, R. (2017). Genetic analysis of the lambda spanins Rz and Rz1: Identification of functional domains. *G3: Genes, Genomes, Genetics*, *7*(2), 741–753. https://doi.org/10.1534/g3.116.037192

Campbell, A. (2003). The future of bacteriophage biology. *Nature Reviews Genetics*, *4*(6), 471–477. https://doi.org/10.1038/nrg1089

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. https://doi.org/10.1021/ACSSYNBIO.9B00075/ASSET/IMAGES/MEDIUM/SB-2019-000754_M001.GIF

Cetinkaya, Y., Falk, P., & Mayhall, C. G. (2000). Vancomycin-Resistant Enterococci. *Clinical Microbiology Reviews*, *13*(4), 686–707. https://doi.org/10.1128/CMR.13.4.686

Chan, B. K., Sistrom, M., Wertz, J. E., Kortright, K. E., Narayan, D., & Turner, P. E. (2016). Phage selection restores antibiotic sensitivity in MDR Pseudomonas aeruginosa. *Scientific Reports 2016 6:1*, *6*(1), 1–8. https://doi.org/10.1038/srep26717

Chaudhary, N., Singh, D., Narayan, C., Samui, B., Mohan, B., Mavuduru, R. S., & Taneja, N. (2021). Complete Genome Sequence of Escherichia Phage 590B, Active against an Extensively Drug-Resistant Uropathogenic Escherichia coli Isolate. *Microbiology Resource Announcements*, *10*(38). https://doi.org/10.1128/MRA.00550-21

Clokie, M., & Kropinski, A. M. (2009). Bacteriophages: Methods and Protocols - Volume 2. In *Life Sciences*. https://doi.org/10.1007/978-1-62703-239-1_1

Clokie, M. R. J., Kropinski, A. M., & Ackermann, H.-W. (2009). Bacteriophages: Volume I. In *Life Sciences*. https://doi.org/10.1007/978-1-62703-239-1_1

Clostridioides difficile ribotyping network (CDRN) report - GOV.UK. (2014, January 1). Public Health England. https://www.gov.uk/government/publications/clostridium-difficile-ribotyping-network-cdrn-report

Colavecchio, A., Cadieux, B., Lo, A., & Goodridge, L. D. (2017). Bacteriophages contribute to the spread of antibiotic resistance genes among foodborne pathogens of the Enterobacteriaceae family - A review. *Frontiers in Microbiology*, *8*(JUN), 1108. https://doi.org/10.3389/FMICB.2017.01108/BIBTEX

Collery, M. M., Kuehne, S. A., McBride, S. M., Kelly, M. L., Monot, M., Cockayne, A., Dupuy, B., & Minton, N. P. (2017). What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of Clostridium difficile strain 630 and derivatives. *Virulence*, *8*(6), 767. https://doi.org/10.1080/21505594.2016.1237333

Collins, J., Danhof, H., & Britton, R. A. (2019). The role of trehalose in the global spread of epidemic Clostridium difficile. *Gut Microbes*, *10*(2), 204–209. https://doi.org/10.1080/19490976.2018.1491266

Collins, J., Robinson, C., Danhof, H., Knetsch, C. W., Van Leeuwen, H. C., Lawley, T. D., Auchtung, J. M., & Britton, R. A. (2018). Dietary trehalose enhances virulence of epidemic Clostridium difficile. *Nature*, *553*(7688), 291. https://doi.org/10.1038/NATURE25178

Committee, H. I. C. P. A. (1995). Recommendations for preventing the spread of vancomycin resistance. *Emerging Infectious Diseases*, *1*(2), 66. https://doi.org/10.3201/eid0102.952010

Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V., & Bikard, D. (2018). A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. *Nature Communications*, *9*(1), 1–10. https://doi.org/10.1038/s41467-018-04209-5

Czepiel, J., Dróżdż, M., Pituch, H., Kuijper, E. J., Perucki, W., Mielimonka, A., Goldman, S., Wultańska, D., Garlicki, A., & Biesiada, G. (2019). Clostridium difficile infection: review. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*, *38*(7). https://doi.org/10.1007/S10096-019-03539-6

de Jonge, P. A., Nobrega, F. L., Brouns, S. J. J., & Dutilh, B. E. (2019). Molecular and Evolutionary Determinants of Bacteriophage Host Range. *Trends in Microbiology*, *27*(1), 51–63. https://doi.org/10.1016/J.TIM.2018.08.006

Deakin, L. J., Clare, S., Fagan, R. P., Dawson, L. F., Pickard, D. J., West, M. R., Wren, B. W., Fairweather, N. F., Dougan, G., & Lawley, T. D. (2012). The Clostridium difficile spo0A gene is a persistence and transmission factor. *Infection and Immunity*, *80*(8), 2704–2711. https://doi.org/10.1128/IAI.00147-12

DeFilipp, Z., Bloom, P. P., Torres Soto, M., Mansour, M. K., Sater, M. R. A., Huntley, M. H., Turbett, S., Chung, R. T., Chen, Y.-B., & Hohmann, E. L. (2019). Drug-Resistant E. coli Bacteremia Transmitted by Fecal Microbiota Transplant. *New England Journal of Medicine*, *381*(21), 2043–2050. https://doi.org/10.1056/NEJMOA1910437/SUPPL_FILE/NEJMOA1910437_DISCLOSURES.PDF

Dei, R. (1989). Observations on phage-typing of Clostridium difficile: preliminary evaluation of a phage panel. *European Journal of Epidemiology*, *5*(3), 351–354.

del Sol, F. G., Penadés, J. R., & Marina, A. (2019). Deciphering the Molecular Mechanism Underpinning Phage Arbitrium Communication Systems. *Molecular Cell*, *74*(1), 72.e3. https://doi.org/10.1016/j.molcel.2019.01.025

Dieterle, M. G., & Young, V. B. (2017). Reducing Recurrence of C. difficile Infection. *Cell*, *169*(3), 375. https://doi.org/10.1016/J.CELL.2017.03.039

Dodd, I. B., Shearwin, K. E., & Egan, J. B. (2005). Revisited gene regulation in bacteriophage λ. *Current Opinion in Genetics & Development*, *15*(2), 145–152. https://doi.org/10.1016/J.GDE.2005.02.001

Dowah, A. (2018). Identification of the receptor binding proteins for Clostridium difficile Bacteriophages. In *undefined*.

Dowah, A. S. A., & Clokie, M. R. J. (2018). Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. *Biophysical Reviews*, *10*(2), 535. https://doi.org/10.1007/S12551-017-0382-3

Dowah, A. S. A., Xia, G., Ali, A. A. K., Thanki, A. M., Shan, J., Wallis, R., & Clokie, M. R. J. (2021). Identification of the receptor-binding protein of Clostridium difficile phage CDHS-1 reveals a new class of receptor-binding domains. *BioRxiv*, 2021.07.05.451159. https://doi.org/10.1101/2021.07.05.451159

Dubberke, E. R., Carling, P., Carrico, R., Donskey, C. J., Loo, V. G., McDonald, L. C., Maragakis, L. L., Sandora, T. J., Weber, D. J., Yokoe, D. S., & Gerding, D. N. (2014). Strategies to prevent Clostridium difficile infections in acute care hospitals: 2014 Update. *Infection Control and Hospital Epidemiology*, *35*(6), 628–645. https://doi.org/10.1086/676023

Dubberke, E. R., Lee, C. H., Orenstein, R., Khanna, S., Hecht, G., & Gerding, D. N. (2018). Results From a Randomized, Placebo-Controlled Clinical Trial of a RBX2660-A Microbiota-Based Drug for the Prevention of Recurrent Clostridium difficile Infection. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *67*(8), 1198–1204. https://doi.org/10.1093/CID/CIY259

Dunne, M., Denyes, J. M., Arndt, H., Loessner, M. J., Leiman, P. G., & Klumpp, J. (2018). Salmonella Phage S16 Tail Fiber Adhesin Features a Rare Polyglycine Rich Domain for Host Recognition. *Structure*, *26*(12), 1573-1582.e4. https://doi.org/10.1016/J.STR.2018.07.017

Dunne, M., Rupf, B., Tala, M., Qabrati, X., Ernst, P., Shen, Y., Sumrall, E., Heeb, L., Plückthun, A., Loessner, M. J., & Kilcher, S. (2019). Reprogramming Bacteriophage Host Range through Structure-Guided Design of Chimeric Receptor Binding Proteins. *Cell Reports*, *29*(5), 1336-1350.e4. https://doi.org/10.1016/j.celrep.2019.09.062

Ecale Zhou, C. L., Malfatti, S., Kimbrel, J., Philipson, C., McNair, K., Hamilton, T., Edwards, R., & Souza, B. (2019). multiPhATE: bioinformatics pipeline for functional annotation of phage isolates. *Bioinformatics*, *35*(21), 4402. https://doi.org/10.1093/BIOINFORMATICS/BTZ258

Eiseman, B., Silen, W., Bascom, G. S., & Kauvar, A. J. (1958). Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. *Surgery*, *44*(5), 854–859.

