Deletion of prophages in *Clostridioides difficile*

submitted to the University of Hertfordshire in partial fulfilment of the requirement of the degree of Master of Science by Research

Author: Ross Kalarus

Supervisor: Dr Shan Goh

Co-Supervisor: Prof Peter Mullany

School of Life and Medicine, University of Hertfordshire

Keywords

Clostridioides difficile, Prophages, Allelic Couple Exchange, CRISPR-Cas9 and Molecular Microbiology

Submitted: Jun 2021

ABSTRACT

Clostridioides difficile (C. difficile) formally known as *Clostridium difficile,* is a Gram-positive spore forming anaerobe responsible for almost all cases of pseudomembranous colitis and is implicated in 25-33% of antibiotic associated diarrhoea. Many *C. difficile* strains are lysogens; that is, they carry bacterial viruses known as phages, which have been shown to affect bacterial cellular processes. Phages can be virulent or temperate; a temperate phage infects a bacterial cell, then integrates its genome (prophage) into the bacterial chromosome. All phages currently identified within *C. difficile* are temperate. *C. difficile* 630 (CD630) was the first isolate to be sequenced and is a reference strain. CD630 harbours two temperate phages, ΦCD630-1, and ΦCD630-2 of 55,850 bp and 49,178 bp, respectively. Both phage genomes have a similar organisation and above 60% genome nucleotide homology. To study whether prophages contribute to host virulence, persistence, and transmission in CD630, prophage deletion is a first step. Due to the large size of the prophage genomes, precise deletion is challenging. This study aims to delete a prophage of CD630.

Two plasmids, pRK6301 and pRK6302, based on pMTL-SC7315 and *codA*-based counter selection, were generated to delete ΦCD630-1 and ΦCD630-2, respectively from the CD630 genome by allele coupled exchange. Homology arms (HA) of 1kb flanking each prophage genome were cloned, such that pRK6301 and pRK6302 were ~8.8kb. Of 200 transconjugants screened, three colonies lost pRK6301 and two colonies lost pRK6302, hence were double crossover mutants. However, all double crossover mutants were found to retain the prophage integrated junctions, hence no prophage deletion occurred.

Another recombinant plasmid, pHHCD630-2sgRNA_HA, based on pMTL83151 and CRISPR-Cas9 selection of viable mutants, was generated to delete ΦCD630-2 by targeting the phage integrase gene. The single guide RNA (sgRNA) was constitutively expressed, while Cas9 was inducible by anhydrotetracycline (aTC). The same homology arm from pRK6302 was cloned, such that pHHCD630-2sgRNA_HA was ~11.9 kb. Nine *C. difficile* transconjugant colonies containing either pHHCD630-2sgRNA_HA, control plasmid without HA (pHH_sgRNA), or control empty plasmid (pHH) were screened by PCR. All 9 colonies contained pHHCD630-2sgRNA_HA. Six colonies were confirmed for carrying pHH_sgRNA plasmid (without HA), and nine colonies contained pHH.

A transconjugant of pHHCD630-2sgRNA_HA was induced with 30 ng/mL of aTC and screened for viable mutants. Five colonies were negative for the phage integrase gene by colony PCR, suggesting the ΦCD630-2 prophage or integrase gene was lost. However, the colonies remained resistant to thiamphenicol, and all retained the plasmid, confirmed by PCR, after efforts to cure the plasmid by serial passage. Further work is needed to confirm their identity. In conclusion, there are genetic engineering tools currently available which allow precise manipulation of genomes including the deletion of prophages. This study generated plasmids and protocols for prophage deletion in CD630, however, did not confirm the deletion of a prophage.

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my thanks to my principal supervisor Dr Shan Goh for their endless support, guidance, and patience throughout this project. I would also like to thank Prof Peter Mullany my secondary supervisor for their guidance and reassurance at key moments throughout this project and essential materials used in this project. Their support made the entire project possible and has helped my transition into a better researcher.

I would also like to thank my colleagues and Microbiology lab members especially Amer, Graham, Dominik and Saman for their general assistance and encouragement whenever I was feeling overwhelmed and their experience whenever discussing ideas.

I would also like to thank my family and friends for their never-ending encouragement, patience and support throughout this project. Without them I would have never reached my goals.

Finally, I would like to extend my thanks to Mansukhlal Vadalia and Andrew Timms as well as the rest of the laboratory technical staff for their endless assistance during the project period and acknowledge the University of Hertfordshire for allowing me access to their facilities which enabled me to complete this work.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

5-FC	-	5-fluorocytosine
aTC	-	Anhydrotetracycline
BHI	-	Brain Heart Infusion
BHIA	-	BHI agar
BHIB	-	BHI broth
BHIS	-	BHI supplemented with L-cysteine and yeast extract
CC	-	Cycloserine and Cefoxitin
CD630	-	*C. difficile* 630
CDMM	-	*C. difficile* minimal media
Cm	-	Chloramphenicol
Cpf1	-	CRISPR from *Prevotella* and *Francisella* 1
CRISPR	-	Clustered regularly interspaced short palindromic repeats
CRISPR-Cas	-	CRISPR associated system
DSBs	-	Double-strand breaks
h	-	Hours
kb	-	Kilobases
KO	-	Knock-out
LHA	-	Left homology arm
LHS	-	Left hand side prophage flanking region
min	-	Minute
PAM	-	Palindromic adjacent motif
PCR	-	Polymerase chain reaction
RHA	-	Right homology arm
RHS	-	Right hand side prophage flanking region
RNP	-	Ribonucleoprotein
s	-	Seconds
sgRNA	-	Small guide RNA
SoC	-	Super optimal broth
SoM	-	Stable outgrowth media
TA	-	Toxin-Antitoxin
Ta	-	Annealing temperature
Tm	-	Thiamphenicol
tracrRNA	-	Trans activating crRNA

1. Introduction

Clostridioides difficile (C. difficile), formally known as *Clostridium difficile,* is a Gram-positive spore forming anaerobe responsible for almost all cases of pseudomembranous colitis and is implicated in 25-33% of antibiotic associated diarrhoea (Banawas, 2018). The vegetative form of *C. difficile* is unable to survive in an aerobic environment, sporulation is therefore an important part of the life cycle.

Sporulation in *C. difficile* is initiated by a variety of signals or environmental stimuli, such as quorum sensing or nutrient limitation. Spores are the clostridial form of persistence to an oxygen rich environment, exposure to oxygen leads to a build-up of reactive oxygen species which ultimately lead to cell death. The exposure to oxygen leads to physiological changes within the cell, including significant changes to gene expression and metabolic pathways which includes the induction of the sporulation pathway (Morvan, Folgosa, Kint, Teixeira, & Martin-Verstraete, 2021). Sporulation is regulated by the master regulator protein, Spo0A (Higgins & Dworkin, 2012). In CD630, the spores have an electron-dense compact outermost exosporium layer with hair-like projections, and are metabolically dormant, intrinsically resistant to antibiotics, attacks from the immune system, and resistant to bleach-free disinfectants commonly used in hospital settings. *C. difficile* spores germinate in response to the combination of specific bile salts such as cholate and its derivatives, as well as L-glycine acting as a co-germinant (Paredes-Sabja, Shen, & Sorg, 2014). After the spores germinate within a host, the major virulence factors of *C. difficile* are exotoxins A and B. These exotoxins have an effect on the Rho GTPases, which interfere with actin cytoskeletal signalling. Disrupting this pathway leads to a destabilised membrane and the characteristic inflammation of the epithelial seen in colitis (Shen, 2012).

Seventy-six strains of *C. difficile* have complete genome sequences, assembled and annotated (National Center for Biotechnology Information, 2021). CD630 was the first *C. difficile* isolate fully sequenced and annotated and is typically used alongside the M120 strain as reference genomes (Knight, Elliott, Chang, Perkins, & Riley, 2015). In most strains, the integrated genomic DNA of a bacteriophage (phage), phages are viruses that infect and replicate inside of a bacterial host. They are morphologically a single or double stranded piece of DNA encapsulated by proteins which is known as a ‘head’ and a ‘tail’

which is a hollow tube that binds to a cell and injects the viral DNA. Integrated phage genomes are known as prophages (Figure 1). These can be present in a variety of forms which range from inducible prophages, prophages with deleted regions, inserts and rearrangements or remnants which have lost the majority of a phage genome (Canchaya, Proux, Fournous, Bruttin, & Brüssow, 2003). Prophage integration is the result of a temperate phage DNA entering a lysogenic cycle instead of a lytic cycle after cell infection. Virulent phages will only enter the lytic cycle, as shown in Figure 1 below. However, not all prophages will integrate; some are capable of existing extrachromosomally, e.g., ΦCD38-2 (Sekulovic, Meessen-Pinard, & Fortier, 2011). It is believed that a past recombination event between the prophage and a plasmid occurred and led to a chimeric phage that can autonomously replicate as a circular plasmid (Sekulovic *et al*., 2011).

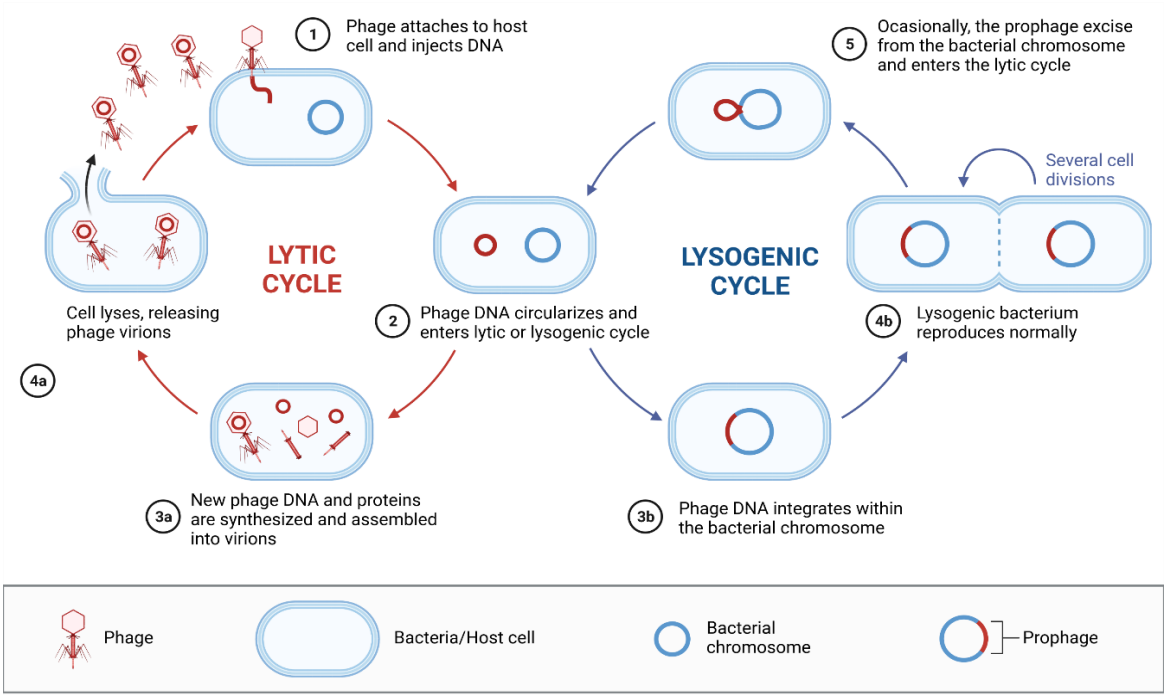


Figure 1. An overview of the lytic and lysogenic cycles which bacteriophages undergo for replication or survival. Created with BioRender.com.

For the lysogenic cycle and integration of phage genome to occur, a phage repressor inhibits transcription of phage genes required for the lytic cycle (Krebs, Goldstein, & Kilpatrick, 2018), and phage-encoded integrases mediate unidirectional site-specific recombination between two DNA recognition sequences, the phage attachment site, *attP*, and the bacterial attachment site, *attB*. The integrated prophage is then replicated concomitantly with the bacterial chromosome during cell division, until a lytic cycle is

triggered (Krebs, Goldstein, & Kilpatrick, 2018). A bacterial cell carrying one or more prophages is called a lysogen and generally, it is immune to lytic or lysogenic infection by other phages of the same group, and therefore factors into phage host range.

Phages, like any other virus, have a specific range of hosts, the range is defined by what bacterial genera, species and strains the phage is capable of infecting. Experimentally this can be identified by screening for lysis within a large collection of bacterial cultures. Bacterial strains which are sensitive to phage infection, shown by the formation of plaques, are known as indicator strains. Identifying phage indicator strains and knowing the host range of a phage is an important step in examining functional phage virions, such as for phage propagation, transduction, and therapy.

The CD630 strain harbours two temperate phages, ΦCD630-1, and ΦCD630-2 (Sebaihia *et al*., 2006; Goh *et al*., 2007; Fortier & Moineau, 2007). These two phages have different prophage genome sizes; ΦCD630-1 has a genome of 55,850 bp and ΦCD630-2 has a genome of 49,178 bp, however they both show a similar genome organisation as well as above 60% genome nucleotide homology (Sebaihia *et al*., 2006; Goh *et al*., 2007). Prophages in some *C. difficile* strains have been shown to affect virulence of the host, such as toxin production (Goh *et al*., 2007; Govind, Vediyappan, Rolfe, Dupuy, & Fralick, 2009; Sekulovic, Meessen-Pinard, & Fortier, 2011). However, there are only a few papers which have studied the two prophages in CD630 in this regard. Soutourina *et al* (2013) identified 511 small noncoding RNA sequences in the CD630 genome, some of which were found in the prophage regions, and some of which were c-di-GMP-I riboswitches that affected the transcription of host genes. Following from this Peltier *et al.* (2020) examined phage-encoded toxin/antitoxin systems contained within prophages which were shown to affect host physiology, for example by confering mobile genetic element maintance such as plasmids or the circularised phage genome.

The interactions between phages and bacteria have been fundamental in the development and evolution of bacteria. This is clear to see by studying the contributions of prophage encoded factors that affect the host bacteria, as seen in *C. difficile* R20291 which was experimentally infected with the ΦCD38-2 phage. The phage downregulated many genes involved with sugar transport and metabolism, such as the phosphotransferase systems (Sekulovic & Fortier, 2015). Another example was in *C.*

difficile strain TW11 infected with the JD032 phage, which affected 17.7% of host gene expression (Li *et al*., 2020). For example, *cwp2*, the gene responsible for a protein expressed on the cell and spore surface involved in adhesion (Bradshaw, Kirby, Roberts, Shone, & Acharya, 2017), was downregulated and increased secretion of toxin A, and a decreased capacity to adhere was recorded (Li *et al*., 2020). The ways in which CD630 phages contribute to virulence, persistence and transmission of their well characterised virulent host strain is yet to be explored. Although there is some evidence, this could be assessed once the prophages are deleted from the bacterial genome, which owing to their large size is a difficult feat.

1.1 Phage contribution to survival in bacteria

The role of phages as contributors to survival in *C. difficile* is an ongoing field of interest. Lytic phages can have a direct effect on host cell colonization, where bacterial cell lysis leads to an accumulation of extracellular DNA that is considered a key component in biofilm formation, as it binds bacteria together as a community. Phage-mediated bacterial cell lysis and subsequent DNA release would then help to establish a biofilm (Slater, Frost, Jossi, Millard, & Unnikrishnan, 2019). In addition, lysogenic phages such as ΦCDHM1 carry gene homologs of quorum sensing, hence have the potential to influence biofilm formation or induction of toxins A and B, shown to be regulated by quorum sensing (Darkoh, DuPont, Norris, & Kaplan, 2015; Hargreaves, Kropinski, & Clokie, 2014).

Phage could contribute to *C. difficile* fitness and transmission. One of the largest phage identified in *C. difficile* so far is ΦCD211, which has a 131 kilobase (kb) genome and was found to have ≥80% nucleotide identity with 149 other prophages in a 2,584 *C. difficile* strain genome screen. Garneau *et al* (2017) found that ΦCD211 has genes with the potential to affect antibiotic resistance, such as ORF8 which encodes a putative AcrB/AcrD/AcrF protein that belongs to a family of multidrug resistance transporters. It was also found that the prophage encodes a EzrA homolog. EzrA is involved in the coordination between cell growth and cell division and in the control of the cell elongation–division cycle by inhibiting the formation of aberrant FtsZ rings (Haeusser, Schwartz, Smith, Oates & Levin, 2004), and that ORF68 of the genome encodes a putative YyaC-like spore germination protease, both of which highlight possible interactions with host sporulation and spore germination respectively. The genomic similarity to 149 other

prophages across such a large number of other *C. difficile* strains highlights that similar genes with similar functions could be carried by multiple prophages across the species.

In the CD630 genome, ΦCD630-2 prophage is integrated into a predicted ABC-type transporter gene, which is normally negatively regulated by the Spo0A protein in wild type CD630. The gene is disrupted by the ΦCD630-2 prophage sequence and has a low level of expression, however in a CD630*Δerm spo0A*::*ermB* mutant the transporter gene was upregulated, indicating that the gene could have a negative effect on sporulation (Pettit *et al*., 2014). Therefore, if the prophage genome was deleted and the transporter gene sequence restored, the ability of CD630 to sporulate could be reduced, which could then affect persistence of the bacterium in hostile conditions.

Considering the potential interactions of phage genomes and the host bacteria, as well as the limited information regarding this subject, it highlights a particular need to track how prophages contribute to their host, using CD630 and its prophages as a model. The most efficient method for understanding the contributions of the separate prophages is to delete them individually and then compare these mutant isolates to a wild type and record any differences.

1.2 Phage excision techniques

There are multiple strategies for deleting a prophage from a bacterial genome. These techniques include curing phages from bacteria by exposing lysogens to mutagens, such as UV or chemicals, however there are limitations with random approaches for deletion. One limitation is that these techniques are not selective for individual phages within a genome, and considering most bacteria have multiple phages integrated into the chromosome this would make tracking and evaluating the loss extremely difficult. Aside from this is the potential to generate unknown and undesirable mutations in the bacterial chromosome using mutagens. Hence, precise techniques based on homologous recombination for gene deletion are preferred. Allelic-coupled exchange plasmids have been a regular tool for genomic editing, such as deleting a gene or a prophage from the genome of the host bacteria (Ng *et al*., 2013).

This system relies on crossover events that result in the uptake by selective single crossover, and then excision by double crossover. Single-crossover events can be mediated by either a left homology arm (LHA) or right homology arm (RHA), which are sequences homologous to regions flanking the target prophage sequence, known as the knock-out (KO) cassette. The crossover event occurs as the regions interact and the plasmid sequence is folded into the chromosome. Selection by antibiotic pressure is how successful single crossover integrants can be recovered, plasmids containing genes such as *catP* which encodes chloramphenicol acetyltransferase, provide resistance to antibiotics and so in theory only those cells which have taken up the plasmid will survive (Figure 2).

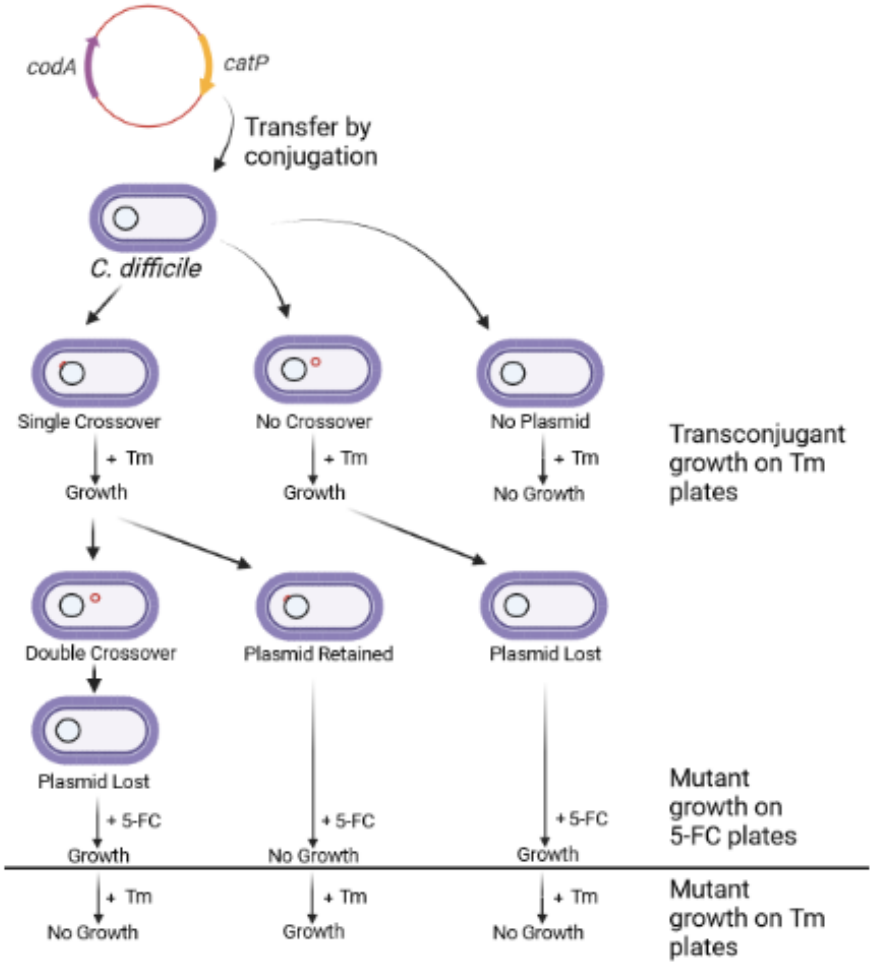


Figure 2. Selection of transconjugants and mutants. Cells containing plasmids bearing resistance genes will survive exposure to antibiotics, for example *catP* (yellow arrow) providing resistance against thiamphenicol (Tm), those cells which then lose the plasmid bearing the *codA* gene (purple arrow) will survive exposure to 5-fluorocytosine (5-FC) as the compound will not be hydrolysed to a toxin.

Double crossovers are rare events where the homologous regions interact once again, however in this case the two regions are exchanged and the nucleotide sequence between the two regions are lost in an excision event. The selection of double crossover events can be based on cell viability after the event has occurred. Genes such as *codA*, which encodes for cytosine deaminase that hydrolyses 5-fluorocytosine into the cytotoxic compound 5-fluorouracil, allows for selection against cells which maintain the plasmid, as they will die in the presence of 5-fluorouracil whereas those which have lost the plasmid will survive (Figure 2). The insertion point changes depending on the arm mediating it, as shown in Figures 3 and 4. Fig 3A.1 shows the insertion mediated by the LHA and 3A.2 shows the cassette inserted upstream of the prophages genome's integrase gene. Fig 3B shows the deletion if mediated by the RHA, where the prophage would excise along with the plasmid bearing the KO cassette.

Figure 4 shows the integration of the KO cassette targeting ΦCD630-2 mediated by the RHA where the cassette would integrate downstream of the integrase gene in the prophage genome. The integration of the cassette can be mediated by either arm when selecting for ΦCD630-2 as with ΦCD630-1. Figure 4A has the RHA mediated integration integrating the cassette downstream of the prophage genome, while Figure 4B shows the excision of the prophage genome when mediated by the LHA.

