

Obtaining a lytic bacteriophage suitable for use in phage therapy against *Clostridioides difficile* infection

Sarah Kerr

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

May 2023



Biotechnology and Biological Sciences Research Council



ROTHAMSTED RESEARCH

Declaration

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

Sarah Kerr

May 2023

Covid Statement

This thesis was conducted during the COVID-19 pandemic. During this time, the laboratory was fully closed for a period of three months, after which access was only available during pre-determined shifts. Long-term shift-working had a substantial impact on the work in this study, as the hours available were at times too short to allow for core assays to be undertaken. Necessary laboratory consumables were also scarce due to government COVID-19 testing being prioritised for supply over research purposes. Adjustments were made to allow work to continue under these conditions, however, not all of the issues could be completely mitigated.

In addition to the practical difficulties in undertaking the scientific work, the impacts of the COVID-19 pandemic took a mental toll which persisted well past the end of proscribed lockdowns. A three month extension to the funding from the BBSRC was therefore welcome and greatly appreciated. The additional time, while unfortunately not sufficient to fully complete the originally proposed work, allowed for a greater amount of progress towards the project objectives and consequently a more solid base from which further work can be developed in future.

Abstract

Clostridioides difficile is the leading cause of hospital-acquired and antibiotic-associated diarrhoea in the developed world. More effective countermeasures are required. Phage therapy, based on lytic bacteriophage, is a particularly attractive option. To date, however, all described *C. difficile* phages have proven to be temperate. In the present study, two possible routes to obtaining the obligatory lytic phage on which a therapeutic treatment could be based were explored. In the one, a screening programme was instigated for naturally occurring lytic phage, while in parallel, rational genomic alterations were explored that could make an existing temperate phage lytic.

In the first approach, novel bacteriophages were isolated from sewage treatment plants and characterised. Genome sequencing revealed that one of the isolates (phiL14WAR): (i) lacked an obvious integrase gene; (ii) exhibited a much higher degree of antimicrobial activity against *C. difficile* than temperate phage, and; (iii) failed to convert its indicator strain to a lysogen. These data indicate that phiL14WAR may be a lytic phage, the first of its kind. Further work, however, is required before phiL14WAR can be definitively declared lytic, including determining whether it possesses other lysogeny-related genes.

In the second approach, genes involved in lysogeny were targeted for removal from the existing temperate phage, phiCD2301. CRISPR-mediated deletion of an integrase gene ablated the ability of the lysogen to form plagues, a phenotype entirely restored by its complementation. Further analysis was compromised by being unable to generate mutant phage particles. Clues as to the identity of a potential repressor gene(s) were sought through the identification of putative homologs to the phage ʎ cI repressor and by determining the highest expressed genes during lysogeny. Candidate genes were unsuccessfully targeted for knockout using CRIPSR/Cas9. Of these, structural predictions made using

AlphaFold 2 suggested that the target identified through homology, rather than those identified by RNA seq, was more likely the repressor.

While it has not been possible to engineer a lytic phage within the time frame of this study, the work undertaken has provided a sound base for the future attainment of this goal. In the meantime, the increased antimicrobial action of the novel integrase-deficient phage makes it a promising candidate for further study while at the same time providing the basis of a phage suitable for therapeutic use against *C. difficile* infection.

Acknowledgments

I would like to thank my primary supervisor Professor Nigel Minton for his support and guidance throughout this project and for giving me the opportunity to undertake this research as part of the Synthetic Biology Research Centre (SBRC). I would also like to thank my secondary supervisory team, Dr. Ruth Griffin and Dr. Terry Bilverstone, for their helpful input and for being there when I needed advice on a range of lab and non-lab based topics. Further thanks go to the BBSRC DTP for providing funding for the research, and to the DTP administrative team for being helpful and supportive throughout.

Special thanks go to the other members of the SBRC who have made it a joy to undertake this work in a friendly environment and have provided a wealth of knowledge that has been incredibly helpful across the course of the research. In particular, I would like to thank Joanna Steczynska for working closely with me as we completed adjacent work and providing friendship and support, as well as my extended friendship group including Temi Daike, Hannah Dallas and Jake Yeboah.

Finally, thanks to my family and friends for their never-ending support, and to my husband Matt, without whom I would never have been able to take the career break necessary to complete my lifelong goal of obtaining a PhD.

Abbreviations

%	Percentage
° C	Degrees Celsius
µg	Micrograms
µl	Microlitres
µm	Micrometre
µM	Micromolar
AMR	Antimicrobial resistance
ATc	Anhydrotetracycline
BHI	Brain Heart Infusion
BHIS	Brain Heart Infusion Supplemented
BLAST	Basic Local Alignment Search Tool
BLASTp	Protein BLAST
bp	Base Pair
CC	*Clostridioides difficile* Selective Supplement
CDC	Centre for Disease Control
CDI	*Clostridioides difficile* Infection
CDRN	*Clostridioides difficile* Ribotyping Network
cfu	Colonies Forming Units
Cm	Chloramphenicol
cm	Centimetres
CRG	Clostridia Research Group
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
dH_20	Deionised Water
DNA	Deoxyribonucleic Acid
dsDNA	Double-stranded DNA
EOP	Efficiency of Plating
ESCMID	European Society of Clinical Microbiology and Infectious Diseases
EU	European Union
FAST	Fluorescence-activating and absorption-shifting tag protein
FDA	Food and Drug Administration
FMT	Faecal Microbiota Transplantation
g	Grams
GC	Guanine-cytosine
HiFi	High Fidelity
ICTV	International Committee on Taxonomy of Viruses
IDSA	Infectious Diseases Society of America
IHF	Integrated Host Factors

IPATH	Centre for Innovative Phage Applications and Therapeutics
kbp	Kilo Base Pairs
kg	Kilograms
L	Litres
LB	Lysogeny Broth
M	Mole
MCS	Multiple cloning site
mg	Milligrams
mins	Minutes
ml	Millilitres
mM	Millimolar
mm	Millimetres
MOI	Multiplicity of Infection
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs
ng	Nanograms
NHS	National Health Service
nm	Nanometre
No.	Number
NTM	Non-tuberculosis mycobacterium
OD	Optical Density
ONT	Oxford Nanopore Technologies
ORF	Open Reading Frame
PAE	Predicted alignment error
PaLoc	Pathogenicity Locus
PATRIC	PathoSystems Resource Integration Center
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pfu	Plaque Forming Units
PGAP	Prokaryotic Genome Annotation Pipeline
pLDDT	Predicted local distance difference test
RAST	Rapid Annotation using Subsystem Technology
RBP	Receptor binding protein
RBS	Ribosome Binding Site
RCSB	Research Collaboratory for Structural Bioinformatics
RDP	Ribosomal Database Project
RNA	Ribonucleic Acid

RNAP	RNA polymerase
rRNA	Ribosomal RNA
rpm	Rotations Per Minute
RT	Ribotype
s	Seconds
SBRC	Synthetic Biology Research Centre
SHEA	Society for Healthcare and Epidemiology of America
SNP	Single Nucleotide Polymorphism
SOC	Super Optimal Broth
SOE	Gene splicing by overlap extension PCR
ssDNA	Single-stranded DNA
TAE	Tris-acetate-EDTA
tBLASTx	Translated BLAST search
TEM	Transmission Electron Microscopy
Th	Theophylline
Tm	Thiamphenicol
TMP	Tail tape measure protein
UK	United Kingdom
UKHSA	United Kingdon Health Security Agency
US	United States
WGS	Whole genome sequencing
WHO	World Health Organisation

Figures

Figure 1.1 - Progression of C. difficile disease cycle in the body. Taken from Isidro et al., 2017. 36

Figure 1.2 - Line drawing of different phage morphologies taken from Turner et. al. (2021). Caudoviricetes represents the viruses previously classified as Caudovirales. 44

Figure 1.3 - Structure of bacteriophages from the three main phage families within Caudovirales, taken from Nobrega et al., 2018. 45

Figure 1.4 - Phage DNA entry mechanisms, adapted from Xu and Xiang, (2017). Green = phage tail or peptidoglycan layers in the cell. Yellow = cell membranes. Long red lines indicate the tail tape measure proteins. Question marks indicate currently unidentified mechanisms. Short, coloured bars indicate additional phage-related proteins believed to facilitate DNA entry. 47

Figure 1.5 - Diagram incorporating the two life cycles available to bacteriophages, adapted from Campbell (2003). Obligatory lytic phages follow only the outermost ring, while temperate phages can enter the central loop when required, maintaining an integrated state until conditions (e.g. damage from UV light) cause the prophage to be induced and re-enter the lytic cycle. 52

Figure 1.6 - Simplification of the genetic toggle switch for the lysis-lysogeny decision in E. coli phage λ. Black lines indicate repression, while coloured arrows indicate activation reactions. 54

Figure 3.1 - a) Siphovirus phage isolated from the WAR-PD sample using strain L12; b) Siphovirus phage isolated from the SWI-PD sample using strain L12; c) Myovirus phage isolated from the WAR-PD sample using

strain L14; d) Myovirus phage isolated from the WAR-PD using strain L15 119

Figure 3.2 – Initial genome of phage L14WAR-PD derived using Illumina short sequencing reads. Yellow = terminase genes; blue = capsid structural genes; dark green = tail structural genes; orange = enzymes; pink = regulatory genes; grey = hypothetical proteins; light green = genes of unknown function containing identifiable domain regions. 121

Figure 3.3 – Benchling alignment of Canu-generated L14WAR-PD contigs to a double-length version of the Illumina assembly. Red indicates a mismatch or, in the case of the ends of contigs 3 and 4, indicates an area outside of the aligned genome. 123

Figure 3.4 - Benchling alignment of the doubled version of the L14WAR-PD Illumina genome assembly to the single copy Flye assembly. a) Alignment before amendments. b) Alignment after insertion of the 472bp section that was missing from upstream of the holin gene in the Illumina assembly and correction of SNPs in the second half of the Flye assembly. 124

Figure 3.5 - Revised genome of phage L14WAR-PD following long-read sequencing using ONT MinION sequencing. The majority of the genome remains the same in terms of the open reading frames, however a hypothetical gene upstream of the holin gene has now been identified as two genes instead of one.....125

Figure 3.6 - Genome of phage L15WAR-PD derived using Illumina short sequencing reads. Yellow = terminase genes; blue = capsid structural genes; green = tail structural genes; orange = enzymes; pink = regulatory genes; grey = hypothetical proteins; light green = genes of unknown function containing identifiable domain regions. 128

Figure 3.7 - Genome of phage L12SWI-PD derived using Illumina short sequencing reads. Yellow = terminase genes (large and small subunits); blue = capsid structural genes; green = tail structural genes; orange = enzymes; pink = regulatory genes; grey = hypothetical proteins; light green = genes of unknown function containing identifiable domain regions.....129

Figure 3.8 - Genome of phage L12WAR-PD derived using Illumina short sequencing reads. Yellow = terminase genes; blue = capsid structural genes; green = tail structural genes; orange = enzymes; pink = regulatory genes; grey = hypothetical proteins; light green = genes of unknown function containing identifiable domain regions. 131

Figure 3.9 - Benchling alignment of phages phiL12S and phiL12W. Several areas of difference are visible between the two phages at the nucleotide level such that these two phages can be considered as different strains. 132

Figure 3.10 – Genomic comparison between L14WAR-PD and L15WAR-PD. Three distinct areas of difference are visible with no other inconsistencies in the nucleotide sequence.....132

Figure 3.11 - Benchling comparison of Sanger sequencing results obtained from PCR screening of L15WAR-PD. Top – comparison of the fragments generated from L15WAR-PD to the Illumina genome sequencing results of the same phage. Bottom – comparison of the same fragments to the L14WAR-PD genome. 133

Figure 3.12 - Proteomic tree generated by VIPtree showing alignment of novel phage isolations against a reference database generated from RefSeq, UniProt, Genbank and ViralZone. Red stars mark the positions of the three phages within the tree. 134

Figure 3.13 – VIPtree alignment and dot plot comparisons of phiL14WAR phage with phiCD119 and phiCDHM19. 136

Figure 3.14 – Alignment of phiL14WAR, phiCD119 and phiCDHM19 using the progressiveMauve algorithm. Purple = shared by all three phages; green = shared by phiCD119 and phiCDHM19; red = shared by phiL14WAR and phiCD119; amber = shared by phiL14WAR and phiCDHM19. Red boxes show the main areas of difference. 137

Figure 3.15 – VIPtree alignment and dot plot comparisons of phiL12S and phiL12W phages against phiCD6356. 138

Figure 3.16 – progressiveMauve alignment of phiL12S, phiL12W and phiCD6356. Red boxes are shared elements between phiL12S and phiCD6356; green boxes are areas shared by phiL12W and phiCD6356. 139

Figure 3.17 - Genomic characterisation of the integrase genes from phiL12S and phiL12W using NCBI BLASTp service. Arrows indicate the active catalytic sites that characterise the Int_C_like_3 subfamily.....140

Figure 3.18 - ClustalW alignment of the tyrosine site-specific integrase genes from phiL12S and phiL12W alongside Int_C_like_3 genes from Clostridium perfringens (C.p) and Salmonella enterica enterica (S.e.e). Black boxes indicate the 5 active catalytic residues for this subfamily of the DNA breaking-rejoining superfamily. * indicates conserved residues between all four genes. 141

Figure 3.19 - Example of a plate generated by the spot-on-the-lawn technique. 20µl of neat and 1/10 dilution of phage preparation was spotted onto a quadrant of the bacterial lawn plate at the time of pouring. The presence of phages interfered with bacterial growth, creating defined areas of clearing in the lawn. The degree of infection was represented by

the clarity of the area but this method was deemed too inaccurate to use for estimating the efficiency of plating for each phage.....142

Figure 3.20 - Gel image following PCR of 18 potential lysogen colonies using primers targeting the DNA of phiL14WAR.....145

Figure 3.21 – Left - BHIS-CC agar plate streaked with C. difficile strain L14 and spotted with neat phiL14WAR phage samples. Right - plate streaked with CD2301 and spotted with neat phiCD2301 samples. X indicates the locations at which the neat phage preparation was spotted. Zones of clearing on the L14 plate showed only single colony growth in two spots, whereas the phiCD2301 spots were almost indistinguishable from the surrounding lawn except for tide marks caused by the addition of the phage samples. 146

Figure 3.22 - a) Screening PCR of the potential L14 lysogen colonies using both phage specific primers (top) and C. difficile genome-specific primers (bottom). b) As per (a) but CD2301 lysogen colonies. 147

Figure 3.23 – Mean culture densities (OD_{600nm}) of triplicate L14 cultures infected with varying MOIs of phage phiL14WAR. Blanks were only included in duplicate for the analysis due to unreliable results in the final triplicate which was excluded. Error bars indicate standard deviations of the means.....148

Figure 3.24 – Mean culture densities (OD_{600nm}) of four additional C. difficile strains (in triplicate) which are susceptible to phiL14WAR - CD196 and R12087 (RT 027), L28 (RT 001) and CD8601 (RT 078). Error bars indicate standard deviations of the means. 150

Figure 3.25 – Mean culture densities (OD_{600nm}) of triplicate CD2301 cultures when infected with known temperate phage phiCD2301 over a 23-hour period. Error bars indicate standard deviations from the means. 152

Figure 3.26 - Comparison of local virulence between susceptible strains of C. difficile using data up to the onset of the stationary phase in the negative culture. v_i represents the biomass in the cultures by calculating the area under the infection curve as a proportion of the area under the uninfected control curve for each MOI tested. 153

Figure 3.27 - Comparison of local virulence between susceptible strains of C. difficile using all available data. v_i represents the biomass in the cultures by calculating the area under the infection curve as a proportion of the area under the uninfected control curve for each MOI tested. ..155

Figure 4.1 - Action of serine integrases, taken from Fogg et. al. (2014). a) Thick blue bands indicate the phage DNA double strand while thick grey lines represent the host DNA. The attachment sites in each organism (attP and attB) are complementary. b) Purple and green circles indicate the four subunits of the integrase enzyme which rotate to bring the complementary ends of the phage genome into contact with the corresponding host genome ends. The breaks are then recombined and the phage is integrated into the host genome, forming new sites attR and attL in preparation for later excision.....172

Figure 4.2 - Action of a phage λ tyrosine integrase enzyme, taken from Jayaram et al. (2015). The Holliday junction recombination intermediate is formed by four subunits of the integrase enzyme (circles) which bind to the bacterial and phage DNA at the appropriate attachment sites. The tyrosine catalytic site (Y) cuts a single strand of each DNA helix, allowing recombination before cutting the other strand to complete the integration. 173

Figure 4.3 - Genome of phiCD2301 initially annotated by Whittle et al. (2022) and re-evaluated as part of this study. Phage genomes are modular in arrangement, with structural genes placed together followed by lysis genes. In addition to those labelled above, pink genes are those believed to be related to lysogeny, while grey genes are those identified

as hypothetical proteins using the Artemis Comparison Tool (Carver et al., 2005) and the NCBI BLASTp database. Orange indicates genes with identified domains but currently unknown function, while other colours represent known function enzymes and miscellaneous proteins. 176

Figure 4.4 - Map of pRECas1_phiCD2301int_sg1 plasmid. The homology cassette and sgRNA1 are under the control of the P_{araE} promotor, while the Cas9 is under the control of the riboswitch $P_{fdx\text{-}E}$. catP is included as an antibiotic resistance marker, and oriT and traJ allow for conjugation between the E. coli and C. difficile. pCB102 and ColE1 act as the replicons in C. difficile and E. coli, respectively. 177

Figure 4.5 - Gel electrophoresis of 24 transformant XL1 Blue colonies from each of the three integrase knockout vectors. * denotes the undigested vector run alongside as negative control. The homology cassette and guide RNA modules were approx. 2kbp in size. Undigested pRECas1_MCS vector run alongside as negative control. 178

Figure 4.6 - a) Initial colony PCR of eight sgRNA1 integrase transconjugants. WT denotes the wild-type CD2301L gDNA. Neg is the negative control. The PCR shows an approximate 500bp difference in size between the transconjugants and the wild type cells, indicating a successful knockout of the 561bp integrase gene. b) Repeated colony PCR of single colonies from integrase 1.1 and 1.2 to confirm mutation of these colonies prior to plasmid loss. 179

Figure 4.7 - Benchling comparison of the phiCD2301 genome sequence to contigs 1 and 12. The compared section of contig 1 extends beyond the phage genes to demonstrate continuation into the bacterial genome sequence. 182

Figure 4.8 - Section of phiCD2301 represented in previous Illumina sequencing but confirmed as not present through the Sanger sequencing. 183

Figure 4.9 – Result of whole genome sequencing of integrase-deficient mutant prophages in comparison to the base lysogen. The results confirmed complete removal of the integrase gene. 184

Figure 4.10 - Restriction digest of PCR products amplified from transconjugant colonies. Expected band sizes: correctly complemented mutants - 2.4kbp and 237bp; wild-type phage: 1.4kbp, 1kbp and 237bp. The original Int 1.2.4.1 mutant and wild-type gDNA were run as comparison samples. Note - the sample bands are approximately 500bp larger in size than in the Int 1.2.4.1 mutant, confirming that the gene has been complemented into the transconjugant genomes. 185

Figure 4.11 - Plasmid map of pSJK1. The native phiCD2301 integrase gene was placed under the control of P_{fdx-E} and inserted into the MCS site of pMTL83151 using restriction enzymes NdeI and NheI. 187

Figure 4.12 - Plasmid map of pSJK2. The native phiCD2301 integrase gene was inserted into pRPF185 in place of the gusA reporter gene using restriction enzymes SacI and BamHI.....188

Figure 4.13 - FAST plasmids generated by inserting a FAST gene downstream of the integrase gene using either NheI in pSJK1 or BamHI in pSJK.....189

Figure 4.14 - Level of expression of the integrase gene by plasmid pSJK2-FAST as tested using TFAmber labelling solution. Three different induction levels were tested - 250, 500 and 1000ng/ml of ATc. Data represented are the means of results from biological triplicate cultures, with error bars indicating the standard deviations. 191

Figure 4.15 - Secondary RBS region fused to the front of the FAST gene to replace a potential RBS region excised when using SacI and BamHI enzymes during the construction of pATcFAST-RL.....192

Figure 5.1- Action of the CI protein in repression of the early lytic gene promotors. Taken from Oppenheim et al (2005). RNAP is the RNA polymerase. Orange dumbbells represent CI protein. a) Operative P_R and P_L promotors when no CI is present, with the P_{RM} promotor inactive and the early lytic genes being expressed. b) Initial binding of the CI protein to oR/L 1 and 2, silencing the lytic promotors and activating P_{RM}. c) Hybridisation of the CI proteins between operator regions to form tetramers strengthening lytic gene repression. d) Deactivation of the P_{RM} promotor as a result of CI proteins binding to the third operator region, initiating the self-regulatory mechanism of CI production to maintain a stable state. 202

Figure 5.2 - Plasmid map of vector pRECas1_phiCD2301c1_sg1. All elements of this plasmid are the same as for pRECas1_phiCD2301int_sg1 (see Figure 4.4) except that the homology cassette and guide RNA were specific to the cI gene.....204

Figure 5.3 - Colony PCR results following the transformation of the cI knockout vectors into XL1 Blue chemically competent E. coli cells. Only six positive results were obtained from 24 colonies tested, indicated with boxes. 205

Figure 5.4 – Screening PCR of four transconjugants from the first two conjugations using CRISPR plasmid pRECas1_phiCD2301cI_sg2. ..206

Figure 5.5 – Benchling comparison of sequencing data from phiCD2301c1sg2 colony 2.5 (bottom) and the expected sequence for a successful cI knockout (top). 207

Figure 5.6 - Plaque screen PCR to determine if the plate flood method had collected phages from the plate and whether they were correct mutants. + denotes the CD2301L gDNA positive control and – denotes the dH_20 negative control. 208

Figure 5.7 - Gel images of phage plaque picks from the liquid induction of colonies 2.17 and 2.18. + denotes CD2301L gDNA used as a positive control.....210

Figure 5.8 - Screening PCR of plaques generated from three induced transconjugant colonies obtained using the new sgRNA1 vector. Fifteen plaques were picked from each of the two colony samples for screening. + and – denote the positive and negative controls, respectively.....211

Figure 5.9 - Quality assessment of RNA samples taken from three overnight cultures of CD2301L. Electropherograms show abundance and fragmentation patterns of RNA extracted from each sample. 213

Figure 5.10 – Agilent BioAnalyzer results showing RNA quality and abundance over time in samples taken from a sub-culture of CD2301L. RNA samples were extracted between hours 3 and 7.....214

Figure 5.11 - a) RNA quality assessment of samples taken from three biological replicate cultures of CD2301L six hours after 1:100 sub-culture; b) Samples were compared pre- and post-DNase treatment using PCR primers to generate a 200bp fragment (top) and a 1kbp fragment (bottom) to prove full gDNA removal; c) PCR of the the post-treatment sample following a second DNase treatment. 215

Figure 5.12 - Results of RNA sequencing comparison to the PGAP-annotated reference for CD2301L, undertaken using CLC Genomics Workbench v. 20. Results are mean values taken from three replicate cultures, with error bars indicating the standard deviations. 216

Figure 5.13 - Results of RNA sequencing comparison to the PATRIC-annotated reference genome of CD2301L, undertaken by DeepSeq. Results are mean values taken from three replicate cultures, with error bars indicating the standard deviation. 217

Figure 5.14 – Queried region of the prophage RNA results obtained from analysis using two different annotation pipelines. a) Results obtained from comparison to the multiPhATE2 pathway; b) DeepSeq's comparison to the PATRIC annotation.....218

Figure 5.15 - Amplification of HiFi fragments for construction of vectors targeting the new potential repressor genes. Top) initial PCR; bottom) repeat PCR of previously failed amplifications. 221

Figure 5.16 - a) PCR check of HiFi ligation mixes for p1341KO2 and p1341KO3 using my initial primer design, b) PCR check of HiFi ligation mix of p1341KO2 using backbone fragments donated by Terry Bilverstone.....223

Figure 5.17 - Colony PCR of NEB® Stable transformants. 224

Figure 5.18 - Left) X-ray crystallisation of an inactive phage λ repressor dimer by Stayrook et al. (2008) and available as structure 3bdn on the RCSB PDB. The purple and pink areas represent the two molecules of the dimer bound to the green and yellow operator region. Right) AlphaFold prediction of the structure of the individual λ repressor molecule, visualised in PyMOL v.2.5.4. 226

Figure 5.19 - Top left) AlphaFold 2 predicted structure of the annotated cI-like repressor in phiCD2301 – the rank 1 (model 3) structure is pictured. Top right) pLDDT scores demonstrating the two domains - pLDDT score is a measure of the local inter-atomic distance between residues, with a high score indicating good confidence in the correct prediction. Bottom) PAE scores for each of the generated models demonstrate the two domain structure that is consistent between models. PAE is a measure of the position error between two residues, with low scores being preferred in this case. Blue indicates low-scoring residues, with red indicating high scores. 227

Figure 5.20 – Left) Ramachandran plot showing the torsion angles of amino acids within the cI-like repressor protein of phiCD2301. Right) Protein structures associated with each area of the plot, taken from Kumar and Arya (2008). Green indicates those results falling in highly preferred conformations, brown are preferred, while red are questionable observations. 228

Figure 5.21 - Alignment of the phage λ repressor molecule structure (green) to the generated cI-like repressor of phiCD2301 (cyan) using PyMOL v.2.5.4. The N-terminal α-helices of the queried molecule show a degree of similarity to the N-terminal of the λ repressor, while the C-terminal is of a very different composition. 229

Figure 5.22 - AlphaFold 2 prediction for the structure of gene 1341. The main α-helix is present in both variations with the models ranked 2 and 4 also containing a smaller α-helix at the C-terminal end. The uncertainty is reflected in the pLDDT scores for the C-terminal, as well as in the PAE graphs where a slight bulging in the blue C-terminal domain area reflects the difference between the models. 231

Figure 5.23 - Ramachandran plot of gene 1341. The single helix model, while represented in 3 out of 5 of the predictions, has a higher rate of questionable results (28%) than the dual helix model (2%) which is only shared by 2 of the predictions. 232

Figure 5.24 - Phobius transmembrane and signal peptide prediction of gene 1341 with the cI-like repressor provided for comparison. Gene 1341 shows a defined transmembrane region in the N-terminal α-helix with both cytoplasmic and non-cytoplasmic regions, while the annotated cI-like gene is fully cytoplasmic. 233

Tables

Table 1.1 - UKHSA definitions of CDI onset. 34

Table 1.2 - Treatment guidelines for ESCMID categories of CDI, summarised from Debast et al. (2014). 38

Table 1.3 - Summary of stages of bacteriophage lytic cycle..... 49

Table 2.1 - Complete list of RT 027 C. difficile strains used in the isolation of novel bacteriophages from environmental sewage samples. 67

Table 2.2 - List of bacterial strains used for gene knockout of cI and int phage genes..... 68

Table 2.3 - List of plasmids used in this study to date 69
Table 2.4 - Oligonucleotide primers used in this study to date 72

Table 2.5 - Q5-mediated PCR protocol used for vector component amplification 86

Table 2.6 - Q5-mediated colony PCR protocol 87

Table 2.7 - DreamTaq-mediated colony PCR protocol 87

Table 2.8 - Antibiotic supplements and inducing chemicals used in this study..... 93

Table 2.9 - Concentrations of inducers tested during FAST assay of integrase expression vectors..... 98

Table 2.10 - Enzymes used to digest DNA for creation of CRISPR vectors 101

Table 3.1- Plaque assay results for isolation of novel bacteriophages from environmental sewage.....118

Table 3.2 - Mean dimensions of phages measured during visualisation using TEM. 120

Table 3.3 - Heat map of efficiency of plating for novel phages tested against sensitive strains of C. difficile 144

Table 3.4 - Virulence indices of C. difficile strains infected with phiL14WAR 156

Table 3.5 - 16s rRNA sequence match results for strain JS50 158

Table 3.6 - 16s rRNA sequence match results for strain L12 158

Table 3.7 - 16s rRNA sequence match results for strain CD1801 159

Table 4.1- C. difficile strains selected for initial conjugation tests, along with their associated isolated phage size 175

Table 5.1 - Translation initiation rates (au) upstream of the queried hypothetical gene identified in the CD2301L PATRIC annotation.....219

Table of Contents

Declaration 2
Covid Statement 3
Abstract 4
Acknowledgments 6
Abbreviations....7
Figures 10
Tables....22
Table of Contents 24
1. Introduction 31
1.1 Clostridioides difficile 32
1.1.1 The organism 32
1.1.2 Ribotypes 33
1.1.3 Disease epidemiology 34
1.1.4 Disease cycle: 35
1.1.5 Current treatments:.... 37
1.1.6 Summary 42
1.2 Bacteriophages.... 42
1.2.1 Taxonomy.... 43
1.2.2 Bacteriophage structure: 44
1.2.3 Life cycles.... 48
1.2.4 Summary 56
1.3 Phage therapy 56
1.3.1 Advantages and disadvantages of phage therapy.... 57
1.3.2 Factors affecting phage therapy 59
1.3.3 Existing cases using phage therapy 61

1.3.4 Phage therapy for CDI 62
1.3.5 Genetically engineered phages for phage therapy against CDI 63
1.3.6 Summary 65
1.4 Project aims 65
2 Materials and Methods 67
2.1 Bacterial strains and plasmids 67
2.2 Oligonucleotide primers 72
2.3 General molecular biology techniques 84
2.3.1 Genomic DNA (gDNA) extraction 84
2.3.2 Oligonucleotide design 85
2.3.3 Amplification of DNA using polymerase chain reaction (PCR) 86
2.3.4 Agarose gel electrophoresis 87
2.3.5 Extraction of DNA from agarose gel 88
2.3.6 PCR clean-up 88
2.3.7 Preparation of plasmid DNA from E. coli strains 89
2.3.8 Restriction enzyme digests 89
2.3.9 RNA extraction 89
2.4 General microbiology techniques 92
2.4.1 Media and buffers 92
2.4.2 Antibiotic supplements and inducing chemicals 93
2.4.3 Growth conditions 94
2.4.4 Strain storage and revival 94
2.4.5 Preparation of E. coli chemically competent cells 95
2.4.6 Transformation of E. coli chemically competent cells 95
2.4.7 Conjugation of plasmid DNA into C. difficile strains 96
2.4.8 Removal of conjugated plasmid DNA 97

2.4.9 Fluorescence-activating and absorption-shifting tag protein (FAST) assay 98
2.5 Vector creation 99
2.5.1 Primer design 99
2.5.2 Vector construction 100
2.6 Submission of samples for Whole Genome Sequencing (WGS) 103
2.6.1 Illumina DNA sequencing 103
2.6.2 Oxford Nanopore Technologies (ONT) DNA sequencing 103
2.6.3 RNA sequencing 104
2.7 Bioinformatics tools 105
2.7.1 Data analysis and presentation 105
2.7.2 Sequencing of PCR fragments and plasmid DNA 105
2.7.3 Assembly and annotation of sequencing data 105
2.7.4 Phylogenomic analysis 106
2.7.5 Protein analysis 107
2.8 Phage-related assays 107
2.8.1 Sewage enrichment 107
2.8.2 Plaque assay for determination of presence of phage in sample 108
2.8.3 Titre increase of new phage samples 109
2.8.4 Transmission electron microscopy (TEM) of new phage samples 110
2.8.5 Determining the host range of bacteriophages 110
2.8.6 Efficiency of Plating (EOP) 111
2.8.7 Lysogen formation for long term storage 111
2.8.8 Induction of phages from lysogens 112
2.8.9 Phage antimicrobial action assay 112

3 Isolation of Novel Bacteriophages Infecting the Clinically Relevant Ribotype 027 Strains 115

3.1 Introduction.....115

3.2 Results 117

3.2.1 Isolation of novel phages.....117

3.2.2 Transmission electron microscopy (TEM) 119

3.2.3 Whole genome sequencing of novel phages.....120

3.2.4 Genetic comparisons between isolated phages 131

3.2.5 Phylogenetic analysis.....133

3.2.6 Siphovirus integrase analyses.....140

3.2.7 Host range testing and efficiency of plating 142

3.2.8 Testing for the ability to form lysogens in phiL14WAR 145

3.2.9 Efficiency of action on a live culture.....148

3.2.10 16S testing of strain L12.....157

3.3 Discussion 160

3.3.1 Siphovirus phages phiJS50S and phiJS50W 161

3.3.2 Myovirus phiL14WAR.....164

3.4 Chapter Summary 168

4 CRISPR knockout of the *int* gene from bacteriophage phiCD2301 171

4.1 Introduction.....171

4.2 Results 175

4.2.1 Selection of initial target bacteriophage.....175

4.2.2 Vector creation and conjugation 177

4.2.3 PCR screening for mutation 179

4.2.4 Induction of mutants for phage release 180

4.2.5 Whole genome sequencing of mutant colonies 180

4.2.6 Complementation of genes back into the knockout strains 184

4.2.7 Assessment of plasmid-borne integrase expression 186

4.3 Discussion 193

4.4 Chapter summary 198

5 CRISPR knockout of the repressor gene from phage phiCD2301 201

5.1 Introduction.....201

5.2 Results 203

5.2.1 Vector creation 203

5.2.2 Trouble-shooting the isolation of cI mutants 205

5.2.3 Identification of the lysogeny control mechanism through RNA sequencing.....212

5.2.4 Targeting the alternative repressor genes 221

5.2.5 Protein folding of the two repressor candidates.....225

5.3 Discussion 233

5.4 Chapter Summary 239

6 Conclusions and further work.....242

6.1 Conclusions 242

6.2 Future work 247

6.2.1 Integrase-deficient mutant 247

6.2.2 Repressor removal 248

6.2.3 phiL14WAR 249

6.2.4 Longer term goals.....250

7 Appendix.....252

7.1 *C. difficile* strains used for host range testing 252

7.2 Phage genome annotations.....257

7.3 Phage infection curve data 263

7.3.1 Average $OD600_{nm}$ of L14 cultures infected with phiL14WAR 263

7.3.2 Average $OD600_{nm}$ of CD196 cultures infected with phiL14WAR 264

7.3.3 Average $OD600_{nm}$ of R12087 cultures infected with phiL14WAR 265

7.3.4 Average $OD600_{nm}$ of L28 cultures infected with phiL14WAR 266

7.3.5 Average $OD600_{nm}$ of CD8601 cultures infected with phiL14WAR 267

7.3.6 Average $OD600_{nm}$ of CD2301 cultures infected with phiCD2301 268

7.3.7 Virulence index results 269

7.4 Whole genome sequencing of CD2301L – non-genomic contig BLAST results (top ten results only) 270

7.4.1 BLAST results 270

7.4.2 Phylogenetic tree (generated by NCBI BLAST).....270

7.5 RNA sequencing results from DeepSeq 271

7.5.1 Quality assessment 271

7.5.2 Mapping summary.....271

7.5.3 Transcripts per million (TPM) of phage-related genes only 272

8 References 274

Chapter 1

Introduction

1. Introduction

Since the mid-twentieth century, antibiotics have formed the core treatment option for combatting bacterial infections in Western medicine. However, multi-drug resistant bacteria were identified as early as the 1950s, and the subsequent over-use of antibiotics has driven an increase in the number and diversity of these resistant organisms which now pose an urgent threat to patients (Bronzwaer et al., 2002). The Centre for Disease Control (CDC) and World Health Organisation (WHO) have now acknowledged the emergency (CDC, 2013; WHO, 2015), and have advised that certain types of antibiotics should be held in reserve to tackle the most difficult infections – so-called "drugs-of-last-resort" such as Vancomycin (WHO, 2019). Despite this, resistance to these drugs is still being reported, with a recent study by Murray et al. (2022) attributing 1.27 million deaths to untreatable antimicrobial-resistant (AMR) pathogens globally in 2019. It is estimated that deaths from AMR infections will rise to around 10 million by 2050 (Department for Health and Social Care, 2019). Alternatives to antibiotics are therefore required to effectively treat infections without aggravating the current state of antimicrobial resistance.

Hospitals use antibiotics to treat a wide variety of infections on a daily basis. Poor hygiene standards, however, can result in these bacteria maintaining a presence in the hospital environment, with the risk of being transferred to other patients (Department for Health, 2009). One such species of bacteria is *Clostridioides difficile* (formerly *Clostridium difficile*), which shows some antibiotic resistance (Andino-Molina et al., 2019; Spigaglia, 2016) but more importantly takes advantage of antibiotic-induced dysbiosis to colonise the gut of infected patients. It has been found to be responsible for more than 15-25% of antibiotic-associated diarrhoea and 95% of cases of pseudomembranous colitis (Eckert et al., 2015). Since the very antibiotics prescribed for the treatment of *C. difficile* leave the patient susceptible to further infection, novel non-antibiotic therapies are required.

One potential treatment to explore is bacteriophage therapy. Bacteriophages were first discovered in the 1910s and were quickly identified as having therapeutic potential (Keen, 2015). The advent of antibiotics, however, resulted in phage therapy being disregarded for much of the following century (Clokie et al., 2011; Lin et al., 2017; Weinbauer, 2004). The renewed demand for novel therapeutics has raised interest in these organisms due to their natural predation on bacteria, and extensive research is underway from a variety of sources to explore their biology and phage-host interactions. The aim of this project was to continue to investigate methods through which phage therapy can be developed to be an effective treatment for *C. difficile* infection (CDI).

1.1 Clostridioides difficile

1.1.1 The organism

C. difficile is an obligately anaerobic, Gram-positive, spore-forming bacillus (Andino-Molina et al., 2019; Androga et al., 2018). First described by Hall and O'Toole in 1935, it was given the name *Bacillus difficilis* (Alcalá Hernández et al., 2017; Lawson et al., 2016) before later being moved into the *Clostridium* genus. A subsequent review of this genus in 2016 resulted in a further name change to *Clostridioides difficile*, distinguishing it from other clostridial species without causing too much confusion in healthcare settings with a whole-scale change to the name (Lawson et al., 2016).

As a strict anaerobe, *C. difficile* vegetative cells are limited in the locations in which they can survive, most notably the colons of animals. However, their ability to form aero-tolerant spores when subjected to a range of environmental stimuli allows them to survive in a dormant state outside of these anaerobic environments, facilitating transmission (Zhu et al., 2018). As a result, *C. difficile* is a ubiquitous species found in soil, freshwater, seawater and animal manure (Al Saif and Brazier, 1996; Candel-Pérez et al., 2019; Pasquale et al., 2012). Animal reservoirs also

exist, including in food animals such as pigs and cows (Dubberke et al., 2011). In humans, it can be found naturally in the gut microbiota of up to 15% of adults, and up to 30% of newborns (Lees et al., 2016).

Once the spores are re-introduced to an anaerobic environment, they can germinate in the presence of bile salts such as taurocholate and infect the new host (Sorg and Sonenshein, 2009). In some cases, there is little effect on the host, such as in newborns where associated symptoms seem rare (Lees et al., 2016). Many strains of *C. difficile* are, however, pathogenic, and it is now known to be the leading cause of nosocomial diarrhoea in many developed countries (Alcalá Hernández et al., 2017; Androga et al., 2018; Brandt et al., 1999).

1.1.2 Ribotypes

C. difficile strains are taxonomically classified through the use of PCR ribotyping. This technique targets polymorphisms in highly conserved areas of the ribosomal RNA, as the low evolutionary selective pressure on ribosomal RNA enables identification of close relatives (Kashyap et al., 2014). Pathogenic strains therefore tend to be clustered into the same ribotype (RT) groups. However, as ribotypying is not directly associated with factors affecting toxin production or control, there is debate as to whether or not the ribotype of a strain is a direct indicator of potential disease severity. For example, RT 027 strains have previously been considered hypervirulent due to their impact during the epidemic in America and the EU in the early 2000s (Alcalá Hernández et al., 2017), however other studies have shown that, out of the context of an epidemic, not all RT 027 strains show similar high levels of toxin production (Walk et al., 2012). Therefore, in clinical settings, it has been recommended that a range of patient and strain-related factors should be used to assess potential for severe disease, as opposed to ribotype alone (Barbut and Rupnik, 2012)

1.1.3 Disease epidemiology

CDI has been identified as largely associated with the treatment of patients with antibiotics, although it is also a serious concern for individuals with existing colonic conditions (Androga et al., 2015; Candel-Pérez et al., 2019). In their 2021/2022 financial year counts, the UK Health Security Agency (UKHSA) defined CDI in the following manner:

Table 1.1 - UKHSA definitions of CDI onset.

Onset	**Association**	**Positive sample provided**
Hospital	Healthcare	≥2 days following hospital admission
Community	Healthcare	<2 days after admission and patient was admitted in the 4 weeks prior to the episode.
	Indeterminate	<2 days after admission and patient was admitted in the previous 12 weeks prior to the episode but not the last 4.
	Community	<2 days after admission and patient was not admitted in the previous 12 weeks prior to the episode.

Due to their extensive use of antibiotics, CDI has mostly been considered an issue in hospitals, with elderly individuals and those undergoing chemotherapy or other prolonged hospital therapies being particularly susceptible (Alcalá Hernández et al., 2017; Badger et al., 2012a; Candel-Pérez et al., 2019). Broad-spectrum antibiotics are known to cause dysbiosis in the gut microbiota, and as a result, niches are left open in which pathogenic strains of *C. difficile* can flourish (Hunt and Ballard, 2013; Lewis and Pamer, 2017). Hospitals, therefore, present a concentrated population of vulnerable individuals.

Although greater rigour in the UK National Health Service (NHS) approach to CDI has aided in gaining some control over infection rates

(Department for Health, 2009), healthcare-associated CDI does not account for all cases, and there has been a documented increase in the incidence of community-associated cases (CDC, 2005). Data around community-associated CDI has only been recorded by the UKHSA since 2017, but it shows a marked increase in cases between 2017/18 (29.4% of cases) and 2019/2020 (40.1% of cases). This could, in part, be due to better recording of these incidences, as the UKHSA changed their guidelines in 2017 to better differentiate between cases attributable to hospital care and those with no relevant healthcare contact; however, this data still shows that at least 39.6% of all CDI cases recorded in the 2021/2022 period occurred in people with no relevant prior hospital treatment, despite the increased hygiene requirements necessitated by the pandemic. Transmission could be via individuals previously admitted to the hospital but who are asymptomatic (Badger et al., 2012a), or could be a reflection of increased spore availability outside of hospital settings. As such, research is ongoing into possible transmission from other reservoirs such as domestic animals and livestock (Andino-Molina et al., 2019; Janezic et al., 2012; Knetsch et al., 2018; Ünlü, 2020; Weese, 2010).

One factor of major concern regarding CDI is the prevalence of recurrent infections (Alcalá Hernández et al., 2017; Androga et al., 2018). Recurrence after initial infection can occur in up to 50% of patients (Aslam et al., 2005), with up to 45% chance of additional recurrences (Johnson, 2009). Many of these recurrences are attributable to a different strain of *C. difficile* than the original infection, implying that incomplete eradication of the initial strain is not the sole cause of recurrence (Androga et al., 2015; Hunt and Ballard, 2013).

1.1.4 Disease cycle:

Understanding the disease cycle is a key component of determining treatment and prevention strategies for any infection. This is especially important for antibiotic-resistant infections such as *C. difficile* where prevention is likely to be easier than treatment. If infection does occur

then the symptoms can range from no visible symptoms or just diarrhoea in mild cases to fever, abdominal pain, vomiting, pseudomembranous colitis, toxic megacolon, and potentially death in the most severe cases (Alcalá Hernández et al., 2017; Badger et al., 2012a; Poxton et al., 2001). It is therefore important to not only understand how transmission occurs but also the mechanisms of disease action in order to identify possible diagnostic tools and intervention methods.

Transmission often occurs through transfer of the aerobically-tolerant spores from their deposition site to vulnerable patients through poor hygiene (Badger et al., 2012a) (Figure 1.1). The spores travel through the stomach, surviving the acidic conditions until they reach the colonic lumen and intersperse with the gut microbiome (Isidro et al., 2017). Exposure to bile salts such as cholate, taurocholate and glycocholate subsequently promotes germination of the spores into vegetative cells (Hunt and Ballard, 2013; Lawler et al., 2020; Sorg and Sonenshein, 2009), and these vegetative cells can occupy niches in the colon left vacant by antibiotic-induced dysbiosis (Lewis and Pamer, 2017).

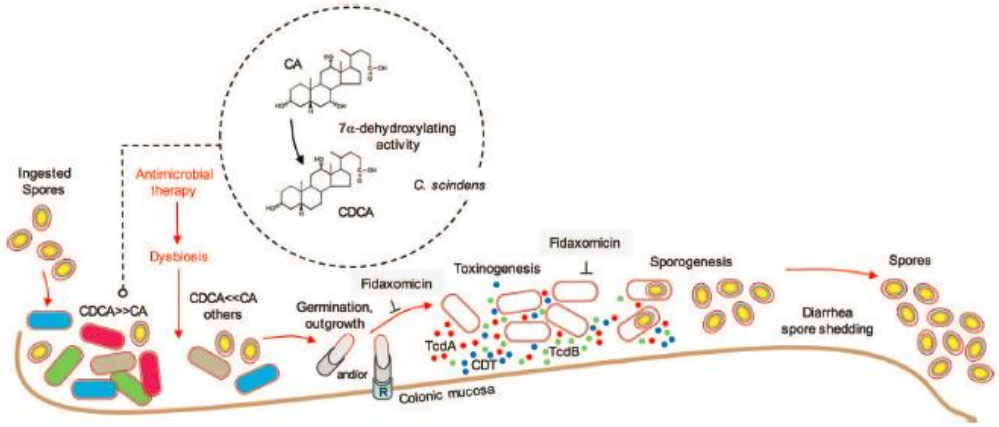


Figure 1.1 - Progression of C. difficile disease cycle in the body. Taken from Isidro et al., 2017.

Once inside the colon, vegetative cells generate a cocktail of up to three species of toxin, depending on the strain of *C. difficile* involved – Toxin A (TcdA), Toxin B (TcdB) and binary toxin (CDT) (Candel-Pérez et al.,

2019; Gerding et al., 2014). All three of these toxins are known to act on the colonic epithelial cells, impairing the cells' ability to maintain their actin cytoskeleton, either through inhibition of Rho GTPase enzymes or through changing the actin monomers themselves to prevent polymerisation (Aktories et al., 2018; Gülke et al., 2001; Hunt and Ballard, 2013). This leads to a reduction in the structural integrity of the cells, and eventually apoptosis and cell death (Badger et al., 2012a; Poxton et al., 2001). This damage to the intestinal barrier then results in the identifiable symptoms of CDI such as inflammation, diarrhoea and pseudomembranous colitis through neutrophil recruitment and migration of granulocytic cells into the intestinal lumen (Alcalá Hernández et al., 2017; Badger et al., 2012a; Poxton et al., 2001).

Alongside toxinogenesis, the vegetative *C. difficile* cells also undergo sporulation, with each cell creating a new spore across the course of about 8 hours (Isidro et al., 2017). The diarrhoeal symptoms elicited by toxinogenesis allow some of these new spores to be shed back into the environment to await transmission to a new host, while others may remain in the original host to await further germination and recurrence of the infection (Isidro et al., 2017).

1.1.5 Current treatments:

In 2014, the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) defined five categories for CDI severity to help health practitioners use the various treatments to best effect, given that it is not possible to recommend a single antibiotic treatment for this type of infection (Debast et al., 2014) (Table 1).

Table 1.2 - Treatment guidelines for ESCMID categories of CDI, summarised from Debast et al. (2014).

Category	Recommended treatments
Initial CDI: non-severe disease	Discontinuation of oral antibiotic treatment; oral metronidazole if symptoms do not resolve
Severe CDI	Oral vancomycin administration; total abdominal colectomy in cases of colonic perforation or systemic inflammation despite maximal antibiotic treatment
First recurrence or (risk of) recurrent CDI	Treatment with oral vancomycin or oral fidaxomicin
Multiple recurrent CDI	Treatment with oral vancomycin or oral fidaxomicin; alternatively, consider faecal transplantation
Treatment of CDI when oral administration is not possible	Combination of parenteral metronidazole with intracolonic or nasogastric vancomycin

The first and simplest treatment, especially where it is suspected that infection is antibiotic-induced, is to remove the antibiotic in question (Debast et al., 2014; Teasley et al., 1983). However, where this does not resolve the symptoms, there are a small number of currently available antibiotics which form the core of treatment options for CDI. These antibiotic treatments are used in nearly all of the above cases and focus on three main candidates: metronidazole, vancomycin and fidaxomicin (Debast et al., 2014; National Institute for Health and Care and Excellence (NICE), 2021).

Metronidazole, derived from a compound isolated from *Streptomyces. sp* and first used in the 1950s (Leitsch, 2019), is an efficient, low-cost first-line treatment for a range of different conditions, including CDI (Alauzet et al., 2019; Debast et al., 2014). However, it does not appear to be as

effective when used against some strains such as the PCR ribotype 027 strains (Freeman et al., 2007) and there is concern that resistance genes to metronidazole may be growing in prevalence (Alauzet et al., 2019; Moura et al., 2014).

Vancomycin is a glycopeptide antibiotic (National Institute for Health and Care, 2019) which is considered a “drug of last resort” against multi-drug resistant, Gram-positive bacterial infections (Boneca and Chiosis, 2005), and the UKHSA have previously recommended it as the preferred treatment for CDI in all but mild-to-moderate cases (Public Health England, 2019). This echoes the 2017/18 update to the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA) clinical practice guidelines, although the more recent IDSA/SHEA guidelines also state a preference for treating mild-to-moderate CDI cases with vancomycin as well unless access to it or its alternative, fidaxomicin, is limited (McDonald et al., 2018).

Fidaxomicin is a more recent addition to the treatment regime for CDI. Approved for use by the Food and Drug Administration (FDA) in 2011, fidaxomicin is a narrow-spectrum macrocylic antibiotic (Bartsch et al., 2013). It is as effective as vancomycin, and in some cases exceeds vancomycin’s treatment ability when patients are treated with concomitant antibiotics for concurrent infections (Mullane et al., 2011). Its narrow-spectrum action has a reduced impact on the commensal gut flora associated with the conversion of primary bile salts to secondary acids which inhibit *C. difficile* spore germination and provide resistance against future spore colonisation and germination (Buckley et al., 2022; Lawler et al., 2020). Fidaxomicin is therefore associated with a reduced risk of recurrence in comparison to broad-spectrum antibiotic treatments (Tannock et al., 2010). However, the price of fidaxomicin has largely relegated this drug to an alternative to vancomycin rather than being the predominant treatment for CDI (Alcalá Hernández et al., 2017; Bartsch et al., 2013). There is now also evidence of fidaxomicin-resistance genes

within some clinical isolates of *C. difficile* in Germany, implying that its effectiveness may be at risk (Schwanbeck et al., 2019).

A non-antibiotic therapy is also available - faecal microbiota transplantation (FMT). Recurrence of CDI is often a result of the antibiotic treatments maintaining dysbiosis in the commensal gut flora which allows for further infection by *C. difficile* spores (van Nood et al., 2013). However, the gut microbiota is largely stable in its composition (Faith et al., 2013) and transplantation of stools from a healthy individual can be effective in helping to rectify the imbalance in the commensal flora (Alcalá Hernández et al., 2017; Debast et al., 2014; Woodworth et al., 2019). In particular, a recent review of FMT treatments identified up to 91% success rates for the treatment of patients with recurrent CDI when FMT is delivered an average of 1.5 times via lower GI endoscopy following antibiotic administration (84% from single treatment - Baunwall et al., 2020). Capsule and enema delivery were also investigated, with capsule delivery providing decent results at a lower cost, but lower GI endoscopy was consistently identified as the most effective means of using FMT (Baunwall et al., 2020).

However, the idea of using human stool samples from healthy patients as a treatment option is understandably one that engenders a degree of distaste in patients, and so uptake of this therapeutic option has been limited (van Nood et al., 2013). Two recent FDA safety alerts also demonstrated potential dangers of untargeted FMT caused by improper screening of donor samples which resulted in several individuals being accidentally infected with strains of antibiotic-resistant *E.coli*, sadly with fatalities (Buckley et al., 2022). Even in cases where patients survived, this transfer raises the additional danger of introducing antibiotic resistance genes to native microbes, perpetuating the issue of AMR. Furthermore, uncertainties surround the long-term effects that transplanting microbiomes between individuals would have, even should the treatment for CDI be successful - for example, changes to the ratio of *Firmicutes* to *Bacteroides* organisms has been implicated in susceptibility

to obesity (Tseng and Wu, 2019) while the effects of adding external bacteria on immune system homeostasis have been associated with autoimmune issues such as irritable bowel syndrome (Buckley et al., 2022).

In an attempt to address some of these concerns, research has been conducted into the potential of more targeted microbiome therapies, with two candidates having recently achieved FDA approval. The first to be authorised was RBX2660 (authorised under the name Rebyota) which was developed by Rebiotix in conjunction with Ferring Pharmaceuticals Inc and consists of a suspension of human donor stool samples for which both the donor and the sample have been extensively screened for any undesirable contents (Dubberke et al., 2018; Orenstein et al., 2016; U.S. Food and Drug Administration, 2022). Unfortunately, this treatment is administered via enema, and so requires a clinician for use (Orenstein et al., 2016). Very recently, however, an oral therapy produced by Seres Therapeutics – test name SER-109 but to be manufactured under the name Vowst – has passed the authorisation board as well (U.S. Food and Drug Administration, 2023). This therapy aims to specifically replenish the populations of *Firmicutes* bacteria in the gut microbiome, a species which converts primary bile acids to secondary acids that inhibit *C. difficile* spore germination, and are therefore integral to host colonisation resistance (Feuerstadt et al., 2022). As an oral therapy, it is likely to be favoured over the enema-based Rebyota due to ease of use; however, it is unknown whether the more targeted nature of the constituent microbes in Vowst will be more or less effective than the broader spectrum of its competitor.

One final therapy that has recently been approved is monoclonal antibodies (MAbs). MAbs target the symptom-inducing toxins secreted by pathogens, neutralising them and reducing the amount of damage the toxins can inflict (Humphreys and Wilcox, 2014). They traditionally have a long serum half-life and low off-target toxicity, making them good candidates for preventative treatments against recurrent CDI (Wang et

al., 2008). Currently, the only MAb which is FDA-approved for *C. difficile* is Bezlotoxumab, which gained authorisation in 2017 in the US and subsequently in Europe in 2018 (Förster et al., 2018; Mullard, 2016). This MAb targets sections of the N-terminal of TcdB and blocks the binding of the toxin to the colonic epithelial cells (Orth et al., 2014) resulting in an approx. 40% decrease in recurrence of CDI after a 10mg/kg of body weight intravenous dose (Navalkele and Chopra, 2018). However, while it is approved for use in conjunction with other treatment methods, such as antibiotics, it is not authorised for use as a sole clinical cure (Navalkele and Chopra, 2018), and its high cost ($3117-$4675 depending on location - Chen et al., 2021) is still somewhat prohibitive with respect to regular use.

Despite established antibiotic and non-antibiotic treatments being available, CDI is still an issue for many healthcare providers. Reducing antibiotic efficacy with recurrence means that CDI continues to affect many people with established infections. Coupled with the increase in community-associated infections, investigations into new treatments for CDI are an urgent requirement.

1.1.6 Summary

CDI has been a known and consistent drain on the NHS for decades, with nearly 47,000 cases registered in the UK between May 2021 and May 2022 alone (UKHSA, 2022). However, the increasing incidence of resistance to standard medications for CDI (Alauzet et al., 2019; Androga et al., 2018) as well as the reduced efficacy of existing treatments as recurrences increase (van Nood et al., 2013) highlights the necessity of investigating alternative potential treatments.

1.2 Bacteriophages

Bacteriophages are viruses which infect bacteria and they represent one of the most diverse groups of organisms on the planet (Dowah and Clokie, 2018) with approximately ten times the number of phages as

bacteria in sea- and freshwater environments alone (Clokie et al., 2011). Discovered in 1915 by Frederick Twort and again by Felix D'Herelle in 1917 (Keen, 2015), they have an extensive history in molecular and medical biology (Dowah and Clokie, 2018), including being used to treat bacterial infections in the post-war, pre-antibiotic era (Summers, 2012). Modern research is continuously demonstrating their importance in all environments, particularly their part in maintaining the human gut microbiota (De Sordi et al., 2019; Scanlan, 2017). Now, with the advent of antimicrobial resistance and the declining effectiveness of antibiotics, these organisms are being re-evaluated for their therapeutic use.

1.2.1 Taxonomy

Phage taxonomy is a constantly evolving area. The existing naming convention for bacteriophages does not follow the traditional Linnaean taxonomic system but is rather based on the morphology of the phages as described by Bradley in 1967 and updated by Ackermann and Eisenstark in 1974. In this system, 95% of known phages fall into the *Caudovirales* order of tailed phages (Prevelige and Cortines, 2018). This order is split into three families based on a shared tail structure – *Siphoviridae* (long-tailed phages), *Myoviridae* (medium-length contractile tails), and *Podoviridae* (short tails). This classification was accepted by the International Committee on the Taxonomy of Viruses (ICTV) by 1984 (Turner et al., 2021b). The other 5% of known phages were spread between seven other families, also based on morphology - *Tubulavirales*, *Microviridae*, *Tectiviridae*, *Corticoviridae*, *Plasmaviridae*, *Leviviridae*, and *Cystoviridae* (Ackermann, 2011; Turner et al., 2021b) (Figure 1.2).

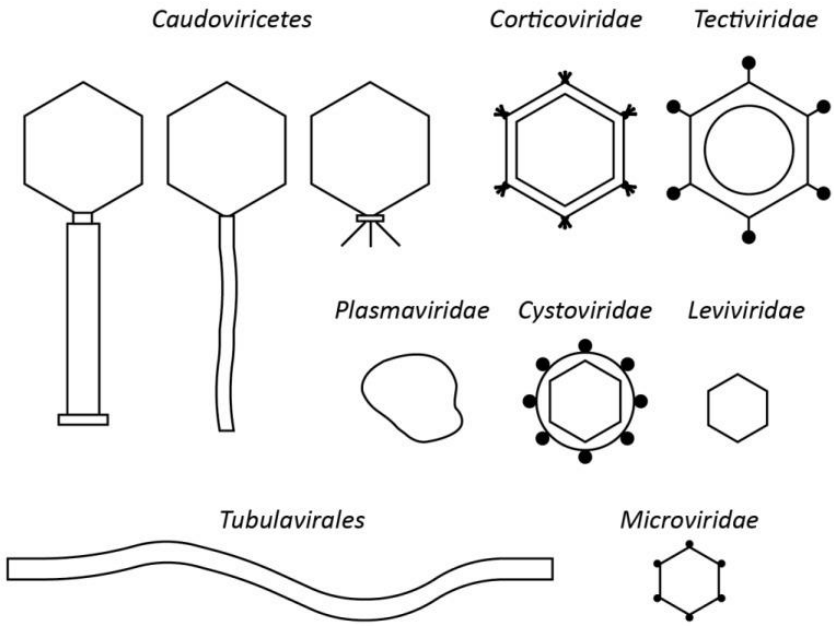


Figure 1.2 - Line drawing of different phage morphologies taken from Turner et. al. (2021). Caudoviricetes represents the viruses previously classified as Caudovirales.

There has long been debate over how (or whether) to convert the taxonomy of phages to the Linnaean binomial form including a genus and species, but the advent of genomic sequencing technologies has allowed the true relationships between phages to start to be untangled. New orders, families and sub-families of phages are regularly being introduced by the ICTV (Adriaenssens et al., 2020), and phages can now also be classified by their genetic structure, i.e. double-stranded (ds) or single-stranded (ss) DNA or RNA. However, despite recommendations from the ICTV and other scientists (Siddell et al., 2020; Turner et al., 2021b), the old classification has so far persisted. It will likely take some time to move past the morphological definition to a Latinised version. Given that these changes are currently in their infancy, I will reference the older, broader taxonomic groupings of phages for this review, with a focus on the order *Caudovirales*.

1.2.2 Bacteriophage structure:

As previously mentioned, the most studied group of phages is the order *Caudovirales*, which constitutes approximately 95% of known phages.

This order is split into three families based on the length of the tail and all contain dsDNA (Dowah and Clokie, 2018) (Figure 1.3).

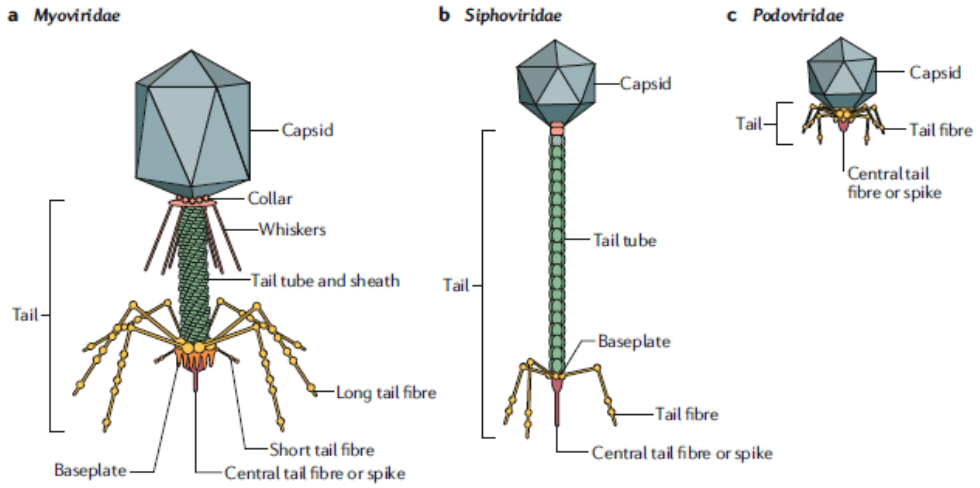


Figure 1.3 - Structure of bacteriophages from the three main phage families within Caudovirales, taken from Nobrega et al., 2018.

The virions of the *Caudovirales* share a common structure consisting of a protein capsid head connected to a multi-part tail comprising of the tail tube, tail fibres, and baseplate/tail spikes. Each of these components plays a role in the infection of the target bacteria.

1.2.2.1 Capsid

The capsid is an icosahedral, proteinaceous head made up of capsomeres sub-units arranged into a repetitive pattern (Madigan and Brock, 2018), and it acts to house the genome of the phage prior to infection of new host bacteria. It has a limited capacity of slightly greater than one full phage genome (Prevelige and Cortines, 2018) but has sufficient elasticity to provide not only chemical protection for the DNA but also mechanical protection (Ivanovska et al., 2004).

1.2.2.2 Tail tube

The tails of phages are an important factor in genome delivery, acting as a conduit between the genome-housing capsid and the tail fibres/tail

spike which facilitate entry of the DNA into the host cell. They are split into three categories which define the *Caudovirales* families:

- Siphoviruses (e.g. *Lactococcus lactis phage* TP901-1) – a long, flexible tail comprising of stacked groups of six protein sub-units through which the DNA from the capsid can pass out of the phage (Fokine and Rossmann, 2014; King et al., 2011);

- Myoviruses (e.g. *Escherichia coli* phage T4) - medium-length tail of similar composition to the siphoviruses which is encapsulated by a contractile sheath (Fokine and Rossmann, 2014).

- Podoviruses (e.g. *E. coli* phage T7) – short, non-contractile tails ending in a “nozzle” protein which lack the tail tube proteins of the sipho- and myoviruses (Zinke et al., 2021).

The differences in the tail structure are responsible for, or are potentially a product of, the different methods by which the phage injects its DNA into the host cell. Siphoviruses, for example, inject their tail tape measure protein (TMP) through the cell wall and membrane without any involvement by the tail tube itself, creating a conduit for the phage DNA to pass through into the cell. The full mechanism of this is not entirely understood but is believed to involve host factors and tail-associated enzymes to facilitate pore formation in the cell wall and DNA transit (Cumby et al., 2015; Fernandes and São-José, 2018; Xu and Xiang, 2017). Myoviruses, however, use the contractile sheath around the tail tube to mechanically drive the TMP through the cell’s protective layers (Leiman et al., 2004) (Figure 1.4).

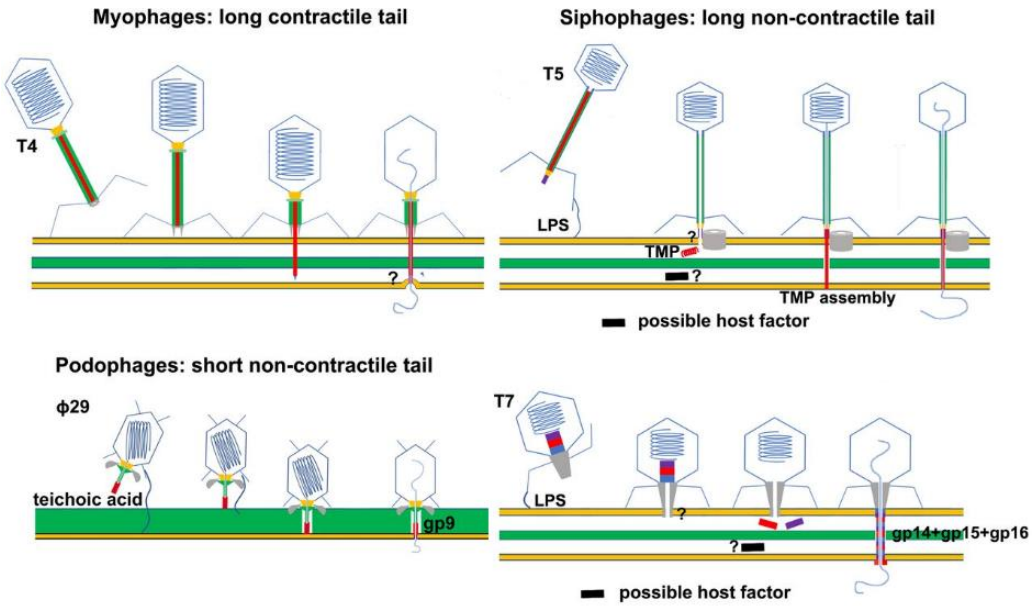


Figure 1.4 - Phage DNA entry mechanisms, adapted from Xu and Xiang, (2017). Green = phage tail or peptidoglycan layers in the cell. Yellow = cell membranes. Long red lines indicate the tail tape measure proteins. Question marks indicate currently unidentified mechanisms. Short, coloured bars indicate additional phage-related proteins believed to facilitate DNA entry.

Podoviral entry mechanisms are slightly more diverse and are dependent on the type of cell being entered. In some cases, the tail of the phage is sufficiently long to traverse the cell wall without the need for additional help. Others, such as *B. subtilis* bacteriophage Φ29, require additional phage-encoded proteins to help build a channel for the DNA to pass through (Xu and Xiang, 2017).

1.2.2.3 Baseplate and tail spikes

In the case of both sipho- and myoviruses, the tail tube culminates in a section known as the baseplate. This multi-protein structure has several functions – it acts as a hub for the host-recognition proteins of the phage, it has a role in cell wall penetration, and it has a central pore through which the phage DNA passes, helping to control DNA injection (Kizziah et al., 2020). Given the importance of the baseplate in the infection mechanisms of phages, the structure has been the focus of several studies across different species, including *E. coli* phages T4 and Mu

(Büttner et al., 2016; Kostyuchenko et al., 2003; Yap et al., 2016) and *Staphylococcus aureus* phages 80α and Φ11 (Kizziah et al., 2020; Li et al., 2016).

1.2.2.4 Tail fibres

The tail fibres of bacteriophages are the receptor binding proteins (RBPs) which are responsible for identifying, and binding to, the host bacterial cell (Le et al., 2013). The variation in these proteins is what determines host specificity (Dams et al., 2019; Dowah and Clokie, 2018). It is quite rare to find a single phage capable of infecting most or all strains of a bacteria (De Sordi et al., 2019), but there are several phages able to infect more than one strain of the target organism – this is called the "host range" (Kutter, 2009).

From a human perspective, the specificity of RBPs is the reason for the consideration of phage therapy as a treatment for bacterial infections. While not as efficient as broad-spectrum antibiotics in that each phage may only be able to be used for a specific bacterial infection, the lack of extraneous targets can allow an infection to be treated without creating the dysbiosis that is the source of issues such as CDI.

1.2.3 Life cycles

Bacteriophages demonstrate a variety of different life cycles depending on which phage is involved and/or the conditions under which the host cell resides. The primary life cycles associated with phages are the lytic and lysogenic cycles (Clokie et al., 2011), however, there are also several less-considered mechanisms of phage survival, such as pseudolysogeny and chronic infection (Clokie et al., 2011; Mäntynen et al., 2021). Each of these life cycles has a place in influencing microbiotal ecosystems but for the purposes of this review, we are primarily concerned with the lytic and lysogenic cycles.

1.2.3.1 Lytic cycle

The core life cycle of a virus involves entering the cell, usurpation of the host's genetic machinery to replicate itself, and then release of the viral progeny by lysis of the cell wall. Viral infection during a lytic cycle therefore consists of five stages – attachment, penetration, synthesis, assembly and release (Madigan and Brock, 2018) (Table 1.3).

Table 1.3 - Summary of stages of bacteriophage lytic cycle

Cycle stage	**Mechanism**
Attachment	RBPs initiate a two-step adsorption process on identification of the host receptor – a reversible binding to bring the particle in closer proximity to the host (Bertozzi Silva et al., 2016; Garen and Puck, 1951), followed by an irreversible binding which triggers conformational changes to facilitate entry into the cell (Molineux and Panja, 2013)
Penetration	The penetration mechanism is dependent on tail type. Once the tail tube has breached the cytoplasm, the viral genome is ejected from the capsid (Madigan and Brock, 2018).
Synthesis	Two methods of synthesis are used by phage: • commandeering of the host's genetic machinery, and/or • encoding their own machinery which the host cell transcribes (Ayora et al., 2002; Mensa-Wilmot et al., 1989; Seco et al., 2013; Weigel and Seitz, 2006). A combination of these methods is often used to initiate transcription and translation of the encoded structural proteins and replication of the genome.
Assembly	Assembly of the new virions is sequential: • an icosahedral capsid is constructed around a portal protein and packaged with the newly-synthesized genome (Aksyuk and Rossmann,

	2011; Fujisawa and Morita, 1997; Murialdo and Becker, 1978; Prevelige and Cortines, 2018) • the tail is constructed beginning with the base plate and tail fibres, above which the tail tube (and sheath for myoviruses) is built (Aksyuk and Rossmann, 2011; Mahony et al., 2016) • terminator proteins at the neck of the tail then interact with the portal protein to join the two sections together (Pell et al., 2009), completing the new virion particle.
Release	ssDNA and all RNA phages – single lysis enzyme encoded on the phage used in conjunction with host-encoded factors (Callanan et al., 2018; Yang et al., 2018; Young et al., 2000) dsDNA phage – a two-enzyme system consisting of a holin protein which permeabilises the inner membrane combined with a muralytic endolysin enzyme which degrades the cell wall (Lin et al., 2017)

The lytic cycle can occur very quickly, with cycle length varying from 20 to 60 minutes and so it can have a large effect on bacterial populations in a short space of time (Madigan and Brock, 2018). To combat this, bacteria have evolved several different defence mechanisms against phage infection; for example, bacteria can develop a genomic immunity to phage infection by modifying the target receptor structure or temporarily rendering it unavailable (Labrie et al., 2010), a process called receptor masking. This prevents the adsorption of the phage to the cell, preventing injection of the phage DNA. Clustered regularly interspaced short palindromic repeats (CRISPR) (Abedon, 2011) are another well-known adaptive bacterial immunity mechanism in which a cell copies short sections of a phage's DNA, creating a library of "spacers" which it can use to recognise and quickly target any further invading phages of that type (Abedon, 2012). Such mechanisms create a constant evolutionary arms race between phage and host, with phages adapting

within as little as 21 days to overcome receptor masking in a mouse model experiment, rendering the bacteria susceptible again (De Sordi et al., 2019). This ability to quickly adapt also facilitates host range expansion in the phage, and experiments by De Sordi et al. (2017) demonstrated how the *E. coli* phage P10 was able to infect a previously immune *E. coli* strain when incubated with its original host and an intermediary strain. This natural coevolution is what makes bacteriophages so interesting as a potential therapeutic option, as it reduces the likelihood that bacteria will evolve a permanent defence against infection by a particular phage in the way that AMR genes have evolved (Koskella and Brockhurst, 2014).