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*, 48. https://doi.org/10.1186/1743-422X-10-48

Enault, F., Briet, A., Bouteille, L., Roux, S., Sullivan, M. B., & Petit, M. A. (2016). Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *The ISME Journal 2017 11:1*, *11*(1), 237–247. https://doi.org/10.1038/ismej.2016.90

Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., Melamed, S., Leavitt, A., Savidor, A., Albeck, S., Amitai, G., & Sorek, R. (2017). Communication between viruses guides lysis-lysogeny decisions. *Nature*, *541*(7638), 488–493. https://doi.org/10.1038/nature21049

Eyre, D. W., Didelot, X., Buckley, A. M., Freeman, J., Moura, I. B., Crook, D. W., Peto, T. E. A., Walker, A. S., Wilcox, M. H., & Dingle, K. E. (2019). Clostridium difficile trehalose metabolism variants are common and not associated with adverse patient outcomes when variably present in the same lineage. *EBioMedicine*, *43*, 347–355. https://doi.org/10.1016/j.ebiom.2019.04.038

Fagan, R. P., & Fairweather, N. F. (2014). Biogenesis and functions of bacterial S-layers. *Nature Reviews. Microbiology*, *12*(3), 211–222. https://doi.org/10.1038/NRMICRO3213

Feuerstadt, P., Louie, T. J., Lashner, B., Wang, E. E. L., Diao, L., Bryant, J. A., Sims, M., Kraft, C. S., Cohen, S. H., Berenson, C. S., Korman, L. Y., Ford, C. B., Litcofsky, K. D., Lombardo, M.-J., Wortman, J. R., Wu, H., Auniņš, J. G., McChalicher, C. W. J., Winkler, J. A., … von Moltke, L. (2022). SER-109, an Oral Microbiome Therapy for Recurrent Clostridioides difficile Infection. *New England Journal of Medicine*, *386*(3), 220–229. https://doi.org/10.1056/NEJMOA2106516/SUPPL_FILE/NEJMOA2106516_DATA-SHARING.PDF

Fokine, A., Leiman, P. G., Shneider, M. M., Ahvazi, B., Boeshans, K. M., Steven, A. C., Black, L. W., Mesyanzhinov, V. v, & Rossmann, M. G. (2005). Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proceedings of the National Academy of Sciences*, *102*(20), 7163–7168.

Fortier, L.-C., & Moineau, S. (2007). Morphological and Genetic Diversity of Temperate Phages in Clostridium difficile. *Applied and Environmental Microbiology*, *73*(22), 7358–7366. https://doi.org/10.1128/AEM.00582-07

Fortier, L.-C., & Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, *4*(5), 354–365. https://doi.org/10.4161/viru.24498

Foster, T. J. (2005). Immune evasion by staphylococci. *Nature Reviews Microbiology 2005 3:12*, *3*(12), 948–958. https://doi.org/10.1038/nrmicro1289

Garcia, E., Chain, P., Elliott, J. M., Bobrov, A. G., Motin, V. L., Kirillina, O., Lao, V., Calendar, R., & Filippov, A. A. (2008). Molecular characterization of L-413C, a P2-related plague diagnostic bacteriophage. *Virology*, *372*(1), 85–96. https://doi.org/10.1016/J.VIROL.2007.10.032

Garcia-Doval, C., & van Raaij, M. J. (2012). Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. *Proceedings of the National*

Academy of Sciences of the United States of America, *109*(24), 9390–9395. https://doi.org/10.1073/PNAS.1119719109/SUPPL_FILE/PNAS.201119719SI.PDF

Garneau, J. R., Depardieu, F., Fortier, L.-C., Bikard, D., & Monot, M. (2017). PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Scientific Reports*, *7*(1), 8292. https://doi.org/10.1038/s41598-017-07910-5

Ge, P., Scholl, D., Leiman, P. G., Yu, X., Miller, J. F., & Zhou, Z. H. (2015). Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nature Structural & Molecular Biology 2015 22:5*, *22*(5), 377–382. https://doi.org/10.1038/nsmb.2995

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. https://doi.org/10.1016/J.VIROL.2017.02.025

Gelderblom, H. R. (1996). *Structure and Classification of Viruses* (S. Baron, Ed.; 4th ed.). University of Texas Medical Branch at Galveston. http://www.ncbi.nlm.nih.gov/books/NBK8174/

George, W. L., Sutter, V. L., Goldstein, E. J., Ludwig, S. L., & Finegold, S. M. (1978). Aetiology of antimicrobial-agent-associated colitis. *Lancet (London, England)*, *1*(8068), 802–803.

Gerding, D. N., Kelly, C. P., Rahav, G., Lee, C., Dubberke, E. R., Kumar, P. N., Yacyshyn, B., Kao, D., Eves, K., Ellison, M. C., Hanson, M. E., Guris, D., & Dorr, M. B. (2018). Bezlotoxumab for Prevention of Recurrent Clostridium difficile Infection in Patients at Increased Risk for Recurrence. *Clinical Infectious Diseases*, *67*(5), 649–656. https://doi.org/10.1093/CID/CIY171

Geric, B., Carman, R. J., Rupnik, M., Genheimer, C. W., Sambol, S. P., Lyerly, D. M., Gerding, D. N., & Johnson, S. (2006). Binary toxin-producing, large clostridial toxin-negative Clostridium difficile strains are enterotoxic but do not cause disease in hamsters. *The Journal of Infectious Diseases*, *193*(8), 1143–1150. https://doi.org/10.1086/501368

Gianotti, R. J., & Moss, A. C. (2017). Fecal Microbiota Transplantation: From Clostridium difficile to Inflammatory Bowel Disease. *Gastroenterology & Hepatology*, *13*(4), 209. https://www.ncbi.nlm.nih.gov/pubmed/28546791

Girinathan, B. P., Monot, M., Boyle, D., McAllister, K. N., Sorg, J. A., Dupuy, B., & Govind, R. (2017). Effect of tcdR Mutation on Sporulation in the Epidemic Clostridium difficile Strain R20291. *MSphere*, *2*(1). https://doi.org/10.1128/mSphere.00383-16

Goh, S., Ong, P. F., Song, K. P., Rily, T. V., & Chang, B. J. (2007). The complete genome sequence of Clostridium difficile phage φC2 and comparisons to φCD119 and

inducible prophages of CD630. *Microbiology*, *153*(3), 676–685. https://doi.org/10.1099/MIC.0.2006/002436-0

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. https://doi.org/10.1128/AEM.71.2.1079-1083.2005

Goorhuis, A., Bakker, D., Corver, J., Debast, S. B., Harmanus, C., Notermans, D. W., Bergwerff, A. A., Dekker, F. W., & Kuijper, E. J. (2008). Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *47*(9), 1162–1170. https://doi.org/10.1086/592257

Gorbach, S. L. (1992). Drugs for your mother-in-law, not your mother. *Infect Dis Clin Pract*, *1*, 46.

Gordillo Altamirano, F. L., & Barr, J. J. (2019). Phage therapy in the postantibiotic era. In *Clinical Microbiology Reviews* (Vol. 32, Issue 2). American Society for Microbiology. https://doi.org/10.1128/CMR.00066-18

Goulet, A., & Cambillau, C. (2021). Structure and topology prediction of phage adhesion devices using alphafold2: The case of two oenococcus oeni phages. *Microorganisms*, *9*(10). https://doi.org/10.3390/MICROORGANISMS9102151/S1

Goulet, A., & Cambillau, C. (2022). Present Impact of AlphaFold2 Revolution on Structural Biology, and an Illustration With the Structure Prediction of the Bacteriophage J-1 Host Adhesion Device. *Frontiers in Molecular Biosciences*, *9*, 470. https://doi.org/10.3389/FMOLB.2022.907452/BIBTEX

Govind, R., & Dupuy, B. (2012). Secretion of Clostridium difficile Toxins A and B Requires the Holin-like Protein TcdE. *PLoS Pathogens*, *8*(6), e1002727. https://doi.org/10.1371/journal.ppat.1002727

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. https://doi.org/10.1128/JB.188.7.2568-2577.2006

Govind, R., Fralick, J. A., & Rolfe, R. D. (2011). In vivo lysogenization of a Clostridium difficile bacteriophage ΦCD119. *Anaerobe*, *17*(3), 125–129. https://doi.org/10.1016/j.anaerobe.2011.05.012