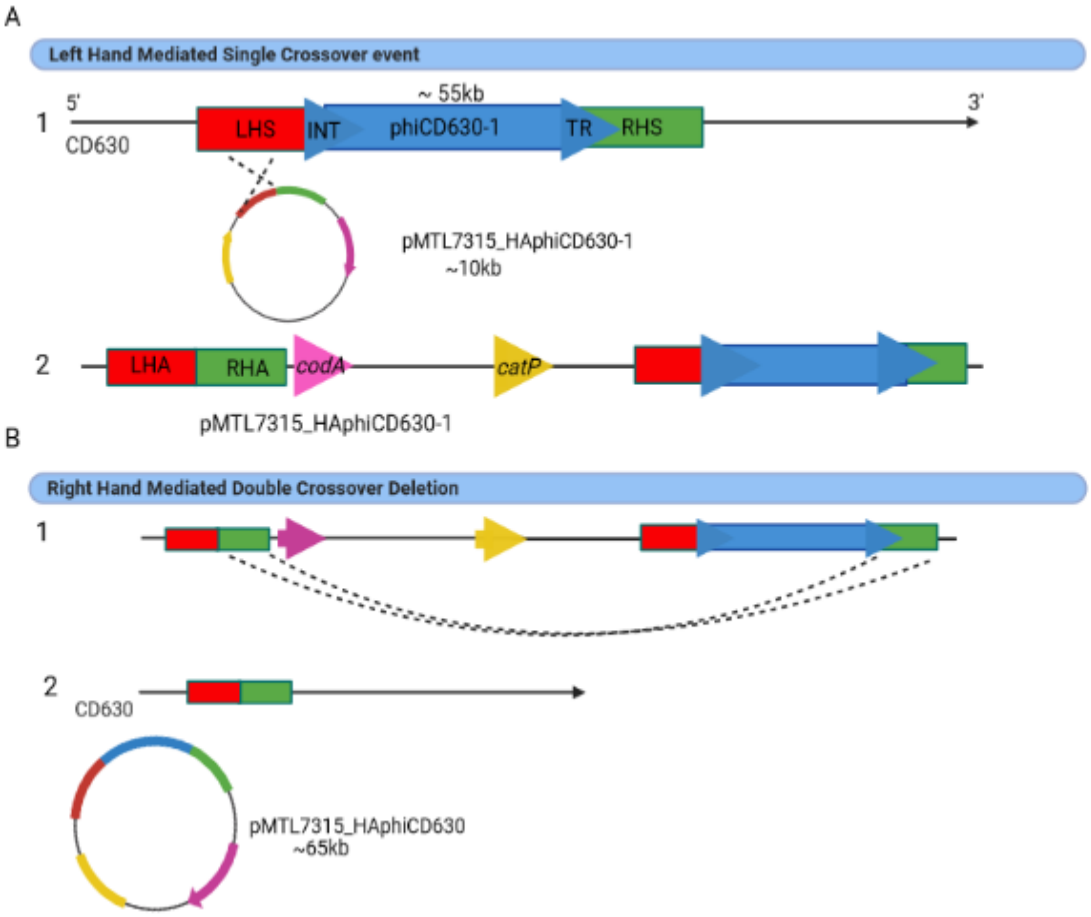


Figure 3. Possible integration and deletion events for a plasmid bearing homologous arms flanking ΦCD630-1. ΦCD630-1 prophage (blue block), INT= integrase (blue arrowhead), TR= transcriptional regulator (blue arrowhead), *catP*= chloramphenicol acetyltransferase gene (yellow block arrow) and *codA*= cytosine deaminase gene (purple block arrow). (A) This is the integration event if mediated by the LHA (red block). (B) This is the deletion if mediated by the RHA (green block). Created with BioRender.com

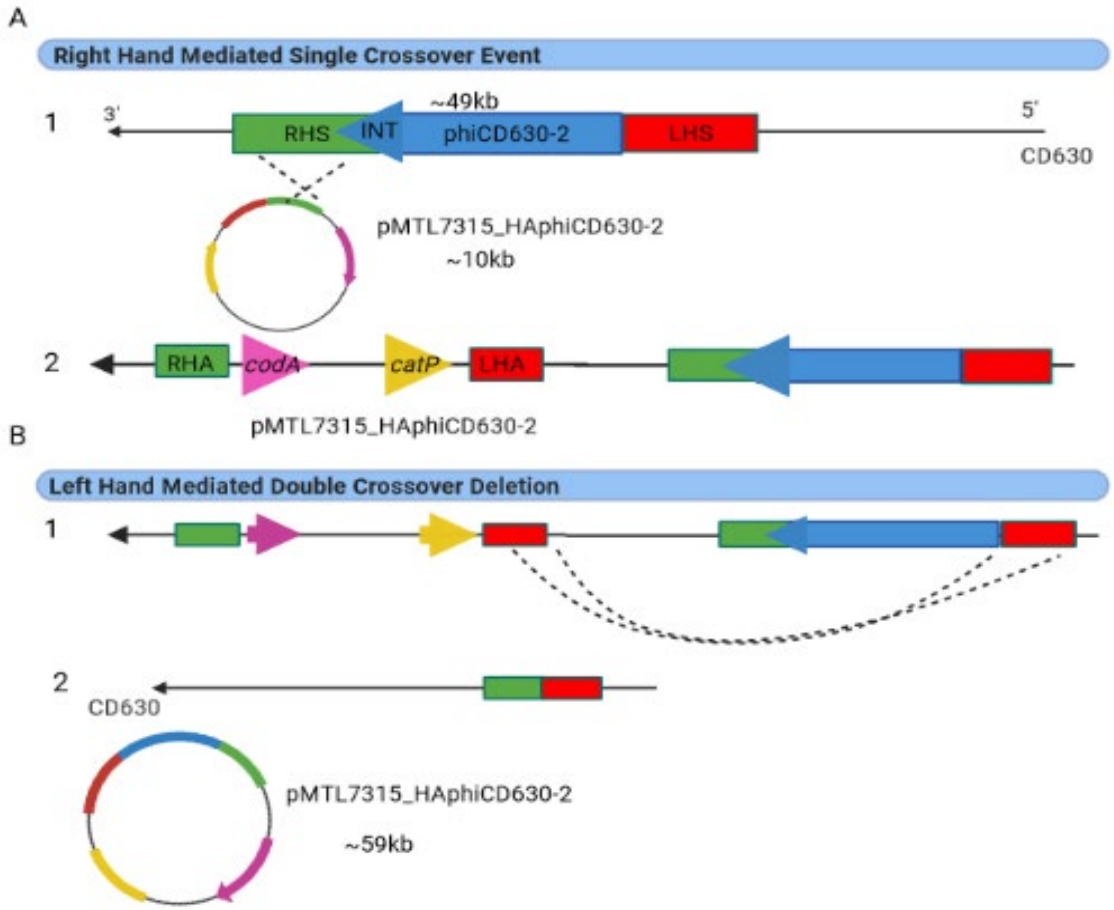


Figure 4. Possible integration and deletions events for a plasmid bearing homologous arms flanking ΦCD630-2. The ΦCD630-2 prophage genome (blue block) is integrated in reverse complement within the CD630 genome. INT= integrase (blue arrowhead), *catP*= chloramphenicol acetyltransferase gene (yellow block arrow) and *codA*= cytosine deaminase gene (purple block arrow). (A) This is the integration event if mediated by the RHA (green block). (B) This is the deletion if mediated by the LHA (red block). Created with BioRender.com.

The deletion events shown in the Figures 3 and 4 are not the only possible outcomes of a double crossover event; it is possible for the same arms that mediated integration to cross again which would lead to the excision of the plasmid sequence without the prophage genome which is known as reversion.

The deletion of genes and prophages using this method has been done in many bacterial species such as *Enterococcus faecalis*, *Staphylococcus aureus*, *Escherichia coli* and *Vibrio natriegens* (Bae, Baba, Hiramatsu, & Schneewind, 2006; Matos *et al*., 2013; Pfeifer *et al*., 2019; Thomason *et al*., 2007) – although prophage genomes tend to be very large genetic fragments that are usually outside the limits of these systems. In *E. faecalis* the deletion

of prophages highlighted the importance of phage-encoded PBL proteins, which are important for bacteria-platelet adhesion in the beginning step of infective endocarditis (Bae, Baba, Hiramatsu, & Schneewind, 2006; Matos *et al*., 2013). Hypothetically the deletion of prophages from *C. difficile* genomes could highlight potential important phage-encoded proteins with major contributions to disease propagation, much like the deletion of prophages highlighted the phage-encoded proteins of *E. faecalis*. There are several plasmids that have been used for gene deletion in the *C. difficile* genome, including plasmids pMTL-SC7215 and pMTL-SC7315, which were used to delete the *tdcC* gene from both R20291 and CD630 strains, respectively (Cartman, Kelly, Heeg, Heap, & Minton, 2012). The TdcC protein is considered to be putative negative regulator of toxin expression as an anti-sigma factor, however it has been found to be membrane-bound with its C-terminus in an extracellular environment which disputes this action (Oliveira Paiva, de Jong, Friggen, Smits & Corver, 2019). These plasmids, and indeed most plasmids used in *C. difficile* cloning, are based on the pMTL8000 series of plasmids which are *E. coli-C. difficile* shuttle plasmids containing both a Gram-negative and a Gram-positive replicon, a gene for conjugal transfer, an antibiotic marker gene, and a multiple cloning site (Heap *et al*., 2009).

A second approach to modify a bacterial genome is to use a Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (CRISPR-Cas) system. The CRISPR system is an acquired immune response system in bacteria and archaea defending against invading phages and other mobile genetic elements that works by recognising DNA through CRISPR RNA, and DNA cleavage by Cas nucleases. DNA recognition is achieved as foreign invading DNA is processed into small fragments which are inserted in the CRISPR loci and then act as transcriptional templates during further exposure. There are 6 CRISPR types, type I, II, III, IV, V and VI with types I, II and III the most studied in bacteria. These different types of CRISPR systems are distinguished by the different ribonucleoprotein (RNP) complexes and further by a “signature protein” that is responsible for DNA degradation which is, respectively, Cas3, Cas9, and Cas10 for the types I, II, and III (Koonin, Makarova, & Zhang, 2017; Makarova *et al*., 2011).

The type II bacterial CRISPR system of *Streptococcus pyogenes* has become highlighted in genomic engineering; this system is associated with Cas9, which is a single protein capable of generating crRNA and cleaving target DNA. In *C. difficile* the endogenous

CRISPR-Cas system is Class I subtype I-b, which utilises a complex that has helicase and DNase activity. The *C. difficile* CRISPR system has a uniquely high number of CRISPR arrays compared to other bacterial systems in the CRISPR database (Makarova *et al*., 2011; Andersen *et al*., 2016). The high number of CRISPR arrays could partially explain difficulties associated with the transfer of mobile genetic elements to *C. difficile* as the endogenous system provides a high level of protection (Andersen *et al*., 2016). There are other known difficulties such as the restriction modification system of *C. difficile* in which restriction enzymes digest the incoming DNA, this then interferes with the transfer of plasmids and other elements (Purdy *et al*., 2002).

The major advantage of using CRISPR-Cas9 is its efficiency compared to other techniques, particularly those which make use of selection-counterselection systems such as allelic-coupled exchange. The CRISPR-Cas9 system was shown to be an effective tool for deletion, integration, single nucleotide modification and desirable mutant selection in *Clostridium beijerinckii*. Deletions of 1.5 kb in the *pta* and *ack* loci of *C. beijerinckii* were recorded in (Wang *et al*., 2016). Up to 30 kb in *Streptomyces* strains (Cobb, Wang, & Zhao, 2015) has been successfully deleted using the CRISPR-Cas9 system. In *C. difficile* the *selD* gene was deleted using the CRISPR-Cas9 system, this was a 950 bp deletion from the genome of the R20291 strain, and the first use of Cas9 to delete genes in *C. difficile* (McAllister, Bouillaut, Kahn, Self, & Sorg, 2017).

The major disadvantage is that Cas9 can lead to plasmid instability, such that Cas9-carrying plasmids were not transferred or were easily lost after transfer, when expressed heterologously in bacterial species that do not have an endogenous type II system such as *C. difficile* (Hong, Zhang, Cui, Wang, & Wang, 2018; Maikova, Kreis, Boutserin, Severinov, & Soutourina, 2019). The small-guide RNA (sgRNA) essential in CRISPR systems typically guides the Cas9 to the target gene of interest adjacent to the protospacer adjacent motif (PAM), which is 2-6 nucleotides downstream of the target DNA sequence (Figure 5). The sgRNA complexes with Cas9 and induces double-strand breaks (DSBs). Usually, the repair of the DSBs and the required modification of the genome are achieved by adding homology templates to the vector bearing the CRISPR-Cas system (Figure 5). However, this does not guarantee that repair will take place and DNA damage to this degree will lead to cell death (Altenbuchner, 2016). Additionally, large nucleotide fragment deletion

has proven difficult when using a Cas9 based system. Attempts to delete >4.5 kb gene fragments were consistently unsuccessful in *C. difficile* (Hong, Zhang, Cui, Wang, & Wang, 2018).

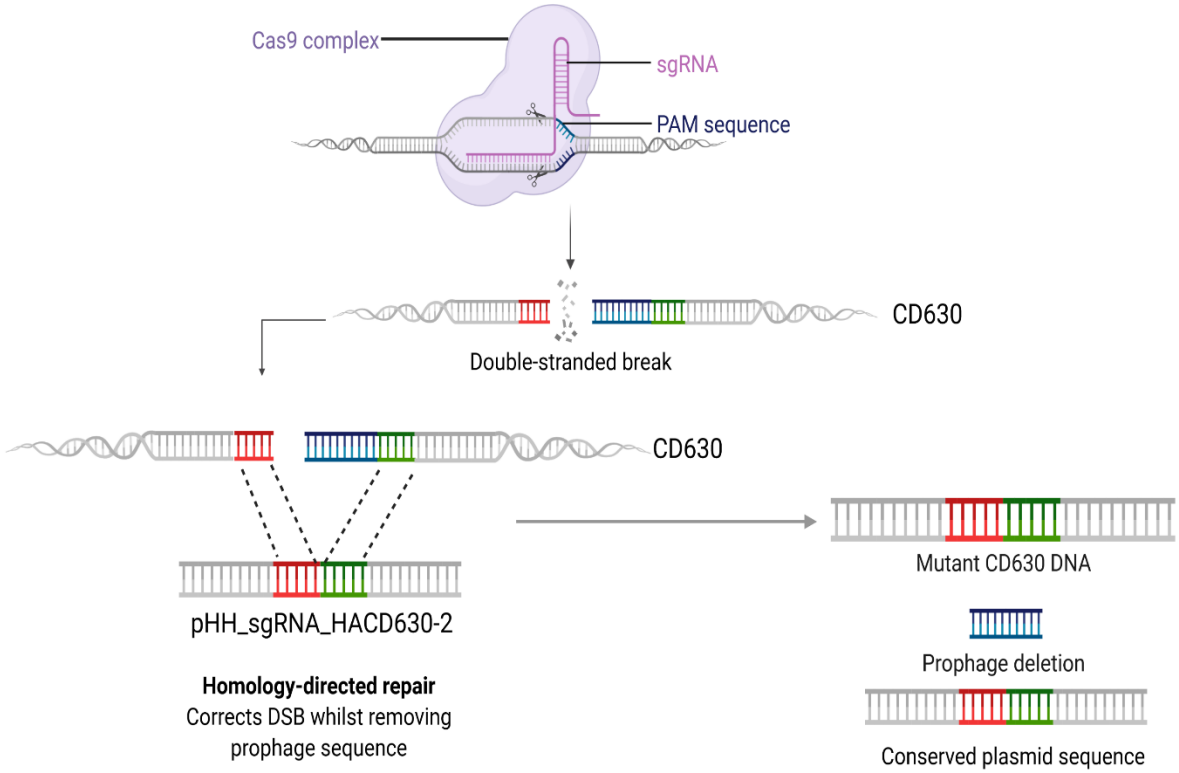


Figure 5. Prophage deletion via CRISPR-Cas9. A double stranded break is generated by sgRNA-specific Cas9 (purple bubble) cleavage, then repaired using the homology arms (red and green blocks) provided heterologously by a recombinant plasmid, leading to the loss of the prophage sequence (blue block). Created with BioRender.com.

A modification to the second approach for the excision of the prophages is to use a CRISPR-Cas12a or CRISPR from *Prevotella* and *Francisella* 1 (Cpf1) system (Fonfara, Richter, Bratoviä, le Rhun, & Charpentier, 2016). This system utilises the Cpf1 nuclease which, unlike Cas9, has highly specific dual RNA and DNA cleavage activity. Furthermore, instead of the trans-activating crRNA (tracrRNA)-Cas9 RNP complex, which recognizes a target site based on PAM sequences, the Cpf1 protein hybridises with a single crRNA to target T rich PAM sequences (Fonfara, Richter, Bratoviä, le Rhun, & Charpentier, 2016). The Cpf1 system is possibly a better approach for the excision process compared to Cas9 in *C. difficile* because it was able to delete the ΦCD630-2 prophage from CD630 (Hong, Zhang, Cui, Wang, & Wang, 2018).

A third approach is to exploit the *C. difficile* endogenous CRISPR-Cas system for genomic editing, by using plasmid vectors containing artificial CRISPR miniarrays and spacers which target a chromosomal gene. Then, crRNAs expressed from a plasmid-borne miniarray utilize the endogenous Cas machinery to form an effector complex which recognises the spacers leading to its cleavage. This was utilised to delete *hfq* gene in strains CD630Δ*erm* and R20291, which both possess a Class I subtype I-b CRISPR-Cas system (Maikova, Kreis, Boutserin, Severinov, & Soutourina, 2019). By targeting a prophage genome with vectors containing phage-specific miniarrays and spacers, the endogenous system could be utilised to delete the prophage from the host genome.

In conclusion, there are genetic engineering tools currently available which allow precise manipulation of genomes including the deletion of prophages. Using these tools would allow the study of prophage contribution to pathogenicity in bacteria and the possible options in how to expand the information about the prophages currently identified such as ΦCD630-1 and ΦCD630-2 in the CD630. Since we have observed that phages in other species contribute to the host by up or downregulating the expression of host genes, we want to delete a prophage from CD630 which should lead to a mutant with different survival capabilities and the effect will be measured by an increased or decreased expression of pathogenic factors such as a reduced ability to sporulate.

This study aimed to delete the ΦCD630-1 or ΦCD630-2 prophage genome from CD630 by a) allele-coupled exchange using the *codA*-based counter-selection method, and b) a CRISPR-Cas9 method to select for viable mutants. The first objective of both aims was to generate a single prophage deletant strain, the second objective was to identify any effects of the deletion and third objective was to generate a double deletant strain where both prophages were removed from the CD630 genome.

2. Methods and Materials

2.1 Bacterial Growth

Escherichia coli cultures were grown aerobically in LB broth or on LB agar (Fisher Bioreagents) at 37°C for 24-48 hours (h) for propagating the plasmids or conjugation with *C. difficile*. *E. coli* broth cultures were incubated with agitation at 180-200rpm. Media were supplemented with chloramphenicol (Cm) at either 12.5µg/mL in broth or 25µg/mL in agar to select for plasmids via the *catP* gene. DH5α, NEB10β and CA434 *E. coli* strains (New England Biolabs) were also grown in super optimal broth (SoC) (New England Biolabs) or stable outgrowth media (SoM) (New England Biolabs) respectively at 37°C, 150-200 rpm for 1h.

CD630 was routinely cultured on Brazier's Agar (Neogen) supplemented with 7% horse blood, cycloserine (Sigma) at 250µg/mL, and cefoxitin (Sigma) at 8µg/mL (CC) at 37°C in an anaerobic atmosphere (10% H_2, 10% CO_2, 80% N_2) for 72h, and used to prepare broth cultures in pre-reduced Brain-Heart Infusion (BHI) media (Oxoid) or BHI supplemented with L-cysteine 1g/L (Sigma-Aldrich) and yeast extract 5g/L (Oxoid) (BHIS) for 24h. Conjugative media was BHI or BHIS, followed by media supplemented with thiamphenicol (Tm) at 12.5µg/mL or 15µg/mL to select for the plasmid, and CC at 250µg/mL and 8µg/mL to select for *C. difficile*. Anhydrotetracycline (aTC, IBA Solutions) at 30 ng/mL was used when inducing the P_{tet} promoter which controls expression of *cas9*.

The defined *C. difficile* minimal medium (CDMM) prepared as previously described (Ehsaan, Kuehne, & Minton, 2016) was used for the selection of cells in which the second crossover event had taken place, supplemented with 5-fluorocytosine (5-FC, Alfa Aesar) at 50 µg/mL in the case of allelic exchange using *codA* as a negative selection marker. All strains and plasmids used in the project are shown in Table 1.

Table 1. Bacterial strains and plasmid including characteristics and sources.

Strains/Plasmids	**Characteristics**	**Source***
Strains		
CD630	Wild type; PCR-ribotype 012	(Wust & Hardegger, 1983)

E. coli DH5α C2987	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
E. coli NEB10β	*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14-* φ*80*d*lacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	
E. coli CA434	*hsd20(rB-, mB-), recA13, rpsL20, leu, proA2*, with IncPb conjugative plasmid R702	(Purdy *et al*., 2002; Williams, Young, & Young, 1990)
Plasmids		
pMTL-SC7315	*E. coli-C. difficile* shuttle vector (pCB102 *catP* ColE1 *traJ*) with *codA*	Prof. Nigel Minton (Cartman, Kelly, Heeg, Heap, & Minton, 2012)
pMTL-SC7315_ΦCD630-1HA (named pRK6301 in this study)	pMTL-SC7315 with 2kb homologous arm cassette for ΦCD630-1 flanking regions	This study
pMTL-SC7315_ΦCD630-2HA (named pRK6302 in this study)	pMTL-SC7315 with 2kb homologous arm cassette for ΦCD630-2 flanking regions	This study
PtetM,LacZ Scaf,Cas9,pMTL83151 (renamed pHH in this study)	*E. coli-C. difficile* shuttle vector (pCB102 *catP* ColE1 *traJ*) with *codA lacZ ptetM pxyl/tetO cas9*	Prof. Peter Mullany
pHH_sgRNA	pHH with addition of annealed sgRNA primer pair (f630-2gRNA2_top/ f630-2gRNA2_bot)	This study

pHH_sgRNA_HA	pHH_sgRNA with the addition of 2kb homologous arms for ΦCD630-2 flanking regions	This study

** Original reference provided where applicable.*

2.2 Allelic Couple Exchange

2.2.1 Homology arm amplification

Assembled KO cassettes comprised of LHA and RHA both approximately 1kb flanking sequences to be deleted (Table 2). To produce the LHA and RHA by PCR, primers were specifically designed using a DNA assembly tool (https://nebuilder.neb.com/#!/) to amplify the desired regions flanking the targeted prophage with overlapping sequences to the plasmid pMTL-SC7315 (Cartman *et al*., 2012) linearised at PmeI (MssI, Fastdigest, Thermo Fisher Scientific), and assembled by Gibson assembly (Figure 6). Gibson Assembly employs three enzymes, an exonuclease, a DNA polymerase and a DNA ligase. The exonuclease acts on the 5′ end of sequences and exposes the complementary sequence for annealing. The polymerase then fills in the gaps of the complementary regions and finally the DNA ligase seals and covalently links the DNA fragments. The design of these primers was carried out before the start of this study by Dr Shan Goh. All primer information is listed in Table 3.

Table 2. Prophage and homology arm locations within the CD630 chromosome (AM180355).

Subject	Location on CD630 genome			Length (bp)
	Strand	Start	End	
ΦCD630-1	Pos	1088024	1143873	55850
LHA flanking ΦCD630-1	Pos	1087024	1088023	1000
RHA flanking ΦCD630-1	Pos	1143874	1144873	1000
ΦCD630-2	Neg	3379705	3428881	49177
LHA flanking ΦCD630-2	Neg	3428882	3429881	1000
RHA flanking ΦCD630-2	Neg	3378705	3379704	1000

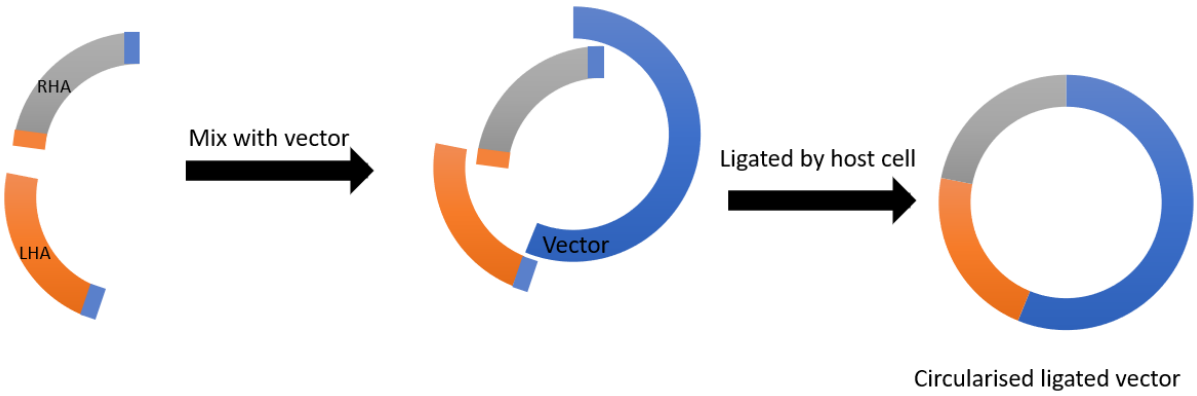


Figure 6. A simplified diagram of Gibson assembly for recombinant plasmid cloning. Purified LHA and RHA amplicons were mixed with the linearized vector backbone and underwent ligation within *E. coli* cells due to homologous regions at the end of the amplicons.

Table 3. Primer sequences. Annealing temperatures (Ta °C) for PCR using either Phusion DNA polymerase (New England Biolabs) or OneTaq DNA Polymerase (New England Biolabs) are indicated.

Primer	Sequence (5' -> 3')	Ta (Phusion) (°C)	Ta (OneTaq) (°C)
630-1LHSGA_F3	TTTTTTGTTACCCTAAGTTTGAAATTGTGTTTGAATTTAATGAGG	59.1	N/A*
630-1LHSGA_R2	GAAAGTTCTTTATTATACTTTGGAATAGTTCCAC		
630-1RHSGA_F2	AAGTATAATAAAGAACTTTCTTTTCTATCAAAAGATG	62.4	N/A
630-1RHSGA_R3	AGATTATCAAAAAGGAGTTTTAGTCTTAGAAAAAAATATCTTAGTTATTATTAAG		
630-2 LHSGA_F2	TTTTTTGTTACCCTAAGTTTTGCTAATAAAAATATTACTAAAGACATAACTAATAC	64	N/A
630-2 LHSGA_R1	TCGGACCTAGTGGATGTGGTAAAACTACATTAC		
630-2 RHSGA_F1	CATCCACTAGGTCCGAGTAAACATAGAAATTC	63.4	N/A
630-2 RHSGA_R2	AGATTATCAAAAAGGAGTTTCAGATATTATATGGGGATTATCAGC		
pMTLSC7215 seqF	GGCAATTCAAGTTCATCACGC	N/A	54
pMTLSC7215 seqR	CTTTTCTACGGGGTCTGACG		

SC7-F	GACGGATTTCACATTTGCCGTTTTGTAAACGAATTGCAGG	72	N/A
SC7-R	AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG		
630-1LHShomrec_F1	GCTATAGAGGCTGTAAAAAAATCA	60.2	N/A
phi630-1LHS_R	AGAAGAACATTGTCAGCTCCTG		
phi630-2RHS_F	GATTTAAGCAGGGTTGAACGC	62	N/A
630-2RHShomrec_R1	GGTTCAGGTTGGGACTTAGA		
f630-2gRNA2_top	GCGCCAAAATTGATAGACTAAGTCGG	N/A	N/A
f630-2gRNA2_bot	GCGCCCGACTTAGTCTATCAATTTTG		
pMTL83151bb_99	CGACTCGGTGCCACTTTT	N/A	55.4
PtetM_191	AAATATGCGGCAAGGTATTCTT		
pHH_f630-2GA_F	CAGGCTTCTTATTTTTATGGTGCTAATAAAAATATTACTAAAGACATAACTAATAC	69.4	N/A
pHH_f630-2GA_R	GTTCAAAAAAATAATGGCGGCAGATATTATATGGGGATTATCAGC		
pHHCas9_5F	AGCTTGTCGTAATAATGGCGG	N/A	53
pHHCas9_5R	TTGATACTGTGGCGGTCTGT		
phiCD630-2int_F2	ACATTGATGAAGGTTGCTCTGC	66	51
phiCD630-2int_R2	ACCAGCAGCAGTGCTAGAAT		

**N/A = Not Applicable as the oligos were not used with that polymerase in the course of the study.*

Polymerase chain reaction (PCR) was carried out to generate the homology arm amplicons using 1x Phusion Mastermix (New England Biolabs), 0.5µM of the appropriate primers, 0.2µM CD630 genomic DNA (gDNA), and nuclease-free water to a final volume of 50µL. Primer pairs used were: 630-1LHSGA_F3 and 630-1LHSGA_R2 for a ΦCD630-1 LHA amplicon; primers 630-1RHSGA_F2 and 630-1RHSGA_R3 for a ΦCD630-1 RHA amplicon; primers 630-2 LHSGA_F2 and 630-2 LHSGA_R1 for a ΦCD630-2 LHA amplicon; primers 630-2 RHSGA_F1 and 630-2 RHSGA_R2 for a ΦCD630-2 RHA amplicon. Thermocycling conditions were 98°C for 30 seconds (s), followed by 30 cycles of 98°C for 5s,

anappropriate annealing temperature in a gradient per primer pair being used for 1 minute (min), 72°C for 30s, and a final extension of 72°C for 5 min.