1.2.3.2 Lysogenic cycle

While all phages can enter the lytic cycle, some phages (known as “temperate” phages) can also undergo an intermediate stage in the life cycle called lysogeny, which falls between penetration and synthesis (Figure 1.5). Rather than beginning the synthesis of new virion proteins once in the cytoplasm, the viral genome instead integrates itself into the bacterial genome through the use of its encoded integrase enzyme (Clokie et al., 2011; Haq et al., 2012). This “prophage” is then replicated alongside the host’s genome during its natural life cycle until such a time as stressful conditions (e.g. UV light) cause the prophage to “pop out”, immediately resuming the lytic life cycle (Bertani, 2004).

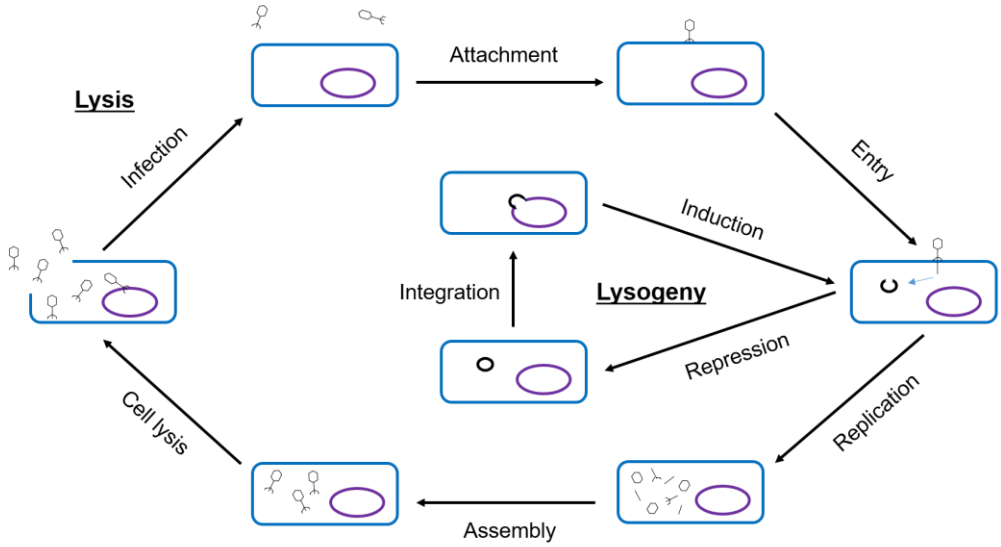


Figure 1.5 - Diagram incorporating the two life cycles available to bacteriophages, adapted from Campbell (2003). Obligatory lytic phages follow only the outermost ring, while temperate phages can enter the central loop when required, maintaining an integrated state until conditions (e.g. damage from UV light) cause the prophage to be induced and re-enter the lytic cycle.

Understanding the cost/benefit analysis underlying temperate strategies for bacteria and phages is complex. For the host bacteria, increased cellular fitness is provided by the prophage in several ways – conveyance of super-infection immunity results in both protection of the cell from lysis by similar phages and lysogens collectively benefiting from the reduction in competition due to the death of other susceptible cells in the population (Bertani, 2004; Lin et al., 2017; Wendling et al., 2021); bacterial evolution is facilitated through horizontal gene transfer, such as the transfer of antibiotic resistance genes (Touchon et al., 2017); and in some cases, the cells rely on prophages for regulation of their own genes to aid in the survival of certain environmental conditions (Govind et al., 2009; Paul, 2008). All of this is balanced by the fact that each cell is carrying a ticking time bomb which will eventually result in its own death once environmental conditions prompt prophage induction.

For phages, on the other hand, the major benefit of lysogeny is the facilitation of long-term population survival, especially in cases where

their hosts are in low abundance, such as in aquatic environments (Li et al., 2020). Temperate phages therefore have a more complex impact on microbial ecologies than strictly lytic phages due to their persistent effect on host populations.

1.2.3.3 Alternative life cycles

As well as the two main lifecycles, there are two less well-documented alternative cycles which sit somewhat between the extremes – chronic infection and pseudolysogeny.

Chronic infection is employed by some ssDNA filamentous phages such as *E. coli* K-12-infecting Ff phages (Rakonjac, 2012). These phages mostly reside episomally within the cell and can produce small numbers of virions without activating the lytic cycle (Mäntynen et al., 2021). These progeny phages are then excreted through the membrane of the host cell, rather than the host being lysed. This difference in strategy is possibly due to the dimensions of filamentous phages – the Ff phage, for example, is approximately 1µm in length, while a typical *E. coli* cell is around 2-3µm. To overcome this, filamentous phage such as these assemble their virions at the cell envelope and release each virion individually by secretion, rather than producing large quantities of viral particles in the cytoplasm that are released simultaneously through lysis (Hay and Lithgow, 2019). Secretion mechanisms for releasing the phage particles therefore allow the phage to produce a greater number of virions without damaging the host cell.

Pseudolysogeny is another less well-characterised state which appears to occur in cases of nutrient depletion where there is a risk that a lytic or lysogenic cycle may need to be aborted (Ripp and Miller, 1998). Essentially a “hibernation” state, pseudolysogeny is the infection of a host cell during which the viral genome initially neither replicates nor integrates, but instead forms a stable “phage carrier cell” which resides in the cytoplasm (Cenens et al., 2013). The genome can stay in this state indefinitely, transferring to daughter cells of the host in either a symmetric

or asymmetric fashion through several generations (Cenens et al., 2013). Once the nutrients are restored or the harsh conditions rectified, the phage genome can resume the lytic or lysogenic cycles as normal (Łoś and Węgrzyn, 2012).

1.2.3.4 The lysis-lysogeny decision

Our knowledge of the mechanisms underpinning the lysis-lysogeny decision in phages originates from extensive characterisation work undertaken on *E. coli* phage λ that started in the 1960s, uncovering a bi-stable cI/Cro switch mechanism that has become a model for gene regulation through the use of genetic toggle switches (Dodd et al., 2005; Oppenheim et al., 2005; Ptashne, 1967). This mechanism can be broken down into a simplified three-component system (just *cI*, *cro* and *cII*) or a more complex five-component system (*cI, cro, cII, cIII* and *N*), but it truly forms part of a much larger regulatory network which encompasses information from the host cell as well as the phage itself (Dodd et al., 2005; Thieffry and Thomas, 1995). Figure 1.6 details the five components involved, however for simplicity I will only describe the core switching components at this time

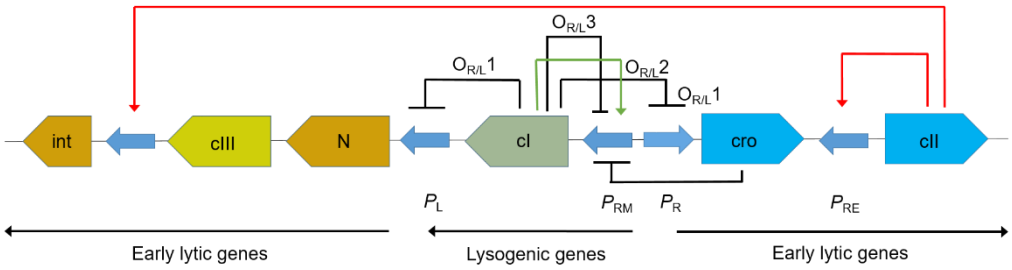


Figure 1.6 - Simplification of the genetic toggle switch for the lysis-lysogeny decision in E. coli phage λ. Black lines indicate repression, while coloured arrows indicate activation reactions.

The toggle switch operates through a series of repression and activation mechanisms which maintain a stable state in one direction or the other,

with *cI* and *cro* being the primary genes of interest. Both show affinities for operator regions associated with the promotors P_R and P_L which control expression of the early phase lytic cycle genes. Each of these operators has three binding sites ($O_{R/L}1$, 2 and 3) for which the two genes have varying binding affinities – CI affinity is $O_{R/L}1 > O_{R/L}2 > O_{R/L}3$, while Cro is the reverse with $O_{R/L}3 > O_{R/L}2 > O_{R/L}1$ (Oppenheim et al., 2005), and it is these varying affinities which help to provide detailed control over the gene regulation.

The process begins with the expression of the CII molecule soon after infection (Oppenheim et al., 2005). CII production has two functions – first, it promotes transcription of the *int* gene to produce the integrase enzyme. This enzyme is responsible for facilitating the integration of the phage DNA into the host genome in order to become a prophage (Cho et al., 2002). Secondly, CII activates the P_{RE} promotor to enable the expression of CI. As CI accumulates, it binds strongly to the O_R1 and O_L1 sites, repressing P_R and P_L and preventing further lytic protein formation (Dodd et al., 2005). After a time, further binding of CI to O_R2 and O_L2 increases the P_{RM} promotor activity, pushing the concentration of CI to approximately 10-fold higher than the threshold required for transitioning back to the lytic cycle (Aurell et al., 2002). Once the concentration of CI exceeds this level, there is sufficient protein to bind more weakly to the O_R3 and O_L3. These operators inhibit P_{RM} activity, thereby working in conjunction with the $O_{R/L}2$ operators to enable a system of self-regulation by modulating CI production (Anderson and Yang, 2008).

When stressful conditions require DNA repair by the host cell, the activity of the cell's RecA proteins results in CI being broken down (Dodd et al., 2005). This reduces the concentration of CI molecules available for repression of P_R and P_L, which triggers transcription of the early lytic genes, including Cro (Anderson and Yang, 2008). Increasing concentrations of Cro inhibit P_{RM} through preferential binding to the O_R3 and O_L3 sites, further reducing CI concentrations and completing the

switch to the lytic life cycle, although the degree to which this binding is necessary requires further study (Oppenheim et al., 2005).

This system has become a model for genetic toggle switches across molecular biology and has therefore been assumed to be the generalised method through which other bacteriophages control the lysis-lysogeny decision. Research has confirmed this to be the case in several species, such as lactococcal phages TP901-1 and Tuc2009 (Johansen et al., 2003; Kenny et al., 2006), however, divergences from this system have been found in other phages (Brady et al., 2021a), such as *B. subtilis* phages Φ3T (quorum-sensing - Brady et al., 2021b; Erez et al., 2017), Φ105 (protease-encoded lysogeny - Bose et al., 2008), and P2 phage of *Shigella* (environmental temperature changes - Bertani, 2004). Improvements in genetic technologies are allowing for more in-depth analysis of these control systems all the time, but a huge amount of work is still required to fully characterise even the narrow band of medically-relevant phages in which we are interested.

1.2.4 Summary

Bacteriophages are ubiquitous, varied and abundant viruses which play a substantive role in the maintenance of microbial ecologies across the planet. They have also contributed extensively to knowledge of molecular biology, including viral gene regulation and the development of CRISPR technologies (Keen, 2015). As we now enter an era where infections such as MRSA are becoming an increasing threat to the world's health, the focus is now turning to whether these naturally-occurring bacterial predators could form part of the solution to antimicrobial resistance.

1.3 Phage therapy

Phage therapy is the treatment of a bacterial infection by the provision of a medical dose of a phage (or cocktail of phages) known to kill the strain responsible for the illness. It has been extensively used in eastern states since World War II but has largely been ignored in the west since the

advent of antibiotics in the mid-twentieth century (Brüssow, 2012; Clokie et al., 2011). However, the increasing global issues with AMR pathogens have led to a resurgence in research into alternative treatments, including revisiting phage therapy. Unfortunately, the lack of rigorous clinical trial data generated from eastern applications has prevented the immediate implementation of phage therapy in western healthcare (Barron, 2022) and so there has been a resurgence of phage research in the west over the last two decades in the hope that phage therapy can be employed as quickly as possible.

1.3.1 Advantages and disadvantages of phage therapy

Bacteriophages hold several advantages over antibiotics, including host specificity, self-amplification and biofilm degradation (Lin et al., 2017). High host specificity allows for much more targeted treatment of infections without promoting the gut dysbiosis which is inevitable when using broad-spectrum antibiotics (although some phage cocktails have demonstrated dysbiosis, but to a lesser degree than antibiotic treatment) (Galtier et al., 2016; Hunt and Ballard, 2013; Lin et al., 2017; Mai et al., 2015). Self-amplification allows for a small dose of phage to be highly efficient in the treatment of infections, as the nature of the phage-host interaction results in its replication *in vivo* at the relevant target site, overcoming the common issue of ensuring that the antimicrobial agent reaches the infection site in sufficient quantity to have an effect (Chan et al., 2013). Issues of developed resistance are also overcome naturally by the phages, as the co-evolution of phage and host results in quick adaptation of the phage to any emerging resistance mechanisms (De Sordi et al., 2019). Finally, biofilms, which consistently cause an issue in chronic infections due to their increased antimicrobial resistance (Musk and Hergenrother, 2006) are naturally disrupted by bacteriophages and so can be used either alone or in conjunction with antibiotics to treat persistent and complex biofilms (Ferriol-González and Domingo-Calap, 2020). This feature also shows promise in aiding in the sterilisation of medical equipment, helping to minimise unnecessary infections by

ensuring that hard-to-eradicate bacterial biofilms are fully broken down prior to use (Donlan, 2009; Motlagh et al., 2016).

A further point to note regarding treatment with antibiotics compared with phage therapy is that antibiotics are known to cause side effects, with as high as 80% of adverse effects being attributable to allergic reactions to the class of antibiotic (Shehab et al., 2008). By contrast, little is known regarding the effects of phages on human biology *in vivo* (Lin et al., 2017). Extensive animal model studies have already proved the efficacy of phage therapy on human pathogens, with studies completed on a range of resistant bacteria including important nosocomial infections such as *Enterobacter faecium* and *C. difficile* (Biswas et al., 2002; Lin et al., 2017), but very little progress has been made regarding clinical trials for phage therapies despite its existing use in the food and dairy industries (Goodridge and Bisha, 2011; Gutiérrez et al., 2019; Hagens and Loessner, 2010; Lucía Fernández et al., 2017). However, as of October 2022, twenty clinical trials are currently listed on the National Institutes of Health (NIH) Clinical Trial directory as either being prepared or actively recruiting participants (ClinicalTrials.gov, 2022) and so potential side effects may soon be able to be analysed.

Phage therapy does also present some disadvantages over antibiotics which must be considered. For instance, the host specificity, while beneficial in minimising secondary effects on the microbiome, means that therapies cannot have a generalised use like broad-spectrum antibiotics as each phage will only infect certain strains. Personalised phage cocktails can be time-consuming and costly to make but developing standardised cocktails for different bacterial types still risks the cocktail not being suitable for the treatment of all strains, while increasing the possibility of dysbiosis occurring (Chan et al., 2013). Bacteria also have strategies for evading phage action thanks to their natural co-evolution, with mechanisms such as CRISPR designed to eliminate phage DNA on entry, and receptor masking preventing entry at all (Abedon, 2012). It takes time for the phage to overcome these immune responses, and so

treatment times may be prolonged. Additionally, some phages are known to harbour AMR genes, and they naturally undertake horizontal gene transfer with their hosts (Colomer-Lluch et al., 2011; von Wintersdorff et al., 2016). There is, therefore, a risk of some transference of further resistance genes (Lin et al., 2017; Penadés et al., 2015). Gut microbiomes, in particular, will need careful assessment in relation to the transfer of AMR genes, as the increased exposure to antibiotics in the gut as opposed to the natural environment may facilitate the transfer of these genes through selective pressure (Touchon et al., 2017; Wendling et al., 2021).

Despite these disadvantages, the ability to target specific bacteria without creating gut dysbiosis is a key advantage in addressing the issue of antibiotic-resistant infections, in particular those such as CDI where antibiotic use is a pre-disposing factor.

1.3.2 Factors affecting phage therapy

The degree of biological variation inherent in bacteriophages complicates the implementation of phage therapy, as there are many factors to consider in terms of ensuring that a safe and consistent treatment can be developed. In an attempt to begin the process of untangling the issues, the US Food and Drug Administration (US FDA) set out a list of requirements for a phage to be considered suitable for therapeutic use (Brüssow, 2012):

- only obligatory lytic phages to be used which do not avoid clearance from the body (not temperate phages)
- the phage must preferentially use virulence factors as receptors
- the phage must be derived from the natural environment, and,
- it must have been fully characterised, including DNA sequencing, toxicity testing in animal systems, as well as tested for stability and sterility

Almost all of these requirements are understandable within the scope of safe treatment regimens – lytic phages will always kill their host, while temperate phage risk creating immune sub-populations and result in the persistence of the infection; virulence factors as receptors means that the target bacteria would need to modify (and potentially reduce) its own virulence to evade phage entry (De Sordi et al., 2019); and full characterisation of any phage is sensible to ensure that the effects of the phage can be fully assessed prior to medical use.

The final issue of the phage being isolated from the natural environment, however, is more complex. As previously stated, personalised phage cocktails are time-consuming to create, due to the need to identify the infecting bacterial strain and then isolate appropriate phages for therapeutic use. This leads to a range of issues: 1) it can cause substantial delays in treatment; 2) it complicates regulatory mechanisms which are not fit for purpose in this case (Fauconnier, 2017; Ferriol-González and Domingo-Calap, 2020); and 3) some bacteria have few to no obligatory lytic phages available so far. Genetic engineering of existing phages, or the creation of wholly synthetic phages, would provide a framework through which therapeutically-suitable phages can be created without the time investment of isolations, as well as producing a more controllable product. Therefore, in my opinion, genetic engineering of phages is more likely to result in standardised treatments being made available than relying on natural phages.

The existing global reticence surrounding genetic modifications presents a barrier to this route, however, there have been recent movements towards re-assessing genetically-modified organisms in both the UK (Department of the Environment, Food and Rural Affairs, 2021) and the EU (European Commission, 2022) which may allow future advances in engineering phages to be considered for medical applications. In the meantime, though, there are several initiatives across the world devoted to growing our library of available natural phages for research and therapeutic uses, including the SEA-PHAGES programme in the US

(Jordan et al., 2014) and the Citizen Phage Library in the UK (Temperton, 2021).

1.3.3 Existing cases using phage therapy

Despite a lack of clinical trials, there are several cases in which phage therapy has been completed successfully. In Tbilisi, Georgia, for instance, the Eliava Phage Therapy Centre is a leading provider of phage therapy cocktails to treat a range of medical conditions. Even in the west, there are documented cases where phage therapy has been allowed for "compassionate use" – the use of unauthorised medicines in cases where all other options have been exhausted. These include the 2015 case of Tom Patterson, a patient suffering from antibiotic-resistant *Acinetobacter baumannii* infection whose wife, Dr. Steffanie Strathdee, later set up the Centre for Innovative Phage Applications and Therapeutics (IPATH) at the University of California San Diego School of Medicine (Schooley et al., 2017), and a 15-year old cystic fibrosis patient who was treated for a pernicious infection of *Mycobacterium abscessus massiliense* (Dedrick et al., 2019). A further instance of phage therapy was also recently released in *Nature Communications*, documenting the use of phage therapy to treat a pandrug-resistant *Klebsiella pneumonia* infection in a survivor of the 2016 Brussels Airport bombing (Eskenazi et al., 2022).

Despite these well-publicised incidences of successful treatment, there have been too few individual cases for an in-depth review of the efficacy of phage therapy in humans. However, Dedrick et al. (2022) recently completed a study of 20 compassionate use cases of phage therapy to treat a range of non-tuberculosis *Mycobacterium* infections (NTM) in which variation could be observed in the efficacy of the treatments. Many of the patients were affected by co-morbidities which complicated the overall interpretation of the negative or inconclusive results, but a positive response to treatment was observed in more than half of the patients. Equally important observations were made in that no phage resistance was developed, indicating the potential to use single phage therapies, and that no reactions to the administration of the treatments were

observed in the patients. While limited to the *Mycobacterium* infections in question, the study was positive in its more consistent evaluation of the effects of phage therapy across multiple individuals, and it raises important considerations for future, more generalised use of phage therapy for these infections and others.

1.3.4 Phage therapy for CDI

Phage therapy has the potential to be a more effective treatment for CDI than antibiotics for several reasons – 1) it removes the chance of leaving an antibiotic-resistant sub-population behind that could continue to cause morbidity; 2) there is a reduced risk of selecting for greater antibiotic resistance amongst the infecting strain which could be passed on to other patients; and 3) it avoids renewing the gut dysbiosis that allowed the *C. difficile* to colonise in the first place, thereby reducing the chance of either the same or a new strain causing recurrence.

However, the implementation of phage therapy for CDI is hampered by the fact that no known purely lytic phages have been isolated for *C. difficile* to date (Heuler et al., 2021), meaning lysogen formation is always a risk. Despite this, some studies have been undertaken using the existing phages to test efficacy.

The earliest documented attempt at treating CDI with phages was that of Ramesh et al. in 1999, who used temperate phage CD140 in a mouse model. Phage-treated mice largely survived the first challenge with *C. difficile* strain 135I where control animals died within 72 hours. However, one treated animal died in this time frame and analysis of the *C. difficile* post-mortem showed resistance to the phage. A later batch fermentation study by Meader et al. (2010) demonstrated that the formation of lysogens in *C. difficile* NCTC 11204, when infected by phage ΦCD27, may be affected by dosage, as resistant lysogens were recovered after using a multiplicity of infection (MOI) of 7 for the experiment but not when using an MOI of 10 (the multiplicity of infection is the ratio of phage to

bacteria – MOI 10 refers to there being 10 phages for every one bacterial cell).

There is some question as to whether a cocktail of temperate phages could provide sufficient safety as long as there was enough difference between the phages to not be affected by super-immunity. Studies have been conducted to this effect which have proved to have a high potential for prophylactic and remedial use, including in infections where biofilms are already established (Nale et al., 2018, 2016b, 2016a). However, the increased chance of transduction remains, so it is unlikely that a temperate phage cocktail would be able to satisfy medical safety requirements.

It is possible that lytic phages for *C. difficile* do exist in the environment but that isolation exercises have been too limited to obtain them. It is quite common for new isolation activities to be focussed on inducing existing prophages rather than isolating from environmental samples, as isolation can be a time-consuming activity (Heuler et al., 2021) but this almost guarantees that these attempts will isolate temperate phages (although there are documented instances of genetic events resulting in a lytic phage from an induced prophage - Pope et al., 2011). In my opinion, isolation from environmental samples is therefore a more realistic approach if we wish to isolate any potentially lytic *C. difficile* phages.

1.3.5 Genetically engineered phages for phage therapy against CDI

Given the scarcity of lytic *C. difficile* phages so far experienced, it is necessary to investigate avenues through which the available phage can be modified to better suit therapeutic needs. Genetic engineering of phages is becoming increasingly common, with examples demonstrated in *E. coli* phages T4 (Tao et al., 2017) and T7 (Kiro et al., 2014), *S. thermophilius* phage 2792 (Martel and Moineau, 2014) and *M. abscessus* phages ZoeJ and BPs (Dedrick et al., 2019), although only the latter have been used therapeutically. There is even work towards the creation of wholly synthetic phages which, if successful, would allow complete

control to eliminate any undesirable genes within phage for phage therapy (Kilcher and Loessner, 2019).

However, the biological variation between phages means that specifics that could be used to guide the engineering of *C. difficile* phages are few and far between, and there is only one example of attempts to engineer a lytic *C. difficile* phage from a temperate one in the recent literature, a study by Selle et al. in 2020. In this study, the authors were primarily interested in using phages to re-purpose the endogenous type I-B CRISPR-cas system of their target strain to kill the host cell, however, they make mention during the article of creating a lytic phage by the deletion of “a region of the genome encoding the *cI* repressor and integrase gene”. Unfortunately, no details as to how this was accomplished are provided, and the genome of the phage is not freely available for review. During *in vivo* mouse modelling of their engineered phages, a minimum of a 2-log reduction in the animals’ faecal *C. difficile* levels was recorded after 48 hours, which is a promising result. However, when the *C. difficile* cells were tested, lysogens were detected, despite no lysogens having been created during *in vitro* tests. Without additional information regarding the genome of the phage or what area was deleted, it is difficult to reliably hypothesise why this may have occurred, although the authors posited that it could be due to complementation by existing prophages in the *in vivo* target cells.

Altogether, genetic engineering of phages has promise in providing phages suitable for use in phage therapy, but this approach requires rigorous study to ensure that no adverse effects are created. A combined approach of environmental isolation and genetic engineering is therefore required in order to obtain a lytic *C. difficile* phage as quickly as possible, especially given that personalised phage therapy medications are highly specific and therefore each patient may need a different phage. Fully synthetic alternatives would be optimal, but these do not currently appear to be being addressed for *C. difficile*.

1.3.6 Summary

Phage therapy is an area of growing interest given the increasing issues with antibiotic-resistant infections, with benefits including high specificity that prevents secondary microbiome damage, and avoidance of AMR gene selection. It has already been used successfully in several well-publicised compassionate use cases, however, the lack of clinical trial data hampers its wider use. Work is ongoing to move forward with these trials in areas such as the treatment of chronic cystic fibrosis secondary infections and the treatment of post-operative infections in prosthetics, but its use for CDI is limited by a lack of phages suitable for therapeutic use. Given that phage therapy has the potential to be a highly effective way of treating CDI (due to removing the risk of further dysbiosis leading to reinfection and recurrence), obtaining lytic phages which infect *C. difficile* is a major priority, and therefore the focus of this research project.

1.4 Project aims

The aim of this project is to obtain a lytic phage infecting *C. difficile* which could be suitable for use in phage therapy. To accomplish this, the project has two primary objectives to address, one of which will be split into two parts:

1) Attempt to isolate and characterise further *C. difficile* phages from environmental samples in the hope of obtaining a naturally-occurring lytic phage.

2) Employ CRISPR-Cas9 engineering techniques to modify available temperate *C. difficile* phages to remove the ability to form lysogens by:

    a. deleting the *cI* repressor equivalent gene to prevent lytic gene repression, and
    b. deleting the integrase gene to remove the potential for transduction by prophage integration.

Chapter 2

Materials and Methods

2 Materials and Methods

2.1 Bacterial strains and plasmids

Table 2.1 contains a complete list of the RT 027 *C. difficile* strains used for the isolation of novel bacteriophages in this study. All strains used for host range testing of bacteriophages are listed in Appendix 8.1. Table 2.2 details all strains used in the knockout of the *cI* and *int* genes in the test phage. Table 2.3 shows all plasmids used as part of this study to date.

Table 2.1 - Complete list of RT 027 C. difficile strains used in the isolation of novel bacteriophages from environmental sewage samples.

In-house strain name	Official strain name	Location of Isolation	Source/Reference
R20291	-	Stoke Mandeville, UK	Val Hall, Anaerobic Reference Unit
DH1834	-	Ipswich	Val Hall, Anaerobic Reference Unit
DH1466	-	Northampton	
DH1858	-	Sunderland	
DH1751	-	Bradford	
R24988	-	Antrim, N. Ireland	
R24694	-	Haverfordwest	
DH1342	-	Macclesfield	
DH1329	-	Coventry	
R23970	-	Dublin, Ireland	

L2	81568	France	Ed Kuijper, LUMC for EACCAD grant
L3	81566	France	
L4	05-1223-046	Belgium	
L5	Libramont 1	Belgium	
L6	5108111	The Netherlands	
L8	32219	Luxembourg	
L9	M246	Ireland	
L10	2191	Ireland	
L11	51556	Germany	
L12	51557	Germany	
L14	60902	Switzerland	
L15	T31694	Denmark	
R12087	-	Historical EU strain (mid-80s)	Michel Popoff

Table 2.2 - List of bacterial strains used for gene knockout of cI and int phage genes

Bacteria	**Strain name**	**Description/PCR ribotype**	**Source/reference**
Clostridium difficile	CD2301	Wild-type *C. difficile* 014 strain	UoN SBRC (Minton group)
Clostridium difficile	CD2301L	Lysogen strain of CD2301 containing the phiCD2301 prophage	UoN SBRC (Minton group)

Escherichia coli	XL1 Blue	Chemically competent strain for long-term storage of vectors.	New England Biolabs (NEB)
Escherichia coli	sExpress	Conjugative transfer strain of NEB Express *E. coli* containing the R702 plasmid (Woods et al., 2019)	UoN SBRC (Minton group)
Escherichia coli	NEB® 5-alpha	Chemically competent derivative of *E. coli* DH5α.	New England Biolabs
Escherichia coli	NEB® 10-beta	Chemically competent derivative of *E. coli* DH10β.	New England Biolabs
Escherichia coli	NEB® Stable™	Chemically competent strain designed for cloning unstable or toxic elements	New England Biolabs

Table 2.3 - List of plasmids used in this study to date

Plasmid	**Description**	**Source**
pMTL83151	Modular plasmid containing pCB102, ColE1+*tra*, CM^R/Tm^R and a multiple cloning site	Heap et al., 2009
pRECas1_MCS	CRISPR-Cas9 plasmid containing a theophylline-inducible riboswitch and multiple cloning site	Cañadas et al., 2019
pRECas1_phiCD2301 *int*_sg1	pRECas1_MCS backbone containing homology arms and single guide RNAs 1, 2 and 3 respectively targeting the *int* gene of phiCD2301 in CD2301L	This study
pRECas1_phiCD2301 *int*_sg2		
pRECas1_phiCD2301		

*int*_sg3		
pRECas1_phiCD2301 intcomp	pRECas1_MCS backbone containing homology arms for the complementation of the native integrase gene	This study
pRECas1_phiCD2301 c1_sg1	Variant of pRECas1_MCS containing homology arms and single guide RNAs 1, 2 and 3 respectively targeting the *cI* gene of phiCD2301 in CD2301L	This study
pRECas1_phiCD2301 c1_sg2		
pRECas1_phiCD2301 c1_sg3		
pRPF185	Expression plasmid containing a *gusA* reporter gene under the control of the inducible anhydrotetracycline (ATc) promotor system.	Fagan and Fairweather, 2011
pMTL-CW21	In-house transposon plasmid containing the ATc promotor system and using an additional ribosome binding site added upstream of the Himar1C9 gene.	UoN SBRC (Minton group)
pSJK1	pMTL83151 backbone containing $P_{fdx\text{-}E}$ and the native phiCD2301 integrase gene inserted as an NdeI/NheI fragment	This study
pSJK1-FAST	As per pSJK1 but with an additional FAST reporter gene fused to the end of the integrase gene.	This study
pRibo-FAST	pMTL83151 backbone containing FAST reporter gene under the control of $P_{fdx\text{-}E.}$ Designed for construction using HiFi ligation.	This study

pRiboFAST-RL	As per pRibo-FAST but designed for construction using restriction enzymes.	This study
pRibo-*int*-FAST_RL	As per pSJK1 but with a modified integrase gene fragment.	This study
pSJK2	pRPF185 backbone with the original *gusA* gene replaced with a SacI/BamHI module containing the native phiCD2301 integrase gene.	This study
pSJK2-FAST	As per pATc-*int* but with an additional FAST reporter gene fused to the end of the integrase gene.	This study
pATc-FAST	pRPF185 with the *gusA* gene replaced by a fragment containing a ribosome binding site taken from pMTL-CW21 and a FAST reporter gene using SacI and BamHI restriction enzymes. Designed for construction using HiFi ligation.	This study
pATcFAST-RL	As per pAtc-FAST but deisgned for construction using restriction enzymes.	This study
pATc-*int*-FAST_RL	As per SJK2-FAST but with the additional ribosome binding site from pMTL-CW21 added upstream of the modified integrase gene used in pRibo-*int*-FAST_RL.	This study
p1341KO2	pRECas1_MCS backbone containing homology arms targeting the hypothetical phiCD2301 gene identified as a potential repressor protein.	This study

p1341TASing	TargetAid vector vFS77 with single guide RNA seed replaced with a seed targeting the potential phiCD2301 repressor gene.	This study
P1341TAMulti	TargetAid vector vFS74 with single guide RNA seeds replaced with three alternative seeds targeting the phiCD2301 hypothetical repressor gene.	This study
pMTL83151y-Ccarb-v2_FAST	In-house confirmed FAST plasmid used as positive control for FAST assay.	UoN SBRC (Minton group)
pJ8221	In-house plasmid containing the FAST gene for cloning purposes	UoN SBRC (Minton group)

2.2 Oligonucleotide primers

Table 2.4 details all oligonucleotide primers used in this study so far for the construction of the CRISPR knockout vectors.

Table 2.4 - Oligonucleotide primers used in this study to date

Primer name	**Primer sequence (5'-3')**	**Function**
phiCD2301int_LHA_F3	TATATAGACGTCGAACTTAAAAAAATGAAGTCTTTAG	Amplification of homology arms targeting the phiD2301 *int* gene.
phiCD2301int_LHA_R1	ATTAAACGTACACAAGAAGTGTTAACCGATACCGTTTACGAAATAGG	
phiCD2301int_RHA_F1	ACCGATACCGTTTACGAAATAGGTGTAAGGTGATATTTTTTTATTTG	
phiCD2301int_RHA_R1	TATATAGGCGCGCCGATATTTTTAAAAAAGCTGCTTC	
phiCD2301c1_LHA_F2	TATATAGACGTCTGAATTTATAGGAAATGACAGG	Amplification of homology arms

phiCD2301c1_LHA_R1	CTAAATAGCGCTTGTTTTTATTTACCGATACCGTTTACGAAATAGG	targeting the phiD2301 *cI* gene.
phiCD2301c1_RHA_F1	ACCGATACCGTTTACGAAATAGGGTTCGTGCAAGTTTTACCA	
phiCD2301c1_RHA_R1	TATATAGGCGCGCCGTTCCCACTTAATTAAATCAGC	
phiCD2301int_sgRNA1_F	TTTTCGTCGACAAAGAATAGGAAGCCTCGGAGTTTTAGAGCTAGAAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	Forward primers for amplification of single guide RNAs targeting the phiD2301 *int* gene.
phiCD2301int_sgRNA2_F	TTTTCGTCGACATATAACAGCACATAGCATGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	
phiCD2301int_sgRNA3_F	TTTTCGTCGACAAAAGAAAATATCCTCACATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	
phiCD2301c1_sgRNA1_F	TTTTCGTCGACAATTGGTTAAGATTCGGAGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	Forward primers for amplification of sgRNAs targeting the c*I* gene.
phiCD2301c1_sgRNA2_F	TTTTCGTCGACTTAATAAATAAATTAATCGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	
phiCD2301c1_sgRNA3_F	TTTTCGTCGACGGATAAGAGATGCGACAGACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	
sgRNAUniv_R	GTATCAGACGTCATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTATAAAAAAAGCACCG	Universal reverse primer for

	ACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCC	amplification of sgRNAs.
sgUniv_RiboHiFi	GACGTCATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTATAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCC	Universal reverse primer for amplification of sgRNAs used in NEBuilder HiFi cloning.
pMTLCas9_F	CCTATCTACATCTACAAAATAATCATC	Screening to either side of the homology cassette in pRECas1_MCS vector.
pCB102_R1	CTGTTATGCCTTTTGACTATC	
phiCD2301intchk_F7	CCTTATTTAAAATAATTGGTATACATA	Screening to confirm knockout of *int* gene from the CD2301L genome.
phiCD2301intchk_M1	CTCATGAGCTTTTTTTATTCTC	
phiCD2301intchk_R1	TTGCCTTGTTTAAAAATTTC	
phiCD2301c1chk_F3	TTTACAGGAGCTAAGTGACG	Screening to confirm knockout of *cI* gene from the CD2301L genome.
phiCD2301c1chk_M1	TGAAAGAGGACGAATTTAGAG	
phiCD2301c1chk_R1	CCTCCTGCTCTTCTATTTTT	
phiCD2301intWM_F1	ACAGTGGAAAAGATATTGTCGCAGTCAAGGAATTGTTG	Overhang primers to create watermarked integrase gene for complementation.
phiCD2301intWM_R1	GACAATATCTTTTCCACTGTCAATATAAATTTTATAAGCATAAGTTTTC	
BM6sgRNA	TTTTCGTCGACACCGATACCGTTTACGAAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	Forward primer for sgRNA targeting the BM6 bookmark.

pfdxeNde1_F	TATATACATATGGTGTAGTAGCCTGTGAAATAAGTAAG	Construction of expression cassette for pSJK1 with P_{fdx-E} upstream of the integrase gene.
Pfdxe-2301intOH_R	GAATATCTAACACCTTGTTGTTACCTCCTTAGCAGGGTG	
2301int-pfdxeOH_F	AACAAGGTGTTAGATATTCAAGATTATCTAAAGTATAAAAGTTACAGAAAC	
Nde1-ATc_F	TATATACATATGATAAAAATAAGAAGCCTGCATTTG	Construction of expression cassette consisting of the ATc promotor system placed upstream of the integrase gene in pMTL83151.
ATc-2301intOH_R	TATCTAACACAAAATTTTCTCCTTTACTGCAGGAGCTCAG	
230int-ATcOH_F	GAGAAAATTTTGTGTTAGATATTCAAGATTATCTAAAGTATAAAAGTTACAG	
FAST+Nde1_F	TATATAGCTAGCATGGAACACGTAGCATTTG	FAST reporter fragment for insertion into pSJK1 to create pSJK1-FAST
FAST+Nhe1_R	TATATAGCTAGCTCATACCCTCTTAACGAAAAC	
SacI-2301int_F1	TATATAGAGCTCACACAAGAAGTGTTAGATATTCAAG	Insertion of *int* gene into pRPF185 to form pSJK2.
2301int-BamHI_R1	TATATAGGATCCTCACCTTACAAAGTCGTCTAATG	
FAST-BamHI_F1	TATATAGGATCCATGGAACACGTAGCATTTG	FAST reporter fragment for insertion into pSJK2 to create pSJK2-FAST.
FAST-BamHI_R1	TATATAGGATCCTCATACCCTCTTAACGAAAAC	
ColE1+tra_F1	CTGTGGATAACCGTATTACC	Check construction of pSJK1 (used with pCB102_R)
Atc-check_F1	GACTTAAGGGCGAATTCTG	

Atc-check_R1	AAGTACATCACCGACGAGC	Primers to check construction of pSJK2
RiboCas1_fwd	GGCGCGCCGCCATTATTT	Amplification of RiboCas backbone fragments for HiFi construction of CRISPR-Cas9 vectors.
RiboCas1_rev	AGGAGGTGACTGAGAATTCGTAATCATGGTCATATGGATACAGC	
RiboCas2_fwd	GATTACGAATTCTCAGTCACCTCCTAGCTG	
RiboCas2_rev	GAAAACTCCTCCTTAAGATTTATATATG	
Atc1_fwd	TTTTCGTTAAGAGGGTATGAATCAACAACTCTCCTGGC	Amplification of pRPF185 backbone for HiFi construction of expression vectors.
Atc1_rev	TTGAGATCCTTTTTTTCTGC	
Atc2_fwd	GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTAC	
Atc2_rev	AAAATTTTCTCCTTTACTGCAG	
pMTL83151_F2	GCTAGCATAAAAATAAGAAGCC	Amplification of pMTL83151 backbone for HiFi construction of expression vectors.
pMTL83151_R2	CATATGGATACAGCGGCC	
2301L-1339HA_fwd	GCAGGCTTCTTATTTTTATGACGTCGGAGCATTAACAACGTCTTC	Amplification of homology arms targeting gene 1339.
2301_1339_LHA_R1	CCTATTTCGTAAACGGTATCGGTTTTCATTTTAAGTCCCCCTTATTTCACAG	
2301L-1339_RHA_F1	ACCGATACCGTTTACGAAATAGGTTGCATTTAAACAAAAAGTCAGTC	
2301L-1339HA_rev	TTCAAAAAAATAATGGCGGCGCGCCTGGTAAAACTTGCACGAAC	
1339_sg1Di	GTCGACCGATTTAGCAGTAATCCTGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	Amplification of sgRNAs

	CGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC	targeting gene 1339.
1339_sg1Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACCGATTTAGCAGTAA TC	
1339_sg2Di	GTCGACAATCTATACTGAATGG AACAGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC	
1339_sg2Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACAATCTATACTGAAT G	
1339_sg3Di	GTCGACACGATTTAGCAGTAAT CCTGGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC	
1339_sg3Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACACGATTTAGCAG	
2301L-1340_HA_fwd	GCAGGCTTCTTATTTTTATGACG TCGTACCCATAGAGCCACTTC	Amplification of homology arms targeting gene 1340.
2301L-1340_LHA_R1	CCTATTTCGTAAACGGTATCGG TTATTAGTTAGTTGTTAAAATAT GTTATAATAGTTTTAGC	
2301L-1340_RHA_F1	ACCGATACCGTTTACGAAATAG GAATTAAACACAAAGTCATTCCC	
2301L-1340_HA_rev	TTCAAAAAAATAATGGCGGCGC GCCGTTGGAGTTGCTGTTATG	
1340_sg1Di	GTCGACTCATAACGCTATTTTTG ACAGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCC GTTATCAACTTGAAAAAGTGGC ACCGAGTCGGTGC	Amplification of sgRNAs targeting gene 1340.
1340_sg1Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACTCATAACGCTATTTT TG	
1340_sg2Di	GTCGACAAAATATAACTGACATA ACAGTTTTAGAGCTAGAAATAG	

	CAAGTTAAAATAAGGCTAGTCC GTTATCAACTTGAAAAAGTGGC ACCGAGTCGGTGC	
1340_sg2Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACAAAATATAACTGAC ATAAC	
1340_sg3Di	GTCGACATGTGCTAAAAATTTCA CTTGTTTTAGAGCTAGAAATAGC AAGTTAAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGTGGCA CCGAGTCGGTGC	
1340_sg3Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACATGTGCTAAAAATTT C	
2301L-1341_HA_fwd	GCAGGCTTCTTATTTTTATGACG TCCTTCTGTCAGTTTCGGTTAC	Amplification of homology arms targeting gene 1341.
2301L-1341_LHA_R1	CCTATTTCGTAAACGGTATCGG TAAATAATTTTTTTATTTCGGAC AAAAACTC	
2301L-1341_RHA_F1	ACCGATACCGTTTACGAAATAG GAATTATCCATACATATTCACCT CCCTTCG	
2301L-1341_HA_rev	TTCAAAAAAATAATGGCGGCGC GCCGAGTTCGATATTGCAAAAC	
1341_sg1Di	GTCGACAACCACTCTGCGCAAA AGAGGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC	Amplification of sgRNAs targeting gene 1340.
1341_sg1Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACAACCACTCTGCG	Amplification of sgRNAs targeting gene 1341.
1341_sg2Di	GTCGACAACCACTCTTTTGCGC AGAGGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC	
1341_sg2Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACAACCACTCTTTTG	

1341_sg3Di	GTCGACACTCTGCGCAAAAGAG TGGTGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC	
1341_sg3Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACACTCTGCGCAAAAG	
2301L-1343_HA_fwd	GCAGGCTTCTTATTTTTATGACG TCGCAAATGCAAGTGTCACAG	Amplification of homology arms targeting gene 1343.
2301L-1343_LHA_R1	CCTATTTCGTAAACGGTATCGG TTTATAATATGAAATTTGATTTA GAAGAAAACG	
2301L-1343_RHA_F1	ACCGATACCGTTTACGAAATAG GTTGTACTCAATTAAATTACCCC	
2301L-1343_HA_rev	TTCAAAAAAATAATGGCGGCGC GCCCAGTTCTCTTGCTTTTATTA TATCAC	
1343_sg1Di	GTCGACGATGTCTTGATATTCC AGTGGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC	Amplification of sgRNAs targeting gene 1343.
1343_sg1Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACGATGTCTTGATATT C	
1343_sg2Di	GTCGACATACATAATACACTATA CGAGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCC GTTATCAACTTGAAAAAGTGGC ACCGAGTCGGTGC	
1343_sg2Hi	ATATAAATCTTAAGGAGGAGTTT TCGTCGACATACATAATACACTA TAC	
1343_sg3Di	GTCGACGCGATGTCTTGATATT CCAGGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC	

1343_sg3Hi	ATATAAATCTTAAGGAGGAGTTTTCGTCGACGCGATGTCTTGATATTC	
FAST-Atc_fwd	GCAGTAAAGGAGAAAATTTTATGGAACACGTAGCATTTG	HiFi construction of pATc-FAST.
FAST-Atc_rev	TCATACCCTCTTAACGAAAAC	
Ribo-FAST_fwd	GCGGCCGCTGTATCCATATGGTGTAGTAGCCTGTGAAATAAG	HiFi construction of pRibo-FAST.
Ribo-FAST-OH_R1	CTACGTGTTCCATCTTGTTGTTACCTCCTTAGCAG	
FAST-RiboOH_F1	CAACAAGATGGAACACGTAGCATTTG	
Ribo-FAST_rev	TCATACCCTCTTAACGAAAAC	
1339-LHA_R2	GTTTAAATGCAACCTATTTCGTAAACGGTATCGGTTTTCATTTTAAGTCCCCC	Second attempt at HiFi construction of expression vectors targeting genes 1339, 1340 and 1341.
1339-RHA_F2	ACTTAAAATGAAAACCGATACCGTTTACGAAATAGGTTGCATTTAAACAAAAAGTCAG	
1340-LHA_R2	TTGTGTTTAATTCCTATTTCGTAAACGGTATCGGTTATTAGTTAGTTGTTAAAATATGTTATAATAGTTTTAG	
1340-RHA_F2	CAACTAACTAATAACCGATACCGTTTACGAAATAGGAATTAAACACAAAGTCATTCC	
1341-LHA_R2	GTATGGATAATTCCTATTTCGTAAACGGTATCGGTAAATAATTTTTTTATTTCGGACAAAAAC	
1341-RHA_F2	AAAAAAATTATTTACCGATACCGTTTACGAAATAGGAATTATCCATACATATTCACCTC	
1339RL_LHA_F1	TATATAGACGTCGGAGCATTAACAACGTCTTC	Primers for restriction and

1339RL_LHA_R1	CCTATTTCGTAAACGGTATCGGTTTTCATTTTAAGTCCCCCTTATTTC	ligation construction of expression vectors targeting genes 1339, 1340 and 1341.
1339RL_RHA_R1	TATATAGGCGCGCCTGGTAAAACTTGCACGAAC	
1340RL_LHA_F1	TATATAGACGTCGTACCCATAGAGCCACTTC	
1340RL_RHA_R1	TATATAGGCGCGCCCACTGTTTGGTAATATAAATAATGC	
1341RL_LHA_F1	TATATAGACGTCGTACCCATAGAGCCACTTC	
1341RL_RHA_F1	ACCGATACCGTTTACGAAATAGGAATTATCCATACATATTCACCTCC	
1341RL_RHA_R1	TATATAGGCGCGCCGAGTTCGATATTGCAAAAC	
SacI-RBS_F1	TATATAGAGCTCGACGTCCGCGAGAGA	Construction of pATc-FAST.
RBS-FASTOH_R1	ACGTGTTCCATATGTATTTCCTCCTCTTCAATATATTTA	
FAST-RBSOH_F1	AGGAAATACATATGGAACACGTAGCATTTG	
FAST-BamHI_R1	TATATAGGATCCTCATACCCTCTTAACGAAAAC	
1341KO2-LHAHi_F1	GCAGGCTTCTTATTTTTATGACGTCGAAAGAAGTCACTAAGAAGG	HiFi construction of p1340KO2.
1341KO2-RHAHi_F1	AAAAAAATTATTTACCGATACCGTTTACGAAATAGGAATTATCCATACATATTCACCTC	
1341KO2-RHAHi_R1	TTCAAAAAAATAATGGCGGCGCGCCCTCATAACGCTATTTTTGAC	
1341KO2-LHARL_F1	TATATAGACGTCGAAAGAAGTCACTAAGAAGGAAG	Restriction and ligation construction of p1340KO2.
1341KO2-RHARL_R1	TATATAGGCGCGCCCTCATAACGCTATTTTTGACA	

1341sg1RL	TTTTCGTCGACAACCACTCTGCGCAAAAGAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	Forward primer to amplify sgRNA for p1340KO2.
1341KO3-LHAHi_F1	GCAGGCTTCTTATTTTTATGACGTCCAAAACTCTTGTTTCTTCC	HiFi construction of p1340KO3.
1341KO3-RHAHi_R1	TTCAAAAAAATAATGGCGGCGCGCCGCTTCCAAAGAAGATATTATCG	
1341KO3-LHARL_F1	TATATAGACGTCCAAAACTCTTGTTTCTTCC	
1341KO3-RHARL_R1	TATATAGGCGCGCCGCTTCCAAAGAAGATATTATCGAG	
Pfdxe-intOHRL_R1	ATCCCTAACATCTTGTTGTTACCTCCTTAGCAG	Restriction and ligation construction of pRibo-*int*-FAST.
Int-PfdxeOH-RL_F1	GTAACAACAAGATGTTAGGGATAACGACAGG	
Ribo-int-FASTOH-RL_R1	ACGTGTTCCATTCACCTTACAAAGTCGTCTAATG	
FAST-intOH-RL_F1	TTGTAAGGTGAATGGAACACGTAGCATTTG	
RBS-intOH_R1	ATCCCTAACATATGTATTTCCTCCTCTTCAATATATTTA	Construction of pATc-*int*-FAST using additional RBS region.
Int-RBSOH_F1	AGGAAATACATATGTTAGGGATAACGACAGGA	
Atc-FAST-intOH_F1	TTGTAAGGTGAATGGAACACGTAGCATTTG	
Ribo1T_F	GGCGCGCCGCCATTATTTTTTTG	Amplification of RiboCas backbone fragments (designed by Terry Bilverstone)
Ribo1T_R	TTTATCCTGCAGGGGGCCCGATCGGTCTTGC	
Ribo2T_F	CCGATCGGGCCCCCTGCAGGATAAAAAAATTG	

Ribo2T_R	GAAAACTCCTCCTTAAGATTTATATATG	
1341TA1_F	ATCTTAAGGAGGAGTTTTCGTCGACAGACAAGTAAAAAACCACTCGTTTTAGAGCTAGAAATAGCAAGTT	Single sgRNA TargetAID vector targeting gene 1341.
1341TA1_R	AACTTGCTATTTCTAGCTCTAAAACGAGTGGTTTTTTACTTGTCTGTCGACGAAAACTCCTCCTTAAGAT	
1341TAMulti1_F	ATCTTAAGGAGGAGTTTTCGTCGACACAAGTAAAAAACCACTCTGGTTTTAGAGCTAGAAATAGCAAGTT	Multiplex TargetAID vector targeting gene 1341.
1341TAMulti1_R	GCGCAAAAGAGTGGTTGGGAGACGTCCAAAAATCTATATTTTTTCTAACTGATTTAATTATAACCATAAATTTACTAGCTTTC	
1341TAMulti2_F	GATTTTTGGACGTCTCCCAACCACTCTTTTGCGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC	
1341TAMulti2_R	TATTTTAACTTGCTATTTCTAGCTCTAAAACCTGCGCAAAAGAGTGGTTGGCTCGAGTACCATATTTATATTATCATATTTTTGCTAATTTTTAAAGTATTTAATATCTC	
rpsJ-F	GATCACAAGTTTCAGGACCTG	Screening primers targeting *rpsJ* in *C. difficile.*
rpsJ-R	GTCTTAGGTGTTGGATTAGC	
2301L-circchk_F1	AGCATTATTACCCCACTGG	Confirmation of phiCD2301 genome in CD2301L, including ends of prophage where they meet the
2301L-loc1_F1	CAAAACCTGTTTTTATTTTCTCTG	
2301L-loc1_F2	ACTCTTAGCATAATTTTCAGCAC	
2301L-loc1_R1	TTTAACAGTTGCAACACCTTC	

2301L-loc1_R2	GATGGAGCAGGAACAGTC	bacterial genome.
2301Lq1_F1	GAACTTGTGTTAAAAGATATACCAGTC	
2301Lq1_F2	AGTCAGGAAAAATATGGCTTC	
2301Lq1_R1	ATGCTCCACTTTTTAAAATACTATG	
2301Lq1_R2	AAATCCGATTCCCAACAC	
2301L-loc2_F1	AATTCTTTATTCATCATATAACTCCTC	
2301L-loc2_F2	TACTATCAGGTTTCCCGTTC	
2301L-loc2_R1	AGCAGAGTTAAAAGCATTGTG	
2301L-loc2_R2	CAATATAACAAATCTCTTGCACC	
2301L-loc2_F1	AATTCTTTATTCATCATATAACTCCTC	Confirmation of prophage location in the CD2301L genome.
2301L-loc2_R1	AGCAGAGTTAAAAGCATTGTG	

2.3 General molecular biology techniques

2.3.1 Genomic DNA (gDNA) extraction

2.3.1.1 Bacteriophage DNA

Phage gDNA extractions were conducted using the Norgen BioTek Phage DNA Isolation kit and 20ml of fresh phage preparations (pfu in excess of 10^8/ml). The manufacturer's instructions were used with the following amendments to improve the quality and quantity of DNA extracted:

- Step 1:
    - Incubation at 65°C was extended to 20 mins from 15 mins

- Step 2:
    - The DNA was combined and bound to 6 columns total rather than 1 column per ml of starting volume
    - Centrifugation at 6000xg was undertaken for 30 secs instead of 1 min

- Step 4:
    - The 6 DNA columns used were combined into 3 samples by eluting one column using 35µl of nuclease-free water, the product of which was used to elute a second column in order to concentrate the sample. The three elutions were then combined into a single sample for testing.

2.3.1.2 Bacterial DNA

Bacterial gDNA was extracted from at least 5ml of overnight culture using the DNeasy Blood and Tissue Kit Gram Positive Bacteria protocol (Qiagen).

2.3.2 Oligonucleotide design

Oligonucleotide primers for restriction/ligation construction of plasmids were manually designed using Benchling.com to visualise the target DNA sequence. Primer melting temperature, secondary structures and primer dimer potential were assessed using a combination of OligoEvaluator™ (http://www.oligoevaluator.com/LoginServlet) and the NEB online Tm calculator (https://tmcalculator.neb.com/) tools. Primers suitable for a HiFi reaction were designed using the NEBuilder online tool (https://nebuilder.neb.com/).

The accepted oligonucleotides were synthesised to a concentration of 100µM by Sigma-Aldrich or Integrated DNA Technologies.

2.3.3 Amplification of DNA using polymerase chain reaction (PCR)

10µM working stocks of designed primers (section 2.3.2) were made for daily use and stored at -20ºC. PCR reactions were undertaken using either Dream*Taq* Green PCR Master Mix (2X, Thermo Scientific) or Q5® High Fidelity 2X Master Mix (New England Biolabs). Both reagents contain all required components for the PCR, and so only the appropriate primers, DNA template and distilled water were added to the reaction mix. The Dream*Taq* Green Master Mix also contains a gel loading dye for subsequent gel electrophoresis. Reactions were undertaken on an Eppendorf Mastercycler nexus X2e using the protocols detailed below. Annealing temperatures were adjusted to reflect the melting temperatures of the primer pairs in use.

Table 2.5 - Q5-mediated PCR protocol used for vector component amplification

Reaction mix	**Quantity per reaction**
Mastermix	10µl
FWD primer	1µl
REV primer	1µl
DNA template	1µl
dH_20	7µl

Step	**Temp (ºC)**	**Time**	**No. of cycles**
Initial denaturation	98ºC	30 secs	1
Denaturation	98ºC	30 secs	30
Annealing	Variable	30 secs	
Extension	72ºC	30 secs	
Final extension	72ºC	2 min	1

Table 2.6 - Q5-mediated colony PCR protocol

Reaction mix	**Quantity per reaction**
Mastermix	10µl
FWD primer	1µl
REV primer	1µl
DNA template	1µl
dH_20	7µl

Step	**Temp (ºC)**	**Time**	**No. of cycles**
Initial denaturation	98ºC	10 mins	1
Denaturation	98ºC	30 secs	30
Annealing	Variable	30 secs	
Extension	72ºC	2 mins	
Final extension	72ºC	10 mins	1

Table 2.7 - DreamTaq-mediated colony PCR protocol

Reaction mix	**Quantity per reaction**
Mastermix	10µl
FWD primer	1µl
REV primer	1µl
Colony resuspension	1µl
dH_20	7µl

Step	**Temp (ºC)**	**Time**	**No. of cycles**
Initial denaturation	95ºC	10 mins	1
Denaturation	95ºC	30 secs	30
Annealing	Variable	30 secs	
Extension	72ºC	3 mins	
Final extension	72ºC	10 mins	1

2.3.4 Agarose gel electrophoresis

Molten agarose was made by adding 1% (w/v) agarose powder (Sigma) to Tris-Acetate-EDTA buffer (BioRad - 50X TAE solution diluted to 1X

concentration using distilled water). SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific) was added to the molten agarose (1:10,000 final concentration) and the mixture poured into an agarose gel mould with a well comb to set. Once set, 4µl of 2-log DNA ladder (New England Biolabs) was added to one well for comparison purposes. Q5-generated PCR products were mixed with 6X gel loading dye (New England Biolabs) and each sample added to the remaining wells in the gel (Dream*Taq* products were added straight to the wells with no dye addition). DNA fragments were subjected to 100-120V for 30-50 mins allowing sufficient separation for visualisation and excision. Once completed, the gels were imaged using a Gel Doc™ XR+ Gel Documentation System (BioRad). Images were processed using ImageLab (BioRad).

2.3.5 Extraction of DNA from agarose gel

Once bands of the correct size for the expected DNA fragment were identified using the Gel Doc™ XR+, the gel was transferred to a UV transilluminator. The DNA bands on the gel fluoresced under UV light, allowing excision of the target band using a sterile scalpel. The gel fragment was treated using a QIAquick Gel Extraction kit (Qiagen). For this, 3 volumes of dissolution buffer was added to the excised band and incubated for 10 mins at 50-55°C. Once dissolved, 1 volume of isopropanol was added and the mixture vortexed before being added to a column and centrifuged (16,000xg, 1min) to bind the DNA. The column was washed and centrifuged several times using the provided wash buffers and the DNA eluted using 12-15µl of nuclease-free water. DNA samples were stored at -20°C until required.

2.3.6 PCR clean-up

For PCR products which did not require separation by electrophoresis, undesirable compounds were removed using a QIAquick PCR Purification kit (Qiagen) using the manufacturer's instructions. In brief, 5 volumes of PCR binding buffer were added to each sample and mixed thoroughly before addition to a binding column. The column was centrifuged at 16,000xg for 1min to capture the DNA before being

washed several times using the provided wash buffers. DNA was eluted using 12-15µl of nuclease-free water and stored at -20ºC until required.

2.3.7 Preparation of plasmid DNA from E. coli strains

E. coli strains harbouring the required plasmid were treated using the QIAprep Spin Miniprep kit (Qiagen) as per the manufacturer's instructions. In brief, 1-5ml of overnight culture was pelleted by centrifugation and the supernatant discarded. The pellet was re-suspended and lysed using the provided reagents before being passed through a binding column. The bound DNA was washed and the column eluted using 12-15µl of nuclease-free DNA. Samples were stored at -20ºC until required.

2.3.8 Restriction enzyme digests

Restriction enzymes and their associated buffers were purchased from New England Biolabs. 1µg of target DNA was added to a 1.5ml Eppendorf tube with 2µl of the required enzyme (2µl each if two enzymes used) and the recommended amount of buffer. Reaction volumes were made up to 20µl using distilled water and incubated at 37ºC overnight.

Digested plasmids were purified by agarose gel electrophoresis, the correct band excised and the DNA extracted for use. The digested insert DNA was subjected to a PCR clean-up.

2.3.9 RNA extraction

All pipettes, surfaces and gloves were pre-treated with RNase Zap (Invitrogen), and only filter tips were used for pipetting of solutions. Microcentrifuge tubes were either double autoclaved in advance or were taken from sterile RNase- and DNase-free supply provided in the relevant kits.

Three cultures of CD2301L were created and incubated overnight at 37ºC in the anaerobic cabinet, after which 1% inoculum sub-cultures were made and allowed to grow for 6-7 hours in anaerobic conditions.

Once the OD_{600nm} reached 0.8-1.2, three 2ml samples were taken and placed into 4ml of RNA Protect (Qiagen) and allowed to rest at room temperature for 5 mins. The samples were centrifuged at 5,000xg, 4ºC for 10 mins, the supernatant removed, and the pellet frozen at -80ºC until needed. One sample from each replicate was used for RNA extraction, with the other two samples stored for repeat work, if required.

To extract the RNA, the pellets were removed from the freezer and allowed to defrost on ice before being re-suspended in FASTRNA Pro Blue solution (MP Biomedicals). The suspension was transferred to the provided matrix lysis tube and homogenised at 4000xg for 45 seconds using a Bertin Technologies Precellys 24 tissue homogeniser. Once finished, the samples were centrifuged at 13,000xg, 4ºC for 15 mins and then transferred to clean 1.5ml Eppendorf microcentrifuge tubes to rest at room temperature for 5 mins. 300ml of ice cold chloroform was added to each tube in the fume hood, vortexed for 6 secs, and allowed to rest at room temperature for 5 mins to dissociate any nucleoprotein complexes that may be present. This was followed by another centrifugation under the same conditions as previously. The upper phase of the solution was carefully transferred to a fresh microcentrifuge tube containing 500µl of 100% ethanol using a 200µl pipette, mixed through inversion, and then incubated at -80ºC for 1-2 hours to allow precipitation of RNA.

Following the incubation, the samples were spun at 13,000xg, 4ºC for 15 mins, the supernatant removed, and the pellet washed using 70% ethanol made with DEPC water (Thermo Fischer Scientific). The sample was then spun again under the same conditions, after which the supernatant was removed and the pellet allowed to air dry for no longer than 30 mins to fully remove any remaining ethanol, before being re-suspended in 50µl of DEPC water. The sample was stored at -80ºC to maintain RNA stability until needed.

The next day, 1µl of each sample was removed to act as a negative control. The remainder was treated with 5µl of Ambion™ Turbo DNase buffer and 4µl of Ambion™ Turbo DNase and incubated at 37ºC for 30 mins to remove any potential remaining gDNA. The samples were then treated using the RNeasy Mini kit (Qiagen) to remove any further impurities. In brief, this consisted of each sample being treated with 350µl of the proprietary RLT buffer to which 3.5µl of 2-mercaptoethanol had been added, as well as 200µl of 100% ethanol. The solution was mixed by inversion and applied to an RNeasy spin column and centrifuged at 9,500xg for 15 seconds. The flow-through was discarded and the column washed twice with 500µl of buffer RPE with further spin steps in between. The column was then further spun at 9,500xg for 2 mins to dry the column and remove any remaining ethanol. To elute, 17µl of DEPC water was added to the column, incubated at room temperature for 3 mins and the column spun at 9,500xg for 1 min. This was repeated twice more to give an approximate elution volume of 50µl. The sample elutions were then separated into 10µl aliquots and stored at -80ºC.

To check the quality of the RNA samples, the 1µl of non-DNase treated samples and 1µl of the treated samples were used as template for a DreamTaq PCR reaction using primer pairs 2301Lq1_F1/R1 (1kb fragment) and rpsJ-F/R (200bp fragment) to confirm full DNA clearance. The PCR products were then subjected to electrophoresis using a 1% (w/v) agarose gel. Any treated samples still showing gDNA contamination were re-processed from the DNase step and tested again.

To measure the concentration and obtain an RNA Index Number (RIN), the samples were analysed using the Agilent 2100 Bioanalyzer. An RIN above 8 was considered suitable for RNA sequencing.

2.4 General microbiology techniques

2.4.1 Media and buffers

Medium/buffer	Constituents	Quantity (g/L)
Brain Heart Infusion Supplemented (BHIS) Broth	• BHI • Yeast Extract • L-Cysteine Hydrochloride Monohydrate (Fisher Scientific)	37.00 5.00 1.00
BHIS Agar	• BHI Agar • Yeast Extract • L-Cysteine Hydrochloride Monohydrate	47.00 5.00 1.00
BHIS Top Agar (0.5 %)	• BHI • Yeast Extract • L-Cysteine Hydrochloride Monohydrate • Agar Bacteriological No. 1	37.00 5.00 1.00 5.00
BHIS Top Agar (0.2 %)	• BHI • Yeast Extract • L-Cysteine Hydrochloride Monohydrate • Agar Bacteriological 1	37.00 5.00 1.00 2.00
BHIS sample enrichment	• BHI • Yeast Extract • L-Cysteine Hydrochloride Monohydrate • 0.1% taurocholate (Sigma, catalogue number 91223) • 1% $MgCl_2$ (Sigma, catalogue number M8266)	37.00 5.00 1.00
Lysogeny Broth (LB)	• Tryptone • Yeast Extract • Sodium Chloride	10.00 5.00 10.00

LB agar	• Tryptone • Yeast Extract • Sodium Chloride • Agar Bacteriological No. 1	10.00 5.00 10.00 15.00
Phosphate Buffered Saline (PBS), pH 7.4	• NaCl • KCl • $NaHPO_4$ • KH_2PO_4	8.00 0.20 1.44 0.24
Colony PCR buffer	• 20mM Tris-HCl • 2mM EDTA • 1.2% Triton-X (v/v)	2.42 0.74

2.4.2 Antibiotic supplements and inducing chemicals

Antibiotic supplements and inducing chemicals were used in the concentrations listed in Table 2.8.

Table 2.8 - Antibiotic supplements and inducing chemicals used in this study

Antibiotic	**Stock Conc. (mg/ml)**	**Dissolved in**	**Working Conc. (µg/ml) (*E. coli*)**	**Working Conc. (µg/ml) (*Clostridia*)**
Chloramphenicol (Cm)	25	100% Ethanol	12.5 (broth) 25 (plates)	
Anhydrotetracycline (ATc)	5	70% Ethanol	10	10
Thiamphenicol (Tm)	15	50% Ethanol		15

2.4.3 Growth conditions

2.4.3.1 *E. coli* strains

E. coli strains were grown overnight at 30-37ºC either in liquid LB media or on a LB agar plate. Liquid cultures were additionally subjected to agitation at 160-200 rotations per minute (rpm). Where appropriate, selective antibiotic supplements were added, most commonly chloramphenicol (CM), the concentration of which varied depending on the growth method (see Table 2.8).

2.4.3.2 *C. difficile* strains

C. difficile strains were grown at 37ºC in anaerobic conditions (80 % N_2, 10 % CO_2 and 10 % H_2) in a Don Whitley MACS MG-1000 anaerobic workstation, either in liquid BHIS media or on plates containing BHIS agar supplemented with a *C. difficile* selective supplement (CC) consisting of 250µg/ml D-cycloserine and 8µg/ml cefoxitin (Oxoid). If required, liquid overnight cultures were sub-cultured into fresh BHIS media using a 1% (v/v) inoculum.

2.4.4 Strain storage and revival

2.4.4.1 *E. coli*

E. coli strains were stored in Microbank™ Long Term Bacterial and Fungus Storage System tubes (Pro-lab Diagnostics) into which 700µl of the appropriate culture was pipetted and the tube shaken to allow thorough mixing of the cells with the beads. Once mixed, the liquid was removed and the tube labelled and stored in the -80ºC freezer. When required, a bead was removed from the appropriate stock and placed in liquid LB media containing 12.5µg/ml CM. This culture was incubated in a shaker overnight at 37ºC in aerobic conditions prior to use.

2.4.4.2 *C. difficile*

All strains of *C. difficile* were stored at -80ºC in 1ml aliquots containing 800µl of bacterial culture and 200µl of 50% glycerol solution. On removal from storage, a quantity of the frozen culture was scraped from the tube

using a sterile 10µl loop and streaked onto a BHIS-CC agar plate. The plate was then incubated overnight at 37ºC in the anaerobic cabinet.

2.4.4.3 Bacteriophages

Filter-sterilised bacteriophage preparations were stored in sterile universals at 4ºC. No revival was required, but the preparations were periodically refreshed using the methods outlined in section 2.8.

2.4.5 Preparation of E. coli chemically competent cells

Preparation of chemically competent *E. coli* strains was undertaken by inoculating 5ml of LB-CM media with the required *E. coli* strain. This culture was incubated overnight at 37ºC. The following day 1ml of culture was sub-cultured at 1% inoculum into 100ml of fresh LB media in a 500ml conical flask. The sub-culture was allowed to grow at 37ºC (200rpm agitation) to an OD_{600nm} between 0.3-0.5 before the flask was placed on ice for 15 mins. The culture was then aliquoted into six pre-chilled 50ml falcon tubes and centrifuged at 2700xg, 4ºC for 10 mins. The supernatant was carefully removed and discarded, and the pellets re-suspended in 10ml of 0.1M $CaCl_2$/20% glycerol (v/v) per falcon tube before being placed on ice for another 15 mins. Centrifugation and re-suspension was repeated once more, followed by combination of the suspensions into two falcon tubes (30ml in each). The falcons were chilled again for 15 mins, centrifuged and the supernatant removed. Final re-suspension was undertaken in 500µl of 0.1M $CaCl_2$ with 20% glycerol (v/v) before the mixture was separated into 40ul aliquots in pre-chilled 1.5ml Eppendorf tubes. The cells were stored at -80ºC until required.

2.4.6 Transformation of E. coli chemically competent cells

Plasmid vectors were transformed into *E. coli* cells following New England Biolabs' High Efficiency Transformation protocol (New England Biolabs, 2020). In brief, 50µl of chemically competent *E. coli* cells were thawed on ice for 10 mins. Once thawed, approximately 5ng of ligation mixture was added to the cells, inverted to mix, and incubated on ice for 30 mins. The mixture was heat-shocked at 42ºC for 30 seconds and

950µl of room temperature SOC media (Sigma) added for recovery. The mixture was incubated in a 37ºC shaker for at least 1 hour before plating onto 3 pre-warmed LB-CM plates per vector - 100µl neat, 100µl 1/10 dilution, and finally the remaining cells pelleted and re-suspended in 100µl SOC media. These plates were incubated at 30ºC for 48 hours. Once grown, individual colonies were patch-plated onto a fresh LB-CM agar plate and incubated at 30ºC for a further 24 hours prior to storage at 4ºC. Colonies were screened for the presence of a correct plasmid through PCR and agarose gel electrophoresis before being submitted for Sanger sequencing to confirm the correct construction of the vectors.

2.4.7 Conjugation of plasmid DNA into C. difficile *strains*

An overnight culture was made of sExpress *E. coli* harbouring the relevant plasmid in 5ml of LB-CM media, as well as a 1ml BHIS-only overnight culture of CD2301L. BHIS non-selective plates were also added to the anaerobic cabinet to reduce.