Govind, R., Vediyappan, G., Rolfe, R. D., Dupuy, B., & Fralick, J. A. (2009). Bacteriophage-mediated toxin gene regulation in Clostridium difficile. *Journal of Virology*, *83*(23), 12037–12045. https://doi.org/10.1128/JVI.01256-09

Guerrero-Ferreira, R. C., Hupfeld, M., Nazarov, S., Taylor, N. M., Shneider, M. M., Obbineni, J. M., Loessner, M. J., Ishikawa, T., Klumpp, J., & Leiman, P. G. (2019). Structure and transformation of bacteriophage A511 baseplate and tail upon

infection of Listeria cells. *The EMBO Journal*, *38*(3), e99455. https://doi.org/10.15252/EMBJ.201899455

Guh, A. Y., Mu, Y., Winston, L. G., Johnston, H., Olson, D., Farley, M. M., Wilson, L. E., Holzbauer, S. M., Phipps, E. C., Dumyati, G. K., Beldavs, Z. G., Kainer, M. A., Karlsson, M., Gerding, D. N., & McDonald, L. C. (2020). Trends in U.S. Burden of Clostridioides difficile Infection and Outcomes. *New England Journal of Medicine*, *382*(14), 1320–1330. https://doi.org/10.1056/NEJMOA1910215/SUPPL_FILE/NEJMOA1910215_DISCLOSURES.PDF

Ha, E., Chun, J., Kim, M., & Ryu, S. (2019). Capsular Polysaccharide Is a Receptor of a Clostridium perfringens Bacteriophage CPS1. *Viruses*, *11*(11). https://doi.org/10.3390/V11111002

Hall, I. C., & O'toole, Elizabeth. (1935). Intestinal Flora in New-born Infants with a Description of New Pathogenic Anaerobe, Bacillus diffieilis. *American Journal of Diseases of Children*, *49*, 390–402.

Hall, J., & Berger, D. (2008). Outcome of colectomy for Clostridium difficile colitis: a plea for early surgical management. *American Journal of Surgery*, *196*(3), 384–388. https://doi.org/10.1016/j.amjsurg.2007.11.017

Hargreaves, K. R., & Clokie, M. R. J. (2014). Clostridium difficile phages: still difficult? *Frontiers in Microbiology*, *5*, 184. https://doi.org/10.3389/fmicb.2014.00184/full

Hargreaves, K. R., Colvin, H. v, Patel, K. v, Clokie, J. J. P., & Clokie, M. R. J. (2013). Genetically diverse Clostridium difficile strains harboring abundant prophages in an estuarine environment. *Applied and Environmental Microbiology*, *79*(20), 6236–6243. https://doi.org/10.1128/AEM.01849-13

Hargreaves, K. R., Flores, C. O., Lawley, T. D., & Clokie, M. R. J. (2014). Abundant and Diverse Clustered Regularly Interspaced Short Palindromic Repeat Spacers in Clostridium difficile Strains and Prophages Target Multiple Phage Types within This Pathogen. *MBio*, *5*(5). https://doi.org/10.1128/MBIO.01045-13

Hashemolhosseini, S., Stierhof, Y. D., Hindennach, I., & Henning, U. (1996). Characterization of the helper proteins for the assembly of tail fibers of coliphages T4 and lambda. *Journal of Bacteriology*, *178*(21), 6258–6265. https://doi.org/10.1128/JB.178.21.6258-6265.1996

Hatfull, G. F. (2008). Bacteriophage genomics. *Current Opinion in Microbiology*, *11*(5), 447–453. https://doi.org/10.1016/j.mib.2008.09.004

Hatfull, G. F., & Hendrix, R. W. (2011). Bacteriophages and their genomes. *Current Opinion in Virology*, *1*(4), 298–303. https://doi.org/10.1016/j.coviro.2011.06.009

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. *The Journal of Infectious Diseases*, *149*(5), 775–780. https://doi.org/10.1093/infdis/149.5.775

Hayes, S., Duhoo, Y., Neve, H., Murphy, J., Noben, J. P., Franz, C. M. A. P., Cambillau, C., Mahony, J., Nauta, A., & van Sinderen, D. (2018). Identification of dual receptor binding protein systems in lactococcal 936 group phages. *Viruses*, *10*(12). https://doi.org/10.3390/v10120668

Hazleton, K. Z., Martin, C. G., Orlicky, D. J., Arnolds, K. L., Nusbacher, N. M., Moreno-Huizar, N., Armstrong, M., Reisdorph, N., & Lozupone, C. A. (2022). Dietary fat promotes antibiotic-induced Clostridioides difficile mortality in mice. *Npj Biofilms and Microbiomes 2022 8:1*, *8*(1), 1–13. https://doi.org/10.1038/s41522-022-00276-1

He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D. J., Martin, M. J., Connor, T. R., Harris, S. R., Fairley, D., Bamford, K. B., D'Arc, S., Brazier, J., Brown, D., Coia, J. E., Douce, G., Gerding, D., Kim, H. J., Koh, T. H., Kato, H., … Lawley, T. D. (2012). Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nature Genetics 2012 45:1*, *45*(1), 109–113. https://doi.org/10.1038/ng.2478

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. https://doi.org/10.1016/j.mimet.2009.05.004

Hecht, G., Pothoulakis, C., LaMont, J. T., & Madara, J. L. (1988). Clostridium difficile toxin A perturbs cytoskeletal structure and tight junction permeability of cultured human intestinal epithelial monolayers. *The Journal of Clinical Investigation*, *82*(5), 1516–1524. https://doi.org/10.1172/JCI113760

Hendrix, R. W., & Duda, R. L. (1992). Bacteriophage λPaPa: Not the Mother of All λ Phages. *Science*, *258*(5085), 1145–1148. https://doi.org/10.1126/SCIENCE.1439823

Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E., & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(5), 2192–2197. https://doi.org/10.1073/pnas.96.5.2192

Herrmann, R., & Wood, W. B. (1981). Assembly of bacteriophage T4 tail fibers: identification and characterization of the nonstructural protein gp57. *Molecular & General Genetics: MGG*, *184*(1), 125–132. https://doi.org/10.1007/BF00271208

Heuler, J., Fortier, L. C., & Sun, X. (2021). Clostridioides difficile phage biology and application. *FEMS Microbiology Reviews*, *45*(5), 1–23. https://doi.org/10.1093/FEMSRE/FUAB012

Hofer, U. (2016). The sting is in the phage's tail. *Nature Reviews Microbiology*, *14*(8), 476. https://doi.org/10.1038/nmicro.2016.97

Horgan, M., O'Sullivan, O., Coffey, A., Fitzgerald, G. F., van Sinderen, D., McAuliffe, O., & Ross, R. P. (2010a). Genome analysis of the Clostridium difficile phage PhiCD6356, a temperate phage of the Siphoviridae family. *Gene*, *462*(1–2), 34–43. https://doi.org/10.1016/j.gene.2010.04.010

Horgan, M., O'Sullivan, O., Coffey, A., Fitzgerald, G. F., van Sinderen, D., McAuliffe, O., & Ross, R. P. (2010b). Genome analysis of the Clostridium difficile phage ΦCD6356, a temperate phage of the Siphoviridae family. *Gene*, *462*(1–2), 34–43. https://doi.org/10.1016/J.GENE.2010.04.010

Horton, R. M. (1995). PCR-mediated recombination and mutagenesis. *Molecular Biotechnology*, *3*(2), 93–99. https://doi.org/10.1007/BF02789105

Hoshiga, F., Yoshizaki, K., Takao, N., Miyanaga, K., & Tanji, Y. (2019). Modification of T2 phage infectivity toward Escherichia coli O157:H7 via using CRISPR/Cas9. *FEMS Microbiology Letters*, *366*(4), 41. https://doi.org/10.1093/FEMSLE/FNZ041

Howell, M. D., Novack, V., Grgurich, P., Soulliard, D., Novack, L., Pencina, M., & Talmor, D. (2010). Iatrogenic gastric acid suppression and the risk of nosocomial Clostridium difficile infection. *Archives of Internal Medicine*, *170*(9), 784–790. https://doi.org/10.1001/archinternmed.2010.89

Hyman, P., & Abedon, S. T. (2010). Bacteriophage host range and bacterial resistance. *Advances in Applied Microbiology*, *70*, 217–248. https://doi.org/10.1016/S0065-2164(10)70007-1

Hyman, P., & van Raaij, M. (2018). Bacteriophage T4 long tail fiber domains. In *Biophysical Reviews* (Vol. 10, Issue 2, pp. 463–471). Springer Verlag. https://doi.org/10.1007/s12551-017-0348-5

ICTV. (2012). Family Podoviridae. In *Virus Taxonomy* (pp. 63–85). http://www.sciencedirect.com/science/article/pii/B9780123846846000033

Janezic, S. (2016). Direct PCR-Ribotyping of Clostridium difficile. *Methods in Molecular Biology (Clifton, N.J.)*, *1476*, 15–21. https://doi.org/10.1007/978-1-4939-6361-4_2