2.2.2 Gel extraction and quantification of PCR amplicons

PCR products were electrophoresed in 1 % agarose (Fisher BioReagents) in 1 x TBE buffer (Invitrogen) and 1 x SYBR Safe dye (Invitrogen), then visualised on a Dark Reader Blue Light Transilluminator (Labgene Scientific). DNA of interest were cut out of the gel using a clean scalpel and purified using Monarch® DNA Gel Extraction Kit following kit instructions (New England Biolabs). DNA concentration was quantified following the protocol for Microvolume Nucleic Acid Quantification Using a NanoDrop 1000 Spectrophotometer as outlined in (Desjardins & Conklin, 2010), where the elution buffer used in DNA purification was applied as a blank.

2.2.3 Restriction digestion of pMTL-SC7315

The pMTL-SC7315 vector (Figure 7) at 1µg was digested with 20U restriction enzyme PmeI (MssI, Fastdigest, Thermo Fisher Scientific), 1x Fast Digest Buffer (Thermo Fisher Scientific) and nuclease-free water to a final volume of 20 µL, incubated at 37°C for 20 min and then 65°C for 10 min to deactivate the enzyme. The linearised plasmid was then electrophoresed, excised, and quantified as described in section 2.2.2.

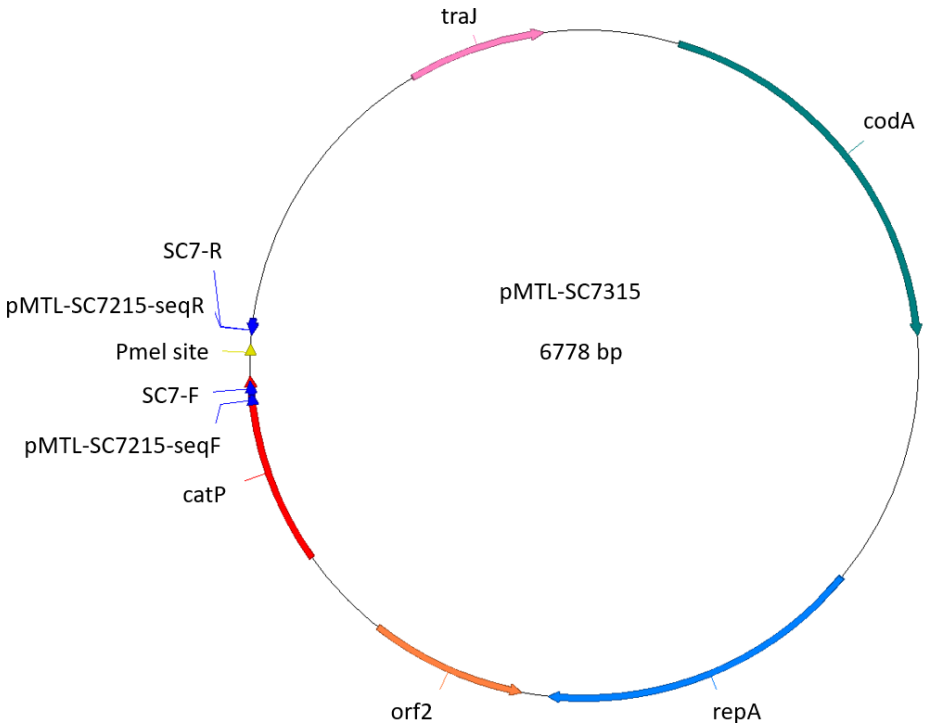


Figure 7. Schematic of pMTL-SC7315. It shows highlighted genes and primer locations represented as blue triangles where relevant.

2.2.4 Ligation of plasmid and insert

Ligation between the linearised pMTL-SC7315 and the two homology arms was performed using a vector: insert ratio of 2:1, with 150ng of plasmid and 50 ng of insert (http://www.insilico.uni-duesseldorf.de/Lig_Input.html). Linearised plasmid DNA and purified homology arm amplicons were mixed with 1X Gibson Assembly Master Mix (New England Biolabs) and nuclease-free water to a final volume of 20µL and incubated in a thermocycler at 50°C for 15 min, as recommended by the manufacturer.

2.2.5. Transformation of ligation reaction

Ligation mixtures were transformed to *E. coli* DH5α or NEB10β chemically competent cells (New England Biolabs). Cells were thawed in ice and to it was added 2µL of ligation mix from 2.2.4 and iced for 30 min. Transformation by heat shocked was carried out at 42°C for 30s exactly followed by placing in ice for 5 min. SoC medium was added (950µL) and cells were placed in an agitating incubator at 37°C, 150-200 rpm for 1h. After recovery, cells were plated as 100µL neat and a 1/10 dilution in SoC onto LB agar (Fisher BioReagents) supplemented with Cm (25µg/mL) to select for transformants and incubated at 37°C for 24h.

2.2.6. Colony PCR screening of positive *E. coli* transformants

Transformants were screened and verified by colony PCR using primers, pMTLSC7215seqF/pMTLSC7215seqR or SC7_F/SC7_R (Table 2) for positive transformants with a 2kb product, compared to negative transformants with a 250bp product. The PCR reaction of nuclease-free water, 1X One Taq Buffer (New England Biolabs), 4µM deoxynucleotide triphosphate (dNTP), 1µM each of F and R primers, 0.75U One Taq polymerase (New England Biolabs) and 10µL boiled colony in a total volume of 25µL. The empty pMTL-SC7315 was used as a PCR control which would produce a 250bp band. Cycling conditions were 94°C for 30s followed by 30 cycles of 94°C for 15s, annealing at 54°C for 15s, and 68°C for 30s with a final extension at 68°C for 5 min. PCR products were electrophoresed in a 0.8% agarose gel in 1 x TBE buffer and 1 X SYBR Safe DNA stain. PCR positive colonies were re-streaked onto fresh LB agar medium supplemented with Cm (25 µg/ml) and incubated aerobically for 24h at 37°C, for safekeeping.

2.2.7. Verification of positive *E. coli* transformants by restriction digestion

PCR positive colonies were inoculated in 5 ml LB broth supplemented with Cm (12.5 µg/ml) and incubated 16 at 37 °C, 180rpm. Plasmid DNA was extracted from 1.5 mL of the culture using Monarch® Plasmid Miniprep Kit (New England Biolabs), according to manufacturer instructions. Plasmids were verified by restriction digestion using 10U of the restriction enzyme HindIII, up to 1µg of plasmid DNA, 1X FastDigest Green Buffer (New England Biolabs) and nuclease-free water to a final volume of 20µL. The reaction mixture was incubated at 37°C for 20 min and then heat inactivated at 80°C for 10 min. The digested DNA was electrophoresed as described in 2.2.6. Finally, the verified clones that gave the correct restriction pattern were authenticated by performing Sanger nucleotide sequencing on the extracted plasmid DNA. The sanger sequencing service of Source Bioscience Ltd (UK) was used for this on which, a total of 5µL of 100ng/µL for each plasmid and 5µL of 3.2pmol/µL for each of the SC7_F/SC7_R primers were used.

2.2.8 Sequence analysis

Sanger sequence data were analysed using the DNASTAR Lasergene software. Bases were manually inspected and manually called when automated base calling was ambiguous. Following from this the forward and reverse sequencing data were analysed via ApE and combined to form a contig, which was aligned with the in silico theoretical sequence of the recombinant plasmids using Clustal-Omega (Sievers *et al*., 2011).

2.3 Conjugation of *C. difficile*

The following protocol was taken from Cartman *et al*., (2012), and Ehsaan, Kuehne, and Minton., (2016).

2.3.1 Single-Crossover Integrants

The prepared chemically competent *E. coli* donor strain CA434 was thawed on ice and transformed with 2µL of pRK6301 and pRK6302, then iced for 30 min. Transformation by heat shocked was carried out at 42°C for 45s exactly followed by placing in ice for 2 min. SoC medium at 860µL was added, and cells were incubated at 37°C, with agitation at 150 rpm for 1h. After recovery, cells were plated as 100µL neat and a 1/10 dilution in LB broth onto LB agar supplemented with Cm (25µg/mL) and incubated at 37°C for 24h to select for transformants. A positive transformant colony was grown aerobically 16-18h at 37°C, 180rpm in 5mL of LB broth supplemented with Cm (12.5µg/mL) in preparation of

conjugation. In parallel the recipient CD630 strain was grown anaerobically 16-18h in 1mL BHIS broth, at 37°C. A 1mL volume of the *E. coli* transformant was pelleted at 1500xg for 1 min, all the culture media was removed, and the pellet was washed twice in 0.5mL 1 x PBS. The cell pellet was anaerobically resuspended in 200µl of CD630, then spotted in 8 distinct droplets of 25µl onto BHIS agar for 16-18h incubation, at 37°C in anaerobic conditions. A sterile disposable loop was used to scrape all growth into 1mL of pre-reduced 1 x PBS, mixed, and 200µl of the cells was spread-plated on 5 BHIS agar supplemented with Tm (15µg/mL), cycloserine (250µg/mL), and cefoxitin (8µg/mL) (BHIS,Tm,CC) to select for growth of *C. difficile* transconjugants after anaerobic incubation at 37°C for 3 days. Putative transconjugants were re-streaked onto BHIS,Tm,CC and incubated anaerobically at 37°C. Transconjugants regrew within 1-2 days.

Visibly larger colonies were identified as more likely to be single crossover integrants which had a growth advantage over cells carrying the plasmid extrachromosomally. A random selection of these larger colonies was streaked onto BHIS,Tm and incubated anaerobically at 37°C for 24h, the clones were re-streaked onto BHIS,Tm,CC and incubated as above for safekeeping.

2.3.2 Double-Crossover Clones

Single-crossover integrants were streaked onto non-selective BHIS and incubated for 4 days at 37°C in anaerobic conditions. This allowed the cells to undergo a rare second recombination event which resulted in the excision of the plasmid integrate. All the growth was collected using a sterile loop and suspended in 500µL of 1 x PBS, then 10-fold serial dilutions in 1 x PBS to 10^{-6} were made. Each dilution was spread-plated in 100µL onto CDMM agar plates supplemented with 50µg/mL 5-FC (Alfa Aesar) and incubated for 2 days at 37°C in anaerobic conditions.

After 2 days of incubation, the 5-FC^R clones were patch plated onto both BHIS,Tm and non-selective BHIS agar plates. 5-FC^R and Tm^S colonies were inoculated in 5mL BHIS broth and anaerobically incubated for 16-18h at 37°C. Each clone was streaked onto BHI agar for safekeeping, while the 16-18h culture was used to extract gDNA, which was analysed by PCR using a flanking primer and a prophage internal primer for the intended deletion site with wild-type CD630 gDNA acting as a control. Primers to detect LHS of ΦCD630-1

were 630-1LHShomrec_F1 and phi630-1LHS_R for a 897bp amplicon in wild type revertants. Primers to detect RHS of ΦCD630-2 were phi630-2RHS_F and 630-2RHShomrec_R1, for a 598bp amplicon in wild type revertants. No amplicons were expected in mutants as the prophage internal primers would have no binding site.

PCR was carried out with 1x Phusion Master Mix, 0.5µM of appropriate forward and reverse primers, 3.05-4.085 ng gDNA and nuclease-free water in 20µL. Cycling conditions were 98°C for 30s followed by 30 cycles of 98°C for 5s, 60.2°C for ΦCD630-1 primers, or 62°C for ΦCD630-2 primers for 15s, and 72°C for 30s with a final extension at 72°C for 5 min. PCR products were electrophoresed as described in 2.2.6.

2.4 Crispr-Cas9 System

2.4.1. Guide RNA sequence prediction and selection

Cas9 sgRNA sequences were found using a CRISPR Guide tool (https://benchling.com/). DNA sequence of CD630_29520, which is the ordered locus name of the ΦCD630-2 integrase gene, was uploaded and processed with default parameters, i.e., guide length of 20 nucleotides specified, and PAM sequences restricted to NGG, where N could be any nucleotide. These parameters generated a selection of candidates with different nucleotide sequences, PAM targets, efficiency scores (Doench *et al*., 2016) and specificity scores (Hsu *et al*., 2013). The efficiency score is essentially the prediction of whether the sgRNA will be an appropriate guide to the target site for the Cas9 protein, the specificity score is essentially the prediction of whether the sgRNA contains mismatches which could lead to off-target activity. According to Doench *et al* (2016) and Hsu *et al* (2013) scores above 50 and 60 for efficiency and specificity target are considered to be good guides. A sgRNA sequence was selected for cloning based on the following parameters: it was close to the 5' end of the integrase gene, had a PAM of TGG with good efficiency and specificity scores.

Self- annealing oligos were designed to contain the sgRNA sequence within a "top" oligo (f630-2gRNA2_top) and "bottom" oligo (f630-2gRNA2_bot). Both primers had GCGCC added to the 5' end and a G added to the 3' end, such that the self-annealed oligos would have appropriate sticky ends for ligation to KasI digested ends, described below in 2.4.2.

2.4.2. Cloning of single guide RNA

2.4.2.1 Self-annealing of sgRNA oligos

Each sgRNA oligo was dissolved separately in annealing buffer (10 mM Tris pH8, 50 mM NaCl, 1 mM EDTA pH 8) to a concentration of 100µM. Following this 50µL of each oligo were mixed together and incubated at 95°C in a hot block for 5 min, the hot block was then removed from the apparatus and allowed to cool to room temperature. To check via gel electrophoresis that the oligos were successfully annealed, 10µL of the reaction and 10µL of each unannealed primer was electrophoresed on a 1.5% agarose gel prepared in 1 x TBE and 1 x SYBR Safe DNA stain for 1h at 100 volts.

To phosphorylate the annealed oligos and make ready for ligation, a reaction mixture of 0.08µM annealed oligos, 1.5U NEB T4 Polynucleotide Kinase (PNK), 1x NEB T4 PNK buffer, 1µM NEB ATP (New England Biolabs) and nuclease free water were combined to a total volume of 30µL and incubated at 37°C for 30 min, then 65°C for 20 min to inactivate the kinase. The phosphorylated oligos were then stored on ice.

2.4.2.2 Digestion of pHH with KasI

Plasmid pHH (Figure 8) was digested with KasI. The digestion reaction was prepared using 1.5 µg of pHH, 20U SspDI (KasI; Thermo Fisher Scientific), 1x Buffer Tango (Thermo Fisher Scientific) and nuclease free water to a final volume of 20µL. Digestion was carried out at 37°C for 16h, then heat-inactivated at 80°C for 20 min. The entire reaction was electrophoresed, gel extracted, and quantified as described in 2.2.2.

2.4.2.3. Ligation of pHH and sgRNA to create pHH_sgRNA

To create pHH_sgRNA, ligation of KasI digested pHH and phosphorylated sgRNA was carried out with 2.17 pmol phosphorylated sgRNA oligos, 100ng of pHH (i.e., vector: insert ration of 1:9), 1x Thermo T4 DNA ligase buffer, 10U T4 DNA ligase (Thermo Fisher Scientific), and nuclease-free water to a total volume of 20µL. The mixture was incubated at room temperature for 10-15 mins, then stored on ice before transformation into *E. coli* C2987 (New England Biolabs), a DH5α cell derivative.

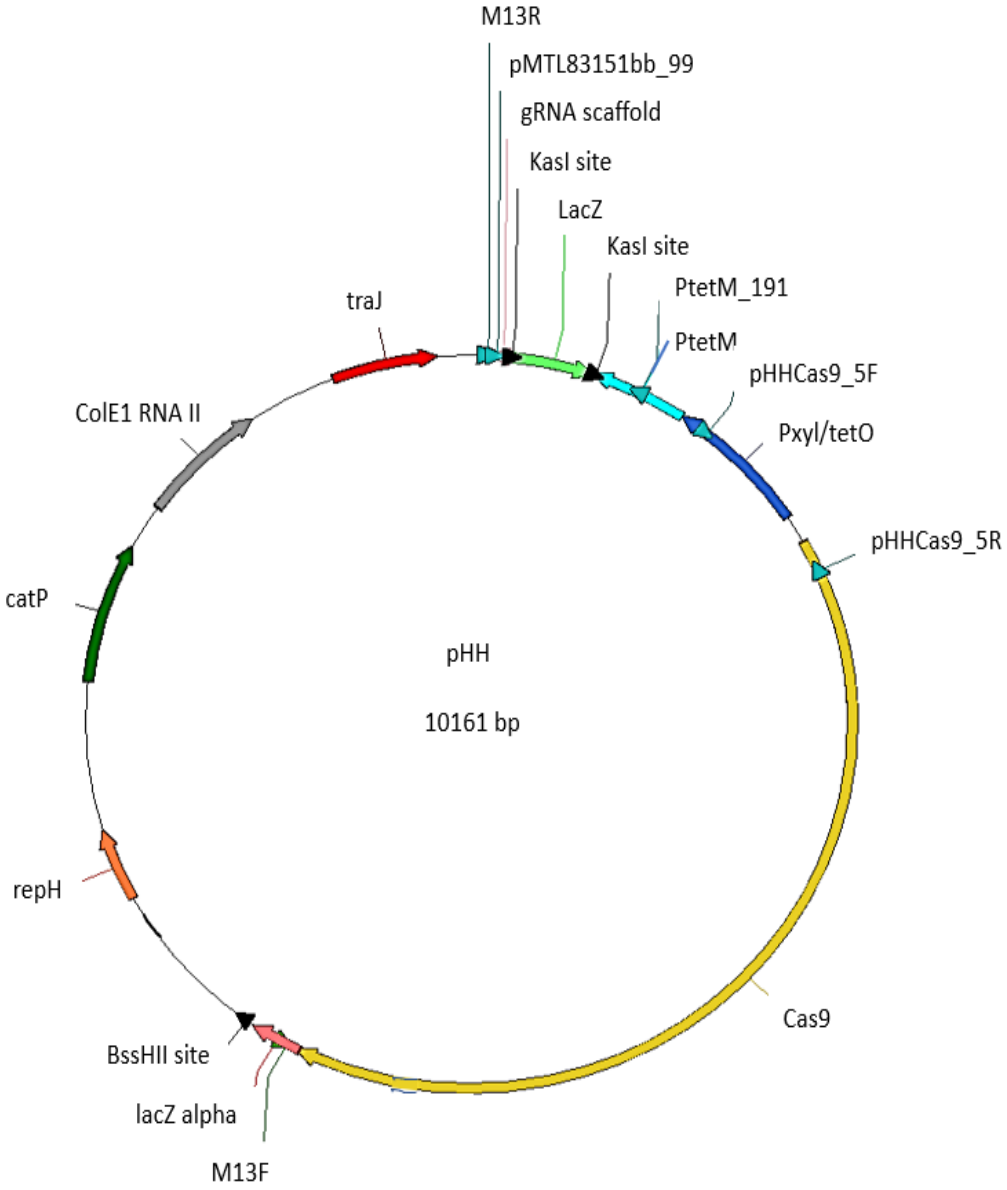


Figure 8. Schematic of pHH. Highlighted are the gene locations displayed as arrows and primer binding sites displayed as triangles.

2.4.2.4. Transformation of pHH_sgRNA ligation reaction to *E. coli* DH5α

Transformation was carried out as in 2.2.5. Positive transformed colonies were confirmed via colony PCR with primers, pMTL83151bb_99 and ptetM_191 (Table 2). The primers targeted regions flanking the inserted sgRNA sequence, with an expected amplicon of 351bp when sgRNA was present, or an expected amplicon of 297bp if the sgRNA was absent.

PCR was carried out using 1X OneTaq Buffer (New England Biolabs) , 4µM dNTP, 1µM primers pMTL83151bb_99 and ptetM_191, 10µL boiled colony, 0.75U OneTaq Polymerase (New England Biolabs) and nuclease-free water up to a final volume of 25µL

per rxn. The thermocycler conditions were 94°C for 30s followed by 30 cycles of 94°C for 15s, annealing at 55.4°C for 15s, and 68°C for 40s with a final extension at 68°C for 5 min. The PCR product was electrophoresed on a 2% agarose gel in 1 X TBE and 1X SYBR Safe DNA strain at 60 V for 2h.

PCR positive transformants were used to inoculate 10mL LB with and incubated 16-18h at 37°C, at 150-200rpm. Plasmid DNA was prepared from 5mL 16-18h culture using Monarch® Plasmid Miniprep Kit (New England Biolabs) and quantified as described in 2.2.2. Plasmid pHH_sgRNA sequence was confirmed by Sanger sequencing using primers pMTL83151bb_99 and ptetM_191, and sequencing analysis using CLC Workbench (Qiagen).

2.4.3 Cloning of homology arms into pHH_sgRNA

2.4.3.1 Homology Arm (HA) Amplification

The 2kb LHA and RHA KO cassette cloned into pMTL-SC7315 as described in 2.2.1 – 2.2.8 was used as a template to generate a similar KO cassette for cloning into pHH_sgRNA. Primers were designed using the DNA assembly tool (https://nebuilder.neb.com/#!/) to amplify the desired regions flanking the targeted prophage from pRK6302 (created in 2.2.1-2.2.8) with overlapping sequences to the pHH_sgRNA plasmid linearised at BssHII restriction site. Primer information is listed in Table 2.

PCR was carried out using 1x Phusion Mastermix (New England Biolabs), 0.5µM of the forward primer pHH_f630-2GA_F, 0.5µM of the reverse primer pHH_f630-2GA_R, 7.78ng of plasmid pRK6302 DNA, and nuclease free water up to 50µL. Cycling conditions were 98°C for 30s, followed by 30 cycles of 98°C for 5s, annealing at 69.4°c for 15s, 72°C for 1 min, and a final extension of 72°C for 5 min. The PCR product was then electrophoresed, gel extracted and quantified as described in 2.2.2. The expected amplicon was 2kb.

2.4.3.2 Digestion of pHH_sgRNA with BssHII

The vector, pHH_sgRNA to a final concentration 1.5ng of DNA was digested with 20U restriction enzyme BssHII (Thermo), 2µL 1x Thermo buffer R and nuclease-free water to a final volume of 20 µL, and incubated 16-18h at 37°C. The reaction was incubated at 65°C to heat inactivate the restriction enzyme, then electrophoresed, gel extracted, and quantified as described in 2.2.2.

2.4.3.3 Ligation of pHH_sgRNA and HA to create pHH_sgRNA_HA

The linearised pHH_sgRNA plasmid and the 2kb HA cassette were ligated at a 1:2 vector to insert ratio to 0.2 pmols, with 1x NEBuilder HiFi DNA Assembly Master Mix (NEB, UK) and nuclease-free water to a final volume of 20µL. The reaction mixture was incubated in thermocycler at 50°C for 1h, and 10µL was gel electrophoresed in a 1.5% agarose gel in 1x TBE and 1x SYBR Safe DNA stain (Invitrogen). The reaction mixture was then transformed into *E. coli*.

2.4.3.4 Transformation of pHH_sgRNA_HA ligation reaction to NEB10β

The transformation was carried out as described in 2.4.2.4 with a few alterations. NEB10β, a DH10β cell derivative was used, the *E. coli* cells were grown in SoM, and the volume of the transformation mixture was 100µL neat and 1/10 dilution in SoM.

Successful transformants were confirmed via colony PCR with primers, 630-2LHShomrec_F1 and 630-2RHShomrec_R1, which generated a 1kb amplicon that represented the junction between the LHS and RHS, which was be visualised via gel electrophoresis. PCR was carried out in 25µL consisting of 1x OneTaq Buffer, 4µM dNTP, 1µM of primers 630-2LHShomrec_F1 and 630-2RHShomrec_R1, 10µL boiled colony, 0.75U OneTaq Polymerase and nuclease-free water. The thermocycling conditions were 94°C for 30s, followed by 34 cycles of 94°C for 15s, 50°c for 15s, 68°C for 1 min, and a final extension of 68°C for 5 min. The resulting amplicons were then electrophoresed as described in 2.2.2 Plasmid from PCR positive transformants were prepared as described in 2.4.2.3.

2.4.4 Conjugation of *Clostridium difficile*

2.4.4.1 Transformation of conjugative *E. coli*

The *E. coli* donor strain CA434 was transformed as described in 2.3.1 with the pHH_sgRNA_HA recombinant plasmid, the pHH_sgRNA and pHH control plasmids. The successful transformants were selected for on LB agar with Cm (25µg/ml). The resultant transformants were then grown 16-18h in 5 mL of LB broth supplemented with Cm (12.5µg/mL) to retain the plasmid. In parallel, the CD630 recipient strain were grown 16-18h at 37°C in pre-reduced 10 mL of BHIB under anaerobic conditions.

2.4.4.2 *Conjugation with CD630*

A 1 mL volume of CA434_pHHsgRNA_HA clone and control clones were pelleted at 2841 x g for 3 min to remove all culture media and washed in 0.5mL sterile LB medium then re-pelleted. An 16-18h CD630 culture (250µL) was heat shocked at 50°C for 15 min (Kirk & Fagan, 2016), then mixed with the *E. coli* pellet in the anaerobic workstation. The mating mixture was spotted onto a dry, nonselective BHIA plate with eight individual drops, 20-25 µL per spot. The plates incubated right side up at 37 °C for between 1-3 days. OD_{600} of *E. coli* donor and CD630 recipient cultures were measured, and 10-fold serial dilutions were carried out then spread-plated onto LB agar with Cm (25µg/ml) and BHIA respectively. Colonies were counted after 24h aerobic incubation for *E. coli* donor and 48h anaerobic incubation for CD630 recipient, and the CFU/mL calculated. These mated spots were then collected and processed as described In 2.3.1.