The next day, 1ml of sExpress cells from each overnight culture was centrifuged at 1,500xg and the supernatant removed. The pellets were washed twice with 400µl of PBS and centrifuged in between, after which the pellet was transferred to the anaerobic cabinet. The *E. coli* cells were re-suspended in 200µl of the CD2301L overnight culture and 20µl spots of the suspension added to a BHIS plate for each vector (hereafter referred to as mating plates). The plate was incubated at 37ºC for 24 hours. Five BHIS-CC plates per vector containing thiamphenicol (Tm) were added to the anaerobic cabinet to reduce overnight.

After 24 hours, the growth was collected from each mating plate using a 10µl loop and re-suspended in 500µl of PBS. A 100µl aliquot of this re-suspension was plated onto each of the 5 BHIS-CC-Tm plates and the plates incubated for 48-72 hours. A selection of transconjugant colonies was then re-streaked onto a second BHIS-CC-Tm plate and incubated overnight. Any colonies which grew during the second subculture on

BHIS-CC-Tm were deemed to be *C. difficile* isolates in which the plasmid had been successfully conjugated.

Once putative transconjugants were confirmed, these were further re-streaked onto BHIS-CC-Tm plates containing theophylline (Th – final concentration 5mM) in order to induce plasmid-borne CRISPR activity. These induction plates were incubated overnight. Following incubation, single colonies were screened by colony PCR using primers flanking the left and right homology arms on the bacterial chromosome. Putative mutants were then taken forward for plasmid loss.

2.4.8 Removal of conjugated plasmid DNA

Successful transconjugants confirmed by colony PCR were streaked from stocks and used to create a 20ml BHIS culture which was incubated anaerobically overnight. A 1% inoculum sub-culture was made the next day from the overnight culture and left to incubate anaerobically for 24 hours. Sequential sub-cultures were made over 5 days before 100µl of the final sub-culture was spread using a sterile angled spreader onto a pre-reduced BHIS-CC agar plate and incubated anaerobically for 24 hours. Colonies from this plate were re-streaked onto a fresh BHIS-CC plate as well as a BHIS-CC-Tm plate.

After 24 hours, the two plates were compared to determine which, if any, of the colonies had grown. Successful growth on the former but not the latter indicated the loss of the plasmid (containing the antibiotic resistance gene) from the bacteria. Mutation of the genomic DNA was then confirmed through PCR, agarose gel electrophoresis, and DNA sequencing. Stocks were made of confirmed mutants and stored at -80°C for onward testing.

2.4.9 Fluorescence-activating and absorption-shifting tag protein (FAST) assay

Biological replicate cultures of CD2301 containing integrase expression vectors were made in 20ml of BHIS media, alongside the following controls:

- pMTL8315y-Ccarb-v2_FAST in CD2301 as a positive control,
- pRPF185 and pMTL83151 in CD2301 as negative controls,
- promoter-only FAST plasmids in CD2301 to check promoter activity,
- empty CD2301 cells as a true negative.

All overnight cultures were sub-cultured at 1:100. Once the sub-cultures had grown to approximately OD_{600nm} of 0.4, each of the expression vector cultures were treated with the appropriate inducing chemical in the amounts detailed in Table 2.9. Test concentrations were selected in line with those used in Cañadas et al. (2019) and Fagan and Fairweather (2011).

Table 2.9 - Concentrations of inducers tested during FAST assay of integrase expression vectors

P_{fdx-E} vector		**ATc vector**	
Concentration (mM)	**Amount of theophylline added (µl)**	**Concentration (ng/ml)**	**Amount of ATc added (µl)**
0.5	2.06	100	0.22
1	4.13	200	0.55
2	8.27	500	1.1
5	20.7	1000	2.2

The positive controls were treated with 2mM of theophylline or 500ng/ml of ATc only. The cultures were allowed to grow for another 3.5 hours

before being centrifuged, the supernatant removed, and the pellet stored at -20ºC. The next day, pellets were re-suspended in PBS to an OD_{600nm} of 1.0. 1ml of this suspension was then centrifuged again and the pellets washed twice with PBS to remove any potential fluorescing components before being re-suspended to a final volume of 250µl. The suspensions were added to a 96-well, flat-bottomed, black-walled Greiner plate, treated with 50µl of a 10µM TFAmber labelling solution (The Twinkle Factory Ltd.) and transferred to the BMG Labtech Clariostar microplate reader. The fluorescence of each well at excitation wavelengths between 499nm and 558nm was recorded to determine the degree of FAST gene expression of each vector.

2.5 Vector creation

2.5.1 Primer design

2.5.1.1 CRISPR vector primer design

A homology arm upstream of the target gene was designed which included an AatII restriction site in the forward primer and bookmark BM6 in the reverse primer (Seys et al., 2020). Primers were also designed for a homology arm downstream of the gene with the forward primer containing the complementary sequence for BM6 and the reverse primer containing an AscI restriction site.

Primers to amplify single guide RNAs (sgRNAs) were also designed using three different 20nt sequences identified as suitable CRISPR targets by Benchling.com. The primers contained a SalI restriction site in the forward primer and an AatII site in the reverse primer, allowing for ligation alongside the homology arms. The sgRNAs were amplified by PCR.

2.5.1.2 Integrase expression vector primer design

Primers were designed for the construction of vectors using two different backbones – pMTL83151 and pRPF185.

For the pMTL83151 vector, primers were designed to amplify the P_{fdx-E} promotor from pRECas1_MCS. The forward primer for this fragment contained an NdeI restriction site to allow ligation into the pMTL83151 MCS region, while the reverse primer contained an overhang homologous to the integrase gene. The second fragment primers were designed to amplify the integrase gene from phiCD2301, with the forward primer containing the complementary overhang to P_{fdx-E} and the reverse primer containing an NheI restriction site.

Primers to amplify the integrase gene for incorporation into the pRPF185 plasmid were initially designed with a SacI site in the forward primer and BamHI site in the reverse primer. However, following further work, new primers were designed to amplify the ribosome binding site (RBS) region upstream of the Himar1C9 gene in pMTL-CW21 with the SacI site and an overhang corresponding to the integrase gene, and the integrase gene forward primer was re-designed to contain the complementary overhang.

Two different variants of primers to amplify the FAST gene from pJ8221 were designed – the P_{fdx-E} vector primers were flanked by NdeI and NheI restriction sites, while the ATc vector primers contained a BamHI site in each primer.

2.5.2 Vector construction

2.5.2.1 CRISPR vectors

DNA fragments for the appropriate vector, including homology arms and single guide RNAs, were amplified by Q5 PCR. The homology arms were then combined using a gene splicing by overlap extension (SOE) PCR (Horton et al., 2013) to form a homology cassette consisting of DNA homologous to the upstream and downstream sections of the phage genome but with the target gene replaced by the BM6 “bookmark” spacer to allow future CRISPR targeting. A stock of pRECas1_MCS (RiboCas) was extracted from cultures of the appropriate *E. coli* storage strain, and

all fragments and vector samples digested using 2µl each of the enzymes listed in Table 2.10.

Table 2.10 - Enzymes used to digest DNA for creation of CRISPR vectors

Component	**Enzyme 1**	**Enzyme 2**
RiboCas	AscI	SalI
Homology arms	AscI	AatII
sgRNAs	AatII	SalI

The digested RiboCas fragment was purified using gel electrophoresis and the remaining DNA fragments were subjected to a PCR clean-up. Once pure, a total of 100ng of digested plasmid DNA was added to a 1.5ml Eppendorf tube along with the digested homology arm cassette (1:3 ratio) and 1µl of the appropriate single guide RNA. 2µl of T4 ligase and 4µl of T4 ligase buffer were added and the reaction mix volumes made up to 50µl with distilled water before incubation on ice overnight. The next day, ligation mixes were transformed into chemically competent XL1 Blue *E. coli* cells, plated and incubated at 30ºC for 48 hours.

2.5.2.2 Expression vectors

The pMTL83151 and pRPF185 vector backbones were extracted from their respective storage *E. coli* strains and digested using NdeI/NheI and SacI/BamHI respectively. The digested plasmid was then purified by gel electrophoresis, ready for vector construction.

For pSJK1, primers were designed using Benchling to amplify the $P_{fdx\text{-}E}$ promoter from RiboCas with an upstream SacI restriction site and an overhang corresponding to the native phiCD2301 integrase gene. The integrase gene primers contained the complementary overhang and a downstream BamHI site for PCR amplification of the appropriate fragment from CD2301L gDNA. The two fragments were fused into a single insert using a SOEing PCR which was then digested with NdeI and NheI and ligated into the digested pMTL83151 backbone. The ligation

mix was transformed into chemically competent XL1 Blue *E. coli* cells for cloning.

For pSJK2, only the integrase fragment was initially required for amplification, and so primers were designed which were flanked by a SacI and BamHI restriction site. The resultant DNA fragment was digested with these enzymes and ligated into the pRPF185 vector backbone, with transformation being undertaken into chemically competent XL1 Blue *E. coli* cells. Subsequently, a further ribosome binding site was required upstream of the integrase gene fragment. Primers were designed to amplify this site from pMTL-CW21 and contained a SacI restriction site in the forward primer and an overhanging section in the reverse primer corresponding to the integrase gene. A new forward primer for the integrase gene was designed to contain the complementary overhang, and the new fragment amplified. The two fragments were again fused together using a SOEing PCR and ligated into the pRPF185 backbone as described previously.

FAST gene fragments were also amplified using plasmid pJ8221, one flanked by NheI sites and the other by BamHI sites. pSJK1 and pSJK2 were digested using the appropriate enzymes and the FAST genes inserted to form pSJK1-FAST and pSJK2-FAST. Further FAST genes were later required which could be fused downstream of $P_{fdx\text{-}E}$ and to the new ribosome binding site in the pSJK2-FAST variant in order to use promoter-only vectors as controls in a FAST assay. In both cases, primers were designed with the appropriate overhangs, the fragments subjected to a SOEing PCR, and ligated into their respective backbones to form pRiboFAST-RL and pATcFAST_RL.

2.5.2.3 NEB HiFi assembly of vectors

Primers designed using NEBuilder were used to amplify the same fragments specified above for pRiboFAST and pATcFAST. A 0.5µl aliquot of each fragment was added to a reaction mix containing 5µl of NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) with

1µl of the sgRNA fragment in the case of CRISPR-Cas9 vectors. The reaction mix volume was made up to 10µl using distilled water and incubated at 50°C for 1 hour before being transformed into an appropriate *E. coli* cloning strain.

2.6 Submission of samples for Whole Genome Sequencing (WGS)

2.6.1 Illumina DNA sequencing

Prepared gDNA samples were measured for quantity using a Qubit dsDNA HS (High Sensitivity) Assay Kit (Thermo Fisher Scientific) and analysed using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific), with a Thermo Scientific™ NanoDrop™ Lite spectrophotometer used to determine the A260/A280 ratio. Approximately 1µg of suitable quality DNA was then submitted to the Deep Seq facility at the University of Nottingham Medical School who prepared the DNA libraries using the Nextera DNA Flex kit (Illumina Inc.), followed by paired-end Illumina sequencing using a MiSeq600 cycle kit with approximately 250x coverage.

2.6.2 Oxford Nanopore Technologies (ONT) DNA sequencing

3µg of prepared gDNA at a concentration of 100ng/µl was run on an agarose gel to check for the presence of RNA. Clear samples were submitted to DeepSeq for long read sequencing using ONT methods. The following details of the sample preparation and processing were provided by DeepSeq:

"The DNA samples were quantified by DeepSeq using the Qubit 4 Fluorometer (Thermo Fisher Scientific) and the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific; Q32853). Sequencing libraries were prepared using the Native barcoding genomic DNA protocol (Oxford Nanopore Technologies; Version: NBE_9065_v109_revS_14Aug2019). This protocol uses the Ligation Sequencing Kit (Oxford Nanopore Technologies; SQK-LSK109) and the Native Barcoding Expansion 1-12 kit (Oxford Nanopore Technologies; EXP-NBD104). Barcoded samples

were then pooled for the final sequencing adapter ligation. All purification steps were performed using AMPure XP beads (Beckman Coulter; A63882). The final barcoded library pool was loaded onto a MinION flow cell (Oxford Nanopore Technologies; FLO-MIN106 R9.4.1) and run on the GridION X5 Mk1."

2.6.3 RNA sequencing

Aliquots of the required RNA samples (2µg each) were submitted to DeepSeq for RNA sequencing. The following details of the methodology were provided by DeepSeq:

"RNA concentrations were measured using the Qubit Fluorometer and the Qubit RNA BR Assay Kit (Thermo Fisher Scientific; Q10211) and RNA integrity was assessed using the Agilent 4200 TapeStation and the RNA ScreenTape Assay (Agilent; 5067-5576 and 5067-5577). Ribodepletion was performed on 1 ug of total RNA, using the Qiagen FastSelect – 5S/16S/23S ribodepletion kit (Qiagen; 335925) using the protocol: FastSelect – 5S/16S/23S with NEBNext Ultra II Directional Library Prep Kit.

Indexed sequencing libraries were then prepared using the NEBNext Ultra II Directional RNA Library Preparation Kit for Illumina (NEB; E7760) and NEBNext Multiplex Oligos for Illumina (96 Unique Dual Index Pairs) (NEBNext; E6440). Libraries were quantified using the Qubit Fluorometer and the Qubit dsDNA HS Kit (Thermo Fisher Scientific; Q32854). Library fragment-length distributions were analysed using the Agilent 4200 TapeStation and the Agilent High Sensitivity D1000 ScreenTape Assay (Agilent; 5067-5584 and 5067-5585). Libraries were pooled in equimolar amounts and final library quantification performed using the KAPA Library Quantification Kit for Illumina (Roche; KK4824). The library pool was sequenced on the Illlumina NextSeq500 on a NextSeq500 High Output 150 cycle kits (Illumina; FC-404-2002), to generate over 10 million pairs of 75-bp paired-end reads per sample."

2.7 Bioinformatics tools

2.7.1 Data analysis and presentation

Data generated from experiments was added to Microsoft Excel for tabulation before being moved to GraphPad Prism for analysis and visualisation. Plasmid maps for publication were created using SnapGene®.

2.7.2 Sequencing of PCR fragments and plasmid DNA

Isolated DNA fragments and plasmids were submitted to either Eurofins TubeSeq service, Source Bioscience or Azenta Genewiz for Sanger sequencing. Benchling.com was used to visualise and analyse the results.

2.7.3 Assembly and annotation of sequencing data

2.7.3.1 Illumina sequencing

Assembly of sequencing results was completed using CLC Genomics Workbench v.20 (Qiagen). Reads were trimmed and the quality assessed using the pre-provided “Prepare Sequencing Data” workflow, followed by *de novo* assembly using the “De novo sequencing” workflow under the default settings. Assessment and further assembly of contigs into full genomes was undertaken manually using Benchling.com.

2.7.3.2 ONT sequencing

As per the Illumina sequencing, the ONT reads were trimmed and quality-checked using CLC Genomics Workbench v.20. Once completed, the trimmed reads were assembled using two different long read-specific programmes – Canu v.2.2 (Koren et al., 2017) and Flye v.2.9 (Kolmogorov et al., 2019). The assembled contigs were subjected to a “polishing” step using the NextPolish v. 1.3.1 software in which the Illumina reads were used to correct any discrepancies in the error-prone long read sequencing (Hu et al., 2020). The contigs generated from the two assemblies were then used in conjunction with Benchling.com to manually complete the assemblies into full genomes.

2.7.3.3 Annotation of genomes

For the novel isolations, phage-related contigs were subjected to a Python-based gene-calling and annotation pipeline called multiPhATE2 (Ecale Zhou et al., 2021; Zhou et al., 2019). This pipeline uses the Prodigal and Glimmer3 annotation algorithms but also includes the Phanotate algorithm (McNair et al., 2019) which has been optimised for use with small phage genomes. Phantome (Zhou, 2021) and pVOGs (Grazziotin et al., 2017) databases were included in the pipeline for local BLAST and HMM searching, and additional databases were downloaded including the NCBI Virus Genome and Protein databases (Hatcher et al., 2017), Swissprot (The UniProt Consortium, 2021), Carbohydrate Active Enzymes database (CAZy - Lombard et al., 2014) and VOGs (https://vogdb.org/). All databases downloaded were the latest versions available as of 5th Jan 2021, except the CAZy database which required a file from 2019 (CAZyDB.07312019.fa). BLAST searches of the NCBI NR and Uniprot web-interface databases were conducted separately due to the large size of these databases precluding download for local searching.

Bacterial assemblies were submitted to the online PATRIC portal (Davis et al., 2020) and to the NCBI Prokaryotic Genome Annotation Pipeline (PGAP - Tatusova et al., 2016) for annotation.

2.7.4 Phylogenomic analysis

The novel phage genome sequences were submitted to the online tool VIPtree (Nishimura et al., 2017) for phylogenetic comparison against dsDNA prokaryotic viruses in the RefSeq database. Direct comparisons were then undertaken using the VIPtree-generated phylogenetic tree, dot plot and alignment results. Additional comparisons were undertaken using the progressiveMauve alignment software (Darling et al., 2010) in order to establish the specific genes that were shared between the novel phages and those of their closest identified relatives.

2.7.5 Protein analysis

Protein structure analysis was conducted using ColabFold (Mirdita et al., 2022), a Colab release of the protein-folding programme AlphaFold 2 (Jumper et al., 2021). Sequences were analysed using the default settings, including an auto-generated multiple sequence alignment. The AlphaFold Database (https://alphafold.ebi.ac.uk/ - Varadi et al., 2022) was used to obtain the X-ray crystallisation result for the λ repressor dimer generated by Stayrook et al. (2008) - structure 3bdn - while a monomer version was obtained by submitting the λ repressor DNA sequence to ColabFold.

Quality control of the AlphaFold 2 predictions was undertaken using analysis generated by the same programme in the form of pLDDT values and predicted alignment errors (PAE), and was further confirmed through the use of Ramachandran plots (Kumar and Arya, 2018). For the latter, the Protein Data Bank (PDB) file generated by AlphaFold 2 was input into the Ramachandran Plot Server (https://zlab.umassmed.edu/bu/rama/) using the default no-glycine settings. PyMOL v.2.5.4 (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) was also used to overlay the 5 possible alternative structures for each molecule that was generated by AlphaFold2 to determine a consensus structure.

Finally, Phobius (Käll et al., 2004) was used on its default settings to determine whether any of the sequences contained transmembrane or signal peptide regions.

2.8 Phage-related assays

2.8.1 Sewage enrichment

NB - *Preparation of sewage samples was undertaken in conjunction with my colleague Joanna Steczynska*

Fourteen samples were collected from seven Thames Water sewage treatment plants across Oxfordshire – Aylesbury, Swindon, Banbury,

Reading, Wargrave, Oxford and Didcot. Each site contributed a 500ml sample of sewage pre- and post- anaerobic digestion. Initially, aliquots of nine of these samples (post digestion samples from all sites, as well as pre digestion samples from Aylesbury and Oxford) were enriched through the addition of brain and heart infusion (BHI) dry media, 5mg/ml yeast extract, 0.1% (w/v) L-cysteine, 0.1% taurocholate and 1% $MgCl_2$ and incubated overnight. The samples were subsequently centrifuged at 10,000xg, 4°C for 10 mins to remove cell debris and the supernatant was filter-sterilised.

2.8.2 Plaque assay for determination of presence of phage in sample

To test for the presence of phage in the sample, *C. difficile* RT 027 strains stored in the University of Nottingham Synthetic Biology Research Centre (SBRC) culture collection were grown on BHIS-CC agar plates, incubated for 24-72 hours, and used to make an overnight culture in 20ml of fresh BHIS broth. A further 20ml of broth per strain was placed in the cabinet to reduce which was then used to make 1:100 sub-cultures the next day.

After 6 hours, a plaque assay was undertaken based on the double agar overlay method detailed by Clokie and Kropinski (2009). 200µl of the enriched sewage supernatant was added to a 15ml falcon tube, along with 1ml of the appropriate *C. difficile* sub-culture and 3ml of 50°C BHIS 0.5% (w/v) top agar. One sample was also made with no sewage as a negative control. These samples were poured onto pre-made BHIS-CC plates, allowed to set on the bench for 10 mins and then moved to the anaerobic cabinet for overnight incubation. The plates were removed from the cabinet the next morning and checked for circular zones of clearing (plaques) indicative of the presence of phage.

If zones of clearing were present, the plates were returned to the cabinet and up to three representative clear zones were scraped from the top agar layer using a pipette tip (avoiding the lower layer of agar). The

picked sample was re-suspended in 900µl of BHIS broth and incubated for one hour in the anaerobic cabinet to allow any phages to diffuse out of the agar. The broth was then filter-sterilised using a 0.22µm filter and stored at 4ºC until required. These samples were used for a further plaque assay as detailed above. If no plaques were obtained from the second assay, no further action was taken. If plaques were seen, the process was repeated twice more to ensure purity of any phage present. Where necessary, a 1/10 and 1/100 dilution was also used to allow individual plaques to be identified.

2.8.3 Titre increase of new phage samples

Due to the isolated phage samples being low in titre, further plaque assays were undertaken to increase the concentration. Plaque assays were completed as above but using BHIS 0.2% (w/v) agar instead of 0.5% to allow diffusion of the phage. To allow the determination of the point at which confluent lysis occurs, dilutions of the phage prep were used down to 10^{-4}.

Following overnight incubation, the plate showing the closest to confluent lysis was selected and the top layer of agar scraped into a 50ml falcon tube containing 10ml pre-reduced BHIS broth. The sample was allowed to incubate for a further hour in the anaerobic cabinet before being centrifuged at 10,000xg, 4ºC for 10 mins. The supernatant was filter-sterilised using a 0.22µm filter and stored at 4ºC.

To check the new titre of the preparation, a plaque assay using BHIS 0.5% (w/v) top agar was undertaken with appropriate dilutions (as low as 10^{-10} depending on the previous titre). Once incubated overnight, the plates were compared and the number of plaques on each plate counted. The approximate number of plaque forming units (pfu) per ml of sample was then calculated using the following equation:

$$pfu/ml = (no.\ of\ plaques\ x\ 5)\ x\ the\ dilution\ factor\ of\ the\ plate$$

The assays were repeated until the phage samples reached a titre of 10^8–10^{10} pfu/ml to ensure sufficient phage was available for ongoing characterisation experiments. All phage preparations were then stored at 4ºC until required.

2.8.4 Transmission electron microscopy (TEM) of new phage samples

Visualisation of the phage was undertaken with a Technai T12 BioTWIN Transmission Electron Microscope (100KV) using a Megaview II Soft Imaging System. To prepare the sample, 1.5ml of phage preparation was centrifuged at 21,000 xg, 4ºC, for 75 mins in 1.5ml Eppendorf tubes. 1.4ml of the supernatant was carefully discarded, ensuring that the pellet was not removed, and 1ml of 0.1M ammonium acetate was added. The sample was centrifuged again under the same conditions and these steps repeated twice more. Once complete, the sample was stored at 4ºC until required.

Once the samples were prepared, a drop of the phage prep was added to a 200 mesh copper grid (EM resolutions) for 10 mins. This allowed the phage particles to precipitate onto the grid for visualisation. The liquid was carefully wicked away using a section of Whatman filter paper held to the edge of the grid. 2% uranyl acetate was added to the grid surface as a negative stain and left for 30 seconds to 1 min, then was wicked away. The grid was transferred into the sample holder rod and inserted into the Tomograde Compustage for visualisation.

2.8.5 Determining the host range of bacteriophages

A complete list of all *C. difficile* strains in the SBRC culture collection was compiled (see Appendix 7.1). Each strain was streaked onto a BHIS-CC agar plate and incubated overnight at 37ºC in the anaerobic cabinet before being used to inoculate 10ml of BHIS broth for an overnight culture. A 1:100 sub-culture was made in pre-reduced BHIS media after 16-18 hours and allowed to incubate for a further 6-8 hours depending on growth rate. Once well-grown (in excess of OD_{600nm} 0.8), 1ml of the

culture was added to a 15ml falcon tube and removed from the cabinet. A 3ml aliquot of molten BHIS 0.5% (w/v) top agar was added to the falcon tube and the mixture poured onto a BHIS-CC plate. The top layer was allowed to set and the plate returned to the cabinet within 20mins. Aliquots were taken from the relevant phage preparations and a 1/10 dilution made for each phage. 20µl samples of neat and 1/10 dilutions of phages were spotted onto quadrants of the bacterial plates and allowed to absorb into the top agar layer. Plates were then incubated in the anaerobic cabinet overnight. After 24 hours, the plates were checked for zones of clearing to determine whether the phages were able to infect that bacterial strain.

2.8.6 Efficiency of Plating (EOP)

The EOP was used to discern the degree of infectivity of a strain in comparison to the isolation strain. For this, phage preparations were made as previously described and their titres adjusted to reflect the titer of the lowest preparation to allow for more direct comparisons of infectivity between phages. These preparations were subjected to a standard plaque assay using the strains identified as positive for infection (the propagating strains). Positive control plates using the isolation strains were also made. Following the assays, the efficiency of plating was determined using the following equation:

$$Efficiency\ of\ plating = \frac{titer\ derived\ from\ propagating\ strain}{titer\ derived\ from\ isolation\ strain}$$

2.8.7 Lysogen formation for long term storage

Three methods of testing for lysogen formation were conducted during this study.

First, a standard plaque assay was undertaken using BHIS 0.5% (w/v) top agar from which individual plaques were identified. Three plaques were selected and a pipette tip used to scrape the agar from the centre of the clear zone and streak it onto a fresh BHIS-CC plate. After overnight

incubation, 20 colonies were selected from the plate and re-streaked 3 times to purity. These colonies were then screened for the presence of phage in the genome using colony PCR.

Due to the plaques for these phages being routinely too small to sample reliably without also collecting uninfected lawn, another method was tested in which a plaque assay was conducted using just the neat plate, which routinely showed total cell death. The plate was incubated for 48 hours rather than 24, and colonies that grew after 48 hours were passaged three times and screened for phage.

Finally, a spotting technique was attempted in which bacterial culture was streaked onto a fresh BHIS-CC plate and 20µl spots of neat phage preparation added on top. After 24 hours, any visible colonies were sampled from the spotted areas, or if the plate was clear then a small plastic loop was used to sample the centre of the zone of clearing. The colonies or loop were then streaked onto fresh BHIS-CC plates and the same passage and screening process applied as normal.

2.8.8 Induction of phages from lysogens

A 20ml overnight culture of the lysogen containing the desired phage was made as previously described. After incubation, the culture was treated with 120µl of 0.5mg/ml Mitomycin C (Sigma-Aldrich) and left to incubate for a further 24 hours, after which it was centrifuged and filter-sterilised as per section 2.8.3. The titre was then increased until it reached 10^8-10^{10} pfu/ml.

2.8.9 Phage antimicrobial action assay

Biological triplicate cultures of the appropriate *C. difficile* strains were made in 20ml BHIS broth and incubated overnight. A 1% (v/v) inoculum sub-culture was created the next morning and allowed to grow to an $OD600_{nm}$ of 0.2-0.3 (approximately 10^7 cfu), at which point 200µl of culture were added to wells in a Greiner 96-well black-walled, flat-bottomed plate in technical triplicate. 20µl of phage dilutions were added

to each make MOIs of 0, 0.01, 0.1, 1 and 10 (again, in technical triplicate) and the plate sealed with a clear film to maintain the anaerobic environment. The plate was transferred to a Clariostar plate reader and the $OD600_{nm}$ of all wells measured every hour for 23 hours with agitation before each measurement.

Chapter 3

Isolation of Novel Bacteriophages Infecting the Clinically Relevant Ribotype 027 Strains

3 Isolation of Novel Bacteriophages Infecting the Clinically Relevant Ribotype 027 Strains

3.1 Introduction

Phage therapy requires suitable (i.e. obligatory lytic) phages to be available that target the bacterial strain responsible for a patient's infection. However, high phage specificity means that isolates acquired to treat a certain infection are unlikely to target many other strains. While this is beneficial in that it limits the degree of gut dysbiosis incurred during treatment, it has limited the usefulness of phage therapy in the past as it necessitated long wait periods for new phages to be identified for each new infection at a time when broad-spectrum antibiotics were regularly and quickly available (Summers, 2012). To combat this restriction, new phages must be isolated at every opportunity to create an established library for testing against new infections, thereby increasing the chance of a suitable phage already being available. This is especially the case for CDI given that no lytic phages have yet been isolated, hampering our ability to use phage therapy for treatment.

Phages are reliant on the presence of their hosts for propagation, and so phages infecting a particular species of bacteria are most likely to be found in areas where those host organisms are abundant. Methods of isolation therefore vary from strain to strain depending on the environment in which those bacteria live (e.g. water, soil or sediment - Clokie and Kropinski, 2009). In the case of *C. difficile* human-pathogenic ribotypes, the bacteria spread through the release of aero-tolerant spores in the patient's faeces which can then be transferred onto other surfaces and picked up by another host (Badger et al., 2012b). This makes sewage a good source for isolating both *C. difficile* bacteria and the phages which infect those strains (although slurry from known animal reservoirs such as pigs is also an option - Usui et al., 2017). In particular, sewage samples from treatment plants that provide services to hospitals are likely sources of *C. difficile* bacteria, as the sewage will contain

spores shed from active infections at the hospital. The sewage can then be incubated with an enrichment medium which contains compounds that encourage the growth of the target bacterial species. The addition of taurocholate to the enrichment medium for *C. difficile* encourages the germination of spores present in sewage samples (Shrestha et al., 2019), which facilitates the induction of any prophages within the genomes of those bacteria. All free phages within the medium can then be harvested through filtration. The supernatant containing these phages is tested against the target bacteria (in the case of phage therapy) or against a collection of strains (when isolating for a library) to determine whether there are any phages present which are capable of infecting the samples.

While historically *C. difficile* phages have been difficult to isolate (Hargreaves and Clokie, 2014; Horgan et al., 2010), improvements in methods over time have resulted in 34 *C. difficile* phage genomes being listed on NCBI as of July 2022. However, the lack of a known lytic phage still presents a barrier to the use of phage therapy to treat CDI. Goh et al. (2005) suggested that this lack may be due to the anaerobic, spore-forming nature of *C. difficile*, as phages require cells to be in a vegetative state to replicate themselves. If environmental populations of *C. difficile* largely comprise of these aero-tolerant spores which do not exhibit the SlpA protein (recently confirmed as the phage receptor binding protein - Kirk et al., 2017; Phothichaisri et al., 2018; Whittle et al., 2022) then lytic varieties of the phages would be hindered by a dearth of viable hosts. This would create a permanent selection pressure favouring temperate phages. However, the inherent difficulty of isolating these phages from the environment may also have provided a factor, driving researchers to target easier methods of isolating phages such as inducing existing prophages from *C. difficile* isolates. This method, while suitable for obtaining phages for study quickly, limits acquisitions to only temperate phages, therefore making it seem from the outside like lytic phages are a rarity.

Given the potential for the latter situation, coupled with the USFDA's 2012 requirement for therapeutic phages to be "natural" (Brüssow, 2012), an attempt to isolate lytic phages from the environment should always be considered in the first instance when searching for therapeutically-usable phages. This chapter therefore details the results of isolation experiments undertaken by challenging epidemic PCR ribotype 027 (RT027) strains of *C. difficile* with samples of sewage taken from wastewater treatment plants in the Thames Water region.

Please note that several elements of this work have been conducted in conjunction with my colleague Joanna Steczynska due to the size of the assays involved. These elements are specifically indicated, and work was split equally between us.

3.2 Results

3.2.1 Isolation of novel phages

NB – This section was completed in conjunction with my colleague Joanna Steczynska

For this study, 36 RT 027 strains of *C. difficile* (sourced from various locations) were chosen from the culture collection of the Clostridia Research Group (CRG) at the SBRC. RT 027 strains were responsible for a worldwide outbreak of CDI in the early 2000s (Alcalá Hernández et al., 2017) making them clinically relevant strains. However, the CRG does not currently hold examples of phages infecting these strains within its collection, and it would be beneficial to have phages for these strains to test for therapeutic purposes.

Of the 36 strains selected, 23 UK- and EU-sourced strains were initially tested by plaque assay against nine samples of filtrate taken from enriched sewage (Table 3.1). Four combinations produced plaques in both initial and confirmatory assays:

Table 3.1- Plaque assay results for isolation of novel bacteriophages from environmental sewage.

	Sewage sample								
Strain	AYL-DD	AYL-PD	OXF-DD	OXF-PD	SWI-PD	DID-PD	WAR-PD	RDG-PD	BAN-PD
R20291									
L9									
L10									
DH1834									
DH1466									
DH1858									
DH1751									
R24988									
R24694									
DH1342									
DH1329									
R23970									
L2									
L3									
L6									
L8									
L11									
L14									
L12									
R12087									
L15									
L4									
L5									

Locations: AYL = Aylesbury, OXF = Oxford, SWI = Swindon, DID = Didcot, WAR = Wargrave, RDG = Reading, BAN = Banbury. Sample types: DD = during anaerobic digestion, PD = post anaerobic digestion. Grey indicates no plaques observed. Green indicates plaques observed and confirmed across two assays.

For the purposes of initial identification, each isolated phage was labelled using its isolation strain (L14, L15 or L12) followed by the sample from which it was isolated (WAR-PD, SWI-PD).

Having obtained putative phages with activity, we sought to ensure that the isolation strains were indeed RT 027 isolates. To do so, the *C. difficile* strains which generated positive results were sent to the *Clostridium* Difficile Ribotyping Network (CDRN) at the Public Health England Microbiology Services laboratory in Leeds. The results confirmed that L14 and L15 were RT 027 strains; however, the strain believed to be L12 was identified as an 078 ribotype (another hypervirulent strain of *C.*

difficile which is also of interest to the CRG). Given the clinical importance of RT 078 strains, testing was continued on these phages.

Following purification by sequential BHIS 0.5% (w/v) agar overlay plaque assays and plaque picks, the titres of the phages were increased through successive BHIS 0.2% (w/v) agar plaque assays until they exceeded 10^8 pfu/ml, at which point preparations were made which were suitable for visualisation using transmission electron microscopy.

3.2.2 Transmission electron microscopy (TEM)

Two different morphologies of phages were identified through TEM, with both phages isolated from L12 confirmed as siphoviruses (characterised by a long thin tail) and the phages isolated from L14 and L15 identified as myoviruses (a medium-length tail and visible contractile tail sheath) (Figure 3.1). The icosahedral capsids were clearly visible, and in some cases the individual proteins of the myovirus tail sheath were evident, creating a textured appearance to the tail. Unfortunately, smaller features such as the tail fibres were unable to be determined even at the lowest resolutions available.

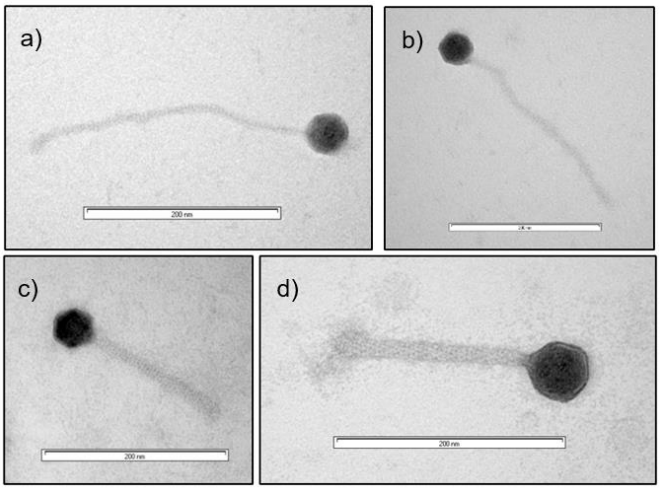


Figure 3.1 - a) Siphovirus phage isolated from the WAR-PD sample using strain L12; b) Siphovirus phage isolated from the SWI-PD sample using strain L12; c) Myovirus phage isolated from the WAR-PD sample using strain L14; d) Myovirus phage isolated from the WAR-PD using strain L15

During visualisation, measurements were taken from several examples of each phage for which the edges of the phage particles could be reasonably well defined. The averages of these measurements were calculated to give the approximate dimensions of each novel phage (Table 3.2).

Table 3.2 - Mean dimensions of phages measured during visualisation using TEM.

Phage	Family	Capsid (nm)	Tail length (nm)	N
L14 WAR-PD	Myoviridae	52.27 ± 5.17	168.43 ± 11.86	14
L15 WAR-PD	Myoviridae	56.10 ± 4.83	168.2 ± 7.20	10
L12 WAR-PD	Siphoviridae	46.00 ± 3.55	314.75 ± 13.65	8
L12 SWI-PD	Siphoviridae	47.30 ± 1.62	257.40 ± 43.10	10

3.2.3 Whole genome sequencing of novel phages

3.2.3.1 L14WAR-PD

Initial results for L14WAR-PD determined the genome to be 45,415bp in size with a GC content of 28.57%. The multiPhATE2 annotation pathway identified seventy-five open reading frames (ORFs) with homology to genes in the provided databases, the majority of which were identified by all three of the gene-calling algorithms in use (Glimmer3, Prodigal and Phanotate). After manual review and cross-checking against the online NCBI NR and Uniprot databases, 69 genes were accepted (Figure 3.2).

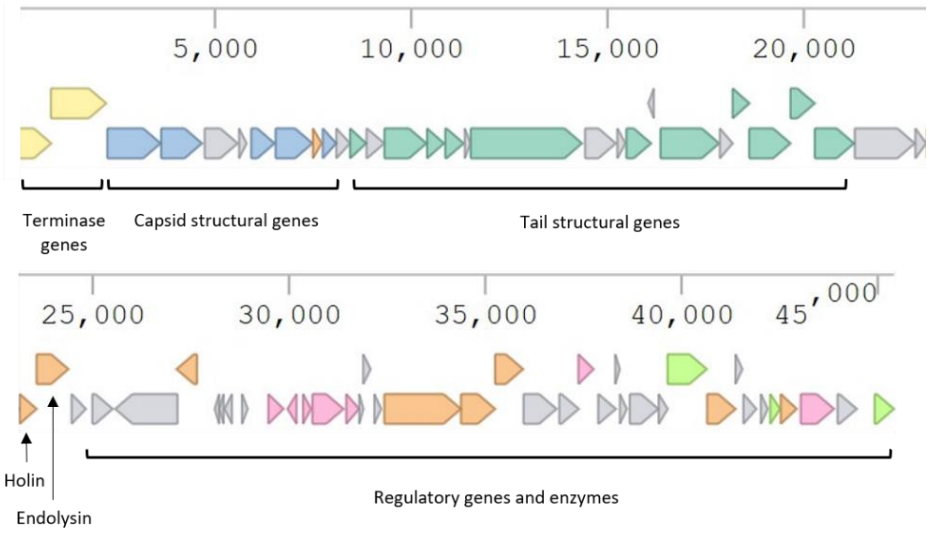


Figure 3.2 – Initial genome of phage L14WAR-PD derived using Illumina short sequencing reads. Yellow = terminase genes; blue = capsid structural genes; dark green = tail structural genes; orange = enzymes; pink = regulatory genes; grey = hypothetical proteins; light green = genes of unknown function containing identifiable domain regions.

During the review of the genome, it was noted that no integrase gene had been identified. Given that this is the defining feature of a temperate phage, and therefore would be expected to be present given past experience, it was determined that ONT long-read sequencing should be used to confirm that no assembly errors had occurred in the Illumina results and thereby confirm the presence or absence of this gene.

The specialised long-read assembly software Canu (using the default settings) generated four contigs from the ONT fastq files which were polished with NextPolish v1.3.1 (polishing is the use of the Illumina data to improve the accuracy of the ONT reads). As the original Illumina sequencing result had been rearranged to start with the terminase gene (as is standard when publishing phage genomes) the contigs were aligned against two consecutive copies of the original Illumina sequencing result (Figure 3.3). This allowed for the contig ends to be aligned to the correct section of the genome irrespective of where the contig started.

All but one of the contigs was larger than a single genome in length, implying that this phage either has a circular genome and/or uses a headful packaging method of DNA packaging (i.e. each capsid contains more than one full genome). However, despite polishing with the Illumina reads, this assembly still showed several areas of discrepancy when compared to the assembled Illumina genome. Closer inspection of the sequences in these areas showed that in the majority of cases the smaller discrepancies were single nucleotide differences in one or two of the contigs, but the other contigs were consistent with the Illumina sequences in these areas.

What this assembly did highlight was that a gap was present in the Illumina sequencing around gp30, a hypothetical gene just upstream of the holin gene. This was consistently represented across all four contigs, indicating a potential misassembly in the original Illumina analysis.

However, aside from this gap and the other small discrepancies in the error-prone long reads, the Canu assembly confirmed the majority of the Illumina sequencing to be correct. This also confirmed that an identifiable integrase gene was not present within the regulatory region. The pre-holin gap was unlikely to contain an integrase gene as it was present in the structural region of the phage. The chance of the lack of an integrase gene being the result of a misassembly is therefore minimal.

As additional confirmation of the genome (in particular the correct sequence of the gap identified by Canu) a further analysis was conducted using the Flye sequencing package in order to form a consensus assembly. Flye assembled the .fastq files into 20 discrete contigs. Examination of the assembly information showed that the majority of these contigs had low coverage (3-5 reads), while two further contigs had slightly greater coverage (118 and 115). One contig, however, was 45,891bp in length with a coverage of 10,128 reads, signifying that this was likely to be the full phage genome assembled as a single contig. As with the Canu assembly, this contig was polished using the Illumina reads

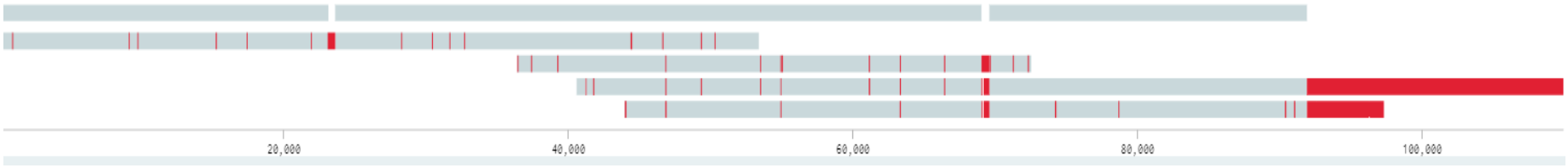


Figure 3.3 – Benchling alignment of Canu-generated L14WAR-PD contigs to a double-length version of the Illumina assembly. Red indicates a mismatch or, in the case of the ends of contigs 3 and 4, indicates an area outside of the aligned genome.



Figure 3.4 - Benchling alignment of the doubled version of the L14WAR-PD Illumina genome assembly to the single copy Flye assembly. a) Alignment before amendments. b) Alignment after insertion of the 472bp section that was missing from upstream of the holin gene in the Illumina assembly and correction of SNPs in the second half of the Flye assembly.

and aligned against two consecutive copies of the original Illumina sequencing result (Figure 3.4a). The polished Flye assembly aligned almost perfectly with the original Illumina sequence except for a few SNPs and an extra 472bp section of nucleotides within the same hypothetical gene just upstream of the holin gene that was identified in the Canu assembly (resulting in the misaligned area indicated in red). Comparison between the Canu and Flye assemblies confirmed these additional base pairs, which were subsequently added to the L14WAR-PD phage genome (Figure 3.4b).

This area of the genome was then checked for any change in the ORFs, and it was determined that the additional base pairs had resulted in the identification of a further hypothetical gene in this region (the annotation of which was confirmed through searching against the NCBI BLASTp database). This revised version of the L14WAR-PD genome (45,887bp, 70 ORFs) has therefore been accepted as the most accurate representation available (Figure 3.5).

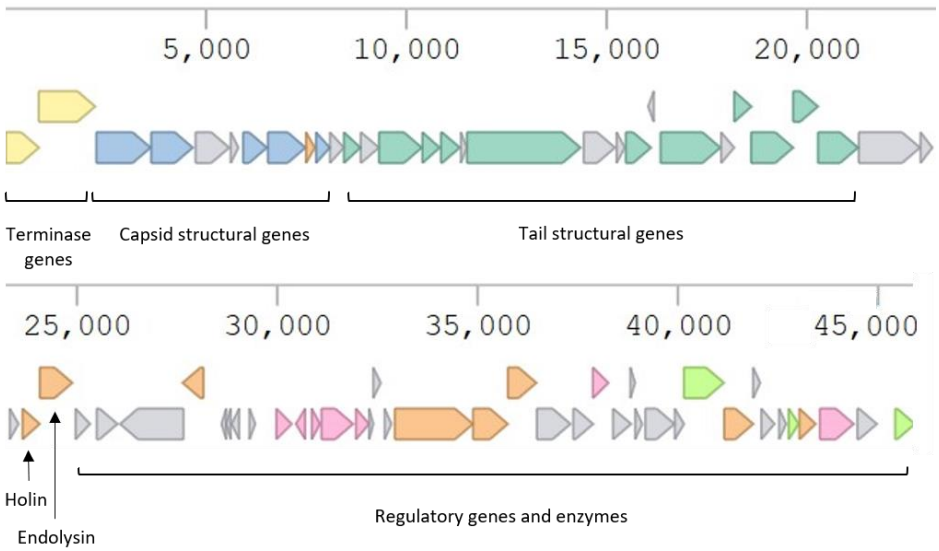


Figure 3.5 - Revised genome of phage L14WAR-PD following long-read sequencing using ONT MinION sequencing. The majority of the genome remains the same in terms of the open reading frames, however a hypothetical gene upstream of the holin gene has now been identified as two genes instead of one.

With the most accurate genome established, the composition of the genome can be investigated in more detail. As with most, if not all, *C. difficile* phages, the terminase genes responsible for DNA packaging precede the structural genes and consist of a small and a large subunit. The structural genes are then grouped into those coding for eight capsid-related proteins and eighteen tail-related proteins, including a 2,835bp tail tape measure protein, a cell-wall hydrolase and two potential tail fibre genes. These genes are reasonably well-characterised in the literature and so are often the most easily annotated genes in a bacteriophage. Together, these structural genes and the terminase genes make up 49% of the phage genome.

The remainder of the genome is less easily characterised, as it is dominated by genes encoding hypothetical proteins with no identified function. A series of transcriptional regulator genes cluster between 29,948bp and 32,261bp and genes coding for several enzymes have been identified with functions contributing to DNA replication/repair (Rec-like proteins and a RusA family crossover junction endodeoxyribonuclease). The latter additionally have functions in recombination, so may have been associated with the integration of the genome into the host prior to the loss of the integrase gene.

A closer inspection of the transcriptional regulators also showed that while an XRE-family regulator (the same family as the *cI*/*cro* genes in phage λ) is present, it was not identified as a potential repressor protein (unlike the adjacent gp45 which has been annotated as an anti-repressor). It is therefore difficult to ascertain from the initial annotation whether only the integrase is missing from this phage or whether it is missing further lysogeny-related genes.

Accordingly, the genome was subjected to a new online tool PhageLeads (Yukgehnaish et al., 2022) in an attempt to identify other potential lysogeny genes. PhageLeads uses a machine-learning approach to assess genes in a queried phage for any features which may indicate a

temperate lifestyle, allowing researchers to quickly assess which phages may be suitable for use in phage therapy. The output of this tool identified two potential lysogeny genes in this genome – gp4 and the aforementioned gp45. While gp45 is understandable and the annotation is consistent (identified by both PhageLeads and multiPhATE2 as an antirepressor), gp4 was identified by PhageLeads as an integrase gene. However, on comparing this gene to the previous annotations and position within the genome, gp4 is located at the start of the structural genes and has been annotated by multiPhATE2 as encoding a minor capsid protein. In my opinion, this is potentially a misidentification by the PhageLeads programme, which would not be unexpected given that it is a new tool. It does, however, bear further examination at a later date to determine what elements of gp4 have been identified by PhageLeads as being sufficiently homologous to integrase genes such that the machine-learning approach has ascribed it a similar function.

While the lack of readily identifiable lysogeny machinery (such as a repressor) was promising, this analysis did not entirely confirm or deny the presence of relevant genes. As such, further experimentation (detailed later in this chapter) was planned to demonstrate this phage's ability (or lack thereof) to form a lysogen.

3.2.3.2 L15WAR-PD

The phage designated as L15WAR-PD was initially determined to be 49,507bp in size with a GC content of 28.23%. Annotation via the multiPhATE2 pathway identified 84 ORFs with homology to ORFs in the provided databases. Again, the majority of the genes were called by all three algorithms, with a few identified by only one or two. Following cross-checking in a similar fashion to L14WAR-PD, 75 genes were accepted (Figure 3.6).

Similar to L14WAR-PD, this phage has genes coding for a small and large terminase subunit followed by genes encoding for eight capsid-related proteins and fifteen tail proteins (with the same length of tail tape

measure protein, consistent with being a myovirus). The structural genes make up 47% of the genome.

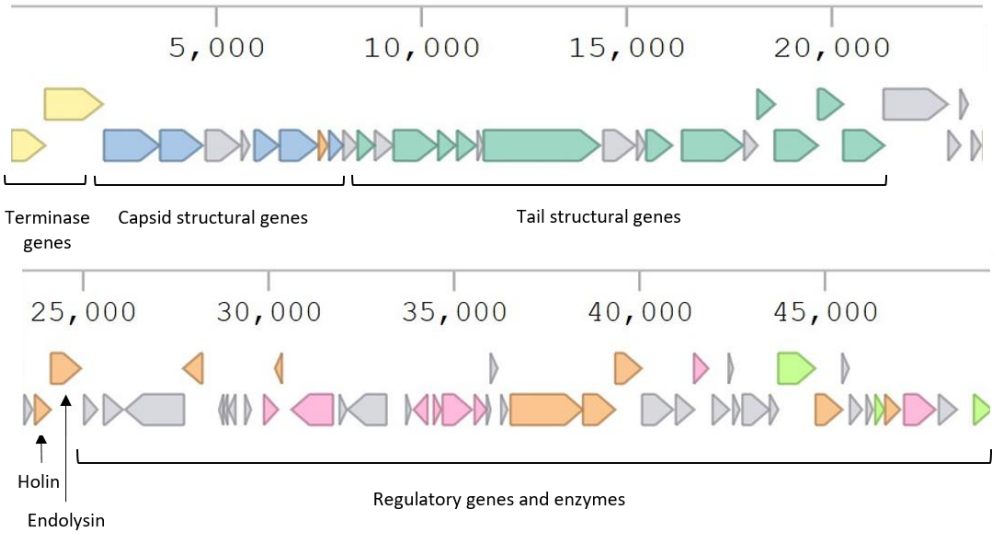


Figure 3.6 - Genome of phage L15WAR-PD derived using Illumina short sequencing reads. Yellow = terminase genes; blue = capsid structural genes; green = tail structural genes; orange = enzymes; pink = regulatory genes; grey = hypothetical proteins; light green = genes of unknown function containing identifiable domain regions.

The remaining genes also show a similar composition, with the same clustering of transcriptional regulator genes further downstream (33,910bp – 35,881bp) and genes coding for the same replication enzymes. However, unlike L14WAR-PD, this phage contains an additional segment of five genes which includes a gene encoding a putative tyrosine site-specific integrase at position gp44, identifying this as a temperate bacteriophage. As such, no additional DNA sequencing was undertaken on this phage since it conformed to our original expectation of *C. difficile* phages being primarily, if not exclusively, temperate in nature.

Unfortunately, as with L14WAR-PD, none of the encoded transcriptional regulators were shown to have a significant degree of homology with either *cI* or *cro* beyond being helix-turn-helix domain proteins, although one has been identified as potentially being from the same XRE family.

It is therefore not possible to surmise from the genome whether this phage uses a similar mechanism of lysogeny control to phage λ.

3.2.3.3 L12SWI-PD

The phage designated as L12SWI-PD was determined to be 35,682bp in size, GC content 28.86%. The multiPhATE2 pathway identified 75 ORFs with homology to ORFs in the provided databases, 16 of which were called by just one of the algorithms (usually Phanotate). 73 genes were confirmed following manual review (Figure 3.7).

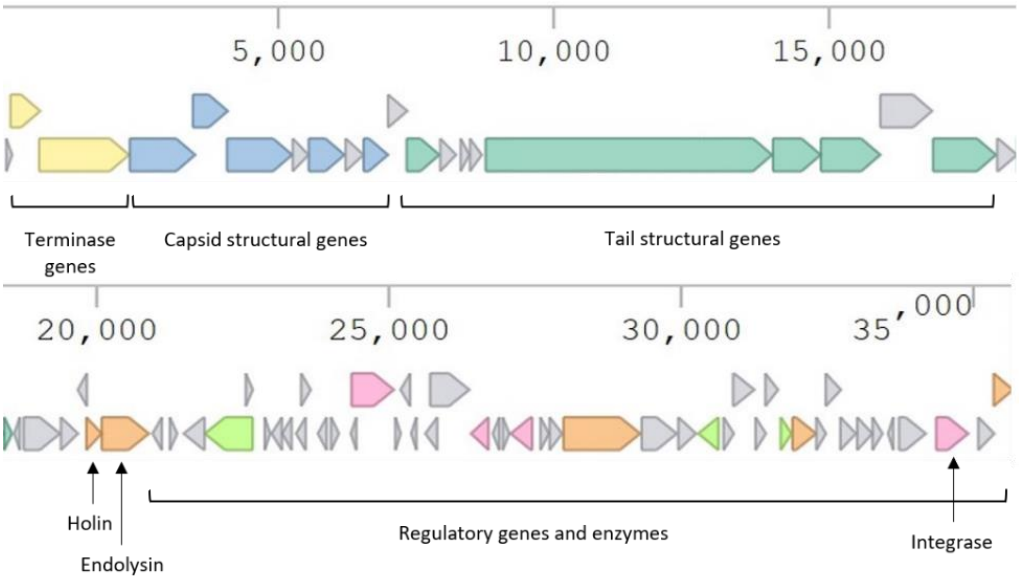


Figure 3.7 - Genome of phage L12SWI-PD derived using Illumina short sequencing reads. Yellow = terminase genes (large and small subunits); blue = capsid structural genes; green = tail structural genes; orange = enzymes; pink = regulatory genes; grey = hypothetical proteins; light green = genes of unknown function containing identifiable domain regions.

As with the myovirus phages previously described, the genome of this siphovirus phage contains genes encoding two terminase subunits, eight capsid proteins and fifteen tail proteins, demonstrating a degree of consistency in the composition of structural proteins despite little to no similarity at the nucleotide level. In contrast, however, the tail tape

measure gene is 5,202bp in size, reflecting the characteristic longer tail of siphovirus phages.

Past this point, the genome contains genes coding for thirty-seven hypothetical proteins, a ssDNA binding protein, a replication protein and several transcriptional regulators, although the latter are not as clustered as they are in L14WAR-PD and L15WAR-PD. A tyrosine site-specific integrase is present at position gp71 indicating that this is a temperate phage, and the phage also contains an HNH endonuclease, which is a type of gene associated with the recombination of selfish genetic elements and therefore may have a role in the transduction of genes between the phage and host genome.

As with L15WAR-PD, no additional DNA sequencing has been undertaken on this phage at this time due to the presence of an integrase gene.

3.2.3.4 L12WAR-PD

The phage designated as L12WAR-PD was determined to be 35,132bp in size, GC content 28.82%. 73 ORFS were identified by the multiPhATE2 pathway, including 16 which were identified only by the Phanotate algorithm. All 73 genes were confirmed during a manual review (Figure 3.8).

The gene-by-gene annotations for this phage are almost identical to L12SWI-PD (including a site-specific tyrosine recombinase at position gp71, indicating that this is also a temperate phage), implying that L12SWI-PD and L12WAR-PD may be the same or very similar phages. The only exceptions are the two genes encoding XRE family transcriptional regulators identified at gp50 and gp52 in L12WAR-PD which are labelled as hypothetical genes in L12SWI-PD. This may be due to a minor misassembly in one of the two phages, especially given the 500bp difference in overall size between them.

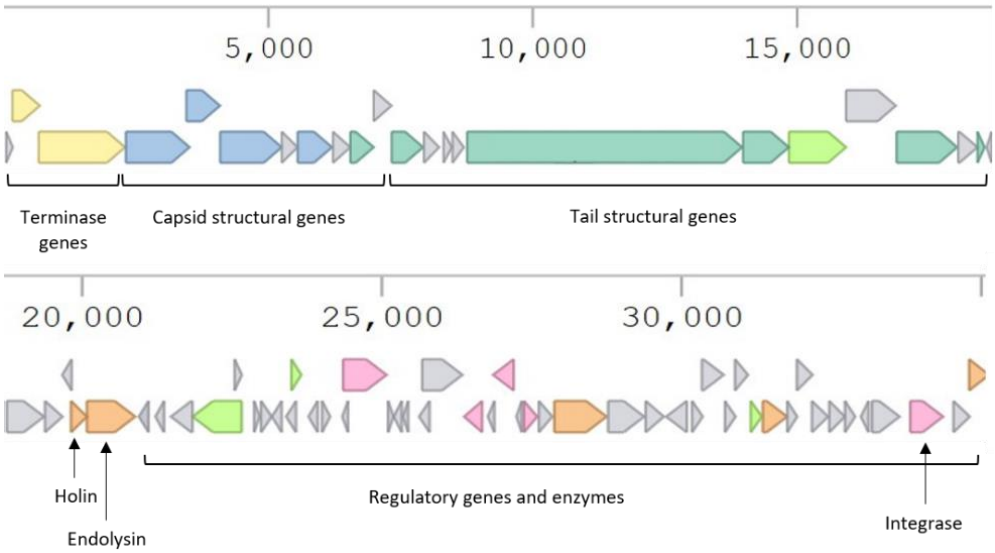


Figure 3.8 - Genome of phage L12WAR-PD derived using Illumina short sequencing reads. Yellow = terminase genes; blue = capsid structural genes; green = tail structural genes; orange = enzymes; pink = regulatory genes; grey = hypothetical proteins; light green = genes of unknown function containing identifiable domain regions.

To determine whether these are the same or similar phages, further genomic analyses were required. Direct comparisons were therefore undertaken between L12SWI-PD and L12WAR-PD. Comparisons were also undertaken between L14WAR-PD and L15WAR-PD given that they are both myovirus phages and have been isolated from the same wastewater sample.

3.2.4 Genetic comparisons between isolated phages

3.2.4.1 Siphovirus phage comparison

Comparison between siphovirus phages L12SWI-PD and L12WAR-PD showed a substantial number of differences at a nucleotide level (Figure 3.9), as well as the aforementioned differences at the proteome level. These differences were sufficient to be confident that these were two different phages, and as such, they were given the designations phiL12S and phiL12W.

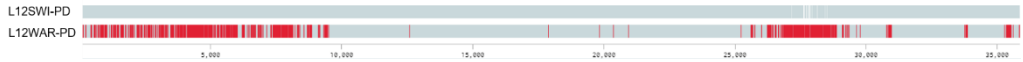


Figure 3.9 - Benchling alignment of phages phiL12S and phiL12W. Several areas of difference are visible between the two phages at the nucleotide level such that these two phages can be considered as different strains.

3.2.4.2 Myovirus phage comparison

The comparison between L14WAR-PD and L15WAR-PD showed a high degree of homology between the two phages, with only three discrete areas of difference – an insert in L14WAR-PD (163bp inserted at base 29,323), and two inserts in L15WAR-PD (an 88bp insert at base 23,169 and a larger 3,695bp insert at base 30,458) (Figure 3.10).

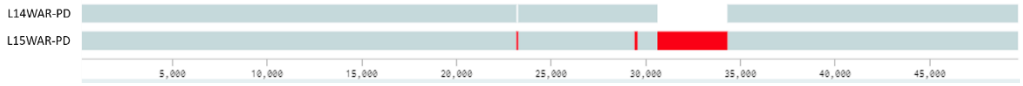


Figure 3.10 – Genomic comparison between L14WAR-PD and L15WAR-PD. Three distinct areas of difference are visible with no other inconsistencies in the nucleotide sequence.

Each of the smaller areas of difference could be attributed to errors in the whole genome sequencing. Given the lack of any other SNPs and the discrete nature of these areas, it was determined that further Sanger sequencing of these areas was not required at this time but could be undertaken in future.

The 3,695bp insert in L15WAR-PD, however, notably contains the integrase gene, and so further PCRs were undertaken (using the phage preparations as a template) to determine whether these differences were being accurately reflected by the Illumina sequencing – i.e. whether the L14WAR-PD and L15WAR-PD phages were the same or whether they were very closely related phages with a mutation event separating them. The PCR confirmed that this area was missing from both phage

preparations tested (L14WAR-PD and L15WAR-PD) and was therefore consistent with the original sequencing of L14WAR-PD (Figure 3.11).

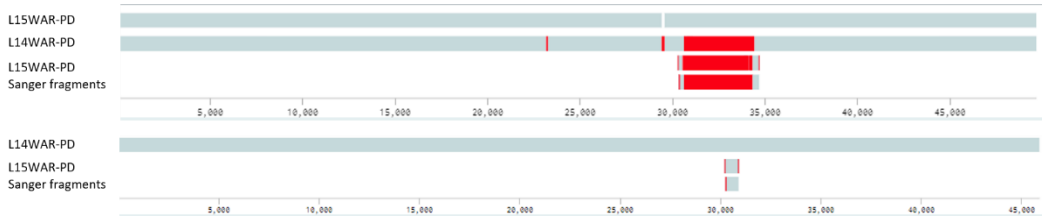


Figure 3.11 - Benchling comparison of Sanger sequencing results obtained from PCR screening of L15WAR-PD. Top – comparison of the fragments generated from L15WAR-PD to the Illumina genome sequencing results of the same phage. Bottom – comparison of the same fragments to the L14WAR-PD genome.

This section of work was conducted during the early post-lockdown shift phase of the project when conditions were difficult and lab time scarce. It is therefore possible that the presence of an integrase-containing section of DNA in L15WAR-PD may be accurate, with this truly being a second phage, but that human error at some point resulted in a sample swap/mixed phage prep. This could only be checked by returning to the original sample, creating a fresh phage preparation and re-submitting for sequencing, which could be undertaken at a later date if required. However, given the results obtained from this comparison work, the existing L14WAR-PD and L15WAR-PD samples were deemed to contain the same phage. Subsequent work was completed under the name phiL14WAR, with labelling checks and aseptic technique applied rigorously during later experiments.

3.2.5 Phylogenetic analysis

The genomes of all confirmed phages were uploaded to the phylogenomic analysis website VIPtree (Nishimura et al., 2017) to establish their closest known relatives (Figure 3.12). VIPtree uses NCBI's tBLASTx package to generate a proteome for the queried sequences. It then compares these sequences to the proteomes of a reference set using the genome-based "Phage Proteomic Tree" classification method

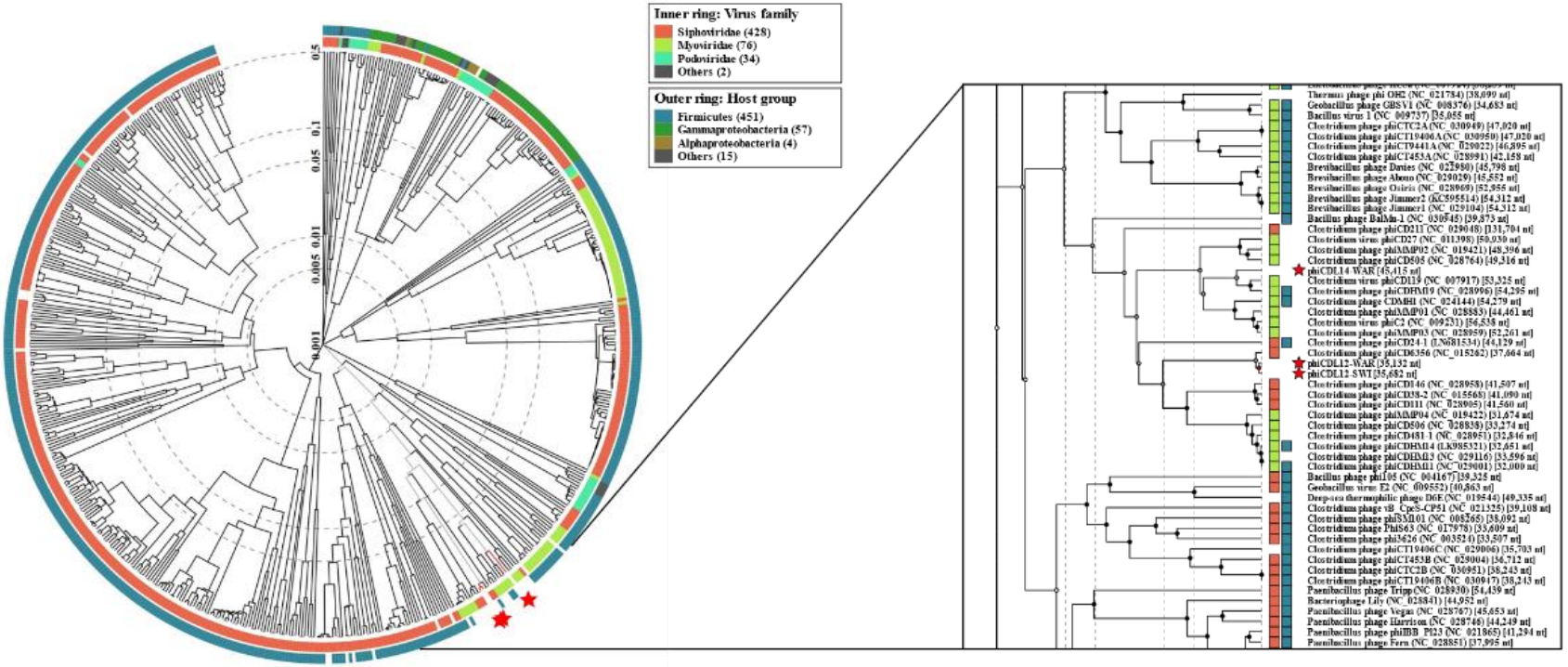


Figure 3.12 - Proteomic tree generated by VIPtree showing alignment of novel phage isolations against a reference database generated from RefSeq, UniProt, Genbank and ViralZone. Red stars mark the positions of the three phages within the tree.

proposed by Rohwer and Edwards (2002). For this analysis, comparisons were limited to dsDNA prokaryotic viruses. As expected, all three phages fall within a large block of viruses which infect strains of *C. difficile*. The closest relatives of phiL14WAR are phiCD119 (Govind et al., 2006) and phiCDHM19 (Hargreaves and Clokie, 2015), while the closest relative to phiL12S and phiL12W is phiCD6356 (Horgan et al., 2010).

3.2.5.1 phiL14WAR

Closer inspection of similarity scores for phiL14WAR shows that while phiCD119 and phiCDHM19 are the closest relatives, the degree of similarity is fairly low in both cases (similarity scores of 0.62 and 0.49 respectively), which is further demonstrated by direct alignment of the three phages (Figure 3.13).

The area of primary similarity is centred on the structural proteins, as would be expected given that these are all Myoviridae phages. However, beyond the end of the tail genes (approximately the 20kb mark in Figure 3.13) there appear to be very few genes in common between phiL14WAR and its closest identified relatives. Accordingly, there are likely other phages which are closer relatives which have yet to be isolated.

While closer relatives have not yet been identified, examination of the differences between phiL14WAR, phiCD119 and phiCDHM19 had the potential to allow the identification of what genes other than the integrase may be missing from the phiL14WAR (a *cI*-equivalent, for example). An additional alignment of these three genomes using the progressiveMauve alignment algorithm (Darling et al., 2010) gave more detail as to which genes are shared between the phages and which are not (Figure 3.14). It should be noted that progressiveMauve identifies homology at the nucleotide level, while the previous VIPtree analysis was at the proteome level, which may account for any discrepancies in the output.

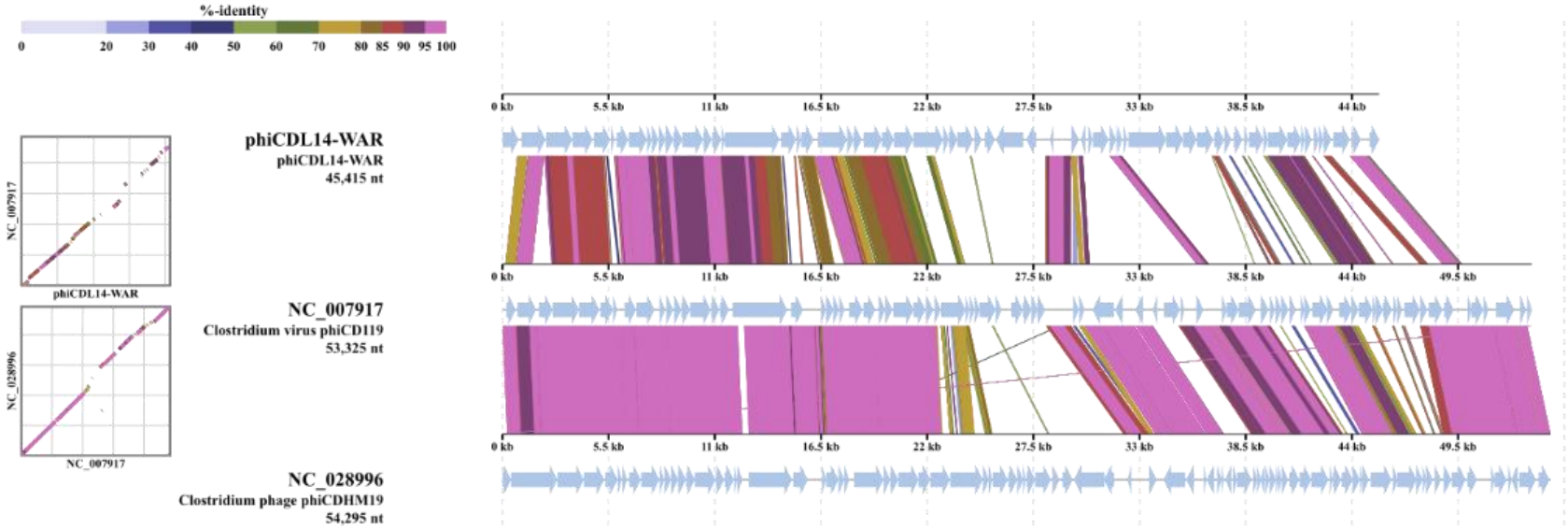


Figure 3.13 – VIPtree alignment and dot plot comparisons of phiL14WAR phage with phiCD119 and phiCDHM19.

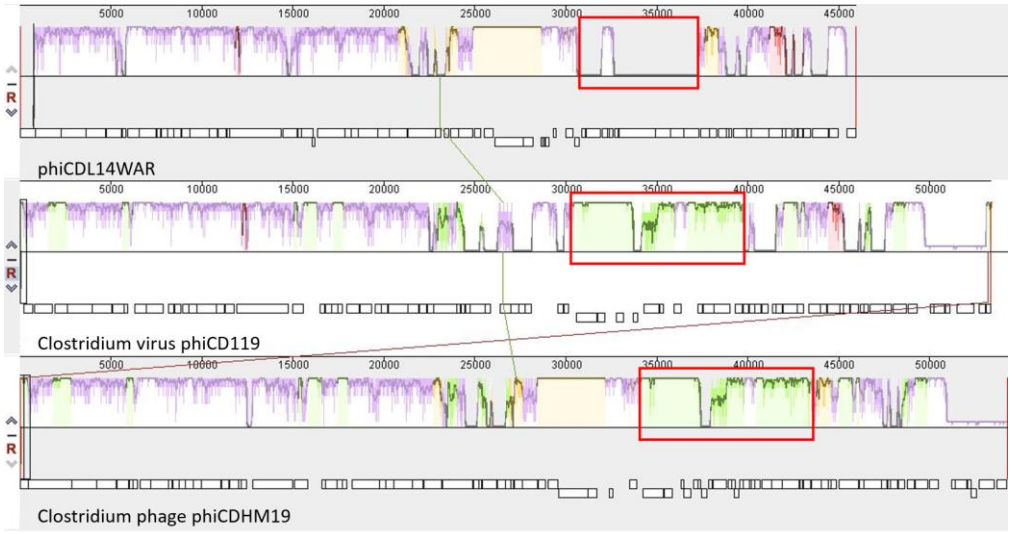


Figure 3.14 – Alignment of phiL14WAR, phiCD119 and phiCDHM19 using the progressiveMauve algorithm. Purple = shared by all three phages; green = shared by phiCD119 and phiCDHM19; red = shared by phiL14WAR and phiCD119; amber = shared by phiL14WAR and phiCDHM19. Red boxes show the main areas of difference.