Jangi, S., & Lamont, J. T. (2010). Asymptomatic colonization by Clostridium difficile in infants: implications for disease in later life. *Journal of Pediatric Gastroenterology and Nutrition*, *51*(1), 2–7. https://doi.org/10.1097/MPG.0b013e3181d29767

Joshi, T., & Xu, D. (2007). Quantitative assessment of relationship between sequence similarity and function similarity. *BMC Genomics*, *8*, 222. https://doi.org/10.1186/1471-2164-8-222

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/S41586-021-03819-2

Kang, H. S., McNair, K., Cuevas, D., Bailey, B., Segall, A., & Edwards, R. A. (2017). Prophage genomics reveals patterns in phage genome organization and replication. *BioRxiv*, 114819. https://doi.org/10.1101/114819

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

King, A., Mullish, B. H., Williams, H. R. T., & Aylin, P. (2017). Comparative epidemiology of Clostridium difficile infection: England and the USA. *International Journal for Quality in Health Care*, *29*(6), 785–791. https://doi.org/10.1093/intqhc/mzx120

King, J., & Laemmli, U. K. (1971). Polypeptides of the tail fibres of bacteriophage T4. *Journal of Molecular Biology*, *62*(3), 465–477. https://doi.org/10.1016/0022-2836(71)90148-3

Kirk, J. A., & Fagan, R. P. (2016). Heat shock increases conjugation efficiency in Clostridium difficile. *Anaerobe*, *42*, 1–5. https://doi.org/10.1016/j.anaerobe.2016.06.009

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). https://doi.org/10.1128/JB.00218-18

Kordus, S. L., Thomas, A. K., & Lacy, D. B. (2021). Clostridioides difficile toxins: mechanisms of action and antitoxin therapeutics. *Nature Reviews Microbiology 2021 20:5*, *20*(5), 285–298. https://doi.org/10.1038/s41579-021-00660-2

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. https://doi.org/10.1016/J.CHOM.2019.01.014

Koskella, B., & Brockhurst, M. A. (2014). Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *Fems Microbiology Reviews*, *38*(5), 916. https://doi.org/10.1111/1574-6976.12072

Kuehne, S. A., Cartman, S. T., Heap, J. T., Kelly, M. L., Cockayne, A., & Minton, N. P. (2010). The role of toxin A and toxin B in Clostridium difficile infection. *Nature*, *467*(7316), 711–713. https://doi.org/10.1038/nature09397

Kutter, E., & Sulakvelidze, A. (2005). *Bacteriophages: biology and applications / edited by Elizabeth Kutter, Alexander Sulakvelidze*. Boca Raton, Fla; London: CRC Press.

Kyne, L., Hamel, M. B., Polavaram, R., & Kelly, C. P. (2002). Health care costs and mortality associated with nosocomial diarrhea due to Clostridium difficile. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *34*(3), 346–353. https://doi.org/10.1086/338260

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews. Microbiology*, *8*(5), 317–327. https://doi.org/10.1038/NRMICRO2315

LaFrance, M. E., Farrow, M. A., Chandrasekaran, R., Sheng, J., Rubin, D. H., & Lacy, D. B. (2015). Identification of an epithelial cell receptor responsible for Clostridium difficile TcdB-induced cytotoxicity. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(22), 7073–7078. https://doi.org/10.1073/PNAS.1500791112

Latka, A., Leiman, P. G., Drulis-Kawa, Z., & Briers, Y. (2019). Modeling the Architecture of Depolymerase-Containing Receptor Binding Proteins in Klebsiella Phages. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/FMICB.2019.02649

Latka, A., Lemire, S., Grimon, D., Dams, D., Maciejewska, B., Lu, T., Drulis-Kawa, Z., & Briers, Y. (2021). Engineering the modular receptor-binding proteins of klebsiella phages switches their capsule serotype specificity. *MBio*, *12*(3). https://doi.org/10.1128/MBIO.00455-21/SUPPL_FILE/MBIO.00455-21-ST002.PDF

Lauer, P., Chow, M. Y. N., Loessner, M. J., Portnoy, D. A., & Calendar, R. (2002). Construction, Characterization, and Use of Two Listeria monocytogenes Site-Specific Phage Integration Vectors. *Journal of Bacteriology*, *184*(15), 4177. https://doi.org/10.1128/JB.184.15.4177-4186.2002

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. https://doi.org/10.1016/j.anaerobe.2016.06.008

Le, S., He, X., Tan, Y., Huang, G., Zhang, L., Lux, R., Shi, W., & Hu, F. (2013). Mapping the Tail Fiber as the Receptor Binding Protein Responsible for Differential Host Specificity of Pseudomonas aeruginosa Bacteriophages PaP1 and JG004. *PLoS One*, *8*(7), e68562. https://doi.org/10.1371/journal.pone.0068562

Leclerc, Q. J., Wildfire, J., Gupta, A., Lindsay, J. A., & Knight, G. M. (2022). Growth-Dependent Predation and Generalized Transduction of Antimicrobial Resistance by Bacteriophage. *MSystems*, *7*(2). https://doi.org/10.1128/MSYSTEMS.00135-22

Lee, J. H., Shin, H., Son, B., Heu, S., & Ryu, S. (2013). Characterization and complete genome sequence of a virulent bacteriophage B4 infecting food-borne pathogenic Bacillus cereus. *Archives of Virology*, *158*(10), 2101–2108. https://doi.org/10.1007/S00705-013-1719-2

Lee, Y., Lim, W. I., Bloom, C. I., Moore, S., Chung, E., & Marzella, N. (2017). Bezlotoxumab (Zinplava) for Clostridium Difficile Infection: The First Monoclonal Antibody Approved to Prevent the Recurrence of a Bacterial Infection. *Pharmacy and Therapeutics*, *42*(12), 735. /pmc/articles/PMC5720485/

Lemay, M. L., Tremblay, D. M., & Moineau, S. (2017). Genome Engineering of Virulent Lactococcal Phages Using CRISPR-Cas9. *ACS Synthetic Biology*. https://doi.org/10.1021/acssynbio.6b00388

León, M., & Bastías, R. (2015). Virulence reduction in bacteriophage resistant bacteria. *Frontiers in Microbiology*, *6*(APR), 23. https://doi.org/10.3389/FMICB.2015.00343

Levy, A. R., Szabo, S. M., Lozano-Ortega, G., Lloyd-Smith, E., Leung, V., Lawrence, R., & Romney, M. G. (2015). Incidence and Costs of Clostridium difficile Infections in Canada. *Open Forum Infectious Diseases*, *2*(3), ofv076. https://doi.org/10.1093/ofid/ofv076

Li, K., Wang, G., Andersen, T., Zhou, P., & Pu, W. T. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PloS One*, *9*(8), e105779. https://doi.org/10.1371/journal.pone.0105779

Li, Q., Seys, F. M., Minton, N. P., Yang, J., Jiang, Y., Jiang, W., & Yang, S. (2019). CRISPR-Cas9 D10A nickase-assisted base editing in the solvent producer Clostridium beijerinckii. *Biotechnology and Bioengineering*, *116*(6), 1475–1483. https://doi.org/10.1002/BIT.26949

Li, X., Koç, C., Kühner, P., Stierhof, Y. D., Krismer, B., Enright, M. C., Penadés, J. R., Wolz, C., Stehle, T., Cambillau, C., Peschel, A., & Xia, G. (2016). An essential role for the baseplate protein Gp45 in phage adsorption to Staphylococcus aureus. *Scientific Reports*, *6*. https://doi.org/10.1038/srep26455

Luciano, J. A., & Zuckerbraun, B. S. (2014). Clostridium difficile infection: prevention, treatment, and surgical management. *The Surgical Clinics of North America*, *94*(6), 1335–1349. https://doi.org/10.1016/j.suc.2014.08.006

Luzon-Hidalgo, R., Risso, V. A., Delgado, A., Ibarra-Molero, B., & Sanchez-Ruiz, J. M. (2021). A protocol to study bacteriophage adaptation to new hosts. *STAR Protocols*, *2*(3). https://doi.org/10.1016/J.XPRO.2021.100784

Magill, S. S., O'Leary, E., Janelle, S. J., Thompson, D. L., Dumyati, G., Nadle, J., Wilson, L. E., Kainer, M. A., Lynfield, R., Greissman, S., Ray, S. M., Beldavs, Z., Gross, C., Bamberg, W., Sievers, M., Concannon, C., Buhr, N., Warnke, L., Maloney, M., … Edwards, J. R. (2018). Changes in Prevalence of Health Care-Associated Infections in U.S. Hospitals. *The New England Journal of Medicine*, *379*(18), 1732–1744. https://doi.org/10.1056/NEJMOA1801550

Mahichi, F., Synnott, A. J., Yamamichi, K., Osada, T., & Tanji, Y. (2009). Site-specific recombination of T2 phage using IP008 long tail fiber genes provides a targeted method for expanding host range while retaining lytic activity. *FEMS Microbiology Letters*, *295*(2), 211–217. https://doi.org/10.1111/J.1574-6968.2009.01588.X

Maniloff, J., & Ackermann, H. W. (1998). Taxonomy of bacterial viruses: establishment of tailed virus genera and the order Caudovirales. *Archives of Virology*, *143*(10), 2051–2063.