Randomly selected Tm^R colonies were analysed via PCR with primers pHHCas9_5F and pHHCas9_5R that targeted *cas9*. The expected amplicon size was 818bp for both the recombinant and control plasmids. The PCR was carried out in 25µL comprising 1x OneTaq Buffer, 4µM dNTP, 1µM of each primer, 10µL boiled colony, 0.75U OneTaq Polymerase and nuclease-free water. Cycling conditions were 94°C for 30s, followed by 30 cycles of 94°C for 15s, annealed at 53°c for 15s, 68°C for 1 min, and a final extension of 68°C for 5 min. The resulting amplicons were then electrophoresed on a 1% agarose gel to confirm their size. Plates containing transconjugants were parafilmed and stored inside an anaerobic workstation.

2.4.5 Induction of *cas9* gene and selection of mutants

A single CD630 colony containing either pHH_sgRNA_HA, pHH_sgRNA or pHH was inoculated in 5mL BHIB supplemented with 15µg/mL Tm, incubated anaerobically 16-18h at 37°C. The culture was serially diluted 10-fold six times in pre-reduced BHIB and 100µL of the neat and 1/10 dilutions were then spread-plated onto reduced dry BHIA supplemented with 15-100ng/mL aTC, which induced the *cas9* gene. After 72h viable counts were recorded and CFU/mL was calculated.

A single colony for each PCR positive transconjugants was inoculated in 5mL BHIB supplemented with 15µg/mL Tm, incubated anaerobically 16-18h at 37°C. The culture

was serially diluted 10-fold six times in pre-reduced BHIB, 100µL of the neat and 1/10 dilutions were then spread-plated onto reduced dry BHIA supplemented with 30ng/mL aTC, which induced the *cas9* gene. Colonies (15-20) were randomly picked for colony PCR to detect the presence of ΦCD630-2 prophage by PCR using primers, phiCD630-2int_F2 and phiCD630-2int_R2 (Table 2), which targeted the prophage integrase gene producing a 221bp amplicon in WT CD630. Hence PCR negative colonies were putative prophage mutants.

PCR was carried out in 25µL runs assembled with 1X OneTaq Buffer, 4µM dNTP, 1µM of primers phiCD630-2int_F2 and phiCD630-2int_R2, 10µL boiled colony, 0.75U OneTaq Polymerase and nuclease-free water. Cycling conditions were 94°C for 30s, followed by 34 cycles of 94°C for 15s, annealed at 51°c for 15s, 68°C for 15s, and a final extension of 68°C for 5 min. Gel electrophoresis was carried out as described in 2.2.2.

PCR negative colonies were subcultured to BHIA and BHIB and incubated 16-18h at 37°C in anaerobic conditions. Transconjugant cultures of 16-18h were then passaged up to 18 times by serial subculture of 0.5mL in 4.5mL BHIB twice a day at 4 hour intervals. After the final passage, cultures were serially diluted 10-fold six times in BHIB, 100µL of each dilution was spread-plated onto BHIA and left to incubate for 2-3 days at 37°C in anaerobic conditions. These plates are used as master plates for replica plating on to BHIA, and BHIA supplemented with 15µg/mL Tm to identify colonies cured of plasmid. Any colonies that had failed to grow on the antibiotic supplemented plate was then screened via PCR using primers pHHCas9_5F and pHHCas9_5R (Table 2), to detect for the presence of the plasmid as described in 2.4.5.2.

3. Results

3.1 Allelic Coupled Exchange

The left and right homology arms (LHA/RHA) were amplified using a temperature gradient PCR ranging from 56-64°C to optimise the reaction at a 20µL final volume (Figure 9A). Bands for the homology arm amplicons were of the expected size of 1kb (Figure 9B).

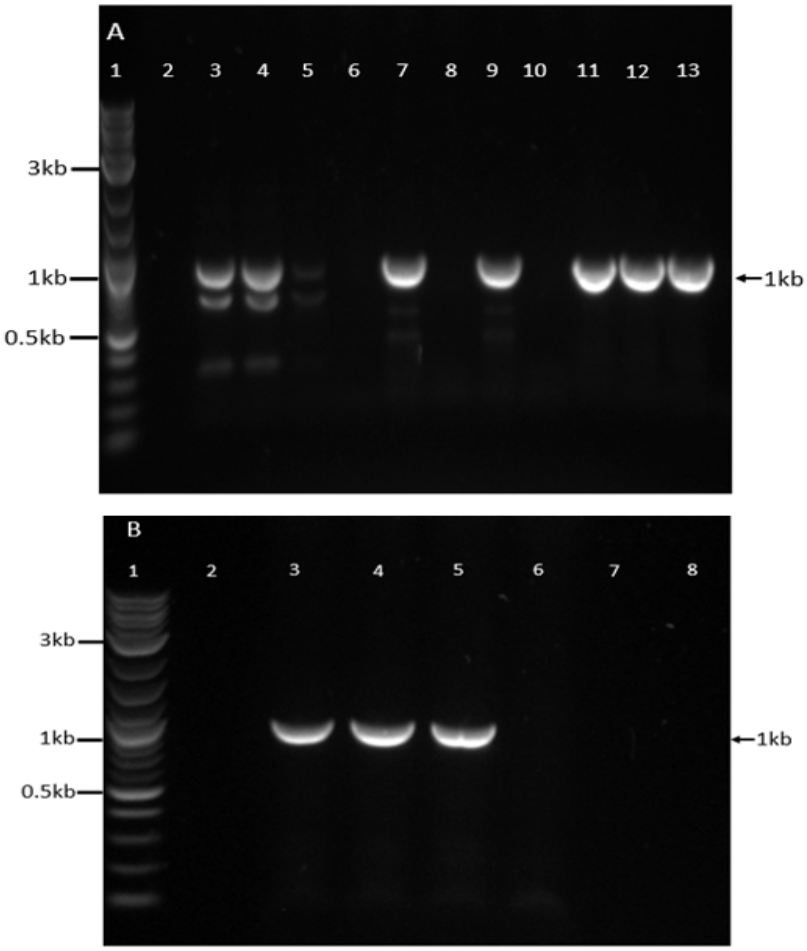


Figure 9. Optimising PCR of homology arms for ΦCD630-1 and ΦCD630-2. An expected amplicon size of 1kb was observed in a 1% agarose gel, electrophoresed for 40 minutes at 100 volts. In well 1 of A and B is the 1kb plus DNA ladder (New England Biolabs). (A) In well 2 is the no template control, in wells 3-5 are the ΦCD630-1 LHA amplicons at annealing temperature 56°C, 59.1°C and 62.4°C respectively, well 6 was a spacer and in wells 7-9 are the ΦCD630-1 RHA amplicons at annealing temperatures 56°C, 59.1°C and 62.4°C respectively. Well 10 was a spacer, in wells 11-13 are the ΦCD630-2 LHA amplicons at annealing temperatures 57.6°C, 60.9°C and 64°C respectively. (B) In well 2 is the no template control, in wells 3-5 are the ΦCD630-2 RHA amplicons at annealing temperatures 56.6°C, 59.1°C and 63.4°C respectively.

Once annealing temperatures were optimised as 59.1°C for ΦCD630-1LHA, 62.4C° for ΦCD630-1RHA, 64°C for ΦCD630-2LHA and 63.4°C for ΦCD630-2RHA, a second round of PCR was carried out at a higher volume of 50µl, electrophoresed, gel extracted, and quantified (Figure 10), yielding 106-219 ng/µL, with satisfactory 260/280 ratios of 1.8-1.9.

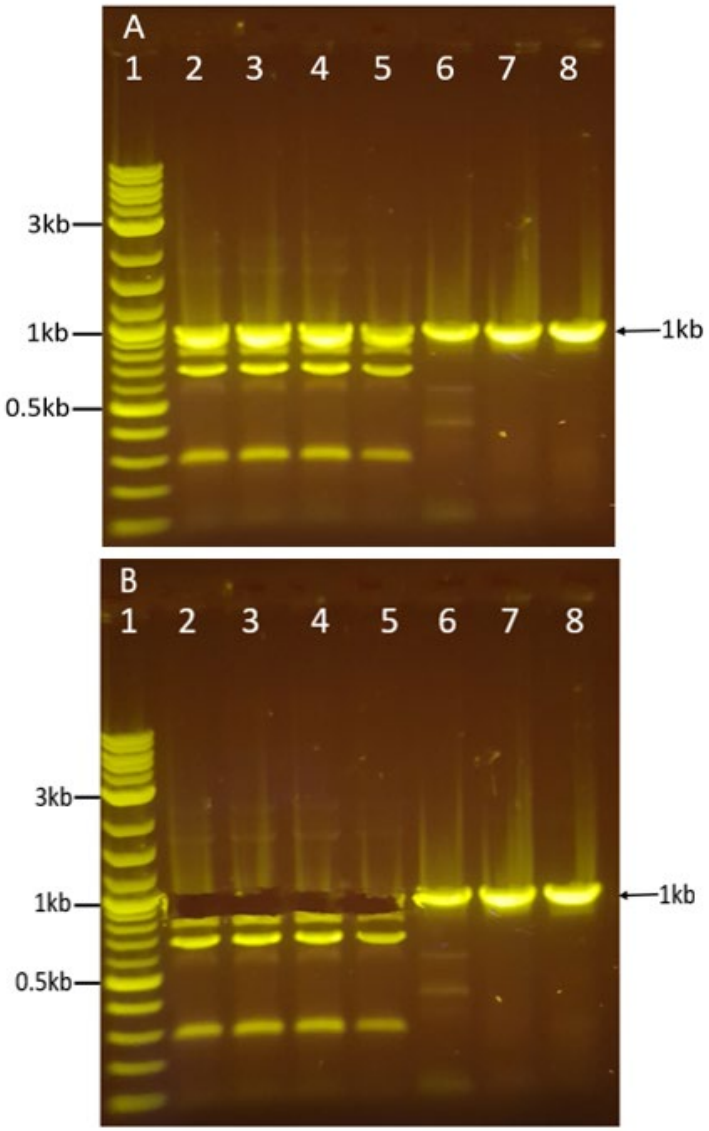


Figure 10. Gel-extraction of LHA and RHA amplicons for both prophages. (A) before gel excision (B) after gel excision. In wells 2-3 were the ΦCD630-1 LHA amplicons which had significant laddering and so were cut from the gel and extracted for purity. In well 6 was the ΦCD630-1 RHA amplicon, in well 7 was the ΦCD630-2 LHA amplicon and in well 8 was the ΦCD630-2 RHA amplicon.

The purified homology arms were ligated into linearised pMTL-SC7315 (55.3ng/µL) and the resulting plasmids are displayed in Figure 11, with Figure 11A showing plasmid

pRK6301, which selects for ΦCD630-1, and Figure 11B showing the plasmid pRK6302, which selects for ΦCD630-2.

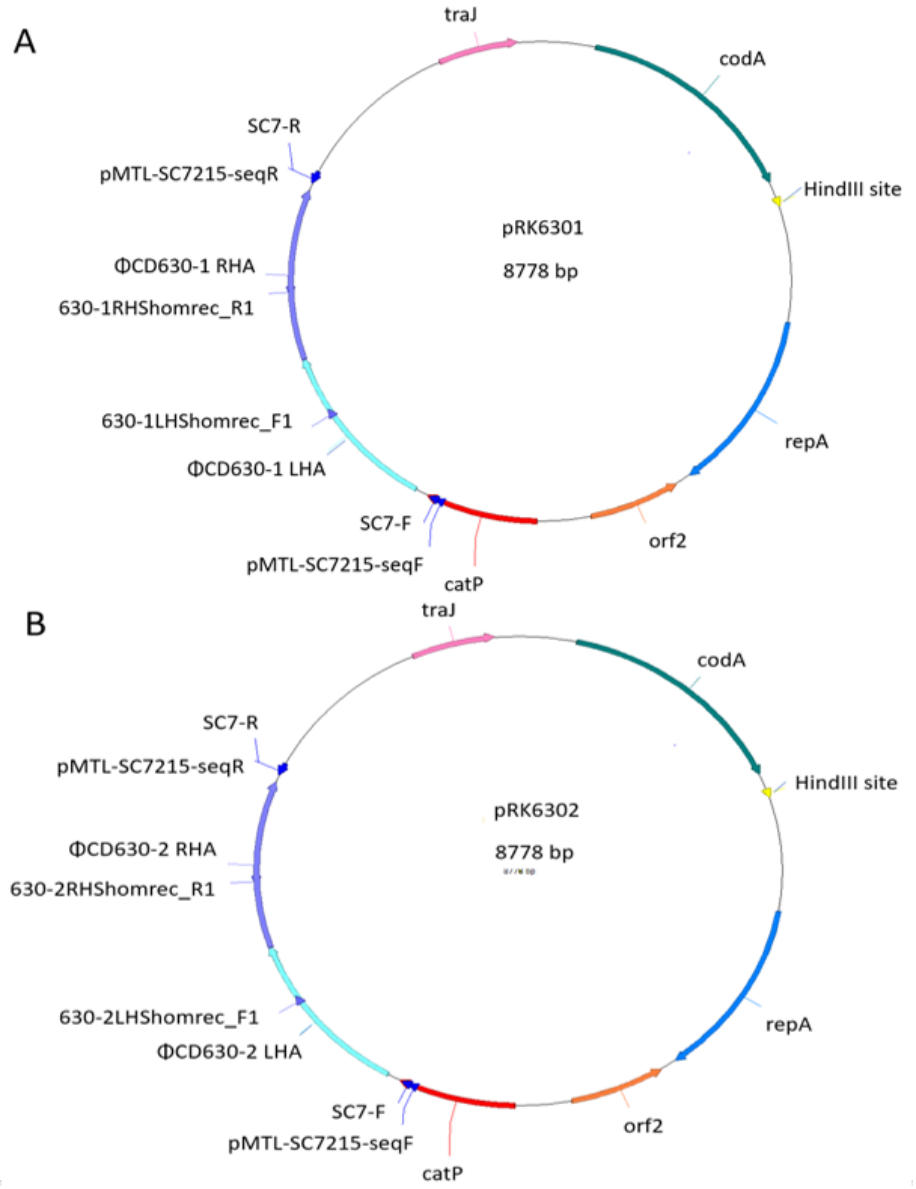


Figure 11. Schematic of ligated pRK6301/2 plasmids. Relevant features shown. (A) This is the circularised plasmid containing the homology arms known as pRK6301 and (B) is the circularised plasmid containing the homology arms known as pRK6302.

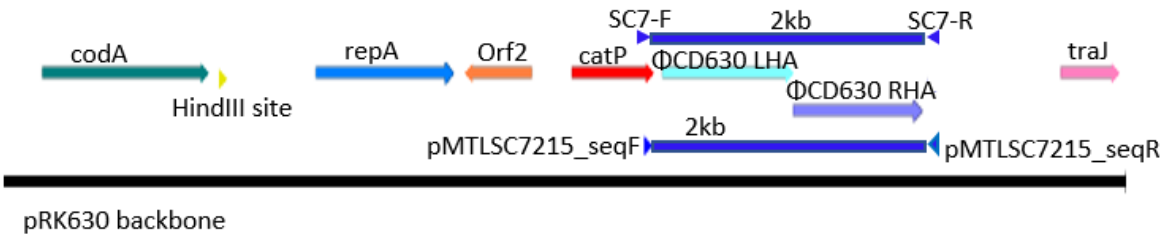


Figure 12. Linear schematic for pMTLSC7215_seqF/R and SC7F/R pairs. Relevant features shown. A representation of expected fragment sizes (solid blue line) and primer annealing sites (blue triangles) for pMTLSC7215_seqF/R and SC7F/R pairs on pRK630 plasmid backbones.

Sixteen *E. coli* transformant colonies screened via colony PCR. Figure 13A, B and C are the result of the pMTLSC7215_seqF/R primer pair and Figure 13D shows the result of the SC7F/R primer pair. Two different primer pairs were used as non-specific amplicons were observed for the pMTLSC7215_seqF/R primer pair. Three of 16 *E. coli* transformants were positive for the expected 2kb amplicon, hence positive for pRK6301. Similarly, four of 16 *E. coli* transformants were positive for the expected 2kb amplicon, hence positive for pRK6302 (Figure 13).

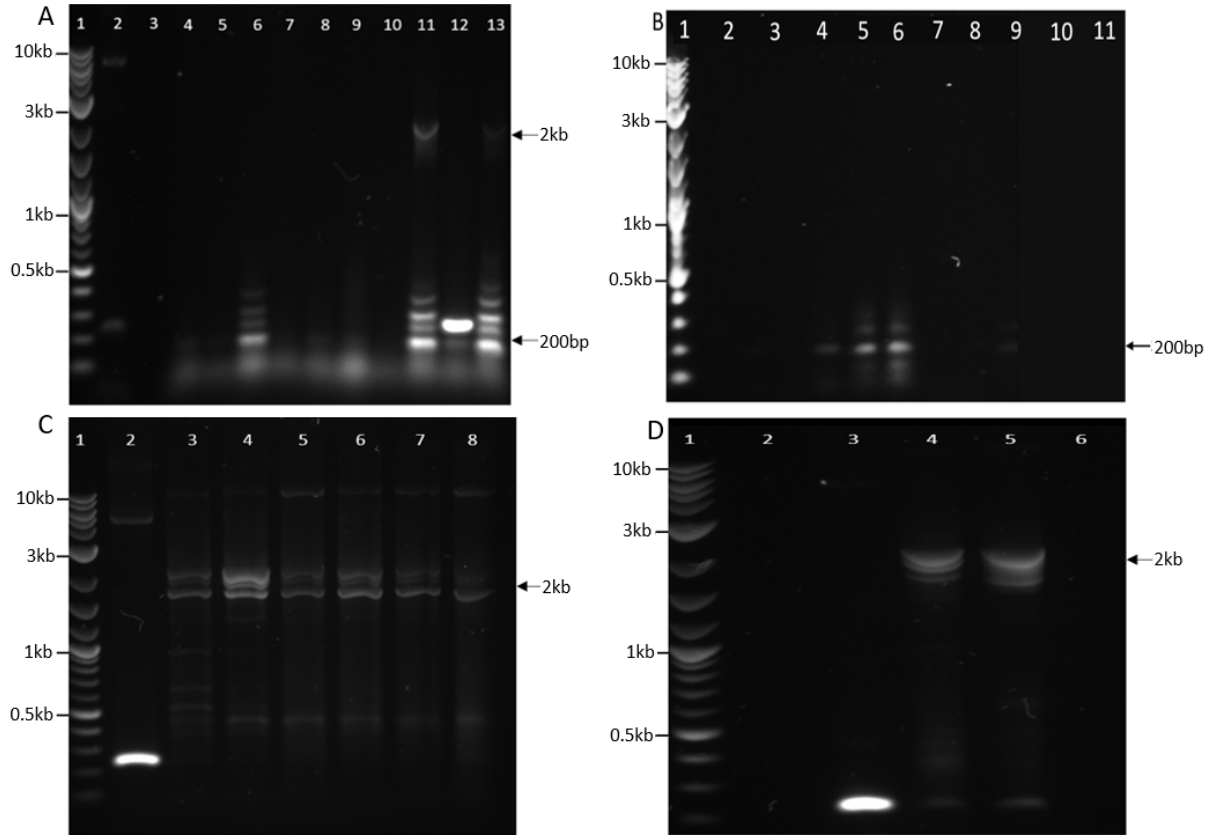


Figure 13. Detection of recombinant pRK6301/2 in *E. coli* transformants via colony PCR. Well 1 in each gel contains the 1kb plus DNA ladder and well 2 contains the positive control (empty pMTL-SC7315 vector). PCR with pMTLSC7215_seqF/R primers are shown, unless stated otherwise. (A) wells 4-13 *E. coli* transformants containing pRK6301 colonies numbered 1-10, (B) wells 4,5 contain *E. coli* transformants containing pRK6302 colonies 1 and 2, wells 6-11 contain *E. coli* transformants containing pRK6302 colonies numbered 11-16. (C) wells 3-4 had *E. coli* transformants containing pRK6301-8 annealing at 60°C and 63°C, wells 5-6 had *E. coli* pRK6301 transformant 11 annealing at 60°C and 63°C, wells 7-8 had *E. coli* pRK6301 transformant 12 annealing at 60°C and 63°C. (D) PCR with primers SC7-F/SC7-R. Well 4 contains *E. coli* pRK6301 transformant 8 with a positive band and *E.*

coli pRK6302 transformant 14 with a positive band, well 6 contained the no template control.

E. coli transformant T8 containing pRK6301 in well 11 of Figure 13A had a weak positive band and so was carried forward, although laddering in lower positions of the same well was noted. *E. coli* transformants T13 and T14 containing pRK6302 were missing a negative amplicon band (Figure 13B) and were carried forward to alternative confirmation with the SC7F/R primer pair which produced a strong positive amplicon (Figure 13D). To optimise the PCR reaction with primers pMTLSC7215_seqF/R, 3% v/v DMSO was added, and this led to the banding pattern seen in Figure 13C.

Putative positive *E. coli* transformants were used for plasmid extraction, yields were 136-208 ng/µL, with satisfactory 260/280 ratios of 1.8-1.9.

To confirm the identity of recombinant plasmids, which have a unique *HindIII* site (Figure 11), pRK6301 from colonies T8, T11 and T12 and pRK6302 from colonies T13 and T14, were digested with *HindIII*. An expected digested pattern was observed (Figure 14). The digestion for the empty control plasmid resulted in a band present at the 6kb position of the ladder (Figure 14B). Plasmids pRK6301 and pRK6302 showed a band at the 8kb position (Figure 14A).

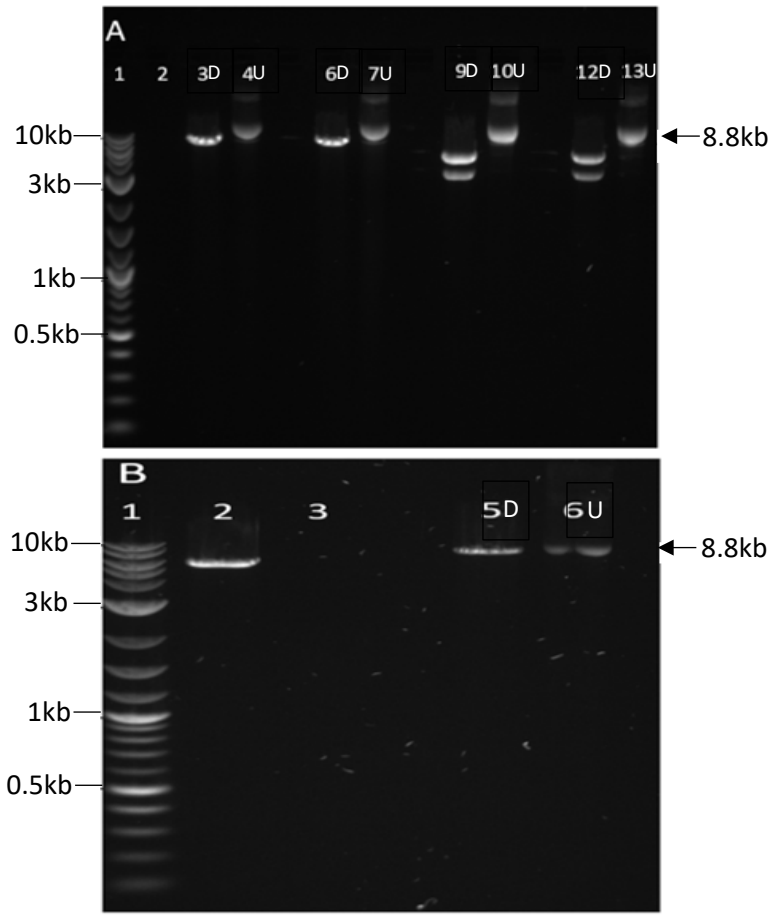


Figure 14. *HindIII* digestion of recombinant plasmids extracted from *E. coli* DH5α transformants T8, T11, T12, T13, T14. (A) Lane 1: 1kb plus DNA ladder (NEB), 2: No Template Control, 3: T11 digested (D), 4: T11 undigested (U), 6: T12D, 7: T12U, 9: T13D, 10: T13U, 12: T14D and 13: T14U.(B) Lane 1: 1kb plus DNA ladder, 2: is the pMTL-SC7315D, 3: No Template Control, 5: T8D and 6: T8U.

Following this, pRK6301 from *E. coli* transformant T8 and pRK6302 from *E. coli* transformant T14 were Sanger sequenced. A contig of 2083 bp from pRK6301-8 aligned with 100% identity to nucleotide 5072-7155 of the plasmid map of pRK6301 (Appendix 1). A contig of 2,134 bp from pRK6302 T14 aligned with 100% identity to nucleotide 5020-7153 of the plasmid map of pRK6302 (Appendix 2).

Two hundred colonies each for *C. difficile* transconjugants containing pRK6301 and pRK6302 and 34 *C. difficile* control transconjugants containing pMTL-SC7315 (i.e., empty plasmid, Figure 7) were replica plated onto BHIS non-selective agar and BHIS,Tm as seen in Figure 15. The transconjugants which grew on BHIS non-selective but not on BHIS,Tm were considered presumptive double crossovers. Three of 200 (1.5%) possible deletants

were found for ϕCD630-1: CD630ΔΦ1A1, CD630ΔΦ1E4 and CD630ΔΦ1E19, while two of 200 (1%) possible deletants were found for ϕCD630-2: CD630ΔΦ2D2 and CD630ΔΦ2C13.

