Transposon mutagenesis in RT 078 *Clostridioides difficile*

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy
March 2022

Marcel te Vrugt

Declaration

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

Marcel te Vrugt, March 2022

Abstract

Clostridioides difficile is a Gram-positive, spore-forming, anaerobic bacterium and a major cause of healthcare associated diarrhoea. Significant increases in the incidence of hypervirulent strains, such as those belonging to PCR ribotype (RT) 027, and increased antibiotic resistance have formed the focus of current *C. difficile* clinical research. Hypervirulent strains belonging to RT 078, in contrast, have received comparative less attention, despite the fact that they are widely recognized as being zoonotic, with a particular association with pigs. A greater understanding of RT 078 strains would benefit from the implementation of forward genetic approaches. Here we sought to implement Transposon directed insertion-site sequencing (TraDIS), a high throughput method able to define gene essentiality under niche-specific conditions, to elucidate physiological changes such as sporulation and germination in RT 078 strains. As effective DNA transfer is a prerequisite for TraDIS implementation, the most efficient strains as both donor and recipient in conjugation were identified. Applying next generation sequencing technologies on 10 clinical isolates and subsequent methylome analysis demonstrated that although the tested strains of RT 078 were genetically similar (up to 99.99%), they possess a variety of potential Restriction-Modification (R-M) barriers. One of these R-M systems was circumvented using the novel *Escherichia coli* donor strain, sExpress. Improved DNA transformability in *C. difficile* RT 078 strain CD9301 made it an optimal target for further genetic manipulations and subsequent TraDIS analysis. Subsequently, several transposon delivery systems were evaluated, based on their potential to mediate random transposon insertion and reliable plasmid loss, to prevent interference of the transposase during downstream experiments in *C. difficile.* The Tet-inducible transposon vector pRF215, performed best in CD9301. Based on this plasmid system, *the* novel vector pMTL-MtV10 was created, which was additionally equipped with I-*SceI* digestion sites, to achieve increased plasmid clearance during library preparation. Using both plasmids, genes essential for growth in rich media were determined. In total, 448 essential genes were predicted. The incorporation of I-*SceI* sites into pMtV-10 did not, however, improves plasmid loss during the TraDIS library preparation. A further 398 genes were predicted to be essential for sporulation. The number of genes identified is most likely an underestimate as the manual cut-off used to predict essentiality lacks sensitivity. The described findings lay the ground work necessary for determining essentiality in RT 078 and improving our understanding of this important ribotype.

Acknowledgements

According to the Oxford English Dictionary the word acknowledgement means: “to recognize or confess someone or something.” I prefer to use the German term “Danksagung”, meaning “expression of gratitude”. During 4 years of working in the Clostridia Research Group, I feel that just recognizing the people who have accompanied, helped, and supported me along the way doesn’t do their relentless efforts justice. Gratitude, however, is more on-point.

Principally my gratitude goes to Prof. Nigel Minton, who gave me the chance to conduct this research in his laboratory and together with Dr. Ruth Griffin supported me along the process.

To describe how grateful, I am for the help and support that I have received from everyone in the Clostridia Research Group is beyond scope. Thanks to Dr Andrew Dempster, Dr Patrick Ingle, Dr Raquel Mesquita Rodrigues, and Dr Terry Bilverstone for your huge endorsement and advice on the lab bench and beyond. Special appreciations must go to Dr. Craig Woods, who inexhaustibly mentored and guided me regardless of the problem, time, or day. Last but not least, I thank my family and my old and new friends that have supported me along the way.

List of Abbreviations

%	Percentage
µL	Microlitres
˚C	Degrees Celsius
µg	Microgram
µM	Micromolar
ACE	Allele-Coupled Exchange
ANI	Average Nucleotide Identity
aTc	Anhydrotetracycline
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CDI	*Clostridioides difficile* Infection
CFU	Colony Forming Unit
Cm	Chloramphenicol
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
Em	Erythromycin
g	Gram
iPCR	inverse Polymerase Chain Reaction
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase pair
LB	Luria-Bertani medium
M	Molar
mg	Milligram
mL	Millilitres
mM	Millimolar
MTase	Methyltransferase
OD	Optical Density
ORF	Open Reading Frame
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PCR-RT	Polymerase Chain Reaction Ribotype
REase	Restriction Endonuclease
RM-system	Restriction Modification System
Tc	Tetracycline
Tm	Thiamphenicol

List of Figures

List of Tables

Table of contents

Chapter 1

Introduction

1.1 *Clostridioides difficile*

Clostridium difficile was first described in 1935, when Hall and O'Toole identified a Gram-positive obligate anaerobic bacterium, isolated from stool samples of healthy infants. The strain was referred to as *Bacillus difficilis* due to difficulties in isolating and culturing the organism. Eventually, they demonstrated progression of disease in Guinea pigs to the isolated bacteria [1]. In 2013, phylogenetic analysis suggested that *C. difficile* should be reclassified as *Peptoclostridium difficile.* While in 2016, Lawson *et al.,* proposed an alternative reclassification of the organism to *Clostridioides difficile* [2] to maintain public and scientific awareness in the form of numerous studies and literature referring to *C. difficile* [3]. In 1962, *C. difficile* was described as a human pathogen for the first time and identified as the causative agent of pseudomembranous colitis, even though the disease was known since 1893 [4, 5]. Today *C. difficile* is the leading cause of nosocomial diarrhoeas world-wide [6, 7]. In 2008, a severe increase of *C. difficile* infections (CDI) was noted in Europe and northern America, linked to the progression of new emerging hyper virulent epidemic, NAPI/B1/027 type strains [8].

The number of *C. difficile* associated infections reached a peak in England and Wales in 2007. Of note was the increased severity of symptoms and disease progression requiring surgical interventions, such as colectomy, resulting in increased mortality [9].

Nowadays, measures such as the isolation of infected patients, alterations in antibiotic prescription policy and improved personal hygiene has led to a decline of CDI and patient related mortality [10]. Nevertheless, *C. difficile* remains a leading cause of antibiotic-associated diarrhoea and a significant burden with 500,000 cases and around 14,000 deaths from CDI in the USA alone [11]. Healthcare providers are confronted with estimated annual costs of $2 billion in Europe and $3 billion in the US caused by extended lengths of patients stay within healthcare settings and increased treatment costs resulting from CDI [12].

1.1.1 *C. difficile* infection

C. difficile can be found in the intestinal tracts of both humans and animals and its environmentally ubiquitous spores can also be isolated from food. Importantly, people with an adequate immune response will either eliminate the infection and/or become asymptomatic carriers. In immunocompromised patients with interrupted protective intestinal microbiota, ingested or already resident *C. difficile* hyper-colonize the gastrointestinal tract and vegetative cells produce toxins and form transmissible spores [13]. Nowadays 15% of patients treated with antibiotics develop antibiotic-associated diarrhoeas. In patients, *C. difficile* hyper colonization is accountable for 20 to 30% of diarrhoea cases during or after

antibiotics therapy [14, 15]. CDI related symptoms are diverse and range from mild to severe diarrhoea. Eventually these complaints manifest in grave patient complaints such as pseudomembranous colitis and toxic megacolon. Consequential mortality rates of ~ 5% have been reported [15].

1.1.2 Molecular Mechanism of disease

The clinical symptoms are caused by the secretion of Clostridia associated toxins. The pathogen produces three protein toxins: *C. difficile* toxin A (TcdA) and B (TcdB), and *C. difficile* transferase toxin (CDT) or binary toxin. Not all *C. difficile* strains produce the latter toxin. The single-chain toxins TcdA and TcdB are the main virulence factors. They bind to cell membrane receptors and are internalized [16]. Both toxins have glucosyltransferase activity and modify the actin skeleton of intestinal epithelial cells through glycosylation of members of the Rho family of small GTPases and thus causing symptoms.

CDIs are a considerable liability to healthcare systems and have overtaken cases of methicillin-resistant *Staphylococcus aureus* (MRSA) infections in specific clinical settings [17]. Laboratory diagnosis of CDI continues to be challenging. Commonly used techniques are enzyme-linked immunoassays (ELISA), detecting the Clostridia toxins A and B, do not have adequate sensitivity to be used alone for

detecting CDI. Conversely, nucleic acid detection tests, targeting chromosomal toxin genes show high sensitivity and specificity, provide rapid results but are not universally recommended for routine use in the recent guidelines. The lack of a gold standard upon diagnosis techniques of CDI exacerbates appropriate treatment. Reliable results may be achieved after 48 to 92 hours to complete, which delays appropriate therapy and critical infection control measures [18]. Nevertheless, epidemiology of *C. difficile* is evolving rapidly. Despite this continued threat, we have a poor understanding of how or why particular variants emerge.