Marino, M., Puppo, F., Del Bo', C., Vinelli, V., Riso, P., Porrini, M., & Martini, D. (2021). A systematic review of worldwide consumption of ultra-processed foods: Findings and criticisms. *Nutrients*, *13*(8). https://doi.org/10.3390/NU13082778/S1

Markham, A. (2016). Bezlotoxumab: First Global Approval. *Drugs*, *76*(18), 1793–1798. https://doi.org/10.1007/S40265-016-0673-1

Martel, B., & Moineau, S. (2014). CRISPR-Cas: An efficient tool for genome engineering of virulent bacteriophages. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gku628

Martin, M. J., Clare, S., Goulding, D., Faulds-Pain, A., Barquist, L., Browne, H. P., Pettit, L., Dougan, G., Lawley, T. D., & Wren, B. W. (2013). The agr Locus Regulates Virulence and Colonization Genes in Clostridium difficile 027. *Journal of Bacteriology*, *195*(16), 3672. https://doi.org/10.1128/JB.00473-13

Martinez, J. E., Kahana, D. D., Ghuman, S., Wilson, H. P., Wilson, J., Kim, S. C. J., Lagishetty, V., Jacobs, J. P., Sinha-Hikim, A. P., & Friedman, T. C. (2021). Unhealthy Lifestyle and Gut Dysbiosis: A Better Understanding of the Effects of Poor Diet and Nicotine on the Intestinal Microbiome. *Frontiers in Endocrinology*, *12*, 649. https://doi.org/10.3389/FENDO.2021.667066/BIBTEX

Matamouros, S., England, P., & Dupuy, B. (2007). Clostridium difficile toxin expression is inhibited by the novel regulator TcdC. *Molecular Microbiology*, *64*(5), 1274–1288. https://doi.org/10.1111/j.1365-2958.2007.05739.x

Matsui, T., Griniuviené, B., Goldberg, E., Tsugita, A., Tanaka, N., & Arisaka, F. (1997). Isolation and characterization of a molecular chaperone, gp57A, of bacteriophage T4. *Journal of Bacteriology*, *179*(6), 1846. https://doi.org/10.1128/JB.179.6.1846-1851.1997

McGovern, B. H., Ford, C. B., Henn, M. R., Pardi, D. S., Khanna, S., Hohmann, E. L., O'Brien, E. J., Desjardins, C. A., Bernardo, P., Wortman, J. R., Lombardo, M. J., Litcofsky, K. D., Winkler, J. A., McChalicher, C. W. J., Li, S. S., Tomlinson, A. D., Nandakumar, M., Cook, D. N., Pomerantz, R. J., … Trucksis, M. (2021). SER-109, an Investigational Microbiome Drug to Reduce Recurrence After Clostridioides difficile Infection: Lessons Learned From a Phase 2 Trial. *Clinical Infectious Diseases*, *72*(12), 2132–2140. https://doi.org/10.1093/CID/CIAA387

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. https://doi.org/10.1016/j.anaerobe.2010.08.006

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. https://doi.org/10.1128/AEM.02275-12

Michal Stevens, A., Esposito, D. H., Stoney, R. J., Hamer, D. H., Flores-Figueroa, J., Bottieau, E., Connor, B. A., Gkrania-Klotsas, E., Goorhuis, A., Hynes, N. A., Libman, M., Lopez-Velez, R., McCarthy, A. E., von Sonnenburg, F., Schwartz, E., van Genderen, P. J. J., Scott Benson, L., Leung, D. T., Kain, K., … Hajek, J. (2017). Clostridium difficile infection in returning travellers. *Journal of Travel Medicine*, *24*(3). https://doi.org/10.1093/JTM/TAW099

Miller, B. A., Chen, L. F., Sexton, D. J., & Anderson, D. J. (2011). Comparison of the burdens of hospital-onset, healthcare facility-associated Clostridium difficile Infection and of healthcare-associated infection due to methicillin-resistant Staphylococcus aureus in community hospitals. *Infection Control and Hospital Epidemiology*, *32*(4), 387–390. https://doi.org/10.1086/659156

Miniño, A. M., Xu, J., & Kochanek, K. D. (2010). Deaths: preliminary data for 2008. *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, *59*(2), 1–52.

Moldovan, R., Chapman-McQuiston, E., & Wu, X. L. (2007). On Kinetics of Phage Adsorption. *Biophysical Journal*, *93*(1), 303. https://doi.org/10.1529/BIOPHYSJ.106.102962

Momani, L. A. al, Abughanimeh, O., Boonpheng, B., Gabriel, J. G., & Young, M. (2018). Fidaxomicin vs Vancomycin for the Treatment of a First Episode of Clostridium Difficile Infection: A Meta-analysis and Systematic Review. *Cureus*, *10*(6), e2778. https://doi.org/10.7759/cureus.2778

Montag, D., & Henning, U. (1987). An open reading frame in the Escherichia coli bacteriophage lambda genome encodes a protein that functions in assembly of

the long tail fibers of bacteriophage T4. *Journal of Bacteriology*, *169*(12), 5884–5886. https://doi.org/10.1128/JB.169.12.5884-5886.1987

Monteford, J., Bilverstone, T. W., Ingle, P., Philip, S., Kuehne, S. A., & Minton, N. P. (2021). What's a SNP between friends: The lineage of Clostridioides difficile R20291 can effect research outcomes. *Anaerobe*, *71*. https://doi.org/10.1016/J.ANAEROBE.2021.102422

Moon, H. G., Jang, Y.-S., Cho, C., Lee, J., Binkley, R., & Lee, S. Y. (2016). One hundred years of clostridial butanol fermentation. *FEMS Microbiology Letters*, *363*(3). https://doi.org/10.1093/femsle/fnw001

Na, X., & Kelly, C. (2011). Probiotics in Clostridium difficile Infection. *Journal of Clinical Gastroenterology*, *45 Yale Un*, S158. https://doi.org/10.1097/MCG.0b013e31822ec787

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: Acta Pathologica, Microbiologica, et Immunologica Scandinavica*, *99*(4), 321–326.

Nale, J. Y., Chutia, M., Carr, P., Hickenbotham, P. T., & Clokie, M. R. J. (2016). '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*(AUG). https://doi.org/10.3389/fmicb.2016.01383

Nale, J. Y., Redgwell, T. A., Millard, A., & Clokie, M. R. J. (2018). Efficacy of an Optimised Bacteriophage Cocktail to Clear Clostridium difficile in a Batch Fermentation Model. *Antibiotics (Basel, Switzerland)*, *7*(1), 1–15. https://doi.org/10.3390/ANTIBIOTICS7010013

Nale, J. Y., Shan, J., Hickenbotham, P. T., Fawley, W. N., Wilcox, M. H., & Clokie, M. R. J. (2012). Diverse Temperate Bacteriophage Carriage in Clostridium difficile 027 Strains. *PLoS One*, *7*(5), e37263. https://doi.org/10.1371/journal.pone.0037263

Nale, J. Y., Spencer, J., Hargreaves, K. R., Trzepiński, P., Douce, G. R., & Clokie, M. R. J. (2016). Bacteriophage Combinations Significantly Reduce Clostridium difficile Growth In Vitro and Proliferation In Vivo. *Antimicrobial Agents and Chemotherapy*, *60*(2), 968. https://doi.org/10.1128/AAC.01774-15

Nilsson, A. S. (2014). Phage therapy-constraints and possibilities. In *Upsala Journal of Medical Sciences* (Vol. 119, Issue 2, pp. 192–198). Informa Healthcare. https://doi.org/10.3109/03009734.2014.902878

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., Shimatani, Z., & Kondo, A. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, *353*(6305). https://doi.org/10.1126/science.aaf8729

Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., & Goto, S. (2017). ViPTree: the viral proteomic tree server. *Bioinformatics*, *33*(15), 2379–2380. https://doi.org/10.1093/BIOINFORMATICS/BTX157

Nobrega, F. L., Vlot, M., de Jonge, P. A., Dreesens, L. L., Beaumont, H. J. E., Lavigne, R., Dutilh, B. E., & Brouns, S. J. J. (2018). Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology*, *16*(12), 760–773. https://doi.org/10.1038/s41579-018-0070-8