The highlighted boxes show the primary area of interest between these three phages. In phiCD119 and phiCDHM19 (the bottom two panes), this area begins with the integrase gene and encompasses several genes encoding transcriptional regulators and proteins related to DNA replication/binding. Examination of this area in phiL14WAR also shows the presence of genes coding for hypothetical, recombination and DNA repair proteins, but with little to no homology to its relatives. From the perspective of establishing whether any other key genes are missing, this comparison was not very informative, as there is no obvious evidence of an encoded repressor protein being present in phiCD119 and phiCDHM19 that is not present in phiL14WAR.

3.2.5.2 phiL12S and phiL12W

PhiL12S and phiL12W showed a higher degree of similarity between themselves (0.93) than to phiCD6356 (0.85 and 0.88 respectively) implying a more recent evolutionary divergence (Figure 3.15).

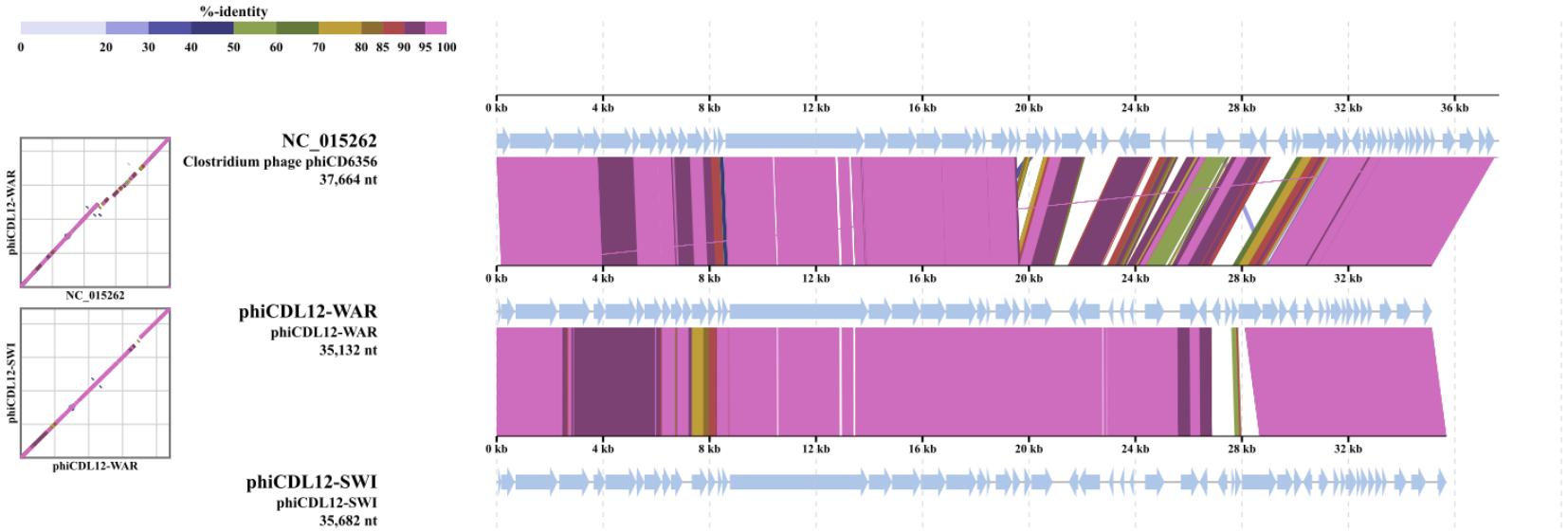


Figure 3.15 – VIPtree alignment and dot plot comparisons of phiL12S and phiL12W phages against phiCD6356.

Subsequent progressiveMauve alignment shows that there are very few differences at the nucleotide level between these two phages (Figure 3.16), however, each shares a different small area of the genome with phiCD6356. In the case of phiL12S, this shared area appears to relate to an encoded helix-turn-helix transcriptional regulator and a hypothetical protein separated by an unshared non-coding area of DNA. PhiL12W, however, appears to share a gene coding for the same variant of putative phage replication protein as phiCD6356, while the equivalent gene in phiL12S is, in fact, 441bp larger than that of its counterparts. The two variants share the same C-terminal, and, given that the two proteins have been annotated in a similar fashion, it is likely that the difference between these variants has not affected the overall function of the protein.

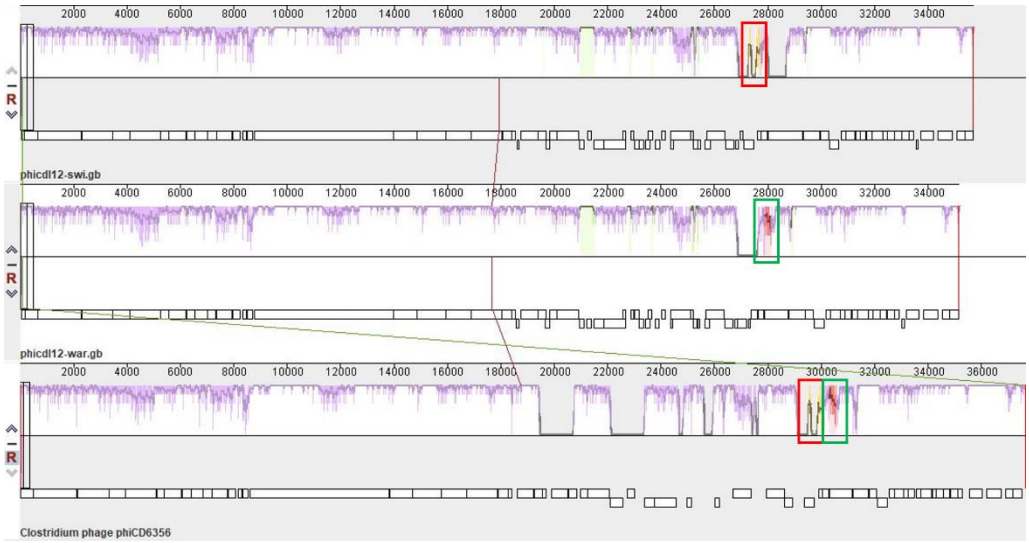


Figure 3.16 – progressiveMauve alignment of phiL12S, phiL12W and phiCD6356. Red boxes are shared elements between phiL12S and phiCD6356; green boxes are areas shared by phiL12W and phiCD6356.

Additionally, phiCD6356 contains two larger sections upstream of the shared regions which show no homology to either of the L12 phages. The first falls within the region encoding the structural proteins of the tail prior to the endolysin gene, where an approx. 1.3kb section of DNA with three hypothetical proteins falls within the space occupied by gp25 in phiL12S and phiL12W, splitting the hypothetical gene there in two. It may be that this section of the genome has been lost from our phages, leaving a

single hypothetical gene in its place, or this may be an insert into phiCD6356 at some point since its evolutionary divergence. Either way this area corresponds with the terminus of the tail of the phage, and therefore may affect the host range.

The other area of difference in phiCD6356 corresponds to the two hypothetical genes immediately downstream of the endolysin gene in each phage. In addition to no nucleotide similarity, direct comparisons show that the encoded hypothetical proteins share no homology to the equivalent proteins in phiL12S and phiL12W, and so may have a different function.

3.2.6 Siphovirus integrase analyses

Further investigation was also undertaken to characterise the site-specific recombinase genes found in both of the L12 phages. As previously stated, both encoded integrases are of the tyrosine type and match the Int_C_like_3 subfamily of the DNA breaking-rejoining enzymes characterised by a C-terminal catalytic domain (Figure 3.17)

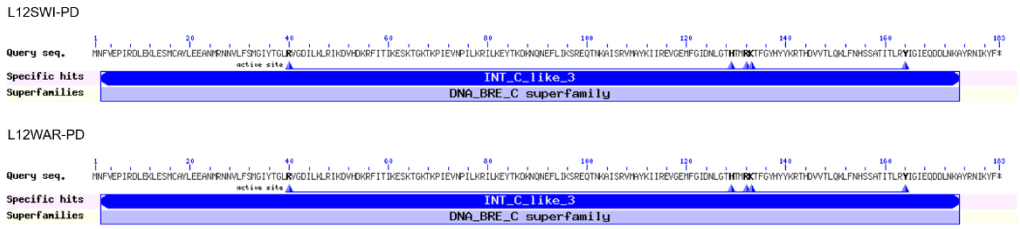


Figure 3.17 - Genomic characterisation of the integrase genes from phiL12S and phiL12W using NCBI BLASTp service. Arrows indicate the active catalytic sites that characterise the Int_C_like_3 subfamily.

Direct alignment of the two isolated integrase genes shows that they are identical despite evolutionary divergence in other areas of the genome. Further alignment to Int_C_like_3 genes from *Clostridium perfringens* and *Salmonella enterica enterica*, in conjunction with cross-referencing against the NCBI database, allows the five active catalytic residues that control the enzyme function to be identified, particularly the C-terminal

active tyrosine residue which is responsible for DNA cleavage (Figure 3.18).

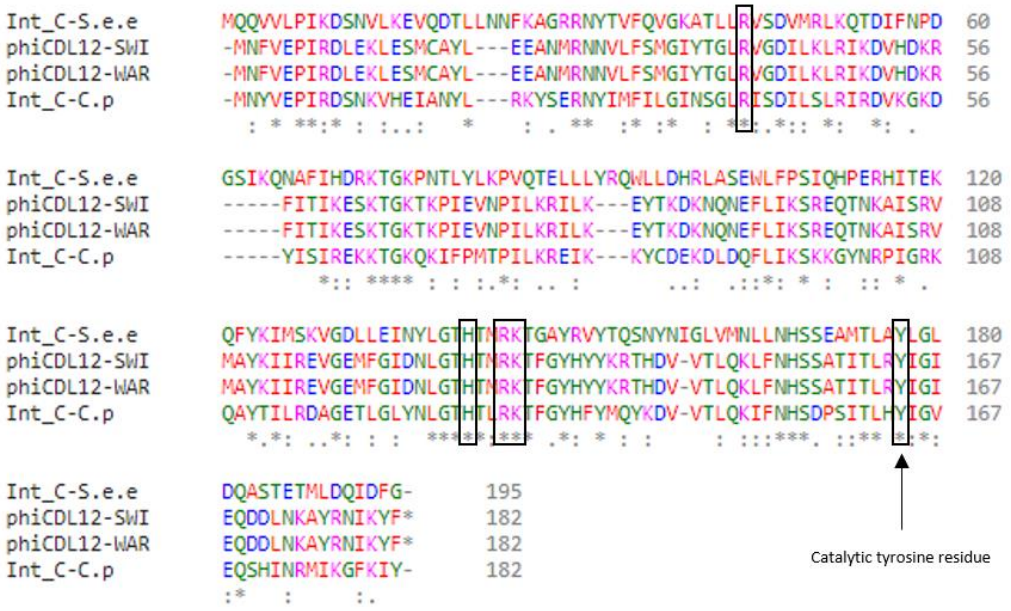


*Figure 3.18 - ClustalW alignment of the tyrosine site-specific integrase genes from phiL12S and phiL12W alongside Int_C_like_3 genes from Clostridium perfringens (C.p) and Salmonella enterica enterica (S.e.e). Black boxes indicate the 5 active catalytic residues for this subfamily of the DNA breaking-rejoining superfamily. * indicates conserved residues between all four genes.*

Genomic determination as to whether these integrase genes are active would be dependent on the identification of appropriate promotor and ribosome binding sites (RBS) upstream of the genes. While a Shine Dalgarno (SD) site is present 10bp upstream of each gene, identification of a specific promotor is more complex, as phage promotors are poorly described with few tools currently available to predict them (Sampaio et al., 2019). As such, lysogeny tests are planned which will provide experimental confirmation as to whether or not these integrase enzymes are active.

3.2.7 Host range testing and efficiency of plating

NB – This section was completed in conjunction with my colleague Joanna Steczynska

Host range testing was undertaken in a step-wise fashion due to the large number of available strains to test within the CRG culture collection. Initial testing was undertaken using a spot-on-the-lawn technique in which 20µl of a neat and a 1/10 dilution of each phage was spotted onto a bacterial lawn of the target *C. difficile* strain (Figure 3.19). A total of 190 strains from 33 different ribotypes were screened using this method for any indication of potential infection by the novel phages. A full list of the strains tested is detailed in Appendix 7.1.

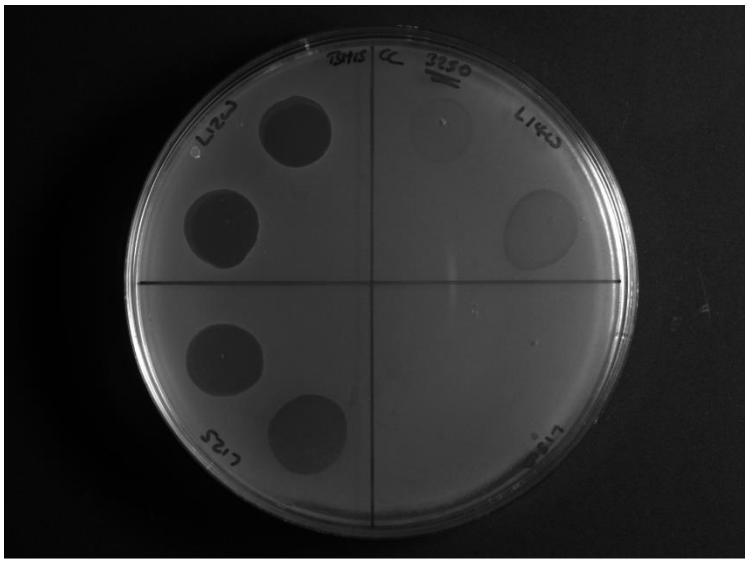


Figure 3.19 - Example of a plate generated by the spot-on-the-lawn technique. 20µl of neat and 1/10 dilution of phage preparation was spotted onto a quadrant of the bacterial lawn plate at the time of pouring. The presence of phages interfered with bacterial growth, creating defined areas of clearing in the lawn. The degree of infection was represented by the clarity of the area but this method was deemed too inaccurate to use for estimating the efficiency of plating for each phage.

The 95 strains (from five ribotypes) which showed some level of sensitivity to the phages were then taken forward for further testing using a full plaque assay to more accurately compare the infectivity against the

test strain (the propagating strain) with that of the isolation strain. This is known as the efficiency of plating – EOP. For this, the phage preparations were adjusted to all have an approximate titre matching the lowest of the three preparations (phiL12S – 4 x 10^7 pfu/ml) to allow for a more direct comparison between the phages. The titre was calculated using plaque counts from each of the test strains, and this was compared to the titre generated from a control plaque assay using the isolation strain. The difference was calculated using the following equation:

$$Efficiency\ of\ plating = \frac{titer\ derived\ from\ propagating\ strain}{titer\ derived\ from\ isolation\ strain}$$

A heat map was then generated from the results of the EOP calculation. The results showed that all three novel phages infect four out of the five ribotypes tested during the EOP (phiL14WAR - 001, 027, 078, 087; phiL12S and phiL12W - 001, 014, 027, 078). As expected, the bulk of the susceptible strains originated from the same ribotype as the isolation strain, however, all three showed some infection of a number of RT 001 strains as well.

Of particular interest is the fact that, while they were isolated using a confirmed RT 078 strain, both phiL12S and phiL12W still infect a few RT 027 strains, with both phages infecting the type-strain R20291 to an equal or higher degree than phiL14WAR. All phages had at least one other strain for which the propagating strain generated a higher titre than the isolation strain, with phiL12W seemingly much more infective in many of the tested cases. This phage may therefore be a good target for genetic modification in future.

Table 3.3 - Heat map of efficiency of plating for novel phages tested against sensitive strains of C. difficile

Strain name	PCR-ribotype	phiL14WAR	phiL12S	phiL12W
001-3	001			
001-2	001			
001-5	001			
001-7	001			
L24	001			
L27	001			
001-6	001			
001-3	001			
R8366	001			
NCTC 11205	001			
NCTC 11207	001			
NCTC 11209	001			
BI-9	001			
LIV29	001			
serogroep K	014			
8085054	014			
TL176	014			
CD3601	015			
CD1902	015			
CD8501	015			
81568	027			
81566	027			
5108111	027			
32219	027			
M246	027			
2191	027			
51556	027			
60902	027			
26131	027			
26173	027			
CDC 32	027			
CDC 38	027			
M13042	027			
DH349	027			
DH320	027			
DH1916	027			
DH1396	027			
DH1432	027			
DH1834	027			
DH1466	027			
DH1858	027			
DH1751	027			
027 Alexander	027			
R20291	027			
DH361	027			
DH131	027			
R20352	027			

Strain name	PCR-ribotype	phiL14WAR	phiL12S	phiL12W
DH478	027			
R12087	027			
T31694	027			
DH482	027			
R24694	027			
DH326	027			
DH1342	027			
DH1329	027			
DH85	027			
R23970	027			
O27	027			
L1	027			
R24988	027			
Libramont 1	027			
R20298	027			
R20291	027			
CD196	027			
210	027			
CD2801	027			
CD5101	027			
CD9201	027			
51557	078			
Wilcox 078	078			
31662	078			
2016	078			
7004578	078			
7009825	078			
Type 078	078			
CD 2315	078			
7045389	078			
CD3801	078			
CD4401	078			
CD8601	078			
CD9001	078			
CD9301	078			
CD9501	078			
CD1801	078			
VPI10463	087			
NCTC 11204	not known			
E101				
B14				
B32				
B35				
BZ47				
V23				
6443 REA type BI 17				

Key

	Cannot be determined
	Resistant isolates
	<0.1
	1-0.1
	>1
	Isolation strain

There was also a substantial number of the tested strains, mostly RT 027, for which the titre was unable to be determined. For these strains, there was evidence of infection from neat samples down to a dilution of 1/100 in the form of small plaque-like zones of clearance concentrated around the edges of the plates. However, further dilutions showed little to no evidence of plaques. Repeating the assay for these strains showed no change to the result, and so while these strains can be noted as positive for infection, it was not possible to determine the titre of the phage using the propagating strain. While most of the strains that this applied to were RT 027 strains, one RT 001 and one RT 078 strain also showed this feature. Comparisons could be undertaken in future to determine whether these strains had any common biological features with RT 027 strains

(e.g. restriction-modification systems) which may be responsible for this result.

3.2.8 Testing for the ability to form lysogens in phiL14WAR

NB - Lysogen formation for the temperate L12 phages was completed by Joanna Steczynska and is therefore not covered here.

The lack of an integrase gene, while usually sufficient to declare a phage as lytic, is unusual enough in the case of *C. difficile* phages that experimental confirmation was desirable.

Initially, agar from the centre of a zone of clearing (generated from a plaque assay) was collected using a pipette tip and streaked on fresh BHIS-CC plates. In theory, any live bacteria present within this agar should be lysogens, as any sensitive bacteria would have been killed by the phages. 20 colonies were selected from those that grew and these colonies were streaked three times over successive days to ensure the removal of any left-over phage particles. The 18 of these colonies which grew across all three purity streaks were then tested for the presence of a prophage by PCR using primers which target the phage DNA.

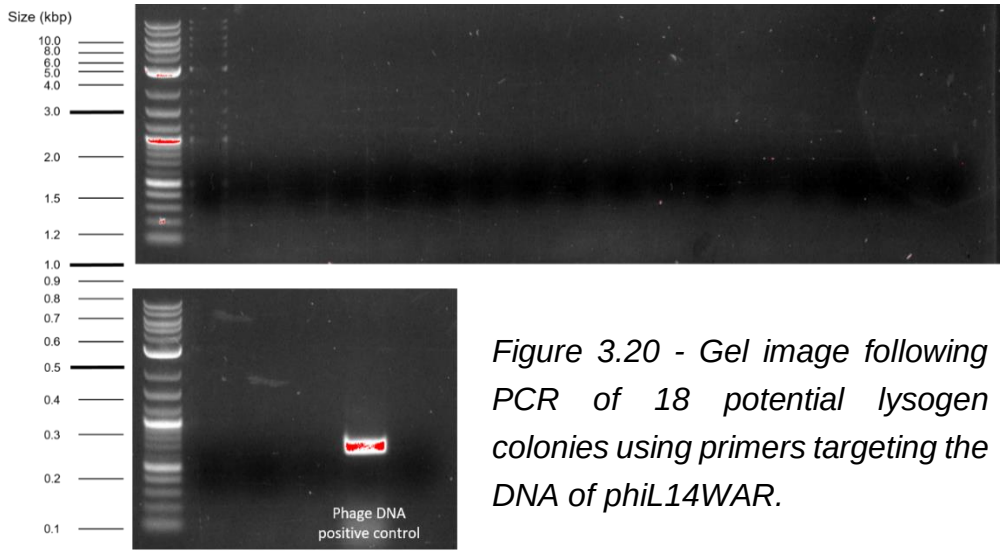


Figure 3.20 - Gel image following PCR of 18 potential lysogen colonies using primers targeting the DNA of phiL14WAR.

No bands were obtained except for the positive control sample created using phage DNA from a current preparation of phiL14WAR. This is consistent with phiL14WAR having lost the integrase gene and therefore being unable to form a lysogen. However, ease of lysogen-formation varies between phages (personal communication, Dr. J Steczynska), and so it was decided that additional tests would be advisable to further strengthen the evidence in this case.

To this end, a modified version of the original assay was conducted where L14 was streaked in lines onto a fresh BHIS-CC agar plate and samples of phiL14WAR were spotted onto the streaks at predetermined points. A positive control was created using the known temperate phage phiCD2301 under the same method. After overnight incubation, there was a distinct difference in bacterial growth within the spotted zones, as demonstrated in Figure 3.21.

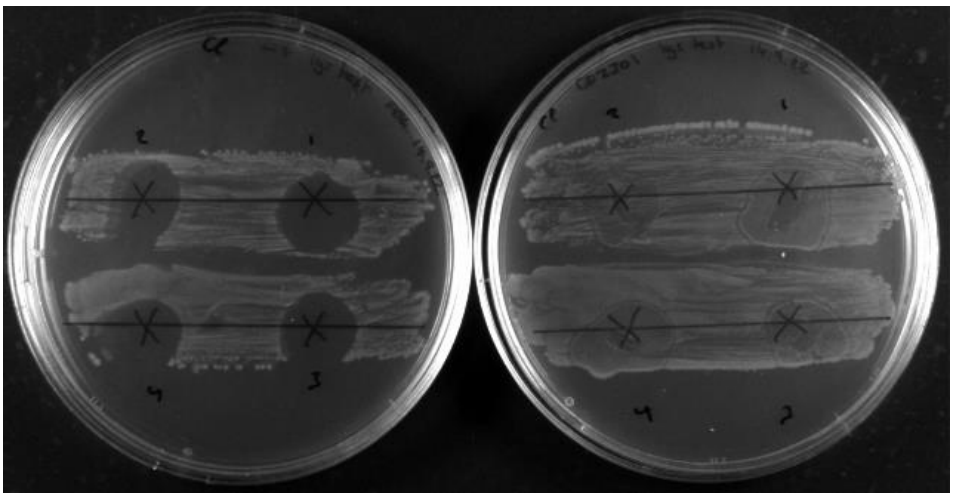


Figure 3.21 – Left - BHIS-CC agar plate streaked with C. difficile strain L14 and spotted with neat phiL14WAR phage samples. Right - plate streaked with CD2301 and spotted with neat phiCD2301 samples. X indicates the locations at which the neat phage preparation was spotted. Zones of clearing on the L14 plate showed only single colony growth in two spots, whereas the phiCD2301 spots were almost indistinguishable from the surrounding lawn except for tide marks caused by the addition of the phage samples.

Single colonies were found within the phiL14WAR zones of clearing which were streaked onto a fresh BHIS-CC agar plate. Samples were also taken from the turbid phage zones of the CD2301 plate as confirmation of the method's efficacy. 20 colonies from each set were re-streaked to purity and PCR-screened for lysogens using primers specific to the phage DNA for each condition (Figure 3.22). The same colony lysates were also tested using primers targeting the *rpsJ* gene, which is common to *C. difficile* genomes to confirm that lysis of the cells had occurred prior to the PCR.

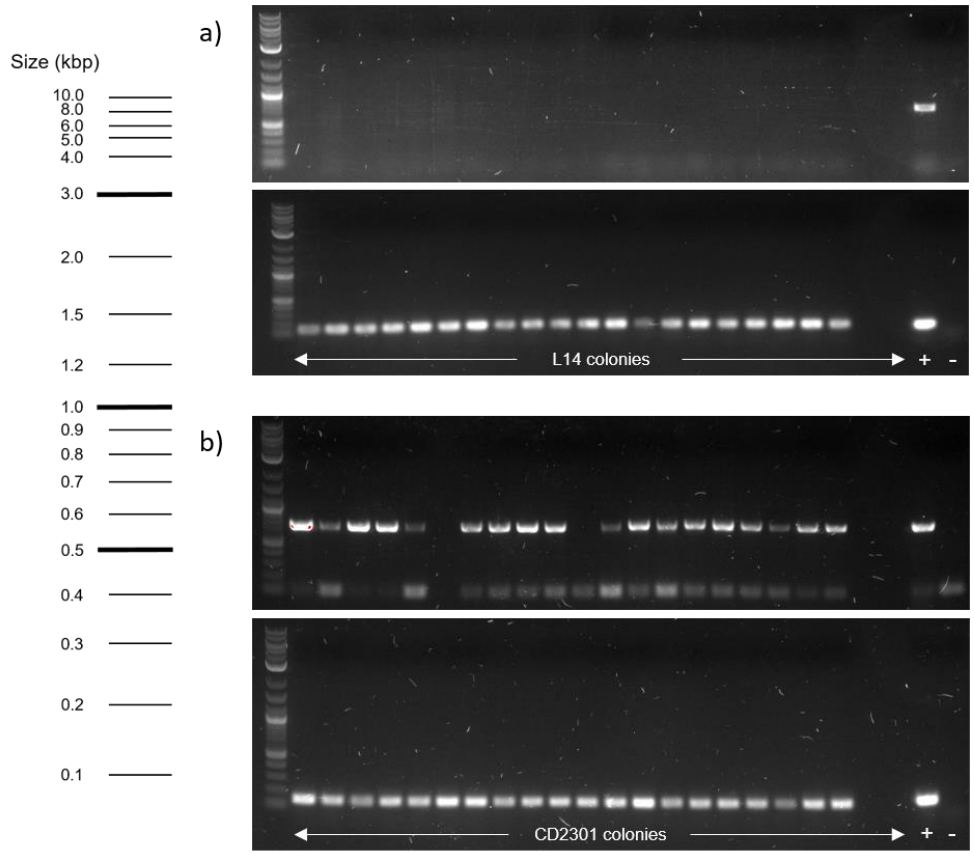


Figure 3.22 - a) Screening PCR of the potential L14 lysogen colonies using both phage specific primers (top) and C. difficile genome-specific primers (bottom). b) As per (a) but CD2301 lysogen colonies.

As expected given the result from plating, no lysogens were obtained for the L14 colonies challenged with phiL14WAR using this method. The CD2301 colonies, on the other hand, were positive for lysogens for 18 of

the 20 colonies selected, demonstrating that this is an efficient method of obtaining lysogens for temperate phages.

These experiments have been replicated independently by both myself and my colleague Joanna with no lysogens as yet identified from more than 100 colonies selected post-infection. These results support the notion that phiL14WAR is unable to form a lysogen.

3.2.9 Efficiency of action on a live culture

As well as establishing the inability of phiL14WAR to create stable lysogens by plating methods, the action of the phage on live bacterial cultures was assessed to quantitatively establish the degree of antimicrobial action against different *C. difficile* strains. The phage was initially tested by measuring the change in density of a culture of L14 over time when infected with phiL14WAR (Figure 3.23).

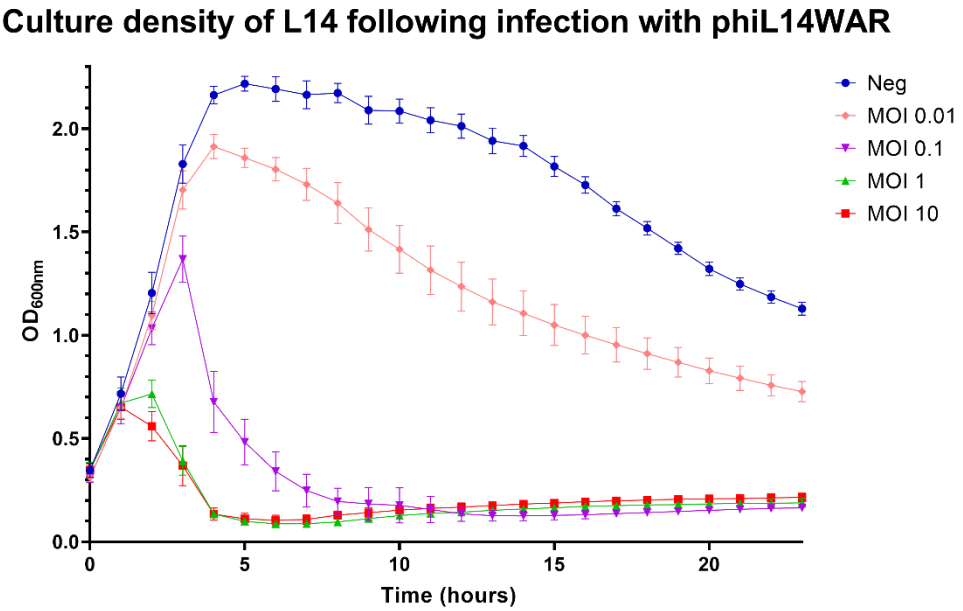


Figure 3.23 – Mean culture densities (OD_{600nm}) of triplicate L14 cultures infected with varying MOIs of phage phiL14WAR. Blanks were only included in duplicate for the analysis due to unreliable results in the final triplicate which was excluded. Error bars indicate standard deviations of the means.

The cultures acted as expected, with the MOI 0.01 treatment showing a small amount of reduction in culture density in comparison to the negative control, while MOIs of 1 and 10 elicited the greatest cell death in the shortest time (almost total culture death within 4 hours). An MOI of 0.1 also caused almost complete culture death, but the lowest point was not reached until approx. 10 hours post-infection, indicating that this dosage would have a similar impact but would need a much longer period to be completely effective. Unexpectedly, there was very little difference in effect when comparing the MOI 1 and MOI 10 results, indicating that an MOI of 1 at the target site should be sufficient to obtain the maximum benefit in any potential phage therapy treatment. Marginal culture growth was observed after the lowest point, but this could be attributable to naturally resistant colonies (i.e. through developed immunity) rather than any lysogen formation.

For comparison purposes, four further *C. difficile* strains were tested using the same method, as well as phiCD2301 to demonstrate the action of a known temperate phage on a live culture. These strains covered a range of EOP and ribotype options from those tested during host range testing – CD196 and R12087 are both RT027 strains which gave a higher EOP than the isolation strain, L28 is an RT001 with EOP of 4×10^{-2}, and CD8601 is an RT078, EOP 6×10^{-4} (Figure 3.24).

The strains treated with phiL14WAR gave a range of results. Both of the high EOP strains acted in a similar fashion to the isolation strain, across roughly the same time scale. In each case, MOI 1 and MOI 10 resulted in almost complete culture death, with a subsequent very gradual increase in OD_{600nm} across the course of the remaining assay period. The MOI 0.1 conditions varied more, however, with a mid-range reaction in CD196, but with a much lower effect in R12087. The MOI 0.01 conditions showed little variation from the negative control. While it is surprising that the culture death was not faster in the higher EOP strains, these results do support the idea that an MOI of 1 would be a suitable target dosage to consider for phage therapy.

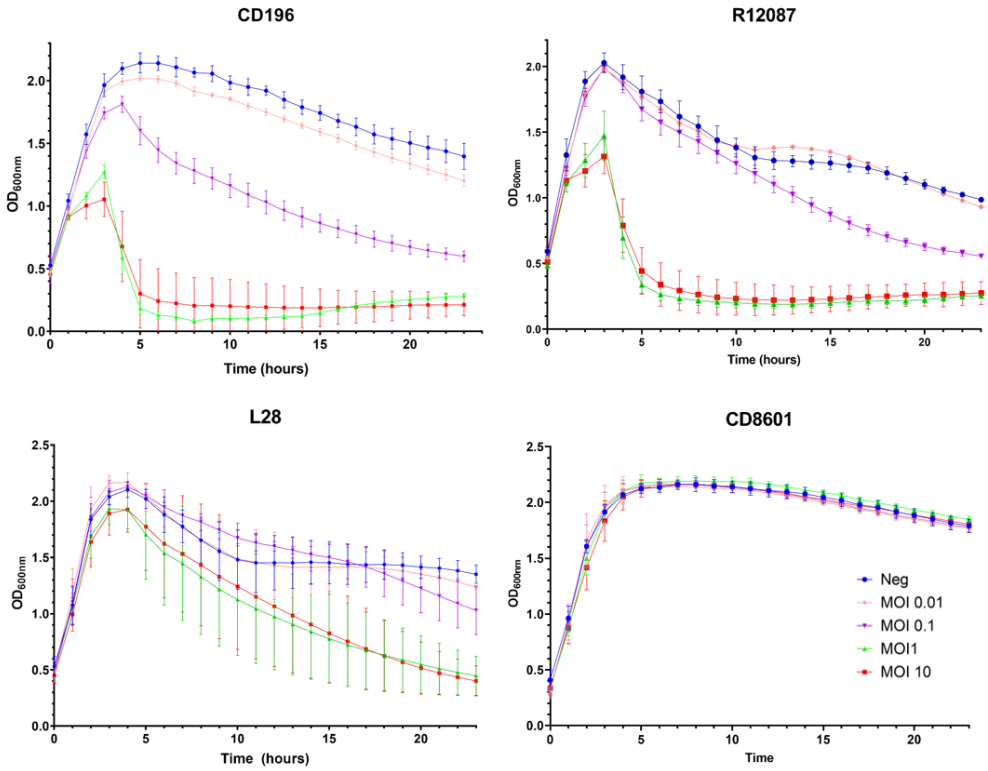


Figure 3.24 – Mean culture densities (OD_{600nm}) of four additional C. difficile strains (in triplicate) which are susceptible to phiL14WAR - CD196 and R12087 (RT 027), L28 (RT 001) and CD8601 (RT 078). Error bars indicate standard deviations of the means.

L28, however, gave a much slower reaction to infection with phiL14WAR with a greater degree of inter-replicate variation - one culture showed a clear effect from MOI 1 and 10 akin to that seen in L14, CD196 and R12087, but only reaching the lowest culture density around 20 hours after infection (although the majority of the reaction occurred within the first 10 hours) while MOIs 0.1 and 0.01 showed little to no variation from the negative control; the other two replicates also demonstrated an overall decrease in culture density that could be attributed to the action of the phage at MOIs 1 and 10, but at a much slower rate to the point where culture death was not achieved within the 23 hours of the assay period. This implies that, again, MOI 1 and 10 are the most effective dosages, but that in this case, the lower infectivity of phiL14WAR for this strain may be inhibiting the speed at which the phage can kill the culture

(and thereby increase the time it would take to clear an infection during phage therapy).

The above possibility is further supported by the impact of phiL14WAR on CD8601. At an EOP of 6 x 10^{-4} (essentially nearly 10,000 times less productive than in the isolation strain), it was expected that there may be a minimal effect of the phage on this strain. This was borne out when there was no difference in culture density between conditions across the assay period. A second attempt at the assay confirmed the result. This result either implies an error during the host range testing or that an EOP of 10^{-4} is too low to make a meaningful impact on an *in vitro* culture of this nature. In the latter case, an impact may be more likely for *in vivo* studies where culture conditions are more stable and results could be assessed over a longer period.

The final strain tested (CD2301) demonstrates our expectations of the action of a temperate phage (Figure 3.25). While the MOI 0.01 treatment follows a similar path to the negative control, as seen in the other strains, the MOI 0.1, 1 and 10 conditions do not eradicate the culture to the same extent as phiL14WAR. For each of these conditions, the lower limit at which culture death appears to cease is higher (OD_{600nm} of 0.4 as opposed to 0.1 in L14 infected with phiL14WAR), with both MOI 1 and MOI 10 demonstrating marked increases in culture density almost immediately. The degree of growth in the MOI 10 cultures then exceeded the culture densities of the MOI 0.1 and 1 conditions within approximately 6 hours. This reduced culture death and subsequent increased growth are consistent with lysogens taking over the culture, with the MOI 10 test supporting previous conclusions that higher MOIs may promote lysogen formation in some phages (Golding et al., 2019).

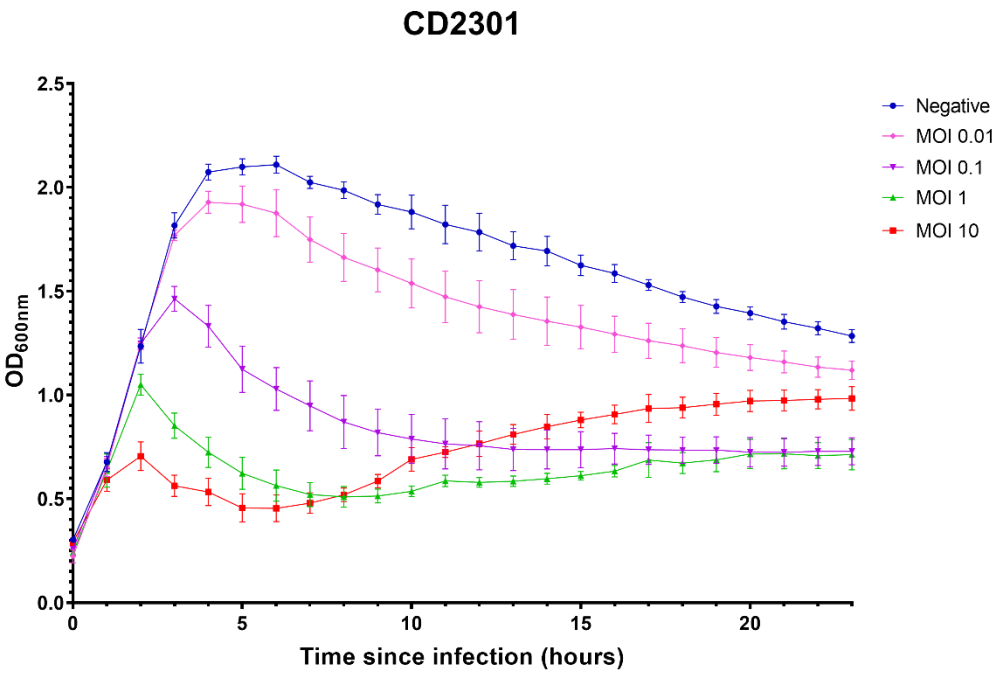


Figure 3.25 – Mean culture densities (OD_{600nm}) of triplicate CD2301 cultures when infected with known temperate phage phiCD2301 over a 23-hour period. Error bars indicate standard deviations from the means.

In order to quantitatively evaluate the differences in infectivity between strains, as well as allow some comparison to the infectivity of phiCD2301 without a shared host, the local virulence (Storms et al., 2020) of phiL14WAR in each strain was calculated, which is also known as the Phage Score (Konopacki et al., 2020). This method assesses the change in cell biomass obtained from infected samples in comparison to the uninfected negative control using the areas under the infectivity curves. It is a precursor to calculating the virulence index (V_p) which has been identified as one of several ways in which phages could begin to be quantitatively compared to assess their suitability for therapeutic use against a target pathogen. The local virulence is calculated using the following equation:

$$v_i = 1 - \frac{A_i}{A_0}$$

where A_i is the area under the infected curve and A_o is the area under the negative control curve. This provides a measure of the virulence between 0 and 1, with 1 representing instant death of cells within the culture and 0 being no death at all.

For the purposes of this analysis, two evaluations were considered. First, a subset of the data was used that encompasses the results up to the point at which the negative culture reached stationary phase (5 hours for L14, CD196, CD8601 and CD2301, 4 hours for L28 and 3 hours for R12087 - Figure 3.26). This limitation was recommended by Storms et al. (2020) in order to allow cross-comparison of phages by only considering the infectivity during the growth stage of the culture, the important stage for reproduction of phages. The calculated v_i for each MOI tested in the strains is available in Appendix 7.3.

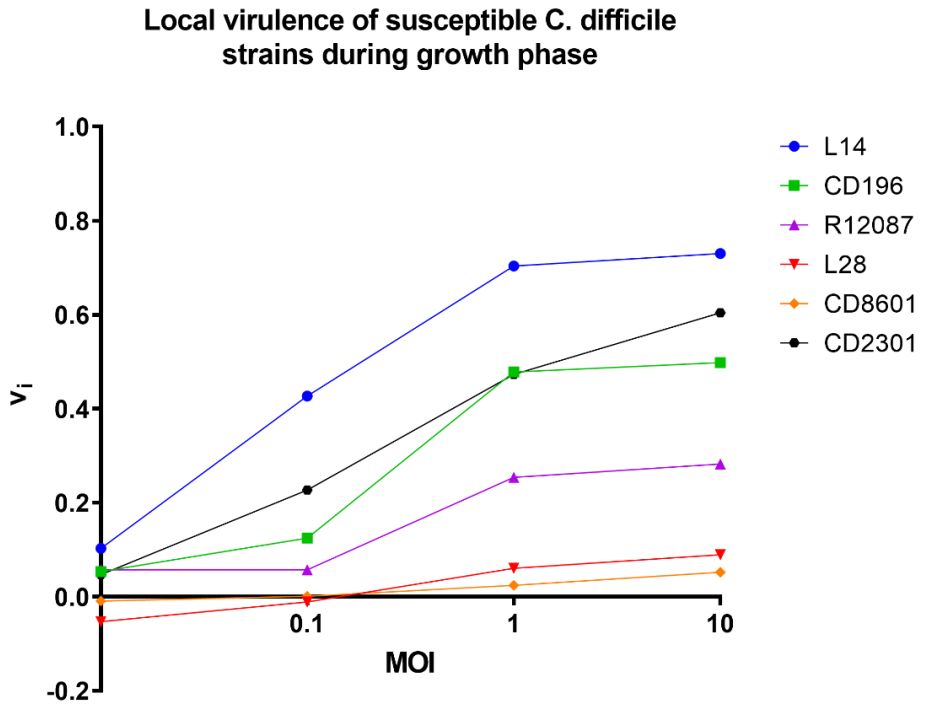


Figure 3.26 - Comparison of local virulence between susceptible strains of C. difficile using data up to the onset of the stationary phase in the negative culture. v_i represents the biomass in the cultures by calculating the area under the infection curve as a proportion of the area under the uninfected control curve for each MOI tested.

The results of this analysis support the previous observations that higher MOIs are more effective at reducing cell biomass quickly, with the MOI 1 sample in L14 generating an approximately 70% reduction in biomass within 5 hours in comparison to the negative control, as opposed to just 10% in the MOI 0.01 condition. The MOI 10 condition, however, only increased biomass reduction to 73%, indicating that anything higher than MOI 1 does not provide a great deal of further bacterial clearance. This pattern is reflected in most strains to varying degrees, including in L28 (MOI 10 – 9%, MOI 1 – 6%) and CD8601 (MOI 10 – 5%, MOI 1 – 2%).

PhiCD2301 infecting CD2301 also shows a 60% reduction in biomass using MOI 10 across this limited dataset, and 47% at MOI 1. However, comparison of this result to the initial infection curves highlights the disadvantage of using the stationary phase limitation when assessing the suitability of temperate phages for phage therapy, as it ignores the subsequent stabilisation and then increase in biomass resulting from lysogeny in this strain at higher MOIs. To account for this, an additional evaluation was conducted using all available data (Figure 3.27).

Using the full 23 hours of data produces results which more closely reflect the observations made from the initial infectivity curves, and there are a few particular differences of note. First, the effects of MOI 0.1 are more pronounced, showing between 2- and 5-fold increases in biomass reduction for all strains except L28 when compared to that captured by the limited analysis. R12087 also shows an overall increase in biomass reduction across all MOIs, bringing it more closely in line with CD196 and L14 for effectiveness over a longer time period. This is a product of the high rate of cell death in the early stationary phase of the culture, and is a valuable point to note when considering this phage for phage therapy against R12087, as the limited data set has essentially masked the effectiveness of this strain to a degree.

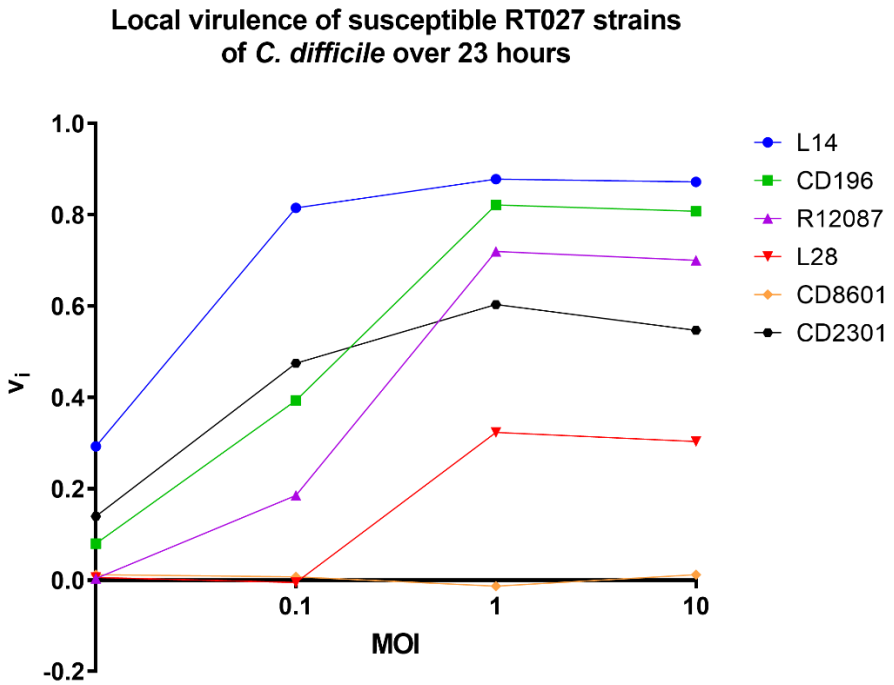


Figure 3.27 - Comparison of local virulence between susceptible strains of C. difficile using all available data. v_i represents the biomass in the cultures by calculating the area under the infection curve as a proportion of the area under the uninfected control curve for each MOI tested.

Similarly, the expanded data set showed a decrease in effectiveness of CD2301 relative to the other strains, with the v_i holding steady at 50-60% despite the increase in time points. This is a reflection of the increase in cell growth once lysogens began to dominate the culture and is an important characteristic of the infection curve to capture when considering phage therapy. Therefore, in my opinion, the local virulence in its limited form is only an informative statistic for phage therapy when applied to obligatory lytic phages, not temperate varieties, as it only considers the initial lytic activity of the phage but doesn't allow time for development of the lysogens of which medical professionals are justifiably wary.

However, while the local virulence is of limited use for assessing therapeutic potential with respect to temperate phages, it does provide a more easily interpreted analysis of the effect of phiL14WAR on the various tested *C. difficile* strains, and confirms the high antimicrobial

activity in L14, CD196 and R12087, particularly at MOIs of 1 and 10. For additional confirmation, the virulence index was calculated for all strains infected with phiL14WAR using the following equation, as detailed by Storms et al. (2020):

$$V_p = \frac{A_p}{A_{max}}$$

where A_p is the area under the local virulence curve of the phage and A_{max} is the area under the curve of theoretical maximum virulence (i.e. a v_i of 1 at all MOIs tested). For this calculation, the $\log^{10}$ of the MOI was used, as recommended, to normalise the weighting of the MOI values in the calculation. The area under the curves was calculated using GraphPad and the results are shown in Table 3.4:

Table 3.4 - Virulence indices of C. difficile strains infected with phiL14WAR

	Strain				
Virulence index	**L14**	**CD196**	**R12087**	**L28**	**CD8601**
Limited analysis	0.517	0.293	0.160	0.044	0.018
Unlimited analysis	0.760	0.553	0.420	0.158	0.007

Once again, the virulence index confirms quantitatively what the infection curves expressed but now in a single value which encompasses all of the MOIs tested. Interestingly, however, L14 is calculated as the highest virulence of the strains tested despite CD196 and R12087 providing higher EOP results for phiL14WAR. This could be a facet of the variability of the EOP testing by nature, and further testing would be needed to investigate this discrepancy further, if required.

The action of phiL14WAR on live cultures of susceptible *C. difficile* strains shows a high degree of lytic activity despite only the absence of

the integrase gene having been confirmed, which is supported by high virulence indices in 3 of the 5 strains tested. The overall profile of the phiL14WAR infections in comparison to phiCD2301 is also highly supportive of the notion that phiL14WAR is lytic, with subsequent small increases in culture density of phiL14WAR being attributable to a small number of naturally resistant colonies as opposed to the more widespread lysogeny expected from phiCD2301. Overall, phiL14WAR appears to be a reasonable candidate for further work to optimise it for phage therapy purposes.

3.2.10 16S testing of strain L12

As previously indicated, the original isolation experiments targeted RT 027 strains in the SBRC culture collection. As part of this work, we sought to ensure that the isolation strains were indeed RT 027 by sending them to the *Clostridioides Difficile* Ribotyping Network (CDRN) at the UKHSA (previously Public Health England) Microbiology Services laboratory in Leeds. The results confirmed that L14 and L15 were RT 027 strains; however, the strain believed to be L12 was identified as an 078 ribotype (another epidemic strain of *C. difficile* which is of interest to the CRG). Given the clinical importance of RT 078 strains, testing was continued on these phages.

For most of the project, the work on the L12 phages was conducted using the same personal strain stock created for the initial isolation experiment but, in the final year of the research, a new phage preparation was required for further test assays. A fresh stock of L12 was grown on a BHIS-CC plate from the culture collection and used for a plaque assay to begin increasing the titre of the existing phage stocks. However, when the plaque assay was conducted, no plaques were seen. The assay was repeated with another fresh stock from the culture collection, with the same result. Plaques were only obtained when the phages were tested against the old personal -80°C stock of the bacteria (named JS50 thereafter) which had been created during the isolation process.

JS50 (personal stock), L12 (from the culture collection) and a known RT 078 strain, CD1801, were then screened using 16S universal primers and the fragments sent for Sanger sequencing. The resulting sequences were submitted to the Ribosomal Database Project (RDP) website for analysis, which provided a list of the strains that the amplified fragments matched most closely (Table 3.5 to Table 3.7).

Table 3.5 - 16s rRNA sequence match results for strain JS50

S_ab score range	**Organism**	**No. of matches**
0.856	*Peptoclostridium difficile*[1] 630; AM180355	4
0.853	*Peptoclostridium difficile* CF5; FN665652	1
0.853 - 0.859	*Peptoclostridium difficile* M120; FN665653	6
0.853	*Peptoclostridium difficile* M68; FN668375	3
0.856	*Peptoclostridium difficile* BI9; FN668944	1
0.856	*Peptoclostridium difficile* 630; CP010905	2
0.856	*Peptoclostridium difficile*; 630Derm; LN614756	1
0.856	*Peptoclostridium difficile* ATCC 9689 = DSM 1296; DSM1296; CP011968	2

Table 3.6 - 16s rRNA sequence match results for strain L12

S_ab score range	**Organism**	**No. of matches**
0.843	*Clostridium difficile*; VPI 10463; AF072473	1
0.843	*Clostridium difficile*; 79685; AF072474	1
0.859	uncultured bacterium; TS51_a02d06; FJ369592	1

[1] *Peptoclostridium difficile* was an interim name for *C. difficile* established by Yutin and Galperin following a 16s rRNA-based reassessment of the Clostridia genus in 2013. This name predates the reclassification to *Clostridioides difficile* suggested by Lawson et al., (2016)

0.844	*Peptoclostridium difficile* CD196; FN538970	3
0.843 - 0.844	*Peptoclostridium difficile* R20291; FN545816	7
0.843	[*Clostridium*] *difficile*; JCM 1296; AB548672	1
0.906	*Clostridium difficile*; DSM 12056; HM007570	1
0.843	*Peptoclostridium difficile* 630; AM180355	1
0.87	uncultured bacterium; 16slp5_06d05.p1k; FJ505390	1
0.844	*Clostridium difficile*; 35; JF693896	1
0.843	*Clostridium difficile*; 38; JF693897	1
0.844	*Peptoclostridium difficile* BI1; FN668941	1

Table 3.7 - 16s rRNA sequence match results for strain CD1801

S_ab score range	**Organism**	**No. of matches**
0.979	*Clostridium difficile*; DSM 12056; HM007570	1
0.981	*Peptoclostridium difficile* 630; AM180355	4
0.978	*Clostridium difficile*; 35; JF693896	1
0.981	*Clostridium difficile*; 38; JF693897	1
0.977	*Peptoclostridium difficile* CF5; FN665652	1
0.982 - 0.984	*Peptoclostridium difficile* M120; FN665653	5
0.977	*Peptoclostridium difficile* M68; FN668375	1
0.981	*Peptoclostridium difficile* BI9; FN668944	1
0.981	*Peptoclostridium difficile* 630; CP010905	2
0.981	*Peptoclostridium difficile*; 630Derm; LN614756	1
0.980 - 0.981	*Peptoclostridium difficile* ATCC 9689 = DSM 1296; DSM1296; CP011968	2

Strain JS50 and CD1801 shared several strains in common, including M120, the type strain for RT 078. The fragment from L12, however, was almost completely different, sharing only *C. difficile* 630 with both JS50 and CD1801, and *C. difficile* 35 and 38 with CD1801 alone. Notably, L12 shares similarities with R20291 and CD196, both of which are known RT 027 strains.

This analysis indicated that the strain which was used to isolate phages phiL12S and phiL12W was not, in fact, L12 as it was believed to be, and confirms the results of the original ribotyping analysis undertaken in 2019. At this stage, it is not possible to determine which strain of *C. difficile* was used for the isolations beyond that it is one of the susceptible strains listed in the host range analysis. The matter was discussed with Joanna and my supervisory team, and it was decided to re-make the lysogens of these phages in a known susceptible strain (L35), but that the original JS50 lysogens would also be retained. The EOP data previously generated is correct in that the EOP was calculated through comparison to the isolation strain, however, it is advised that strain L35 be used for future titre increase of these phages. These phages have now been renamed phiJS50S and phiJS50W for clarity.

3.3 Discussion

Bacteriophage therapy holds a great deal of potential as an alternative treatment for antibiotic-resistant infections such as *C. difficile*. However, while the specificity of phages is beneficial for medical applications, it also complicates their use. Treatment of an infection requires possession of a lytic phage that targets the bacteria in question, which is unlikely unless that strain or a closely related one has been treated previously. This can lead to delays while such a phage is identified. Regular and consistent isolation of novel phages to form libraries of potential therapeutic options is therefore a key route to making phage therapy viable.

Phage therapy for CDI, in particular, requires substantial investment in isolating novel phages, as the existing options are all temperate in nature. The ability to integrate into the bacterial genome and provide super-immunity from subsequent phage attack undermines their use as a therapeutic, as a subpopulation of bacteria is left behind to continue to cause symptoms. In 2018, Nale et al. conducted a study into whether a suitably diverse cocktail of temperate phages would produce an acceptable result, under the assumption that where one phage caused super-immunity, one of the others would result in lysis of the bacteria. However, this approach still carries the risk of an escape population, and so may struggle to be accepted as a solution. As such, further isolation attempts are necessary to try and identify new *C. difficile* phages which are obligatory lytic.

In this chapter, samples from seven wastewater treatment plants across the Thames Water service area were used in an attempt to isolate new phages infecting RT 027 strains of *C. difficile*. These clinically relevant strains were responsible for an epidemic across the USA, Canada and EU at the start of the 21st century (Alcalá Hernández et al., 2017) but the SBRC culture collection did not contain any examples of phages infecting this ribotype. Three novel phages were isolated using these samples.

3.3.1 Siphovirus phages phiJS50S and phiJS50W

Two siphovirus phages (characterised by a long, non-contractile tail) were isolated from wastewater samples taken from two different locations – Wargrave and Swindon. DNA sequencing showed both phages to be approximately 35kbp in size with a high degree of identity between them (over 90%), such that they are likely to be close evolutionary relatives. However, the degree of difference at a nucleotide level between the phages, alongside differences in the host range and the fact that they were isolated from locations approximately fifty miles apart, imply that these can be considered as distinctly different phages and so have been named phiJS50S (previously phiL12S) and phiJS50W (previously phiL12W) after their strain and location of recovery.

Sequencing also identified the presence of tyrosine integrase genes in each phage genome, indicating that these are temperate phages as would normally be expected for phages infecting *C. difficile*. These integrase genes are genetically identical, further supporting the idea that the two phages are the product of recent evolutionary divergence. However, while both phages possess a series of transcriptional regulators that could be involved in the initiation and maintenance of lysogeny, only two of the regulators in phiJS50W have been identified as being of the same family as the phage λ *cI*/*cro*-type regulators (XRE family transcriptional regulators). These may prove to be initial targets for genetic modification in future, although further investigation would be required to confirm whether or not these two genes were involved in the maintenance of lysogeny.

Other than the integrase and transcriptional regulators, the genomes of both of these phages consist of genes encoding highly homologous structural genes and enzymes, as well as hypothetical proteins for which the function has yet to be determined. It is possible to infer the function of some hypothetical proteins by identifying homologous protein domains in other organisms, but many show no such similarities and therefore remain unidentified. This is characteristic of annotated phage genomes in online databases with approximately 70% of phage genes being identified as “hypothetical” (Van den Bossche et al., 2014).

Phylogenetic analysis of the siphovirus phages shows that their closest identified relative is phiCD6356 which was isolated at the Teagasc Food Research Centre in Cork, Ireland (Horgan et al., 2010). While it is not specifically stated in the literature what samples were used to isolate the host strain of *C.difficle* (DPC6356) it is likely given the location that this was isolated from livestock. If so, then two further points of interest should be noted. Firstly, these phages may have been isolated from two different *C. difficile* reservoirs (humans and livestock). If phiJS50S and phiJS50W were capable of infecting DPC6356 as well then this would be an example of how phages of interest to human medical science may also

have uses in other arenas and vice versa. The crossover between reservoirs is a point that should always be kept in mind when considering medical phage applications, as it provides a greater breadth of phage to consider for therapeutics. A good example of this is the recently developed Actiphage® system which uses the same D29 *Mycobacterium* phage to help diagnose tuberculosis in both bovine (Swift et al., 2020) and human blood samples (Verma et al., 2020). The second point of interest is the substantial distance between the isolation locations of our phages and their relative (sewage samples from southern England as opposed to lysogens in Ireland). If DPC6356 was isolated from cattle then the distance could be a result of the carriage of the lysogens by the livestock as they moved locations. This would be a good example of one of the benefits of lysogeny to the virus – carriage to new susceptible populations without risk to the virus itself.

As would be expected based on their genetic similarity, the host ranges of these two phages appeared comparable, with each infecting four ribotypes (001, 014, 027 and 078). PhiJS50S gave positive results for 43 of the tested strains, 11 of which were at an equal or higher level than the propagating strain. PhiJS50W infected 40 of the tested strains, and these were mostly the same strains infected by phiJS50S. 24 of these strains (mostly the clinically-relevant RT 078) showed a higher EOP for the propagating strain than the isolation strain.

However, even with a reasonable host range, temperate phages such as these would not be considered ideal candidates for therapeutic use. They may be of use in future work to study phage biology further, or as potential subjects for genetic modification, particularly in the case of phiJS50W given its larger host range and higher infectivity. Lysogens have therefore been created using these phages which have been stored in the SBRC culture collection for future study.

3.3.2 Myovirus phiL14WAR

The final phage obtained was a myovirus-type phage isolated using strain L14 (also known as CD60902) and the wastewater sample from Wargrave. Genetic analysis showed this phage to be just under 46kbp in length with a genome mostly consistent with other phages of *C. difficile*. A phylogenetic analysis identified phiCD119 (Govind et al., 2006) and phiCDHM19 (Hargreaves et al., 2014) as the closest relatives. The identity between phiL14WAR and these two phages is, however, reasonably low (approximately 50-60%) indicating that there are likely many intermediary phages which have not been identified/published. Direct comparison between the three phages shows that most of the homology occurs in the structural genes. Beyond this, there is a large section of DNA in the regulatory region which is shared between phiCD119 and phiCDHM19 but shows limited homology with phiL14WAR. Examination of this area shows that the gene-types in this region are similar to the ones in phiCD119 and phiCDHM19 but they lack the nucleotide/protein homology of closely related phages. The host range was also measured and showed that phiL14WAR infected all but four of the 52 RT 027 strains tested from the SBRC culture collection, including six which showed equal or higher titre from the propagating strain than the L14 isolation strain. It also infected strains from three other ribotypes (001, 078, 087), but to varying degrees.

Unlike the other phages isolated in this work, the sequenced genome of phiL14WAR contained no identifiable integrase gene when compared to seven different online protein and nucleotide databases, even with HMM searching. Natural loss of lysogeny genes from a temperate parent resulting in a lytic phage is not unheard of (Pope et al., 2011), but due to the complexity of *de novo* genome assembly, a misassembly might have occurred at the section of the genome containing an integrase gene. As such, ONT MinION sequencing was undertaken in an attempt to use longer reads to cover larger sections of the genome, double-checking the original Illumina sequencing and minimising the chance of mis-assemblies.

The long read sequencing, alongside subsequent Sanger sequencing of the original phage preparation, confirmed the result of the Illumina short read sequencing (i.e. that there was no *int* gene), although doubt remained due to the potential for an unidentified integrase variant as suggested by the subsequent PhageLeads bioinformatics search. It was therefore necessary to test for lysogen formation experimentally. These tests resulted in no lysogens despite several attempts by two different researchers using two different methods, confirming that no integrase variant appears to be present.

The lack of an integrase gene, however, does not necessarily equate to a phage being lytic if it had a temperate parent or was previously temperate itself, as a study by Gottesman showed in 1974 where an integrase-deficient temperate phage was still able to enact its repression mechanisms because the rest of the lysogeny genes remained in place. A closer examination of the genome of phiL14WAR does show genes which could potentially be a part of a lysogeny mechanism, in the form of the DUF955 *IrrE*/*ImmA* metalloendopeptidase. The *ImmR*/*ImmA* system has previously been linked to protease-encoded lysogeny, as found in the *B. subtilis* mobile genetic element ICE*Bs*1 (Brady et al., 2021a), however, the presence of a gene with a similar domain does not necessarily mean that the system is in place in this phage, and so experimental testing was needed to determine whether this phage acted in a similar fashion to a known temperate phage or more like a lytic one.

A series of bacterial reduction curves were conducted in which growing cultures of susceptible strains were challenged with different concentrations of phiL14WAR (MOIs of 0.01, 0.1, 1 and 10). Temperate phages would be expected to kill some of the culture but show an uptick in growth again once newly-formed lysogens were able to grow in place of the sensitive bacteria (Fillol-Salom et al., 2019), which was demonstrated in the CD2301 control culture that was infected with temperate phage phiCD2301. Lytic phages, however, should completely kill the culture over time. While complete culture death was not quite

achieved, infection with an MOI of 0.1 or higher did indeed result in rapid death of the culture with no significant resurgence in optical density within the 23 hours of the experiment. This result was replicated across three susceptible RT 027 strains (L14, CD196 and R12087), although a small amount of culture recovery was observed which is likely a result of developed immunity (Stern and Sorek, 2011). The overall profile of these three strains was markedly different from the temperate strain, with a greater degree of cell death and slower overall recovery lending weight to the hypothesis that phiL14WAR is either obligatory lytic or may have lost its lysogeny equipment during a mutation event. From a phage therapy perspective, the best MOI for this phage seemed to be a minimum of MOI 1, with 10 not delivering a substantially different result in the *in vitro* tests. However, recent modelling has indicated that a failure to obtain an MOI of 10 or more at the target site during live therapy may impair treatment effectiveness (Suh et al., 2022).

Two additional strains which were tested, L28 and CD8601, were also informative. Both of these strains had a significantly lower EOP than the isolation strain (10^{-2} and 10^{-4} respectively) and there was a markedly slower effect in the former, with maximum cell death occurring around 20 hours after infection rather than the 5 hours observed in L14, CD196 and R12087. CD8601 showed no effect at all of the infection with phiL14WAR. This reduced effectiveness in strains with a lower EOP implies a lower limit to the EOP that should be considered when assessing a phage for therapeutic purposes, which is consistent with other studies where the EOP lower limit for use in phage therapy has been defined as 0.01 (Haines et al., 2021).

However, the EOP should not be the only consideration. Evaluation of the tested strains using the local virulence/virulence index methods (Storms et al., 2020) demonstrated that higher or lower EOP does not necessarily impact directly on virulence, with CD196 and R12087 both producing lower biomass levels of reduction than L14 despite higher experimental EOPs. This may have been a facet of the variability

observed while attempting to measure the EOP thereby making it a poor reflection of infectivity, but there could be other elements involved which would require further investigation. In any case, L14 was deemed the most susceptible of the strains tested, generating a virulence index of 0.760, with CD196 (0.553) and R12087 (0.420) both showing reasonable susceptibility but nowhere near as high. Should this quantitative method of assessing virulence be more widely adopted, thresholds will likely be necessary in order to establish what level of infectivity is appropriate in order for a particular phage to be deemed a good candidate for therapeutic use. This assessment would not only need to include the virulence index of the phage, but also the toxicity of the targeted *C. difficile* strain, as a low to moderate reduction in infection by hypervirulent strains may still cause a beneficial clinical effect for the patient, even if it doesn't clear the infection completely. A phage may therefore not need to have a high virulence index in order to be deemed suitable for use against a particular strain.

Another consideration in terms of using this phage for phage therapy lies in some of the inconclusive results obtained from the host range testing. While 48 of the local RT 027 strains were identified as susceptible to phiL14WAR, some of these strains only showed a reaction at low dilutions. In these cases, lysis was observed down to a dilution of approximately 1/100, but at higher dilution factors the plaques were no longer discernible on the plates so it was not possible to accurately determine the EOP for these strains. This phenomenon has been observed in previous in-house work on *C. difficile* phages (Whittle, 2018), as well as in other species of phages. It was posited that the lysis observed at lower dilutions was a result of an abortive infection or lysis from without, rather than from a completed phage cycle releasing progeny virions (Kutter, 2009). Several explanations were suggested for this including prophages causing interruptions to cellular mechanics as a reaction to infecting lytic phage, an unrelated bacteriocin, or potentially an incompatible restriction-modification system (Kutter, 2009).

This may cause complications for the use of phiL14WAR against infections by some RT 027 strains, as it would require additional investigation as to whether the phage could still be effective for strains where this phenomenon is observed; for example, additional culture density tests such as those used for CD196 and R12087. Given the speed with which phage therapy would be desired to be ready, it is unlikely that such investigations would be undertaken at the point of need. *In vitro* characterisation of these anomalous strains to discern the mechanism causing abnormal infection in advance of such a requirement occurring would be beneficial but was outside the scope of the current project.

Despite this complication, isolation of phiL14WAR achieves the two goals set out within this chapter – 1) acquisition of a phage infecting the clinically relevant RT 027 strains of *C. difficile*, and 2) identification of a potentially lytic phage infecting this bacteria through environmental isolation. Further testing is required to definitively determine whether other lysogeny-maintenance genes are still present in the genome or have been lost alongside the integrase gene. However, given that the phage seems to be incapable of creating lysogens, alongside the greater culture lysis observed in comparison to a known temperate strain, this phage represents an important first step in realising the potential of bacteriophage therapy as a treatment for a clinically-relevant strain of *C. difficile*.

3.4 Chapter Summary

Isolation experiments using sewage samples taken from Thames Water wastewater treatment plants resulted in the isolation of three novel bacteriophages. Two of these (phiJS50S and phiJS50W) were characterised as temperate siphovirus phages primarily infecting the 078 ribotype strains of *C. difficile*. Each of these phages contains an identical tyrosine site-specific recombinase, and lysogens have been created and stored for future study of these phages.

The remaining novel phage (phiL14WAR) was characterised as a myovirus phage primarily infecting RT 027 strains of *C. difficile*. However, whole genome sequencing using both short Illumina and long ONT reads showed no identifiable integrase gene. While it is difficult to definitively prove that this phage is unable to form a lysogen (as it is not possible to rule out the potential for very low-efficiency integration), repeated experimental attempts were all negative for lysogen formation, lending weight to the idea that this phage does not contain even an unidentified variant of the integrase gene. One-step growth curves using various levels of phage infection were also consistent with what would be expected of a lytic phage. Further experiments can be undertaken in future to more fully characterise this phage, however, the current evidence is suggestive of this being the first documented non-temperate phage infecting *C. difficile*. It is therefore also a step towards realising the use of phage therapy for the treatment of CDI.

Chapter 4

CRISPR knockout of the *int* gene from bacteriophage phiCD2301

4 CRISPR knockout of the *int* gene from bacteriophage phiCD2301

4.1 Introduction

While we have had some success in isolating a potential lytic phage from the environment (Chapter 3), the lack of available and confirmed obligatory lytic phages for *C. difficile* (Heuler et al., 2021; Nale et al., 2016b) implies that they are exceedingly rare; potentially too rare for viable therapeutic use given their host specificity. As such, if phage therapy is to be implemented for CDI then consideration needs to be given as to how the currently available temperate phages may be used. As mentioned previously, some studies have taken place investigating whether cocktails of temperate phages could produce satisfactory clearance of infection (Nale et al., 2018, 2016b), but while there was some success in minimising the risk of transduction, it cannot be entirely negated while the ability to integrate remains. However, evolving techniques for genetic engineering may allow for existing phages to be modified to make them suitable for therapeutic use.

From the perspective of phage therapy, the most important genes in the lysis-lysogeny control system are the repressor (e.g. *cI* in *E. coli* phage λ) and the integrase genes. While it could be argued that removing the repressor gene alone could render a phage obligatory lytic (as the lytic genes would not become repressed) the continued presence of the integrase gene could cause issues.

Also known as a site-specific recombinase (Fogg et al., 2011; Groth and Calos, 2004), the integrase enzyme mediates both the integration and excision of the phage DNA into and out of the host genome (Gottesman, 1974) through the exchange of complimentary sites on the phage and bacteria (known as *attP* and *attB*, respectively). Before the advent of CRISPR mutagenesis, integrase was investigated extensively as a potential method of achieving targeted recombination (Fogg et al., 2011;

Platt et al., 2003; St-Pierre et al., 2013) and novel uses are still being identified today (Huang et al., 2019). The method through which it accomplishes integration varies depending on the type of integrase.