1.1.3 *C. difficile* PCR RT 078

As infectious diseases such as CDI, represent a major problem in hospital settings, it is important to identify virulent strains as quickly as possible. Genotyping is a method to quickly evaluate the genetic composition of various bacterial strains and isolates.

In 1993, Gürtler firstly introduced a process known as PCR ribotyping [19]. Three years later, O'Neill et al., introduced modifications to the protocol and it became the global standard for *C. difficile* genotyping [20]. PCR ribotyping of *C. difficile* strains involves analysis of the intergenic spacer region (ISR) between the 16S and 23S rRNA genes. The space between the ISR varies in length and contains variable

alleles. PCR amplification of this region with a single pair of primer results in a band pattern formation, individual for each specific RT [21].

Worldwide, numerous *C. difficile* PCR RTs dominating temporally and geographically have been linked to patient complaint. In 2003, the previously rare RT 027 first emerged to be epidemic [21]. From its first outbreak in Canada it has spread across Northern America to Europe [22, 23]. In earlier years RT 027 has spread globally and accounted for the majority of isolates in United Kingdom and North American hospitals. Recently, there has been a decrease in prevalence of RT 027 and the patterns of Ribotypes in the UK have become more heterogenous. While in some areas of Britain, RT 027 has almost completely disappeared as novel strains emerged [10]. Since 2018, it has been reported, that the patterns of RT predominance in England have been relatively stable, with RTs 001, 017, and 078 have been commonly present in hospital settings [8, 24-26].

Despite the global efforts to keep *C. difficile* strain 027 under control, it remains difficult to eliminate CDI incidence completely. Infection control barriers and education of healthcare personal and patients resulted in a 78% reduction in CDI incidence and a decrease in severity. The notorious epidemic *C. difficile* strain 027 has decreased from 51% of clinical isolates associated with CDI in 2001 to 13% in 2005. A reduction of the predominant epidemic strain represents a facility for new epidemic strains with a different profile to emerge. Wilcox et al., already described

the emergence of *C. difficile* PCR RT 078 predominant in parts of Europe where type 027 has rapidly decreased from 55% of all detected strains in 2007 to 21% in 2010 [27]. *C. difficile* PCR RT 078 is the most common PCR RT among CDI positive isolates from swine (86%) in Korea and isolates from calves (94%) in the United States [28, 29]. Older studies from Europe and the United States revealed that in CDI related human patients, 078 was found at a prevalence ranging from 3% to 11%. More recently, a study revealed an increase of the epidemic 078 strains in the Netherlands from 3% to 13% during February 2005–2008 and thus can be considered hypervirulent [8, 30]. PCR RT 078 possesses binary toxins and has a deletion and stop codon in the *tcdC* gene, resulting in increased toxin production and thus increased disease severity [31]. Patients infected with 078 or 027 develop similar symptoms, while a younger population is more frequently associated with 078.

It remains questionable, if a zoonotic transmission of RT 078 from livestock towards immunocompromised patients is likely. However, the increasing evidence of the hyper virulent 078 strain in clinical settings in Northern America and Europe suggests that further analysis of this RT and its method of transmission and infection is necessary [32].

1.2 *C. difficile* sporulation

A characteristic feature of *Clostridium* and *Bacillus* species is the ability to form dormant spores during a process called sporulation (Figure 1.1. The spore morphotype protects the bacteria from environmental stress and ensures long term survival. In the dormant spore form, clostridia can survive circumstances which would otherwise have killed the vegetative organism, like the exposure to oxygen, heat, alcohol, antibiotics and certain disinfectants or a lack of nutrition in the environment [33]. As during CDI, *C. difficile* cells leave the body in faecal matter they are unable to survive in the aerobic environment and due to nutrient starvation. Thus, the affected vegetative form initiates the sporulation pathway to persist in hostile environments. This allows *C. difficile* to contaminate infected surfaces and also disseminate through patient-to-patient contact. Spores are metabolically dormant and intrinsically resistant to antibiotics, attacks from the host's immune system and once shed into the environment, they are also resistant to bleach-free disinfectants commonly used in hospital settings [34, 35]. Ingested spores migrate into the anaerobic environment of the large intestine and germinate under appropriate stimuli to form the toxin-producing vegetative cells which can proliferate in susceptible individuals [36].

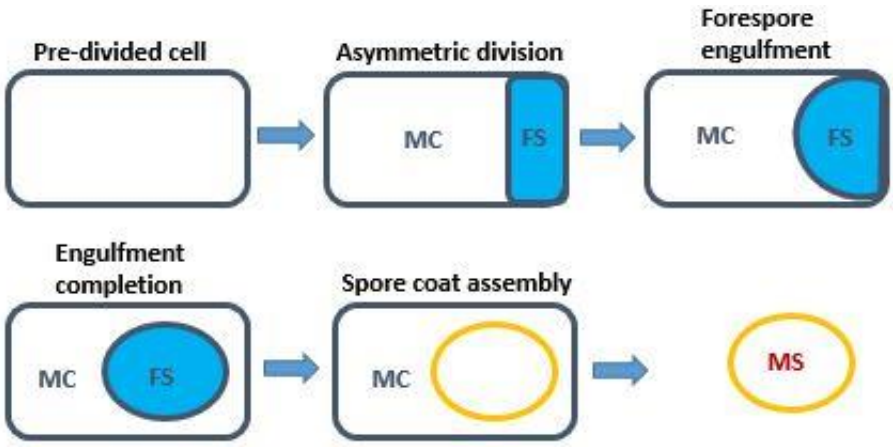


Figure 1.1: Main morphogenetic stages of the process of sporulation.
In chase of nutrient starvation, bacilli and clostridia initiate endospore formation. Initially an asymmetric division of the cytosol forms a small forespore and a mother compartment. The cell segregates its DNA and additionally, the mother cell engulfs the forespore. The forespore compartment remains metabolically dormant. Thus, the mother cell produces the spore cortex and the inner and outer coat. This leads to spore maturation and eventually the mother cell lyses and releases the mature spore. Abbreviations: MC, mother cell compartment; FS, forespore compartment; MS, mature spore. Electron Microscopy pictures have been used and adapted from Paredes-Sabja et al [14].

1.2.1 Endospore structure

The steps during spore formation are similar in bacilli and clostridia and are characterised by the initial asymmetric division of the cell to produce the larger mother cell and the smaller forespore, the latter being the eventual spore (Figure 1.1). The course of *C. difficile* sporulation can be divided into seven distinct steps. The first step, termed stage 0, describes the vegetative mother cell before the

beginning of the sporulation. During step one and two sporulation is initiated via the introduction of environmental stress. The asymmetrically positioned spore septum begins to form while the mother cells replicates 25 chromosome copy of its genome. Eventually, the cell divides into a larger mother cell compartment and a smaller pre-spore compartment via the formation of a septum in the cell. In stage three, the pre-spore compartment is engulfed by the mother cell compartment, forming the forespore with an inner and an outer membrane. Stage four describes the synthesis of a peptidoglycan layer, forming the primordial germ cell wall and the cortex between the two spore membranes. In step five, the spore coat, surrounding the outer membrane of the forespore, is produced. This is accompanied by transportation of pyridine-2, 6-dicarboxylic acid (DPA) into the developing forespore by products of the *spoVA* operon. Eventually the spore matures in step six. It undergoes dehydration and increases the density of the spore coat (Figure 1.2).

Programmed cell death of the mother cell acumulates in the release of the mature endospore into the surrounding environment [33, 37, 38]. The formation of this multi-layered protein structure mainly protects the spore from environmental interaction but is not a significant permeability barrier. In *B. subtilis*, the coat is the outermost spore structure, while in *C. difficile*, *Bacillus anthracis* and *Bacillus cereus*, the spore is surrounded by an additional layer, termed exosporium. Both structures are developed from the sporangium by the mother cell, utilizing a series

of core proteins, morphogenetic proteins, and several types of post-translational modifications [33].

The function of the exosporium remains unclear but recently has been suggested to participate as virulence factor in *C. difficile* strain 630 [39, 40]. Located underneath these structures is the spore cortex, expressing species specific peptidoglycan cross-linkages [38]. The central core region, is separated from the outer core by the germ cell wall, composed of peptidoglycan and the spore inner membrane. The germ cell wall also serves as the cell wall for vegetative bacteria after outgrowth of the spore. The spore core contains all the spore's DNA, RNA, proteins and ribosomes. The dormant spore core has an extremely low water content as in comparison to the protoplast. Ratios can vary between 35% of water content wild type in well studied *B. subtilis* species spores to 80% of wet wild type protoplasts. Although the spore inner membrane has a similar lipid composition to the mother cell, small molecules like water cannot diffuse, causing core dehydration and respective loss in weight [41].