North, O. I., & Davidson, A. R. (2021). Phage Proteins Required for Tail Fiber Assembly Also Bind Specifically to the Surface of Host Bacterial Strains. *Journal of Bacteriology*, *203*(3). https://doi.org/10.1128/JB.00406-20

North, O. I., Sakai, K., Yamashita, E., Nakagawa, A., Iwazaki, T., Büttner, C. R., Takeda, S., & Davidson, A. R. (2019). Phage tail fibre assembly proteins employ a modular structure to drive the correct folding of diverse fibres. *Nature Microbiology*, *4*(10), 1645–1653. https://doi.org/10.1038/S41564-019-0477-7

O'Connor, J. R., Johnson, S., & Gerding, D. N. (2009). Clostridium difficile infection caused by the epidemic BI/NAP1/027 strain. *Gastroenterology*, *136*(6), 1913–1924. https://doi.org/10.1053/J.GASTRO.2009.02.073

Oechslin, F. (2018). Resistance Development to Bacteriophages Occurring during Bacteriophage Therapy. *Viruses*, *10*(7). https://doi.org/10.3390/V10070351

Ofori, E., Ramai, D., Dhawan, M., Mustafa, F., Gasperino, J., & Reddy, M. (2018). Community-acquired Clostridium difficile: epidemiology, ribotype, risk factors, hospital and intensive care unit outcomes, and current and emerging therapies. *The Journal of Hospital Infection*, *99*(4), 436–442. https://doi.org/10.1016/j.jhin.2018.01.015

Peng, Z., Jin, D., Kim, H. B., Stratton, C. W., Wu, B., Tang, Y.-W., & Sun, X. (2017). Update on Antimicrobial Resistance in Clostridium difficile: Resistance Mechanisms and Antimicrobial Susceptibility Testing. *Journal of Clinical Microbiology*, *55*(7), 1998–2008. https://doi.org/10.1128/JCM.02250-16

Pépin, J., Saheb, N., Coulombe, M.-A., Alary, M.-E., Corriveau, M.-P., Authier, S., Leblanc, M., Rivard, G., Bettez, M., Primeau, V., Nguyen, M., Jacob, C.-E., & Lanthier, L. (2005). Emergence of fluoroquinolones as the predominant risk factor for Clostridium difficile-associated diarrhea: a cohort study during an epidemic in Quebec. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *41*(9), 1254–1260. https://doi.org/10.1086/496986

Planche, T., & Wilcox, M. (2011). Reference assays for Clostridium difficile infection: one or two gold standards? *Journal of Clinical Pathology*, *64*(1), 1–5. https://doi.org/10.1136/jcp.2010.080135

Popoff, M. R., Rubin, E. J., Gill, D. M., & Boquet, P. (1988). Actin-specific ADP-ribosyltransferase produced by a Clostridium difficile strain. *Infection and Immunity*, *56*(9), 2299–2306.

Prabhu, V. S., Dubberke, E. R., Dorr, M. B., Elbasha, E., Cossrow, N., Jiang, Y., & Marcella, S. (2018). Cost-effectiveness of Bezlotoxumab Compared With Placebo for the Prevention of Recurrent Clostridium difficile Infection. *Clinical Infectious Diseases*, *66*(3), 355–362. https://doi.org/10.1093/CID/CIX809

Prag, S., & Adams, J. C. (2003). Molecular phylogeny of the kelch-repeat superfamily reveals an expansion of BTB/kelch proteins in animals. *BMC Bioinformatics*, *4*, 42. https://doi.org/10.1186/1471-2105-4-42

Ptashne, M. (2004). *A genetic switch: phage lambda revisited*. https://library.naist.jp/opac/en/volume/40783

Ptashne, M., & Watson, J. D. (1967). ISOLATION OF THE λ PHAGE REPRESSOR. *Proceedings of the National Academy of Sciences of the United States of America*, *57*(2), 306. https://doi.org/10.1073/PNAS.57.2.306

Purdy, D., O'Keeffe, T. A. T., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., & Minton, N. P. (2002a). Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier. *Molecular Microbiology*, *46*(2), 439–452. https://doi.org/10.1046/J.1365-2958.2002.03134.X

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

Quiberoni, A., Guglielmotti, D., Binetti, A., & Reinheimer, J. (2004). Characterization of three Lactobacillus delbrueckii subsp. bulgaricus phages and the physicochemical analysis of phage adsorption. *Journal of Applied Microbiology*, *96*(2), 340–351. https://doi.org/10.1046/J.1365-2672.2003.02147.X

Rakhuba, D. v., Kolomiets, E. I., Szwajcer Dey, E., & Novik, G. I. (2010). Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. In *Polish Journal of Microbiology* (Vol. 59, Issue 3). https://doi.org/10.33073/pjm-2010-023

Ramesh, V., Fralick, J. A., & Rolfe, R. D. (1999). Prevention of Clostridium difficile - induced ileocecitis with Bacteriophage. *Anaerobe*, *5*(2), 69–78. https://doi.org/10.1006/anae.1999.0192

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. https://doi.org/10.1177/1178633718758023

Reynolds, C. B., Emerson, J. E., de la Riva, L., Fagan, R. P., & Fairweather, N. F. (2011). The Clostridium difficile Cell Wall Protein CwpV is Antigenically Variable between Strains, but Exhibits Conserved Aggregation-Promoting Function. *PLoS Pathogens*, *7*(4). https://doi.org/10.1371/JOURNAL.PPAT.1002024

Riede, I. (1987). Receptor specificity of the short tail fibres (gp12) of T-even type Escherichia coli phages. *Molecular & General Genetics: MGG*, *206*(1), 110–115. https://doi.org/10.1007/BF00326544

Riede, I., Drexler, K., Schwarz, H., & Henning, U. (1987). T-even-type bacteriophages use an adhesin for recognition of cellular receptors. *Journal of Molecular Biology*, *194*(1), 23–30. https://doi.org/10.1016/0022-2836(87)90712-1

Rohwer, F., Prangishvili, D., & Lindell, D. (2009). Roles of viruses in the environment. *Environmental Microbiology*, *11*(11), 2771–2774. https://doi.org/10.1111/J.1462-2920.2009.02101.X

Rothenberg, E., Sepúlveda, L. A., Skinner, S. O., Zeng, L., Selvin, P. R., & Golding, I. (2011). Single-virus tracking reveals a spatial receptor-dependent search mechanism. *Biophysical Journal*, *100*(12), 2875–2882. https://doi.org/10.1016/j.bpj.2011.05.014

Royer, A. L. M., Umansky, A. A., Allen, M.-M., Garneau, J. R., Ospina-Bedoya, M., Kirk, J. A., Govoni, G., Fagan, R. P., Soutourina, O., & Fortier, L.-C. (2022). The Clostridioides difficile S-Layer Protein A (SlpA) serves as a general phage receptor. *BioRxiv*, 2022.09.19.508581. https://doi.org/10.1101/2022.09.19.508581

Saginur, R., Hawley, C. R., & Bartlett, J. G. (1980). Colitis Associated with Metronidazole Therapy. *The Journal of Infectious Diseases*, *141*(6), 772–774. https://doi.org/10.1093/infdis/141.6.772

Saund, K., Rao, K., Young, V. B., & Snitkin, E. S. (2020). Genetic Determinants of Trehalose Utilization Are Not Associated With Severe Clostridium difficile Infection Outcome. *Open Forum Infectious Diseases*, *7*(1). https://doi.org/10.1093/OFID/OFZ548

Schilling, T., Dietrich, S., Hoppert, M., & Hertel, R. (2018). A CRISPR-cas9-based toolkit for fast and precise in vivo genetic engineering of Bacillus subtilis phages. *Viruses*. https://doi.org/10.3390/v10050241

Schmidt, C. (2019). Phage therapy's latest makeover. *Nature Biotechnology*, *37*(6), 581–586. https://doi.org/10.1038/s41587-019-0133-z

Schutze, G. E., & Willoughby, R. E. (2013). Clostridium difficile infection in infants and children. *Pediatrics*, *131*(1), 196–200. https://doi.org/10.1542/peds.2012-2992

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., Thomson, N. R., Roberts, A. P., Cerdeño-Tárraga, A. M., Wang, H., Holden, M. T.

G., Wright, A., Churcher, C., Quail, M. A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., … Parkhill, J. (2006). The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature Genetics*, *38*(7), 779–786. https://doi.org/10.1038/NG1830

Seki, H., Shiohara, M., Matsumura, T., Miyagawa, N., Tanaka, M., Komiyama, A., & Kurata, S. (2003). Prevention of antibiotic-associated diarrhea in children by Clostridium butyricum MIYAIRI. *Pediatrics International: Official Journal of the Japan Pediatric Society*, *45*(1), 86–90.