Serine integrases, such as that found in lactococcal phage TP901-1, have simple attachment sites consisting of inverted repeats. The integrase enzyme consists of four subunits, each of which attaches to a strand of DNA on the phage and host genomes at the attachment sites. The enzyme then cuts all four strands of DNA simultaneously before two of the subunits rotate, bringing the attachment sites from opposing DNA strands into alignment. These ends are then recombined, resulting in the phage DNA being bound to the host DNA (Figure 4.1). Subsequent excision requires a small recombination directionality factor (RDF) which changes the confirmation of the integrase enzyme, allowing it to recognise the new attachment sites (Bibb et al., 2005).

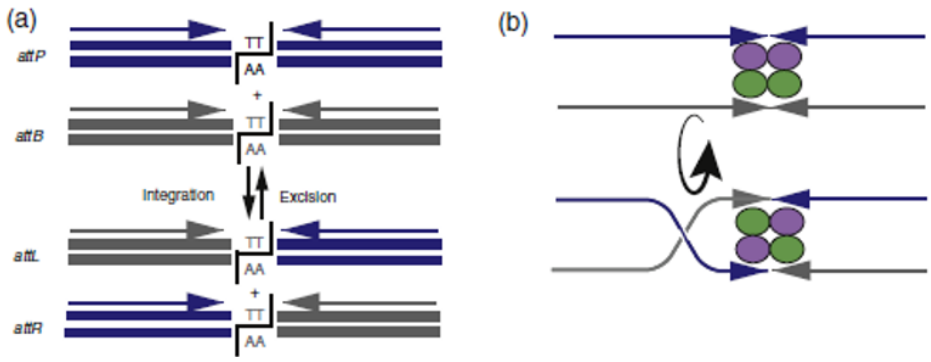


Figure 4.1 - Action of serine integrases, taken from Fogg et. al. (2014). a) Thick blue bands indicate the phage DNA double strand while thick grey lines represent the host DNA. The attachment sites in each organism (attP and attB) are complementary. b) Purple and green circles indicate the four subunits of the integrase enzyme which rotate to bring the complementary ends of the phage genome into contact with the corresponding host genome ends. The breaks are then recombined and the phage is integrated into the host genome, forming new sites attR and attL in preparation for later excision.

Tyrosine integrases, such as that found in phage λ, act in a more complicated manner. The attachment sites are composites of repeat units and asymmetrical core sequences, and the reaction relies on additional host-encoded integration factors (IHF) to bring the host DNA into contact with the integrase-*attP* complex (Cho et al., 2002; Fogg et al., 2014; Groth and Calos, 2004; Han et al., 1993). Once there, the integrase cleaves the DNA in two steps (Jayaram et al., 2015) – first, a single strand of the bacterial DNA is cleaved, allowing the free ends to bond covalently with the cleaved phage DNA to form a Holliday junction; a second cleavage and recombination of the remaining strands then occurs, resulting in the phage DNA becoming fully inserted into the bacterial genome (Figure 4.2). Reversal of the process relies on the help of a different enzyme called an excisionase, as well as IHFs, to bind to the changed attachment sites in order to allow induction of the prophage.

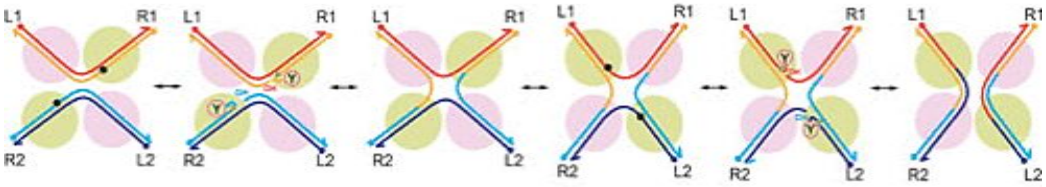


Figure 4.2 - Action of a phage λ tyrosine integrase enzyme, taken from Jayaram et al. (2015). The Holliday junction recombination intermediate is formed by four subunits of the integrase enzyme (circles) which bind to the bacterial and phage DNA at the appropriate attachment sites. The tyrosine catalytic site (Y) cuts a single strand of each DNA helix, allowing recombination before cutting the other strand to complete the integration.

Irrespective of the method through which the integrase operates, this enzyme is a key concern when considering the suitability of a phage for use in phage therapy. Integration into the host genome makes bacteriophages an important mediator of genetic mobility between bacteria, and they are known to be responsible for carrying harmful genes such as antibiotic resistance and toxin genes to new strains (a process known as transduction).

A primary example of this is Shiga toxin-producing *E. coli* (STEC).The toxin in this case is encoded on prophages within the genome, and production of the toxin is associated with induction of the prophage and subsequent transcription of the *stx* genes rather than through any action of the *E. coli* cell itself (Węgrzyn and Muniesa, 2021). A further example is the phage-mediated interplay between toxigenic *Clostridium novyii* type A strains and *Clostridium botulinum* types C and D. Bacteriophages are integral to the production of the characteristic toxin for each strain (NA1 for *C. novyii* type A, phage 3C for *C. botulinum* type C and phage 1D for *C. botulinum* type D), but work by Eklund et al. (1974, 1976) demonstrated that a non-toxigenic, phage-sensitive strain of *C. botulinum* type C (HS37) could be converted into either *C. novyii* type A through infection with NA1, toxigenic *C. botulinum* type C using phage 3C or type D when infected with phage 1D. Furthermore, the conversion was reversible, and once cured of the prophage, the same strain could be re-infected with one of the other phages to convert it to another strain.

Examples of the transduction of AMR genes can be found in food-borne pathogens as a result of either an integrated prophage providing resistances, or through pseudolysogeny where the phage can exist as a plasmid in the cell during a period of unfavourable growth conditions (Colavecchio et al., 2017). In addition, errors in excision of the prophage from the genome can result in bacterial genes in close proximity to the attachment sites being transcribed as if they were part of the phage genome, resulting in them being packaged alongside the phage DNA into the capsid and potentially integrated into the next host that the virion infects (Borodovich et al., 2022).

This transduction capability makes the integrase a required target for removal if we are to engineer an existing temperate phage with a view to phage therapy. While this gene remains intact, there is the chance that temporary integration could occur which would still enable transduction of undesirable genes even without a stable lysogen being formed (e.g. if the phage only integrates temporarily but still excises imprecisely). As

such, removal of the integrase gene is likely to be a necessity in order to obtain medical approval for the therapeutic use of any genetically modified phages, should this be allowed in future. This chapter therefore details the work undertaken to remove the integrase gene from an existing temperate phage using CRISPR-based genome editing.

4.2 Results

4.2.1 Selection of initial target bacteriophage

Easy conjugation of host strains was deemed important in ensuring the success of the planned work. Selection of a proof-of-concept phage was therefore undertaken by testing the conjugation capability of five *C. difficile* strains for which small temperate phages were available in the SBRC culture collection. The test was conducted by transforming the CRISPR test plasmid pMTL83151 (Heap et al., 2009) into both *E. coli* CA434 and sExpress, a chemically competent NEB Express derivative which contains the R702 conjugation plasmid (Woods et al., 2019). Following a standard 24 hour conjugation protocol, the number of transconjugant colonies obtained with five *C. difficile* recipients was counted across the five selection plates (Table 4.1).

Table 4.1- C. difficile strains selected for initial conjugation tests, along with their associated isolated phage size

Strain name	PCR Ribotype	Phage size (kbp)	No. of transconjugants	
			CA434	SExpress
08011	002	31.4	374	TNTC*
2301	014	38.7	268	TNTC*
2315	078	35	9	49
3801	078	35	5	68
8601	078	35.5	4	92

*TNTC = Too numerous to count.

CD08011 and CD2301 appeared to be the most efficient recipients out of the strains selected, however, the colonies obtained with CD08011 exhibited different morphology to those of the other four strains. Re-

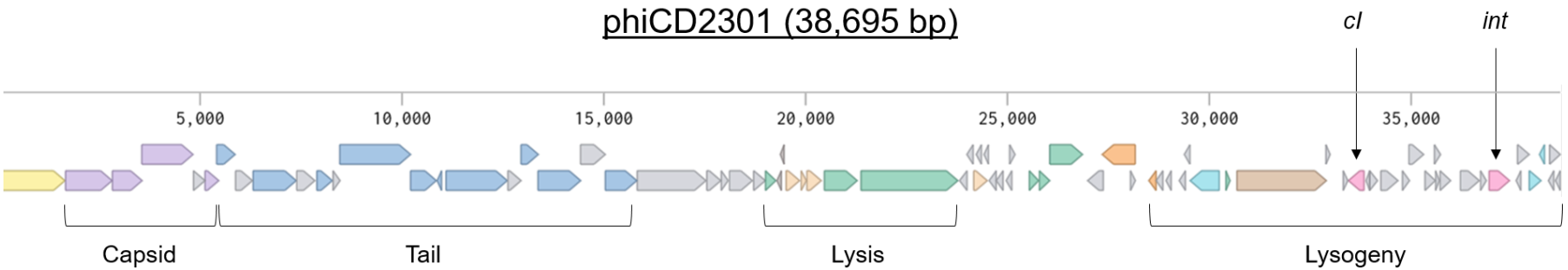


Figure 4.3 - Genome of phiCD2301 initially annotated by Whittle et al. (2022) and re-evaluated as part of this study. Phage genomes are modular in arrangement, with structural genes placed together followed by lysis genes. In addition to those labelled above, pink genes are those believed to be related to lysogeny, while grey genes are those identified as hypothetical proteins using the Artemis Comparison Tool (Carver et al., 2005) and the NCBI BLASTp database. Orange indicates genes with identified domains but currently unknown function, while other colours represent known function enzymes and miscellaneous proteins.

streaks of these transconjugants failed to grow on fresh BHIS-CC-Tm and so it was determined that these were likely to be background *E. coli* contaminants. As such, it was determined that the phage infecting CD2301 (phiCD2301 - Whittle et al., 2022) would be the optimal test candidate for this study (Figure 4.3). The published annotations for this phage clearly identified the integrase gene near the end of the genome, allowing easy targeting.

4.2.2 Vector creation and conjugation

CRISPR vectors were created targeting the integrase gene using three sgRNA fragments. These vectors were named pRECas1_phiCD2301int_sg1, pRECas1_phiCD2301int_sg2 and pRECas1_phiCD2301int_sg3 (Figure 4.4).

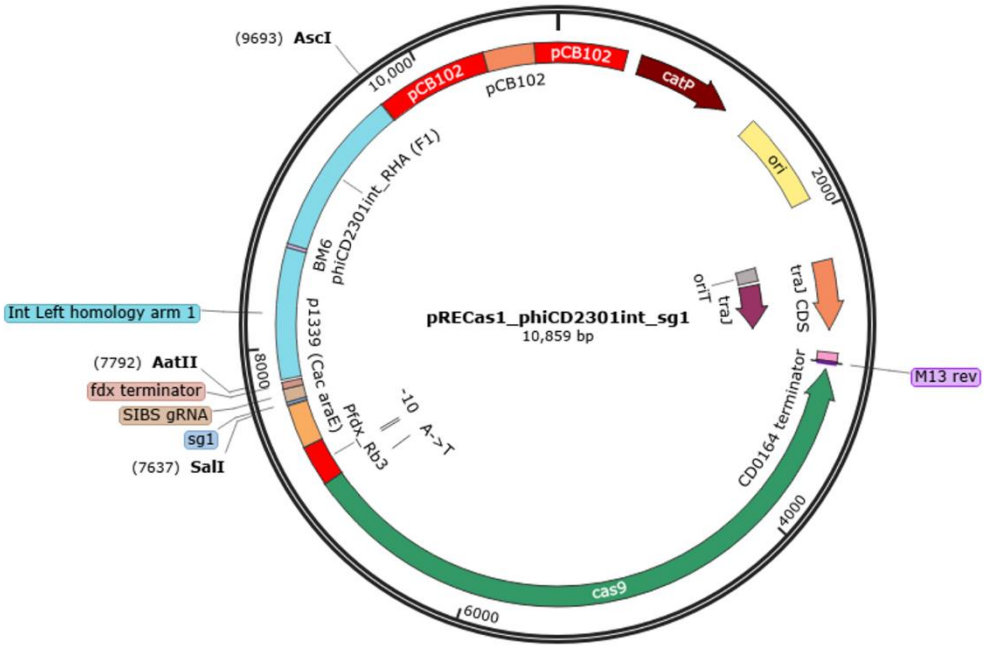


Figure 4.4 - Map of pRECas1_phiCD2301int_sg1 plasmid. The homology cassette and sgRNA1 are under the control of the P_{araE} promotor, while the Cas9 is under the control of the riboswitch $P_{fdx\text{-}E}$. catP is included as an antibiotic resistance marker, and oriT and traJ allow for conjugation between the E. coli and C. difficile. pCB102 and ColE1 act as the replicons in C. difficile and E. coli, respectively.

Once the three knockout vectors had been created, they were each transformed into chemically competent *E. coli* XL1 Blue cells. 24 transformants from each vector were checked for the presence of the insert using a DreamTaq colony PCR with primers flanking the insert site on the plasmid (Figure 4.5).

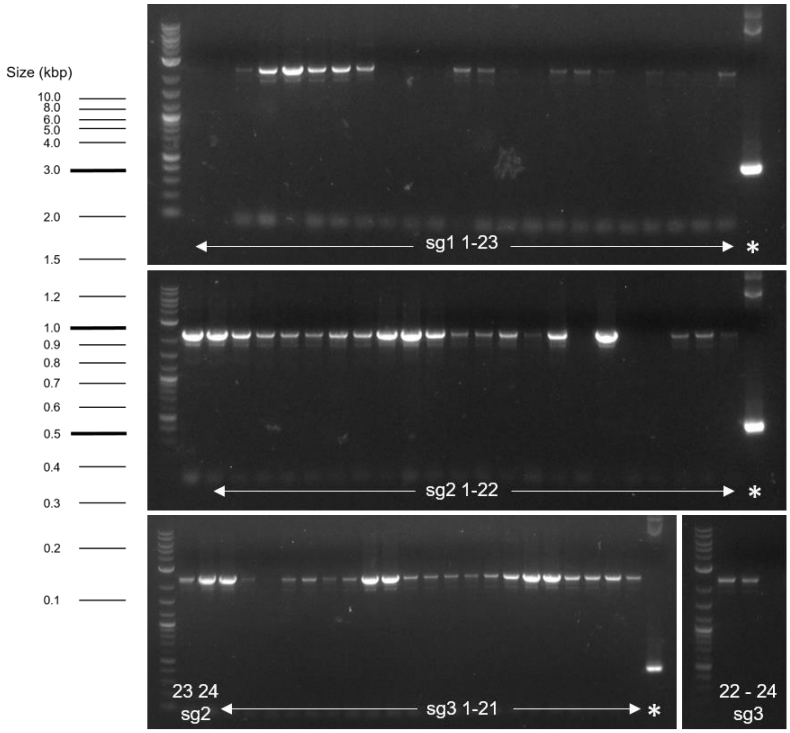


*Figure 4.5 - Gel electrophoresis of 24 transformant XL1 Blue colonies from each of the three integrase knockout vectors. * denotes the undigested vector run alongside as negative control. The homology cassette and guide RNA modules were approx. 2kbp in size. Undigested pRECas1_MCS vector run alongside as negative control.*

The amplicons of the two positive colonies with the strongest results for each vector were excised and the DNA extracted. The fragments were sent for Sanger DNA sequencing, and the results aligned to an expected plasmid map using Benchling.com to determine which colonies contained accurately constructed knockout vectors. The correct vectors were then extracted from new cultures of the XL1 Blue cells and transformed into chemically competent sExpress *E. coli* cells. Unfortunately, transformation of the sg2 variant plasmid failed at this stage, and so only

variants sg1 and sg3 were carried forward, with further -80ºC stocks made from the new transformants. The sExpress cells were then used to conjugate the knockout vectors into the lysogen CD2301L.

4.2.3 PCR screening for mutation

The conjugations appeared to be effective, with over 100 transformant colonies observed per plate. Eight medium to large colonies were selected for each of the vectors and re-streaked onto BHIS-CC-Tm-Th plates for CRISPR induction. All colonies grew well on the induction plates and stocks were made from a single colony of each for storage at -80ºC and onward processing.

A single colony from each of the eight sgRNA1 transconjugants (Int 1.1-1.8) was tested using a colony PCR to confirm that all were successful knockout clones (Figure 4.6a). It was later realised that, due to human error, the colonies which were tested were not the same ones used for making the -80ºC stocks. The stocks for colonies Int1.1 and Int1.2 were therefore re-checked for the knockout (Figure 4.6b) before being used for plasmid loss.

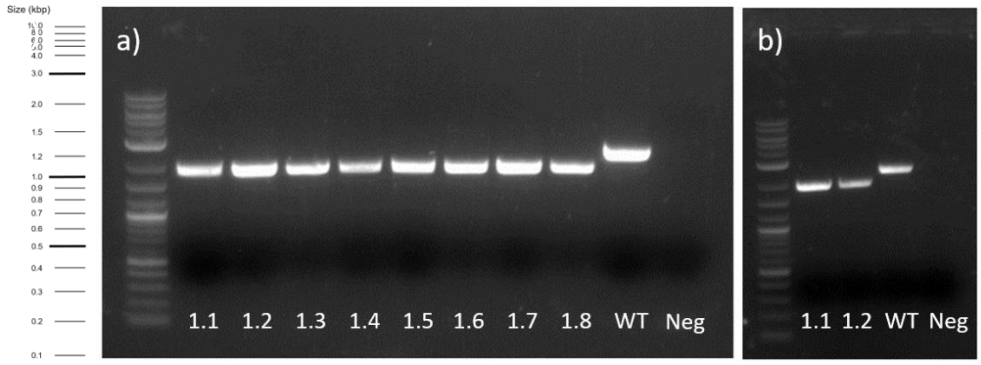


Figure 4.6 - a) Initial colony PCR of eight sgRNA1 integrase transconjugants. WT denotes the wild-type CD2301L gDNA. Neg is the negative control. The PCR shows an approximate 500bp difference in size between the transconjugants and the wild type cells, indicating a successful knockout of the 561bp integrase gene. b) Repeated colony PCR of single colonies from integrase 1.1 and 1.2 to confirm mutation of these colonies prior to plasmid loss.

Liquid cultures of the two samples were sub-cultured over five successive days, after which 100µl of the final culture was spread on a fresh BHIS-CC plate and allowed to incubate overnight. In total, 50 of the resulting colonies were screened through patch plating on BHIS-CC plates and BHIS-CC-Tm plates to determine which colonies had lost the plasmid based on their resistance to thiamphenicol, which was plasmid-borne. The removal of the target gene was confirmed to be as intended through Sanger sequencing, resulting in the selection of colony 1.2.4 for downstream applications.

4.2.4 Induction of mutants for phage release

Overnight cultures of the Δ*int* mutant and CD2301L (positive control) were made in 20ml of BHIS media and induced with Mitomycin C after 24 hours. Following a further 24 hour incubation, 100µl of each culture was spread onto a fresh BHIS-CC plate to allow for post-induction colony stocks to be made, while the remaining culture was centrifuged and filter-sterilised to produce phage preparations for use in plaque assays with the susceptible wild-type CD2301.

The plaque assay of the 1.2.4 colony preparation showed no zones of clearing. This was not an unexpected result, given that we know site-specific recombinases are also responsible for excision of the prophage from the genome. To confirm that the prophage that remained in the genome post-induction was intact and contained no additional changes that could explain the lack of induction (e.g. SNP in an essential gene that may result in a non-viable phage), samples of gDNA from three of the previously saved post-induction colonies (named 1.2.4.1 – 1.2.4.3) were sent for whole genome sequencing alongside a sample of CD2301L for comparison purposes.

4.2.5 Whole genome sequencing of mutant colonies

Short paired-end reads generated from an Illumina MiSeq600 analysis (250X coverage) were assembled using CLC Genomics Workbench v.20. Initial *de novo* assembly resulted in between 60 and 80

disconnected contigs, and it was not possible to manually compile such a large number of contigs into a single genome. Much of the prophage was identified on a single contig, however, it was difficult to locate the remainder of the phage genes amongst the rest of the data. As such, it was determined that further sequencing using long-read technologies would be advantageous in helping to fully assemble the genome sequence. Due to a potential mutation visible in the results obtained from sample 1.2.4.3, this sample was discontinued and only samples of the lysogen and colonies 1.2.4.1 and 1.2.4.2 were submitted for further analysis.

Long-read sequencing of the samples using ONT MinION methods, and subsequent assembly using the Flye and Canu software programmes, assembled the data into a much more manageable eight contigs for each of the three submitted samples. The Illumina short reads were then used to “polish” the assembly using the NextPolish software in order to improve the sequence accuracy. To make comparisons easier, the data from the lysogen was analysed first to facilitate analysis of the mutants.

4.2.5.1 CD2301L

The two largest assembled contigs measured 1.4Mbp and 2.7Mbp in size, and together represent most of the expected genome size for a strain of *C. difficile* (approximately 4.5Mbp). The remaining six contigs ranged in size from 5.2kbp to 124kbp for a total of 219kbp. Comparison of the existing genome for phiCD2301 to these contigs showed that the majority of the phage DNA was contained in a single contig of size 27,362bp. Examination of homologous regions allowed seven of the eight contigs to be assembled into a full genome measuring 4.2Mbp, with the remaining 124,116bp length contig identified by BLAST search as potentially originating from a plasmid (Appendix 7.4). The genome was then annotated using both PATRIC (Davis et al., 2020) and the NCBI PGAP (Li et al., 2021; Tatusova et al., 2016).

Identifying the location of the phage contig within this genome was complicated by the presence of several phage genes also identified elsewhere within the lysogen sequence. For example, the previous version of the phiCD2301 genome obtained from Whittle et al. (2022) contained two holin and two endolysin genes, but the sequences of these genes matched very different locations within the assembled contigs. Additional Sanger sequencing using primers binding in the confirmed areas of the phage genome and in the implicated integration site of the bacterial genome was required in order to confirm the phage contig's placement at the end of contig 1 (Figure 4.7).

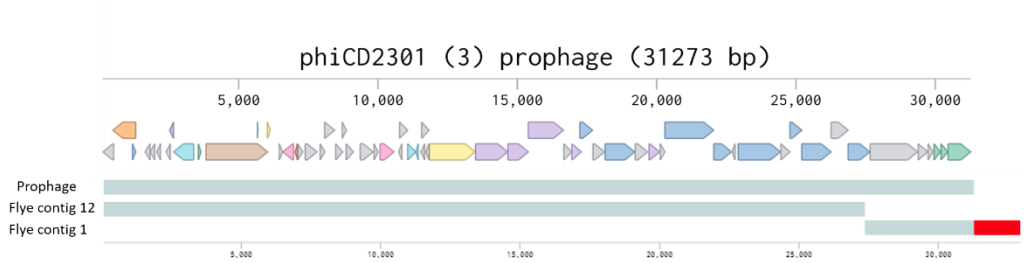


Figure 4.7 - Benchling comparison of the phiCD2301 genome sequence to contigs 1 and 12. The compared section of contig 1 extends beyond the phage genes to demonstrate continuation into the bacterial genome sequence.

In addition to the location of the prophage within the lysogen genome, a further product of this analysis was clarification around the size of the phage. The original genome sequence for phiCD2301 stated the size as 38,695bp. However, the combined long-read/short-read approach adopted here confirmed that this phage is only 31,273bp in size. Further Sanger sequencing using assembled phages as DNA template for the PCRs confirmed that an approx. 7.5kbp section of DNA upstream of the Haemolysin XhlA gene in the previous sequencing (including one of each of the holin and endolysin genes) was not present within the phage itself (Figure 4.8). Given that the genes in this section matched phage genes found in other areas of the lysogen genome, it is possible that the original DNA sample sent for sequencing by Whittle et al. contained some level of bacterial genome contamination which has been amplified alongside the phage DNA and has resulted in a mis-assembly during the analysis.



Figure 4.8 - Section of phiCD2301 represented in previous Illumina sequencing but confirmed as not present through the Sanger sequencing.

Having confirmed the genome sequence of the lysogen, this sequence was then used as a reference to which the sequences from mutant strains Int 1.2.4.1 and Int 1.2.4.2 could be aligned.

4.2.5.2 Integrase mutants

As with the lysogen, the majority of the phage DNA was contained within a single contig, with the remainder located on one of the larger contigs. Following assembly of the contigs in a similar fashion to the lysogen, comparison of the 1.2.4.1 and 1.2.4.2 sequences both showed that the integrase gene had been removed from the mutants with no off-target effects within the phage genome (Figure 4.9). However, while mutant 1.2.4.1 showed no issues with the remainder of the sequence, 1.2.4.2 showed a small deletion in the tail fibre gene and a SNP in the endolysin gene. Work on this colony was therefore discontinued and only 1.2.4.1 taken forward for further work.

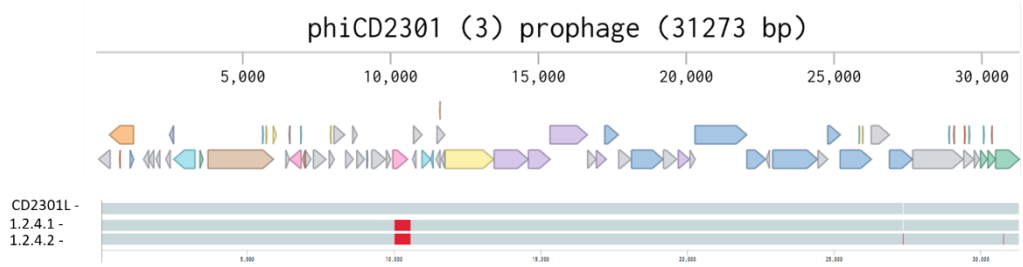


Figure 4.9 – Result of whole genome sequencing of integrase-deficient mutant prophages in comparison to the base lysogen. The results confirmed complete removal of the integrase gene.

4.2.6 Complementation of genes back into the knockout strains

In order to corroborate the hypothesis that the lack of phage induction was a result of the removal of the *int* gene, an additional CRISPR vector was constructed to replace the gene in an attempt to restore wild-type behaviour. This vector used the same homology arms as the original CRISPR knockout vector, however, this time the homology cassette was designed to include a modified copy of the original *int* gene.

To distinguish complemented mutants from potential wild-type contamination, a “watermark” was introduced into the complementation gene, as described in Ng et al. (2019). To accomplish this, a single nucleotide was changed in codon 144 of the gene from C to T. This change maintained the Serine amino acid at that position but removed the AciI restriction site that was present in the wild-type gene. Removal of this site allowed easy identification of wild-type cells, as the localised PCR product would be cleaved twice when digested with AciI in comparison to complemented DNA which would only be cleaved once.

This complementation homology cassette was combined with a sgRNA targeting the BM6 bookmark, and the new cassette was ligated into the RiboCas backbone. The vector was conjugated into the Int 1.2.4.1 integrase-deficient mutant strain and the resulting transconjugants were screened using a combination of PCR to amplify the targeted region followed by restriction digest of the PCR products using AciI. Results of the screening process demonstrated complementation of the integrase gene into the integrase-deficient mutant (Figure 4.10).

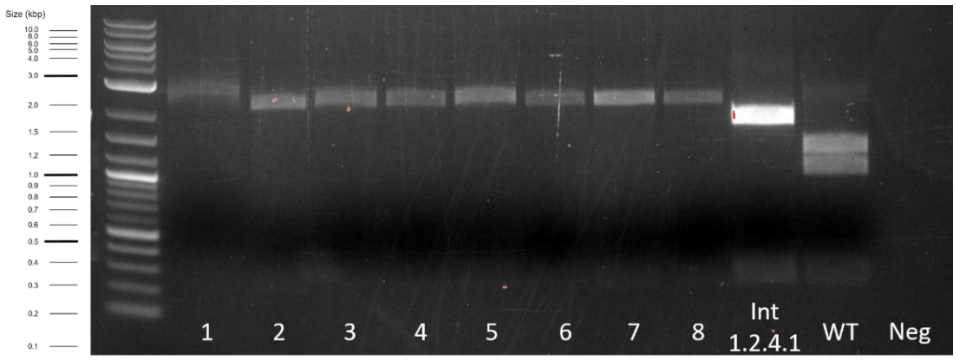


Figure 4.10 - Restriction digest of PCR products amplified from transconjugant colonies. Expected band sizes: correctly complemented mutants - 2.4kbp and 237bp; wild-type phage: 1.4kbp, 1kbp and 237bp. The original Int 1.2.4.1 mutant and wild-type gDNA were run as comparison samples. Note - the sample bands are approximately 500bp larger in size than in the Int 1.2.4.1 mutant, confirming that the gene has been complemented into the transconjugant genomes.

To test whether the excision function had been restored, the complemented strains were grown as overnight cultures in the anaerobic cabinet and then treated with Mitomycin C the next day. Following incubation, the cultures were centrifuged and filter-sterilised and the supernatant used in a plaque assay to test for the presence of phages. The plaque assay confirmed the presence of phages at the expected titre for normal induction for this phage (10^3-10^4 pful/ml), confirming that the removal of the integrase gene was responsible for the loss of induction capability. This is consistent with past work indicating that tyrosine integrases are responsible for both integration into and excision out of the bacterial genome (Cho et al., 2002; Fogg et al., 2014; Groth and Calos, 2004).

While this confirmation was welcome, the issue of harvesting the integrase-deficient mutant remained, as without an integrase to allow excision, the phage is locked within the lysogen and therefore not available for further use. However, previous work by Platt et al. (2003) demonstrated that as long as the requisite phage attachment sites exist within the genome, an exogenously-expressed integrase can function to both integrate and excise DNA. As such, it was determined that one method of extracting the phage could be to deliver a plasmid-borne integrase which could express sufficient enzyme to allow the mutant phage to excise from the genome.

4.2.7 Assessment of plasmid-borne integrase expression

Two plasmids were designed to episomally express the native integrase gene in the hope that the mutated phage could use this protein to excise from the lysogen. In each case, the integrase gene was placed under the control of one of two inducible promotor systems – $P_{fdx\text{-}E}$ from the RiboCas system or the anhydrous tetracycline system from pRPF185 (ATc - Fagan and Fairweather, 2011). Both systems have been shown to be tightly controlled in *C. difficile*, although only the ATc system has currently been tested as an expression vector in this organism.

For the P_{fdx-E} plasmid, two insert fragments were created – 1) the promoter, which was amplified using a forward primer with an NdeI restriction site and a reverse primer with an overhang corresponding to the integrase gene, and 2) the integrase gene which was amplified from CD2301L using a forward primer with a corresponding overhang to the promoter fragment and an NheI restriction site in the reverse primer. These fragments were combined using a SOEing PCR, digested with the appropriate restriction enzymes, and ligated into a digested pMTL83151 plasmid to create pSJK1 (Figure 4.11).

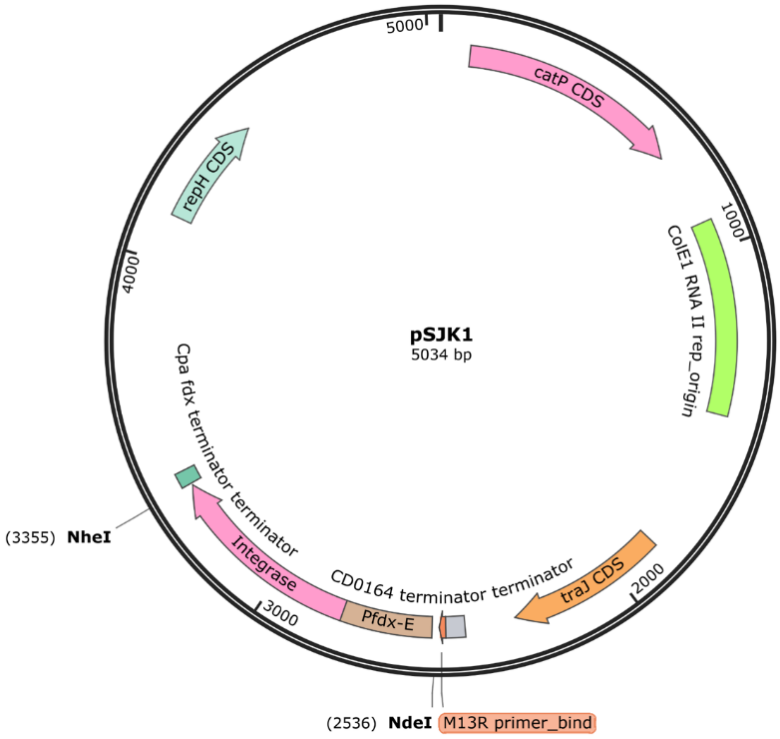


Figure 4.11 - Plasmid map of pSJK1. The native phiCD2301 integrase gene was placed under the control of P_{fdx-E} and inserted into the MCS site of pMTL83151 using restriction enzymes NdeI and NheI.

The same procedure was attempted for creating the ATc plasmid, however, there was some difficulty in amplifying the promotor system from the original pRPF185 plasmid. Given that pRPF185 is an established expression plasmid for use in *C. difficile*, it was decided that it would be preferable to replace the existing *gusA* reporter gene in the

plasmid with the integrase gene rather than transferring the system into a different backbone. This was accomplished through amplification of the integrase gene from CD2301L using a forward primer containing a SacI restriction site and a reverse primer with a BamHI site, the enzymes used by the authors for the original insertion of the *gusA* gene into the plasmid. The plasmid and fragment were then digested using these two enzymes and ligated together as with previous vectors to create pSJK2 (Figure 4.12).

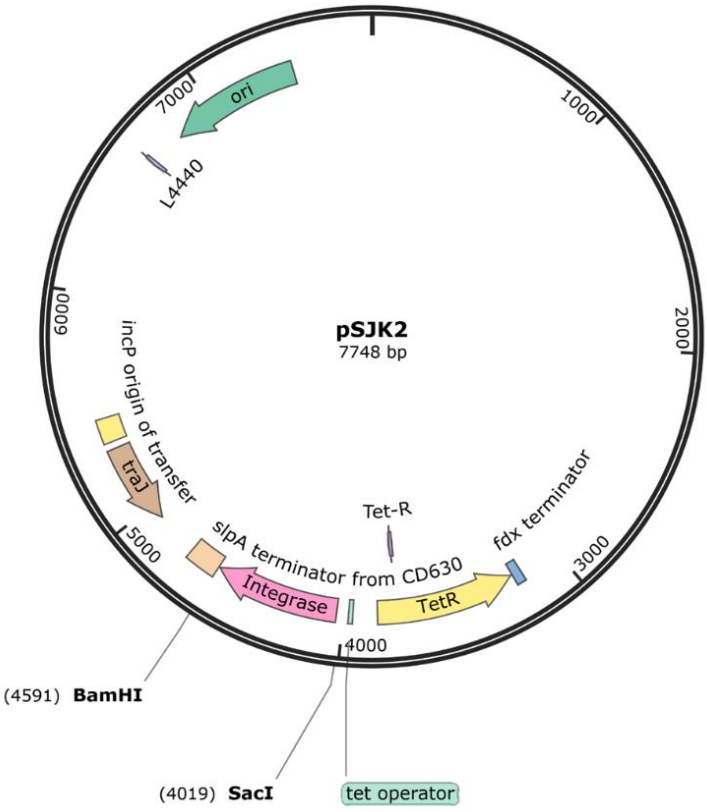


Figure 4.12 - Plasmid map of pSJK2. The native phiCD2301 integrase gene was inserted into pRPF185 in place of the gusA reporter gene using restriction enzymes SacI and BamHI.

Two additional fragments were also created containing a FAST reporter gene (Streett et al., 2019), one of which was flanked by NheI sites and another flanked by BamHI sites. These were digested and ligated into pSJK1 and pSJK2, respectively, to create pSJK1-FAST and pSJK2-FAST (Figure 4.13). Both of these plasmids were conjugated into wild-

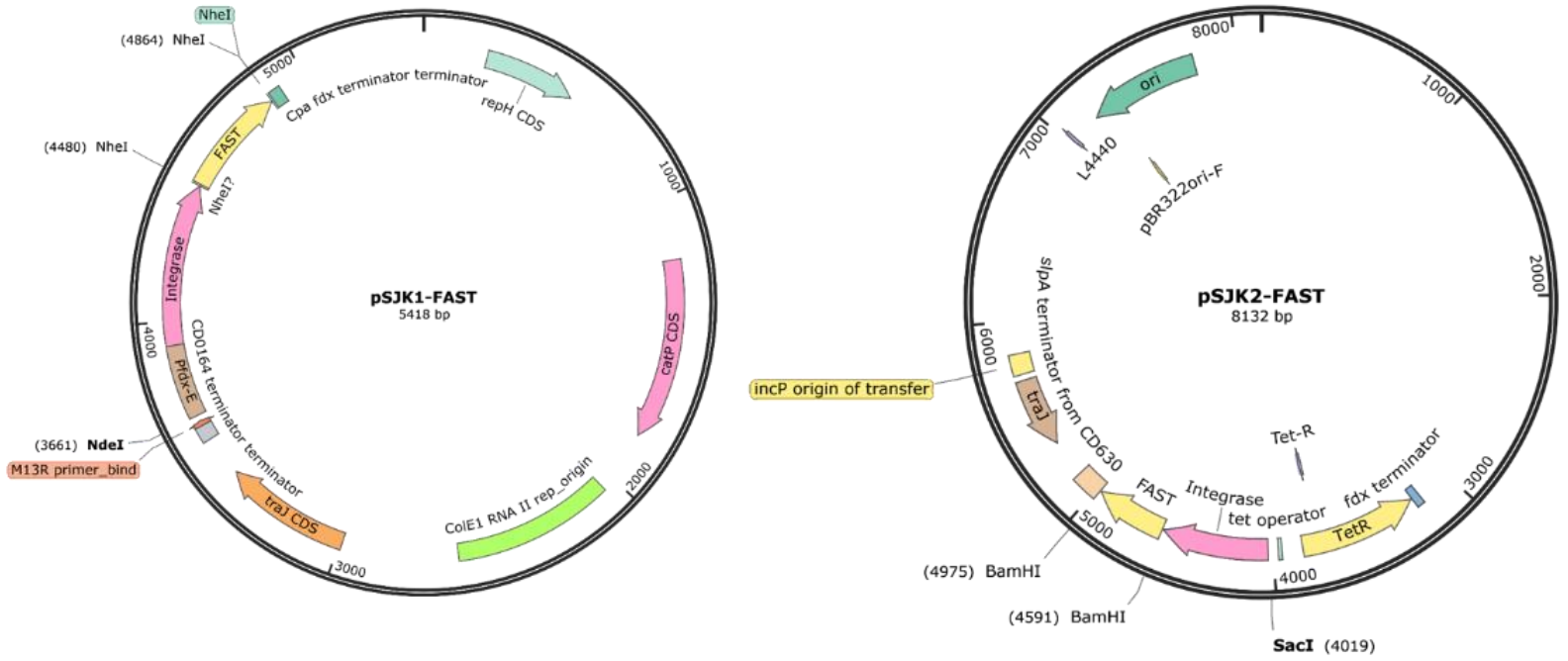


Figure 4.13 - FAST plasmids generated by inserting a FAST gene downstream of the integrase gene using either NheI in pSJK1 or BamHI in pSJK

type CD2301 to allow quantification of the expression levels of the vectors prior to conjugation into the lysogen. An empty version of pMTL83151 and the original pRPF185 plasmid were also conjugated into CD2301 to act as negative controls for the FAST assay, as was a confirmed in-house positive control plasmid (pMTL83151y-Ccarb-v2_FAST) which has the FAST gene under the control of a P_{fdx} promotor.

Following two streaks on BHIS-CC-Tm plates for selection purposes, eight colonies were selected from each set of transconjugants and stored at -80°C for later testing.

Triplicate overnight cultures of all strains were made in BHIS-Tm broth in preparation for the FAST assay but, unfortunately, only the cultures containing pSJK2-FAST grew. Subsequent streaking and colony PCR screening showed that none of the saved transconjugants for pSJK1-FAST had retained the plasmid past this third generation (two confirmation streaks during conjugation and a further streak from the prepared -80°C stocks). In hindsight, this may have been due to an error in not maintaining antibiotic selection when creating the -80°C stocks for these strains. As such, strains containing the pSJK1-FAST plasmid could not be tested at this time.

In the absence of the pSJK1-FAST condition, a preliminary assay was undertaken using the pSJK2-FAST variants to test the FAST system. The cultures of CD2301 containing one of three plasmids (the positive control, the original pRPF185 plasmid, or pSJK2-FAST) were sub-cultured at 1:100. Once the sub-cultures had grown to approximately OD_{600nm} of 0.4, the pSJK2-FAST cultures were treated with 250, 500 or 1000ng/ml of anhydrotetracycline. The cultures were allowed to grow for another 3.5 hours before being treated with TFAmber ligand solution and the fluorescence measured using a Clariostar plate reader (Figure 4.14).

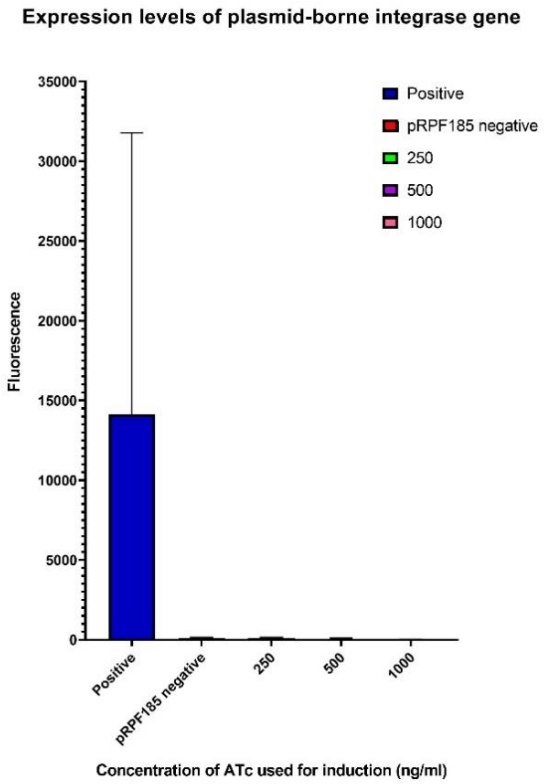


Figure 4.14 - Level of expression of the integrase gene by plasmid pSJK2-FAST as tested using TFAmber labelling solution. Three different induction levels were tested - 250, 500 and 1000ng/ml of ATc. Data represented are the means of results from biological triplicate cultures, with error bars indicating the standard deviations.

Only the positive control gave any level of fluorescence during this assay. To test whether this was due to human error in completing the FAST assay, a cursory test involving conjugation of the non-FAST vector variants into CD2301L and induction with the same concentrations of ATc alongside Mitomycin C was undertaken. No plaques were seen during the subsequent plaque assay, confirming that the issue was likely to lie either with the construction of the vectors themselves or with the expression of integrase potentially being toxic to the cells. It was decided that further control plasmids could be constructed in the form of a plasmid containing only the FAST gene under the control of the promotor system (pATc-FAST and pRibo-FAST). This would demonstrate that the promotor systems were able to express the FAST gene, thereby providing insight into whether the gene itself was the issue.

During the process of designing this new plasmid, it was noted that removal of the *gusA* gene from pRPF185 using SacI may have resulted in the removal of the RBS, as there was a potential SD variation that lay between the start codon of the gene and the upstream restriction site. Further discussion with colleagues experienced in using this system noted that previous in-house plasmids had been designed to contain an additional RBS site added upstream of their target genes, implying that this may have been identified previously as an issue. This could explain the negative results from the pSJK2-FAST plasmid in the previous experiment if no RBS site was available for the polymerase to bind to during induction. To avoid this occurring in the new plasmid, this same extra RBS system was amplified out of the in-house plasmid pMTL-CW21 and fused to the front of the FAST gene (Figure 4.15). New integrase plasmids were also designed to include this extra RBS region.

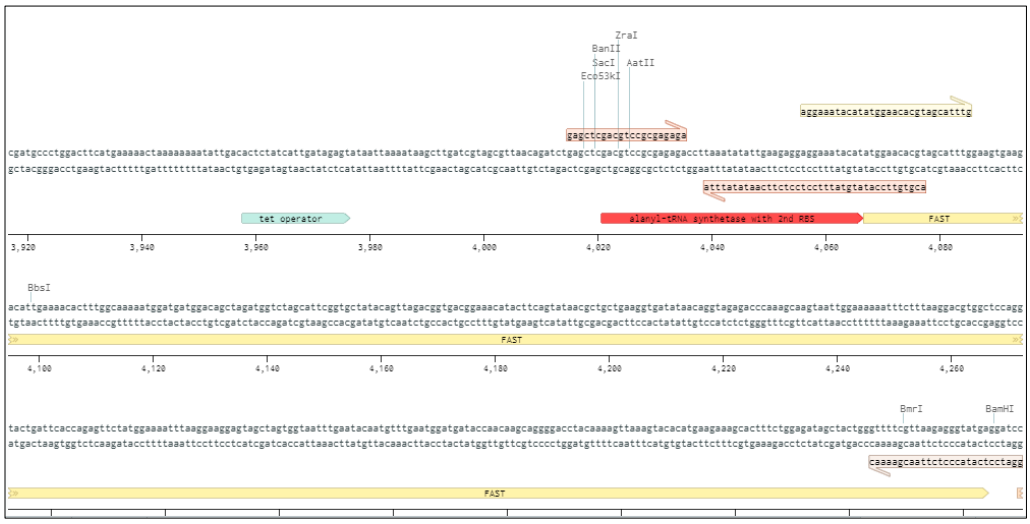


Figure 4.15 - Secondary RBS region fused to the front of the FAST gene to replace a potential RBS region excised when using SacI and BamHI enzymes during the construction of pATcFAST-RL

In addition to the proposed RBS insertion, it was noted during a review of the phiCD2301 genome that the integrase sequence used for the plasmid design was based upon the one identified in the original annotations of the genome (561bp in size) prior to my re-analysis during the whole genome sequencing (504bp). While the sequence remained

mostly the same, the start codon of the updated gene was determined as the first ATG codon downstream of the original proposed GTG codon. While GTG is a common start codon for bacteriophages (Turner et al., 2021a), a multiPhATE2 analysis of the new phiCD2301 genome indicated a consensus between all three gene-calling algorithms for the integrase gene to start with the later ATG codon. Therefore, in order to conform to this updated analysis, new vectors for pSJK1 were also required and so new integrase fragments were created for all vector variants.

Unfortunately, initial attempts to construct these vectors using NEB HiFi assembly failed repeatedly, and I eventually returned to using a restriction/ligation method. Both promotor-FAST variants (pRiboFAST-RL and pAtcFAST-RL) were constructed and transformed into CD2301 for testing but it was not possible to complete the integrase-containing variants before the project deadline was reached. These constructs remain available for testing in future, if required.

4.3 Discussion

The lack of obligatory lytic phages suitable for use as therapeutics is a major roadblock to phage therapy as a treatment for CDI. Even given the promising results of the work undertaken on phiL14WAR (see chapter 3), high specificity means that a diverse library of suitable phages is required in order to readily treat the varying strains of *C. difficile* which are responsible for infections and recurrences. While other aspects of phages have been investigated as alternative solutions, such as the use of endolysins to lyse cells (Mayer et al., 2011; Peng et al., 2019) and the use of phages as a delivery mechanism for a different bactericidal compound (Selle et al., 2020), the evolution of modern molecular biology techniques presents a further avenue to explore – engineering of the existing temperate phages to make them lytic.

One of the main elements of phage genomes to consider for engineering is the integrase gene, which is a characteristic of all temperate phages. This gene is responsible for the integration and excision of the phage from its host's genome. While it does not play a direct role in the repression of lytic genes to enable lysogeny, this integration is thought to facilitate horizontal transfer (also called transduction) of undesirable genes such as those related to antibiotic resistance (Jebri et al., 2021). Therefore, while its removal would not be expected to render a phage strictly lytic, it would be necessary in order to assuage medical concerns regarding the transfer of harmful genes during therapeutic use.

Very little research is available in the current literature quantifying the effect of targeted removal of the integrase gene, with most dating back to the initial studies of phage λ in the late 20th century. Gingery and Echols (1967), for example, demonstrated that the integrase-deficient mutant λb2 developed by Kellenberger et al. (1961) could not form a stable lysogen but did still show some abortive lysogeny due to the continued presence of the repression machinery. This was supported by a 1974 study by Gottesman, who showed that integrase-deficient λ mutants could only temporarily convey super-immunity on a culture because the inability to integrate into the bacterial genome prevented passive replication of the prophage, thereby resulting in loss of the phage from the culture as immune cells died off and were replaced with non-immune cells. However, these studies originate from a time prior to the availability of sequencing technologies, and so it is not possible to definitively confirm that the only difference in the phages was the loss of the integrase gene. More recent studies, such as those described by (Selle et al., 2020) and Pope et al. (2011), focus on the effects of the removal of both the integrase gene and repressor gene from the phage simultaneously. It was therefore decided that characterisation of a confirmed integrase-deficient phage would be preferable in the first instance to confirm the conclusions of the earlier studies that removing the integrase gene alone is insufficient to render the phage lytic.

In this study, I attempted to remove the integrase gene from the known temperate phage phiCD2301 using CRISPR-Cas9 methods. Vectors were constructed which targeted the integrase gene, replacing it with a “bookmark” region within the CD2301L lysogen. The resulting mutant was then induced using Mitomycin C in an attempt to remove the phage for testing. As expected, the phage could not be excised from the genome in the absence of the integrase gene.

To confirm that the lack of an integrase was responsible for this effect, the mutant cells were sent for whole genome sequencing to check for off-target effects. Of the three clones sent, one was confirmed as having the correct sequence. Complementation vectors were then used to target the bookmark region and replace it with a modified version of the integrase gene from which the internal AciI restriction site had been removed. Following conjugation and plasmid loss, digestion of the targeted PCR fragment using AciI confirmed the successful complementation of the gene, and these cells were induced with Mitomycin C in the normal way. A plaque assay of the supernatant confirmed the presence of phages and therefore the reversion to wild-type behaviour.

The relatively easy removal and complementation of the integrase gene from the lysogen was promising, however the question as to how to excise the mutant phage in the absence of the integrase gene remained. Multiple studies have shown that integrase variants expressed on a plasmid are able to enact integration and excision of DNA from a genome as long as the relevant attachment sites were present (Groth et al., 2000; Makhija et al., 2018; Platt et al., 2003; Saviola, 2004). It was therefore hypothesised that this could work to integrate or excise an integrase-deficient phage mutant as well, as long as no other changes were present in the genome that might affect any required integrated host factors.

Expression vectors were designed with the endogenous *int* gene under the control of one of two potential inducible promotor systems proved to be effective in *C. difficile* – $P_{fdx\text{-}E}$ (Cañadas et al., 2019) and the ATc

system (Fagan and Fairweather, 2011). Each vector also had a variant constructed in the same manner but with a FAST reporter gene inserted immediately after the *int* in order to test the expression levels of each vector. The constructed vectors were then conjugated into wild-type CD2301 for testing prior to use in the lysogen itself alongside appropriate controls.

When cultures of the transconjugants were made in order to undertake the FAST assay to check expression levels, the $P_{fdx\text{-}E}$ condition cultures failed to grow. Later investigation determined that antibiotic selection had not been maintained sufficiently following the initial conjugation, and the plasmid had been lost at some point prior to the -80°C stocks being made. This variant was unable to be tested at this time. However, the ATc cultures grew appropriately, and so an initial FAST assay was conducted using these cultures as practice and to ensure that these vectors had been constructed correctly. Unfortunately, this assay confirmed that while the positive control was active, the ATc constructs themselves did not show any FAST expression.

The failure of the FAST assay led to the identification of several issues with the initial plasmid design, including a missing RBS in the ATc variants and a difference in the composition of the integrase gene between the initial plasmid design and the reanalysis of the sequencing results. The construction of new vectors to fix these issues was then hampered by failed amplifications, ligations and a change of method for construction. It therefore took several months for correct vectors to be conjugated into CD2301 for testing. Unfortunately, due to the time taken and the prioritisation of other work within this project, a new round of FAST assays was unable to be completed. These vectors remain available for testing in future.

While it has not been possible to fully investigate the effects of removing the integrase gene from a temperate phage, the work undertaken in this chapter is a stepping stone towards that goal. The results obtained in the

previous chapter regarding phiL14WAR allow us to infer what we might expect of an engineered phage from which the integrase gene had been removed (with the caveat that the presence or absence of other lysogeny mechanisms in phiL14WAR remains unconfirmed), but being able to test this in a much more controlled manner would be preferable. As such, completing the FAST assay and inducing the mutant phage using the integrase expression vectors would be a high priority in future, should time and funding become available. If successful, the phage could then be tested in a similar manner to phiL14WAR, including activity in a live culture and the ability to form lysogens on plate.

However, the lack of induction by the integrase-deficient phage provides an additional, previously unconsidered benefit. If the phage is unable to induce without external input, then it would allow us to make any changes we need to the phage genome without losing the phage or causing issues with harvesting mutants. The engineered phage would remain locked into the host genome until such a time as we are satisfied that all work has been completed safely, at which point the requisite plasmid could be conjugated into the cells to release the phage. In addition, barring unforeseen recombination events, this method would also improve screening time by preventing unwanted spontaneous wild-type induction which could complicate the identification and harvesting of mutants. This would be particularly important in relation to the removal of the repression systems, as without those we would expect the phage to quickly induce, killing its host and making identification of mutants difficult due to a mixed population (engineered phages alongside spontaneously-induced wild-type virions).

If the integrase expression vectors work successfully in the future, or an alternative induction method can be found, then locking the phage into the genome by removal of the integrase gene gives us a much easier subject to work with. In addition to engineering phages for phage therapy, we could also study the fundamental biology of the phage genome without losing time screening for mutants amongst a mixed population.

As an example, TraDIS (Barquist et al., 2016; Langridge et al., 2009) could be implemented using an integrase-deficient mutant lysogen. Attempting to do this using the wild-type lysogen risks loss of data when phage induce spontaneously, killing their hosts. An integrase-deficient lysogen would be unable to induce, thereby preserving all potential transposon mutants to create an accurate library. The expression vectors could then be used to release the phage, allowing us to identify which genes within the phage are essential to virion formation and function.

The work undertaken in this chapter, although incomplete, still holds promise with respect to engineering phages to be suitable for therapeutic use. Removal of the integrase gene is key if we are to minimise the risk of horizontal gene transfer, and while accomplishing this goal in a usable fashion is more complicated than it first appeared, success would provide added benefits that would facilitate other important tasks, such as the removal of the repression systems governing lysogeny. Should time and funding become available, completing this work would, in my opinion, be a high priority.

4.4 Chapter summary

Removal of the integrase gene from a temperate phage is a key requirement for engineering a phage that is suitable for therapeutic use against CDI. This chapter details the first steps taken to remove the integrase gene from phiCD2301 using CRISPR-Cas9 gene knockout technologies, including confirmation of function through complementation of the gene and the issues around recovery of the mutant phage in the absence of an endogenous integrase gene. The hypothesis that a phage could use a plasmid-borne integrase gene to facilitate excision was unable to be fully tested due to complications with plasmid design and construction, however, this remains a promising line of enquiry for obtaining the integrase-deficient mutant for testing. While removal of the integrase would not be expected to render the phage completely lytic in nature, it would remove a main avenue of horizontal gene transfer which

is a primary concern when using temperate phages therapeutically. If successful, further work would include testing the ability of the phage to form lysogens in the absence of the integrase gene, the impact of the mutant phage on a live culture of *C. difficile* in comparison to the wild-type phage, and the effects of further gene knockouts, such as the repressor gene.

Chapter 5

CRISPR knockout of the repressor gene from phage phiCD2301

5 CRISPR knockout of the repressor gene from phage phiCD2301

5.1 Introduction

The major stumbling block preventing the use of temperate phages for phage therapy is the ability to suppress their lytic genes, allowing them to persist within the host genome indefinitely and provide super-immunity to the cell. Super-immunity creates a resistant sub-population of the bacteria which cannot be treated using the same or similar phages, and it is therefore not possible to guarantee full infection clearance when using temperate varieties. Identification and removal of the mechanisms controlling lytic gene repression is therefore a key goal in order to adapt temperate phages for therapeutic use.

Investigations into the genetic controls underpinning the lysis-lysogeny decision have been ongoing since the middle of the 20th century, and several repression mechanisms have been posited. The best characterised of these is the *cI*/*cro* bi-stable switch identified in *E. coli* phage λ, as described in section 1.2.3.4 (Figure 5.1). This system relies on the expression of the repressor gene *cI* and its self-regulation at levels above the threshold required to maintain repression of the lytic genes. If levels of the repressor drop below that threshold, the anti-repressor gene *cro* is activated, competing for the same operator regions used by the repressor and depressing its production. Eventually, levels of the repressor drop below the level needed to maintain the lytic gene repression and the phage re-enters the lytic life cycle (Oppenheim et al., 2005).

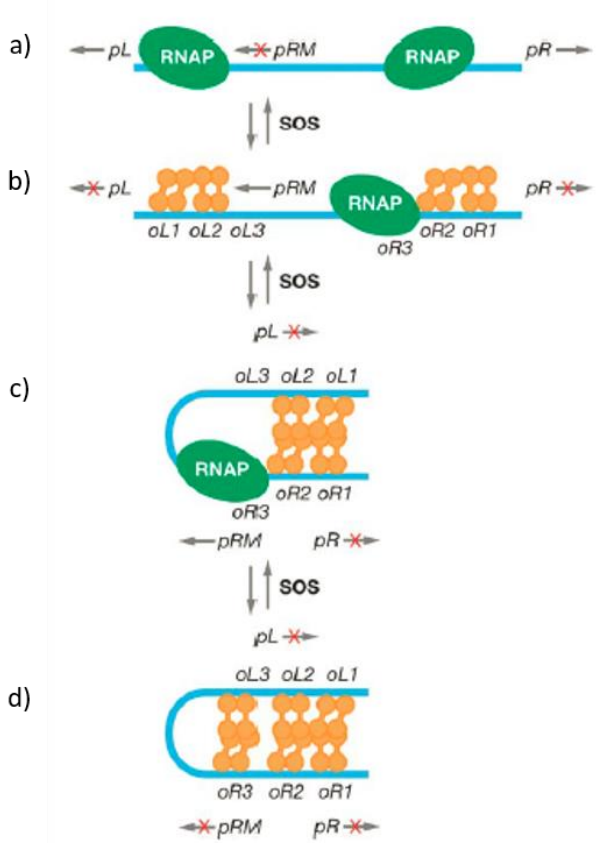


Figure 5.1- Action of the CI protein in repression of the early lytic gene promotors. Taken from Oppenheim et al (2005). RNAP is the RNA polymerase. Orange dumbbells represent CI protein. a) Operative P_R and P_L promotors when no CI is present, with the P_{RM} promotor inactive and the early lytic genes being expressed. b) Initial binding of the CI protein to oR/L 1 and 2, silencing the lytic promotors and activating P_{RM}. c) Hybridisation of the CI proteins between operator regions to form tetramers strengthening lytic gene repression. d) Deactivation of the P_{RM} promotor as a result of CI proteins binding to the third operator region, initiating the self-regulatory mechanism of CI production to maintain a stable state.

Examination of the full genetic control of this system shows that *cI* transcription is initiated when levels of an early phase gene, *cII*, are high enough to activate the P_{RE} promotor (see Figure 1.6) (Hao et al., 2021; Oppenheim et al., 2005), and so removal of *cII* was initially considered to prevent the lysogeny system from being activated. However, Gingery and Echols (1967) quote a study by Kaiser (1957) which showed that

removal of *cII* did not entirely prevent lysogeny from occurring, whereas removal of the *cI* gene did (although it should be noted that these experiments were based solely on the appearance of plaques in comparison to the wild-type λ phage rather than on dedicated lysogen-formation tests).

More recent studies have also shown that the removal or truncation of homologs of the *cI* gene has successfully created lytic versions of their parent temperate phage. Phage TM4 of *Mycobacterium avium*, for example, was initially isolated from a lysogenic strain but appears to have lost two sections of the genome during the induction process, one containing the integrase and the other a potential repressor gene. Studies by Pope et al. (2011) have found that TM4 is now unable to form lysogens and it has been classified as a lytic phage despite its temperate origins. Another example is the *Pseudomonas aeruginosa* phage DMS3, from which Cady et al. (2012) deleted the first 76 amino acids of the *cI*-equivalent gene to create the lytic variant DMS3 *vir*. These examples support the hypothesis that any gene which shows homology to the phage λ *cI* gene should be a primary target for consideration when engineering temperate phages to make them lytic.

While other lysogeny mechanisms have been identified in phages, such as protease-encoding and quorum-sensing mechanisms (Brady et al., 2021a, 2021b), the phage selected for engineering in this project (phiCD2301) contains a gene which has previously been identified through bioinformatics as homologous to *cI*. As such, this gene was used as an initial target for knockout using CRISPR techniques, in the hope that its removal would render this phage lytic.

5.2 Results

5.2.1 Vector creation

Vectors for this knockout were created using primers targeting the proposed *cI*-like gene to create the three vectors

pRECas1_phiCD2301c1_sg1, pRECas1_phiCD2301c1_sg2 and pRECas1_phiCD2301c1_sg3 (Figure 5.2).

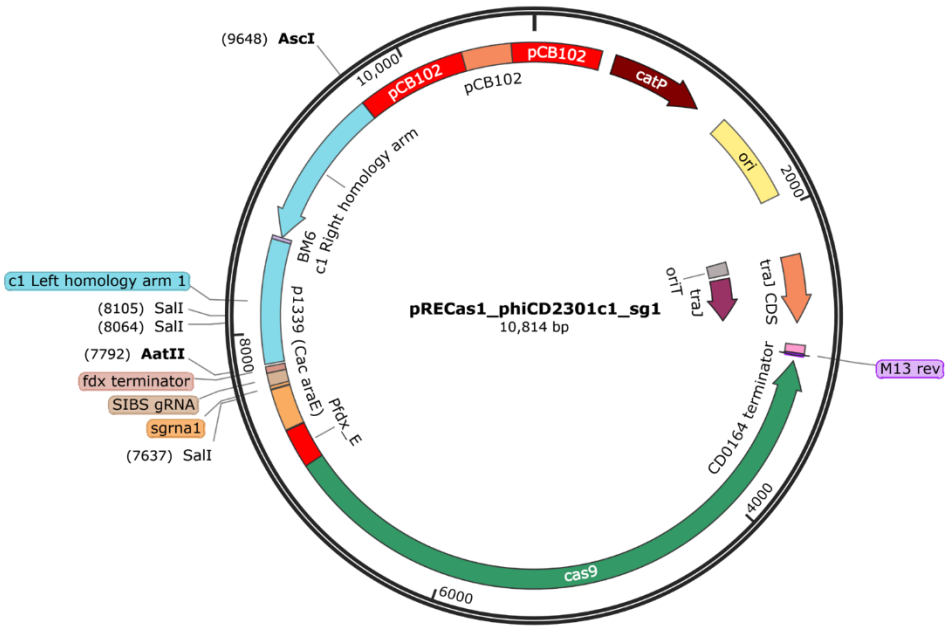


Figure 5.2 - Plasmid map of vector pRECas1_phiCD2301c1_sg1. All elements of this plasmid are the same as for pRECas1_phiCD2301int_sg1 (see Figure 4.4) except that the homology cassette and guide RNA were specific to the cI gene.

However, unlike the integrase knockout, obtaining correctly assembled vectors for this gene was difficult. Transformation of the ligated vectors into XL1 Blue cells gave limited positive results during colony PCR (Figure 5.3), with only one colony from the pRECas1_phiCD2301c1_sg2 transformation showing a correctly ligated vector after Sanger sequencing (colony 8). This vector was carried forward for transformation into chemically competent sExpress cells and was conjugated into CD2301L, with bead stocks of the requisite *E. coli* strains made and stored at -80ºC.

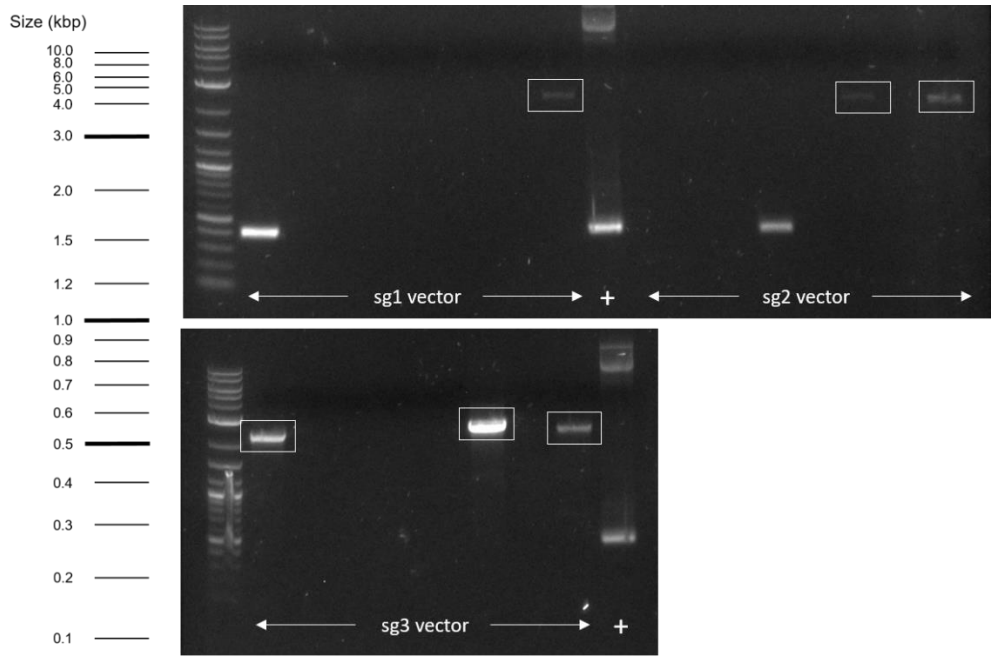


Figure 5.3 - Colony PCR results following the transformation of the cI knockout vectors into XL1 Blue chemically competent E. coli cells. Only six positive results were obtained from 24 colonies tested, indicated with boxes.

Following conjugation, eight transconjugant colonies were selected and re-streaked onto BHIS-CC-Tm-Th plates for induction, as per standard protocol. However, only two of these colonies (2.2 and 2.5) grew on the initial conjugation attempt, with a further two growing on a second attempt (colonies 2.9 and 2.12). The lack of successful transconjugants indicated that many of the tested colonies were either *E. coli* or wild-type CD2301L contaminants that did not survive the second application of thiamphenicol and theophylline, and that successful mutants may have been lost due to knockout of the *cI*-like gene resulting in prophage induction and host cell death. This raised the question as to whether the standard conjugation plate protocol could be used if there was a risk that successful induction would result in no colony for onward testing and confirmation.

5.2.2 Trouble-shooting the isolation of cI *mutants*

If the identified *cI*-homolog is the repressor responsible for silencing the early lytic genes in phiCD2301 then it is possible that successful knockout of this gene would result in the lytic genes no longer being repressed, the

phage excising from the bacterial genome and the host being killed through the phage lytic activity. Harvesting mutants for testing, therefore, had the potential to be more complex than it was for the *int* mutant, as successfully mutated transconjugants would not survive induction. To account for this, three methods of harvesting mutants were trialled.

5.2.2.1 Identification through colony PCR screening of transconjugant colonies

To demonstrate whether the standard plate procedure for conjugation could still work, the four transconjugants obtained from the first two conjugation attempts were screened using colony PCR, as per the standard protocol (Figure 5.4).

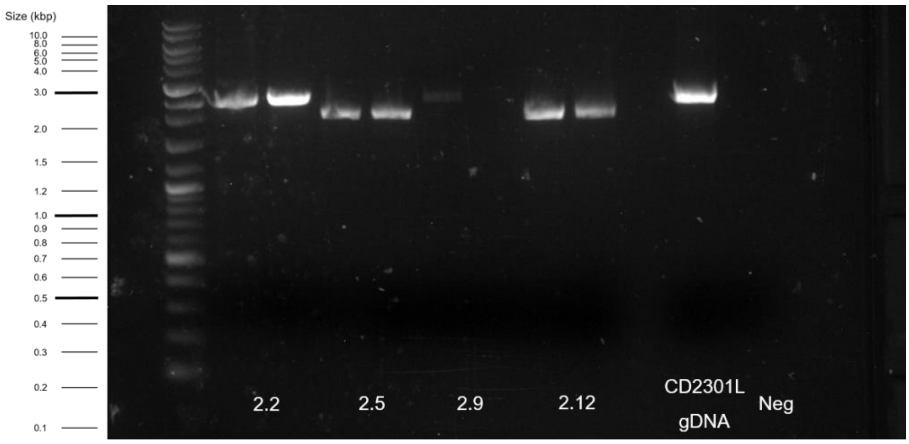


Figure 5.4 – Screening PCR of four transconjugants from the first two conjugations using CRISPR plasmid pRECas1_phiCD2301cI_sg2.

Of the tested colonies, only 2.5 and 2.12 were confirmed as having had any change in the sequence when compared to the wild-type lysogen. However, subsequent Sanger sequencing of PCR-amplified DNA from these two colonies showed that while the gene did appear to have been removed, a large section of the right homology arm in both had also been deleted (Figure 5.5). Given its size, it is likely that this deletion resulted in downstream effects that prevented the prophage from being induced, and so work on these colonies was discontinued.



Figure 5.5 – Benchling comparison of sequencing data from phiCD2301c1sg2 colony 2.5 (bottom) and the expected sequence for a successful cI knockout (top).

A third conjugation yielded eight further transconjugants which successfully grew on the induction plate (2.17-2.24). Stocks were made of all colonies from both the BHIS-CC-Tm-Th plate (for screening) and the BHIS-CC-Tm plate (for later induction using a different method) and the stocks were stored at -80ºC. However, screening of these colonies using the standard method showed no result due to issues with screening primers. Given the premise of these tests (i.e. that visible transconjugants were unlikely to contain correctly mutated phages), this test was discontinued at this point.

5.2.2.2 Identification through plate flooding

Following on from the hypothesis that successful mutations should result in host cell death, a second harvesting method was devised in which transconjugants were induced on a BHIS-CC-Tm-Th plate (as per standard protocol) but the plate was subsequently flooded with fresh BHIS to collect any free phage released by cell lysis from the plate surface.

As the previous test had not confirmed the mutation status of colonies 2.9 or 2.17 – 2.24, these samples were streaked onto a BHIS-CC-Tm-Th plate from previous un-induced stocks and incubated for 24 hours, after which the plate was flooded with 5ml of BHIS. A sterile plastic spreader was used to scrape the surface of the agar and the slurry was transferred to a 50ml falcon tube. Following centrifugation and filter-sterilisation to

remove bacterial cells, the supernatant was used in a plaque assay and any resultant plaques were picked for PCR screening.

The plaque assay results differed greatly between the samples with colony 2.9 producing more than 200 plaques when used neat, and 23 plaques when diluted to 1/10, while colonies 2.18 and 2.22 from the third conjugation showed only single plaques. The other colonies produced none at all. On later review, an element of this difference may be due to the stock of colony 2.9 having been made from multiple streaked colonies and thereby containing a greater diversity of bacteria, where 2.17 – 2.24 stocks were made from single colony samples pre-induction. However, at the time, the discrepancy was noted but the test was continued, as the aim was to obtain a successfully mutated phage.