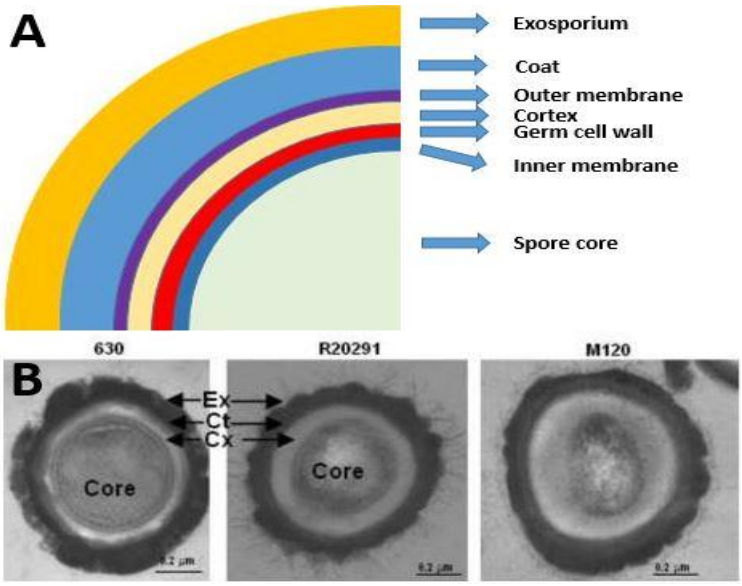


Figure 1.2: Structural layers in bacterial spores.

A) Structural layers of a bacterial endospore. The structural layers are represented and respectively labelled with arrows. Especially the exosporium is just expressed in some species and may contain sublayers, which are not represented. B) Micrographs of the spore from *C. difficile* strains 630, 027 and M120. Interestingly different ultrastructural phenotypes can be observed upon the different lineages, while the major compartments remain the same. Abbreviations: EX-exosporium, Ct- coat, Cx-cortex [14].

1.2.2 Sporulation

Master Regulator of Sporulation: Spo0A

Extensive past investigation of sporulation in *B. subtilis* has led to its adoption as the model organism for this process. The shared morphological features and other

similarities between sporulation in *B. subtilis* and *Clostridium*, has led to a number of predictive assumptions of the mechanisms in operation in *C. difficile*. However, more detailed analysis has shown that during the early stages of sporulation significant differences occur between the species [42].

Unlike in *B. subtilis*, the environmental stimuli causing *C. difficile* to induce sporulation are largely unknown. However, the downstream protein activation, has been studied in detail in *B. subtilis*. Once *B. subtilis* undergoes nutrient starvation it has been demonstrated, to activate the orphan master histidine kinase, Spo0A (Figure 1.3). Further this inhibits the *abrB* repressor, eventually allowing expression of genes essential for sporulation [43]. The key regulator Spo0A is activated following phosphorylation by Spo0B.

Further, *spo0A* transcription is indirectly modulated by a double repression system in which Spo0A represses *abrB*, causing increased *sigH* transcription. Higher levels of the stationary phase sigma factor SigH promotes expression of several genes, including *spo0A*. While both bacilli and clostridia activate Spo0A, the Spo0B pathway is restricted to *Bacilli*, while just a Spo0B homologue has been found in *Clostridium tetani,* whose function has yet to be determined [44]. The mechanism by which Spo0A is activated during the initiation of *C. difficile* sporulation is unclear, although its ablation completely abolishes spore formation in the laboratory *C. difficile* strain 630 [45]. For Spo0A activation it is expected that a

phosphoryl group is transferred directly from the orphan histidine kinases to Spo0A, as observed in *Clostridium acetobutylicum* [43]. *C. difficile* strain 630 genome encodes five orphan histidine kinases (CD1352, CD1492, CD1579, CD1949 and CD2492)[45]. Only CD1579 has been demonstrated to autophosphorylate and transfer a phosphate directly to Spo0A [45, 46]. The exact mechanism of how *C. difficile* activates Spo0A remains questionable (Figure 1.3). It is speculated that it is either activated directly by the histidine kinases or a yet as unidentified phosphorylation system is responsible [41].

The importance of Spo0A in clostridial sporulation has been demonstrated, as its inactivation results in loss of production of sporulation specific sigma factors and thus sporulation. Moreover, transcriptomic and proteomic analysis in *C. difficile* revealed, that Spo0A takes an active role in regulation of virulence factors, biofilm formation and toxin production and overall CDI progression [45, 47-49].

Sigma factors

In the well-studied *B. subtilis*, sporulation pathway, activation of RNA polymerase sigma factors remains crucial for sporulation [50]. It has been demonstrated that 314 genes are induced during the sporulation process in *C. difficile*. Of these, 224 genes are primarily controlled by 4 sigma factors σ^F, σ^E, σ^G, and σ^K. A total of 183 were σ^F-dependent, 169 were σ^E-dependent, 34 were σ^G-dependent and 31 were

σ^K-dependent [51]. This similarity to *B. subtilis* initially led to the acceptance, that the sporulation in both organisms is conserved and would be regulated by the same pattern [52]. However, genome-wide analyses demonstrated that the *C. difficile* sporulation process lacks communication between sporulation-specific sigma factors observed in *B. subtilis* (shown in Figure 1.3) [51]. Moreover, the morphological phenotypes of the individual sigma factor mutants in *C. difficile* is different to the *B. subtilis* mutants. Overall, this indicated a different mode of regulation of sporulation between these two species [53]. Furthermore, the regulatory pathway, mediating the activation of specific sigma factors in *C. acetobutylicum* and *C. perfringens*, is highly variable from the one in *C. difficile* [54, 55]. However, sigma factor regulation in most *Clostridium spp*., such as *C. perfringens* and *C. acetobutylicum,* have not been thoroughly investigated. Thus, it is likely that *C. difficile* sigma factor activation might be even more differentiated from other *Clostridium* and *Bacillus* species [56].

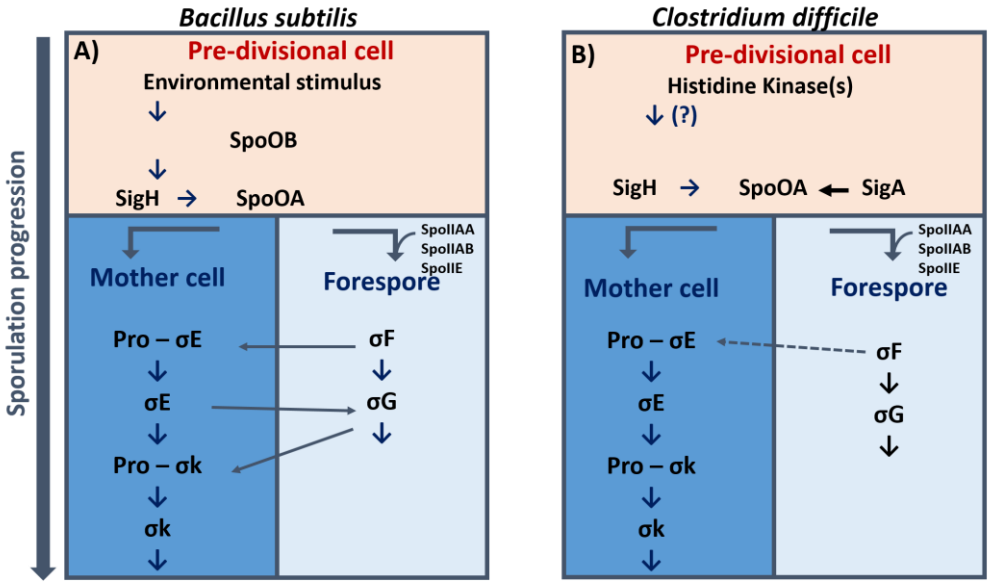


Figure 1.3: Basic order of *B. subtilis* and *C. difficile* sporulation.

A) *B. subtilis* sporulation is regulated through the cross- regulation of four sporulation-specific RNA polymerase sigma factors (s), of which σ^E and σ^K are specific to the mother cell and σ^F and σ^G are unique to the developing forespore. Under normal conditions, σF is supressed by anti-σ factor. Once septum formation is completed it undergoes activation and activated genes, necessary for cleavage of an inhibitory pro-peptide anti-σ^E. Active σ^E activates genes, necessary for induction of σ^G in the forespore. Further, σ^G directs gene expression and activates σ^K in the mother cell via. B) In *C. difficile*, σ^F is also consulted for activation of σ^G and partially for the activation of σ^E. Further downstream, σ^E is required for the production and activation of σ^K. Interestingly in *C. difficile* σ^K does not require proteolytic activation as it does in *B. subtilis*. Eventually, σ^E is not relevant for σ^G activation in the *C. difficile* forespore [14, 41].