Sekulovic, O., & Fortier, L. C. (2015). Global transcriptional response of Clostridium difficile carrying the CD38 prophage. *Applied and Environmental Microbiology*, *81*(4), 1364–1374. https://doi.org/10.1128/AEM.03656-14

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. https://doi.org/10.1128/AEM.00237-14

Sekulovic, O., Meessen-Pinard, M., & Fortier, L.-C. (2011). Prophage-Stimulated Toxin Production in Clostridium difficile NAP1/027 Lysogens. In *Journal of Bacteriology* (Vol. 193, Issue 11, pp. 2726–2734). https://doi.org/10.1128/JB.00787-10

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.

Selle, K., Fletcher, J. R., Tuson, H., Schmitt, D. S., McMillan, L., Vridhambal, G. S., Rivera, A. J., Montgomery, S. A., Fortier, L. C., Barrangou, R., Theriot, C. M., & Ousterout, D. G. (2020). In Vivo Targeting of Clostridioides difficile Using Phage-Delivered CRISPR-Cas3 Antimicrobials. *MBio*, *11*(2). https://doi.org/10.1128/MBIO.00019-20

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, *577*(7792), 706–710. https://doi.org/10.1038/S41586-019-1923-7

Seys, F. M., Rowe, P., Bolt, E. L., Humphreys, C. M., & Minton, N. P. (2020). *A Gold Standard, CRISPR/Cas9-Based Complementation Strategy Reliant on 24 Nucleotide Bookmark Sequences*. https://doi.org/10.20944/PREPRINTS202003.0321.V1

Shan, J., Patel, K. V, Hickenbotham, P. T., Nale, J. Y., Hargreaves, K. R., & Clokie, M. R. J. (2012). Prophage carriage and diversity within clinically relevant strains of

Clostridium difficile. *Applied and Environmental Microbiology*, *78*(17), 6027–6034. https://doi.org/10.1128/AEM.01311-12

Shetty, N., Wren, M. W. D., & Coen, P. G. (2011). The role of glutamate dehydrogenase for the detection of Clostridium difficile in faecal samples: a meta-analysis. *The Journal of Hospital Infection*, *77*(1), 1–6. https://doi.org/10.1016/j.jhin.2010.07.024

Silpe, J. E., & Bassler, B. L. (2019). A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. *Cell*, *176*(1–2), 280.e13. https://doi.org/10.1016/j.cell.2018.10.059

Sims, M. D., Khanna, S., Feuerstadt, P., Louie, T. J., Kelly, C. R., Huang, E. S., Hohmann, E. L., Wang, E. E. L., Oneto, C., Cohen, S. H., Berenson, C. S., Korman, L., Lee, C., Lashner, B., Kraft, C. S., Ramesh, M., Silverman, M., Pardi, D. S., De, A., … Onwueme, K. (2023). Safety and Tolerability of SER-109 as an Investigational Microbiome Therapeutic in Adults With Recurrent Clostridioides difficile Infection: A Phase 3, Open-Label, Single-Arm Trial. *JAMA Network Open*, *6*(2), e2255758–e2255758. https://doi.org/10.1001/JAMANETWORKOPEN.2022.55758

Sinha, V., Goyal, A., Svenningsen, S. L., Semsey, S., & Krishna, S. (2017). In silico evolution of lysis-lysogeny strategies reproduces observed lysogeny propensities in temperate bacteriophages. *Frontiers in Microbiology*, *8*(JUL), 1386. https://doi.org/10.3389/FMICB.2017.01386/BIBTEX

Smith, H. W., & Huggins, M. B. (1982). Successful treatment of experimental Escherichia coli infections in mice using phage: its general superiority over antibiotics. *Journal of General Microbiology*, *128*(2), 307–318. https://doi.org/10.1099/00221287-128-2-307/CITE/REFWORKS

Solomon, K. (2013). The host immune response to Clostridium difficile infection. In *Therapeutic Advances in Infectious Disease* (Vol. 1, Issue 1, pp. 19–35). https://doi.org/10.1177/2049936112472173

Sonenshein, A. L. (2006). Bacteriophages: how bacterial spores capture and protect phage DNA. *Current Biology: CB*, *16*(1), 14. https://doi.org/10.1016/j.cub.2005.12.007

Sorg, J. A., & Sonenshein, A. L. (2008). Bile salts and glycine as cogerminants for Clostridium difficile spores. *Journal of Bacteriology*, *190*(7), 2505–2512. https://doi.org/10.1128/JB.01765-07

Soutourina, O. A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., Semenova, E., Severinov, K., le Bouguenec, C., Coppée, J. Y., Dupuy, B., & Martin-Verstraete, I. (2013). Genome-Wide Identification of Regulatory RNAs in the Human Pathogen Clostridium difficile. *PLoS Genetics*, *9*(5), 1003493. https://doi.org/10.1371/JOURNAL.PGEN.1003493

Spinelli, S., Bebeacua, C., Orlov, I., Tremblay, D., Klaholz, B. P., Moineau, S., & Cambillau, C. (2014). Cryo-Electron Microscopy Structure of Lactococcal Siphophage 1358 Virion. *Journal of Virology*, *88*(16), 8900–8910. https://doi.org/10.1128/JVI.01040-14/ASSET/6A3BEE41-C20D-47C3-B0CA-9016C5C62669/ASSETS/GRAPHIC/ZJV9990993300008.JPEG

Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T. D., Sebaihia, M., Quail, M. A., Rose, G., Gerding, D. N., Gibert, M., Popoff, M. R., Parkhill, J., Dougan, G., & Wren, B. W. (2009). Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biology*, *10*(9), R102. https://doi.org/10.1186/GB-2009-10-9-R102

Stern, A., & Sorek, R. (2011). The phage-host arms race: shaping the evolution of microbes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *33*(1), 43–51. https://doi.org/10.1002/bies.201000071

Sulakvelidze, A., Alavidze, Z., & Morris, J. G. J. (2001). Bacteriophage Therapy. *Antimicrobial Agents & Chemotherapy*, *45*(3), 649–659. https://doi.org/10.1128/AAC.45.3.649-659.2001

Summer, E. J., Berry, J., Tran, T. A. T., Niu, L., Struck, D. K., & Young, R. (2007). Rz/Rz1 Lysis Gene Equivalents in Phages of Gram-negative Hosts. *Journal of Molecular Biology*, *373*(5), 1098–1112. https://doi.org/10.1016/j.jmb.2007.08.045

Takeuchi, I., Osada, K., Azam, A. H., Asakawa, H., Miyanaga, K., & Tanji, Y. (2016). The presence of two receptor-binding proteins contributes to the wide host range of staphylococcal twort-like phages. *Applied and Environmental Microbiology*, *82*(19), 5763–5774. https://doi.org/10.1128/AEM.01385-16

Tao, L., Tian, S., Zhang, J., Liu, Z., Robinson-McCarthy, L., Miyashita, S. I., Breault, D. T., Gerhard, R., Oottamasathien, S., Whelan, S. P. J., & Dong, M. (2019). Sulfated glycosaminoglycans and low-density lipoprotein receptor contribute to Clostridium difficile toxin A entry into cells. *Nature Microbiology*, *4*(10), 1760–1769. https://doi.org/10.1038/S41564-019-0464-Z

Tao, L., Zhang, J., Meraner, P., Tovaglieri, A., Wu, X., Gerhard, R., Zhang, X., Stallcup, W. B., Miao, J., He, X., Hurdle, J. G., Breault, D. T., Brass, A. L., & Dong, M. (2016). Frizzled proteins are colonic epithelial receptors for C. difficile toxin B. *Nature*, *538*(7625), 350–355. https://doi.org/10.1038/NATURE19799

Tao, P., Wu, X., Tang, W. C., Zhu, J., & Rao, V. (2017). Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. *ACS Synthetic Biology*. https://doi.org/10.1021/acssynbio.7b00179

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

Tétart, F., Desplats, C., & Krisch, H. M. (1998). Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. *Journal of Molecular Biology*, *282*(3), 543–556. https://doi.org/10.1006/JMBI.1998.2047

Thanki, A. M., Taylor-Joyce, G., Dowah, A., Nale, J. Y., Malik, D., & Clokie, M. R. J. (2018). Unravelling the Links between Phage Adsorption and Successful Infection in Clostridium difficile. *Viruses*, *10*(8). https://doi.org/10.3390/V10080411

Theriot, C. M., & Young, V. B. (2015). Interactions Between the Gastrointestinal Microbiome and Clostridium difficile. *Https://Doi.Org/10.1146/Annurev-Micro-091014-104115*, *69*(1), 445–461. https://doi.org/10.1146/ANNUREV-MICRO-091014-104115