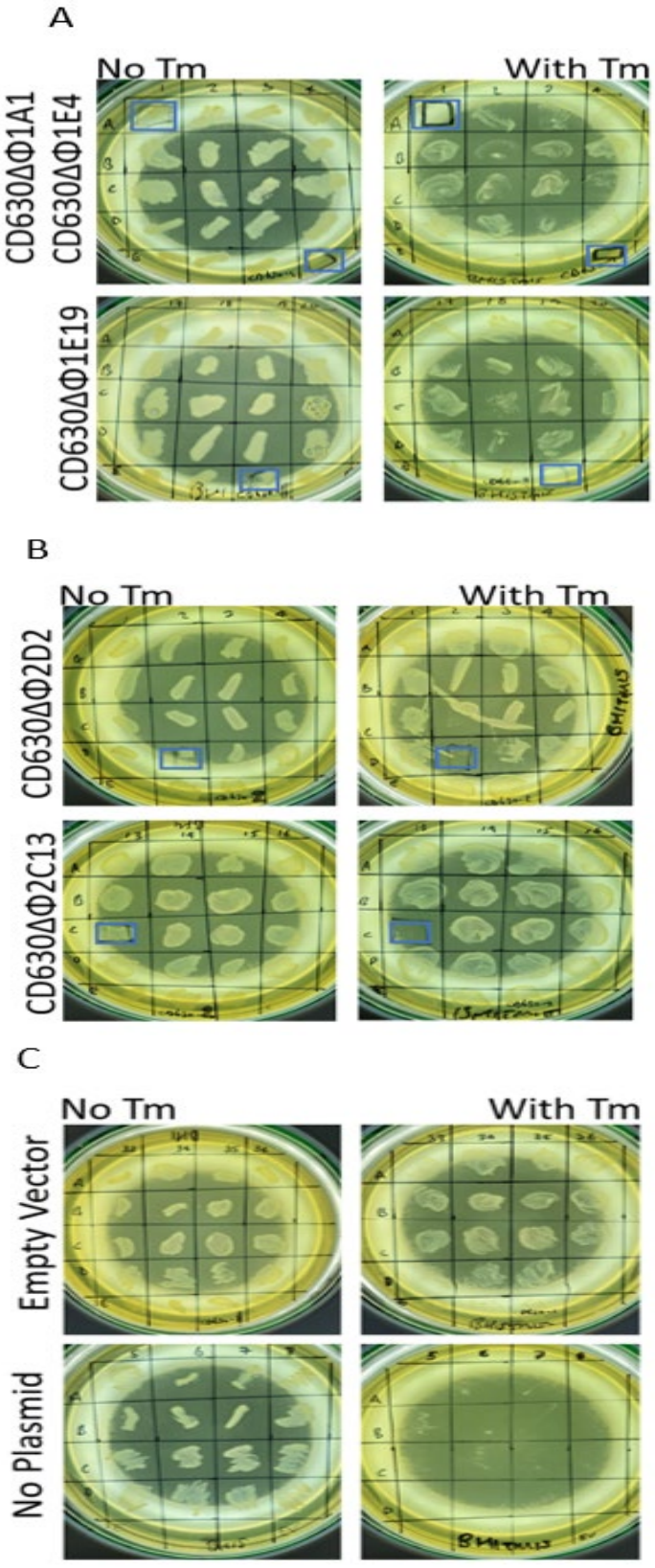


Figure 15. Replica plates for double crossover selection of CD630 transconjugants. Colonies of interest are highlighted in blue boxes. (A) These were the replica plates for the double crossovers CD630ΔΦ1A1, CD630ΔΦ1E4 and CD630ΔΦ1E19. (B) These were

the replica plates for the double crossover CD630ΔΦ2D2 and CD630ΔΦ2C13. (C) These were the replica plates for the empty vector control and CD630 no plasmid control.

The extracted genomic DNA for potential deletants yielded concentrations ranging from 63.1-81.7 ng/µL, with satisfactory 260/280 ratios of 1.76-1.86. CD630ΔΦ1A1, CD630ΔΦ1E4, CD630ΔΦ1E19, CD630ΔΦ2D2 and CD630ΔΦ2C13, were then subjected to PCR with primers 630-1LHShomrec_F1 and phi630-1LHS_R, as described in 2.3.2. A 897bp product was expected if ΦCD630-1 was present, while a 598bp product was expected if ΦCD630-2 was present. No amplification would occur if prophages were deleted. Figure 16 showed that both prophages were not deleted. This could be because the plasmid was lost with no recombination, since we did not confirm single crossover in our transconjugants, or it could be because double crossover occurred without prophage deletion. This will be discussed later in Chapter 4.

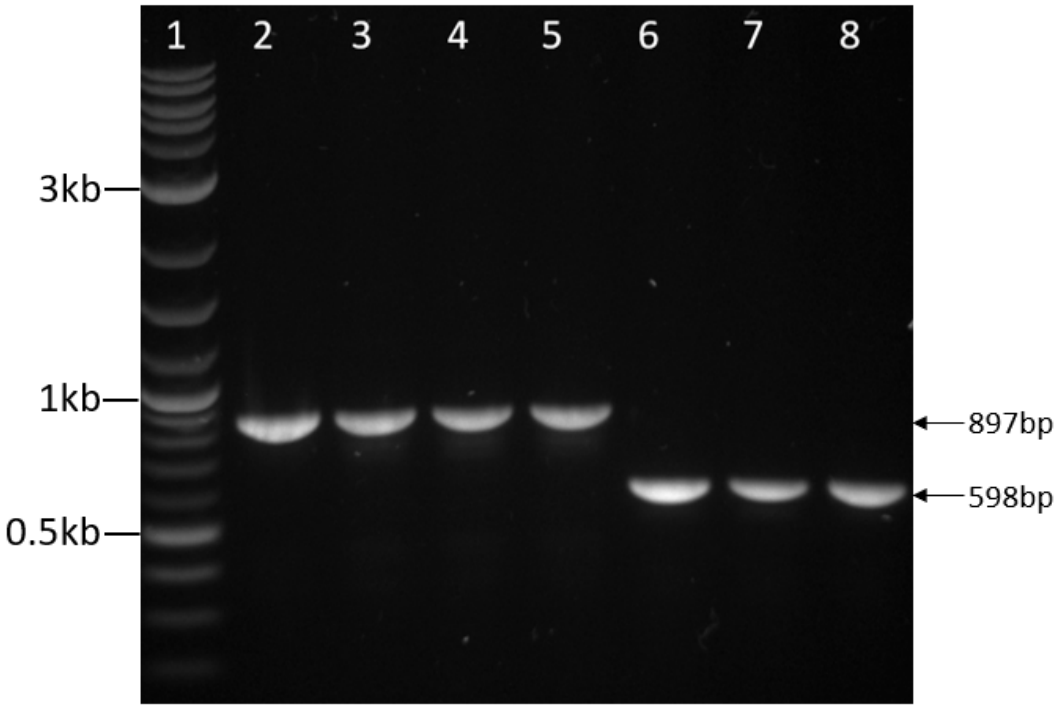


Figure 16. PCR detection of putative prophage deletion in CD630 putative mutants. Gel electrophoresis was done in 0.8% agarose gel at 60 volts for 2 h. Lane 1: 1kb plus DNA ladder, lanes 2-5: 630-1LHShomrec_F1/phi630-1LHS_R PCR of CD630ΔΦ1A1, CD630ΔΦ1E4, CD630ΔΦ1E19. Lanes 6-8: phi630-2RHS_F/630-2RHShomrec_R1 PCR of CD630ΔΦ2D2 and CD630ΔΦ2C13.

3.2 Crispr-Cas9 System

Another strategy was employed to delete ΦCD630-2 in CD630. This was based on successful gene deletions using CRISRP-Cas9 (McAllister *et al*., 2019) and the availability of a suitable vector constructed in Prof Peter Mullany's lab. The vector pHH consisted of *E. coli-C. difficile* shuttle vector (pCB102 *catP* ColE1 *traJ*) with *codA lacZ ptetM* Pxyl*/tetO cas9* (Figure 8). The *catP* is a resistance marker coding for chloramphenicol acetyltransferase, which provides resistance to chloramphenicol and thiamphenicol, and the Pxyl/TetO promoter is an inducible promoter by aTC which is in control of the expression of the *cas9* which codes for the Cas9 protein (Figure 8).

Candidate guide RNA sequences found by the CRISPR guide tool (https://benchling.com/) are shown in Appendix 3. The sgRNA sequence chosen was 5' AAAATTGATAGACTAAGTCG 3', which targeted nt. 3427732- 3427751 of the CD630 genome, and PAM at nt 3427749 which lies within the ϕCD630-2 integrase gene. The sgRNA sequence had efficiency and specificity scores of 67.2 and 99.9, respectively. The integrase gene was chosen for cleavage because its deletion was hypothesized to lead to prophage excision and inability to re-integrate into the bacterial genome, shown in other studies (Chen *et al*., 2019; Hong, Zhang, Cui, Wang, & Wang, 2018; Selle *et al*., 2020; Wang *et al*., 2010).

The sgRNA oligos were successfully annealed (Figure 17), pHH was successfully linearised with KasI and gel extracted (Figure 18) and it was either self-ligated to generate pHH (no *lacZ*) as shown in Figure 19 or it was ligated with the annealed oligos (Figure 20) to generate pHH_sgRNA (Figure 21).

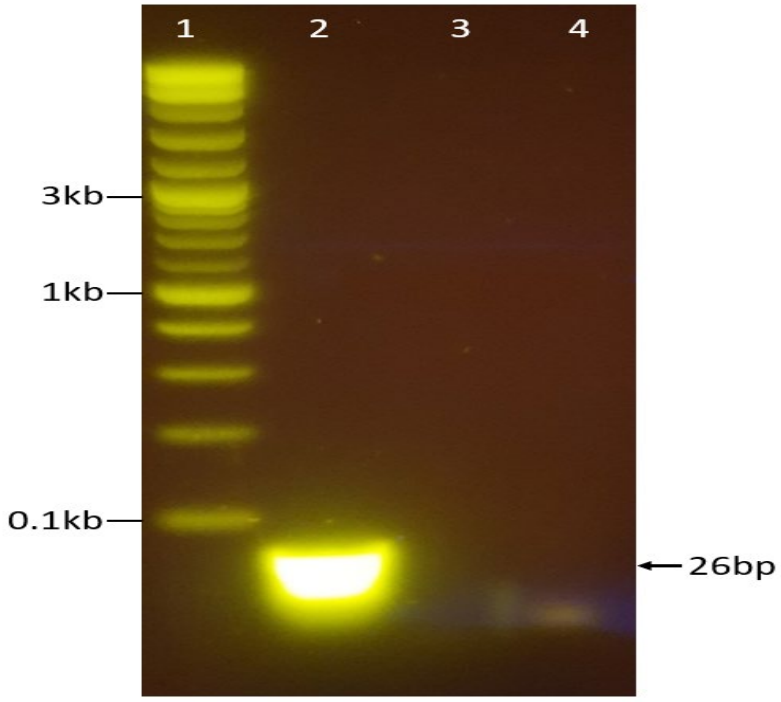


Figure 17. Annealing of sgRNA oligos. gel electrophoresis was carried out in 2% agarose and 1X SYBR Safe, at 80 volts for 90 minutes. Lane 1: 1kb plus DNA ladder, 2: annealed oligos f630-2gRNA2_top and f630-2gRNA2_bot, 3: Oligo f630-2gRNA2_top, 4: oligo f630-2gRNA2_bot.

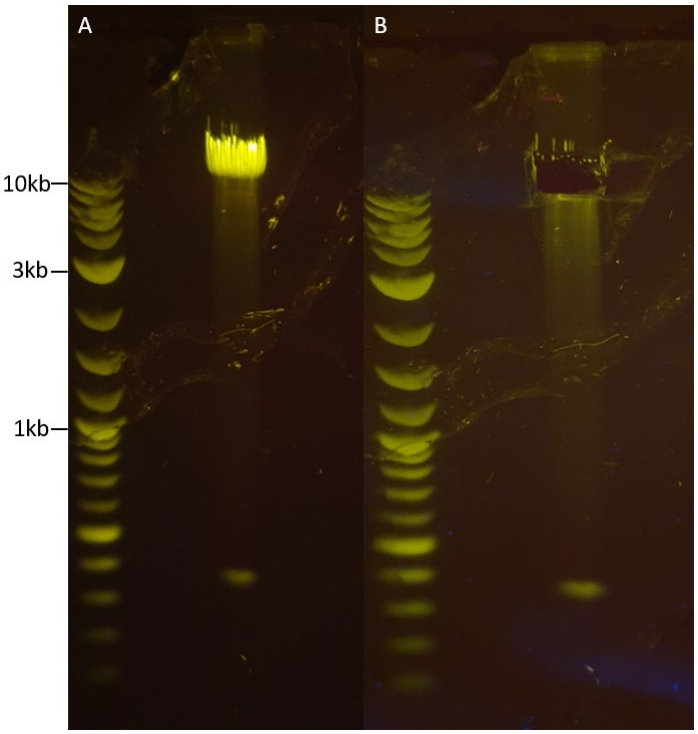


Figure 18. Gel extraction of linearised pHH backbone. DNA was separated in 0.8% agarose and 1 X SYBR Safe in 1X TBE at 60 volts for 2 hours. (A) Linearised plasmid at approximately 10 kb and excised lacZ at 350 bp before gel extraction. (B) Agarose gel after linearised plasmid was removed.

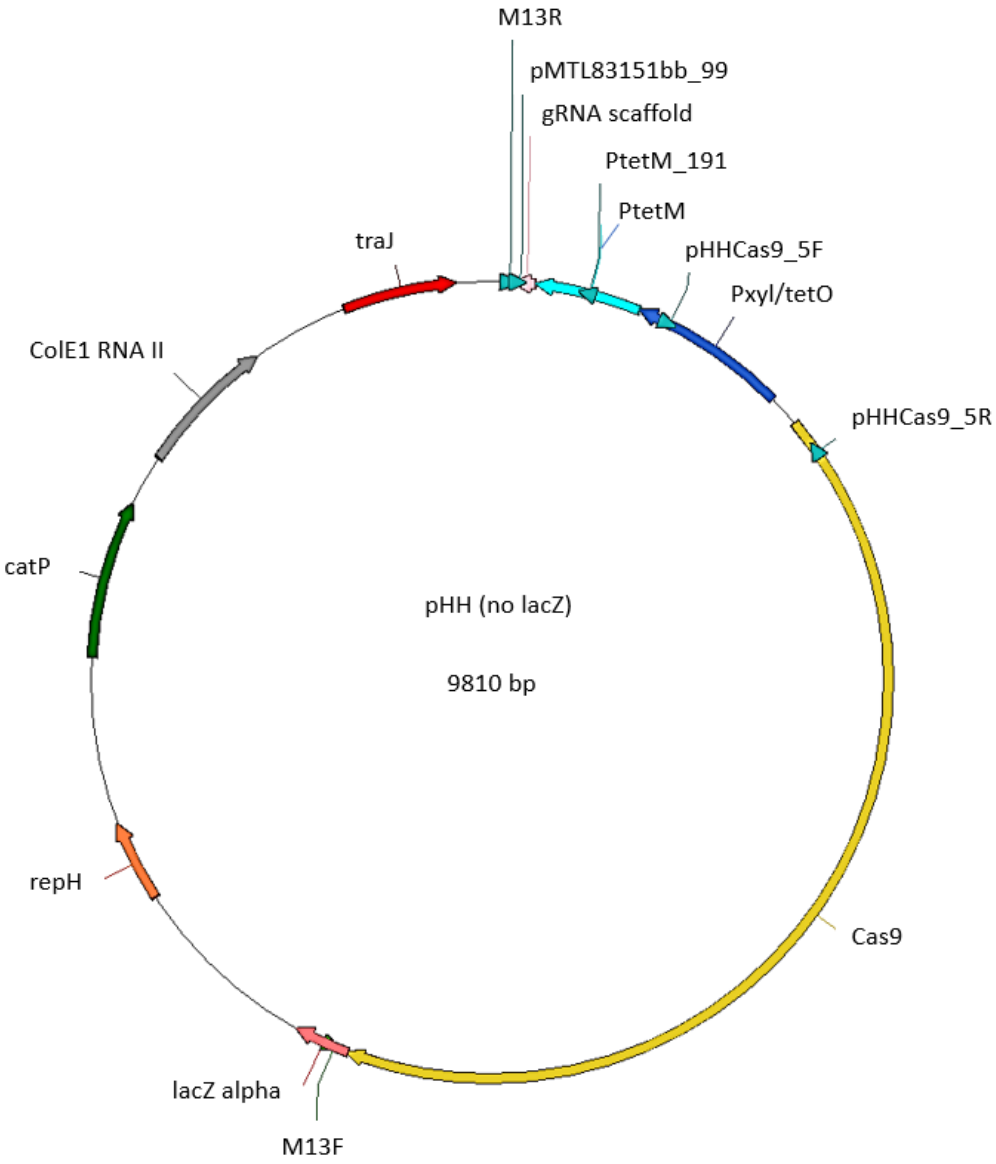


Figure 19. Schematic of pHH (no *lacZ*). Highlighted genes represented as arrows and primer locations where important represented as triangles.

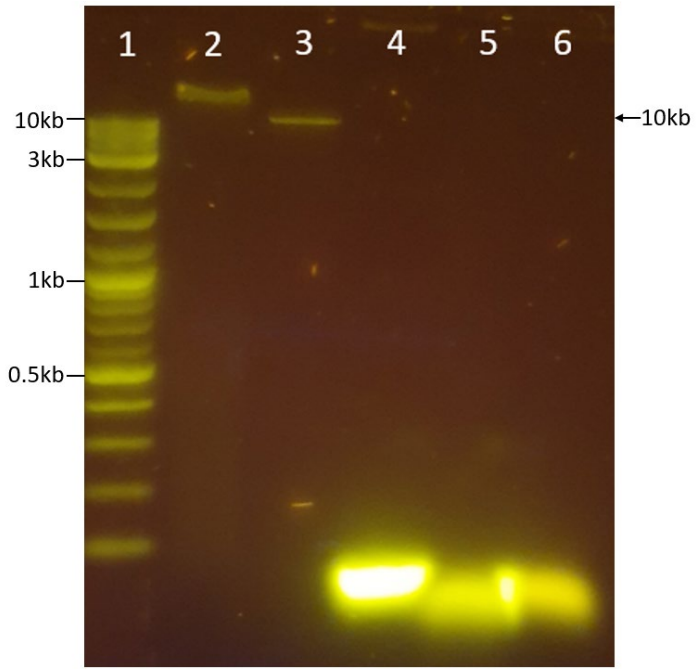


Figure 20. pHH backbone ligation with annealed sgRNA oligos. 1.5% agarose gel ran at 100 volts for 1 hour which were visualised using a dark viewer. Lane 1: 1kb plus DNA ladder, 2: ligated pHH_sgRNA recombinant, 3: linearised pHH backbone, 4: annealed sgRNA oligos, 5: f630-2gRNA2_top and 6: f630-2gRNA2_bot.

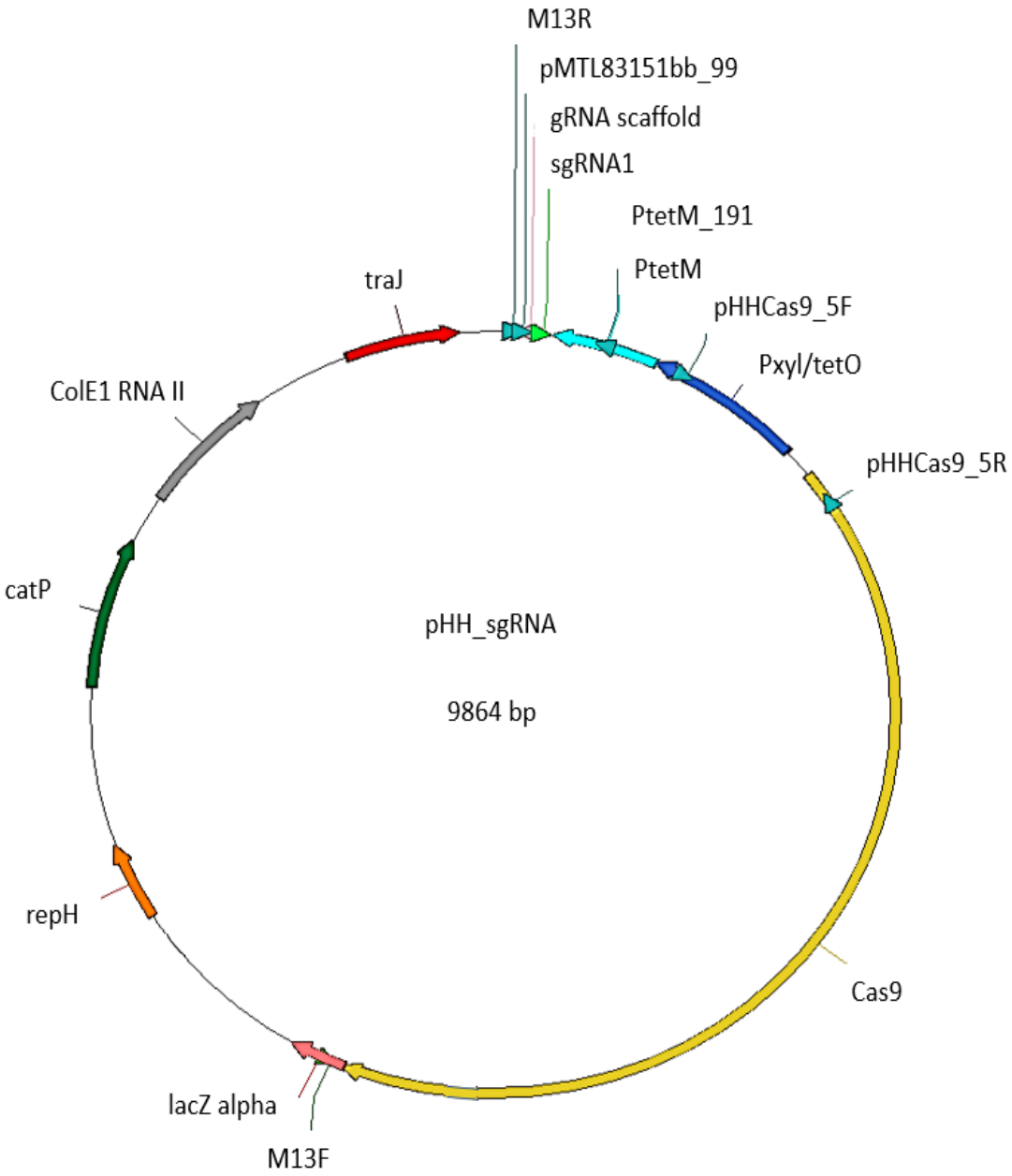


Figure 21. Schematic of pHH_sgRNA. Highlighted genes represented as arrows and primer locations where important represented as triangles.

A total of 36 *E. coli* transformants were screened and 7 were positive for pHH_sgRNA (Figure 22).

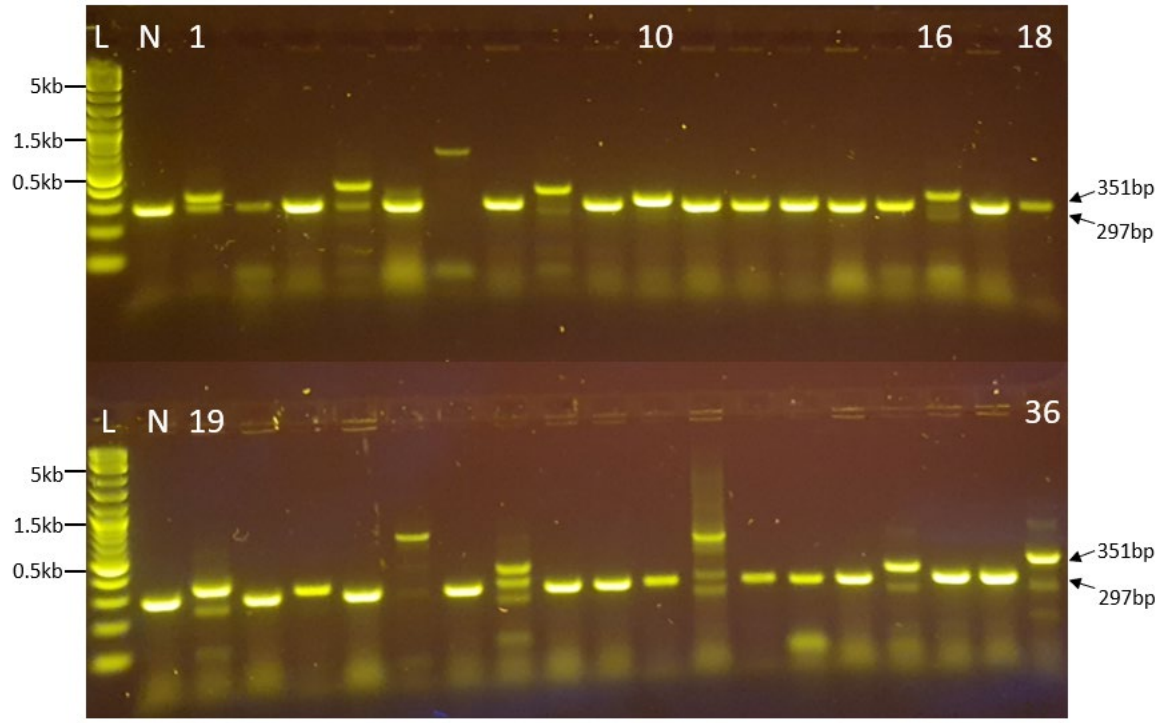


Figure 22. Colony PCR screening for the DH5α transformants containing the pHH_sgRNA recombinant plasmid. Amplicons were separated in a 2% gel for 2 hours at 60 volts. Wells L contained the DNA ladder marker, wells N was a colony transformed with linearised pHH plasmid acting a negative amplicon control, numbered wells provide a guide for the corresponding colony number, whilst colonies of interest were 10 and 16.