1.3 Endospore germination

Spores are considered dormant and thus have little or no metabolic activity. The morphological change, that returns a spore to the vegetative state is called

germination. There are numerous agents, triggering spore germination in various organisms.

These so called germinants of *Bacillus* and *Clostridium* spores are mostly nucleosides, sugars, amino acids, and ions [14]. However *in vitro* evidence of purified germination receptors binding to specific germinants are currently lacking [36]. Nevertheless, clostridia and bacilli have multiple germination receptors that interact with various germinants. These eventually trigger the release of monovalent cations activating a downstream signalling cascade in the spore core. In *B. subtilis* this eventuates in core re-hydration and resumption of metabolism in the spore core. *C. difficile* germination is triggered by a combination of specific bile salts like cholate, its derivatives (like taurocholate) and L-glycine acting as a co-germinant [57]. The mechanisms of germination receptors are complex and some *C. difficile* strains also germinate in rich media without taurocholate. Additionally high specificity rates upon receptor homology in different *C. difficile* strains hamper the full understanding of receptor activation [14]. Despite the fact that *C. difficile* germination diverges considerably from the well-studied *B. subtilis* model there are some similarities between nutrient germination of spores of clostridia and bacilli, including multiple SpoVA proteins, essential at least for ion uptake during sporulation. While, many clostridial spores contain receptor components, similar to those discovered in spores of bacilli, *C. difficile* receptors lack any similarity to *Bacillus* [36].

1.3.1 *Bacillus subtilis* endospore germination

Various germinants trigger spore germination in *Bacillus*. Many of these agents like, calcium dipicolinic acid (CaDPA), and cationic surfactants like dodecylamine are of industrial importance and just considered to play a role in the laboratory. Naturally it is considered that the presence of nutrients triggers spore germination by binding to multiple germinant receptors in the inner membrane of the spore. Germinant receptor activation can be substrate or concentration specific [39]. Within the well-studied spore model *B. subtilis,* specifically L amino-acids (L-alanine, L-valine, and L-asparagine), but not the D-amino acids, trigger germination [58]. Germinant receptor activation results in an initial lag period. In "superdormant" spores this lag phase might last up to 24 hours and alters the inter membrane permeability and structure. Consequently, monovalent cations, (H^+, K^+, and Na^+) are released. This causes a CaDPA release via channels composed of the multiple spore-specific SpoVA proteins [37]. CaDPA release triggers degradation of the spore cortex, allowing the germ cell wall and the spore core to expand and rehydrate. Increased core water content allows bacterial metabolism in the core to reinitiate. Macromolecular synthesis, produces proteins converting the germinated spore which undergoes outgrowth to produce a growing vegetative cell [45].

1.3.2 *C. difficile* endospore germination

Also, for *C. difficile*, spore germination is an important step to induce CDI. Unlike any other spore forming organisms, germination in *C. difficile* is regulated by the presence of specific bile salts and for some strains the co-germinant L-glycine in the spore's environment [57]. Recently, the interplay between different bile salts and their role for disease progression have been observed in the mouse model, which has been described as the most accurate model in mimicking human disease [59]. It has been demonstrated that the environment of the mouse small intestine appears to support germination, while outgrowth in the mouse large intestine could only be observed after excessive antibiotic treatment with cefoperazone, clindamycin and vancomycin [60]. Additionally, a decline of secondary bile salts like deoxycholate, ursodeoxycholate and hyodeoxycholate has been observed in the affected regions. How exactly the different muricholic acids influence spore germination remains unclear. Although their different balance appears to directly affect *C. difficile* spore germination hence chenodeoxycholate is known to be a competitive inhibitor of the germinant taurocholate. *In vitro* however, primary bile acids such as cholic- and taurochenodeoxycholic acid are seen as germinants and cause spore outgrowth, regardless of antibiotic treatment [61]. Thus, the concentration of bile acids is considered as an important factor during CDI. Especially, as spore outgrowth in healthy individuals is suppressed by secondary

bile acids, such as lithocholic acid, in the colon. However, increasingly virulent *C. difficile* strains have been more tolerant towards secondary bile acids [62].

C. difficile strains react to different germinants and are not known to contain germinant receptor homologues to the canonical GerA receptor family in well-studied *B. subtilis* model. Francis *et al*., discovered that a different mechanism is responsible for *C. difficile* outgrowth. The bile acid specific pseudoprotease, C-spC, has been demonstrated as a functional germinant receptor in the hamster model [63].

The exact signalling pathway, starting from the activation of the germination receptor C-spC via, for example, taurocholate remains unclear. C-spC is expected to activate the downstream serine protease C-spB. C-spB is encoded in the *C-spBA* gene locus and transcribed into the C-spAB protease. Hence the A subunit of this molecule is catalytically inactive, interdomain cleavage of C-spBA via C-spC releases activated C-spB, entering the spore. In the spore, C-spB may processe pro-S-leC into the mature cortex hydrolase S-leC [61].

Lack of current knowledge highlights the need for further *in vivo* studies that investigate the role of other factors, for example, Ca-DPA release from the spore core, essential in *B. subtilis* or *C. perfringens* to eventually result in full spore outgrowth. Interestingly, Gutelius *et al*., have provided evidence that S-leC cortex

hydration *in vitro* does not require proteolytic cleavage [64]. These results are supported by different studies suggesting that neither the inactivation of S-leC or C-spC prevents cortex hydrolysis and DPA release. These results underline, the importance of DPA, inducing osmotic swelling, rehydration of the spore core and thus eventually spore germination [64].

1.4 A roadmap in *C. difficile* genetics

In *C. difficile,* the sigma factors have been identified as the main sporulation regulators. Each of these sigma factors contains a representative homologue in *B. subtilis* pathways. The 228 key genes involved in sigma factor directed sporulation of *C. difficile* correspond to about half of the genes activated in *B. subtilis*. Given that the genus *Clostridium* comprises not only disease-causing organisms of humans and animals but also important targets for food, fuel or biotechnological industry, their genetic and phenotypical characterization becomes more important. However, clostridial species are complicated targets for genetic modification. Thus, several methods and approaches for improving the specific clostridial genetics have been suggested by Minton et al. and presented as a roadmap [65]. This roadmap has been implemented for numerous *Clostridium* species and partially implemented in *Clostridium pasteurianum* [66].

The basis of this roadmap builds upon the availability of a fully sequenced genome of the respective target organism and the identification and negation of the species-specific restriction/modification (RM) systems in order to optimize DNA transfer into the organism [67]. In the case of *C. difficile*, DNA is introduced by its conjugative transfer from *E. coli* donor strains that bear an appropriate methylase for host specific DNA modification [68, 69]. Once effective DNA transfer is obtained, genome editing tools such as CRISPR or ACE are exploited to bring about defined changes in the target organism.

1.4.1 Transformation

Implementing genetic tools in clostridia for industrial or therapeutical reasons makes gene transfer into the organism necessary. C*. acetobutylicum* represents an important research target for gene transfer. In 1988 the first successful gene transfer into *C. acetobutylicum* via electroporation was reported [70]. This method offers rapid and easy insertion of DNA fragments and was then implemented in several other *Clostridium spp*. including *C. pasteurianum* [71] and *C. perfringens* [72]. Ackermann described potential electroporation of DNA fragments from pathogenic isolates into the non-toxigenic *C. difficile* strain P-881 [73]. Purdy *et al*., however, concluded that the design of Ackermann's protocol was flawed and could not be replicated in the laboratory [74].

More recently, Bhattacharjee and Sorg described another electroporation protocol *for C. difficile*, achieving 20 and 200 colonies per microgram of DNA. They focused their studies on C. *difficile* strains R20291, CD630 and JSC10. The method, however, is difficult to conduct and needs long recovery times to obtain transformants. The authors further suggest, that the success and performance of the protocol relies heavily on the environmental conditions of each lab [75]. *C. difficile* turned out to be complicated to transform by electroporation or chemical integration. Based on the conjugation method, developed by Liyanage *et al*., Purdy described the successful transfer of DNA via conjugation into *C. difficile* strains CD3 and CD6 from *E. coli* donor strains [74, 76]. In 2016, Kirk et al., proved, that heat treatment of recipient *C. difficile* strain R20291 of up to 52 °C, prior to conjugation causes an increase of conjugation efficiency [77]. Until now conjugation remains the gold standard DNA transfer method in this organism.