Toor, D., Wasson, M. K., Kumar, P., Karthikeyan, G., Kaushik, N. K., Goel, C., Singh, S., Kumar, A., & Prakash, H. (2019). Dysbiosis Disrupts Gut Immune Homeostasis and Promotes Gastric Diseases. *International Journal of Molecular Sciences 2019, Vol. 20, Page 2432*, *20*(10), 2432. https://doi.org/10.3390/IJMS20102432

Trojet, S. N., Caumont-Sarcos, A., Perrody, E., Comeau, A. M., & Krisch, H. M. (2011). The gp38 adhesins of the T4 superfamily: a complex modular determinant of the phage's host specificity. *Genome Biology and Evolution*, *3*(1), 674–686. https://doi.org/10.1093/GBE/EVR059

Uchiyama, J., Takemura, I., Satoh, M., Kato, S. ichiro, Ujihara, T., Akechi, K., Matsuzaki, S., & Daibata, M. (2011). Improved Adsorption of an Enterococcus faecalis Bacteriophage ΦEF24C with a Spontaneous Point Mutation. *PLoS ONE*, *6*(10). https://doi.org/10.1371/JOURNAL.PONE.0026648

Vale, P. F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S., & Gandon, S. (2015). Costs of CRISPR-Cas-mediated resistance in Streptococcus thermophilus. *Proceedings. Biological Sciences*, *282*(1812). https://doi.org/10.1098/RSPB.2015.1270

Vasu, K., & Nagaraja, V. (2013). Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiology and Molecular Biology Reviews: MMBR*, *77*(1), 53. https://doi.org/10.1128/MMBR.00044-12

Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *P & T: A Peer-Reviewed Journal for Formulary Management*, *40*(4), 277. https://www.ncbi.nlm.nih.gov/pubmed/25859123

Volozhantsev, N. v, Oakley, B. B., Morales, C. A., Verevkin, V. v, Bannov, V. A., Krasilnikova, V. M., Popova, A. v, Zhilenkov, E. L., Garrish, J. K., Schegg, K. M., Woolsey, R., Quilici, D. R., Line, J. E., Hiett, K. L., Siragusa, G. R., Svetoch, E. A., & Seal, B. S. (2012). Molecular characterization of podoviral bacteriophages virulent for Clostridium perfringens and their comparison with members of the

Picovirinae. *PloS One*, *7*(5), e38283. https://doi.org/10.1371/journal.pone.0038283

Voona, S., Abdic, H., Montgomery, R., Clarkson, A., Twitchell, H., Hills, T., Briggs, S., Crooks, C., & Monaghan, T. M. (2022). Impact of COVID-19 pandemic on prevalence of Clostridioides difficile infection in a UK tertiary centre. *Anaerobe*, *73*, 102479. https://doi.org/10.1016/J.ANAEROBE.2021.102479

Voth, D. E., & Ballard, J. D. (2005). Clostridium difficile toxins: mechanism of action and role in disease. *Clinical Microbiology Reviews*, *18*(2), 247–263. https://doi.org/10.1128/CMR.18.2.247-263.2005

Vuotto, C., Donelli, G., Buckley, A., & Chilton, C. (2018). Clostridium difficile Biofilm. *Advances in Experimental Medicine and Biology*, *1050*, 97–115. https://doi.org/10.1007/978-3-319-72799-8_7

Webb, B. J., Subramanian, A., Lopansri, B., Goodman, B., Jones, P. B., Ferraro, J., Stenehjem, E., & Brown, S. M. (2020). Antibiotic Exposure and Risk for Hospital-Associated Clostridioides difficile Infection. *Antimicrobial Agents and Chemotherapy*, *64*(4). https://doi.org/10.1128/AAC.02169-19

Weinbauer, M. G., & Suttle, C. A. (1996). Potential Significance of Lysogeny to Bacteriophage Production and Bacterial Mortality in Coastal Waters of the Gulf of Mexico. *Applied and Environmental Microbiology*, *62*(12), 4374. https://doi.org/10.1128/AEM.62.12.4374-4380.1996

Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., & Brouns, S. J. J. (2013). Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLoS Genetics*, *9*(9). https://doi.org/10.1371/JOURNAL.PGEN.1003742

White, H. E., São-José, C., Tavares, P., Zafarani, A., Auzat, I., Plisson, C., Orlova, E. v, & Lhuillier, S. (2007). Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. *The EMBO Journal*, *26*(15), 3720–3728. https://doi.org/10.1038/sj.emboj.7601786

Whittle, M. J., Bilverstone, T. W., van Esveld, R. J., Lücke, A. C., Lister, M. M., Kuehne, S. A., & Minton, N. P. (2022). A Novel Bacteriophage with Broad Host Range against Clostridioides difficile Ribotype 078 Supports SlpA as the Likely Phage Receptor. *Microbiology Spectrum*, *10*(1). https://doi.org/10.1128/spectrum.02295-21

Wilcox, M. H., Gerding, D. N., Poxton, I. R., Kelly, C., Nathan, R., Birch, T., Cornely, O. A., Rahav, G., Bouza, E., Lee, C., Jenkin, G., Jensen, W., Kim, Y. S., Yoshida, J., Gabryelski, L., Pedley, A., Eves, K., Tipping, R., Guris, D., Kartsonis, N., … MODIFY I and MODIFY II Investigators (2017). Bezlotoxumab for Prevention of Recurrent

Clostridium difficile Infection. The New England journal of medicine, 376(4), 305–317. https://doi.org/10.1056/NEJMoa1602615

Wilkinson, L. (2001). Felix d'Herelle and the origins of molecular biology. In *Medical history; Med.Hist.* (Vol. 45, Issue 2, pp. 294–295).

Williams, R., Meader, E., Mayer, M., Narbad, A., Roberts, A. P., & Mullany, P. (2013). Determination of the attP and attB sites of phage CD27 from Clostridium difficile NCTC 12727. *Journal of Medical Microbiology*, *62*(Pt 9), 1439–1443. https://doi.org/10.1099/jmm.0.058651-0

Williams Smith, H., & Huggins, M. B. (1983). Effectiveness of phages in treating experimental Escherichia coli diarrhoea in calves, piglets and lambs. *Journal of General Microbiology*, *129*(8), 2659–2675. https://doi.org/10.1099/00221287-129-8-2659

Wilson, K. H., Kennedy, M. J., & Fekety, F. R. (1982). Use of sodium taurocholate to enhance spore recovery on a medium selective for Clostridium difficile. *Journal of Clinical Microbiology*, *15*(3), 443–446.

Yang, H., Ma, Y., Wang, Y., Yang, H., Shen, W., & Chen, X. (2014). Transcription regulation mechanisms of bacteriophages. *Bioengineered*, *5*(5), 300–304. https://doi.org/10.4161/bioe.32110

Yap, M. L., Mio, K., Ali, S., Minton, A., Kanamaru, S., & Arisaka, F. (2010). Sequential assembly of the wedge of the baseplate of phage T4 in the presence and absence of gp11 as monitored by analytical ultracentrifugation. *Macromolecular Bioscience*, *10*(7), 808–813. https://doi.org/10.1002/MABI.201000042

Yazdi, M., Bouzari, M., & Ghaemi, E. A. (2018). Isolation and Characterization of a Lytic Bacteriophage (vB_PmiS-TH) and Its Application in Combination with Ampicillin against Planktonic and Biofilm Forms of Proteus mirabilis Isolated from Urinary Tract Infection. *Journal of Molecular Microbiology and Biotechnology*, *28*(1), 37–46. https://doi.org/10.1159/000487137

Yehl, K., Lemire, S., Yang, A. C., Ando, H., Mimee, M., Torres, M. D. T., de la Fuente-Nunez, C., & Lu, T. K. (2019). Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. *Cell*. https://doi.org/10.1016/j.cell.2019.09.015

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. https://doi.org/10.1016/J.JBIOTEC.2004.08.003

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*. https://doi.org/10.1016/j.molcel.2017.04.025

Young, I., Wang, I., & Roof, W. D. (2000). Phages will out: strategies of host cell lysis. *Trends in Microbiology*, *8*(3), 120–128.

Yuan, P., Zhang, H., Cai, C., Zhu, S., Zhou, Y., Yang, X., He, R., Li, C., Guo, S., Li, S., Huang, T., Perez-Cordon, G., Feng, H., & Wei, W. (2015). Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for Clostridium difficile toxin B. *Cell Research*, *25*(2), 157–168. https://doi.org/10.1038/CR.2014.169

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. https://doi.org/10.1111/1462-2920.12173

Zhou, C. L. E., Kimbrel, J., Edwards, R., McNair, K., Souza, B. A., & Malfatti, S. (2021). MultiPhATE2: code for functional annotation and comparison of phage genomes. *G3 Genes|Genomes|Genetics*, *11*(5). https://doi.org/10.1093/G3JOURNAL/JKAB074