Two positive colonies, named D5a-pHH_sgRNA_C10 and D5a-pHH_sgRNA_C16, were subsequently used. The negative control colonies contained pHH (no lacZ) (Figure 19). Plasmids extracted from D5a-pHHsgRNA_C10 and D5a-pHHsgRNA_C16 were then sequenced to confirm the PCR results, as well as the orientation of the sgRNA. D5a-pHHsgRNA_C10 had the sgRNA cloned in the correct orientation (Appendix 4), while the sequence of the sgRNA in D5a-pHHsgRNA_C16 could not be confirmed because of premature termination of base calling signals (Appendix 5).

The 2kb homology arm cassette consisting of both the LHA and RHA were amplified from pRK6302 as a template using primers pHH-f630-2_GA_F and pHH-f630-2_GA_R, by temperature gradient PCR ranging from 58-66°C (Figure 23).

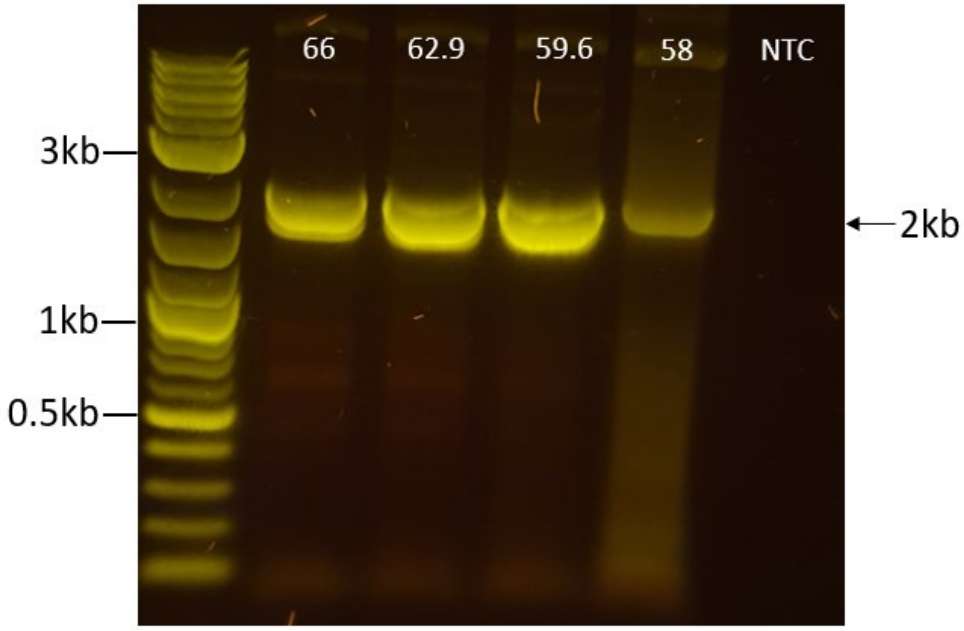


Figure 23. ΦCD630-2 KO cassette PCR amplification. pRK6302 was used as a template, the amplicons were generated at the expected position of 2kb using primers pHH-f630-2_GA_F and pHH-f630-2_GA_R. The labelling is the annealing temperatures of the reaction, the reactions at 66°C and 62.9°C had two bands at similar sizing at 2kb, the reaction at 59.6°C had 1 band at 2kb and the final reaction at 58°C had a single band at 2kb as well as smearing across the well.

Once the annealing temperature was optimised as 59.6°C due to lack of smearing and a single band, a second round of PCR was carried out in a 50µl reaction, and gel extracted (Figure 24) to yield 184.5ng/µL and a 260/280 ratio of 1.89.

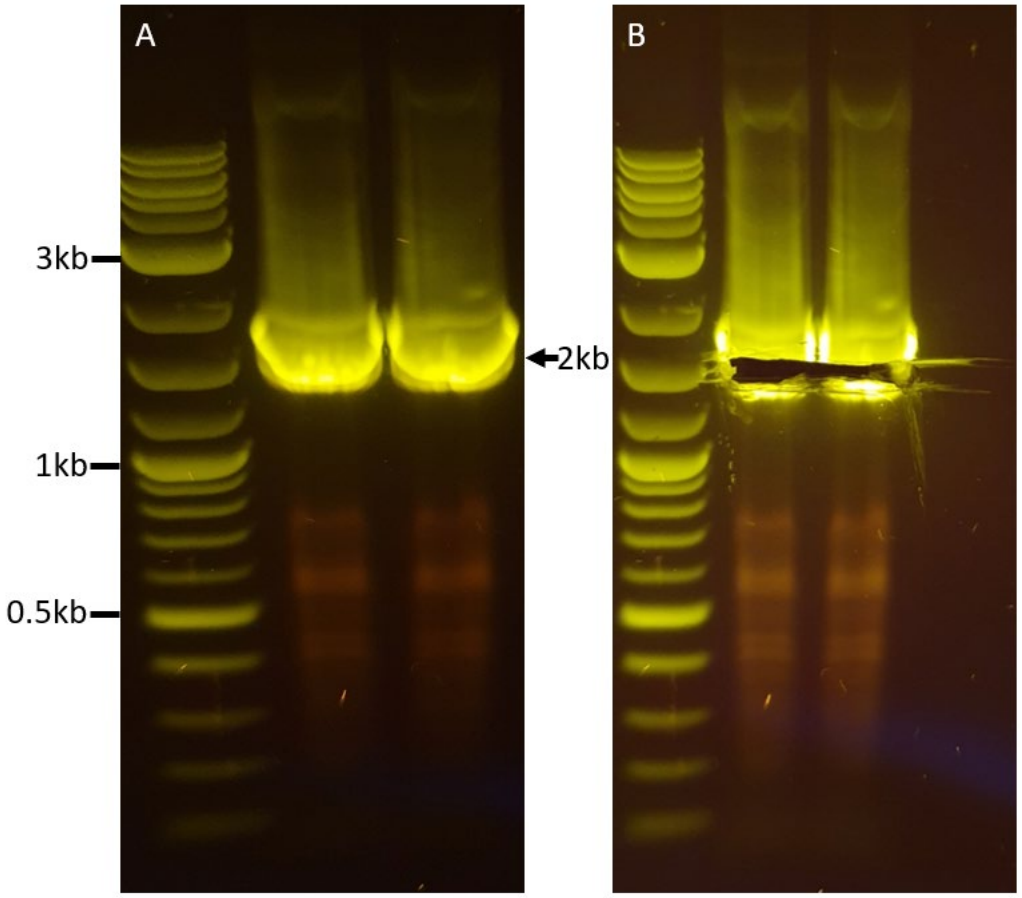


Figure 24. Gel extraction of the KO cassette homologous to ΦCD630-2 flanking regions. A) Two wells with 25µL amplicon before cutting (B) Two wells with the band at 2kb size cut from the agarose.

D5a-pHHsgRNA_C10 recombinant plasmid was successfully linearised with BssHII, and gel extracted (Figure 25). The KO cassette was successfully ligated to pHH_sgRNA (Figure 26).

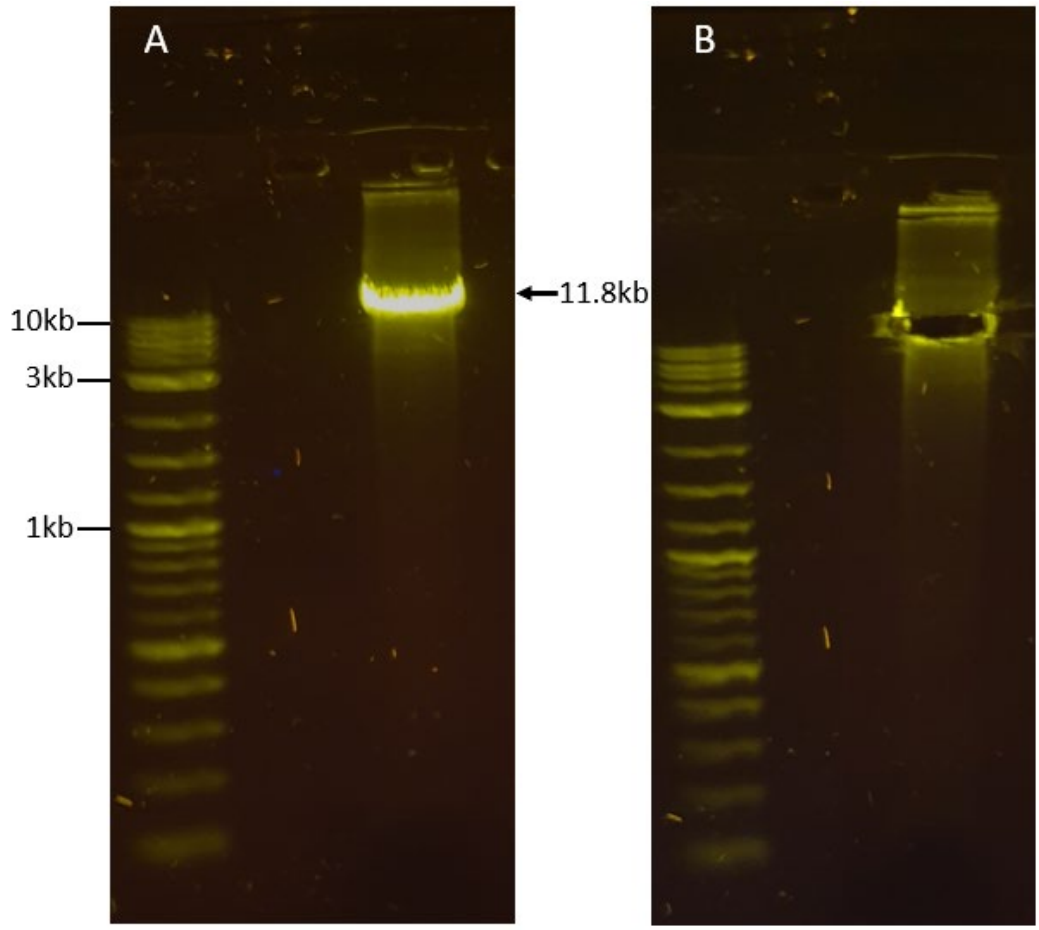


Figure 25. Gel extraction of BssHII digested pHH_sgRNA from D5a-pHHsgRNA_C10. Gel electrophoresis was carried out in 1% agarose gel for 1 hour at 80 volts. (A) A band of approximately 11kb, close to the expected size of the linearised plasmid, (B) The linearised plasmid band cut from the agarose gel.

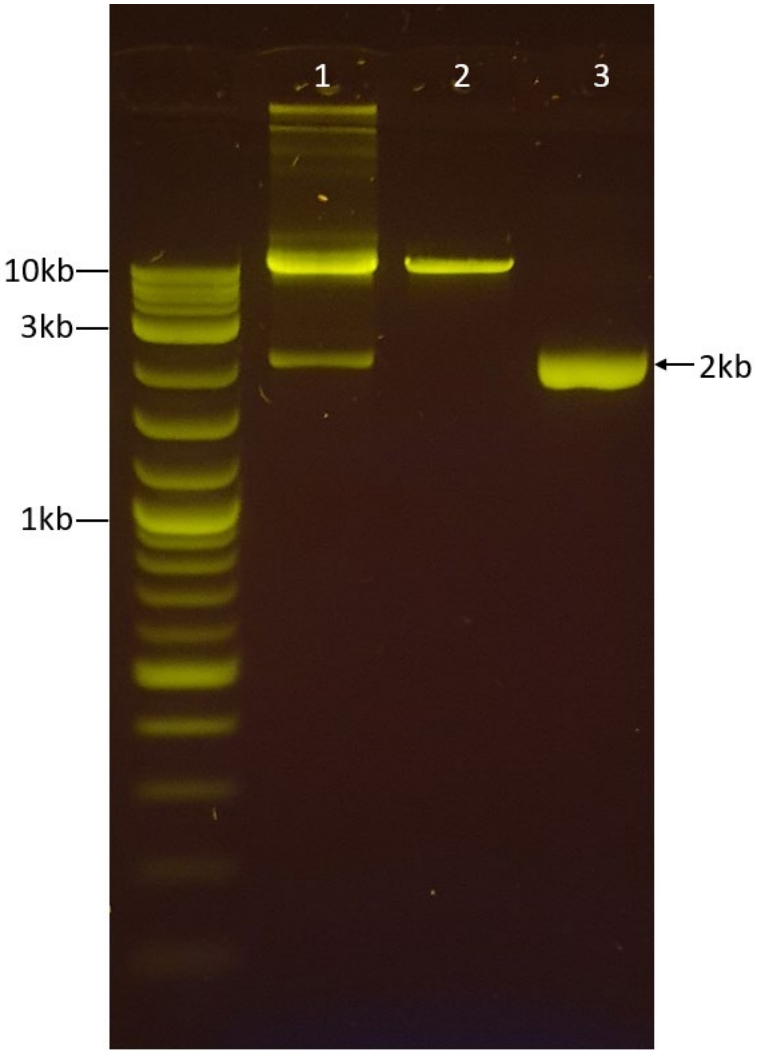


Figure 26. Gibson assembly of pHH_sgRNA and the KO cassette. Gel electrophoresis was carried out in 1.5% gel for 1 h at 100 volts . Lane 1: ligation mixture, lane 2: digested plasmid, lane 3: KO cassette amplicon.

The ligated plasmid, pHH_sgRNA_HA (Figure 27), was transformed into *E. coli* NEB10β. Of 14 colonies screened by colony PCR with primers 630-2LHShomrec_F1 and 630-2RHShomrec_R1, two colonies were positive for the expected 1kb amplicon (Figure 28).

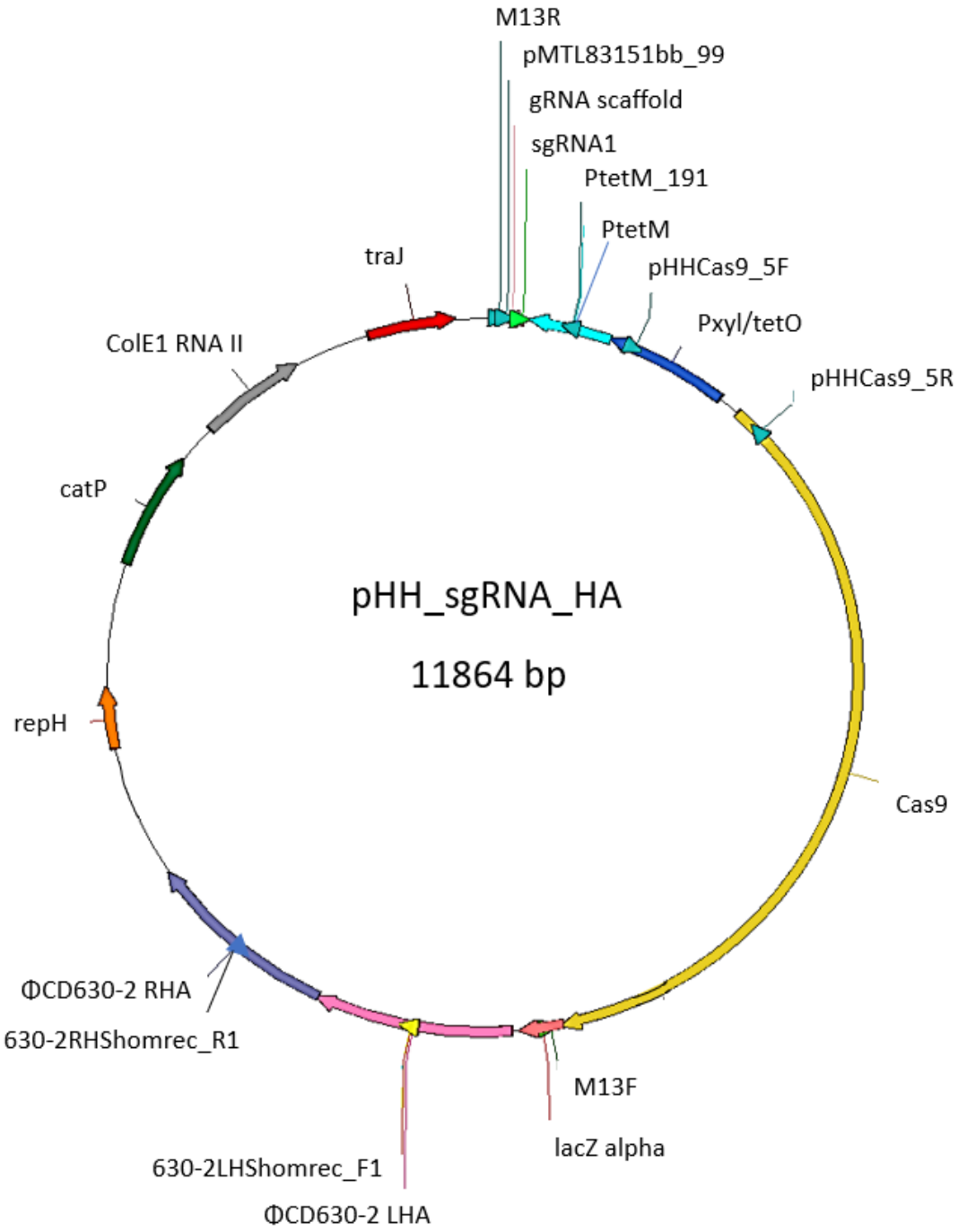


Figure 27. Schematic of pHH_sgRNA_HA for ΦCD630-2 deletion. Highlighted genes represented as arrows and primer locations where important represented as triangles.

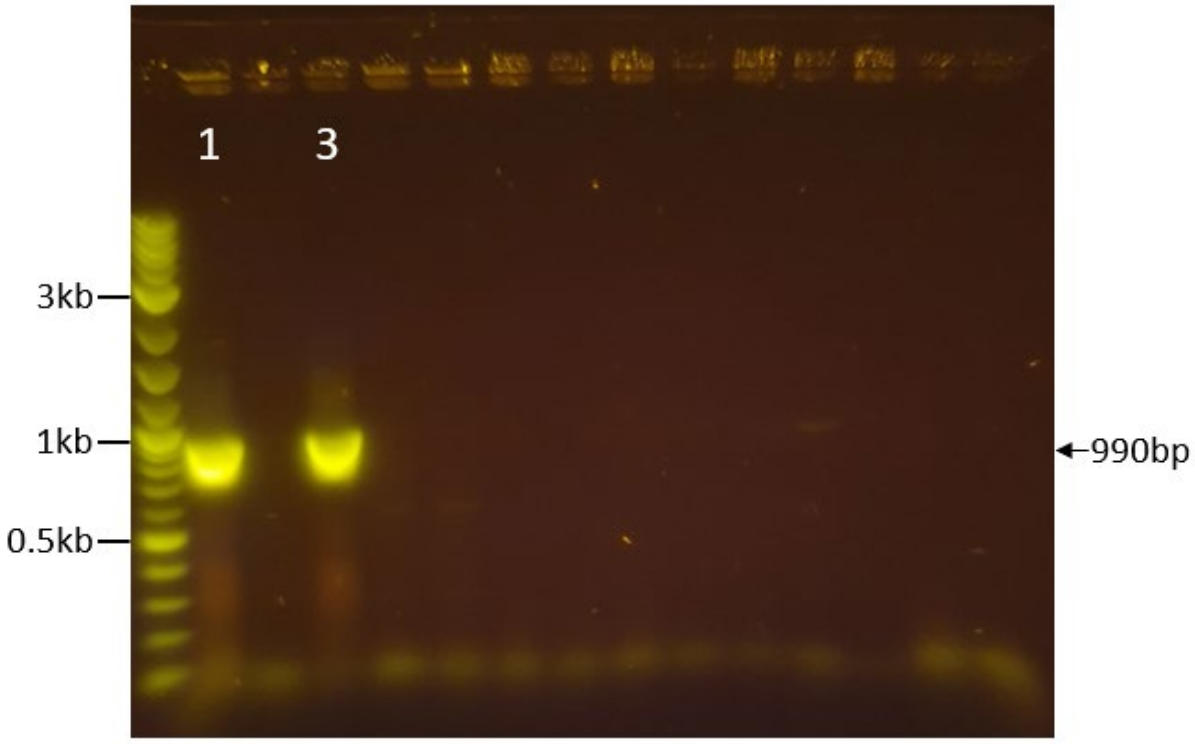


Figure 28. Colony PCR screening of transformed NEB10β for pHH_sgRNA_HA. The expected amplicon size was approximately 1kb which is indicated on the 1kb plus DNA ladder, the amplicons were run on a 1% gel at 80 volts for 1 h. Numbered wells indicate colonies which display bands of interest.

The NEB10_pHH_sgRNA_HA 1 and 3 colonies were subcultured and gDNA was extracted for a repeat PCR, which confirmed the above results (Figure 29).

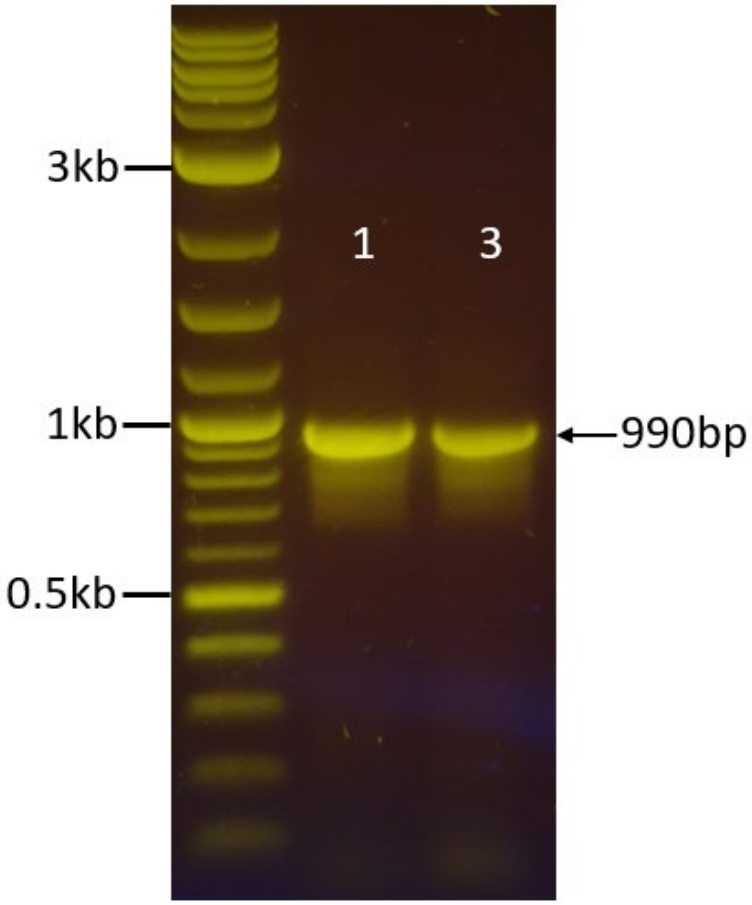


Figure 29. Re-screening of NEB10β colonies transformed with pHH_sgRNA_HA. Colonies 1 and 3 amplicons produced band that was more accurate to the expected 1kb amplicon size when electrophoresed on 1% agarose gel at 60 volts for 2 h and compared to the 1kb plus DNA ladder.

These two colonies were then named D10-pHHsgRNAHA6302_C1 and D10-pHHsgRNAHA6302_C3. Plasmid yield from D10-pHHsgRNAHA6302_C1 was 218.8 ng/uL. The pHH_sgRNA_HA recombinant plasmid (Figure 27), and control plasmids pHH_sgRNA (Figure 21), and pHH (Figure 8) were transformed into *E. coli* CA434. *E. coli* CA434 transformants were conjugated with CD630 in three experiments. For one of the experiments the viable count for the CA434 donor and CD630 recipient were $5.2x10^5$ and $3x10^7$ cfu/mL, respectively. Unfortunately, viable counts of transconjugants were not determined. Nine colonies of each recombinant and controls were screened for a vector-specific sequence via colony PCR using primers pHHCas9_5F and pHHCas9_5R, which target *cas9* to yield a 818 bp amplicon (Figure 30).

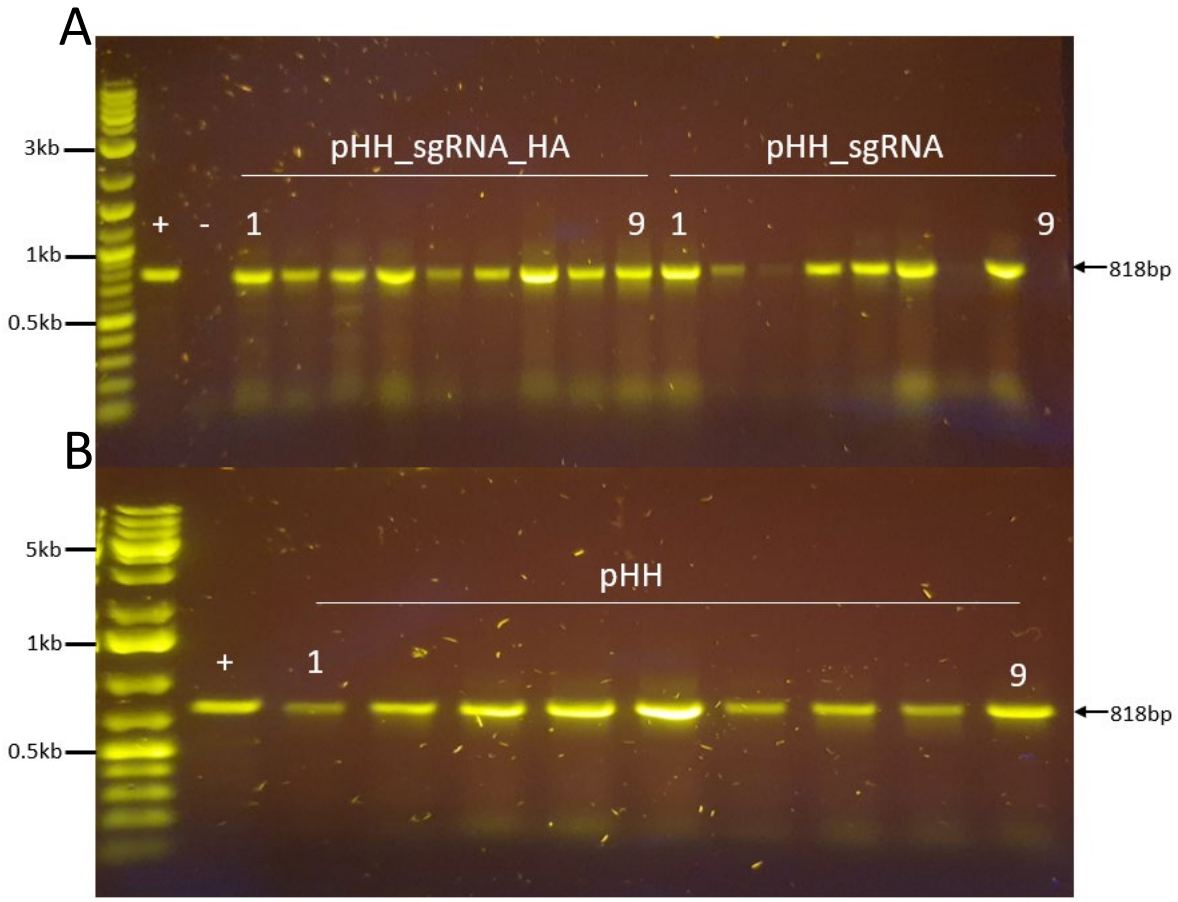


Figure 30. Colony PCR screening of CD630 transconjugants for the pHH derived plasmids. The expected amplicon size was 818bp which the positive control of pHH_sgRNA_HA displayed and the negative control CD630 genomic DNA did not produce a band as expected. The amplicons were run on an 1% agarose gel at 60 volts for 2 h, the DNA ladder was 1kb plus DNA ladder.