Plaques were picked (three from 2.9 and the single plaques from 2.18 and 2.22), purified and screened for mutation using PCR (Figure 5.6). Of these, a PCR product was obtained from two of the 2.9 plaques (1 and 2), however, the amplicon generated was that of a wild-type.

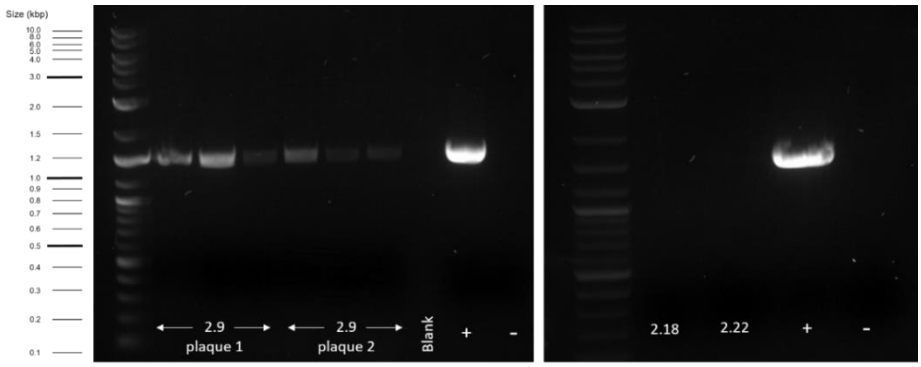


Figure 5.6 - Plaque screen PCR to determine if the plate flood method had collected phages from the plate and whether they were correct mutants. + denotes the CD2301L gDNA positive control and – denotes the dH_20 negative control.

At the time, it was the intention to repeat the inconclusive results obtained from colonies 2.19 (3), 2.18 and 2.22. However, while this method proved successful in harvesting phages from the plate surface, the additional

manual handling steps of flooding the plate and transferring the slurry were sometimes awkward, risking spillages and sample loss. It was therefore determined that, in light of the results from the 2.9 plaques, this method was not optimal for harvesting mutant phages and therefore the repeats of 2.9 (3), 2.18 and 2.22 were not undertaken.

5.2.2.3 Isolation through induction of transconjugant cultures

The third method of harvesting mutants involved the induction of transconjugants in culture, as opposed to on a plate. It was expected that the addition of theophylline to a culture of a confirmed transconjugant would release the mutant phages into the surrounding media, which could then be centrifuged, filter-sterilised and PCR-screened for the presence of mutant phages.

For this, two overnight cultures were made of each of colonies 2.17 and 2.18 (in the first instance), as well as a culture of CD2301L. These cultures were made in 20ml BHIS with 20µl of 15mg/ml thiamphenicol added to select for correct transconjugants. After a 24-hour anaerobic incubation, one culture from each colony was treated with 400µl of 48mg/ml theophylline (5mM final concentration) for CRISPR-Cas induction while the remaining culture was left untreated to act as a negative control. The CD2301L culture was treated with the same concentration of theophylline to control for the effect of theophylline on spontaneous (i.e. non-CRISPR induced) phage production. Following 24-hour incubation, each of these cultures were centrifuged and filter-sterilised to create a phage preparation.

Plaque assays were undertaken to sample the phages in the culture and as a preliminary gauge as to whether this method of induction had been successful. If successful, the expectation was that the induced samples would contain a much higher phage titre than the un-induced sample. However, contrary to expectation, the assay using colony 2.17 only identified 15 phages in the neat induced sample compared to 138 phages on the 10-fold dilution of the un-induced culture (with too many plaques

to count on the neat plate). The 2.18 sample identified even fewer plaques across both induced and un-induced cultures, but the two conditions were relatively equal in number (13 plaques on the neat sample of the former, and 9 plaques on the latter). Given the previous results obtained using this vector, it was deemed that any negative result was more likely to originate from an issue with the vector rather than as a failing of the method itself.

As with the previous identification method, plaques were selected for screening using PCR to determine whether the knockout had been successful. A total of 16 plaques were screened from colony 2.17 and 15 plaques from colony 2.18, all of which were confirmed as wild-type phages (Figure 5.7).

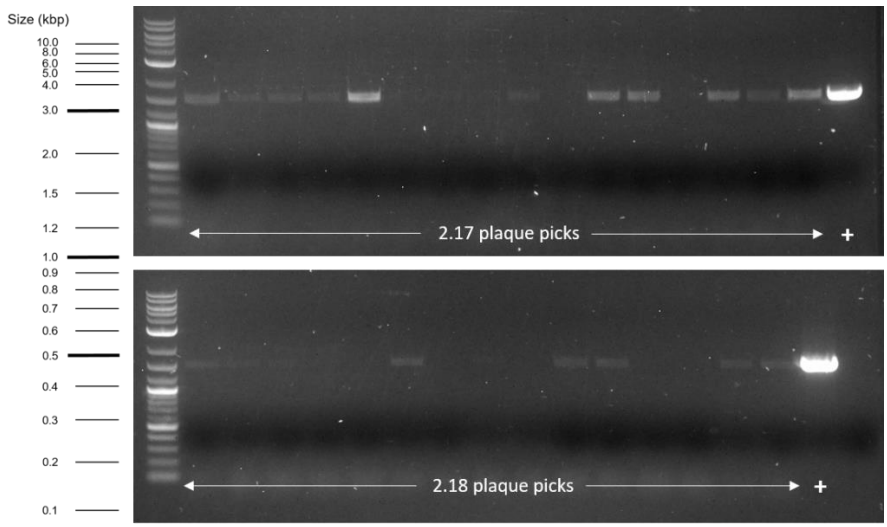


Figure 5.7 - Gel images of phage plaque picks from the liquid induction of colonies 2.17 and 2.18. + denotes CD2301L gDNA used as a positive control.

Colonies 2.19 – 2.24 remained to be tested, and so cultures of colony 2.19 were subjected to the same method as a final test of the efficacy of this knockout vector. The plaque assay results for this colony were slightly more promising, producing 40 plaques from the induced neat sample and 19 on the un-induced neat sample. However, the PCR

screening confirmed that all 25 plaques tested from the induced sample were wild-type phages.

Given the repeated negative results for colonies using pRECas_phiCD2301_sg2, it was decided to discontinue work with this vector and make a second attempt to construct vectors using single guide RNAs 1 and 3, which had previously failed to transform correctly. The second attempt at transforming the pRECas_phiCD2301c1_sg1 vector was successful, although the sg3 vector continued to fail. The sg1 vector was conjugated into CD2301L as previously described, and the culture method of induction was used for testing. Again, all screened plaques originated from wild-type phages.

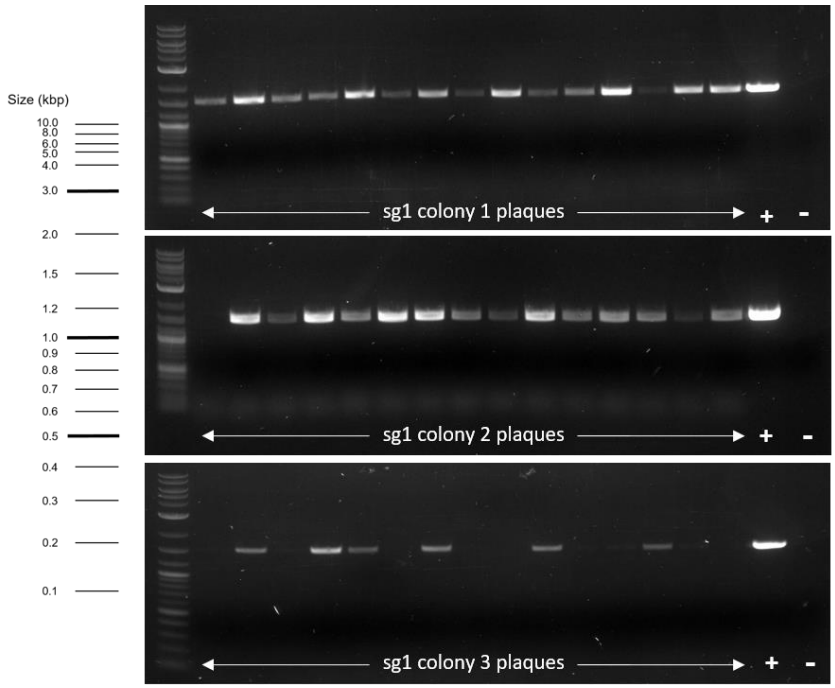


Figure 5.8 - Screening PCR of plaques generated from three induced transconjugant colonies obtained using the new sgRNA1 vector. Fifteen plaques were picked from each of the two colony samples for screening. + and – denote the positive and negative controls, respectively.

While knockouts of this gene continued to be elusive, even using a second guide RNA, in my opinion, this method of induction was the best method of harvesting potential repressor mutants, as phages could be

easily collected for testing with minimal manual handling and without risk of sample loss. One planned improvement which was not implemented at this stage was to centrifuge the culture before the addition of theophylline and then wash the cell pellet with PBS. This would remove any spontaneously induced wild-type phages that may be present in the culture after overnight growth. The cell pellet could then be re-suspended in fresh pre-reduced BHIS media. Adding these steps would minimise the number of extraneous phages in the sample, thereby reducing screening time.

Given the integral nature of the repressor gene to maintaining lysogeny, it was inevitable that identifying successfully mutated phages would be a complicated task. However, while the lack of positive mutants observed so far could be a result of the number of escapees outweighing mutants in the tested samples, it could also be a result of targeting the wrong gene, with this phage adopting a completely different repressor system from phage λ. As such, it was decided that further analysis was required to confirm that the *cI*-homolog was the correct target for this work and, if not, what potential alternative targets existed.

5.2.3 Identification of the lysogeny control mechanism through RNA sequencing

The gene that has been targeted so far in this work was identified through bioinformatics by a colleague as a potential *cI*-homolog (Whittle, 2018). However, no experimental data is available to support this conclusion. Previous work on *Salmonella* prophages by Owen et. al (2020) demonstrated through RNA sequencing that only a limited number of prophage genes are active while the phage is stably integrated, primarily those related to lysogeny maintenance, immunity and improving bacterial fitness. This included the *cI* gene in phage BTP1, which should be intermittently transcribed to maintain the lytic gene repression. It was therefore decided that RNA sequencing of the prophage phiCD2301 in a stable culture of CD2301L could be used to confirm whether the

annotated *cI* gene is the correct target or provide alternative potential repressor targets if needed.

The genome of CD2301L detailed in section 4.2.5.1 was used as a reference during this analysis.

5.2.3.1 RNA sample selection

Given the necessity of obtaining a high concentration of RNA for sequencing, a preliminary test was conducted using three cultures of CD2301L (biological replicates). After overnight incubation, 2ml samples of the culture were taken and extracted before being analysed using the Agilent BioAnalyzer (Figure 5.9).

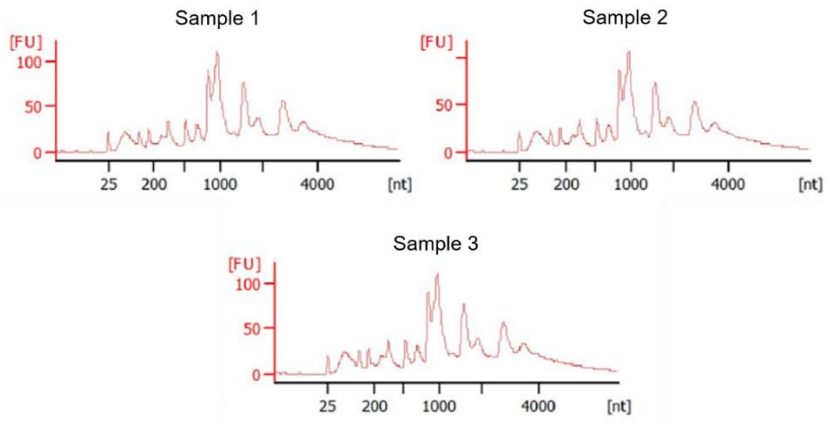


Figure 5.9 - Quality assessment of RNA samples taken from three overnight cultures of CD2301L. Electropherograms show abundance and fragmentation patterns of RNA extracted from each sample.

The RNA taken from these samples was found to be too fragmented to be suitable for sequencing, which was determined to be due to the fragility of RNA over time. As such, a growth curve was conducted using a single 1% inoculum sub-culture of CD2301L to determine the best point at which to obtain RNA of sufficient quality and abundance for sequencing analysis. OD_{600nm} readings were taken every hour for seven hours after inoculation, with 2ml RNA samples taken at each time point between hours three and seven (Figure 5.10).

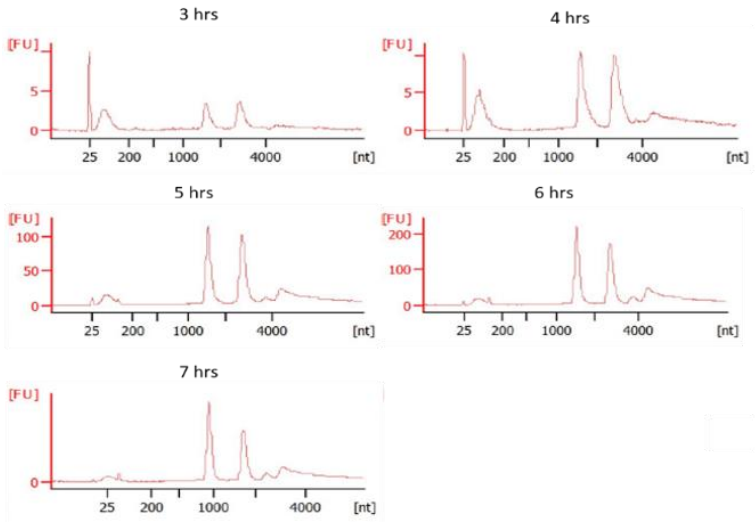


Figure 5.10 – Agilent BioAnalyzer results showing RNA quality and abundance over time in samples taken from a sub-culture of CD2301L. RNA samples were extracted between hours 3 and 7.

RNA quality was found to be much better in the sub-culture than in the overnight culture, showing only the two expected 16S and 23S RNA peaks as opposed to the fragmented pattern previously seen (although a RIN value could not be calculated initially due to unexpected baseline readings). As the quantity of RNA required was 1-2µg total, it was determined that samples taken after six hours would provide the required quality and abundance to satisfy the RNA sequencing requirements.

Three biological replicate cultures of CD2301L were created to repeat the RNA sampling, this time with 2ml samples taken from each culture after 6 hours of growth. These samples were taken in technical triplicate to ensure that samples would be available for additional testing if required. The OD_{600nm} reading was also taken at this time to ensure that culture growth was at an equivalent level to that of the test culture. One sample from each culture was then extracted and assessed using the Agilent BioAnalyzer before submission for sequencing. However, the first attempt at clearing the sample of gDNA through DNase treatment was not as successful as originally expected. An additional DNase treatment and RNA clean-up step was required to ensure full gDNA removal. Following re-treatment, assessment using the BioAnalyzer showed that,

while some RNA had been lost, the sample remained of sufficient quality and abundance to be suitable for RNA sequencing (Figure 5.11). The samples were then submitted to the DeepSeq facility for analysis.

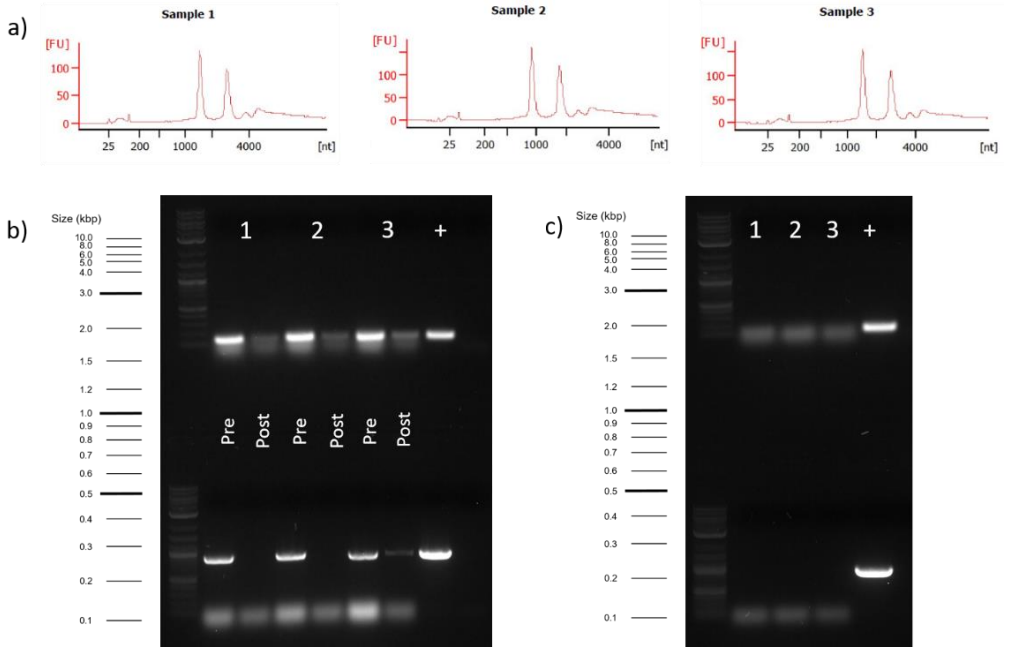


Figure 5.11 - a) RNA quality assessment of samples taken from three biological replicate cultures of CD2301L six hours after 1:100 sub-culture; b) Samples were compared pre- and post-DNase treatment using PCR primers to generate a 200bp fragment (top) and a 1kbp fragment (bottom) to prove full gDNA removal; c) PCR of the the post-treatment sample following a second DNase treatment.

5.2.3.2 RNA sequencing analysis

The raw .fastq files received from DeepSeq were initially assembled using the RNA Seq Analysis function of CLC Genomics Workbench v.21 while awaiting the official analysis from DeepSeq's bioinformatics team. This workflow mapped the Illumina sequencing reads to the CD2301L reference genome and generated counts of transcripts per million (TPM) for each of the annotated genes in the reference. A subset of the data relating only to the prophage was extracted and analysed to determine which of the genes were expressed while the prophage was stable within the lysogen (Figure 5.12). Once DeepSeq had completed their analysis,

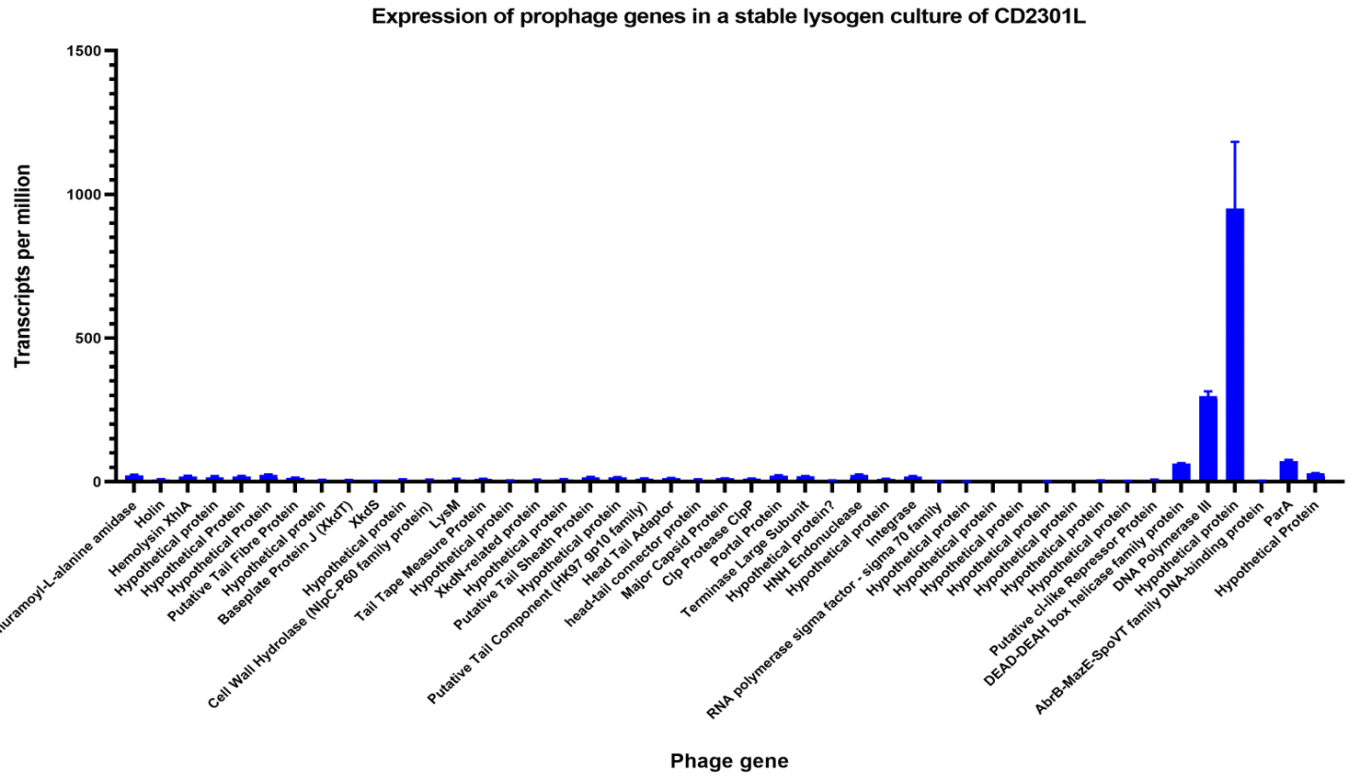


Figure 5.12 - Results of RNA sequencing comparison to the PGAP-annotated reference for CD2301L, undertaken using CLC Genomics Workbench v. 20. Results are mean values taken from three replicate cultures, with error bars indicating the standard deviations.

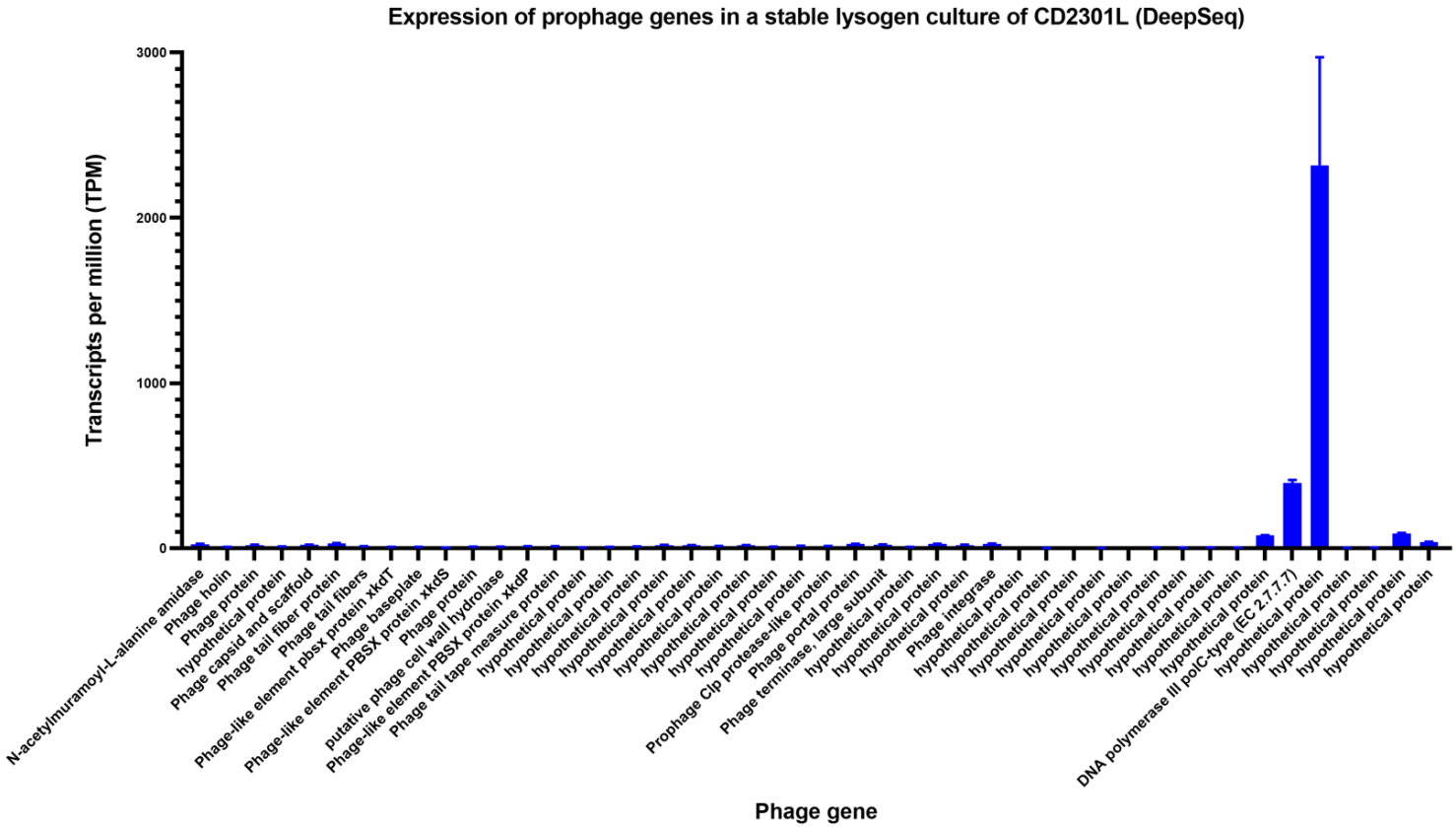


Figure 5.13 - Results of RNA sequencing comparison to the PATRIC-annotated reference genome of CD2301L, undertaken by DeepSeq. Results are mean values taken from three replicate cultures, with error bars indicating the standard deviation.

the same subset of results was extracted (Figure 5.13) and compared to the CLC results previously obtained.

The two sets of analyses show almost identical profiles except in one location where an additional low-expression hypothetical gene was observed just after the most highly expressed gene in the analysis (Figure 5.14). On review, the RNA reads had been compared to references annotated using two different annotation pipelines, with my analysis conducted using the PGAP pipeline reference for the lysogen, with gene names manually added based on the multiPhATE2 phage results, while the DeepSeq analysis used the PATRIC pipeline result which did not annotate functions for many of the genes. Further inspection showed that the two pipeline results were not identical with respect to their gene-calling, with PGAP using GeneMarkS+ (Besemer et al., 2001; Tatusova et al., 2016) while PATRIC uses the RASTk system which includes a range of different gene-calling algorithms (Brettin et al., 2015). This has resulted in additional potential genes being identified by the PATRIC pipeline which may not have been identified by PGAP.

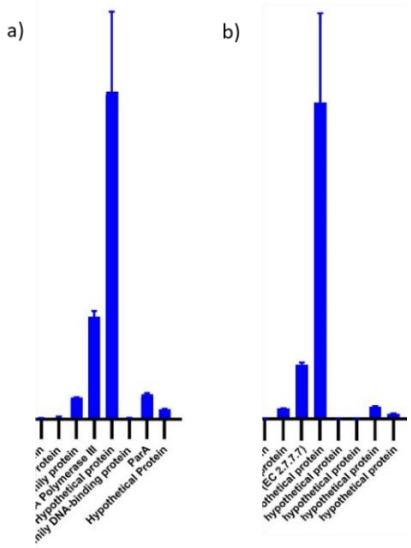


Figure 5.14 – Queried region of the prophage RNA results obtained from analysis using two different annotation pipelines. a) Results obtained from comparison to the multiPhATE2 pathway; b) DeepSeq's comparison to the PATRIC annotation.

A review of the DNA sequence in question did show an SD site 12bp upstream of this additional gene. This was confirmed through analysis using the online tool RBS calculator (Espah Borujeni et al., 2014; Salis et al., 2009) in which the DNA upstream of the queried hypothetical gene showed translation initiation rates of more than 94,000 au at position 58 (the start of the hypothetical gene - Table 5.1) indicating an active translation site. This result implies that the queried gene in the PATRIC annotation is a functioning gene, although not one with a function involved in lysogeny, given the low level of RNA expression observed. As this analysis indicated a correct annotation using PATRIC, I have used the DeepSeq results for the ongoing analysis, which are given in full in Appendix 7.5.

Table 5.1 - Translation initiation rates (au) upstream of the queried hypothetical gene identified in the CD2301L PATRIC annotation.

Position in queried sequence	ORF_number	Translation Initiation Rate (au)
0	0	0.059
6	0	0.000
19	1	6.256
58	1	94223.194

Examining the remaining results, low-level expression was observed across the majority of the phage genes. This could be attributed to background noise in cases of extremely low TPM counts, but it could also represent natural spontaneous induction of a small proportion of the phages during the 6-hour incubation period prior to RNA sampling, resulting in expression of all the lytic genes while these spontaneous phages are assembled.

Four genes had a notably higher level of expression than the apparent baseline, with one small hypothetical gene in particular (referenced as gene 1341 in the PGAP annotation of the lysogen) being highly

expressed in comparison to the others (average of 2318.1 TPM for gene 1341 as opposed to an average of 396.81 TPM for the DNA polymerase III). Given that the RNA for this experiment originates from phages which should be stably integrated into their hosts, this high level of expression implied a potentially significant role in lysogeny maintenance. This was therefore deemed to be the most likely alternative candidate for a repressor gene in this strain.

Of the remaining notably-expressed genes, *parA* (gene 1343 in the PGAP annotation – 89.41 TPM) has been highlighted as a gene which allows access to the temperate lifecycle and is used in the recently created PhageLeads programme (Yukgehnaish et al., 2022) as a determinant of whether a phage is suitable for use in phage therapy or not. The primary documented function of the encoded ParA, however, is as a DNA partitioning protein, rather than a repressor, so it is unlikely to be responsible for aiding in the maintenance of lysogeny.

The DNA polymerase III (gene 1340) and DEAD/DEAH box helicase (gene 1339) are both enzymes with documented functions involved in DNA transcription and translation; the former as a mediator of DNA replication, and the latter being involved in RNA metabolism. The level of expression of both of the encoded genes could be a cumulative result of transcription of the structural genes caused by spontaneous induction, and, as such, these are also less obvious candidates for lysogeny maintenance functions.

Notably, the *cI*-homog identified in the original annotation process (which was the initial target of this work) was not one of those which showed a high level of expression during stable lysogeny, with an average of just 2.59 TPM in the DeepSeq data. This result supported the idea that the *cI*-homolog may not be the relevant repressor in this case, and that this phage may instead use a different lysogeny system than phage λ. Given this, it was decided to target those genes which showed high expression levels during stable lysogeny (genes 1339 – 1341 and 1343), in the hope

that deletion of one or more of these genes would remove the phage's ability to form a lysogen.

5.2.4 Targeting the alternative repressor genes

5.2.4.1 CRISPR knockout vector creation

To expedite this new stage of work, the construction of the knockout vectors for the alternative repressor targets was designed to use the NEBuilder HiFi assembly system rather than the restriction and ligation method. Overlapping primers were designed using the NEBuilder online tool to amplify a homology cassette (with bookmark) and a single guide RNA targeting each of the relevant genes, however, amplification of these fragments proved difficult (Figure 5.15), and ultimately it was only possible to obtain homology cassettes for vectors targeting genes 1339 (DEAD/DEAH box helicase) and 1341 (hypothetical gene).

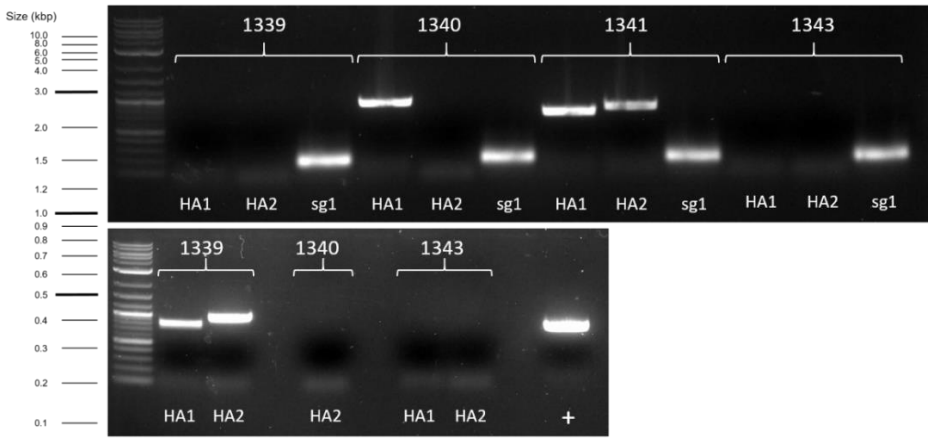


Figure 5.15 - Amplification of HiFi fragments for construction of vectors targeting the new potential repressor genes. Top) initial PCR; bottom) repeat PCR of previously failed amplifications.

A HiFi reaction was set up to insert these cassettes into the pRECas1_MCS backbone using an NEBuilder HiFi DNA Assembly Master Mix kit. The resultant ligation mixes were then transformed into XL1 Blue chemically competent cells. Following 48 hr incubation, only a single colony grew for each vector which, on screening by PCR, proved to be self-ligated vectors and therefore false positives.

5.2.4.2 Troubleshooting vector creation

Given the difficulties in amplifying homology arms alongside the relatively low TPM observed during the RNA sequencing, it was decided to discontinue work targeting gene 1343 (ParA) at this stage and focus on the vectors for which there was evidence of previous amplification. New primers were designed to amplify the homology arms for vectors targeting genes 1339, 1340 and 1341, with both HiFi and restriction/ligation variants to allow for simultaneous construction in case there was an issue with the HiFi method. However, while homology arms were amplified for both 1339 and 1341, 1340 amplifications failed again with both methods, and so work on this vector was ended at this stage. Given the time constraints of nearing the end of the project period, and in light of its recognisable function, work was also discontinued for gene 1339, leaving only 1341 (the most highly expressed gene) as a target for knockout. Unfortunately, ligations of the new 1341 vector fragments (both HiFi and T4) failed again when transformed into XL1 Blue cells, indicating that additional troubleshooting was required.

It was unclear what stage of the cloning process was being impeded during the construction of these vectors, however, discussion with colleagues identified a potential cause for concern in the presence of a large portion of the DNA polymerase III gene C-terminal in the homology arms of the 1341 vector. Polymerases are known to cause issues of toxicity during the cloning process, so it was decided to attempt to reduce the size of the homology arms to minimise the amount of polymerase gene that could interfere. Two new vectors were therefore designed to have only 400bp (p1341KO2) and 600bp (p1341KO3) homology arms respectively on each side of the gene. Two versions of each vector were designed, one suitable for HiFi and the other for restriction/ligation cloning. The new fragments were PCR amplified successfully, the restriction/ligation fragments used in a SOEing PCR, and then both variants were ligated using the appropriate method.

As a precursor check that the ligation had occurred correctly, the p1341KO2 HiFi ligation mixes were screened with PCR in advance of transformation (Figure 5.16a). The results of this check showed that the ligation stage itself could be part of the issue, with both vectors producing bands equal to the control backbone without inserts. At this time, new universal pRECas1_MCS backbone fragments were donated by my colleague Terry Bilverstone, who had been advising me on the HiFi process throughout this stage of the work. I completed another ligation for p1341KO2 using these fragments as a test and screened the ligation mix using PCR. This time, the ligation was successful (Figure 5.16b), implying that my designed primers were in some way producing incorrect fragments for the backbone that were impeding the HiFi reaction.

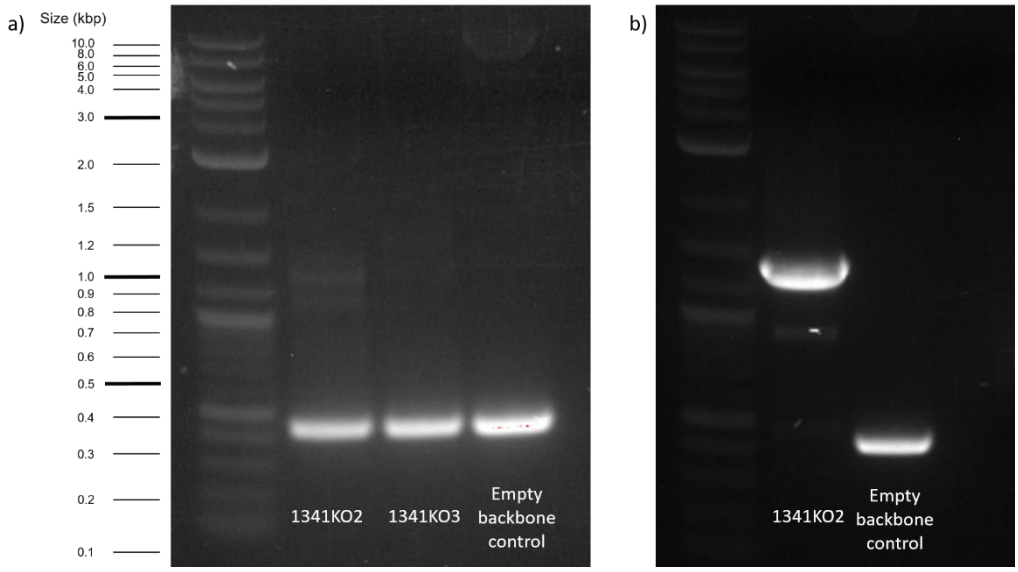


Figure 5.16 - a) PCR check of HiFi ligation mixes for p1341KO2 and p1341KO3 using my initial primer design, b) PCR check of HiFi ligation mix of p1341KO2 using backbone fragments donated by Terry Bilverstone.

The confirmed ligation mix was used to transform the vector into XL1 Blue cells, but failed again, indicating that there was also an issue at this stage of the process. While the homology arms had been amended to minimise the presence of the DNA polymerase III, the potential for the vector to still be toxic could not be ruled out. To test this, transformation was attempted

into three further sets of chemically competent cells – NEB® 5α (the recommended cells for HiFi), NEB® 10β (a more specialised NEB cell designed for larger plasmids and more troublesome transformations), and NEB® Stable (a cell specifically designed to deal with more toxic plasmids). Of these, only the NEB® Stable produced colonies, and the subsequent colony PCR to check the vectors showed a great deal of variation (Figure 5.17). Only four out of the eight colonies screened were approximately the right band size (colonies 1, 3, 5 and 8), indicating that this vector is likely very unstable in *E. coli*. Sanger sequencing of these colonies showed multiple SNPs and deletions, confirming that these vectors would be unsuitable for use.

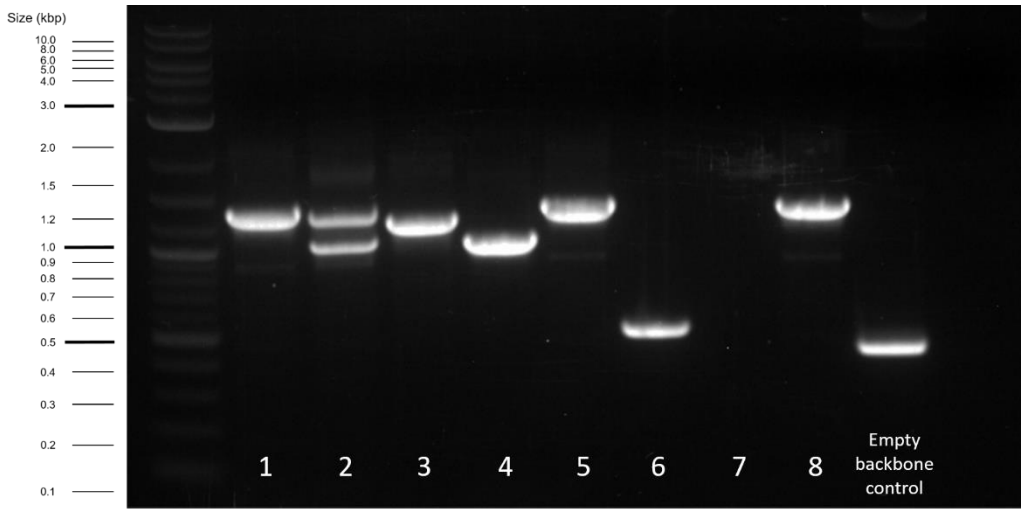


Figure 5.17 - Colony PCR of NEB® Stable transformants.

The instability of the vector during the first transformation was concerning, as inserting vectors into *C. difficile* is reliant on conjugation. A correct vector at the NEB® Stable stage would therefore still need to be transformed into sExpress cells, which would likely also struggle with vector toxicity. Given the likelihood of failure at that stage, it was decided to try a different method of testing the gene's function by silencing it through the insertion of a stop codon using targeted activation-induced cytidine deaminase (TargetAID - Nishida et al., 2016). Unfortunately, attempts to create a TargetAID vector to change the C in position 79 to a T, thereby introducing a stop codon early in the gene, were begun too

late in the project to yield any results, with vector creation left incomplete. This route would be a high priority for future work on this topic, should further funding become available.

5.2.5 Protein folding of the two repressor candidates

As a final attempt to discern which of the two potential target genes could act as the lysogeny repressor in phiCD2301, the protein sequences were compared to the phage λ repressor protein structure using AlphaFold 2 (Jumper et al., 2021) and PyMOL (The PyMOL Molecular Graphics System, v.2.5.4, Schrödinger LLC).

While the individual λ repressor protein does not have an experimentally confirmed structure available on the RCSB Protein Data Bank (https://www.rcsb.org/ - Berman et al., 2000; Burley et al., 2020), an inactive dimer bound to an operator site has been confirmed by X-ray crystallisation (Stayrook et al., 2008) the structure of which is consistent with the individual molecule structure predicted using AlphaFold (Figure 5.18). The individual molecule consists of two main domains – an N-terminal binding domain made up of 5 small α-helices, and a twisted seven-stranded anti-parallel β-sheet C-terminal domain which facilitates dimerization and interacts with other repressor molecules (Bell et al., 2000). The two domains are connected by a linkage segment.

To compare the project's target genes to this structure, the amino acid sequences of each of the potential repressor genes were input into ColabFold, a Google colab integration of AlphaFold 2 (Mirdita et al., 2022) using the default settings to generate a predicted structure for each gene. ColabFold generated 5 models for each gene which were then mapped onto each other using PyMOL to assess for similarity, and the quality checked using the predicted local distance difference (pLDDT), predicted aligned error (PAE) values, and a Ramachandran plot.

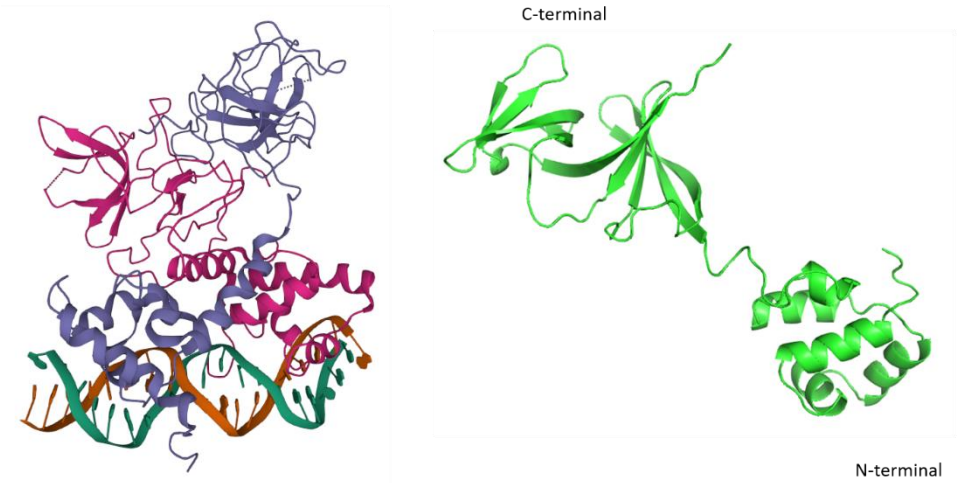


Figure 5.18 - Left) X-ray crystallisation of an inactive phage λ repressor dimer by Stayrook et al. (2008) and available as structure 3bdn on the RCSB PDB. The purple and pink areas represent the two molecules of the dimer bound to the green and yellow operator region. Right) AlphaFold prediction of the structure of the individual λ repressor molecule, visualised in PyMOL v.2.5.4.

5.2.5.1 Annotated *cI* gene

The 5 models generated for the *cI*-like gene were almost identical when aligned, except for model 5 in which the C-terminal was rotated by 180°, although the key features remained the same. pLDDT values were consistently high across the active domains, indicating a high degree of confidence in the local distances between residues, with drops in quality only observed in those areas which are likely disordered linkage segments. The PAE scores for each model were also consistent, showing two domains in which the inter-residue positional error was low and thereby lending confidence that the predicted structure is correct (Figure 5.19).

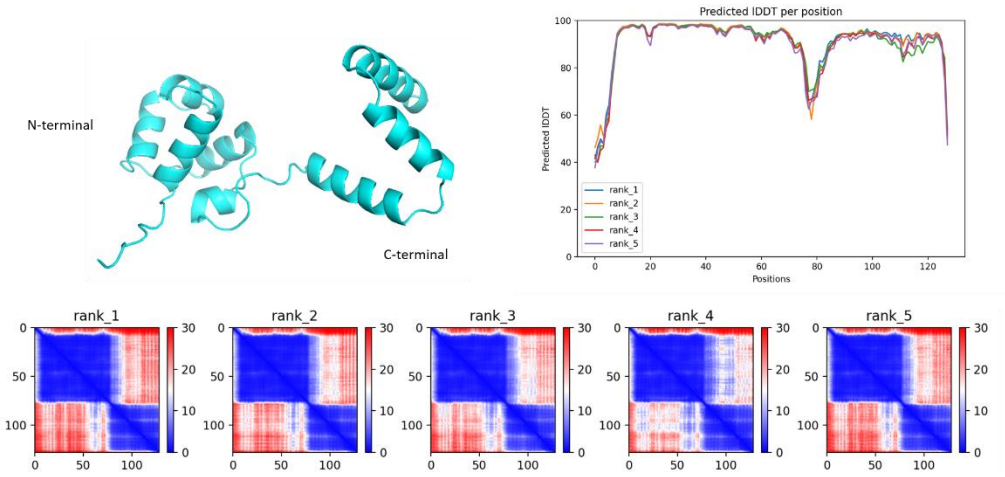


Figure 5.19 - Top left) AlphaFold 2 predicted structure of the annotated cI-like repressor in phiCD2301 – the rank 1 (model 3) structure is pictured. Top right) pLDDT scores demonstrating the two domains - pLDDT score is a measure of the local inter-atomic distance between residues, with a high score indicating good confidence in the correct prediction. Bottom) PAE scores for each of the generated models demonstrate the two domain structure that is consistent between models. PAE is a measure of the position error between two residues, with low scores being preferred in this case. Blue indicates low-scoring residues, with red indicating high scores.

As a further confirmation of the accuracy of the prediction, the amino acid sequence was input into an online Ramachandran Plot generator run by Zlab (https://zlab.umassmed.edu/bu/rama/ - Anderson et al., 2005). This plot maps the torsion angle (Ψ or Φ) of each of the amino acid residues in the sequence with respect to angles which are physically possible, and the location on the graph in which the amino acids fall indicates the likely structure which the protein would take (Kumar and Arya, 2018). Shaded areas on the plot indicate allowable torsion angles, with darker areas indicating the most highly preferable conformations. White areas are those where the angles could not exist within a viable protein due to steric interference or the fact that they would require the atoms to occupy the same physical space (Figure 5.20).

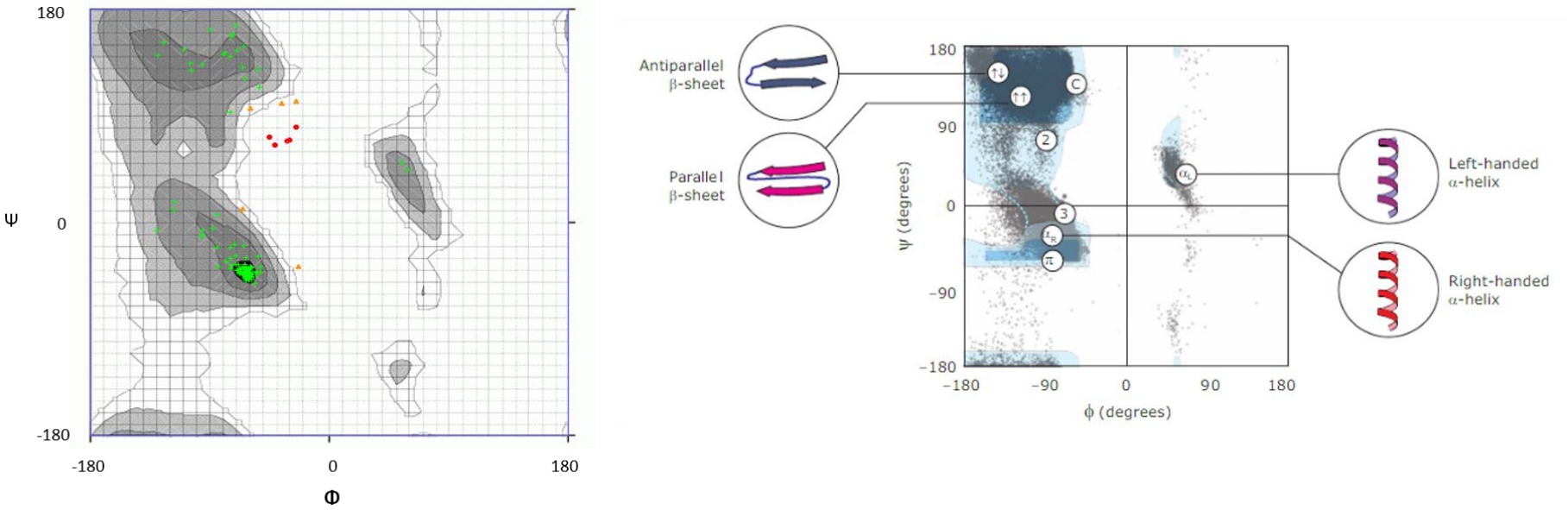


Figure 5.20 – Left) Ramachandran plot showing the torsion angles of amino acids within the cI-like repressor protein of phiCD2301. Right) Protein structures associated with each area of the plot, taken from Kumar and Arya (2008). Green indicates those results falling in highly preferred conformations, brown are preferred, while red are questionable observations.

The Ramachandran plot for the annotated *cI*-like repressor indicates that 92% of the angles fall within the highly preferred regions of the graph, with the majority clustered within the area indicating right-handed α-helices. This is consistent with the structure predicted using AlphaFold 2. Of the remaining 8%, 4% are still within preferred zones, with only the final 4% being questionable observations. Combined with the previous quality scores one can have a good degree of confidence that the AlphaFold 2 prediction is very close to the true structure, although it would not be possible to confirm that without experimental analysis.

With the quality of the predicted structure being acceptable for use, PyMOL v.2.5.4 was employed to align the *cI*-like gene against the previously-established structure of the λ repressor (Figure 5.21).

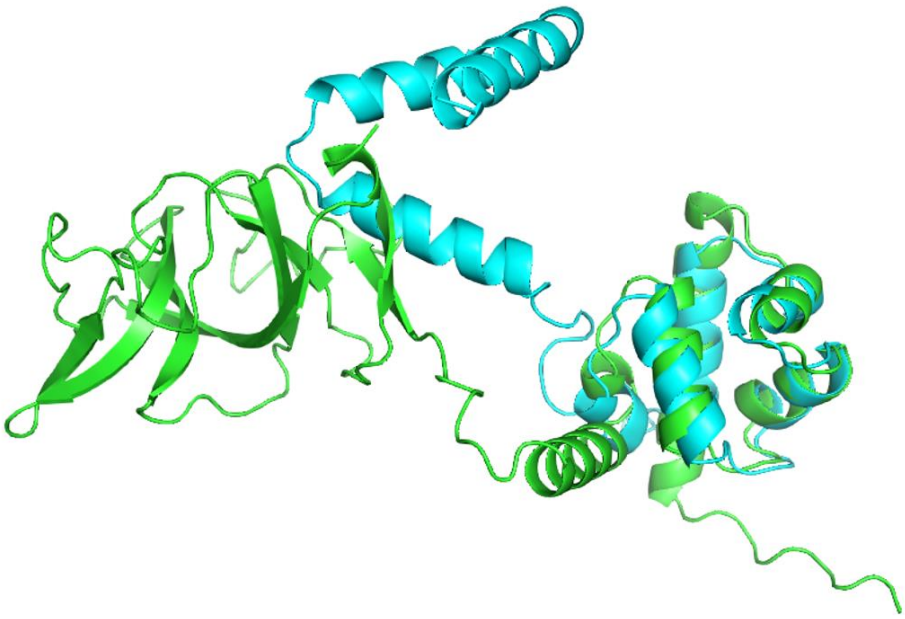


Figure 5.21 - Alignment of the phage λ repressor molecule structure (green) to the generated cI-like repressor of phiCD2301 (cyan) using PyMOL v.2.5.4. The N-terminal α-helices of the queried molecule show a degree of similarity to the N-terminal of the λ repressor, while the C-terminal is of a very different composition.

As expected given the obvious differences in the C-terminal domain (β-sheets for λ repressor but α-helices for the *cI*-like gene), the alignment

was not very good in this region. However, while not a perfect match, the N-terminal domain does show some similarity, with all five of the *cI*-like molecule's α-helices aligning to some extent with the λ repressor N-terminal. The degree of similarity implies that the *cI*-like gene in phiCD2301 could indeed have a similar binding function to that of the λ repressor, and therefore should be re-instated as the primary target for gene knockout in this phage.

5.2.5.2 Gene 1341

Establishing that the *cI*-like gene may have been correctly identified as the repressor for phiCD2301 raised the question of why gene 1341 demonstrated such a high degree of expression during stable lysogeny if it was not the repressor, and so the structure of this molecule was also investigated in the same manner.

AlphaFold 2 predicted two slightly different structures for gene 1341 which consisted of an N-terminal α-helix with a disordered tail for all, but with two models (rank 2 and 4) also showing a small C-terminal α-helix. The pLDDT and PAE scores provide a reasonable degree of confidence in this structure, with the N-terminal α-helix showing above 90% pLDDT for all 5 models generated and with a low PAE score in all cases, while the C-terminal uncertainty is also reflected in the variable scores at the tail end of the graph (Figure 5.22).

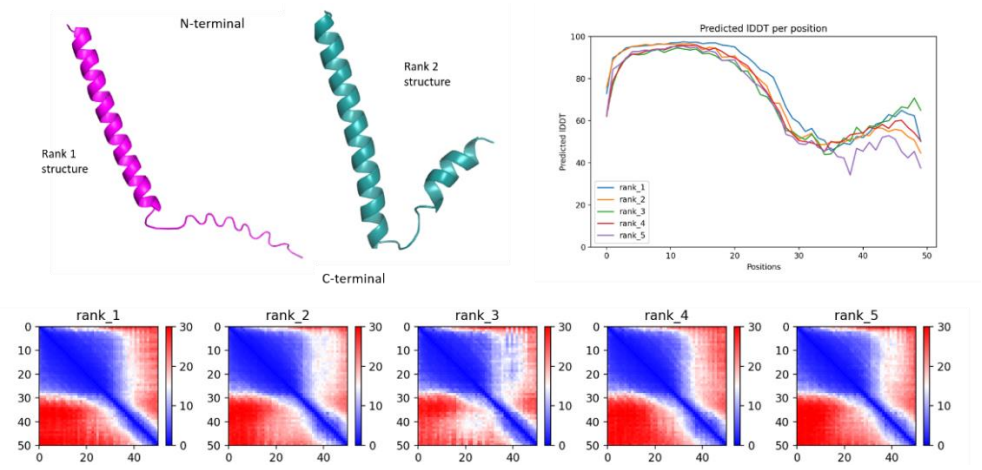


Figure 5.22 - AlphaFold 2 prediction for the structure of gene 1341. The main α-helix is present in both variations with the models ranked 2 and 4 also containing a smaller α-helix at the C-terminal end. The uncertainty is reflected in the pLDDT scores for the C-terminal, as well as in the PAE graphs where a slight bulging in the blue C-terminal domain area reflects the difference between the models.

The Ramachandran plots (Figure 5.23) also reflect this uncertainty in the C-terminal. For the rank 1 model, only 72% of the results fell in either the highly preferred or preferred conformations, with 28% marked as questionable, while the rank 2 model had 98% of the results in preferred or highly preferred regions. For both models, the majority of the results are clustered in the right-handed α-helix portion of the graph, which is consistent with the predicted structures insofar as both variants show this major domain feature. Given the higher level of preferred conformation results in the Ramachandran plot, it is likely that the second prediction with the additional α-helix is the more accurate structure.

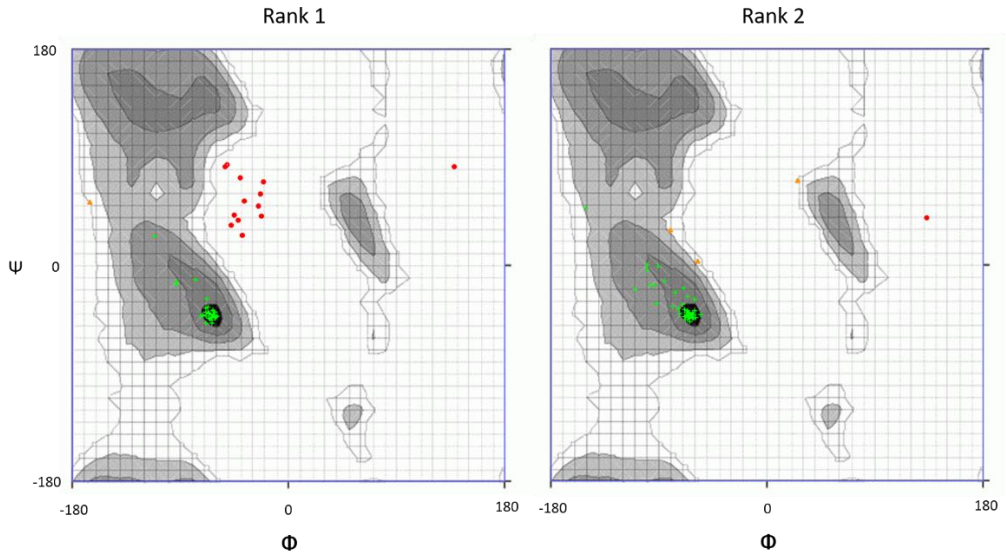


Figure 5.23 - Ramachandran plot of gene 1341. The single helix model, while represented in 3 out of 5 of the predictions, has a higher rate of questionable results (28%) than the dual helix model (2%) which is only shared by 2 of the predictions.

The predicted structure for gene 1341 shows no similarity to that of the phage λ repressor. The presence of the secondary helix, however, along with the small size of the molecule (only 50 residues in size), suggested a possible function as a signal, transmembrane or even quorum-sensing peptide. AlphaFold 2 was not trained using transmembrane proteins, so is unlikely to model transmembrane areas with any degree of accuracy. To account for this, the amino acid sequence was analysed using Phobius (https://phobius.sbc.su.se/ - Käll et al., 2004) on the default settings to determine whether there were any transmembrane or signal peptide elements to the molecule.

The Phobius results clearly showed that, based on the amino acid sequence rather than structure predictions, the 1341 gene does appear to have a transmembrane region (Figure 5.24). Contrary to my initial opinion, this transmembrane region seems to be within the primary α-helix rather than the disordered end section where AlphaFold 2 showed the questionable results. There is no indication from this analysis that there is a signalling region involved in the molecule.

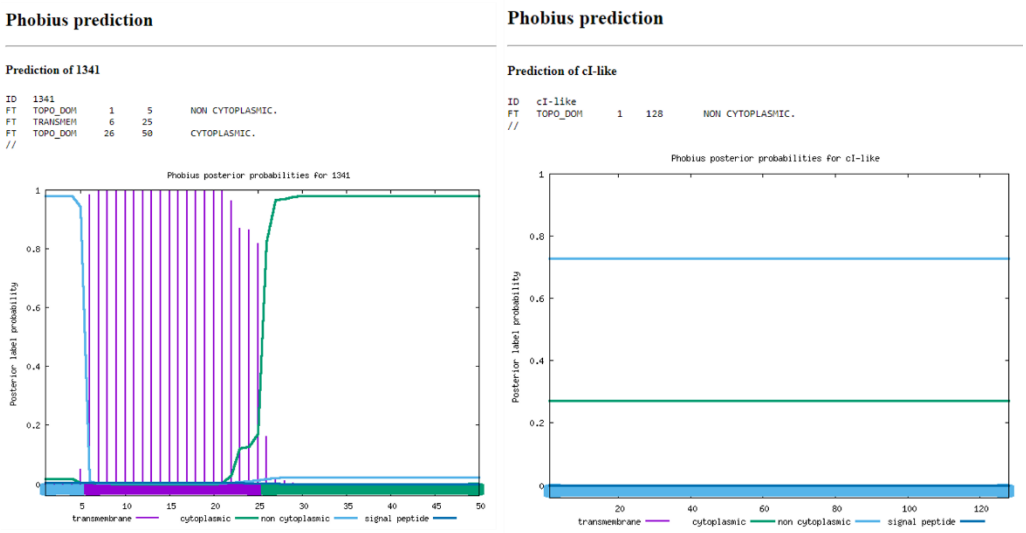


Figure 5.24 - Phobius transmembrane and signal peptide prediction of gene 1341 with the cI-like repressor provided for comparison. Gene 1341 shows a defined transmembrane region in the N-terminal α-helix with both cytoplasmic and non-cytoplasmic regions, while the annotated cI-like gene is fully cytoplasmic.

Transmembrane proteins are not uncommon in bacteriophages, however, their functions are poorly understood (Kristensen et al., 2015). Therefore, it is not possible to discern from the structure what role this molecule may play in the phage lifecycle. However, based on the lack of structural similarity to the λ repressor, it is unlikely that this protein serves the same function, therefore it should not be considered a target for any future repressor knockout attempts. Further investigation into the function of this protein in future may lead to insights into the biology of this phage, especially given its high degree of expression during stable lysogeny.

5.3 Discussion

The main impediment to the use of temperate phages in phage therapy is their ability to repress their lytic genes, allowing the phage to successfully integrate into the bacterial genome and remain dormant instead of immediately killing the host. Removing the repression mechanism is therefore a key requirement for engineering phages to be suitable for therapeutic use.

Identification of the genetic control systems underlying lytic repression began in the mid to late- 20th century using *E. coli* phage λ. This work uncovered the *cI*/*cro* bi-stable switch that became a model for genetic switching systems (Hochschild et al., 1986; Ptashne, 1967), and focus on this system has continued into the 21st century (Anderson and Yang, 2008; Oppenheim et al., 2005; Vohradsky, 2017). Similar mechanisms have now been identified in several other unrelated species of phages such as lactococcal phages (Kenny et al., 2006), and although alternative mechanisms have been uncovered in phages infecting other bacterial types such as Gram-positive bacteria (e.g. quorum-sensing and protease-encoded lysogeny - Brady et al., 2021a) the bi-stable switch remains the predominant consideration when determining whether a phage is lytic or temperate.

Examination of the genome of phiCD2301 in this study uncovered an annotated *cI*-like helix-turn-helix transcriptional regulator near the tail end of the phage, upstream of the previously identified *int* gene. It was therefore decided to target this gene to determine whether its removal resulted in a reduction or elimination of lysogen formation. Unfortunately, initial attempts to create appropriate CRISPR-Cas9 vectors had limited success, with only one correctly constructed vector being obtained. Conjugation of this vector into CD2301L then uncovered an issue in that the standard plate-based method of recovering transconjugants may not be appropriate for this experiment, given the expectation that successful removal of the *cI*-like repressor would elicit cell death through phage induction as it would no longer be able to repress its lytic genes. This was borne out by the fact that the only colonies which could be recovered from the plate conjugations were either incorrect mutations or contained wild-type prophages.

Given the issues acquiring a successful mutant and the time constraints imposed on the project by the Covid-19 pandemic, RNA sequencing was employed to confirm whether this *cI*-like gene could be the repressor in

order to avoid lengthy delays targeting an incorrect gene. The premise for this test was based on a study by Owen et al. (2020) who showed that a stably integrated prophage only expresses genes important for bacterial fitness or maintaining the lysogenic state, including the repressor. RNA sequencing of samples taken from a stable culture of lysogens should therefore yield only a few potential target genes. The hope was that one of those more highly expressed genes would be the annotated *cI*-like repressor.

As expected, the RNA results showed only four genes with a noticeable level of expression, and all but one had been annotated as having recognisable functional domains linked to viral reproduction. The *cI*-like gene, however, was not among them, raising doubts as to whether this gene was the correct target. The gene with the highest expression was a small hypothetical gene with no known function. Given that alternative mechanisms have been identified in Gram-positive phages, it was possible that the *cI*-like gene had some other regulatory function and that this phage employed an alternative mechanism using this highly expressed small protein. New vectors were therefore designed to target this gene for removal instead.

Unfortunately, the creation of this new vector was also difficult, despite two different construction methods (HiFi and restriction/ligation). Through a process of elimination regarding possible impediments, it was eventually determined that the vector was likely to be toxic, interfering with the transformation process. An attempt to silence the gene (instead of removing it) using TargetAID was also hampered by cloning issues and was eventually discontinued due to time constraints.

A final attempt to uncover which of these two targets could be the lytic repressor was prompted following a discussion with phage bioinformatics specialist Prof. Rob Lavigne during a conference attendance near the end of the project period. Prof. Lavigne suggested that the low RNA expression of the *cI*-like gene may not necessarily indicate that this was

not the lytic repressor, but instead could be a reflection of the steady state of the lysogen requiring little repressor expression to maintain the dormant state. He suggested taking a bioinformatics approach to determine whether there were any structural similarities to the phage λ repressor which could indicate a shared function.

Protein-folding prediction algorithms have taken huge leaps in accuracy over the last few years with the increases in genome sequencing, experimental confirmation of protein structures and computational power limits. A further significant boost has occurred recently in the form of DeepMind's deep learning modeller AlphaFold 2, which vastly out-performed both its predecessor (AlphaFold) and other software attempts during the CASP14 blind protein-folding contest, generating predictions of unpublished protein structures which closely matched the experimentally confirmed structures (Callaway, 2020). It is therefore now much more feasible to accurately predict protein structures to inform experimental investigations. To test the hypothesis that the *cI*-like gene may still be the repressor despite the RNA results, AlphaFold 2 was used through the online portal ColabFold to predict the structures of both of the target molecules which could then be compared to previously predicted structures for the λ repressor.

Although the monomer form of the λ repressor is only available as an AlphaFold prediction, there are several experimental examples of the structure of modified versions of the repressor available on the PDB for comparison purposes, including an X-ray crystallisation of the repressor dimer bound to an operator region from Stayrook et al. (2008). This data, in particular, confirms that the λ repressor consists of two domains – a series of α-helices in the N-terminal which bind to the operator region, and an antiparallel β-sheet C-terminal responsible for dimerization and inter-dimer interactions (Bell et al., 2000). It was anticipated that any homologues with a similar function would adopt a similar structure.

The predicted structure for the phiCD2301 *cI*-like gene did indeed show some level of similarity to the λ repressor. The overall structure consisted of a bundled N-terminal made up of 5 α-helices connected by a linkage segment to a second domain of 3 α-helices. An alignment of the two molecules demonstrated that the conformation of the N-terminal helices was homologous to the N-terminal of the λ repressor, implying a similar DNA-binding function, although the C-terminal domains were entirely different to each other. While the presence of a DNA binding domain is consistent with our expectation of a CI protein, it should be noted that similar structures have been known to have functions other than in lysogeny, such as the SinR repressor involved in the sporulation of *Bacillus subtilis* (Lewis et al., 1998), and so experimental confirmation is still required before being able to determine whether this gene is the lytic repressor for this phage. It is, however, the more obvious candidate of the two queried genes, especially given its identified homology during BLAST searches.

As for the small hypothetical protein 1341 identified during the RNA sequencing, the structure prediction for this protein using AlphaFold 2 revealed a small protein consisting of either a single long α-helix with a disordered tail or, more likely, a long α-helix linked to a smaller helix by a linkage section. Without the helix-turn-helix motif common to DNA-binding proteins already seen in the λ repressor (Aravind et al., 2005), the function of this protein (or possibly more accurately a peptide, given its small size) is difficult to discern but is unlikely to be similar repression mechanism. However, the dual-helix predicted model may suggest that peptide 1341 could be either a signalling peptide or a transmembrane protein of some sort. AlphaFold 2, while highly accurate in the majority of cases, was not trained using transmembrane proteins, and the developers acknowledge that the software is not able to model these types of protein as accurately, although a recent assessment did show that a cautious approach does provide acceptable results (Hegedűs et al., 2022).

To test whether peptide 1341 does have transmembrane regions, the amino acid sequence was assessed using the transmembrane and signalling peptide prediction programme Phobius, which has been designed not only to identify either type of protein but also to more accurately differentiate between the two (Käll et al., 2004). The results of this analysis did identify a transmembrane region in the N-terminal of the molecule with a small non-cytoplasmic region before and a larger cytoplasmic region after, but no indication of a potential signalling function.

Transmembrane proteins are a peculiarity in phages, given that they do not individually have a membrane on which to act, and they have been thought to primarily fall into one or two classes – those involved in viral scaffolding and those that interact with the host membrane in some way (Kristensen et al., 2015). Given the location in the genome of the phage within the regulatory zone, it is less likely that this peptide could be involved in virion production, but the question remains as to what function it could be providing that involves the host membrane during the dormant prophage period, as evidenced by the RNA sequencing. This peptide may have some function in improving bacterial fitness such as in the provision of super-immunity, however, experimentation is required to narrow down the potential options.

Despite several avenues being addressed and multiple targets identified, it has not been possible to remove the lytic repressor from phiCD2301 at this time. However, the recent bioinformatics analysis does lend more confidence to the idea that the initial target (the *cI*-like helix-turn-helix transcriptional regulator) should be favoured for any future work on this topic. New vectors conjugated using the modified harvesting method would be ideal, with the intention of either full removal of the gene or removal of the N-terminal in a similar fashion to that adopted by Cady et al. (2012) to remove the binding domain. Alternatively, should vectors continue to be difficult to construct then TargetAID could be considered to silence the gene. It is unlikely that silencing the gene would produce a

phage that was acceptable in medical circles, but it would act as a test to confirm the removal of the lytic repression and therefore that this gene is the correct target.

Although unrelated to the main aim of creating a lytic phage, elucidating the function of the small highly expressed peptide would also be an interesting line of enquiry, as the more information that can be gathered on the biology of the phage, the more likely that any medical concerns can be alleviated should an engineered phage be created that is suitable for therapeutic use.

5.4 Chapter Summary

The key element allowing temperate phages to form stable lysogens (and therefore create an immune sub-population of bacteria) is their ability to repress their lytic genes. The removal of the control mechanisms underlying this lytic repression through genetic engineering is therefore a high priority if phages are to be created which are suitable for therapeutic use. This chapter details the efforts made to construct vectors targeting an annotated *cI*-homolog that was identified in the genome of phiCD2301, as well as the RNA work undertaken in an attempt to confirm that this was the correct gene to target. The RNA work led to the identification of another potential target, but knockout vectors were unable to be constructed to test this hypothesis. Latterly, protein structure predictions were undertaken in an attempt to bioinformatically establish which of the two targets was most likely to be the repressor. This analysis showed that the original target (the *cI*-like gene) contained a bundled helix-turn-helix motif in the N-terminal which was homologous to that found in the phage λ repressor, implying a potential shared function. The small highly expressed hypothetical protein identified by the RNA sequencing appears to be a transmembrane protein with no currently known function. The degree of expression of this small peptide by a stable prophage is, however, intriguing and warrants further investigation.