1.5 Forward genetics in *Clostridium*

Forward genetic approaches are based on the isolation of mutants of a defined phenotype and there after determining the genotype responsible. Such approaches help to identify the underlying genetic bases behind specific phenotypes without drawing previous assumptions of the genes involved. This is

normally achieved by the analysis of naturally occurring mutants or production of chemical, radiation or transposon insertion induced random mutations in the bacterial genome. Furthermore, random mutagenesis may be followed by high throughput gene mapping methods and correlation of the phenotype with the affected genes. Followed up by isolation and further analysis of individual mutants, a mutant library allows the screening and analysis of numerous mutants simultaneously and thus optimizes the speed of the analysis [65, 78].

1.5.1 Transposon mutagenesis

Transposable elements, or short transposons, are a DNA segment, which can alter its position within a genome or from one genome to another. Several of these genetic elements are not dependent on host factors (Figure 1.4). A distinction is drawn between class 1 transposons, or retrotransposons, and class 2 transposons. Retrotransposons form an RNA intermediate and are inserted via reverse transcription. Class two transposons are DNA transposons which encode the corresponding transposase, implemented into inverted terminal repeats (ITRs). Transposition of class 2 transposons happens mainly during a non-replicative cut-and-paste mechanism [79]. Transposable elements play an important role in the fitness of both eukaryotic and prokaryotic organisms. They protect against viral attacks or provide environmental fitness against stressors, e.g., antibiotics via

random insertion mutation or dissemination of regulatory factors [80]. However, transposons can participate in disease formation and development of a deleterious phenotype [81].

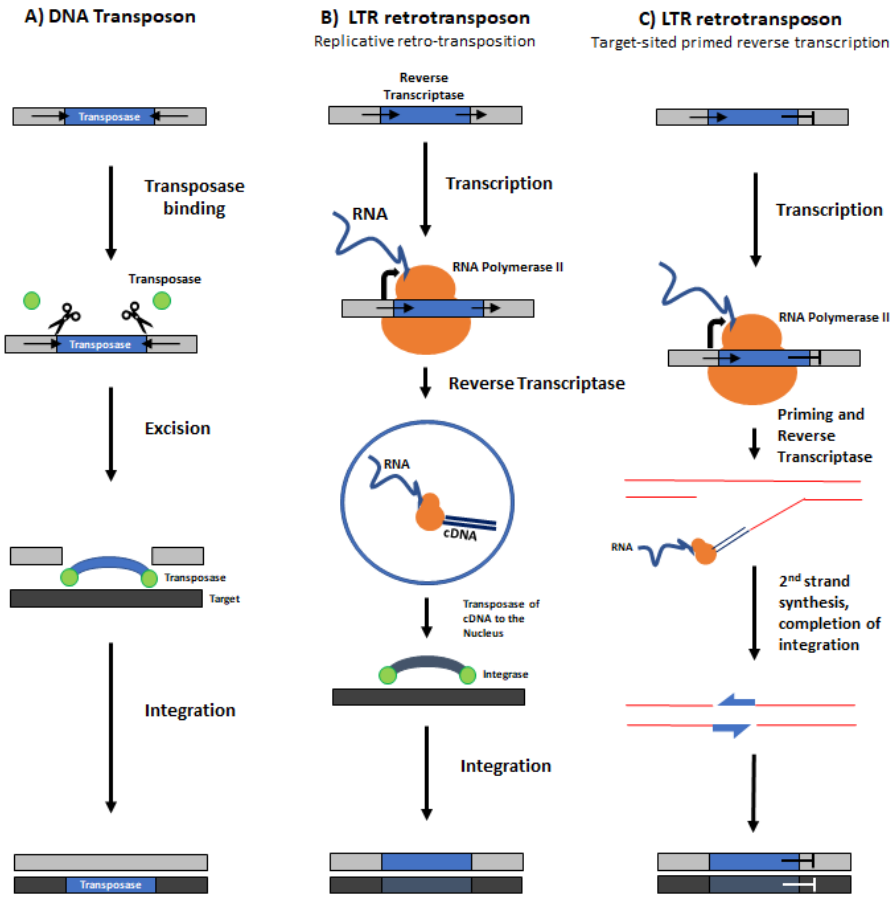


Figure 1.4: Mechanisms of DNA transposons.

A) Many transposons are mobilized by a cut-paste mechanism and are encoding a transposase, flanked by internal repeats (ITR), indicated by the black arrows. The transposase binds near or at the ITRs and transfers the transposon from its initial location on the genome or plasmid into the target. B) and C) Retrotransposons are mobilized via replication, depending on a reverse transcriptase. B) Retrotransposases require very long terminal repeats (LTR) for successful replication. On the 5' end of the transposon is a special promotor, recognized by the hosts, producing transposon encoding mRNA. Subsequently, reverse

transcriptase copied the Transposon mRNA back into full length dsDNA (cDNA). In another step, integrases build in the transposon cDNA into the new target side. C) Non-LTR retrotransposons Lack LTRs and encode either one or two OFRs. Similarly, to B), the transcription of non-LTR retrotransposons generates a full mRNA. Retrotranscription, however, is primed on the target side directly. An endonuclease generates a single-strand nick into the genomic DNA priming reverse transcription of the RNA by non-autonomous and autonomous retrotransposase. The integration of non-LTR retrotransposons can lead to target side depletions (TSDs) or small deletions at the target site in genomic DNA. The Graphic was adapted from Levin et al., [82].

Transposable elements make up a large proportion of the eukaryotic genome and can constitute up to 85% in some plants [83]. In *C. difficile* around 11% of the genome is comprised of mobile genetic elements. This confers a high degree of plasticity to genome. The complete role of these mobile elements upon the organism's biology, evolution and pathogenicity, however, remains to be fully understood [84].

1.5.2 Transposon mutagenesis in *C. difficile*

Transposable elements are a promising tool for performing forward genetics and have been used for mutant production and mutant identification in several clostridial species. Two basic types of transposon system have been described in clostridia: conjugative and non-conjugative transposons.

Integrative Conjugative Elements are circular structures, introduced from a donor cell into a recipient using a conjugation-like mechanism. These DNA fragments do not contain their own replicon and thus integrate into the hosts origin of replication to endure. Non-conjugative elements, however, lack the genetic machinery for conjugation and use the mechanism present in the cell [84]. Most studies trying to produce random insertion mutant libraries rely on the conjugative Tn*5* or the non-conjugative mariner transposon. Tn*5* is one of the first transposons discovered and derived from bacterial origin and has been reported to function in most bacterial species it has been tested in [85]. Tn*5* has been shown not to have a strong preference for a particular insertion site or the GC content [86]. Interestingly, Transposon Directed Insertion Sequencing (TraDIS) performed in *Salmonella enterica* has revealed a small bias upon AT richer sequence regions. Considering the high insertion frequency of Tn*5*, however, this bias has not been seen as a major obstacle [78, 87].

As they are directly transformed into the host, the success of these *in vitro* mutagenesis systems varies based on the transformability of the target organism. As DNA transfer into *C. difficile* is poorly characterized and challenging, these systems are less appropriate for implementation of transposon mutagenesis. Two transposons Tn*916* and Tn*5397* have been tested in *C. difficile*. Both are biased in their integrative motif and have the preference to lead to multiple insertions. As

such, these systems have been neglected as suitable for *C. difficile* mutant library production [88].

On the other hand, the mariner element *HimarI* is derived from the horn fly, *Haematobia irritans*, and reconstructed as a consensus sequence with a weak preferences for insertion site [89, 90]. The *Himar1* transposase is the only protein required for its integration and thus the transposon can be integrated into the host genome without activation of any host factor. Eventually, the transposon can be integrated into the host genome via homologous recombination. This makes it a method of choice for bacteria, that are poorly transformed with foreign DNA [89, 91].

Based on these findings, the mariner transposon system became of scientific interest for application in *C. difficile,* due to the high AT content of its genome. The transposon's bias against AT rich insertion sites but low GC content makes it an optimal choice for clostridia [65]. Two studies have been performed on *C. difficile* utilizing the mariner transposon and indeed, successful library production in several *C. difficle* strains have been reported [86, 91].

The mariner-based transposon system designed by Cartman and Minton was based on the pMTL-SC1 plasmid, adjusted with the Himar1 C9 transposase gene [91]. The *C. difficile* toxin B promoter, P_{tcdB}, was located in front of the *HimarI* transposase gene. A copy of the *catP* gene, encoding chloramphenicol

acetyltransferase, provides the transformed organism with resistance to Thiamphenicol. Transcription from the P_{tcdB} promoter is mediated by the alternative sigma factor TcdR which is absent in Gram-negative background and thus considered inactive in *E. coli*. This lack of a Gram-negative origin of replication is considered to prevent transposition events from happening in the Gram-negative host before conjugation into *C. difficile* [92].