All colonies screened contained the expected plasmids except 3 colonies appeared to be negative for pHH_sgRNA (Figure 30). A selection of the positive colonies was carried forwards for long term storage, 6 out of 9 positive isolates of CD-pHH_sgRNA_HA which were colonies 1, 2, 4, 5, 8 and 9, 1 out of 6 CD-pHH_sgRNA control isolates which was colony 1, and 1 out of 9 positive isolates which was colony 5 for the CD-pHH control. From these isolates CD-pHH_sgRNA_HA_8, CD-pHH_sgRNA_1 and CD-pHH_5 were induced with aTC for *cas9* expression.

Induction of the *cas9* gene using aTC and the recovered cell viability was tested with varying concentrations and an incubation period of 72h, after which the colony forming units (cfu) were recorded in Table 4. There was no significant reduction in growth despite increasing aTC concentration. This was observed in the control plasmid cultures of CD-

pHH (empty plasmid without sgRNA), and CD-pHH_sgRNA (plasmid with sgRNA but without HA), as well as the culture of CD-pHH_sgRNA_HA (plasmid with sgRNA and HA).

Table 4. Viable count CD630 colonies (cfu/mL) containing the recombinant and control plasmids after 72h growth on 15-100 ng/mL aTC in BHIA.

aTC (ng/mL)	**CD-pHH_sgRNA_HA (cfu/mL)**	**CD-pHH_sgRNA (cfu/mL)**	**CD-pHH (cfu/mL)**
15	2.64×10^8	2.12×10^8	1.88×10^8
30	3.0×10^8	1.72×10^8	1.80×10^8
50	3.4×10^8	2.32×10^8	1.88×10^8
80	2.88×10^8	1.92×10^8	2.12×10^8
100	2.36×10^8	1.92×10^8	1.84×10^8

After inducing the *cas9* gene, random colonies were screened for the loss of the prophage by PCR detection of the ϕCD630-2 integrase gene, with an expected 221 bp amplicon if the prophage was present, and no amplicon if prophage was lost. The positive control was CD630 gDNA, and the negative was pHH empty plasmid (Figure 30). Compared to the CD630 gDNA positive control, the empty plasmid yielded a weak band of 221 bp. Five of 16 colonies containing pHH_sgRNA_HA had similarly weak or no amplicons, hence potentially negative for phage integrase. One of three colonies containing pHH_sgRNA, and one of three colonies containing pHH were also potentially negative for the integrase (Figure 31).

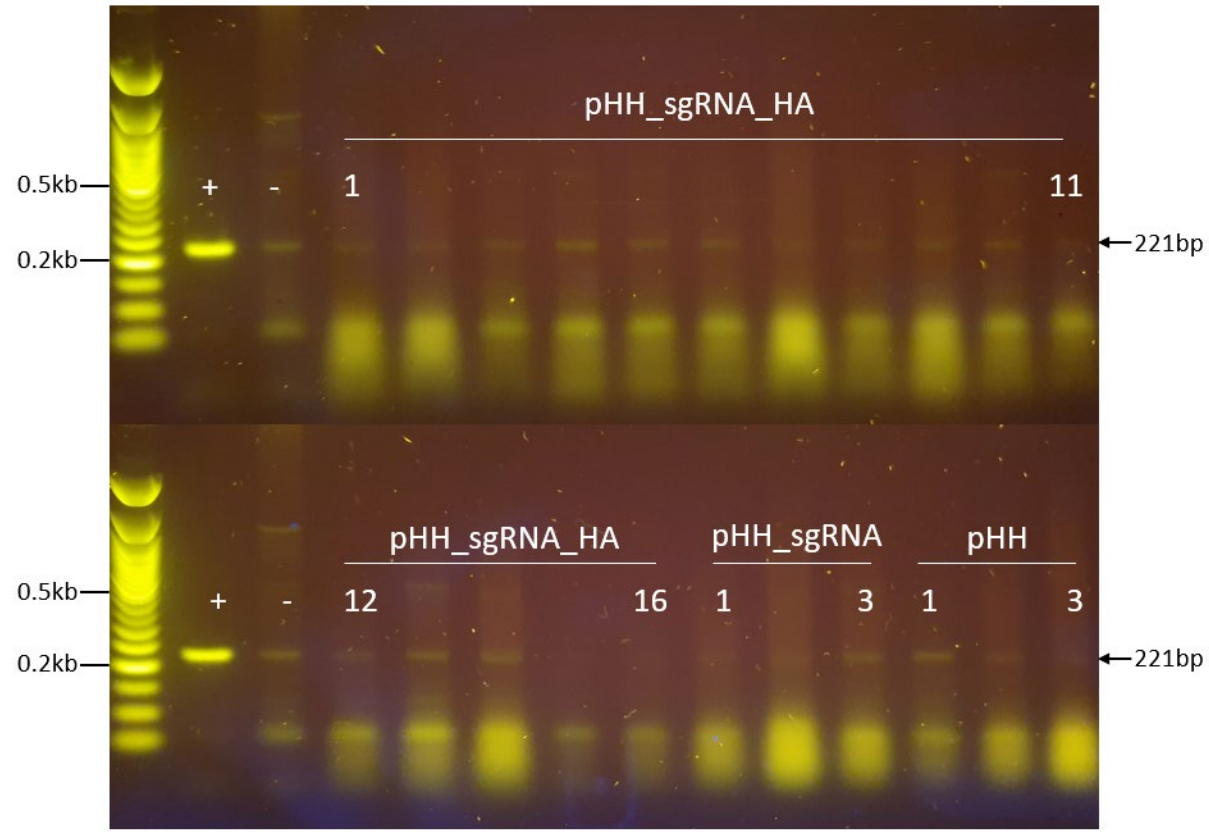


Figure 31. Colony PCR screening for the deletion of the ΦCD630-2 prophage. Positive control was CD630 genomic DNA, and negative control was pHH empty plasmid. The numbered wells under the headings of each plasmid correspond to the colonies that were tested. The phi630-2int_F2 and phi630-2int_R2 primers were expected to produce a 221bp amplicon in WT CD630. Gel electrophoresis was carried out in 2% agarose gel at 60 volts for 2 h.

Colony PCR revealed weak bands in the negative control of pHH. Therefore, to confirm whether the prophage was deleted, it would be necessary to lose the plasmid and re-screen colonies of interest, which were CD-pHH_sgRNA_HA7ΔΦ, 11ΔΦ, 12ΔΦ, 15ΔΦ and 16ΔΦ, CD-pHH_sgRNA1ΔΦ, and CD-pHH3ΔΦ. Although the control plasmids lacked either HA or both HA and sgRNA and were not expected to select for viable prophage deletants, we included it for further analysis based on results in Figure 31.

After six serial passages of these seven colonies in non-selective BHIB, followed by replica plating on non-selective and selective BHIA plates, all remained resistant to thiamphenicol. The plasmids were considered to be stably maintained, as its presence was confirmed by PCR with primers pHHCas9_5F and pHHCas9_5R that targeted the *cas9* gene on the plasmid to generate a 818bp amplicon (Figure 32). An unexpected band of

1.5 kb, presumed to be due to non-specific amplification, was observed in three out of 5 clones containing pHH_sgRNA_HA (11ΔΦ, HA12ΔΦ, HA15ΔΦ) and one clone containing pHH_sgRNA (1ΔΦ) and not detected in the one pHH clone. Further work is needed to confirm their identity.

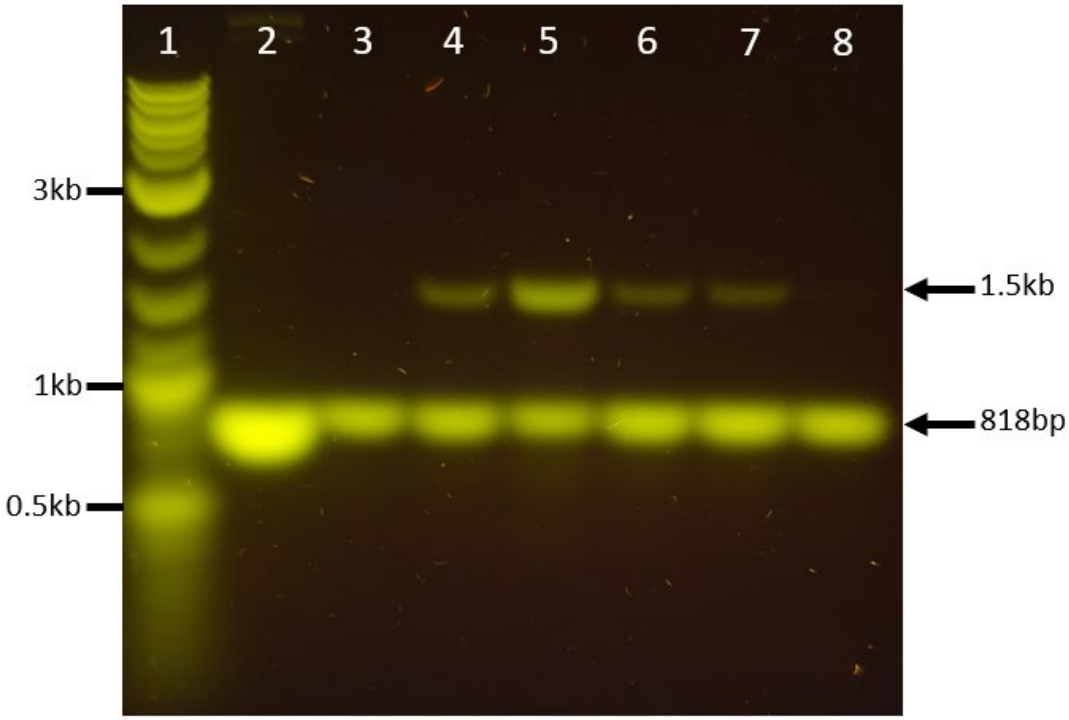


Figure 32. Colony PCR screening for *cas9* in putative prophage deletants. Gel electrophoresis was carried out in 1% agarose gel for 2h at 60v. Lane 1: 1kb plus DNA ladder, 2: positive control pHH, 3: CD-pHH_sgRNA_HA7ΔΦ, 4: CD-pHH_sgRNA_HA11ΔΦ, 5: CD-pHH_sgRNA_HA12ΔΦ, 6: CD-pHH_sgRNA_HA15ΔΦ, 7: CD-pHH_sgRNA1ΔΦ and 8: CD-pHH3ΔΦ.

4. Discussion

The purpose of this study was to delete the ΦCD630-1 and ΦCD630-2 prophages from a virulent CD630 isolate in an attempt to generate isogenic mutants that could be used to investigate the effect of the phage genome on the virulence, persistence, or transmission of the host bacterium. As a human pathogen that is commonly the cause of antibiotic-associated diahorrea, understanding the pathways that affect the bacterium in terms of the ability to cause disease, grow in the environment or pass between susceptible hosts.

In this study the allelic couple exchange system was unsuccessful in excising either of the prophages, and we found only 5 putative double crossover colonies in total; these colonies were presumptive single crossover clones that had lost thiamphenicol resistance upon negative selection by 5-FC. Theoretically, double-crossover clones should comprise a 50:50 mix of wild-type revertant clones, clones which still have the wild-type genome and recombinant clones in which the wild-type genome has been exchanged for the recombinant genome, where the prophage sequence has been deleted (Ehsaan, Kuehne, and Minton, 2016). All of the putative double crossover clones recovered in this study were wild-type revertant clones, combining this with the incredibly low 1-1.5% recovery frequencies seen in the literature would indicate that this method is not appropriate for the deletion of large genomic fragments such as the two CD630 prophages. Potentially this could highlight that the particular homologous region that mediated integration is the preferential recombination site, which would make the recovery of an event leading to a prophage excision particularly infrequent especially when considering the double crossover recovery frequencies of this study.

The method has worked for in-frame deletions of the *spoA*, *cwp84* and *mtlD* genes in both CD630Δ*erm* and R20291 strains (Ng *et al*., 2013). The percentage of wild type revertants for these deletions was between 34-75%, which when compared with the 100% recovery of presumed revertants of this study highlights a low mutation efficiency of the method. The largest fragment so far deleted in *C. difficile* was the 19.6 kb pathogenicity locus from the R20291 strain with a *codA*-based plasmid (Bilverstone, Kinsmore, Minton, & Kuehne, 2017) which is still less than half the size of either the ΦCD630-1 and ΦCD630-2 prophage genome.

An altered approach to the allelic-coupled exchange method used in this study was developed by Peltier *et al* (2020) after the work of this study had begun, in which the counter selection of double crossovers was modified so that cells retaining the altered pMTL-SC7315 plasmid either integrated or extrachromosomally were killed by an inducible overexpression of the CD2517.1 toxin. This toxin is part of a Type I toxin-antitoxin (TA) module which typically cause bacterial cell death or growth stasis by damaging the cell membrane, in CD630 the overexpression of the toxin led to significant growth arrest as well as a 9% increase in cell length (Maikova *et al*., 2018). The method was shown to be effective in deleting ΦCD630-2 from CD630 as the switch facilitated a greater recovery of deletant double crosses as any cells which contained the plasmid after growth allowing for double cross events would die as the expression of the *CD2517.1* gene is at toxic levels. They were also able to isolate ΦCD630-1 deletant mutants by inducing overexpression of the prophage excisionase contained on a plasmid. However, it was an incredibly infrequent event, as in less than 10% of the cell population the ΦCD630-1 prophage was excised from the genome but retained extrachromosomally (Peltier *et al*, 2020).

The Peltier study showed that allelic couple exchange is capable of prophage deletion, but the *CD2517.1* toxin gene, was a better selective marker than *codA*. The counter-selection eliminated both integrated and excised plasmid containing cells, which reduced false positive detection to as low as 0.1%. Finally, it vastly improved the recovery of colonies that had undergone double crossover events as 50 colonies were recovered which is a much higher recovery of colonies than previously recorded in this study or others (Ng *et al*., 2013; Peltier *et al*., 2020).

The CRISPR-Cas9 system has not been shown to delete a prophage from *C. difficile* yet, however it has been used to delete genes from the species. *selD* mutants were generated using an ~14kb plasmid containing *cas9* controlled by the *tetR* promoter which is induced with aTC (McAllister, Bouillaut, Kahn, Self, & Sorg, 2017) similar to the promoter in the pHH plasmid of this study. Aside from the different aTC inducible promoter, the pHH plasmid of this study also differed in replicons, the pJK02 plasmid was based on the pMTL84151 backbone which carries a pCD6 replicon derived from the *C. difficile* CD6 strain instead of a pCB102 replicon derived from the *C. butyricum* in the plasmid of this study. The pCD6 based plasmids were experimentally shown to be transferred at higher frequencies and maintained more effectively (Heap, Pennington, Cartman, & Minton,

2009; Purdy *et al*., 2002). A limitation in this study is that viable counts of pHH-derived transconjugants were not recorded. This would have allowed for the calculation of conjugation frequency; this data would indicate the uptake efficiency of the plasmids and allowed comparison to the literature.

At the start of this research, it was not known whether CRISPR-Cas9 could delete large fragments. In the course of this work, Hong *et al* (2018) reported difficulties when trying to delete large fragments using CRISPR-Cas9 in comparison to CRISPR-Cpf1. Firstly, the Cas9 protein was highly toxic to the host strains, more so than Cpf1. Secondly, their Cas9-harboring plasmids were usually large, resulting in a relatively low conjugation efficiency. It is challenging to detect ΦCD630-2 prophage deletion because the prophage is known to spontaneously excise (unpublished data), hence PCR of the regenerated phage integrated junctions (*attB* or *attP)* would not necessarily indicate a successful mutant by design. However, we could have designed primers to detect phage integrated junctions to confirm prophage deletion by negative evidence. The pHH plasmid was not lost despite passaging and this was unexpected because it is known to be unstable with 76.2 % ± 0.5 segregational stability in CD630 (Heap, Pennington, Cartman, & Minton, 2009). Perhaps a mutation in the replicon could lead to plasmid integration in the genome.

The concentration of aTC used to induce the *tetO* promoter in this study was less than previously used for similar induction in published literature; McAllister *et al*., (2017) induced the *tetR* promoter with 100ng/mL of aTC and in this study 30ng/mL was used. This was because increased inducer concentrations did not show changes to transconjugant viability.The recombinant plasmid in theory should have a higher rate of recovery as the homologous recombination with the homology arms will repair the double strand breaks that resulted from Cas9 cleavage, while the negative control plasmid which has Cas9 and sgRNA but lacking homology arms should result in non-viable cells. In order to understand whether 30 ng/mL aTC was effective, viability of cells containing the recombinant or negative control plasmid at varying aTC concentrations should be repeated. An uninduced control should also be included.

An appropriate method of promoter inducer exposure would also have been a beneficial investigation as the use of broth and the use of agar have both been used in literature. The use of broth for induction was recorded in McAllister *et al*., (2017) which resulted in the deletion of the 951 bp *selD* gene in 13 out of 64 total colonies tested, whilst the use of

agar plates for induction was recorded in Hong *et al*., (2018) for the deletion of multiple genomic elements, the most comparable to *selD* in target size being *fur* and *cwp66*, with 16 colonies out of 16 total were confirmed for having the 390 bp *fur* gene deleted. Based on this it seems that aTC induction on agar plates leads to a lower number of total colonies but a higher mutation rate. Plasmid integration of pRK6301 and pRK6302 into the CD630 genome was not confirmed by PCR due to time constraints. This would be helpful in confirming plasmid-mediated thiamphenicol resistance.

A limitation of the CRISPR-Cas9 system is that only one sgRNA sequence was tested. There is no guarantee that the sgRNA used in this study was an efficient target which would lead to an excision. It would be ideal to test multiple sgRNA sequences next to different PAM (Guo *et al*., 2018; J. Shen, Zhou, Chen, & Xiu, 2018). One method for multisite targeting is to transfer multiple plasmids bearing individual system requirements, for example one vector which bears the multiple sgRNA each under its own promoter and another vector bearing the *cas9* gene under an inducible promoter and homology arms for repair. Using multiple plasmids reduces the overall size of the vectors which should in turn allow for more efficient uptake and maintenance, this method was used to shown to be effective in modification of up to four loci in *E. coli* (Feng, Zhao, Zhang, Ding, & Bi, 2018). Another limitation was that the presence of target sequence for the sgRNA was never experimentally verified by Sanger sequencing; the sgRNA sequence was identified and selected based on sequence information from the GenBank database.

One possible improvement is to generate a plasmid bearing the Cpf1 nuclease; a system utilising this nuclease was capable of deleting the ΦCD630-2 prophage from the CD630 genome (Hong, Zhang, Cui, Wang, & Wang, 2018). Cpf1 was less cytotoxic than Cas9, and the plasmid bearing Cpf1 was smaller in size despite containing two sgRNA sequences at ~12.2kb compared to ~14.7kb for the plasmid bearing Cas9. This is possible as sgRNA expression in Cpf1 plasmid can be placed under a single promoter unlike Cas9 in which separate promoters would be needed for each sgRNA (Fonfara *et al*., 2016). The larger Cas9 plasmid failed to transfer by conjugation in the Hong *et al* (2018) paper, interestingly the plasmid of this study was smaller than either being ~11.9kb and was successfully conjugated without much difficulty, potentially the pBP1 replicon of pJZ180 did not allow for efficient transfer or maintenance by the cells after uptake.

It would have been ideal to delete the ΦCD630-1 prophage as the full extent of interactions this phage has with the host is yet to be studied. This could be achieved through a variety of means, generating a plasmid that is capable of doing so by following the improvements stated above such as multisite targeting. The overexpression of the excisionase gene and deletion of genes involved with prophage maintenance to prevent the retention in an extrachromosomal form could also be an effective method for generating this deletion (Peltier *et al*., 2020). These methods still have their limitations as large plasmids bearing sgRNA arrays are less likely to be successfully up taken by cells and overexpression of excisionase genes is only capable in phages that encode them; phages such as ΦCD630-2 which do not encode an excisionase will have to be deleted using other tools.

In conclusion, there is an expanding range of genetic engineering tools currently available which allow precise deletion of prophages. This is clearly an area of intense research, with two reports on prophage deletion in CD630 published by two different groups since the commencement of this project. Using these tools would allow the study of prophage contribution to pathogenicity in bacteria, not just limited to ΦCD630-1 and ΦCD630-2 in the CD630. Unfortunately, this study was unable to delete a prophage, however, some genetic tools, pHH_sgRNA_HA and pHH_sgRNA were generated and should be useful for further refinement.

5. References

Altenbuchner, J. (2016). Editing of the *Bacillus subtilis* Genome by the CRISPR-Cas9 System. *Applied and Environmental Microbiology*, *82*(17). https://doi.org/10.1128/AEM.01453-16

Andersen, J. M., Shoup, M., Robinson, C., Britton, R., Olsen, K. E. P., & Barrangou, R. (2016). CRISPR Diversity and Microevolution in *Clostridium difficile*. *Genome Biology and Evolution*, *8*(9). https://doi.org/10.1093/gbe/evw203

Bae, T., Baba, T., Hiramatsu, K., & Schneewind, O. (2006). Prophages of *Staphylococcus aureus* Newman and their contribution to virulence. *Molecular Microbiology*, *62*(4), 1035–1047. https://doi.org/10.1111/j.1365-2958.2006.05441.x

Banawas, S. S. (2018). *Clostridium difficile* Infections: A Global Overview of Drug Sensitivity and Resistance Mechanisms. *BioMed Research International*, Vol. 2018. Hindawi Limited. https://doi.org/10.1155/2018/8414257

Bilverstone, T. W., Kinsmore, N. L., Minton, N. P., & Kuehne, S. A. (2017). Development of *Clostridium difficile* R20291ΔPaLoc model strains and in vitro methodologies reveals CdtR is required for the production of CDT to cytotoxic levels. *Anaerobe*, *44*. https://doi.org/10.1016/j.anaerobe.2017.01.009

Bradshaw, W. J., Kirby, J. M., Roberts, A. K., Shone, C. C., & Acharya, K. R. (2017). Cwp2 from Clostridium difficile exhibits an extended three domain fold and cell adhesion in vitro. *The FEBS journal*, *284*(17), 2886–2898. https://doi.org/10.1111/febs.14157

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., & Brüssow, H. (2003). Prophage genomics. *Microbiology and molecular biology reviews : MMBR*, *67*(2), 238–276. https://doi.org/10.1128/MMBR.67.2.238-276.2003

Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T., & Minton, N. P. (2012). Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. *Applied and Environmental Microbiology*, *78*(13), 4683–4690. https://doi.org/10.1128/AEM.00249-12

Chen, Y. Y., Wang, J. T., Lin, T. L., Gong, Y. N., Li, T. H., Huang, Y. Y., & Hsieh, Y. C. (2019). Prophage Excision in *Streptococcus pneumoniae* Serotype 19A ST320 Promote Colonization: Insight Into Its Evolution From the Ancestral Clone Taiwan 19F-14 (ST236). *Frontiers in microbiology*, *10*, 205. https://doi.org/10.3389/fmicb.2019.00205

Cobb, R. E., Wang, Y., & Zhao, H. (2015). High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synthetic Biology*, *4*(6). https://doi.org/10.1021/sb500351f

Darkoh, C., DuPont, H. L., Norris, S. J., & Kaplan, H. B. (2015). Toxin synthesis by *Clostridium difficile* is regulated through quorum signaling. *mBio*, *6*(2), e02569. https://doi.org/10.1128/mBio.02569-14

Desjardins, P., & Conklin, D. (2010). NanoDrop Microvolume Quantitation of Nucleic Acids. *Journal of Visualized Experiments*, (1). https://doi.org/10.3791/2565

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., … Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2). https://doi.org/10.1038/nbt.3437

Ehsaan, M., Kuehne, S. A., & Minton, N. P. (2016). *Clostridium difficile* (Second Edition, Vol. 1476; A. P. Roberts & P. Mullany, Eds.). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-6361-4

Feng, X., Zhao, D., Zhang, X., Ding, X., & Bi, C. (2018). CRISPR/Cas9 Assisted Multiplex Genome Editing Technique in *Escherichia coli*. *Biotechnology Journal*, *13*(9). https://doi.org/10.1002/biot.201700604

Fonfara, I., Richter, H., Bratoviä, M., le Rhun, A., & Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, *532*(7600), 517–521. https://doi.org/10.1038/nature17945

Fortier, L. C., & Moineau, S. (2007). Morphological and genetic diversity of temperate phages in *Clostridium difficile*. *Applied and environmental microbiology*, *73*(22), 7358–7366. https://doi.org/10.1128/AEM.00582-07

Garneau, J. R., Sekulovic, O., Dupuy, B., Soutourina, O., Monot, M., & Fortier, L. C. (2018). High Prevalence and Genetic Diversity of Large phiCD211 (phiCDIF1296T)-Like Prophages in *Clostridioides difficile*. *Applied and environmental microbiology*, *84*(3), e02164-17. https://doi.org/10.1128/AEM.02164-17

Goh, S., Ong, P. F., Song, K. P., Riley, T. v, & Chang, B. J. (2007). The complete genome sequence of *Clostridium difficile* phage φC2 and comparisons to φCD119 and inducible prophages of CD630. *Microbiology*, *153*(3). https://doi.org/10.1099/mic.0.2006/002436-0

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

Guo, J., Wang, T., Guan, C., Liu, B., Luo, C., Xie, Z., … Xing, X.-H. (2018). Improved sgRNA design in bacteria via genome-wide activity profiling. *Nucleic Acids Research*, *46*(14). https://doi.org/10.1093/nar/gky572

Haeusser, D. P., Schwartz, R. L., Smith, A. M., Oates, M. E., & Levin, P. A. (2004). EzrA prevents aberrant cell division by modulating assembly of the cytoskeletal protein FtsZ. *Molecular microbiology*, *52*(3), 801–814. https://doi.org/10.1111/j.1365-2958.2004.04016.x

Hargreaves, K. R., Kropinski, A. M., & Clokie, M. R. (2014). What does the talking?: quorum sensing signalling genes discovered in a bacteriophage genome. *PLoS One*, *9*(1), e85131. https://doi.org/10.1371/journal.pone.0085131

Heap, J. T., Pennington, O. J., Cartman, S. T., & Minton, N. P. (2009). A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods*, *78*(1). https://doi.org/10.1016/j.mimet.2009.05.004

Higgins, D., & Dworkin, J. (2012). Recent progress in *Bacillus subtilis* sporulation. *FEMS microbiology reviews*, *36*(1), 131–148. https://doi.org/10.1111/j.1574-6976.2011.00310.x

Hong, W., Zhang, J., Cui, G., Wang, L., & Wang, Y. (2018). Multiplexed CRISPR-Cpf1-Mediated Genome Editing in *Clostridium difficile* toward the understanding of pathogenesis of *C. difficile* infection. *ACS Synthetic Biology*, *7*(6), 1588–1600. https://doi.org/10.1021/acssynbio.8b00087

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., … Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9). https://doi.org/10.1038/nbt.2647

Krebs, J. E., Goldstein, E. S., & Kilpatrick, S. T. (2018). Lewin's genes XII . In Chapter 27 Phage Strategies (12th ed., pp. 777–803). Burlington: Jones & Bartlett Learning.