Unfortunately, the aim of this chapter (to remove the lytic repression mechanism of phiCD2301) was not attained. The results of the protein structure analysis, however, did lend weight to the original hypothesis that the *cI*-like gene is the desired target for removal.

Chapter 6

Conclusions

6 Conclusions and further work

6.1 Conclusions

Antibiotic resistance in human pathogens is an increasing problem worldwide and a global response is required to combat it. Work is ongoing to discover new antibiotics (Stokes et al., 2020; Thong et al., 2021), improve infection prevention practices (Hascic et al., 2022; Suzuki et al., 2022) and develop alternative treatment regimens to bypass resistance (Allen, 2017; Kumar et al., 2021). Many of these alternatives, however, are still in their infancy and require further research before they can be applied therapeutically.

Bacteriophages are one of those alternatives being considered. As a natural predator of bacteria, they present many advantages over antibiotics, not least their ability to evolve to overcome developed bacterial resistances in a way that static chemical antibiotics cannot (De Sordi et al., 2019). Their therapeutic value has been recognised for decades, with eastern European states such as Georgia using phages for medical purposes since the early 20th century, but a great deal of research is still required before bacteriophage treatments will routinely meet western medical standards for use.

One question that needs to be resolved as part of this is the debate around the use of whole phages vs. phage tail-like particles (PTLP), also known as bacteriocins or tailocins. Bacteriocins are structurally similar to phages in that they consist of the tail and tail fibres but they lack the capsid (and therefore genome) of a viable phage virion (Bhattacharjee et al., 2022). They have a similar capacity to bind to a limited range of target host cells via the tail fibres (Schwemmlein et al., 2018), at which point they have two forms of action – 1) they create an entry pore in the cell membrane through which DNA could enter, but without the capsid to block the tube the bacteriocin acts as a channel through which ions can flow, thereby disrupting the membrane potential and killing the cell; 2) the

bacteriocin can act as a peptidoglycan hydrolase, inhibiting cell wall synthesis and disrupting the cell membrane without forming a pore (Rendueles et al., 2022). Either case results in cell death, but without the production of new bacteriocins, unlike viable phages which will reproduce as part of the lytic cycle.

The lack of a capsid and genome makes bacteriocins a tempting option over complete phages because it removes some of the disadvantages which complicate whole phage therapy; for example, no viral genome means no direct interaction with the bacterial genome and therefore no risk of transduction of undesirable genes. As such, they have been heavily targeted as potential therapeutics, with a natural pyocin having been tested on *P. aeruginosa* (Merrikin and Terry, 1972), and genetically engineered bacteriocins tested on Shiga toxin-producing *E. coli* (Ritchie et al., 2011) and *C. difficile* (Gebhart et al., 2015) with some *in vivo* success. However, the lack of a genome prevents bacteriocins from being able to self-replicate (Bhattacharjee et al., 2022), therefore placing the burden of manufacturing on external sources, with batch fermentation or chemical synthesis methods being costly and slow (Hols et al., 2019). Despite this, some bacteriocins have already been authorised for use as biopreservatives, such as nisins (EFSA Panel on Food Additives and Nutrient Sources added to Food et al., 2017) but not yet as medical treatments.

While bacteriocins show some promise in their use as antimicrobials, phage therapy using whole and viable phages has some advantages as well. Aside from their ability to self-replicate ensuring that only small doses are required to control or eradicate an infection (Chan et al., 2013), phages appear to have a greater capacity to deal with biofilms than bacteriocins, with a bacteriocin isolated from *Bacillus* spp. SH10 being unable to fully clear an established biofilm of a susceptible *Proteus mirabilis* infection even at relatively high concentrations (Shayesteh et al., 2020). This may be due to the bacteriocin being unable to cope with evolved resistances in the biofilm which would protect against phage

attack, such as thickening of the extracellular polymeric matrix to trap particles and protect the constituent bacterial cells (Ferriol-González and Domingo-Calap, 2020). Phages, on the other hand, have a range of mechanisms to overcome these resistances, including the production of lysins and depolymerases to degrade the cell wall and any extracellular substances which form a barrier to cell entry (Ferriol-González and Domingo-Calap, 2020). As such, while bacteriocins may present a potential alternative to full phage therapy, the use of whole and viable virions is preferable when dealing with biofilm-forming infections such as *C. difficile* and the ESKAPE pathogens (*Enterococcus* spp., *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp. - Patil et al., 2021).

However, the major limitation for the use of phage therapy with regards to *C. difficile* is that it requires the use of obligatory lytic phages to maximise the antimicrobial effect while minimising the risk of transduction of undesirable genes. Unfortunately, there are currently no documented obligatory lytic phages available for this species. This project, therefore, had the following aims:

1) Attempt to isolate and characterise novel *C. difficile* phages from environmental samples in the hope of obtaining a naturally-occurring lytic phage.

2) Employ CRISPR-Cas9 genome editing techniques to modify temperate *C. difficile* phages to remove their ability to form lysogens by:

    a. Deleting the potential repressor gene to prevent lytic gene repression, and
    b. Deleting the integrase gene to remove the potential for transduction by prophage integration.

Initially, it was believed that the engineering route was the most likely to succeed, as with sufficient information regarding the phage biology and genome it should be possible to remove the genetic mechanisms controlling lysogeny in the chosen phage. CRISPR-Cas9 vectors were, therefore, constructed to remove the two identified key genes for lysogeny – the *int* (integrase) gene and the *cI*-homolog - from a lysogen of phiCD2301.

Unfortunately, this strategy proved challenging for several reasons. While deletion of the integrase gene was relatively easy, subsequent work was hampered by the need to develop a method of removing the engineered phage from its host, a method which was ultimately unable to be completed due to time constraints. Removal of the repressor was also challenging, with a range of cloning issues coupled with doubt over which gene was, in this case, the correct target. While bioinformatics has provided some direction with regard to the latter issue, these difficulties highlight the need for a thorough understanding of the underlying biology of each phage when undertaking genetic engineering. An efficient pipeline will be needed to gather the required information prior to gene editing if the engineering of therapeutic phages is to become a common approach.

While the engineering approach was more complex than initially intended, it remains a very promising strategy. The issues encountered in this project stem, in part, from an initial lack of experience in the field as well as time constraints and difficulties imposed by the Covid-19 pandemic, and so it is my belief that, with further time, solutions could be found that enable the aims to be met. Also for consideration, phiCD2301 may not have been the ideal phage to use as a proof-of-concept for this method. This phage was selected based on the consistency of conjugation into CD2301L and the overall size of the phage genome, however, vector creation has been consistently problematic. In some cases, this is likely due to the compact nature of the genome requiring adjacent genes (such as DNA polymerases) to be present within the

homology arms of the knockout vector, inhibiting cloning. Other phages with a different gene order may not encounter the same difficulties. It would, therefore, be worth examining other temperate phages of *C. difficile* to see whether cloning of vectors for alternative phages yielded better results.

However, while the engineering approach stalled, there was success in the isolation of novel phages using enrichment techniques. Isolation experiments for *C. difficile* phages are often undertaken through induction of existing prophages from *C. difficile* strains, which may partly explain why no lytic phages have been obtained, as all phages recovered using this method would be temperate. Our isolation experiment using wastewater obtained from around the Thames Valley region, however, identified three novel phages, two of which were temperate, as expected, but one (phiL14WAR) was discovered to be a naturally occurring integrase-deficient mutant.

Experimental tests as to whether phiL14WAR retains any other lysogenic mechanisms were not undertaken, although genes were identified which may indicate the use of a protease-encoded lysogeny system. However, irrespective of the presence or absence of these other genes, this phage did present us with the chance to investigate the effects of integrase deficiency on phage action when compared to temperate phages. These tests showed a marked improvement in antimicrobial action on live cultures in a range of hosts and an absolute lack of ability to form lysogens. While removal of the integrase alone is unlikely to entirely assuage medical concerns regarding the use of temperate phages (given the potential for remaining repression mechanisms to still facilitate super-immunity), the characteristics of this phage do support the idea that targeted removal of just one or two genes could be sufficient to switch a temperate phage to a permanently lytic lifecycle. PhiL14WAR should therefore now be a priority phage for characterisation and engineering, as with its reasonable host range it could become an effective phage for

therapeutic use if any remaining repression mechanisms could be removed.

While it has not been possible to fulfil all of the aims detailed at the start of this project, each strand of work has contributed knowledge which should make it possible to complete these aims in the future. The acquisition of a naturally integrase-deficient phage, however, is a particular success, both because it has high antimicrobial action in itself and because it lends weight to our expectations of genetically engineered integrase mutants. Further work on this phage is key to progressing towards phage therapy for use against CDI, and should therefore be prioritised as part of future work.

6.2 Future work

There is a range of further work which could be conducted, should additional funds be obtained to continue this work in the laboratory.

6.2.1 Integrase-deficient mutant

For this strand of work, the highest priority is the completion and testing of the expression vectors to enable the induction of the integrase-deficient phage from its host. Should these not be successful, an alternative approach would be to insert the integrase gene into the *C. difficile* genome itself, potentially at the *pyr-E* locus using existing in-house procedures, to determine whether endogenous expression of the integrase is sufficient to allow phage induction. As the integrase would not be within the prophage, it should not be included in virion transcription, and therefore may allow for mutant virion production without reversion to wild-type. Once this phage is available, similar experiments to those conducted on phiL14WAR can be undertaken, including testing the antimicrobial action on cultures of susceptible hosts and testing for lysogen formation. Should this prove successful, the mutant lysogen can be used as a base from which to make further mutations.

In addition to further genetic engineering, this mutant may also be suitable for investigations into more general phage biology and the engineering of synthetic phages. An important consideration for developing synthetic phages is the ability to discern which genes are essential to the survival of the organism, which is difficult with phages due to the mosaic nature of their genome and their penchant for horizontal gene transfer with their hosts. TraDIS is often used to establish gene essentiality in bacteria but has so far only been used to investigate host-side interactions with phages rather than being used on the phages themselves (Cain et al., 2020; Cowley et al., 2018; Goh et al., 2017). An integrase-deficient mutant, however, may provide a subject for TraDIS to investigate phage biology, as it avoids the risk of losing any relevant mutants due to spontaneous induction. These mutants could then be tested to determine core genes which are required for virion production, as well as which genes are responsible for phage infection. These genes could then form the base from which to build a synthetic phage suitable for infecting *C. difficile*.

6.2.2 Repressor removal

The result of the protein structure analysis indicates work on the *cI*-homolog gene should be re-visited. Given the issues experienced so far isolating mutants using the existing plasmids, TargetAID could be employed in an attempt to quickly generate mutants in which the gene is truncated with a premature stop codon, allowing confirmation that this gene is the correct target. If the expected phenotype is observed, new vectors could then be constructed to remove the gene in its entirety using CRISPR-Cas9 but with conjugation being undertaken using the modified method detailed in chapter 5.

Once the gene has been deleted and a suitable mutant phage isolated, the antimicrobial action of the phage can be tested using one-step growth curves, as for phiL14WAR. In this case, direct comparisons can be made to the parent temperate phage, providing greater confidence that the

removal of the gene is responsible for the outcomes observed. The phage would also be tested to establish whether it is capable of forming a lysogen.

If the phage shows both increased antimicrobial action and a lack of lysogen formation, then the same vector can be transformed into the integrase-deficient lysogen to create a double knockout. This stage relies on the success of the integrase expression vector in releasing integrase-deficient mutants from the genome, but it is, in my opinion, necessary to obtain a phage with both the integrase and repressor removed if that phage is to be considered for phage therapy. This phage could then be subjected to the same tests as those mentioned previously, as well as a thorough characterisation including phage burst size, stability of the phage preparation, and identification of potential antibiotic resistance genes (Fernández et al., 2019; Suh et al., 2022).

6.2.3 phiL14WAR

The results of tests undertaken on this novel phage are very promising with respect to phage therapy. However, only *in vitro* experiments have been conducted so far. An animal study is now being planned to test this phage in a hamster model to demonstrate that antimicrobial action remains at similar levels to the *in vitro* tests without any lysogen formation. This study will involve challenging hamster models with an appropriate strain of *C. difficile*, after which doses of the phage will be administered every 8 hours. The study will aim to determine the therapeutic potential of phiL14WAR for reducing the symptoms associated with CDI. Faecal samples will be taken during this experiment to monitor the persistence of *C. difficile* in the animals during treatment, and caecum samples will be taken at the end of the experiment to allow for histopathological assessment.

In addition to the animal study, and especially if the results are favourable, this phage should also be characterised further to identify whether other lysogeny-related genes remain intact. To do this, a similar

method to Platt et al., (2003) could be used to integrate phiL14WAR into a potential host, i.e. creation of an integrase expression plasmid to facilitate integration at native attachment sites (if they exist) or introduced sites (if they do not). If a stable lysogen can be formed, then this implies that repression machinery is present, and attempts can be made using CRISPR-Cas9 plasmids to remove candidate repressor genes from the phage. In the case of phiL14WAR, genes linked to protease-encoded lysogeny have been identified as part of the genome sequencing, and so these genes would be the initial targets for such work. Bioinformatic analysis could also be conducted to identify any antibiotic-resistance genes which should also be removed.

6.2.4 Longer term goals

The long-term aim for this work is to develop a cocktail of engineered lytic phages which could be used to treat strains of *C. difficile*. As such, once methods have been determined which can successfully generate a lytic phage from a temperate parent, these techniques can be applied to other phages in the culture collection to prove replicability and consistency. Further isolation attempts can also be undertaken to acquire additional phages targeting a single strain of *C. difficile* that can then be used to test the *in vitro* and *in vivo* efficacy of an engineered phage cocktail.

While the current discourse does not approve of genetically engineered organisms for clinical use, I believe that this will not always be the case. A greater understanding of the biology and uses of bacteriophages across a range of arenas will eventually allow for sufficient control of the process to satisfy medical regulation requirements. It is therefore important to ensure that research is undertaken early and rigorously to facilitate phage therapy implementation for CDI at the earliest opportunity.

Chapter 7

Appendix

7 Appendix

7.1 *C. difficile* strains used for host range testing

CRG Strain name	Real strain name	PCR-ribotype
L26	001-3	001
L25	001-2	001
L28	001-5	001
L30	001-7	001
L24		001
L27		001
L29	001-6	001
001	001-3	001
R8366	R8366	001
NCTC 11205		001
NCTC 11207		001
NCTC 11209		001
BI-9		001
LIV29		001
EK9	MF081988	002
EK2	CD 16839	002
EK3	8083598	002
EK4	8085053	002
EK6	8092419	002
EK7	9001966	002
L33	Subtype A2	002
EK8	9003048	002
002	Subtype A2	002
EK5	8069158	002
TL178		002
CD1101		002
003	G,08,0000780	003
3232		010
7032989		010
012		012
630WT		012
L31	serogroup H	014
L32	serogroup K	014
EK14		014
EK15	8085054	014
EK16		014
014		014
TL176		014
CD2301		014
015	UK1	015

CRG Strain name	Real strain name	PCR-ribotype
TL174		015
CD3601		015
CD1902		015
CD8501		015
CD0301		015
CD4501		015
CD4701		015
CD4202		015
CD5901		015
CD7301		015
CD4502		015
CD8401		015
EK19	675,1	017
L23	Serogroup F	017
L22	GAI 95601	017
EK20	8078003	017
EK21	8079089	017
017	Serogroup F	017
M68		017
CF5		017
020	R10079	020
023	Clos di 21	023
CD305		023
L2	81568	027
L3	81566	027
L6	5108111	027
L8	32219	027
L9	M246	027
L10	2191	027
L11	51556	027
L14	60902	027
L16	26131	027
L17	26173	027
L19	CDC 32	027
L20	CDC 38	027
L21	M13042	027
DH349	DH349	027
DH32	DH320	027
DH196	DH1916	027
DH1396	DH1396	027
DH1432	DH1432	027
DH1834	DH1834	027
DH1466	DH1466	027
DH1858	DH1858	027

CRG Strain name	Real strain name	PCR-ribotype
DH1751	DH1751	027
L13	027 Alexander	027
L12	51557	027
R20291	R20291	027
DH361	DH361	027
DH131	DH131	027
R20352	R20352	027
DH478	DH478	027
R12087	R12087	027
L15	T31694	027
DH482	DH482	027
R24694	R24694	027
DH326	DH326	027
DH183	DH183	027
DH1342	DH1342	027
DH1329	DH1329	027
DH85	DH85	027
R23970	R23970	027
O27		*027*
L1		*027*
R24988	R24988	*027*
L4	05-1223-046	027
L5	Libramont 1	027
R20298	R20298	027
NCTC 12727		027
R20291		027
CD196		027
210		027
CD2801		027
CD5101		027
CD9201		027
029	R8438	029
7322		031
V		045?
046	Subtype E6	046
053	CDC 34	053
056	Subtype S1	056
1351		060
VII	57267	063
070	R9367	070
075	IS93	075
077	----	077
L35	Wilcox 078	078
L34	31662	078

CRG Strain name	Real strain name	PCR-ribotype
EK26	2016	078
EK27	7004578	078
EK28	7009825	078
M120		078
EK23	Type 078	078
EK24	CD 2315	078
EK29	7045389	078
XVI	SUC36	078
M120		078
CD3801		078
CD4401		078
CD8601		078
CD9001		078
CD9301		078
CD9501		078
CD1801		078
III A	SE844	080
081	R9764	081
VPI10463		087
095	R8858	095
EK31		106
EK32		106
EK34		106
EK36		106
LIV22		106
R10432	R10432	106
R12801	R12801	106
R15347	R15347	106
R22079	R22079	106
R23942	R23942	106
R25469	R25469	106
R10459	R10459	106
R27384	R27384	106
106	R10459	106
R108095	R108095	106
117	R10071	117
126		126
131	R16080	131
E101		
B9		
B10		
B14		
B32		
B35		

CRG Strain name	Real strain name	PCR-ribotype
BZ47		
CS3		
CS8		
V23		
V43 54		
6443 REA type BI 17		

7.2 Phage genome annotations

phiL14WAR	Annotation
gp1	Terminase small subunit
gp2	Terminase large subunit
gp3	Putative portal protein
gp4	Minor capsid protein
gp5	Hypothetical protein
gp6	Hypothetical protein
gp7	Putative capsid scaffold protein
gp8	Putative capsid protein
gp9	Putative Rho-like termination factor protein
gp10	Putative head-tail connector protein
gp11	Hypothetical protein
gp12	HK97 gp10 family phage protein
gp13	Hypothetical protein
gp14	Putative tail-sheath (XkdK-like)
gp15	Putative tail core protein (XkdM-like)
gp16	XkdN-like protein
gp17	Hypothetical protein
gp18	Phage tail tape measure protein
gp19	Hypothetical protein
gp20	Hypothetical protein
gp21	LysM protein (XkdP-like)
gp22	Hypothetical protein
gp23	Putative cell wall hydrolase protein
gp24	Hypothetical protein (DUF2577 family)
gp25	XkdS-like protein (DUF2634 family)
gp26	Putative baseplate J protein (XkdT-like)
gp27	DUF2313 domain-containing protein (XkdU-like putative tail fibre)
gp28	Putative tail fiber protein
gp29	Hypothetical protein
gp30	Hypothetical protein
gp31	Hypothetical protein
gp32	Holin
gp33	N-acetylmuramoyl-L-alanine amidase
gp34	Hypothetical protein
gp35	Hypothetical protein
gp36	Hypothetical protein
gp37	ImmA/IrrE family metallo-endopeptidase (DUF955)
gp38	Hypothetical protein
gp39	Hypothetical protein
gp40	Hypothetical protein

gp41	Hypothetical protein
gp42	BlaI/MecI/CopY family transcriptional regulator
gp43	Helix-turn-helix transcriptional regulator
gp44	Helix-turn-helix transcriptional regulator (XREfamily)
gp45	Putative anti-repressor protein (ORF6N)
gp46	Helix-turn-helix domain-containing protein(putative repressor)
gp47	Hypothetical protein
gp48	Hypothetical protein
gp49	Hypothetical protein
gp50	Recombinase RecF (AAA family ATPase)
gp51	Recombination protein RecT
gp52	Metallo-beta-lactamase (MBL) fold metallo-hydrolase
gp53	Hypothetical protein (replication)
gp54	Hypothetical protein
gp55	Putative regulatory protein (NUMOD1 domain protein)
gp56	Hypothetical protein (YopX family)
gp57	Hypothetical protein
gp58	Hypothetical protein
gp59	Hypothetical protein
gp60	Hypothetical protein
gp61	Nucleoid-associated protein
gp62	Phosphoadenosine phosphosulfate (PAPS) reductasefamily protein
gp63	Hypothetical protein
gp64	Hypothetical protein
gp65	Hypothetical protein
gp66	DUF3850 domain-containing protein
gp67	RusA family crossover junction endodeoxyribonuclease
gp68	Putative antirepressor (ORF6 transcriptional regulator?)
gp69	Hypothetical protein
gp70	ParB N-terminal domain-containing protein

phiJS50S	**Annotation**
gp1	Hypothetical protein
gp2	Small terminase subunit
gp3	Putative terminase large subunit
gp4	Phage portal protein

gp5	Phage prohead protease (HK97 family)
gp6	Phage major capsid protein
gp7	Hypothetical protein
gp8	Putative phage head-tail adaptor
gp9	Hypothetical protein
gp10	Putative head-tail joining protein
gp11	Hypothetical protein
gp12	Putative phage major tail protein
gp13	Hypothetical protein
gp14	Hypothetical protein
gp15	Hypothetical protein
gp16	Phage tail tape measure protein
gp17	Putative phage tail protein
gp18	Phage tail protein (possible endopeptidase or glucosyltransferase)
gp19	Hypothetical protein
gp20	Putative tail fiber protein H (kelch repeat-containing protein)
gp21	Hypothetical protein
gp22	XkdX family protein
gp23	Hypothetical protein
gp24	Hypothetical protein
gp25	Hypothetical protein
gp26	Hypothetical protein
gp27	Putative holin
gp28	N-acetylmuramoyl-L-alanine amidase
gp29	Hypothetical protein
gp30	Hypothetical protein
gp31	Hypothetical protein
gp32	ParA family protein
gp33	Hypothetical protein
gp34	Hypothetical protein
gp35	Hypothetical protein
gp36	Hypothetical protein (possible membrane protein)
gp37	Hypothetical protein
gp38	Hypothetical protein
gp39	Hypothetical protein
gp40	Hypothetical protein
gp41	Hypothetical protein
gp42	Helix-turn-helix domain-containing protein
gp43	Hypothetical protein
gp44	Hypothetical protein
gp45	Hypothetical protein
gp46	Hypothetical protein

gp47	Hypothetical protein
gp48	Helix-turn-helix transcriptional regulator
gp49	Hypothetical protein
gp50	Hypothetical protein
gp51	Helix-turn-helix transcriptional regulator
gp52	Hypothetical protein
gp53	Hypothetical protein
gp54	Putative phage replication protein
gp55	Hypothetical protein
gp56	Hypothetical protein
gp57	DUF2513-containing protein
gp58	Hypothetical protein
gp59	Hypothetical protein
gp60	Hypothetical protein
gp61	Hypothetical protein
gp62	DUF3797 domain-containing protein
gp63	Putative single-strand binding protein
gp64	Hypothetical protein
gp65	Hypothetical protein (Interleukin-18-binding protein?)
gp66	Hypothetical protein
gp67	Hypothetical protein
gp68	Hypothetical protein
gp69	Hypothetical protein
gp70	Hypothetical protein
gp71	Site-specific integrase (tyrosine)
gp72	Hypothetical protein (Anti-restriction endonuclease?)
gp73	Putative HNH endonuclease

phiJS50W	**Annotation**
gp1	Hypothetical protein
gp2	Terminase small subunit
gp3	Putative terminase large subunit
gp4	Putative phage portal protein
gp5	HK97 family phage prohead protease
gp6	Phage major capsid protein A
gp7	Hypothetical protein
gp8	Putative phage head-tail adaptor
gp9	Hypothetical protein
gp10	Putative head-tail joining protein
gp11	Hypothetical protein

gp12	Putative phage major tail protein
gp13	Hypothetical protein
gp14	Hypothetical protein
gp15	Hypothetical protein
gp16	Phage tail tape measure protein
gp17	Phage tail family protein
gp18	Siphovirus ReqiPepy6 Gp37-like family protein(possible glucosyltransferase or endopeptidase)
gp19	Hypothetical protein (truncated putative phage structural protein)
gp20	Kelch repeat-containing protein
gp21	Hypothetical protein (possible tail fiber)
gp22	XkdX family protein
gp23	Hypothetical protein
gp24	Hypothetical protein
gp25	Hypothetical protein
gp26	Hypothetical protein
gp27	Putative holin
gp28	N-acetylmuramoyl-L-alanine amidase
gp29	Hypothetical protein
gp30	Hypothetical protein
gp31	Hypothetical protein
gp32	ParA-like protein
gp33	Hypothetical protein
gp34	Hypothetical protein
gp35	Hypothetical protein
gp36	Hypothetical protein (possible membrane protein)
gp37	Hypothetical protein (possible membrane protein)
gp38	Putative membrane protein
gp39	Hypothetical protein
gp40	Hypothetical protein (putative secreted protein)
gp41	Hypothetical protein
gp42	Helix-turn-helix domain-containing protein(possible transcriptional regulator)
gp43	Hypothetical protein
gp44	Hypothetical protein
gp45	Hypothetical protein
gp46	Hypothetical protein
gp47	Hypothetical protein
gp48	Helix-turn-helix transcriptional regulator(XRE-family - possible cI or Cro)

gp49	Hypothetical protein
gp50	Helix-turn-helix transcriptional regulator (possible XRE-family - cI or Cro)
gp51	Hypothetical protein
gp52	XRE transcriptional regulator (possible cI)
gp53	Hypothetical protein
gp54	Phage replication initiation protein (DnaD?)
gp55	Hypothetical protein
gp56	Hypothetical protein
gp57	Hypothetical protein (DUF2513 domain-containing protein)
gp58	Hypothetical protein
gp59	Hypothetical protein
gp60	Hypothetical protein
gp61	Hypothetical protein
gp62	DUF3797 domain-containing protein
gp63	Putative single-strand binding protein
gp64	Hypothetical protein
gp65	Hypothetical protein
gp66	Hypothetical protein
gp67	Hypothetical protein
gp68	Hypothetical protein
gp69	Hypothetical protein
gp70	Hypothetical protein (possible sigma factor)
gp71	Site-specific integrase (tyrosine)
gp72	Hypothetical protein (possible transglycosylase)
gp73	Putative HNH endonuclease

7.3 Phage infection curve data

7.3.1 Average $OD600_{nm}$ of L14 cultures infected with phiL14WAR

	Neg			MOI 0.01			MOI 0.1			MOI 1			MOI 10		
Time (hours)	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
0	0.36	0.31	0.37	0.32	0.29	0.33	0.31	0.29	0.33	0.38	0.31	0.35	0.37	0.31	0.35
1	0.75	0.62	0.78	0.68	0.59	0.68	0.67	0.56	0.72	0.73	0.59	0.70	0.70	0.58	0.67
2	1.27	1.09	1.26	1.14	1.02	1.13	1.05	0.95	1.10	0.77	0.64	0.74	0.58	0.48	0.62
3	1.88	1.72	1.89	1.79	1.61	1.71	1.46	1.24	1.40	0.42	0.31	0.44	0.37	0.27	0.47
4	2.16	2.12	2.20	1.96	1.85	1.93	0.78	0.51	0.74	0.15	0.12	0.13	0.15	0.10	0.15
5	2.21	2.19	2.26	1.90	1.81	1.87	0.52	0.36	0.56	0.12	0.08	0.10	0.13	0.08	0.12
6	2.19	2.13	2.25	1.85	1.74	1.82	0.38	0.23	0.40	0.11	0.07	0.08	0.13	0.08	0.11
7	2.16	2.10	2.23	1.79	1.64	1.76	0.28	0.16	0.31	0.11	0.07	0.08	0.13	0.08	0.11
8	2.16	2.13	2.22	1.72	1.53	1.67	0.21	0.12	0.25	0.12	0.08	0.09	0.15	0.11	0.13
9	2.08	2.03	2.16	1.60	1.40	1.54	0.18	0.11	0.27	0.13	0.11	0.10	0.16	0.12	0.14
10	2.08	2.03	2.14	1.52	1.29	1.44	0.16	0.10	0.27	0.14	0.11	0.12	0.17	0.13	0.16
11	2.03	1.99	2.11	1.42	1.19	1.33	0.15	0.10	0.22	0.15	0.12	0.13	0.18	0.14	0.17
12	2.01	1.96	2.07	1.35	1.11	1.25	0.14	0.10	0.17	0.16	0.13	0.14	0.18	0.14	0.18
13	1.93	1.89	2.01	1.27	1.05	1.17	0.14	0.10	0.14	0.17	0.14	0.15	0.19	0.16	0.18
14	1.91	1.87	1.97	1.21	0.99	1.11	0.14	0.10	0.14	0.17	0.15	0.15	0.20	0.16	0.19
15	1.81	1.77	1.87	1.14	0.95	1.06	0.14	0.10	0.14	0.18	0.15	0.16	0.20	0.17	0.19
16	1.72	1.69	1.77	1.08	0.90	1.01	0.14	0.11	0.14	0.19	0.16	0.17	0.21	0.18	0.20
17	1.60	1.58	1.65	1.03	0.87	0.96	0.15	0.12	0.15	0.19	0.17	0.17	0.21	0.18	0.20
18	1.51	1.49	1.55	0.98	0.83	0.92	0.15	0.13	0.15	0.19	0.17	0.17	0.22	0.19	0.20
19	1.42	1.39	1.45	0.94	0.79	0.88	0.15	0.13	0.16	0.20	0.17	0.17	0.22	0.19	0.21
20	1.32	1.29	1.36	0.89	0.76	0.84	0.16	0.14	0.16	0.20	0.17	0.18	0.22	0.19	0.21
21	1.24	1.22	1.28	0.85	0.73	0.80	0.16	0.14	0.17	0.21	0.17	0.18	0.23	0.19	0.21
22	1.18	1.16	1.22	0.80	0.70	0.77	0.17	0.14	0.17	0.21	0.17	0.18	0.23	0.19	0.21
23	1.12	1.10	1.16	0.77	0.67	0.74	0.17	0.15	0.17	0.21	0.17	0.18	0.23	0.19	0.22

7.3.2 Average $OD600_{nm}$ of CD196 cultures infected with phiL14WAR

	Neg			MOI 0.01			MOI 0.1			MOI1			MOI 10		
Time (hours)	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
0	0.53	0.52	0.53	0.44	0.47	0.51	0.45	0.40	0.64	0.42	0.46	0.48	0.46	0.46	0.45
1	1.09	0.99	1.05	0.91	0.91	1.02	1.00	0.92	0.97	0.90	0.91	0.94	0.89	0.93	0.91
2	1.67	1.51	1.54	1.46	1.44	1.52	1.50	1.39	1.43	1.06	1.07	1.11	0.93	0.96	1.12
3	2.06	1.88	1.96	1.93	1.90	1.91	1.79	1.71	1.72	1.22	1.27	1.33	0.98	0.97	1.21
4	2.15	2.05	2.09	1.99	1.98	2.02	1.89	1.78	1.77	0.49	0.54	0.74	0.51	0.52	1.00
5	2.19	2.05	2.18	2.01	2.00	2.05	1.73	1.52	1.56	0.15	0.16	0.24	0.12	0.16	0.61
6	2.20	2.09	2.12	2.01	1.99	2.04	1.56	1.39	1.40	0.10	0.12	0.16	0.08	0.11	0.54
7	2.18	2.03	2.11	1.99	1.96	2.00	1.44	1.29	1.30	0.09	0.11	0.14	0.08	0.09	0.50
8	2.11	2.04	2.06	1.95	1.88	1.91	1.40	1.22	1.23	0.07	0.08	0.10	0.06	0.08	0.46
9	2.11	1.99	2.08	1.90	1.87	1.88	1.32	1.17	1.17	0.08	0.11	0.12	0.08	0.08	0.46
10	2.05	1.95	1.96	1.87	1.84	1.86	1.27	1.10	1.12	0.08	0.11	0.12	0.08	0.08	0.45
11	2.01	1.88	1.97	1.82	1.78	1.80	1.20	1.03	1.04	0.09	0.11	0.12	0.08	0.08	0.42
12	1.95	1.88	1.93	1.77	1.72	1.76	1.14	0.98	0.98	0.09	0.12	0.12	0.08	0.09	0.41
13	1.90	1.78	1.87	1.71	1.67	1.71	1.05	0.91	0.92	0.10	0.12	0.13	0.09	0.09	0.39
14	1.84	1.73	1.79	1.66	1.62	1.65	1.00	0.86	0.87	0.11	0.13	0.13	0.10	0.09	0.37
15	1.79	1.68	1.76	1.61	1.56	1.60	0.94	0.82	0.82	0.13	0.16	0.15	0.11	0.10	0.36
16	1.74	1.64	1.66	1.56	1.50	1.56	0.90	0.78	0.78	0.16	0.19	0.18	0.12	0.11	0.35
17	1.68	1.55	1.66	1.50	1.44	1.50	0.86	0.73	0.74	0.18	0.22	0.21	0.13	0.12	0.34
18	1.65	1.50	1.56	1.45	1.39	1.45	0.81	0.69	0.70	0.20	0.24	0.24	0.13	0.12	0.34
19	1.62	1.46	1.53	1.41	1.34	1.40	0.77	0.66	0.67	0.21	0.25	0.25	0.15	0.12	0.34
20	1.59	1.41	1.50	1.37	1.29	1.36	0.74	0.64	0.65	0.23	0.26	0.27	0.16	0.14	0.34
21	1.56	1.38	1.46	1.32	1.25	1.31	0.70	0.61	0.62	0.25	0.27	0.28	0.16	0.14	0.33
22	1.54	1.36	1.42	1.28	1.20	1.27	0.67	0.59	0.60	0.26	0.28	0.29	0.17	0.14	0.32
23	1.51	1.30	1.38	1.23	1.15	1.23	0.65	0.57	0.58	0.27	0.29	0.29	0.17	0.15	0.31

7.3.3 Average $OD600_{nm}$ of R12087 cultures infected with phiL14WAR

	Neg			MOI 0.01			MOI 0.1			MOI1			MOI 10		
Time (hours)	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
0	0.51	0.47	0.57	0.50	0.43	0.53	0.47	0.44	0.48	0.50	0.50	0.53	0.47	0.45	0.46
1	1.19	1.02	1.11	1.11	0.98	1.08	1.08	1.01	1.08	1.09	0.97	1.08	1.02	0.95	1.01
2	1.76	1.59	1.68	1.71	1.58	1.74	1.63	1.52	1.82	1.24	1.11	1.26	1.04	0.91	1.21
3	1.99	1.88	1.97	1.94	1.87	2.01	1.81	1.86	2.26	1.48	1.38	1.54	1.12	0.97	1.33
4	1.94	1.88	1.98	1.92	1.84	2.00	1.69	1.70	2.38	1.10	0.55	1.56	0.95	0.31	1.34
5	1.90	1.80	1.95	1.85	1.74	1.97	1.53	1.46	2.46	0.97	0.24	1.61	0.93	0.11	1.37
6	1.84	1.77	1.92	1.80	1.67	1.93	1.47	1.29	2.46	0.98	0.17	1.67	1.04	0.07	1.45
7	1.78	1.70	1.89	1.74	1.61	1.92	1.42	1.16	2.42	1.04	0.15	1.65	1.15	0.06	1.53
8	1.71	1.59	1.86	1.65	1.52	1.88	1.39	1.05	2.34	1.10	0.12	1.64	1.22	0.04	1.58
9	1.66	1.53	1.85	1.67	1.50	1.89	1.37	1.01	2.08	1.16	0.12	1.62	1.33	0.06	1.65
10	1.64	1.50	1.85	1.67	1.48	1.90	1.34	0.94	1.88	1.22	0.11	1.62	1.39	0.06	1.69
11	1.62	1.47	1.82	1.66	1.46	1.89	1.31	0.87	1.73	1.26	0.11	1.63	1.43	0.06	1.70
12	1.60	1.44	1.81	1.65	1.44	1.88	1.27	0.82	1.65	1.30	0.11	1.64	1.47	0.06	1.70
13	1.57	1.43	1.79	1.63	1.43	1.85	1.25	0.75	1.61	1.32	0.11	1.64	1.48	0.06	1.70
14	1.58	1.42	1.77	1.61	1.40	1.83	1.22	0.69	1.57	1.33	0.11	1.63	1.49	0.06	1.69
15	1.56	1.40	1.75	1.58	1.36	1.80	1.20	0.64	1.55	1.34	0.11	1.62	1.49	0.06	1.67
16	1.55	1.38	1.72	1.54	1.31	1.76	1.19	0.60	1.52	1.33	0.12	1.61	1.48	0.07	1.65
17	1.53	1.33	1.69	1.52	1.25	1.73	1.18	0.56	1.49	1.33	0.13	1.59	1.47	0.07	1.63
18	1.51	1.29	1.67	1.48	1.19	1.69	1.17	0.53	1.47	1.32	0.14	1.57	1.47	0.08	1.61
19	1.47	1.25	1.65	1.45	1.13	1.66	1.15	0.51	1.44	1.32	0.15	1.55	1.45	0.08	1.59
20	1.44	1.19	1.62	1.42	1.07	1.62	1.14	0.49	1.41	1.31	0.16	1.53	1.44	0.09	1.57
21	1.42	1.14	1.58	1.38	1.02	1.59	1.13	0.47	1.39	1.30	0.16	1.52	1.43	0.10	1.55
22	1.40	1.10	1.55	1.35	0.98	1.55	1.12	0.45	1.36	1.29	0.17	1.51	1.41	0.10	1.54
23	1.37	1.05	1.52	1.33	0.93	1.52	1.10	0.44	1.33	1.28	0.17	1.49	1.40	0.11	1.53

7.3.4 Average $OD600_{nm}$ of L28 cultures infected with phiL14WAR

	Neg			MOI 0.01			MOI 0.1			MOI1			MOI 10		
Time (hours)	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
0	0.44	0.62	0.54	0.39	0.58	0.49	0.39	0.57	0.47	0.41	0.55	0.47	0.38	0.52	0.45
1	0.90	1.23	1.10	0.94	1.37	1.23	0.88	1.27	1.16	0.88	1.21	1.10	0.83	1.12	1.04
2	1.69	1.97	1.85	1.72	2.10	1.97	1.66	2.01	1.90	1.48	1.87	1.72	1.40	1.84	1.67
3	1.97	2.11	2.05	2.14	2.24	2.11	1.96	2.15	2.13	1.77	2.04	1.99	1.67	2.02	1.99
4	2.19	2.07	2.05	2.27	2.12	2.11	2.14	2.12	2.14	1.73	2.01	2.03	1.70	2.05	2.03
5	2.12	1.97	1.97	2.17	1.99	1.94	2.08	2.03	2.04	1.24	1.95	1.92	1.33	1.98	2.01
6	2.06	1.77	1.81	2.06	1.84	1.82	1.99	1.91	1.94	1.00	1.83	1.78	1.07	1.93	1.86
7	1.94	1.67	1.71	1.95	1.66	1.68	1.92	1.82	1.88	0.86	1.78	1.70	0.94	1.87	1.79
8	1.88	1.50	1.57	1.90	1.51	1.56	1.89	1.74	1.82	0.74	1.66	1.59	0.81	1.79	1.70
9	1.85	1.36	1.45	1.86	1.42	1.45	1.83	1.67	1.74	0.63	1.54	1.47	0.69	1.70	1.59
10	1.78	1.31	1.35	1.79	1.30	1.37	1.74	1.60	1.68	0.55	1.42	1.40	0.60	1.63	1.50
11	1.76	1.29	1.31	1.79	1.27	1.31	1.69	1.55	1.65	0.48	1.32	1.33	0.52	1.54	1.40
12	1.73	1.30	1.33	1.77	1.24	1.27	1.63	1.55	1.62	0.43	1.23	1.26	0.45	1.45	1.30
13	1.71	1.29	1.35	1.73	1.23	1.29	1.54	1.55	1.60	0.39	1.14	1.19	0.40	1.35	1.20
14	1.68	1.33	1.36	1.70	1.25	1.29	1.45	1.55	1.58	0.35	1.04	1.12	0.36	1.25	1.10
15	1.67	1.34	1.35	1.66	1.28	1.31	1.37	1.56	1.58	0.32	0.95	1.06	0.32	1.14	1.01
16	1.64	1.31	1.37	1.62	1.29	1.34	1.29	1.55	1.55	0.30	0.87	0.99	0.29	1.05	0.92
17	1.61	1.30	1.40	1.57	1.31	1.35	1.21	1.52	1.51	0.28	0.80	0.93	0.27	0.95	0.83
18	1.57	1.30	1.44	1.52	1.33	1.35	1.12	1.49	1.46	0.28	0.74	0.88	0.26	0.85	0.75
19	1.52	1.30	1.46	1.46	1.34	1.35	1.05	1.44	1.39	0.27	0.67	0.82	0.26	0.77	0.68
20	1.48	1.30	1.46	1.40	1.33	1.33	0.98	1.39	1.31	0.27	0.61	0.77	0.25	0.69	0.61
21	1.44	1.31	1.47	1.34	1.31	1.31	0.91	1.33	1.24	0.26	0.57	0.71	0.25	0.62	0.55
22	1.38	1.29	1.47	1.28	1.28	1.28	0.85	1.26	1.16	0.26	0.52	0.65	0.25	0.56	0.49
23	1.34	1.27	1.44	1.22	1.25	1.24	0.79	1.20	1.10	0.26	0.48	0.60	0.25	0.51	0.45

7.3.5 Average $OD600_{nm}$ of CD8601 cultures infected with phiL14WAR

	Negative			MOI 0.01			MOI 0.1			MOI 1			MOI 10		
Time (hours)	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
0	0.35	0.36	0.51	0.24	0.33	0.38	0.27	0.35	0.36	0.28	0.34	0.37	0.28	0.33	0.40
1	0.84	1.04	1.01	0.71	1.01	1.02	0.76	0.93	1.07	0.80	0.89	0.89	0.71	0.94	0.98
2	1.54	1.65	1.63	1.39	1.74	1.85	1.40	1.62	1.79	1.33	1.60	1.57	1.19	1.54	1.52
3	1.89	1.87	1.99	1.78	2.07	2.09	1.80	2.00	2.07	1.80	2.00	1.94	1.62	1.93	1.94
4	2.06	2.07	2.07	1.97	2.17	2.19	2.00	2.10	2.20	2.03	2.18	2.09	1.92	2.12	2.11
5	2.16	2.08	2.13	2.07	2.17	2.17	2.09	2.16	2.20	2.13	2.26	2.15	2.04	2.19	2.14
6	2.18	2.09	2.14	2.10	2.19	2.15	2.14	2.14	2.21	2.17	2.22	2.15	2.09	2.20	2.17
7	2.21	2.13	2.14	2.12	2.16	2.14	2.16	2.15	2.18	2.21	2.22	2.14	2.14	2.19	2.15
8	2.23	2.13	2.12	2.14	2.15	2.14	2.18	2.13	2.19	2.22	2.22	2.14	2.14	2.19	2.17
9	2.21	2.11	2.11	2.14	2.13	2.11	2.17	2.11	2.16	2.21	2.21	2.14	2.14	2.17	2.14
10	2.20	2.11	2.11	2.15	2.12	2.10	2.17	2.09	2.14	2.21	2.21	2.15	2.13	2.15	2.14
11	2.19	2.09	2.09	2.12	2.11	2.09	2.15	2.08	2.12	2.20	2.20	2.14	2.12	2.14	2.12
12	2.17	2.07	2.08	2.10	2.08	2.07	2.14	2.06	2.10	2.18	2.18	2.13	2.09	2.12	2.10
13	2.16	2.06	2.06	2.08	2.05	2.05	2.11	2.02	2.08	2.15	2.16	2.11	2.05	2.11	2.09
14	2.15	2.04	2.04	2.06	2.02	2.02	2.09	2.01	2.05	2.12	2.13	2.08	2.02	2.07	2.05
15	2.11	2.01	2.02	2.02	1.99	1.99	2.05	1.97	2.02	2.09	2.11	2.06	2.00	2.05	2.04
16	2.09	1.98	1.99	2.01	1.95	1.96	2.03	1.95	1.99	2.07	2.09	2.03	1.96	2.02	2.01
17	2.05	1.94	1.96	1.96	1.92	1.93	1.99	1.90	1.95	2.03	2.05	2.00	1.93	1.99	1.99
18	2.02	1.91	1.93	1.94	1.89	1.90	1.96	1.88	1.92	2.00	2.02	1.97	1.91	1.96	1.97
19	1.98	1.87	1.89	1.91	1.85	1.86	1.93	1.84	1.89	1.97	1.99	1.94	1.88	1.93	1.93
20	1.95	1.84	1.86	1.88	1.82	1.83	1.90	1.81	1.86	1.94	1.95	1.91	1.86	1.89	1.90
21	1.92	1.81	1.83	1.85	1.80	1.81	1.87	1.79	1.83	1.91	1.93	1.88	1.83	1.87	1.88
22	1.88	1.77	1.79	1.83	1.77	1.78	1.84	1.76	1.80	1.88	1.89	1.85	1.80	1.84	1.85
23	1.86	1.74	1.77	1.80	1.74	1.76	1.82	1.74	1.77	1.86	1.87	1.82	1.78	1.81	1.82

7.3.6 Average $OD600_{nm}$ of CD2301 cultures infected with phiCD2301

	Neg			MOI 0.01			MOI 0.1			MOI 1			MOI 10		
Time (hours)	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
0	0.30	0.30	0.32	0.23	0.19	0.25	0.25	0.26	0.26	0.25	0.24	0.23	0.31	0.27	0.30
1	0.72	0.63	0.69	0.71	0.63	0.63	0.70	0.68	0.64	0.72	0.61	0.57	0.65	0.55	0.57
2	1.31	1.15	1.24	1.28	1.24	1.23	1.28	1.23	1.23	1.11	1.01	1.04	0.78	0.65	0.69
3	1.89	1.77	1.79	1.79	1.78	1.74	1.53	1.41	1.45	0.91	0.79	0.86	0.59	0.50	0.59
4	2.11	2.07	2.04	1.99	1.92	1.88	1.45	1.27	1.28	0.79	0.65	0.73	0.58	0.46	0.57
5	2.12	2.12	2.05	2.00	1.93	1.83	1.25	1.04	1.08	0.71	0.56	0.60	0.46	0.39	0.52
6	2.15	2.11	2.07	1.98	1.88	1.76	1.15	0.96	0.98	0.65	0.51	0.53	0.46	0.39	0.52
7	2.05	2.04	1.99	1.85	1.76	1.63	1.08	0.86	0.89	0.59	0.47	0.51	0.48	0.43	0.53
8	2.01	2.00	1.94	1.78	1.66	1.55	1.02	0.77	0.82	0.56	0.46	0.50	0.51	0.49	0.56
9	1.95	1.94	1.86	1.70	1.61	1.49	0.94	0.73	0.78	0.55	0.48	0.51	0.55	0.59	0.61
10	1.91	1.94	1.79	1.65	1.54	1.42	0.92	0.69	0.76	0.56	0.54	0.51	0.63	0.75	0.69
11	1.87	1.88	1.72	1.60	1.47	1.35	0.90	0.66	0.73	0.56	0.58	0.62	0.70	0.73	0.75
12	1.85	1.82	1.68	1.56	1.40	1.31	0.88	0.65	0.73	0.56	0.57	0.61	0.71	0.76	0.83
13	1.75	1.76	1.64	1.52	1.36	1.28	0.85	0.64	0.72	0.57	0.58	0.61	0.76	0.82	0.85
14	1.69	1.77	1.62	1.49	1.31	1.27	0.83	0.65	0.73	0.57	0.60	0.62	0.78	0.87	0.89
15	1.64	1.67	1.57	1.45	1.29	1.25	0.83	0.65	0.73	0.59	0.62	0.63	0.84	0.90	0.90
16	1.57	1.63	1.55	1.39	1.26	1.23	0.82	0.67	0.74	0.60	0.66	0.64	0.85	0.93	0.94
17	1.50	1.55	1.54	1.36	1.23	1.20	0.81	0.67	0.73	0.61	0.68	0.77	0.86	0.94	1.00
18	1.44	1.49	1.49	1.33	1.20	1.18	0.80	0.67	0.73	0.61	0.69	0.71	0.88	0.97	0.97
19	1.39	1.45	1.45	1.28	1.18	1.15	0.79	0.67	0.74	0.62	0.73	0.71	0.90	0.99	0.98
20	1.36	1.42	1.41	1.25	1.15	1.14	0.79	0.65	0.73	0.64	0.75	0.76	0.91	1.00	1.00
21	1.31	1.38	1.37	1.22	1.13	1.12	0.79	0.65	0.74	0.63	0.76	0.76	0.92	1.00	1.00
22	1.29	1.34	1.34	1.19	1.11	1.10	0.79	0.65	0.74	0.64	0.77	0.72	0.93	1.00	1.01
23	1.25	1.31	1.30	1.17	1.10	1.09	0.78	0.66	0.75	0.64	0.78	0.72	0.92	1.01	1.02

7.3.7 Virulence index results

7.3.7.1 Limited local virulence

	Local virulence					
MOI	**L14**	**CD196**	**R12087**	**L28**	**CD8601**	**CD2301**
0.01	0.103	0.054	0.057	-0.053	-0.009	0.047
0.1	0.427	0.125	0.057	-0.011	0.001	0.227
1	0.704	0.478	0.254	0.061	0.024	0.473
10	0.730	0.498	0.283	0.089	0.052	0.604

7.3.7.2 Unlimited local virulence

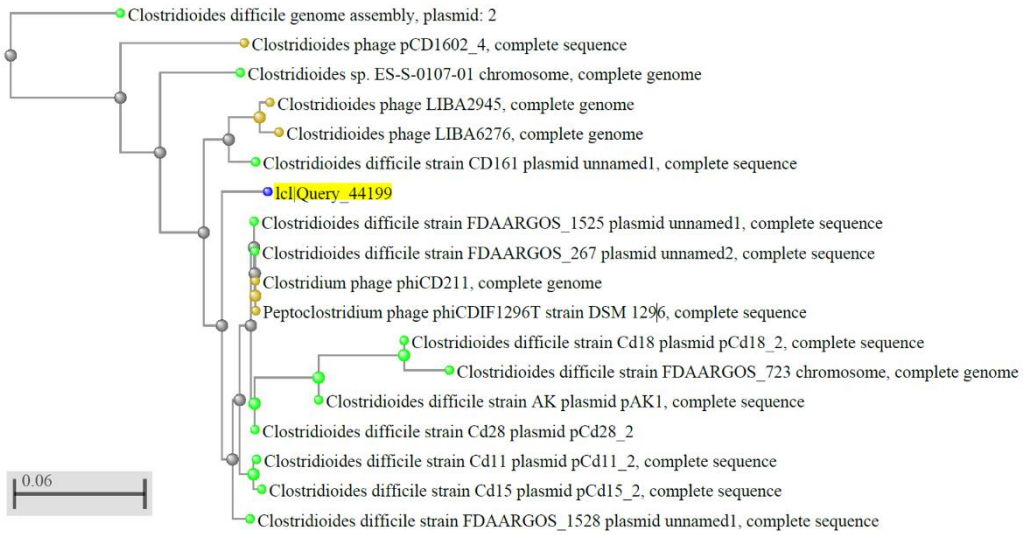
	Local virulence					
MOI	**L14**	**CD196**	**R12087**	**L28**	**CD8601**	**CD2301**
0.01	0.293	0.080	0.003	0.006	0.011	0.139
0.1	0.815	0.393	0.185	-0.006	0.007	0.475
1	0.878	0.821	0.720	0.323	-0.014	0.603
10	0.872	0.808	0.700	0.303	0.011	0.547

7.4 Whole genome sequencing of CD2301L – non-genomic contig BLAST results (top ten results only)

7.4.1 BLAST results

Description	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Clostridioides difficile strain FDAARGOS_1528 plasmid unnamed1, complete sequence	1.44E+05	70%	0	97.48	127406	CP083617.1
Clostridioides difficile strain FDAARGOS_267 plasmid unnamed2, complete sequence	1.25E+05	61%	0	96.4	131326	CP020426.2
Clostridioides difficile strain AK plasmid pAK1, complete sequence	1.27E+05	61%	0	96.4	142753	CP027015.1
Clostridium phage phiCD211, complete genome	1.25E+05	61%	0	96.4	131704	LN681537.2
Clostridioides difficile strain FDAARGOS_1525 plasmid unnamed1, complete sequence	1.25E+05	61%	0	96.4	131326	CP084370.1
Clostridioides difficile strain Cd28 plasmid pCd28_2	1.19E+05	54%	0	96.36	150276	CP037799.1
Clostridioides difficile strain Cd18 plasmid pCd18_2, complete sequence	1.12E+05	56%	0	96.32	136294	CP037807.1
Clostridioides difficile strain Cd11 plasmid pCd11_2, complete sequence	1.56E+05	68%	0	96.2	132448	CP037823.1
Clostridioides difficile strain Cd15 plasmid pCd15_2, complete sequence	1.55E+05	68%	0	96.35	129127	CP037814.1
Clostridioides difficile strain CD161 plasmid unnamed1, complete sequence	1.12E+05	57%	0	96.81	130529	CP029155.1

7.4.2 Phylogenetic tree (generated by NCBI BLAST)



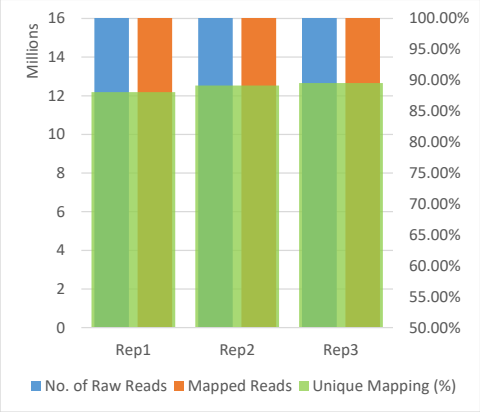
7.5 RNA sequencing results from DeepSeq

7.5.1 Quality assessment

Deep Seq sample number	Customer sample name	Concentration (ng/ul)	Volume (ul)	RIN	TapeStation conc (ng/ul)	Amount for 1 ug	Amount H2O	PCR cycles	Index	Library Qubit	TS av size	nM	Amount for 4 nM	Amount H2O
ds1058_1	Rep_1	~600	10	8.3	664	1.5	11	7	A4	8.86	386	34.8	8	61.6
ds1058_2	Rep_2	~600	10	8.3	646	1.6	10.9	7	B4	9.36	387	36.6	8	65.3
ds1058_3	Rep_3	~600	10	8.5	508	2	10.5	7	C4	7.98	380	31.8	8	55.6

7.5.2 Mapping summary

Sample	No. of Raw Reads	Mapped Reads		Unique Mapping (%)	
		Counts	Percentage	Counts	Percentage
Rep1	19,613,968	19,345,257	98.63%	17274270	88.07%
Rep2	18,111,254	17,877,619	98.71%	16145785	89.15%
Rep3	16,462,062	16,269,456	98.83%	14745144	89.57%



7.5.3 Transcripts per million (TPM) of phage-related genes only

GeneI	GeneName	Rep-1_S1	Rep-2_S2	Rep-3_S3
1323	N-acetylmuramoyl-L-alanine amidase	20.26	26.15	26.50
1324	Phage holin	5.85	7.64	8.65
1325	Phage protein	17.33	21.33	22.31
1326	hypothetical protein	8.44	8.64	11.70
1327	Phage capsid and scaffold	20.56	20.83	22.33
1328	Phage tail fiber protein	26.48	30.90	30.09
1329	Phage tail fibers	11.02	12.72	11.74
1330	Phage-like element pbsx protein xkdT	6.83	8.47	8.01
1331	Phage baseplate	7.71	8.60	6.63
1332	Phage-like element PBSX protein xkdS	3.82	3.61	4.66
1333	Phage protein	9.23	6.67	9.18
1334	putative phage cell wall hydrolase	9.21	10.71	7.96
1335	Phage-like element PBSX protein xkdP	9.35	13.33	10.72
1336	Phage tail tape measure protein	12.12	11.59	13.28
1337	hypothetical protein	6.41	4.85	3.98
1338	hypothetical protein	8.37	9.17	7.01
1339	hypothetical protein	11.96	9.70	10.09
1340	hypothetical protein	17.64	20.50	19.88
1341	hypothetical protein	17.94	19.58	19.19
1342	hypothetical protein	14.11	12.41	12.08
1343	hypothetical protein	17.04	20.35	17.25
1344	hypothetical protein	8.66	10.26	5.61
1345	hypothetical protein	15.36	14.53	14.19
1346	Prophage Clp protease-like protein	12.72	12.59	14.07
1347	Phage portal protein	25.28	24.80	28.24
1348	Phage terminase, large subunit	24.14	23.72	21.49
1349	hypothetical protein	9.24	6.26	3.99
1350	hypothetical protein	27.90	24.90	27.74
1351	hypothetical protein	17.59	20.32	21.68
1352	Phage integrase	27.08	28.77	24.27
1353	hypothetical protein	0.98	0.90	0.99
1354	hypothetical protein	2.51	1.35	0.29
1355	hypothetical protein	1.02	0.00	0.00
1356	hypothetical protein	1.79	1.92	0.63
1357	hypothetical protein	0.00	0.81	0.00
1358	hypothetical protein	4.62	6.11	1.81
1359	hypothetical protein	4.22	4.53	1.35
1360	hypothetical protein	4.41	5.64	3.18
1361	hypothetical protein	3.45	4.32	0.00
1362	hypothetical protein	79.24	79.25	77.63
1363	DNA polymerase III polC-type (EC 2.7.7.7)	375.97	406.27	408.18
1364	hypothetical protein	2580.05	2799.89	1574.36
1365	hypothetical protein	3.52	2.51	2.75
1366	hypothetical protein	2.55	4.69	2.99
1367	hypothetical protein	85.26	94.28	88.68
1368	hypothetical protein	36.46	40.47	38.89

Chapter 8

References

8 References

Abedon, S.T., 2012. Bacterial ‘immunity’ against bacteriophages. Bacteriophage 2, 50–54. https://doi.org/10.4161/bact.18609

Abedon, S.T., 2011. Facilitation of CRISPR adaptation. Bacteriophage 1, 179–181. https://doi.org/10.4161/bact.1.3.16709

Ackermann, H.W., 2011. Bacteriophage taxonomy. Microbiol. Aust. 32, 90. https://doi.org/10.1071/MA11090

Ackermann, H.W., Eisenstark, A., 1974. The present state of phage taxonomy. Intervirology 3, 201–219. https://doi.org/10.1159/000149758

Adriaenssens, E.M., Sullivan, M.B., Knezevic, P., van Zyl, L.J., Sarkar, B.L., Dutilh, B.E., Alfenas-Zerbini, P., Łobocka, M., Tong, Y., Brister, J.R., Moreno Switt, A.I., Klumpp, J., Aziz, R.K., Barylski, J., Uchiyama, J., Edwards, R.A., Kropinski, A.M., Petty, N.K., Clokie, M.R.J., Kushkina, A.I., Morozova, V.V., Duffy, S., Gillis, A., Rumnieks, J., Kurtböke, İ., Chanishvili, N., Goodridge, L., Wittmann, J., Lavigne, R., Jang, H.B., Prangishvili, D., Enault, F., Turner, D., Poranen, M.M., Oksanen, H.M., Krupovic, M., 2020. Taxonomy of prokaryotic viruses: 2018-2019update from the ICTV Bacterial and Archaeal Viruses Subcommittee. Arch Virol 165, 1253–1260. https://doi.org/10.1007/s00705-020-04577-8

Aksyuk, A.A., Rossmann, M.G., 2011. Bacteriophage Assembly. Viruses 3, 172–203. https://doi.org/10.3390/v3030172

Aktories, K., Papatheodorou, P., Schwan, C., 2018. Binary Clostridium difficile toxin (CDT) - A virulence factor disturbing the cytoskeleton. Anaerobe, ClostPath 2017: 10th International Conference on the Molecular Biology and Pathogenesis of the Clostridia 53, 21–29. https://doi.org/10.1016/j.anaerobe.2018.03.001

Al Saif, N., Brazier, J.S.Y. 1996, 1996. The distribution of *Clostridium difficile* in the environment of South Wales. Journal of Medical Microbiology 45, 133–137. https://doi.org/10.1099/00222615-45-2-133

Alauzet, C., Lozniewski, A., Marchandin, H., 2019. Metronidazole resistance and nim genes in anaerobes: A review. Anaerobe 55, 40–53. https://doi.org/10.1016/j.anaerobe.2018.10.004

Alcalá Hernández, L., Reigadas Ramírez, E., Bouza Santiago, E., 2017. *Clostridium difficile* infection. Medicina Clínica (English Edition) 148, 456–463. https://doi.org/10.1016/j.medcle.2017.04.034

Allen, H.K., 2017. Alternatives to Antibiotics: Why and How. NAM Perspectives. https://doi.org/10.31478/201707g

Anderson, L.M., Yang, H., 2008. DNA looping can enhance lysogenic CI transcription in phage lambda. PNAS 105, 5827–5832. https://doi.org/10.1073/pnas.0705570105

Anderson, R.J., Weng, Z., Campbell, R.K., Jiang, X., 2005. Main-chain conformational tendencies of amino acids. Proteins: Structure, Function, and Bioinformatics 60, 679–689. https://doi.org/10.1002/prot.20530

Andino-Molina, M., Barquero-Calvo, E., Seyboldt, C., Schmoock, G., Neubauer, H., Tzoc, E., Rodríguez, C., Quesada-Gómez, C., 2019. Multidrug-resistant *Clostridium difficile* ribotypes 078 and 014/5-FLI01 in piglets from Costa Rica. Anaerobe 55, 78–82. https://doi.org/10.1016/j.anaerobe.2018.11.004

Androga, G.O., Hart, J., Foster, N.F., Charles, A., Forbes, D., Riley, T.V., 2015. Infection with Toxin A-Negative, Toxin B-Negative, Binary Toxin-Positive *Clostridium difficile* in a Young Patient with Ulcerative Colitis. Journal of Clinical Microbiology 53, 3702–3704. https://doi.org/10.1128/JCM.01810-15

Androga, G.O., Knight, D.R., Lim, S.-C., Foster, N.F., Riley, T.V., 2018. Antimicrobial resistance in large clostridial toxin-negative, binary toxin-positive *Clostridium difficile* ribotypes. Anaerobe 54, 55–60. https://doi.org/10.1016/j.anaerobe.2018.07.007

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., Iyer, L.M., 2005. The many faces of the helix-turn-helix domain: Transcription regulation and beyond⋆. FEMS Microbiology Reviews 29, 231–262. https://doi.org/10.1016/j.fmrre.2004.12.008

Aslam, S., Hamill, R.J., Musher, D.M., 2005. Treatment of *Clostridium difficile*-associated disease: old therapies and new strategies. The Lancet Infectious Diseases 5, 549–557. https://doi.org/10.1016/S1473-3099(05)70215-2

Aurell, E., Brown, S., Johanson, J., Sneppen, K., 2002. Stability puzzles in phage lambda. Phys. Rev. E 65, 051914. https://doi.org/10.1103/PhysRevE.65.051914

Ayora, S., Missich, R., Mesa, P., Lurz, R., Yang, S., Egelman, E.H., Alonso, J.C., 2002. Homologous-pairing Activity of the *Bacillus subtilis* Bacteriophage SPP1 Replication Protein G35P. J. Biol. Chem. 277, 35969–35979. https://doi.org/10.1074/jbc.M204467200

Badger, V.O., Ledeboer, N.A., Graham, M.B., Edmiston, C.E., 2012a. *Clostridium difficile*. Journal of Parenteral and Enteral Nutrition 36, 645–662. https://doi.org/10.1177/0148607112446703

Badger, V.O., Ledeboer, N.A., Graham, M.B., Edmiston Jr, C.E., 2012b. *Clostridium difficile*: epidemiology, pathogenesis, management, and prevention of a recalcitrant healthcare-associated pathogen. Journal of parenteral and enteral nutrition 36, 645–662.

Barbut, F., Rupnik, M., 2012. Editorial Commentary: 027, 078, and Others: Going Beyond the Numbers (and Away From the Hypervirulence). Clinical Infectious Diseases 55, 1669–1672. https://doi.org/10.1093/cid/cis790

Barquist, L., Mayho, M., Cummins, C., Cain, A.K., Boinett, C.J., Page, A.J., Langridge, G.C., Quail, M.A., Keane, J.A., Parkhill, J., 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant

libraries. Bioinformatics 32, 1109–1111. https://doi.org/10.1093/bioinformatics/btw022

Barron, M., 2022. Phage Therapy: Past, Present and Future. ASM.org.

Bartsch, S.M., Umscheid, C.A., Fishman, N., Lee, B.Y., 2013. Is Fidaxomicin Worth the Cost? An Economic Analysis. Clin Infect Dis 57, 555–561. https://doi.org/10.1093/cid/cit346

Baunwall, S.M.D., Lee, M.M., Eriksen, M.K., Mullish, B.H., Marchesi, J.R., Dahlerup, J.F., Hvas, C.L., 2020. Faecal microbiota transplantation for recurrent *Clostridioides difficile* infection: An updated systematic review and meta-analysis. eClinicalMedicine 29. https://doi.org/10.1016/j.eclinm.2020.100642

Bell, C.E., Frescura, P., Hochschild, A., Lewis, M., 2000. Crystal Structure of the λ Repressor C-Terminal Domain Provides a Model for Cooperative Operator Binding. Cell 101, 801–811. https://doi.org/10.1016/S0092-8674(00)80891-0

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Research 28, 235–242. https://doi.org/10.1093/nar/28.1.235

Bertani, G., 2004. Lysogeny at Mid-Twentieth Century: P1, P2, and Other Experimental Systems. Journal of Bacteriology 186, 595–600. https://doi.org/10.1128/JB.186.3.595-600.2004

Bertozzi Silva, J., Storms, Z., Sauvageau, D., 2016. Host receptors for bacteriophage adsorption. FEMS Microbiol. Lett. 363. https://doi.org/10.1093/femsle/fnw002

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29, 2607–2618.

Bhattacharjee, R., Nandi, A., Sinha, A., Kumar, H., Mitra, D., Mojumdar, A., Patel, P., Jha, E., Mishra, S., Rout, P.K., Panda, P.K., Suar, M., Verma, S.K., 2022. Phage-tail-like bacteriocins as a biomedical platform to counter anti-microbial resistant pathogens. Biomedicine & Pharmacotherapy 155, 113720. https://doi.org/10.1016/j.biopha.2022.113720

Bibb, L.A., Hancox, M.I., Hatfull, G.F., 2005. Integration and excision by the large serine recombinase φRv1 integrase. Molecular Microbiology 55, 1896–1910. https://doi.org/10.1111/j.1365-2958.2005.04517.x

Biswas, B., Adhya, S., Washart, P., Paul, B., Trostel, A.N., Powell, B., Carlton, R., Merril, C.R., 2002. Bacteriophage therapy rescues mice bacteremic from a clinical isolate of vancomycin-resistant *Enterococcus faecium.* Infect. Immun. 70, 204–210. https://doi.org/10.1128/iai.70.1.204-210.2002

Boneca, I.G., Chiosis, G., 2005. Vancomycin resistance: occurrence, mechanisms and strategies to combat it. Expert Opinion on Therapeutic Targets.

Borodovich, T., Shkoporov, A.N., Ross, R.P., Hill, C., 2022. Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. Gastroenterol Rep (Oxf) 10, goac012. https://doi.org/10.1093/gastro/goac012

Bose, B., Auchtung, J.M., Lee, C.A., Grossman, A.D., 2008. A conserved anti-repressor controls horizontal gene transfer by proteolysis. Molecular Microbiology 70, 570–582. https://doi.org/10.1111/j.1365-2958.2008.06414.x

Bradley, D.E., 1967. Ultrastructure of bacteriophage and bacteriocins. Bacteriol Rev 31, 230–314.

Brady, A., Felipe-Ruiz, A., Gallego del Sol, F., Marina, A., Quiles-Puchalt, N., Penadés, J.R., 2021a. Molecular Basis of Lysis–Lysogeny Decisions

in Gram-Positive Phages. Annu. Rev. Microbiol. 75, 563–581. https://doi.org/10.1146/annurev-micro-033121-020757

Brady, A., Quiles-Puchalt, N., Gallego del Sol, F., Zamora-Caballero, S., Felipe-Ruíz, A., Val-Calvo, J., Meijer, W.J.J., Marina, A., Penadés, J.R., 2021b. The arbitrium system controls prophage induction. Current Biology 31, 5037-5045.e3. https://doi.org/10.1016/j.cub.2021.08.072

Brandt, L.J., Kosche, K.A., Greenwald, D.A., Berkman, D., 1999. *Clostridium difficile*-associated diarrhea in the elderly. The American Journal of Gastroenterology 94, 3263–3266. https://doi.org/10.1016/S0002-9270(99)00593-6

Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A., Stevens, R., Vonstein, V., Wattam, A.R., Xia, F., 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5, 8365. https://doi.org/10.1038/srep08365

Bronzwaer, S.L.A.M., Cars, O., Buchholz, U., Mölstad, S., Goettsch, W., Veldhuijzen, I.K., Kool, J.L., Sprenger, M.J.W., Degener, J.E., 2002. The Relationship between Antimicrobial Use and Antimicrobial Resistance in Europe. Emerg Infect Dis 8, 278–282. https://doi.org/10.3201/eid0803.010192

Brüssow, H., 2012. What is needed for phage therapy to become a reality in Western medicine? Virology, Special issue: viruses of microbes 434, 138–142. https://doi.org/10.1016/j.virol.2012.09.015

Buckley, A.M., Moura, I.B., Wilcox, M.H., 2022. The potential of microbiome replacement therapies for *Clostridium difficile* infection. Current Opinion in Gastroenterology 38, 1. https://doi.org/10.1097/MOG.0000000000000800

Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G.V., Christie, C.H., Dalenberg, K., Di Costanzo, L., Duarte, J.M., Dutta, S.,

Feng, Z., Ganesan, S., Goodsell, D.S., Ghosh, S., Green, R.K., Guranović, V., Guzenko, D., Hudson, B.P., Lawson, C.L., Liang, Y., Lowe, R., Namkoong, H., Peisach, E., Persikova, I., Randle, C., Rose, A., Rose, Y., Sali, A., Segura, J., Sekharan, M., Shao, C., Tao, Y.-P., Voigt, M., Westbrook, J.D., Young, J.Y., Zardecki, C., Zhuravleva, M., 2020. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Res 49, D437–D451. https://doi.org/10.1093/nar/gkaa1038

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 113, 10174–10179. https://doi.org/10.1073/pnas.1607966113

Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., O'Toole, G.A., 2012. The CRISPR/Cas Adaptive Immune System of *Pseudomonas aeruginosa* Mediates Resistance to Naturally Occurring and Engineered Phages. Journal of Bacteriology 194, 5728–5738. https://doi.org/10.1128/JB.01184-12

Cain, A.K., Barquist, L., Goodman, A.L., Paulsen, I.T., Parkhill, J., van Opijnen, T., 2020. A decade of advances in transposon-insertion sequencing. Nat Rev Genet 21, 526–540. https://doi.org/10.1038/s41576-020-0244-x

Callanan, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P., Hill, C., 2018. RNA Phage Biology in a Metagenomic Era. Viruses 10, 386. https://doi.org/10.3390/v10070386

Callaway, E., 2020. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. Nature 588, 203–204. https://doi.org/10.1038/d41586-020-03348-4

Cañadas, I.C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., Minton, N.P., 2019. RiboCas: A Universal CRISPR-Based Editing Tool for Clostridium. ACS Synth. Biol. 8, 1379–1390. https://doi.org/10.1021/acssynbio.9b00075

Candel-Pérez, C., Ros-Berruezo, G., Martínez-Graciá, C., 2019. A review of *Clostridioides* [*Clostridium*] *difficile* occurrence through the food chain. Food Microbiology 77, 118–129. https://doi.org/10.1016/j.fm.2018.08.012

CDC, 2013. Antibiotic resistance threats in the United States, 2013.