A major drawback of this study was found in the segregationally unstable pseudo-suicide vector system. It is emphasized that it takes at least two passages to eventually eradicate the transposon plasmid. This approach improves the time and work efforts required to produce libraries containing thousands of mutants [91].

In 1990, Geissendorfer and colleagues introduced a Tetracycline (Tc)-inducible expression system in *Bacillus*. This Tc-inducible system consists of divergent promoters, each with an overlapping *tet* operator sequence (TetO). When no Tc is present, the P_{tetR}, promotor drives expression of *tetR,* binding to the *tet* operators and eventually represses both promoters. The adjacent promotor P_{tet}, drives expression of any subsequent gene. Tc or its less cytotoxic analogue anhydrotetracycline (aTc) induces a conformational change in TetR that prevents binding to the operator and relieves repression [93]. Fagan and colleagues introduced two studies on a Tc-inducible expression system cloned into the

conditional pRF177 plasmid [86, 94]. Downstream of the P_{tet} promotor, they cloned an optimised *Himar1* transposase gene with customised ITR regions, flanking an *ermB* resistance gene. Plasmid instability was achieved by positioning the P_{tetR} promoter toward the *C. difficile* pCD6 origin of replication keeping the plasmid intact during normal conditions but inducing aTc dependent instability. Using this system, the group was able to generate a large pool of random transposon insertion mutants. In the *C. difficile* strain 630 *Δerm*, around 85,000 Em resistant colonies were pooled, from which 44.102 unique mutants were identified via sequencing. In the clinically relevant R20291 strain approximately 70,000 unique mutants were achieved. This study demonstrated the functionality of the Tc inducible system for *C. difficile* transposon mutant library production and represents a promising target for further research [86, 94].

Monitoring large transposon libraries is crucial for identifying random insertions and thus analysing the fitness of a given phenotype under various environmental conditions. Already in 1999, Hutchison et al, developed a sequencing-based method to screen for transposon insertion sites in around 1000 transposon mutant in reduced *Mycoplasma* genomes [95]. This technique, however, was time intensive and the low accuracy prevented its widespread use.

1.5.3 Mutant sequencing and TraDIS

More recently, large transposon libraries have been analysed by high throughput sequencing technologies like Transposon Directed Insertion Sequencing (TraDIS), Insertion sequencing (INSeq), high-throughput insertion track by deep sequencing HITS and Transposon sequencing (Tn-seq) which with minor variations, generally follow a similar workflow [78]. These techniques are characterised by a high level of reproducibility, up to 90%, and are therefore considered as reliable and fast methods for broad scale genome screening [96, 97]. Langridge et al demonstrated a technique which could be sufficiently used to screen mutant related fitness within a population of single transposon mutants [87]. The workflow (of TraDIS) initiated with the fragmentation of the selected mutant genomic DNA and then the subsequent PCR amplification and enrichment of these transposon containing fragments. The technique is highly adaptable to various transposons by simply redesigning the sequencing primers [98]. To verify transposon insertion, primers for sequencing are designed to anneal within the transposon and thus form a 'transposon-tag, before the read. Sequencing is performed on the chromosomal DNA. A short 10–12 cycle transposon read by Illumina sequencing routinely achieves results of > 90% of sequencing reads. TraDIS that has been implemented in Tn*5*-, Tn*917*- and *Himar1*- based mutant libraries, is anticipated to work in any transposon model, given the appropriate set of sequencing primers is available. The technique has been performed successfully

in many organisms and transposon models, including Tn*5*-based libraries in *Salmonella* [87] and *Escherichia* [99, 100]. Mariner-based libraries in *Mycobacteria* [101] and in *C. difficile* [86] have also been successfully analysed using this high throughput sequencing method.

1.6 Overall Aims of this Work

The broad aim of this project is to gain a deeper understanding of the molecular mechanisms involved in sporulation and germination in *C. difficile* PCR RT078 strains, based on our knowledge on transposon derived mutagenesis. We first, aim to improve DNA transfer into *C. difficle* RT078 and additionally evaluate a variety of transposon delivery systems. Finally, the most seminal construct is depicted for execution of an enhanced TraDIS protocol from transformation to library prep. Constructive on that, this study reveals the currently most successful TraDIS pipeline to elucidate the role of genes involved in in *C. difficile* sporulation and seek to establish reverse genome editing in *C. difficile*.

Chapter 2

Materials and Methods

2.1 Bacterial strains

Table 2.1.1: Bacterial strains used in this study.

Strain	Source/Reference	Description/
***E. coli* TOP 10**	Invitrogen	Plasmid cloning and storage
***E. coli* XL-1 Blue**	Promega UK Ltd	Plasmid cloning and storage
***E. coli* DH5 alpha**	New England Biolabs (NEB)	Plasmid cloning and storage
***E. coli* CA434**	M. Young, UCW, Aberystwyth, UK [74]	Conjugative transfer strain
***E. coli* sExpress**	Craig Woods [102]	Conjugative transfer strain
***C. difficile* 630**	NCTC, Public Health England	PCR-RT 012 (Zurich, Switzerland)
C. difficile* 630Δ*erm	Hussain *et al*., (2005) [103]	Em-sensitive derivative of *C. difficile 630*
C. difficile* 630Δ*erm	Patrick Ingle	Em-sensitive derivative of *C. difficile* 630
***C. difficile* R20291**	Patrick Ingle	Clinical PCR RT 027 strain
***C. difficile* EK24**	Heeg et al., [104]	*C. difficile* PCR RT 078 CD2315
***C. difficile* EK26**	Heeg et al., [104]	*C. difficile* PCR RT 078
***C. difficile* EK28**	Heeg et al., [104]	*C. difficile* PCR RT 078 CD7009825
***C. difficile* 18-01**	Nottinghamshire	Clinical PCR RT 078 isolate
***C. difficile* 44-01**	Nottinghamshire	Clinical PCR RT 078 isolate
***C. difficile* 90-01**	Nottinghamshire	Clinical PCR RT 078 isolate
***C. difficile* 93-01**	Nottinghamshire	Clinical PCR RT 078 isolate
***C. difficile* 95-01**	Nottinghamshire	Clinical PCR RT 078 isolate
***C. difficile* 97-01**	Nottinghamshire	Clinical PCR RT 078 isolate
***C. difficile* M120**	[13]	*C. difficile* PCR RT 078

Table 2.1.2: Plasmids used in this study

Plasmid Name	Description	Reference
pMTL-81151	*Clostridium* shuttle vector with no Gram-positive replicon	SBRC
pMTL-82151	pMTL-81151 containing the pBP1 replicon	SBRC
pMTL-83151	pMTL-81151 containing the pCB102 replicon	SBRC
pMTL-84151	pMTL-81151 containing the pCD6 replicon	SBRC
pMTL-85151	pMTL-81151 containing the pIM13 replicon	SBRC
pMTL-86151	pMTL-81151 containing the pIP404 replicon	SBRC
pMTL-CW17	Himar1C9 transposase under the control of the lactose-inducible bgaR-PbgaL promoter, instead of the orthogonally-controlled P*tcdB* like in CW21, 22, 26 and 27	Craig Woods
pMTL-CW21	Transposon mutagenesis vector harbouring the Himar1 transposase and pcB102 Gram-positive replicon	Craig Woods
pMTL-CW22	CW21 with pCD6 Gram-positive replicon	Craig Woods
pDIG-1	Part 1 of a 2-part transposon mutagenesis vector system, harbouring the Himar1 transposase and pCB102 Gram-positive replicon	James Millard
pDIG-2	Part 2 of a 2-part transposon mutagenesis vector system, harbouring the mariner transposon and no Gram-positive replicon	James Millard
pMTL-GL15	Suicide Transposon mutagenesis vector, mobilizing a catP transposon, driving Himar1C9 transposase via P*tcdB*	[91]
pRF215	Transposon mutagenesis vector harbouring the Himar1 transposase and pcD6 Gram-positive replicon	Robert Fagan[86]
pMTL-MTV10	Transposon mutagenesis vector, harbouring the SBRC backbone and the Himar1 transposase, Em Transposon and Tet-promotor derived from pRF215. Additional I-*SceI, AscI* and *Fse-I* side have been added.	This study

2.2 Growth conditions

Aerobic growth conditions

E. coli strains were cultured aerobically in Luria-Bertani (LB) media at 37°C, in liquid broth with horizontal shaking at 200 revolutions per minute, or on solid agar plates. LB medium was infused with appropriate supplements listed in Table 2.5.2.