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

Knight, D. R., Elliott, B., Chang, B. J., Perkins, T. T., & Riley, T. V. (2015). Diversity and Evolution in the Genome of *Clostridium difficile*. *Clinical microbiology reviews*, *28*(3), 721–741. https://doi.org/10.1128/CMR.00127-14

Koonin, E. v., Makarova, K. S., & Zhang, F. (2017, June 1). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, Vol. 37, pp. 67–78. Elsevier Ltd. https://doi.org/10.1016/j.mib.2017.05.008

Li, T., Zhang, Y., Dong, K., Kuo, C.-J., Li, C., Zhu, Y.-Q., … Zhu, Y. (2020). Isolation and Characterization of the Novel Phage JD032 and Global Transcriptomic Response during JD032 Infection of *Clostridioides difficile* Ribotype 078. *MSystems*, *5*(3). https://doi.org/10.1128/mSystems.00017-20

Maikova, A., Kreis, V., Boutserin, A., Severinov, K., & Soutourina, O. (2019). Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. *Applied and Environmental Microbiology*, *85*(20). https://doi.org/10.1128/AEM.01416-19

Maikova, A., Peltier, J., Boudry, P., Hajnsdorf, E., Kint, N., Monot, M., … Soutourina, O. (2018). Discovery of new type I toxin–antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*. *Nucleic Acids Research*, *46*(9). https://doi.org/10.1093/nar/gky124

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., … Koonin, E. v. (2011, June). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*, Vol. 9, pp. 467–477. https://doi.org/10.1038/nrmicro2577

Matos, R. C., Lapaque, N., Rigottier-Gois, L., Debarbieux, L., Meylheuc, T., Gonzalez-Zorn, B., … Serror, P. (2013). *Enterococcus faecalis* Prophage Dynamics and Contributions to Pathogenic Traits. *PLoS Genetics*, *9*(6). https://doi.org/10.1371/journal.pgen.1003539

McAllister, K. N., Bouillaut, L., Kahn, J. N., Self, W. T., & Sorg, J. A. (2017). Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-15236-5

Morvan, C., Folgosa, F., Kint, N., Teixeira, M. and Martin-Verstraete, I. (2021), Responses of Clostridia to oxygen: from detoxification to adaptive strategies. Environ Microbiol, 23: 4112-4125. https://doi.org/10.1111/1462-2920.15665

Ng, Y. K., Ehsaan, M., Philip, S., Collery, M. M., Janoir, C., Collignon, A., … Minton, N. P. (2013). Expanding the Repertoire of Gene Tools for Precise Manipulation of the *Clostridium difficile* Genome: Allelic Exchange Using *pyrE* Alleles. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0056051

Oliveira Paiva, A., de Jong, L., Friggen, A., Smits, W., & Corver, J. (2019). The C-terminal domain of Clostridioides difficile TcdC is exposed on the bacterial cell surface. doi: 10.1101/872283

Paredes-Sabja, D., Shen, A., & Sorg, J. A. (2014). *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. *Trends in microbiology*, *22*(7), 406–416. https://doi.org/10.1016/j.tim.2014.04.003

Peltier, J., Hamiot, A., Garneau, J. R., Boudry, P., Maikova, A., Hajnsdorf, E., … Soutourina, O. (2020). Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in *Clostridioides difficile*. *Communications Biology*, *3*(1). https://doi.org/10.1038/s42003-020-01448-5

Pettit, L. J., Browne, H. P., Yu, L., Smits, W., Fagan, R. P., Barquist, L., … Lawley, T. D. (2014). Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. *BMC Genomics*, *15*(1). https://doi.org/10.1186/1471-2164-15-160

Pfeifer, E., Michniewski, S., Gätgens, C., Münch, E., Müller, F., Polen, T., … Frunzke, J. (2019). Generation of a Prophage-Free Variant of the Fast-Growing Bacterium *Vibrio natriegens*. *Applied and Environmental Microbiology*, *85*(17). https://doi.org/10.1128/AEM.00853-19

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

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., … Parkhill, J. (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genetics*, *38*(7). https://doi.org/10.1038/ng1830

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

Sekulovic, O., Meessen-Pinard, M., & Fortier, L. C. (2011). Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. *Journal of bacteriology*, *193*(11), 2726–2734. https://doi.org/10.1128/JB.00787-10

Selle, K., Fletcher, J. R., Tuson, H., Schmitt, D. S., McMillan, L., Vridhambal, G. S., Rivera, A. J., Montgomery, S. A., Fortier, L. C., Barrangou, R., Theriot, C. M., & Ousterout, D. G. (2020). *In Vivo* Targeting of *Clostridioides difficile* Using Phage-Delivered CRISPR-Cas3 Antimicrobials. *mBio*, *11*(2), e00019-20. https://doi.org/10.1128/mBio.00019-20

Shen, A. (2012).*Clostridium difficile* Toxins: Mediators of Inflammation. *Journal of Innate Immunity*, *4*(2). https://doi.org/10.1159/000332946

Shen, J., Zhou, J., Chen, G.-Q., & Xiu, Z.-L. (2018). Efficient Genome Engineering of a Virulent *Klebsiella* Bacteriophage Using CRISPR-Cas9. *Journal of Virology*, *92*(17). https://doi.org/10.1128/JVI.00534-18

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1). https://doi.org/10.1038/msb.2011.75

Slater, R.T., Frost, L.R., Jossi, S.E. *et al. Clostridioides difficile* LuxS mediates inter-bacterial interactions within biofilms. *Sci Rep* **9,** 9903 (2019). https://doi.org/10.1038/s41598-019-46143-6

Soutourina, O. A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., … Martin-Verstraete, I. (2013). Genome-Wide Identification of Regulatory RNAs in the Human Pathogen *Clostridium difficile*. *PLoS Genetics*, *9*(5). https://doi.org/10.1371/journal.pgen.1003493

Thomason, L., Court, D. L., Bubunenko, M., Costantino, N., Wilson, H., Datta, S., & Oppenheim, A. (2007). Recombineering: Genetic Engineering in Bacteria Using Homologous Recombination. *Current Protocols in Molecular Biology*, *78*(1). https://doi.org/10.1002/0471142727.mb0116s78

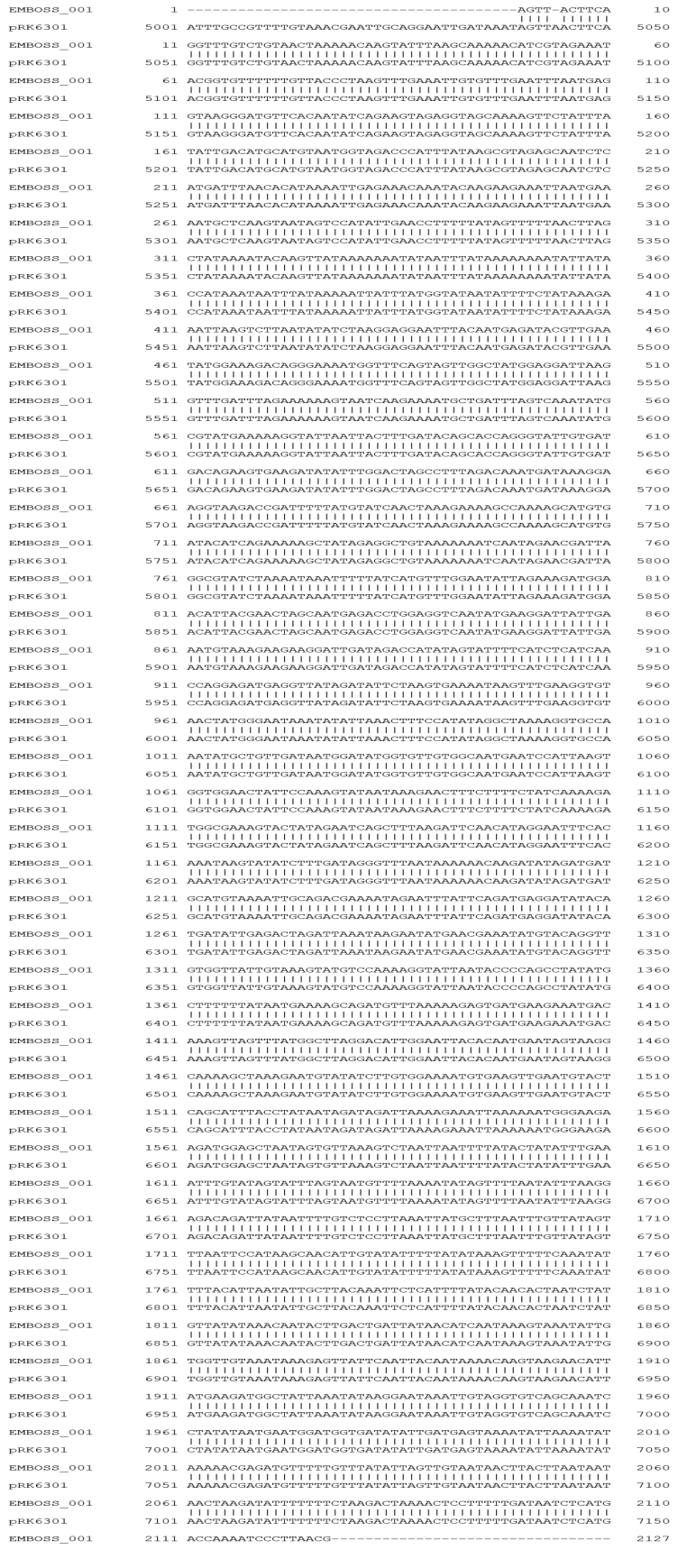
Wang, Y., Zhang, Z.-T., Seo, S.-O., Lynn, P., Lu, T., Jin, Y.-S., & Blaschek, H. P. (2016). Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable "Clean" Mutant Selection in *Clostridium beijerinckii* as an Example. *ACS Synthetic Biology*, *5*(7). https://doi.org/10.1021/acssynbio.6b00060

Wang, X., Kim, Y., Ma, Q. *et al.* Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1,** 147 (2010). https://doi.org/10.1038/ncomms1146

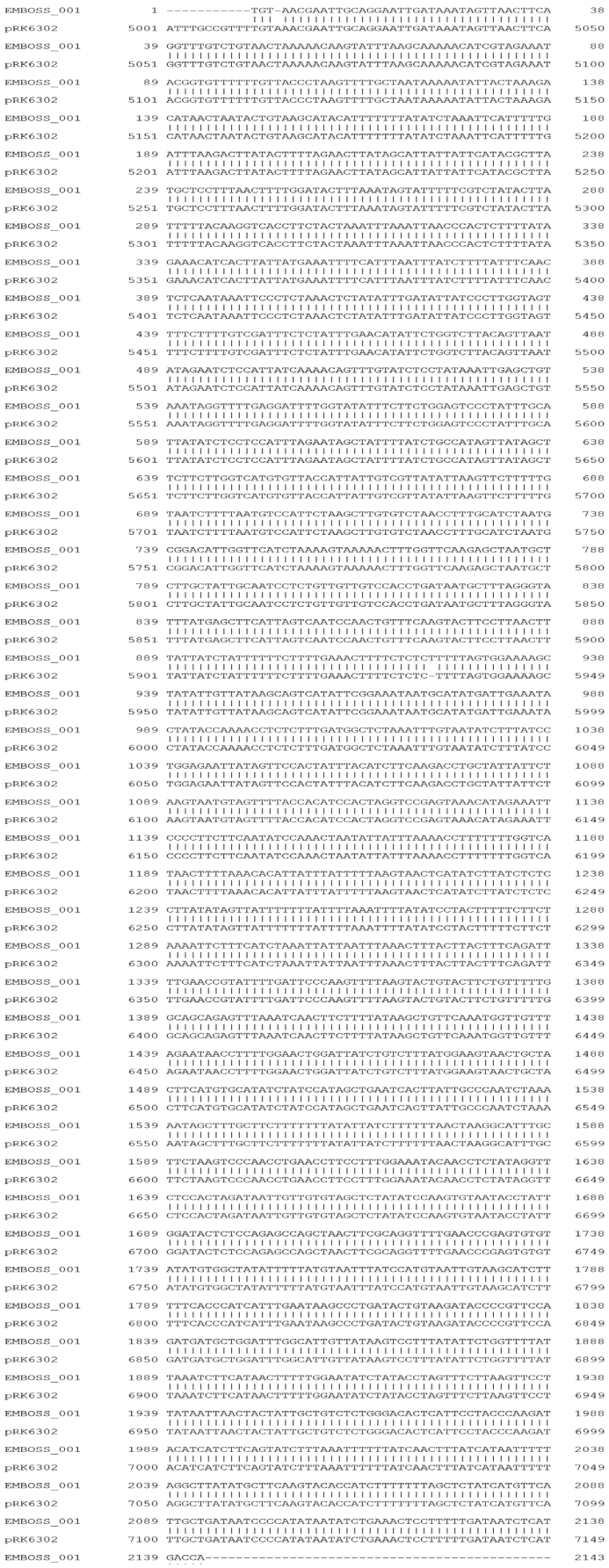
Williams, D. R., Young, D. I., & Young, M. (1990). Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. *Journal of General Microbiology*, *136*(5). https://doi.org/10.1099/00221287-136-5-819

Wust, J., & Hardegger, U. (1983). Transferable Resistance to Clindamycin, Erythromycin, and Tetracycline in *Clostridium difficile*. *Antimicrobial Agents and Chemotherapy*, *23*(5). https://doi.org/10.1128/AAC.23.5.784

6. Appendices



Appendix 1. Clustal-Omega alignment between the Sequenced contig of T8 pRK6301 compared to the theoretical insilico sequence of pRK6301.

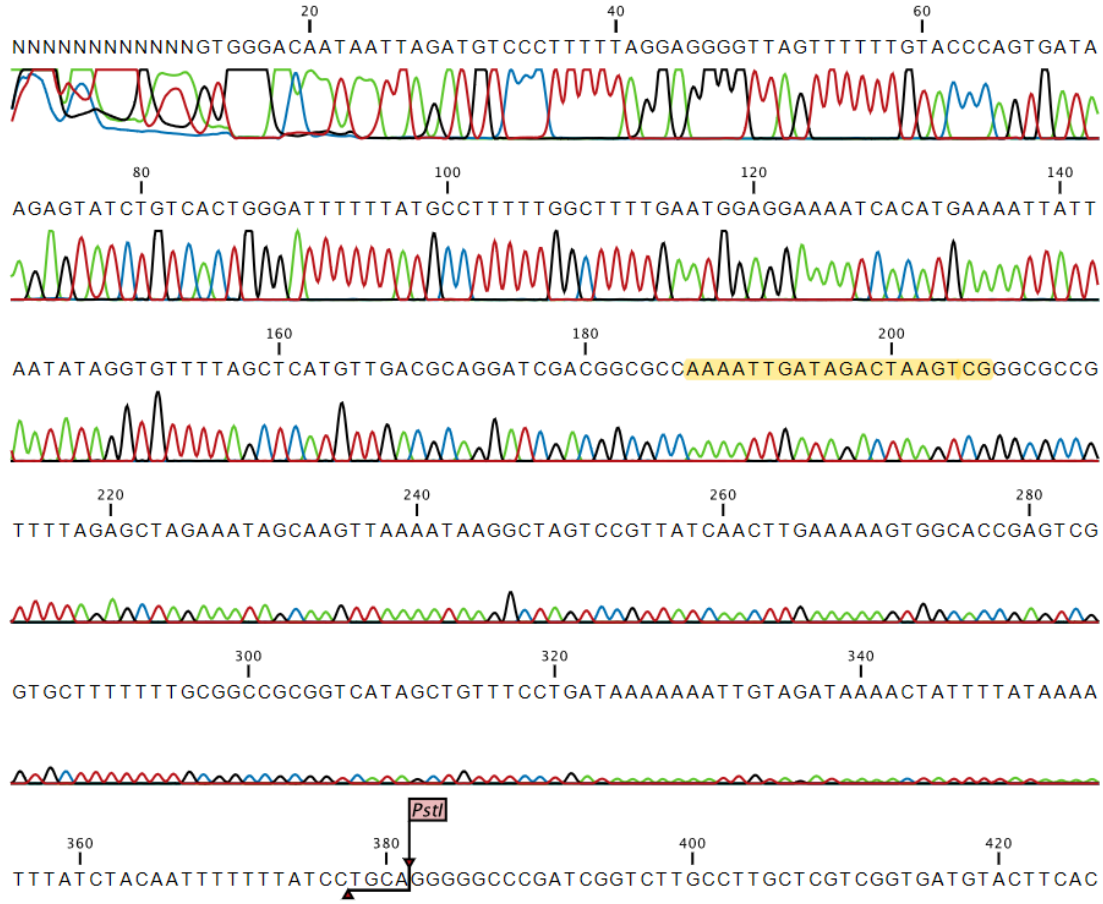


Appendix 2. Clustal-Omega alignment between the Sequenced contig of T14 pRK6302 compared to the theoretical insilico sequence of pRK6302.

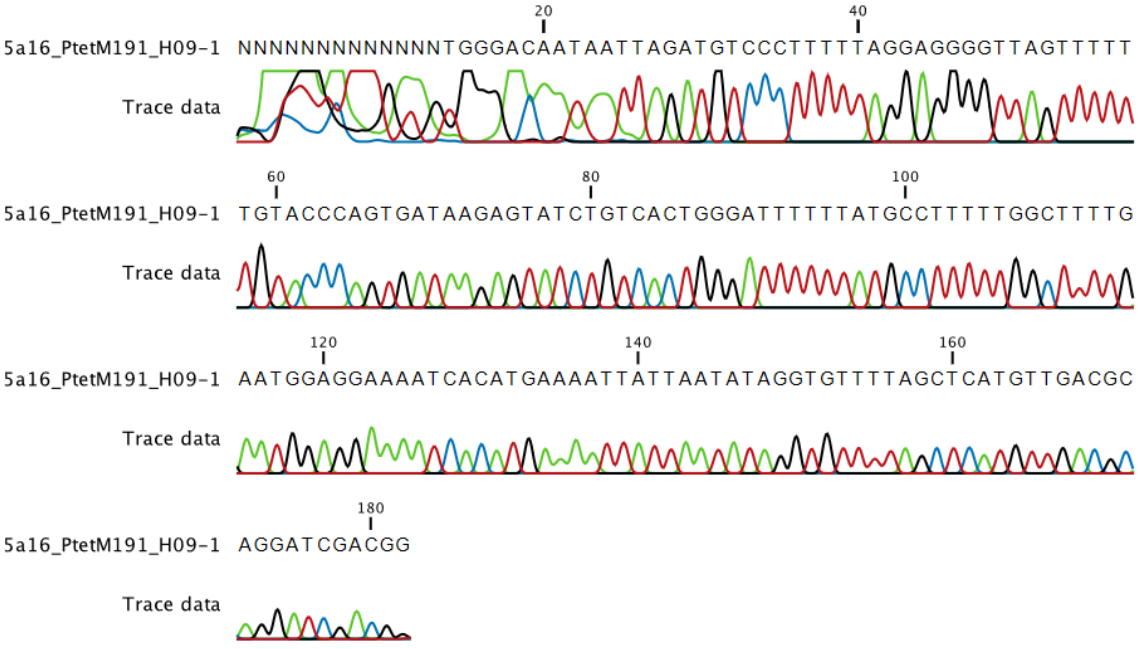
Position	Strand	Sequence	PAM	Specificity Score	Efficiency Score
3428178	1	ATAGGAATCTATAATAAACG	AGG	99.9	76.5
3427749	1	AAAATTGATAGACTAAGTCG	TGG	99.9	67.2
3428328	1	CTTTTAAAATGTAAGGAGTG	TGG	99.6	64.0
3427974	1	ATAAGTGGTACAATTCCCAT	AGG	99.9	62.9
3427577	1	TAGCATAAGGACACAAATAG	AGG	99.4	60.5
3427694	1	ACAACGATTATTAAGAGATG	TGG	100.0	58.1
3427959	1	AGACGTAAAGAAGTTATAAG	TGG	99.6	57.7
3427784	1	TTATAAAATAATGGAAACCT	TGG	100.0	57.4
3427845	1	TACGATTCTAGCACTGCTGC	TGG	100.0	57.4
3427790	-1	ATCACACTTATGTTTTTCCA	AGG	99.8	56.7
3427638	1	GTTGGTAAATACATTGATGA	AGG	99.8	56.2
3427564	1	CTTTAAATGGAGATAGCATA	AGG	99.7	55.1
3428488	-1	ACCATATGTTTATGTAATTG	TGG	99.7	52.9
3428771	1	GATGTATAGTGATTTAAGCA	GGG	99.8	51.7
3428361	1	AAGGGTAATTCTAGTAATGT	TGG	100.0	51.3
3428071	1	ACGAGAAATCAGGTTCTGTT	AGG	99.3	50.7
3427804	1	TGGAAAAACATAAGTGTGAT	TGG	100.0	48.3
3428061	1	TTTGACGAATACGAGAAATC	AGG	100.0	47.7
3427978	-1	TTTTCTATTTTGTAACCTAT	GGG	99.7	47.6
3428103	1	GAAACAATTAATAATCTTCA	TGG	99.5	47.1
3427722	-1	CAATTTTAGTCATTAAAACT	AGG	99.2	46.5
3428498	1	ACCACAATTACATAAACATA	TGG	99.8	46.5
3428160	1	CTAAGAAATGAACTTTATAT	AGG	99.4	43.3
3428770	1	AGATGTATAGTGATTTAAGC	AGG	98.8	42.6
3428321	1	TTCTGGTCTTTTAAAATGTA	AGG	98.8	42.1
3428690	1	GGATAATATAACTGAAGAAC	AGG	99.9	40.4
3428469	-1	GTGGTTTTAAGTTAGTCAAA	AGG	99.6	40.3
3428342	1	GGAGTGTGGTTATACATTAA	AGG	99.0	40.1

3427620	1	GAGAATGATTTTAACATAGT	TGG	99.8	38.6
3427979	-1	ATTTTCTATTTTGTAACCTA	TGG	100.0	38.2
3428612	1	AATCAAAGATTTATATTTAG	AGG	96.3	37.9
3428343	1	GAGTGTGGTTATACATTAAA	GGG	98.5	37.6
3427775	1	TAAAAACTATTATAAAATAA	TGG	89.8	36.5
3428247	-1	TATTTTTTTCTAATATTCTT	TGG	91.4	34.4
3427675	-1	CTTAATAATCGTTGTAAGTT	AGG	99.6	33.7
3428669	1	TGAAAAATTACAATCTCAAC	TGG	99.8	31.9
3428281	-1	TGATAATGTATGTTTTTATT	AGG	99.1	31.5
3428215	-1	AATTGTTTTTTTGATATAAT	AGG	95.5	28.6
3427551	1	ACAGAAGAACAAGCTTTAAA	TGG	99.0	28.0
3428304	1	ATACATTATCATATTTTTTC	TGG	94.9	26.0
3428791	1	GGGTTGAACGCAGACGTTTT	TGG	99.9	24.8
3427820	-1	AGAATCGTAGTTTTCTAAAA	TGG	98.6	18.1
3427928	1	AAGAATAAAATTTGTTTTTC	AGG	97.5	8.4

Appendix 3. Benchling CRISPR single guide sgRNA for Cas9. Highlighted is the sgRNA that was used in this study.



Appendix 4. Sequencing chromatograph for D5a-pHH_sgRNA_C10. Highlighted is the expected sgRNA sequence.



Appendix 5. Sequencing chromatograph for D5a-pHH_sgRNA_C16. The signal was lost just before the sgRNA sequence insertion site, which suggests the presence of a secondary structure.