CDC, 2005. Severe *Clostridium difficile*-associated disease in populations previously at low risk--four states, 2005. MMWR Morb. Mortal. Wkly. Rep. 54, 1201–1205.

Cenens, W., Makumi, A., Mebrhatu, M.T., Lavigne, R., Aertsen, A., 2013. Phage–host interactions during pseudolysogeny. Bacteriophage 3, e25029. https://doi.org/10.4161/bact.25029

Chan, B.K., Abedon, S.T., Loc-Carrillo, C., 2013. Phage cocktails and the future of phage therapy. Future Microbiology 8, 769–783. https://doi.org/10.2217/fmb.13.47

Chen, J., Gong, C.L., Hitchcock, M.M., Holubar, M., Deresinski, S., Hay, J.W., 2021. Cost-effectiveness of bezlotoxumab and fidaxomicin for initial *Clostridioides difficile* infection. Clinical Microbiology and Infection 27, 1448–1454. https://doi.org/10.1016/j.cmi.2021.04.004

Cho, E.H., Gumport, R.I., Gardner, J.F., 2002. Interactions between Integrase and Excisionase in the Phage Lambda Excisive Nucleoprotein Complex. J Bacteriol 184, 5200–5203. https://doi.org/10.1128/JB.184.18.5200-5203.2002

ClinicalTrials.gov [WWW Document], 2022. URL https://clinicaltrials.gov/ (accessed 10.5.22).

Clokie, M.R.J., Kropinski, A.M. (Eds.), 2009. Bacteriophages: methods and protocols vol 1 - Isolation, characterization and interactions, Methods in molecular biology. Humana Press, New York.

Clokie, M.R.J., Millard, A.D., Letarov, A.V., Heaphy, S., 2011. Phages in nature. Bacteriophage 1, 31–45. https://doi.org/10.4161/bact.1.1.14942

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. Front Microbiol 8. https://doi.org/10.3389/fmicb.2017.01108

Colomer-Lluch, M., Jofre, J., Muniesa, M., 2011. Antibiotic Resistance Genes in the Bacteriophage DNA Fraction of Environmental Samples. PLOS ONE 6, e17549. https://doi.org/10.1371/journal.pone.0017549

Cowley, L.A., Low, A.S., Pickard, D., Boinett, C.J., Dallman, T.J., Day, M., Perry, N., Gally, D.L., Parkhill, J., Jenkins, C., Cain, A.K., 2018. Transposon Insertion Sequencing Elucidates Novel Gene Involvement in Susceptibility and Resistance to Phages T4 and T7 in *Escherichia coli* O157. mBio 9, e00705-18. https://doi.org/10.1128/mBio.00705-18

Cumby, N., Reimer, K., Mengin-Lecreulx, D., Davidson, A.R., Maxwell, K.L., 2015. The phage tail tape measure protein, an inner membrane protein and a periplasmic chaperone play connected roles in the genome injection process of *E . coli* phage HK97: Phage and host protein requirements for HK97 genome injection. Molecular Microbiology 96, 437–447. https://doi.org/10.1111/mmi.12918

Dams, D., Brøndsted, L., Drulis-Kawa, Z., Briers, Y., 2019. Engineering of receptor-binding proteins in bacteriophages and phage tail-like bacteriocins. Biochem Soc Trans 47, 449–460. https://doi.org/10.1042/BST20180172

Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLOS ONE 5, e11147. https://doi.org/10.1371/journal.pone.0011147

Davis, J.J., Wattam, A.R., Aziz, R.K., Brettin, T., Butler, R., Butler, R.M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E.M., Gabbard, J.L., Gerdes, S., Guard, A., Kenyon, R.W., Machi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E.K., Olsen, G.J., Olson, R.D., Overbeek, J.C., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomas, C., VanOeffelen, M., Vonstein, V., Warren, A.S., Xia, F., Xie, D., Yoo, H., Stevens, R., 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res 48, D606–D612. https://doi.org/10.1093/nar/gkz943

De Sordi, L., Khanna, V., Debarbieux, L., 2017. The Gut Microbiota Facilitates Drifts in the Genetic Diversity and Infectivity of Bacterial Viruses. Cell Host Microbe 22, 801-808.e3. https://doi.org/10.1016/j.chom.2017.10.010

De Sordi, L., Lourenço, M., Debarbieux, L., 2019. “I will survive”: A tale of bacteriophage-bacteria coevolution in the gut. Gut Microbes 10, 92–99. https://doi.org/10.1080/19490976.2018.1474322

Debast, S.B., Bauer, M.P., Kuijper, E.J., 2014. European Society of Clinical Microbiology and Infectious Diseases: Update of the Treatment Guidance Document for *Clostridium difficile* Infection. Clinical Microbiology and Infection 20, 1–26. https://doi.org/10.1111/1469-0691.12418

Dedrick, R.M., Guerrero-Bustamante, C.A., Garlena, R.A., Russell, D.A., Ford, K., Harris, K., Gilmour, K.C., Soothill, J., Jacobs-Sera, D., Schooley, R.T., Hatfull, G.F., Spencer, H., 2019. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus.* Nat Med 25, 730–733. https://doi.org/10.1038/s41591-019-0437-z

Dedrick, R.M., Smith, B.E., Cristinziano, M., Freeman, K.G., Jacobs-Sera, D., Belessis, Y., Whitney Brown, A., Cohen, K.A., Davidson, R.M., van Duin, D., Gainey, A., Garcia, C.B., Robert George, C.R., Haidar, G., Ip, W., Iredell, J., Khatami, A., Little, J.S., Malmivaara, K., McMullan, B.J.,

Michalik, D.E., Moscatelli, A., Nick, J.A., Tupayachi Ortiz, M.G., Polenakovik, H.M., Robinson, P.D., Skurnik, M., Solomon, D.A., Soothill, J., Spencer, H., Wark, P., Worth, A., Schooley, R.T., Benson, C.A., Hatfull, G.F., 2022. Phage Therapy of *Mycobacterium* Infections: Compassionate-use of Phages in Twenty Patients with Drug-Resistant Mycobacterial Disease. Clinical Infectious Diseases ciac453. https://doi.org/10.1093/cid/ciac453

Department for Health, 2009. *Clostridium difficile* infection: How to deal with the problem 140.

Department for Health and Social Care, 2019. Contained and controlled: the UK's 20-year vision for antimicrobial resistance. UK government.

Department of the Environment, Food and Rural Affairs, 2021. The regulation of genetic technologies [WWW Document]. URL https://consult.defra.gov.uk/agri-food-chain-directorate/the-regulation-of-genetic-technologies/ (accessed 10.5.22).

Dodd, I.B., Shearwin, K.E., Egan, J.B., 2005. Revisited gene regulation in bacteriophage λ. Current Opinion in Genetics & Development, Chromosomes and expression mechanisms 15, 145–152. https://doi.org/10.1016/j.gde.2005.02.001

Donlan, R.M., 2009. Preventing biofilms of clinically relevant organisms using bacteriophage. Trends Microbiol. 17, 66–72. https://doi.org/10.1016/j.tim.2008.11.002

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. Biophys Rev 10, 535–542. https://doi.org/10.1007/s12551-017-0382-3

Dubberke, E.R., Haslam, D.B., Lanzas, C., Bobo, L.D., Burnham, C. -a. D., Gröhn, Y.T., Tarr, P.I., 2011. The Ecology and Pathobiology of *Clostridium difficile* Infections: An Interdisciplinary Challenge. Zoonoses

and Public Health 58, 4–20. https://doi.org/10.1111/j.1863-2378.2010.01352.x

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 67, 1198–1204. https://doi.org/10.1093/cid/ciy259

Ecale Zhou, C.L., 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. https://doi.org/10.1093/g3journal/jkab074

Eckert, C., Emirian, A., Le Monnier, A., Cathala, L., De Montclos, H., Goret, J., Berger, P., Petit, A., De Chevigny, A., Jean-Pierre, H., Nebbad, B., Camiade, S., Meckenstock, R., Lalande, V., Marchandin, H., Barbut, F., 2015. Prevalence and pathogenicity of binary toxin–positive *Clostridium difficile* strains that do not produce toxins A and B. New Microbes and New Infections 3, 12–17. https://doi.org/10.1016/j.nmni.2014.10.003

EFSA Panel on Food Additives and Nutrient Sources added to Food (ANS), Younes, M., Aggett, P., Aguilar, F., Crebelli, R., Dusemund, B., Filipič, M., Frutos, M.J., Galtier, P., Gundert-Remy, U., Kuhnle, G.G., Lambré, C., Leblanc, J.-C., Lillegaard, I.T., Moldeus, P., Mortensen, A., Oskarsson, A., Stankovic, I., Waalkens-Berendsen, I., Woutersen, R.A., Wright, M., Herman, L., Tobback, P., Pizzo, F., Smeraldi, C., Tard, A., Papaioannou, A., Gott, D., 2017. Safety of nisin (E 234) as a food additive in the light of new toxicological data and the proposed extension of use. EFSA Journal 15, e05063. https://doi.org/10.2903/j.efsa.2017.5063

Eklund, M., Poysky, F., Peterson, M., Meyers, J., 1976. Relationship of bacteriophages to alpha toxin production in *Clostridium novyi* types A and B. Infection and immunity 14, 793–803. https://doi.org/10.1128/IAI.14.3.793-803.1976

Eklund, M.W., Poysky, F.T., Meyers, J.A., Pelroy, G.A., 1974. Interspecies Conversion of *Clostridium botulinum* Type C to *Clostridium novyi* Type A by Bacteriophage. Science 186, 456–458. https://doi.org/10.1126/science.186.4162.456

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, 488–493. https://doi.org/10.1038/nature21049

Eskenazi, A., Lood, C., Wubbolts, J., Hites, M., Balarjishvili, N., Leshkasheli, L., Askilashvili, L., Kvachadze, L., van Noort, V., Wagemans, J., Jayankura, M., Chanishvili, N., de Boer, M., Nibbering, P., Kutateladze, M., Lavigne, R., Merabishvili, M., Pirnay, J.-P., 2022. Combination of pre-adapted bacteriophage therapy and antibiotics for treatment of fracture-related infection due to pandrug-resistant *Klebsiella pneumoniae*. Nat Commun 13, 302. https://doi.org/10.1038/s41467-021-27656-z

Espah Borujeni, A., Channarasappa, A.S., Salis, H.M., 2014. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Research 42, 2646–2659. https://doi.org/10.1093/nar/gkt1139

European Commission, 2022. EU consultation on legislation for plants produced by certain new genomic techniques [WWW Document]. Have your say. URL https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13119-Legislation-for-plants-produced-by-certain-new-genomic-techniques/public-consultation_en (accessed 10.5.22).

Fagan, R.P., Fairweather, N.F., 2011. *Clostridium difficile* Has Two Parallel and Essential Sec Secretion Systems. Journal of Biological Chemistry 286, 27483–27493. https://doi.org/10.1074/jbc.M111.263889

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L.,

Rosenbaum, M., Gordon, J.I., 2013. The Long-Term Stability of the Human Gut Microbiota. Science 341, 1237439. https://doi.org/10.1126/science.1237439

Fauconnier, A., 2017. Regulating phage therapy. EMBO Rep 18, 198–200. https://doi.org/10.15252/embr.201643250

Fernandes, S., São-José, C., 2018. Enzymes and Mechanisms Employed by Tailed Bacteriophages to Breach the Bacterial Cell Barriers. Viruses 10. https://doi.org/10.3390/v10080396

Fernández, L., Gutiérrez, D., García, P., Rodríguez, A., 2019. The Perfect Bacteriophage for Therapeutic Applications—A Quick Guide. Antibiotics (Basel) 8, 126. https://doi.org/10.3390/antibiotics8030126

Ferriol-González, C., Domingo-Calap, P., 2020. Phages for Biofilm Removal. Antibiotics (Basel) 9, 268. https://doi.org/10.3390/antibiotics9050268

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., McGovern, B.H., Trucksis, M., Henn, M.R., von Moltke, L., 2022. SER-109, an Oral Microbiome Therapy for Recurrent *Clostridioides difficile* Infection. N Engl J Med 386, 220–229. https://doi.org/10.1056/NEJMoa2106516

Fillol-Salom, A., Alsaadi, A., Sousa, J.A.M. de, Zhong, L., Foster, K.R., Rocha, E.P.C., Penadés, J.R., Ingmer, H., Haaber, J., 2019. Bacteriophages benefit from generalized transduction. PLOS Pathogens 15, e1007888. https://doi.org/10.1371/journal.ppat.1007888

Fogg, P.C.M., Colloms, S., Rosser, S., Stark, M., Smith, M.C.M., 2014. New Applications for Phage Integrases. J Mol Biol 426, 2703–2716. https://doi.org/10.1016/j.jmb.2014.05.014

Fogg, P.C.M., Rigden, D.J., Saunders, J.R., McCarthy, A.J., Allison, H.E., 2011. Characterization of the relationship between integrase, excisionase and antirepressor activities associated with a superinfecting Shiga toxin encoding bacteriophage. Nucleic Acids Res 39, 2116–2129. https://doi.org/10.1093/nar/gkq923

Fokine, A., Rossmann, M.G., 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4, e28281. https://doi.org/10.4161/bact.28281

Förster, B., Chung, P.K., Crobach, M.J.T., Kuijper, E.J., 2018. Application of Antibody-Mediated Therapy for Treatment and Prevention of *Clostridium difficile* Infection. Frontiers in Microbiology 9.

Freeman, J., Baines, S.D., Saxton, K., Wilcox, M.H., 2007. Effect of metronidazole on growth and toxin production by epidemic *Clostridium difficile* PCR ribotypes 001 and 027 in a human gut model. J Antimicrob Chemother 60, 83–91. https://doi.org/10.1093/jac/dkm113

Fujisawa, H., Morita, M., 1997. Phage DNA packaging. Genes to Cells 2, 537–545. https://doi.org/10.1046/j.1365-2443.1997.1450343.x

Galtier, M., De Sordi, L., Maura, D., Arachchi, H., Volant, S., Dillies, M.-A., Debarbieux, L., 2016. Bacteriophages to reduce gut carriage of antibiotic resistant uropathogens with low impact on microbiota composition: Reducing intestinal carriage with bacteriophages. Environmental Microbiology 18, 2237–2245. https://doi.org/10.1111/1462-2920.13284

Garen, A., Puck, T.T., 1951. THE FIRST TWO STEPS OF THE INVASION OF HOST CELLS BY BACTERIAL VIRUSES. II. J Exp Med 94, 177–189.

Gebhart, D., Lok, S., Clare, S., Tomas, M., Stares, M., Scholl, D., Donskey, C.J., Lawley, T.D., Govoni, G.R., 2015. A Modified R-Type Bacteriocin Specifically Targeting *Clostridium difficile* Prevents

Colonization of Mice without Affecting Gut Microbiota Diversity. mBio 6. https://doi.org/10.1128/mBio.02368-14

Gerding, D.N., Johnson, S., Rupnik, M., Aktories, K., 2014. Clostridium difficile binary toxin CDT. Gut Microbes 5, 15–27. https://doi.org/10.4161/gmic.26854

Gingery, R., Echols, H., 1967. Mutants of bacteriophage lambda unable to integrate into the host chromosome. Proc Natl Acad Sci U S A 58, 1507–1514.

Goh, K.G.K., Phan, M.-D., Forde, B.M., Chong, T.M., Yin, W.-F., Chan, K.-G., Ulett, G.C., Sweet, M.J., Beatson, S.A., Schembri, M.A., 2017. Genome-Wide Discovery of Genes Required for Capsule Production by Uropathogenic *Escherichia coli*. mBio 8, e01558-17. https://doi.org/10.1128/mBio.01558-17

Goh, S., Riley, T.V., Chang, B.J., 2005. Isolation and Characterization of Temperate Bacteriophages of *Clostridium difficile*. Appl. Environ. Microbiol. 71, 1079–1083. https://doi.org/10.1128/AEM.71.2.1079-1083.2005

Golding, I., Coleman, S., Nguyen, T.V.P., Yao, T., 2019. Decision Making by Temperate Phages, in: Reference Module in Life Sciences. Elsevier, p. B9780128096338209000. https://doi.org/10.1016/B978-0-12-809633-8.20969-2

Goodridge, L.D., Bisha, B., 2011. Phage-based biocontrol strategies to reduce foodborne pathogens in foods. Bacteriophage 1, 130–137. https://doi.org/10.4161/bact.1.3.17629

Gottesman, M.E., 1974. The integration and excision of bacteriophage lambda. Cell 1, 69–72. https://doi.org/10.1016/0092-8674(74)90062-2

Govind, R., Fralick, J.A., Rolfe, R.D., 2006. Genomic Organization and Molecular Characterization of *Clostridium difficile* Bacteriophage

ΦCD119. Journal of Bacteriology 188, 2568–2577. https://doi.org/10.1128/JB.188.7.2568-2577.2006

Govind, R., Vediyappan, G., Rolfe, R.D., Dupuy, B., Fralick, J.A., 2009. Bacteriophage-Mediated Toxin Gene Regulation in *Clostridium difficile*. J Virol 83, 12037–12045. https://doi.org/10.1128/JVI.01256-09

Grazziotin, A.L., Koonin, E.V., Kristensen, D.M., 2017. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. Nucleic Acids Res 45, D491–D498. https://doi.org/10.1093/nar/gkw975

Groth, A.C., Calos, M.P., 2004. Phage Integrases: Biology and Applications. Journal of Molecular Biology 335, 667–678. https://doi.org/10.1016/j.jmb.2003.09.082

Groth, A.C., Olivares, E.C., Thyagarajan, B., Calos, M.P., 2000. A phage integrase directs efficient site-specific integration in human cells. PNAS 97, 5995–6000. https://doi.org/10.1073/pnas.090527097

Gülke, I., Pfeifer, G., Liese, J., Fritz, M., Hofmann, F., Aktories, K., Barth, H., 2001. Characterization of the Enzymatic Component of the ADP-Ribosyltransferase Toxin CDTa from Clostridium difficile. Infection and Immunity 69, 6004–6011. https://doi.org/10.1128/IAI.69.10.6004-6011.2001

Gutiérrez, D., Fernández, L., Rodríguez, A., García, P., 2019. Role of Bacteriophages in the Implementation of a Sustainable Dairy Chain. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.00012

Hagens, S., Loessner, M.J., 2010. Bacteriophage for biocontrol of foodborne pathogens: calculations and considerations. Current pharmaceutical biotechnology 11, 58–68.

Haines, M.E.K., Hodges, F.E., Nale, J.Y., Mahony, J., van Sinderen, D., Kaczorowska, J., Alrashid, B., Akter, M., Brown, N., Sauvageau, D., Sicheritz-Pontén, T., Thanki, A.M., Millard, A.D., Galyov, E.E., Clokie,

M.R.J., 2021. Analysis of Selection Methods to Develop Novel Phage Therapy Cocktails Against Antimicrobial Resistant Clinical Isolates of Bacteria. Frontiers in Microbiology 12.

Han, Y.W., Gumport, R.I., Gardner, J.F., 1993. Complementation of bacteriophage lambda integrase mutants: evidence for an intersubunit active site. The EMBO Journal 12, 4577–4584. https://doi.org/10.1002/j.1460-2075.1993.tb06146.x

Hao, N., Agnew, D., Krishna, S., Dodd, I.B., Shearwin, K.E., 2021. Analysis of Infection Time Courses Shows CII Levels Determine the Frequency of Lysogeny in Phage 186. Pharmaceuticals 14, 998. https://doi.org/10.3390/ph14100998

Haq, I.U., Chaudhry, W.N., Akhtar, M.N., Andleeb, S., Qadri, I., 2012. Bacteriophages and their implications on future biotechnology: a review. Virology Journal 9, 9. https://doi.org/10.1186/1743-422X-9-9

Hargreaves, K.R., Clokie, M.R.J., 2015. A Taxonomic Review of *Clostridium difficile* Phages and Proposal of a Novel Genus, “Phimmp04likevirus.” Viruses 7, 2534–2541. https://doi.org/10.3390/v7052534

Hargreaves, K.R., Clokie, M.R.J., 2014. *Clostridium difficile* phages: still difficult? Front. Microbiol. 5. https://doi.org/10.3389/fmicb.2014.00184

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, e01045-13. https://doi.org/10.1128/mBio.01045-13

Hascic, A., Wolfensberger, A., Clack, L., Schreiber, P.W., Kuster, S.P., Sax, H., 2022. Documentation of adherence to infection prevention best practice in patient records: a mixed-methods investigation. Antimicrobial Resistance & Infection Control 11, 107. https://doi.org/10.1186/s13756-022-01139-2

Hatcher, E.L., Zhdanov, S.A., Bao, Y., Blinkova, O., Nawrocki, E.P., Ostapchuck, Y., Schäffer, A.A., Brister, J.R., 2017. Virus Variation Resource – improved response to emergent viral outbreaks. Nucleic Acids Res 45, D482–D490. https://doi.org/10.1093/nar/gkw1065

Hay, I.D., Lithgow, T., 2019. Filamentous phages: masters of a microbial sharing economy. EMBO reports 20, e47427. https://doi.org/10.15252/embr.201847427

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, 79–85. https://doi.org/10.1016/j.mimet.2009.05.004

Hegedűs, T., Geisler, M., Lukács, G.L., Farkas, B., 2022. Ins and outs of AlphaFold2 transmembrane protein structure predictions. Cell Mol Life Sci 79, 73. https://doi.org/10.1007/s00018-021-04112-1

Heuler, J., Fortier, L.-C., Sun, X., 2021. *Clostridioides difficile* phage biology and application. FEMS Microbiology Reviews 45, fuab012. https://doi.org/10.1093/femsre/fuab012

Hochschild, A., Douhan, J., Ptashne, M., 1986. How λ repressor and λ Cro distinguish between OR1 and OR3. Cell 47, 807–816. https://doi.org/10.1016/0092-8674(86)90523-4

Hols, P., Ledesma-García, L., Gabant, P., Mignolet, J., 2019. Mobilization of Microbiota Commensals and Their Bacteriocins for Therapeutics. Trends in Microbiology 27, 690–702. https://doi.org/10.1016/j.tim.2019.03.007

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

Horton, R.M., Cai, Z., Ho, S.N., Pease, L.R., 2013. Gene Splicing by Overlap Extension: Tailor-Made Genes Using the Polymerase Chain Reaction. BioTechniques 54, 129–133. https://doi.org/10.2144/000114017

Hu, J., Fan, J., Sun, Z., Liu, S., 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics 36, 2253–2255. https://doi.org/10.1093/bioinformatics/btz891

Huang, H., Chai, C., Yang, S., Jiang, W., Gu, Y., 2019. Phage serine integrase-mediated genome engineering for efficient expression of chemical biosynthetic pathway in gas-fermenting *Clostridium ljungdahlii*. Metabolic Engineering 52, 293–302. https://doi.org/10.1016/j.ymben.2019.01.005

Humphreys, D.P., Wilcox, M.H., 2014. Antibodies for Treatment of *Clostridium difficile* Infection. Clin Vaccine Immunol 21, 913–923. https://doi.org/10.1128/CVI.00116-14

Hunt, J.J., Ballard, J.D., 2013. Variations in Virulence and Molecular Biology among Emerging Strains of *Clostridium difficile*. Microbiol Mol Biol Rev 77, 567–581. https://doi.org/10.1128/MMBR.00017-13

Isidro, J., Mendes, A.L., Serrano, M., O.Henriques, A., Oleastro, M., 2017. Overview of *Clostridium difficile* Infection: Life Cycle, Epidemiology, Antimicrobial Resistance and Treatment. *Clostridium Difficile* - A Comprehensive Overview. https://doi.org/10.5772/intechopen.69053

Ivanovska, I.L., Pablo, P.J. de, Ibarra, B., Sgalari, G., MacKintosh, F.C., Carrascosa, J.L., Schmidt, C.F., Wuite, G.J.L., 2004. Bacteriophage capsids: Tough nanoshells with complex elastic properties. PNAS 101, 7600–7605. https://doi.org/10.1073/pnas.0308198101

Janezic, S., Ocepek, M., Zidaric, V., Rupnik, M., 2012. *Clostridium difficile* genotypes other than ribotype 078 that are prevalent among

human, animal and environmental isolates. BMC Microbiology 12, 48. https://doi.org/10.1186/1471-2180-12-48

Jayaram, M., Ma, C.-H., Kachroo, A.H., Rowley, P.A., Guga, P., Fan, H.-F., Voziyanov, Y., 2015. An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective. Microbiology Spectrum 3, 3.4.12. https://doi.org/10.1128/microbiolspec.MDNA3-0021-2014

Jebri, S., Rahmani, F., Hmaied, F., 2021. Bacteriophages as antibiotic resistance genes carriers in agro-food systems. Journal of Applied Microbiology 130, 688–698. https://doi.org/10.1111/jam.14851

Johansen, A.H., Brøndsted, L., Hammer, K., 2003. Identification of operator sites of the CI repressor of phage TP901-1: evolutionary link to other phages. Virology 311, 144–156. https://doi.org/10.1016/S0042-6822(03)00169-7

Johnson, S., 2009. Recurrent *Clostridium difficile* infection: A review of risk factors, treatments, and outcomes. Journal of Infection 58, 403–410. https://doi.org/10.1016/j.jinf.2009.03.010

Jordan, T.C., Burnett, S.H., Carson, S., Caruso, S.M., Clase, K., DeJong, R.J., Dennehy, J.J., Denver, D.R., Dunbar, D., Elgin, S.C.R., Findley, A.M., Gissendanner, C.R., Golebiewska, U.P., Guild, N., Hartzog, G.A., Grillo, W.H., Hollowell, G.P., Hughes, L.E., Johnson, A., King, R.A., Lewis, L.O., Li, W., Rosenzweig, F., Rubin, M.R., Saha, M.S., Sandoz, J., Shaffer, C.D., Taylor, B., Temple, L., Vazquez, E., Ware, V.C., Barker, L.P., Bradley, K.W., Jacobs-Sera, D., Pope, W.H., Russell, D.A., Cresawn, S.G., Lopatto, D., Bailey, C.P., Hatfull, G.F., 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. mBio 5, e01051-01013. https://doi.org/10.1128/mBio.01051-13

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., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kaiser, A.D., 1957. Mutations in a temperate bacteriophage affecting its ability to lysogenize *Escherichia coli*. Virology 3, 42–61. https://doi.org/10.1016/0042-6822(57)90022-3

Käll, L., Krogh, A., Sonnhammer, E.L.L., 2004. A Combined Transmembrane Topology and Signal Peptide Prediction Method. Journal of Molecular Biology 338, 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016

Kashyap, S.K., Maherchandani, S., Kumar, N., 2014. Chapter 18 - Ribotyping: A Tool for Molecular Taxonomy, in: Verma, A.S., Singh, A. (Eds.), Animal Biotechnology. Academic Press, San Diego, pp. 327–344. https://doi.org/10.1016/B978-0-12-416002-6.00018-3

Keen, E.C., 2015. A century of phage research: Bacteriophages and the shaping of modern biology. BioEssays 37, 6–9. https://doi.org/10.1002/bies.201400152

Kellenberger, G., Zichichi, M.L., Weigle, J., 1961. A mutation affecting the DNA content of bacteriophage lambda and its lysogenizing properties. Journal of Molecular Biology 3, 399-IN1. https://doi.org/10.1016/S0022-2836(61)80053-3

Kenny, J.G., Leach, S., de la Hoz, A.B., Venema, G., Kok, J., Fitzgerald, G.F., Nauta, A., Alonso, J.C., van Sinderen, D., 2006. Characterization of the lytic–lysogenic switch of the lactococcal bacteriophage Tuc2009. Virology 347, 434–446. https://doi.org/10.1016/j.virol.2005.11.041

Kilcher, S., Loessner, M.J., 2019. Engineering Bacteriophages as Versatile Biologics. Trends in Microbiology, Special Issue: Antimicrobial

Resistance and Novel Therapeutics 27, 355–367. https://doi.org/10.1016/j.tim.2018.09.006

King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), 2011. Family - Siphoviridae, in: Virus Taxonomy - Ninth Report of the International Committee on Taxonomy of Viruses. https://doi.org/10.1016/B978-0-12-384684-6.00004-5

Kirk, J.A., Gebhart, D., Buckley, A.M., Lok, S., Scholl, D., Douce, G.R., Govoni, G.R., Fagan, R.P., 2017. New class of precision antimicrobials redefines role of *Clostridium difficile* S-layer in virulence and viability. Sci. Transl. Med. 9, eaah6813. https://doi.org/10.1126/scitranslmed.aah6813

Kiro, R., Shitrit, D., Qimron, U., 2014. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. RNA Biology 11, 42–44. https://doi.org/10.4161/rna.27766

Kizziah, J.L., Manning, K.A., Dearborn, A.D., Dokland, T., 2020. Structure of the host cell recognition and penetration machinery of a *Staphylococcus aureus* bacteriophage. PLOS Pathogens 16, e1008314. https://doi.org/10.1371/journal.ppat.1008314

Knetsch, C.W., Kumar, N., Forster, S.C., Connor, T.R., Browne, H.P., Harmanus, C., Sanders, I.M., Harris, S.R., Turner, L., Morris, T., Perry, M., Miyajima, F., Roberts, P., Pirmohamed, M., Songer, J.G., Weese, J.S., Indra, A., Corver, J., Rupnik, M., Wren, B.W., Riley, T.V., Kuijper, E.J., Lawley, T.D., 2018. Zoonotic Transfer of *Clostridium difficile* Harboring Antimicrobial Resistance between Farm Animals and Humans. Journal of Clinical Microbiology 56, e01384-17. https://doi.org/10.1128/JCM.01384-17

Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37, 540–546. https://doi.org/10.1038/s41587-019-0072-8

Konopacki, M., Grygorcewicz, B., Dołęgowska, B., Kordas, M., Rakoczy, R., 2020. PhageScore: A simple method for comparative evaluation of

bacteriophages lytic activity. Biochemical Engineering Journal 161, 107652. https://doi.org/10.1016/j.bej.2020.107652

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736. https://doi.org/10.1101/gr.215087.116

Koskella, B., Brockhurst, M.A., 2014. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiol. Rev. 38, 916–931. https://doi.org/10.1111/1574-6976.12072

Kostyuchenko, V.A., Leiman, P.G., Chipman, P.R., Kanamaru, S., van Raaij, M.J., Arisaka, F., Mesyanzhinov, V.V., Rossmann, M.G., 2003. Three-dimensional structure of bacteriophage T4 baseplate. Nat Struct Mol Biol 10, 688–693. https://doi.org/10.1038/nsb970

Kristensen, D.M., Saeed, U., Frishman, D., Koonin, E.V., 2015. A census of α-helical membrane proteins in double-stranded DNA viruses infecting bacteria and archaea. BMC Bioinformatics 16, 380. https://doi.org/10.1186/s12859-015-0817-4

Kumar, M., Sarma, D.K., Shubham, S., Kumawat, M., Verma, V., Nina, P.B., JP, D., Kumar, S., Singh, B., Tiwari, R.R., 2021. Futuristic Non-antibiotic Therapies to Combat Antibiotic Resistance: A Review. Front Microbiol 12, 609459. https://doi.org/10.3389/fmicb.2021.609459

Kumar, P., Arya, A., 2018. Ramachandran Plot.

Kutter, E., 2009. Phage Host Range and Efficiency of Plating, in: Clokie, M.R.J., Kropinski, A.M. (Eds.), Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions, Methods in Molecular Biology™. Humana Press, Totowa, NJ, pp. 141–149. https://doi.org/10.1007/978-1-60327-164-6_14

Labrie, S.J., Samson, J.E., Moineau, S., 2010. Bacteriophage resistance mechanisms. Nature Reviews Microbiology 8, 317–327. https://doi.org/10.1038/nrmicro2315

Langridge, G.C., Phan, M.-D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., Wain, J., Parkhill, J., Turner, A.K., 2009. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. Genome Res 19, 2308–2316. https://doi.org/10.1101/gr.097097.109

Lawler, A.J., Lambert, P.A., Worthington, T., 2020. A Revised Understanding of *Clostridioides difficile* Spore Germination. Trends in Microbiology 28, 744–752. https://doi.org/10.1016/j.tim.2020.03.004

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, e68562. https://doi.org/10.1371/journal.pone.0068562

Lees, E.A., Miyajima, F., Pirmohamed, M., Carrol, E.D., 2016. The role of *Clostridium difficile* in the paediatric and neonatal gut — a narrative review. Eur J Clin Microbiol Infect Dis 35, 1047–1057. https://doi.org/10.1007/s10096-016-2639-3

Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V., Rossmann, M.G., 2004. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. Cell 118, 419–429. https://doi.org/10.1016/j.cell.2004.07.022

Leitsch, D., 2019. A review on metronidazole: an old warhorse in antimicrobial chemotherapy. Parasitology 146, 1167–1178. https://doi.org/10.1017/S0031182017002025

Lewis, B.B., Pamer, E.G., 2017. Microbiota-Based Therapies for *Clostridium difficile* and Antibiotic-Resistant Enteric Infections. Annu. Rev. Microbiol. 71, 157–178. https://doi.org/10.1146/annurev-micro-090816-093549

Lewis, R.J., Brannigan, J.A., Offen, W.A., Smith, I., Wilkinson, A.J., 1998. An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex11Edited by J. M. Thornton. Journal of Molecular Biology 283, 907–912. https://doi.org/10.1006/jmbi.1998.2163

Li, G., Cortez, M.H., Dushoff, J., Weitz, J.S., 2020. When to be temperate: on the fitness benefits of lysis vs. lysogeny. Virus Evolution 6, veaa042. https://doi.org/10.1093/ve/veaa042

Li, W., O'Neill, K.R., Haft, D.H., DiCuccio, M., Chetvernin, V., Badretdin, A., Coulouris, G., Chitsaz, F., Derbyshire, M.K., Durkin, A.S., Gonzales, N.R., Gwadz, M., Lanczycki, C.J., Song, J.S., Thanki, N., Wang, J., Yamashita, R.A., Yang, M., Zheng, C., Marchler-Bauer, A., Thibaud-Nissen, F., 2021. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res 49, D1020–D1028. https://doi.org/10.1093/nar/gkaa1105

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*. Sci Rep 6, 26455. https://doi.org/10.1038/srep26455

Lin, D.M., Koskella, B., Lin, H.C., 2017. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. World J Gastrointest Pharmacol Ther 8, 162–173. https://doi.org/10.4292/wjgpt.v8.i3.162

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research 42, D490–D495. https://doi.org/10.1093/nar/gkt1178

Łoś, M., Węgrzyn, G., 2012. Chapter 9 - Pseudolysogeny, in: Łobocka, M., Szybalski, W.T. (Eds.), Advances in Virus Research, Bacteriophages, Part A. Academic Press, pp. 339–349. https://doi.org/10.1016/B978-0-12-394621-8.00019-4

Lucía Fernández, Susana Escobedo, Diana Gutiérrez, Silvia Portilla, Beatriz Martínez, Pilar García, Ana Rodríguez, 2017. Bacteriophages in the Dairy Environment: From Enemies to Allies. Antibiotics 6, 27. https://doi.org/10.3390/antibiotics6040027

Madigan, M.T., Brock, T.D. (Eds.), 2018. Brock Biology of microorganisms, 15th ed. ed. Pearson, New York.

Mahony, J., Alqarni, M., Stockdale, S., Spinelli, S., Feyereisen, M., Cambillau, C., Sinderen, D. van, 2016. Functional and structural dissection of the tape measure protein of lactococcal phage TP901-1. Scientific Reports 6, 36667. https://doi.org/10.1038/srep36667

Mai, V., Ukhanova, M., Reinhard, M.K., Li, M., Sulakvelidze, A., 2015. Bacteriophage administration significantly reduces *Shigella* colonization and shedding by *Shigella*-challenged mice without deleterious side effects and distortions in the gut microbiota. Bacteriophage 5, e1088124. https://doi.org/10.1080/21597081.2015.1088124

Makhija, H., Roy, S., Hoon, S., Ghadessy, F.J., Wong, D., Jaiswal, R., Campana, D., Dröge, P., 2018. A novel λ integrase-mediated seamless vector transgenesis platform for therapeutic protein expression. Nucleic Acids Research 46, e99–e99. https://doi.org/10.1093/nar/gky500

Mäntynen, S., Laanto, E., Oksanen, H.M., Poranen, M.M., Díaz-Muñoz, S.L., 2021. Black box of phage–bacterium interactions: exploring

alternative phage infection strategies. Open Biology 11, 210188. https://doi.org/10.1098/rsob.210188

Martel, B., Moineau, S., 2014. CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. Nucleic Acids Res 42, 9504–9513. https://doi.org/10.1093/nar/gku628

Mayer, M.J., Garefalaki, V., Spoerl, R., Narbad, A., Meijers, R., 2011. Structure-based modification of a *Clostridium difficile*-targeting endolysin affects activity and host range. J Bacteriol 193, 5477–5486. https://doi.org/10.1128/JB.00439-11

McDonald, L.C., Gerding, D.N., Johnson, S., Bakken, J.S., Carroll, K.C., Coffin, S.E., Dubberke, E.R., Garey, K.W., Gould, C.V., Kelly, C., Loo, V., Shaklee Sammons, J., Sandora, T.J., Wilcox, M.H., 2018. Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). Clin Infect Dis 66, e1–e48. https://doi.org/10.1093/cid/cix1085

McNair, K., Zhou, C., Dinsdale, E.A., Souza, B., Edwards, R.A., 2019. PHANOTATE: a novel approach to gene identification in phage genomes. Bioinformatics 35, 4537–4542. https://doi.org/10.1093/bioinformatics/btz265

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, 549–554. https://doi.org/10.1016/j.anaerobe.2010.08.006

Mensa-Wilmot, K., Seaby, R., Alfano, C., Wold, M.C., Gomes, B., McMacken, R., 1989. Reconstitution of a nine-protein system that initiates bacteriophage lambda DNA replication. J. Biol. Chem. 264, 2853–2861.

Merrikin, D.J., Terry, C.S., 1972. Use of Pyocin 78-C2 in the Treatment of *Pseudomonas aeruginosa* Infection in Mice. Appl Microbiol 23, 164–165. https://doi.org/10.1128/am.23.1.164-165.1972

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. Nat Methods 19, 679–682. https://doi.org/10.1038/s41592-022-01488-1

Molineux, I.J., Panja, D., 2013. Popping the cork: mechanisms of phage genome ejection. Nature Reviews Microbiology 11, 194–204. https://doi.org/10.1038/nrmicro2988

Motlagh, A.M., Bhattacharjee, A.S., Goel, R., 2016. Biofilm control with natural and genetically-modified phages. World J Microbiol Biotechnol 32, 67. https://doi.org/10.1007/s11274-016-2009-4

Moura, I., Monot, M., Tani, C., Spigaglia, P., Barbanti, F., Norais, N., Dupuy, B., Bouza, E., Mastrantonio, P., 2014. Multidisciplinary Analysis of a Nontoxigenic *Clostridium difficile* Strain with Stable Resistance to Metronidazole. Antimicrob Agents Chemother 58, 4957–4960. https://doi.org/10.1128/AAC.02350-14

Mullane, K.M., Miller, M.A., Weiss, K., Lentnek, A., Golan, Y., Sears, P.S., Shue, Y.-K., Louie, T.J., Gorbach, S.L., 2011. Efficacy of Fidaxomicin Versus Vancomycin as Therapy for *Clostridium difficile* Infection in Individuals Taking Concomitant Antibiotics for Other Concurrent Infections. Clin Infect Dis 53, 440–447. https://doi.org/10.1093/cid/cir404

Mullard, A., 2016. FDA approves antitoxin antibody. Nature Reviews Drug Discovery 15, 811–811. https://doi.org/10.1038/nrd.2016.257

Murialdo, H., Becker, A., 1978. Head morphogenesis of complex double-stranded deoxyribonucleic acid bacteriophages. Microbiol Rev 42, 529–576.

Murray, C.J., Ikuta, K.S., Sharara, F., Swetschinski, L., Aguilar, G.R., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S.C., Browne, A.J., Chipeta, M.G., Fell, F., Hackett, S., Haines-Woodhouse, G., Hamadani, B.H.K., Kumaran, E.A.P., McManigal, B., Agarwal, R., Akech, S., Albertson, S., Amuasi, J., Andrews, J., Aravkin, A., Ashley, E., Bailey, F., Baker, S., Basnyat, B., Bekker, A., Bender, R., Bethou, A., Bielicki, J., Boonkasidecha, S., Bukosia, J., Carvalheiro, C., Castañeda-Orjuela, C., Chansamouth, V., Chaurasia, S., Chiurchiù, S., Chowdhury, F., Cook, A.J., Cooper, B., Cressey, T.R., Criollo-Mora, E., Cunningham, M., Darboe, S., Day, N.P.J., Luca, M.D., Dokova, K., Dramowski, A., Dunachie, S.J., Eckmanns, T., Eibach, D., Emami, A., Feasey, N., Fisher-Pearson, N., Forrest, K., Garrett, D., Gastmeier, P., Giref, A.Z., Greer, R.C., Gupta, V., Haller, S., Haselbeck, A., Hay, S.I., Holm, M., Hopkins, S., Iregbu, K.C., Jacobs, J., Jarovsky, D., Javanmardi, F., Khorana, M., Kissoon, N., Kobeissi, E., Kostyanev, T., Krapp, F., Krumkamp, R., Kumar, A., Kyu, H.H., Lim, C., Limmathurotsakul, D., Loftus, M.J., Lunn, M., Ma, J., Mturi, N., Munera-Huertas, T., Musicha, P., Mussi-Pinhata, M.M., Nakamura, T., Nanavati, R., Nangia, S., Newton, P., Ngoun, C., Novotney, A., Nwakanma, D., Obiero, C.W., Olivas-Martinez, A., Olliaro, P., Ooko, E., Ortiz-Brizuela, E., Peleg, A.Y., Perrone, C., Plakkal, N., Ponce-de-Leon, A., Raad, M., Ramdin, T., Riddell, A., Roberts, T., Robotham, J.V., Roca, A., Rudd, K.E., Russell, N., Schnall, J., Scott, J.A.G., Shivamallappa, M., Sifuentes-Osornio, J., Steenkeste, N., Stewardson, A.J., Stoeva, T., Tasak, N., Thaiprakong, A., Thwaites, G., Turner, C., Turner, P., Doorn, H.R. van, Velaphi, S., Vongpradith, A., Vu, H., Walsh, T., Waner, S., Wangrangsimakul, T., Wozniak, T., Zheng, P., Sartorius, B., Lopez, A.D., Stergachis, A., Moore, C., Dolecek, C., Naghavi, M., 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet 399, 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0

Musk, D.J., Hergenrother, P.J., 2006. Chemical countermeasures for the control of bacterial biofilms: effective compounds and promising targets.

Curr Med Chem 13, 2163–2177. https://doi.org/10.2174/092986706777935212

Nale, J.Y., Chutia, M., Carr, P., Hickenbotham, P.T., Clokie, M.R.J., 2016a. 'Get in Early'; Biofilm and Wax Moth (Galleria mellonella) Models Reveal New Insights into the Therapeutic Potential of *Clostridium difficile* Bacteriophages. Front. Microbiol. 7. 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 7, 13. https://doi.org/10.3390/antibiotics7010013

Nale, J.Y., Spencer, J., Hargreaves, K.R., Buckley, A.M., Trzepiński, P., Douce, G.R., Clokie, M.R.J., 2016b. Bacteriophage Combinations Significantly Reduce *Clostridium difficile* Growth In Vitro and Proliferation In Vivo. Antimicrob Agents Chemother 60, 968–981. https://doi.org/10.1128/AAC.01774-15

National Institute for Health and Care, 2019. BNF: British National Formulary - NICE [WWW Document]. URL https://bnf.nice.org.uk/drug/vancomycin.html (accessed 7.19.19).

National Institute for Health and Care and Excellence, 2021. *Clostridioides difficile* infection: antimicrobial prescribing.

Navalkele, B.D., Chopra, T., 2018. Bezlotoxumab: an emerging monoclonal antibody therapy for prevention of recurrent *Clostridium difficile* infection. Biologics 12, 11–21. https://doi.org/10.2147/BTT.S127099

New England Biolabs, 2020. High Efficiency Transformation Protocol (C2984) [WWW Document]. High Efficiency Transformation Protocol (C2984). URL https://international.neb.com/protocols/0001/01/01/high-efficiency-transformation-protocol-c2984 (accessed 2.24.20).

Ng, Q.X., Loke, W., Foo, N.X., Mo, Y., Yeo, W.-S., Soh, A.Y.S., 2019. A systematic review of the use of rifaximin for *Clostridium difficile* infections. Anaerobe 55, 35–39. https://doi.org/10.1016/j.anaerobe.2018.10.011

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, aaf8729. 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, 2379–2380. https://doi.org/10.1093/bioinformatics/btx157

Oppenheim, A.B., Kobiler, O., Stavans, J., Court, D.L., Adhya, S., 2005. Switches in Bacteriophage Lambda Development. Annual Review of Genetics 39, 409–429. https://doi.org/10.1146/annurev.genet.39.073003.113656

Orenstein, R., Dubberke, E., Hardi, R., Ray, A., Mullane, K., Pardi, D.S., Ramesh, M.S., 2016. Safety and Durability of RBX2660 (Microbiota Suspension) for Recurrent *Clostridium difficile* Infection: Results of the PUNCH CD Study. Clin Infect Dis. 62, 596–602. https://doi.org/10.1093/cid/civ938

Orth, P., Xiao, L., Hernandez, L.D., Reichert, P., Sheth, P.R., Beaumont, M., Yang, X., Murgolo, N., Ermakov, G., DiNunzio, E., Racine, F., Karczewski, J., Secore, S., Ingram, R.N., Mayhood, T., Strickland, C., Therien, A.G., 2014. Mechanism of action and epitopes of *Clostridium difficile* toxin B-neutralizing antibody bezlotoxumab revealed by X-ray crystallography. J Biol Chem 289, 18008–18021. https://doi.org/10.1074/jbc.M114.560748

Owen, S.V., Canals, R., Wenner, N., Hammarlöf, D.L., Kröger, C., Hinton, J.C.D.Y. 2020, 2020. A window into lysogeny: revealing temperate phage

biology with transcriptomics. Microbial Genomics 6, e000330. https://doi.org/10.1099/mgen.0.000330

Pasquale, V., Romano, V., Rupnik, M., Capuano, F., Bove, D., Aliberti, F., Krovacek, K., Dumontet, S., 2012. Occurrence of toxigenic *Clostridium difficile* in edible bivalve molluscs. Food Microbiology 31, 309–312. https://doi.org/10.1016/j.fm.2012.03.001

Patil, A., Banerji, R., Kanojiya, P., Koratkar, S., Saroj, S., 2021. Bacteriophages for ESKAPE: role in pathogenicity and measures of control. Expert Review of Anti-infective Therapy 19, 845–865. https://doi.org/10.1080/14787210.2021.1858800

Paul, J.H., 2008. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? ISME J 2, 579–589. https://doi.org/10.1038/ismej.2008.35

Pell, L.G., Liu, A., Edmonds, L., Donaldson, L.W., Howell, P.L., Davidson, A.R., 2009. The X-Ray Crystal Structure of the Phage λ Tail Terminator Protein Reveals the Biologically Relevant Hexameric Ring Structure and Demonstrates a Conserved Mechanism of Tail Termination among Diverse Long-Tailed Phages. Journal of Molecular Biology 389, 938–951. https://doi.org/10.1016/j.jmb.2009.04.072

Penadés, J.R., Chen, J., Quiles-Puchalt, N., Carpena, N., Novick, R.P., 2015. Bacteriophage-mediated spread of bacterial virulence genes. Current Opinion in Microbiology, Host–microbe interactions: bacteria • Genomics 23, 171–178. https://doi.org/10.1016/j.mib.2014.11.019

Peng, Z., Wang, S., Gide, M., Zhu, D., Lamabadu Warnakulasuriya Patabendige, H.M., Li, C., Cai, J., Sun, X., 2019. A Novel Bacteriophage Lysin-Human Defensin Fusion Protein Is Effective in Treatment of *Clostridioides difficile* Infection in Mice. Front. Microbiol. 9, 3234. https://doi.org/10.3389/fmicb.2018.03234

Phothichaisri, W., Ounjai, P., Phetruen, T., Janvilisri, T., Khunrae, P., Singhakaew, S., Wangroongsarb, P., Chankhamhaengdecha, S., 2018.

Characterization of Bacteriophages Infecting Clinical Isolates of *Clostridium difficile*. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.01701

Platt, R., Reynolds, D.L., Phillips, G.J., 2003. Development of a novel method of lytic phage delivery by use of a bacteriophage P22 site-specific recombination system. FEMS Microbiol Lett 223, 259–265. https://doi.org/10.1016/S0378-1097(03)00388-4

Pope, W.H., Ferreira, C.M., Jacobs-Sera, D., Benjamin, R.C., Davis, A.J., DeJong, R.J., Elgin, S.C.R., Guilfoile, F.R., Forsyth, M.H., Harris, A.D., Harvey, S.E., Hughes, L.E., Hynes, P.M., Jackson, A.S., Jalal, M.D., MacMurray, E.A., Manley, C.M., McDonough, M.J., Mosier, J.L., Osterbann, L.J., Rabinowitz, H.S., Rhyan, C.N., Russell, D.A., Saha, M.S., Shaffer, C.D., Simon, S.E., Sims, E.F., Tovar, I.G., Weisser, E.G., Wertz, J.T., Weston-Hafer, K.A., Williamson, K.E., Zhang, B., Cresawn, S.G., Jain, P., Piuri, M., Jacobs, W.R., Hendrix, R.W., Hatfull, G.F., 2011. Cluster K Mycobacteriophages: Insights into the Evolutionary Origins of Mycobacteriophage TM4. PLoS One 6. https://doi.org/10.1371/journal.pone.0026750

Poxton, I.R., McCoubrey, J., Blair, G., 2001. The pathogenicity of *Clostridium difficile*. Clinical Microbiology and Infection 7, 421–427. https://doi.org/10.1046/j.1198-743x.2001.00287.x

Prevelige, P.E., Cortines, J.R., 2018. Phage assembly and the special role of the portal protein. Current Opinion in Virology, Virus structure and expression • Viral evolution 31, 66–73. https://doi.org/10.1016/j.coviro.2018.09.004

Ptashne, M., 1967. ISOLATION OF THE λ PHAGE REPRESSOR*. Proc Natl Acad Sci U S A 57, 306–313.

Public Health England, 2019. *Clostridium difficile* Ribotyping Network (CDRN) for England and Northern Ireland (Government).

Rakonjac, J., 2012. Filamentous Bacteriophages: Biology and Applications, in: ELS. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0000777

Ramesh, V., Fralick, J.A., Rolfe, R.D., 1999. Prevention of *Clostridium difficile*-induced ileocecitis with Bacteriophage 10.

Rendueles, C., Duarte, A.C., Escobedo, S., Fernández, L., Rodríguez, A., García, P., Martínez, B., 2022. Combined use of bacteriocins and bacteriophages as food biopreservatives. A review. International Journal of Food Microbiology 368, 109611. https://doi.org/10.1016/j.ijfoodmicro.2022.109611

Ripp, S., Miller, R.V., 1998. Dynamics of the pseudolysogenic response in slowly growing cells of *Pseudomonas aeruginosa*. Microbiology, 144, 2225–2232. https://doi.org/10.1099/00221287-144-8-2225

Ritchie, J.M., Greenwich, J.L., Davis, B.M., Bronson, R.T., Gebhart, D., Williams, S.R., Martin, D., Scholl, D., Waldor, M.K., 2011. An Escherichia coli O157-Specific Engineered Pyocin Prevents and Ameliorates Infection by *E. coli* O157:H7 in an Animal Model of Diarrheal Disease. Antimicrobial Agents and Chemotherapy 55, 5469–5474. https://doi.org/10.1128/AAC.05031-11

Rohwer, F., Edwards, R., 2002. The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. Journal of Bacteriology 184, 4529–4535. https://doi.org/10.1128/JB.184.16.4529-4535.2002

Salis, H.M., Mirsky, E.A., Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27, 946–950. https://doi.org/10.1038/nbt.1568

Sampaio, M., Rocha, M., Oliveira, H., Dias, O., 2019. Predicting promoters in phage genomes using PhagePromoter. Bioinformatics 35, 5301–5302. https://doi.org/10.1093/bioinformatics/btz580

Saviola, B., 2004. Method to integrate multiple plasmids into the mycobacterial chromosome. Nucleic Acids Research 32, 11e–111. https://doi.org/10.1093/nar/gnh005

Scanlan, P.D., 2017. Bacteria–Bacteriophage Coevolution in the Human Gut: Implications for Microbial Diversity and Functionality. Trends in Microbiology 25, 614–623. https://doi.org/10.1016/j.tim.2017.02.012

Schooley, R.T., Biswas, B., Gill, J.J., Hernandez-Morales, A., Lancaster, J., Lessor, L., Barr, J.J., Reed, S.L., Rohwer, F., Benler, S., Segall, A.M., Taplitz, R., Smith, D.M., Kerr, K., Kumaraswamy, M., Nizet, V., Lin, L., McCauley, M.D., Strathdee, S.A., Benson, C.A., Pope, R.K., Leroux, B.M., Picel, A.C., Mateczun, A.J., Cilwa, K.E., Regeimbal, J.M., Estrella, L.A., Wolfe, D.M., Henry, M.S., Quinones, J., Salka, S., Bishop-Lilly, K.A., Young, R., Hamilton, T., 2017. Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant *Acinetobacter baumannii* Infection. Antimicrob. Agents Chemother. 61, e00954-17, /aac/61/10/e00954-17.atom. https://doi.org/10.1128/AAC.00954-17

Schrödinger, L., n.d. PyMOL | pymol.org. The PyMOL Molecular Graphics System.

Schwanbeck, J., Riedel, T., Laukien, F., Schober, I., Oehmig, I., Zimmermann, O., Overmann, J., Groß, U., Zautner, A.E., Bohne, W., 2019. Characterization of a clinical *Clostridioides difficile* isolate with markedly reduced fidaxomicin susceptibility and a V1143D mutation in rpoB. J Antimicrob Chemother 74, 6–10. https://doi.org/10.1093/jac/dky375

Schwemmlein, N., Pippel, J., Gazdag, E.-M., Blankenfeldt, W., 2018. Crystal Structures of R-Type Bacteriocin Sheath and Tube Proteins CD1363 and CD1364 From *Clostridium difficile* in the Pre-assembled State. Frontiers in Microbiology 9.

Seco, E.M., Zinder, J.C., Manhart, C.M., Lo Piano, A., McHenry, C.S., Ayora, S., 2013. Bacteriophage SPP1 DNA replication strategies promote viral and disable host replication in vitro. Nucleic Acids Research 41, 1711–1721. https://doi.org/10.1093/nar/gks1290

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. https://doi.org/10.1128/mBio.00019-20

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. Genes 11, 458. https://doi.org/10.3390/genes11040458

Shayesteh, F., Ahmad, A., Usup, G., 2020. In vitro anti-biofilm activity of bacteriocin from a marine Bacillus sp. strain Sh10 against *Proteus mirabilis*. Iran J Microbiol 12, 52–61.

Shehab, N., Patel, P.R., Srinivasan, A., Budnitz, D.S., 2008. Emergency department visits for antibiotic-associated adverse events. Clin. Infect. Dis. 47, 735–743. https://doi.org/10.1086/591126

Shrestha, R., Cochran, A.M., Sorg, J.A., 2019. The requirement for co-germinants during *Clostridium difficile* spore germination is influenced by mutations in yabG and cspA. PLOS Pathogens 15, e1007681. https://doi.org/10.1371/journal.ppat.1007681

Siddell, S.G., Walker, P.J., Lefkowitz, E.J., Mushegian, A.R., Dutilh, B.E., Harrach, B., Harrison, R.L., Junglen, S., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Nibert, M.L., Rubino, L., Sabanadzovic, S., Simmonds, P., Varsani, A., Zerbini, F.M., Davison, A.J., 2020. Binomial nomenclature for virus species: a consultation. Arch Virol 165, 519–525. https://doi.org/10.1007/s00705-019-04477-6

Sorg, J.A., Sonenshein, A.L., 2009. Chenodeoxycholate Is an Inhibitor of *Clostridium difficile* Spore Germination. Journal of Bacteriology 191, 1115–1117. https://doi.org/10.1128/JB.01260-08

Spigaglia, P., 2016. Recent advances in the understanding of antibiotic resistance in *Clostridium difficile* infection. Therapeutic Advances in Infection 3, 23–42. https://doi.org/10.1177/2049936115622891

Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A., Lewis, M., 2008. Crystal structure of the lambda repressor and a model for pairwise cooperative operator binding. Nature 452, 1022–1025. https://doi.org/10.1038/nature06831

Stern, A., Sorek, R., 2011. The phage-host arms-race: Shaping the evolution of microbes. Bioessays 33, 43–51. https://doi.org/10.1002/bies.201000071

Stokes, J.M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N.M., MacNair, C.R., French, S., Carfrae, L.A., Bloom-Ackermann, Z., Tran, V.M., Chiappino-Pepe, A., Badran, A.H., Andrews, I.W., Chory, E.J., Church, G.M., Brown, E.D., Jaakkola, T.S., Barzilay, R., Collins, J.J., 2020. A Deep Learning Approach to Antibiotic Discovery. Cell 180, 688-702.e13. https://doi.org/10.1016/j.cell.2020.01.021

Storms, Z.J., Teel, M.R., Mercurio, K., Sauvageau, D., 2020. The Virulence Index: A Metric for Quantitative Analysis of Phage Virulence. PHAGE. https://doi.org/10.1089/phage.2019.0001

St-Pierre, F., Cui, L., Priest, D.G., Endy, D., Dodd, I.B., Shearwin, K.E., 2013. One-Step Cloning and Chromosomal Integration of DNA. ACS Synth. Biol. 2, 537–541. https://doi.org/10.1021/sb400021j

Streett, H.E., Kalis, K.M., Papoutsakis, E.T., 2019. A Strongly Fluorescing Anaerobic Reporter and Protein-Tagging System for *Clostridium* Organisms Based on the Fluorescence-Activating and Absorption-Shifting Tag Protein (FAST). Appl Environ Microbiol 85. https://doi.org/10.1128/AEM.00622-19

Suh, G.A., Lodise, T.P., Tamma, P.D., Knisely, J.M., Alexander, J., Aslam, S., Barton, K.D., Bizzell, E., Totten, K.M.C., Campbell, J.L., Chan, B.K., Cunningham, S.A., Goodman, K.E., Greenwood-Quaintance, K.E., Harris, A.D., Hesse, S., Maresso, A., Nussenblatt, V., Pride, D., Rybak, M.J., Sund, Z., van Duin, D., Van Tyne, D., Patel, R., for the Antibacterial Resistance Leadership Group, 2022. Considerations for the Use of Phage Therapy in Clinical Practice. Antimicrobial Agents and Chemotherapy 66, e02071-21. https://doi.org/10.1128/aac.02071-21

Summers, W.C., 2012. The strange history of phage therapy. Bacteriophage 2, 130–133. https://doi.org/10.4161/bact.20757

Suzuki, Y., Morino, M., Morita, I., Ohiro, S., 2022. Comparison of two alcohol hand rubbing techniques regarding hand surface coverage among hospital workers: a quasi-randomized controlled trial. Antimicrobial Resistance & Infection Control 11, 132. https://doi.org/10.1186/s13756-022-01172-1

Swift, B.M.C., Meade, N., Barron, E.S., Bennett, M., Perehenic, T., Hughes, V., Stevenson, K., Rees, C.E.D., 2020. The development and use of Actiphage® to detect viable mycobacteria from bovine tuberculosis and Johne's disease-infected animals. Microb Biotechnol 13, 738–746. https://doi.org/10.1111/1751-7915.13518

Tannock, G.W., Munro, K., Taylor, C., Lawley, B., Young, W., Byrne, B., Emery, J., Louie, T., 2010. A new macrocyclic antibiotic, fidaxomicin (OPT-80), causes less alteration to the bowel microbiota of *Clostridium difficile*-infected patients than does vancomycin. Microbiology 156, 3354–3359. https://doi.org/10.1099/mic.0.042010-0

Tao, P., Wu, X., Tang, W.-C., Zhu, J., Rao, V., 2017. Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. ACS Synth. Biol. 6, 1952–1961. https://doi.org/10.1021/acssynbio.7b00179

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., Ostell, J.,

2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44, 6614–6624. https://doi.org/10.1093/nar/gkw569

Teasley, DavidG., Olson, MaryM., Gebhard, RogerL., Gerding, DaleN., Peterson, LanceR., Schwartz, MichaelJ., Lee, JamesT., 1983. PROSPECTIVE RANDOMISED TRIAL OF METRONIDAZOLE VERSUS VANCOMYCIN FOR *CLOSTRIDIUM-DIFFICILE*-ASSOCIATED DIARRHOEA AND COLITIS. The Lancet, Originally published as Volume 2, Issue 8358 322, 1043–1046. https://doi.org/10.1016/S0140-6736(83)91036-X

Temperton, B., 2021. Hunting for “holes of hope.” RSB.

The UniProt Consortium, 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research 49, D480–D489. https://doi.org/10.1093/nar/gkaa1100

Thieffry, D., Thomas, R., 1995. Dynamical behaviour of biological regulatory networks—II. Immunity control in bacteriophage lambda. Bltn Mathcal Biology 57, 277–297. https://doi.org/10.1007/BF02460619

Thong, W.L., Zhang, Y., Zhuo, Y., Robins, K.J., Fyans, J.K., Herbert, A.J., Law, B.J.C., Micklefield, J., 2021. Gene editing enables rapid engineering of complex antibiotic assembly lines. Nat Commun 12, 6872. https://doi.org/10.1038/s41467-021-27139-1

Touchon, M., Moura de Sousa, J.A., Rocha, E.P., 2017. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. Current Opinion in Microbiology, Mobile genetic elements and HGT in prokaryotes * Microbiota 38, 66–73. https://doi.org/10.1016/j.mib.2017.04.010

Tseng, C.-H., Wu, C.-Y., 2019. The gut microbiome in obesity. Journal of the Formosan Medical Association, Gut microbiome in human health and diseases 118, S3–S9. https://doi.org/10.1016/j.jfma.2018.07.009

Turner, D., Adriaenssens, E.M., Tolstoy, I., Kropinski, A.M., 2021a. Phage Annotation Guide: Guidelines for Assembly and High-Quality Annotation. PHAGE 2, 170–182. https://doi.org/10.1089/phage.2021.0013

Turner, D., Kropinski, A.M., Adriaenssens, E.M., 2021b. A Roadmap for Genome-Based Phage Taxonomy. Viruses 13, 506. https://doi.org/10.3390/v13030506

UKHSA, 2022. *C*. *difficile* infection: monthly data by prior trust exposure [WWW Document]. GOV.UK. URL https://www.gov.uk/government/statistics/c-difficile-infection-monthly-data-by-prior-trust-exposure (accessed 7.21.22).

Ünlü, G., 2020. The Prevalence, Transmission, and Control of C. difficile in Food.

U.S. Food and Drug Administration, 2023. FDA Approves First Orally Administered Fecal Microbiota Product for the Prevention of Recurrence of Clostridioides difficile Infection [WWW Document]. FDA. URL https://www.fda.gov/news-events/press-announcements/fda-approves-first-orally-administered-fecal-microbiota-product-prevention-recurrence-clostridioides (accessed 4.27.23).

U.S. Food and Drug Administration, 2022. Summary basis for regulatory action - REBYOTA.

Usui, M., Kawakura, M., Yoshizawa, N., San, L.L., Nakajima, C., Suzuki, Y., Tamura, Y., 2017. Survival and prevalence of *Clostridium difficile* in manure compost derived from pigs. Anaerobe 43, 15–20. https://doi.org/10.1016/j.anaerobe.2016.11.004

Van den Bossche, A., Ceyssens, P.-J., De Smet, J., Hendrix, H., Bellon, H., Leimer, N., Wagemans, J., Delattre, A.-S., Cenens, W., Aertsen, A., Landuyt, B., Minakhin, L., Severinov, K., Noben, J.-P., Lavigne, R., 2014. Systematic Identification of Hypothetical Bacteriophage Proteins

Targeting Key Protein Complexes of *Pseudomonas aeruginosa*. J. Proteome Res. 13, 4446–4456. https://doi.org/10.1021/pr500796n

van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E.G., de Vos, W.M., Visser, C.E., Kuijper, E.J., Bartelsman, J.F.W.M., Tijssen, J.G.P., Speelman, P., Dijkgraaf, M.G.W., Keller, J.J., 2013. Duodenal Infusion of Donor Feces for Recurrent *Clostridium difficile*. New England Journal of Medicine 368, 407–415. https://doi.org/10.1056/NEJMoa1205037

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., Velankar, S., 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research 50, D439–D444. https://doi.org/10.1093/nar/gkab1061

Verma, R., Swift, B.M.C., Handley-Hartill, W., Lee, J.K., Woltmann, G., Rees, C.E.D., Haldar, P., 2020. A Novel, High-sensitivity, Bacteriophage-based Assay Identifies Low-level *Mycobacterium tuberculosis* Bacteremia in Immunocompetent Patients With Active and Incipient Tuberculosis. Clinical Infectious Diseases ciz548. https://doi.org/10.1093/cid/ciz548

Vohradsky, J., 2017. Lambda phage genetic switch as a system with critical behaviour. Journal of Theoretical Biology 431, 32–38. https://doi.org/10.1016/j.jtbi.2017.07.024

von Wintersdorff, C.J.H., Penders, J., van Niekerk, J.M., Mills, N.D., Majumder, S., van Alphen, L.B., Savelkoul, P.H.M., Wolffs, P.F.G., 2016. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00173

Walk, S.T., Micic, D., Jain, R., Lo, E.S., Trivedi, I., Liu, E.W., Almassalha, L.M., Ewing, S.A., Ring, C., Galecki, A.T., Rogers, M.A.M., Washer, L., Newton, D.W., Malani, P.N., Young, V.B., Aronoff, D.M., 2012. *Clostridium difficile* Ribotype Does Not Predict Severe Infection. Clinical Infectious Diseases 55, 1661–1668. https://doi.org/10.1093/cid/cis786

Wang, W., Wang, E., Balthasar, J., 2008. Monoclonal Antibody Pharmacokinetics and Pharmacodynamics. Clinical Pharmacology & Therapeutics 84, 548–558. https://doi.org/10.1038/clpt.2008.170

Weese, J.S., 2010. *Clostridium difficile* in food--innocent bystander or serious threat? Clin. Microbiol. Infect. 16, 3–10. https://doi.org/10.1111/j.1469-0691.2009.03108.x

Węgrzyn, G., Muniesa, M., 2021. Editorial: Shiga Toxin-Converting Bacteriophages. Frontiers in Microbiology 12.

Weigel, C., Seitz, H., 2006. Bacteriophage replication modules. FEMS Microbiol Rev 30, 321–381. https://doi.org/10.1111/j.1574-6976.2006.00015.x

Weinbauer, M.G., 2004. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. 28, 127–181. https://doi.org/10.1016/j.femsre.2003.08.001

Wendling, C.C., Refardt, D., Hall, A.R., 2021. Fitness benefits to bacteria of carrying prophages and prophage-encoded antibiotic-resistance genes peak in different environments. Evolution 75, 515–528. https://doi.org/10.1111/evo.14153

Whittle, M., 2018. Isolation and characterisation of four novel bacteriophages infecting clinically relevant PCR ribotypes of *Clostridium difficile* (Thesis (University of Nottingham only)).

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. Microbiol Spectr 10, e0229521. https://doi.org/10.1128/spectrum.02295-21

WHO, 2019. Critically important antimicrobials for human medicine.

WHO, 2015. Global action plan on antimicrobial resistance.

Woods, C., Humphreys, C.M., Rodrigues, R.M., Ingle, P., Rowe, P., Henstra, A.M., Köpke, M., Simpson, S.D., Winzer, K., Minton, N.P., 2019. A novel conjugal donor strain for improved DNA transfer into Clostridium spp. Anaerobe 59, 184–191. https://doi.org/10.1016/j.anaerobe.2019.06.020

Woodworth, M.H., Hayden, M.K., Young, V.B., Kwon, J.H., 2019. The Role of Fecal Microbiota Transplantation in Reducing Intestinal Colonization With Antibiotic-Resistant Organisms: The Current Landscape and Future Directions. Open Forum Infect Dis 6, ofz288. https://doi.org/10.1093/ofid/ofz288

Xu, J., Xiang, Y., 2017. Membrane Penetration by Bacterial Viruses. Journal of Virology 91, e00162-17. https://doi.org/10.1128/JVI.00162-17

Yang, Y., Le, S., Shen, W., Chen, Q., Huang, Y., Lu, S., Tan, Y., Li, M., Hu, F., Li, Y., 2018. Antibacterial Activity of a Lytic Enzyme Encoded by *Pseudomonas aeruginosa* Double Stranded RNA Bacteriophage phiYY. Frontiers in Microbiology 9.

Yap, M.L., Klose, T., Arisaka, F., Speir, J.A., Veesler, D., Fokine, A., Rossmann, M.G., 2016. Role of bacteriophage T4 baseplate in regulating assembly and infection. Proc Natl Acad Sci U S A 113, 2654–2659. https://doi.org/10.1073/pnas.1601654113

Young, R., Wang, I.-N., Roof, W.D., 2000. Phages will out: strategies of host cell lysis. Trends in Microbiology 8, 120–128. https://doi.org/10.1016/S0966-842X(00)01705-4

Yukgehnaish, K., Rajandas, H., Parimannan, S., Manickam, R., Marimuthu, K., Petersen, B., Clokie, M.R.J., Millard, A., Sicheritz-Pontén, T., 2022. PhageLeads: Rapid Assessment of Phage Therapeutic Suitability Using an Ensemble Machine Learning Approach. Viruses 14, 342. https://doi.org/10.3390/v14020342

Yutin, N., Galperin, M.Y., 2013. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. Environmental Microbiology 15, 2631–2641. https://doi.org/10.1111/1462-2920.12173

Zhou, C., 2021. multiPhATE2 Github.

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, 4402–4404. https://doi.org/10.1093/bioinformatics/btz258

Zhu, D., Sorg, J.A., Sun, X., 2018. *Clostridioides difficile* Biology: Sporulation, Germination, and Corresponding Therapies for C. difficile Infection. Frontiers in Cellular and Infection Microbiology 8.