Anaerobic growth conditions

C. difficile strains were cultured anaerobically at 37°C within an MG1000 Mark II anaerobic workstation (Don Whitley Scientific Ltd, UK), with an internal atmosphere of Nitrogen (80%), Carbon dioxide (10%) and Hydrogen (10%). Prior to use, all culture media were pre-reduced within the anaerobic workstation, for a minimum of four hours for agar plates or 24 hours for liquid media. *C. difficile* strains were routinely cultured on Brain Heart Infusion Supplemented (BHIS) media with appropriate supplements listed in Table 2.5.2.

Strain storage

E. coli strains were grown on LB agar, before the resulting growth was harvested using a 10 µL plastic loop and resuspended in Microbank™ Long Term Storage tubes (Pro-Lab Diagnostics), for storage at -80°C. *C. difficile* strains were grown on

BHIS agar, before the resulting growth was harvested using a 10 µL plastic loop and resuspended in screw cap tubes containing BHIS broth with 10 % (v/v) glycerol, for storage at -80°C.

2.3 Chemicals and Suppliers

Plasmid mini-prep kits, genomic DNA extraction kits and Gel extraction kits were sourced from New England BioLabs. For polymerases reactions DreamTaq (Sigma-Aldrich), Phusion (New England BioLabs) were used. Oligonucleotides were sourced from Sigma-Aldrich. All other enzymes for molecular biology were sourced from New England BioLabs as well as 2-log and 1kb+ DNA ladders. All chemicals were ordered from Sigma-Aldrich Company Ltd.

2.4 Bioinformatics tools

DNA visualization

DNA was visualized using the web-based tool Benchling (https://www.benchling.com/). Plasmid maps in this study were constructed using SnapGene for preferred visualization.

Sequence data analysis

Sequence data and plasmid maps were routinely viewed using the Benchling online resource accessible at www.benchling.com. The phylogenetic analysis and genome alignments were performed using Qiagen CLC workbench 20. RM-system prediction was performed on the internet platform Rebase using the R-M finder tool (http://rebase.neb.com/rebase/rebase.htmL)

BLAST

For searches of DNA sequence and transposon integration side the databases using the Basic Local Alignment Search Tool (BLAST) were performed using the BLAST algorithm accessible at https://blast.ncbi.nlm.nih.gov/Blast.cgi.

MiSeq data analysis

Reads were filtered for quality and trimmed using TrimGalore (https://github.com/FelixKrueger/TrimGalore). The MiSeq custom recipe already filters for reads that contain the specified transposon tag. Bioinformatic analysis was then undertaken by Dr. Craig Woods, using methods based on those outlined in Barquist et al., 2016 [98].

2.5 General microbiological techniques

2.5.1 Bacterial Growth medium

Table 2.5.1: Growth media used in this study.

Medium	Components	Quantity g/l-1	Target Organisms
Luria-Bertani (LB)	Sodium chloride	10.0	*E. coli* strains
	Tryptone	10.0	
	Yeast extract	5.0	
Brain Heart Infusion Supplemented (BHIS)	Brain Heart Infusion	37.0	*C. difficile* strains
	Yeast extract	5.0	
	L-cysteine	1.0	

For solid media 1.5g/100mL No. 1 Bacteriological Agar was added. Media sterilization proceeded at 121°C for 20 min.

2.5.2 Supplements

Media supplements were prepared as stock solutions according to manufacturer's instructions. Solutions were sterilised by filtration through a 0.2-µm membrane (Minisart®). Stock solutions were stored at the recommended temperatures for no longer than 5 weeks.

2.5.3 Strain Storage and Revival

Bacterial stocks were stored at -80°C. *E. coli* strains were kept in the Microbank™ Long Term Bacterial and Fungal Storage System by Pro-lab Diagnostics. *Clostridia* stocks were preserved by the addition of 10% glycerol to 1mL of *Clostridia* culture.

Table 2.5.2: Supplements for bacterial growth media.

Supplement	Stock (mg/mL^{-1})	Solvent	Working Stock in *E. coli* ($\mu g/mL^{-1}$)	Working Stock in Clostridia ($\mu g/mL^{-1}$)
Chloramphenicol	25	EtOH (100%)	Broth: 12.5 Agar plates: 25	-
Tm	15	EtOH (50%)	-	15
Em	50	EtOH (100%)	500	10
D-cycloserine	50	dH2O	-	250
Cefoxitin	50	dH2O	-	8
Anhydro-tetracycline	50	EtOH (70%)	10	10
IPTG	23.81 =100 (mM)	dH2O	-	476 =2mM
Lactose			-	10
Sodium taurocholate	100	dH20	-	1000

2.5.5 Preparation of electrocompetent *E. coli.*

E. coli sExpress (NEB) were inoculated on an LB plate. Growing colonies were picked from the plate and inoculated in 5 mL LB medium at 37°C with horizontal shacking at 200 rpm. After 24 hours 1 mL of the subsequent culture was incubated in 200 mL of LB broth under the above-mentioned growth conditions, till the culture reached an OD of 0.5 – 0.7 at 600 nm. For optimized handling, the 200 mL culture was divided into four 50 mL falcon tubes. Each falcon was centrifuged at 4000 x *g* for 10 min at 4°C to obtain a pellet. The pellet was resuspended in ice-cold sterile water and centrifuged as above. The washing step was subsequently repeated. Eventually the resulting pellets were washed in 20 mL of sterile ice-cold water and pooled. A final centrifugation step was performed as the previous once and the pellet was resuspended in ice-cold MOPS (1 mL, 1 mM) with 10% v/v glycerol before being pooled in 50 µL aliquots and stored at -80°C.

2.5.6 Preparation of chemical competent *E. coli.*

Relevant *E. coli* strains were inoculated on an LB plate. Growing colonies were picked from the plate and inoculated in 5 mL LB medium at 37°C with horizontal shacking at 200 rpm. After 24 hours 1 mL of the subsequent culture was incubated in 200 mL of LB broth under the above-mentioned growth conditions, till the culture reached an OD of 0.5 – 0.7 at 600 nm. For optimized handling, the 200 mL culture was divided into four 50 mL falcon tubes. Each falcon was centrifuged at

4000 x *g* for 10 min at 4°C to obtain a pellet. The supernatant was discarded, and the pellet was resuspended in 20 mL 100 mM of $MgCl_2$. The cell mixture was placed on ice for 30 min. Centrifugation was repeated and the pellet was resuspended gently in 12.5 mL 100 mM $CaCl_2$ and the mix was inoculated for 30 min on ice. After a final spin, the pellet was resuspended equally in 100 mM of $CaCl_2$ and 20 % glycerol. Aliquot are stored at -80°C.

2.5.7 Transforming electrocompetent *E. coli.*

A 50 µL aliquot of the electro-competent *E. coli* was thawed on ice and subsequently 2 µL of plasmid DNA was added and mixed by resuspending. The mixture was transferred into a chilled a 0.2 cm gap electroporation cuvette (Bio-Rad) and pulsed with an electroporator (Bio-Rad MicroPulser) using pre-set conditions (2.5 kV, 200 Ω, 25 µF capacitance). After a pulse, the cells were mixed with 400 µL SOC broth (Invitrogen) and transferred to a 1.5 mL microcentrifuge and incubated at 37°C and 200 rpm of horizontal shaking for one hour. A 100 µL aliquot of the cell mixture was spread plated in a dilution series of up to 10^{-3} onto LB agar medium containing the appropriate antibiotic and incubated at 37°C overnight. Growing colonies were picked from the plate and inoculated in 10 mL of LB medium, containing the appropriate antibiotic at 37°C with horizontal

shacking at 200 rpm overnight. The growing cultures were harvested in 50 µL aliquots and stored at -80°C.

2.5.8 Transforming chemically competent *E. coli.*

A 50 µL aliquot of the chemical-competent *E. coli* was thawed on ice and subsequently 2 µL of plasmid DNA was added and mixed by resuspending. The mixture thawed for 30 min on ice and then immediately transferred for 45 seconds on a preheated 42°C heating block. After the heat shock, the mixture was transferred on ice again, for 2 min. The cells were mixed with 400 µL SOC broth (Invitrogen) and incubated at 37°C and 200 rpm of horizontal shaking for one hour. A 100 µL aliquot of the cell mixture was spread plated in a dilution series of up to 10^{-3} onto LB agar medium containing the appropriate antibiotic and incubated at 37°C overnight. Growing colonies were picked from the plate and inoculated in 10 mL of LB medium, containing the appropriate antibiotic at 37°C with horizontal shacking at 200 rpm overnight. The growing cultures were harvested in 50 µL aliquots and stored at -80°C.

2.5.9 *C. difficile* transformation via conjugation with *E. coli.*

E. coli cultures, harbouring the plasmid construct of interest, were grown in 5 mL of LB broth, supplemented with the appropriate antibiotics at 30°C with horizontal

shaking at 200 rpm overnight. A 1 mL aliquot of BHIS broth was incubated with the recipient *C. difficile Δerm* strain and incubated anaerobically at 37°C overnight. The following day, the *E. coli* cultures reached the stationary phase and 1 mL was harvested by centrifugation at 1800 x *g* for 1 minute at 25°C. The pellet was washed twice in sterile PBS and transferred into the anaerobic cabinet and resuspended in 200 µL of the *C. difficile* overnight culture. The culture was divided and spread on two non-supplemented BHIS agar plates and incubated anaerobically for a minimum of 8 hours. Subsequent growth was harvested by spreading 200 mL of sterile PBS on the plate and scraped of the plate with sterile cell spreaders. The PBS-bacteria mixture was spread plated onto BHIS agar supplemented with D-cycloserine, Cefoxitin and the plasmid-appropriate antibiotic. Plates were incubated anaerobically at 37°C for 24-72 hours, until distinct transconjugant colonies had appeared. Growth was harvested and incubated in 15 mL of supplemented BHIS broth.

2.5.10 Measurements of bacterial growth.

Growth of *C. difficile* strains was measured via monitoring the changes in optical density at 600 nm (OD600) over a 24-hour period. The optical density of overnight *C. difficile* cultures was measured, and the volume required to produce sub-cultures (50 mL) with OD600 = 0.05 calculated.

2.6. General molecular techniques

2.6.1 Plasmid construction.

All plasmids were cloned using NEB HiFi assembly, the necessary fragments were generated via PCR using the appropriate templates and primers. Transformants were selected for on LB agar supplemented with chloramphenicol and Em and inoculated to overnight cultures with the same supplementation. Plasmid DNA was prepared from overnight cultures and verified with diagnostic digests and Sanger sequencing.

2.6.2 Plasmid DNA extraction from *E. coli.*

E. coli cells harbouring the plasmid of interest were grown over night in 15 mL of LB culture with the corresponding antibiotic at 37°C under horizontal rotation at 200 rpm. After 24 hours the cultures were centrifuged for 2 min at 16000 rpm and room temperature. For harvesting the DNA from the pellet, the GenElute HP Plasmid Miniprep Kit (Sigma) was utilized according to manufacturer's protocol. The plasmid DNA was eluted 50 µL of sterile dH_2O, incubated at 40°C for enhanced elution and stored at -20°C. Quantification of DNA preparations. Extracted plasmid

DNA and genomic DNA samples concentrations were measured by a nanodrop ND-1000 spectrophotometer (Thermo Scientific) according to manufacturer's recommendations as double-stranded DNA has an OD260 value of 1 with a 1 cm path length.

2.6.3 Extraction of clostridial genomic DNA using GenElute kits.

Ten mL of *C. difficile* culture, grown under anaerobic conditions at 37°C overnight, were taken out of the anaerobic workstation and pelleted at 16000 x *g* for 3 min at 25°C. The supernatant was discarded, and the cell pellet was treated according to the protocol of GenElute Bacterial Genomic DNA Kit (Sigma). The cells were resuspended in 200 µL Lysis buffer, containing 10mg.mL^{-1} lysozyme and incubated at 37°C for 30 min. A 20 µL aliquot of RNase a Solution (Sigma) was added and incubated at room temperature for 2 min. 20 µL of Proteinase K in 200 µL Lysis solution (Sigma) were added to the mix and incubated at 55°C for 10 min with intermediate vortexing. To precipitate the genomic DNA, 200 µL of 99, 9% ethanol were added to the mixture and the whole conglomerate was introduced into a silica membrane binding column (Sigma) and centrifuged at 6500 x *g* for 1 minute. The membrane was washed twice with the enclosed washing solution (Sigma) and DNA was eluted from the membrane in 50 µL of sterile dH_2O, incubated at 40°C

for enhanced elution and centrifugation at 16000 x *g* for 3 min. The resulting purified DNA was sorted at -20°C.

2.6.4 Extraction of clostridial genomic DNA using Phenol-Chloroform.

Between 1 to 5 mL of culture was pelleted at 4000 x g for 3 min at room temperature. Supernatant was removed by decanting, followed by thorough pipetting. Bacterial pellets were re-suspended in 200 µL PBS with 10 mg/mL lysozyme. Cells were lysed at 37°C for 30 min with occasional mixing. After degradation of the cell wall with lysozyme, 25 µL proteinase K, 85 µL dH2O and 100 µL 10% SDS solution were added and vortexed. Incubation of the mixture proceeded at 55°C for 30 min with occasional vortexing. A 400 µL volume of Chloroform: Isoamyl alcohol (25:24:1) with 10 mM Tris, pH 8.0, 1 mM EDTA was added to phase-lock gel tubes and after proteinase treatment the samples were added to the prepared phase-lock tubes. Samples were then centrifuged at 14,000x g for 2 min at room temperature allowing the extraction of the top layer to which 20 µL RNAse 65 was added and incubated for 2 min at room temperature. Phase-lock clean-up was repeated a further two times. Eventually, the top layer was mixed with 40 µL of 3 M Sodium acetate and 800 µL of 100% ethanol. After thorough resuspension, the sample mix was inoculated at -80°C for at least 30 min. Sample were pelleted in a 4°C pre-cooled centrifuge at 14,000 x g for 15 min. The

supernatant was carefully removed by aspiration. A 1 mL volume of 70% ethanol was added to the pellet and a further centrifugation step was performed at 14,000 x g for 5 min. The supernatant was poured off and the remainder removed carefully using a P200 pipette. The pellet was air-dried for around 45 min before being resuspended in 50-100 µL sterile dH2O.

2.6.5 Restriction digest of genomic and plasmid DNA.

DNA was cleaved at specific sites using appropriate restriction endonucleases Mix (NEB) based on manufacturer's instructions. Typically, reactions were performed in 25 µL volumes with incubation at 37°C for one hour, followed by 25 min heat inactivation at 80°C. The resulting DNA was measured with a nanodrop ND-1000 spectrophotometer.

Table 2.6.1: Restriction endonuclease master mix on the example of HindIII digestion of *Clostridia* genomic DNA.

Ingredient	µL
gDNA	15
CutSmart Buffer	2.5
HindIII	1
H_2O	6.5
Total Volume	25

2.6.6 DNA ligation.

Fragments with compatible ends resulted from the restriction digested genomic DNA were ligated together using T4 DNA ligase (NEB) in accordance with the manufacturer's instructions. Typically, reactions were performed in 10 µL volumes with incubation at room temperature for one hour, followed by 25 min heat inactivation at 65°C.

2.6.7 Polymerase Chain Reaction amplification of DNA.

Oligonucleotide primers for utilized for DNA amplification, described above, were synthesised by Sigma. Primers were diluted to a working concentration of 10 µM and stored at -20°C. The PCR mix was prepared typically in 25 µL reaction Volume. For amplification of plasmid sequences, DreamTaq (Sigma) master mix was applied according to manufacturer's protocol. For determination of transposon insertion, Phusion (NEB) master mix was applied manufacturer's protocol. The annealing temperature for each individual reaction was calculated using NCBI online-primer BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). While elongation time was determined according to the expected product size, with 30 secs allocated per 1 kb.

2.6.7.1 Inverse PCR.

During this protocol, a possible transposon mutant colony was grown in supplemented LB medium overnight. 1-5 mL of the culture were isolated using the Spin columns or phenol-chloroform. Subsequently, the DNA was digested with HindIII master mix (as described previously). Digestion was conducted for 4 hours, followed by a heat inactivation 80^{o}C for 25:00 min. A total of 30-75 ng of the heat inactivated digest was ligated overnight (on Ice) or for 4 hours at room temperature to form Transposon-Chromosome junction – ring formation. These ring structures were amplified by an intricated PCR run, using (NEB) Q5 high fidelity PCR protocol (Table 2.6.2) and the Forward primer catP_INV_F1 as well as Forward primer catP_INV_R1. The inverse PCR products amplified were visualized on by agarose gel electrophoresis. The fragments were obtained and extracted from the agarose. For sequence identification, the samples were sequenced and the data obtained blasted, using run products on gel, extract fragments, and sequence with catP-INV-R2.

Table2.6.2: Integrated PCR protocol to amplify chromosome transposon junctions. To run this PCR reaction, a "touch-down" protocol has been used, as follows:

Temp (°C)	Time (min)	Description
98	05:00	
98	0:10	
70-55	2:00	Temperature decreases by 1°C per cycle until 60°C is reached.
72	2:00	
98	0:10	
60	0:30	20-25 cycles
72	2:00	
72	5:00	
15	Pause	

2.7 Isolation of *C. difficile* spores.

2.7.1 Spore purification.

Prior to spore isolation, *C. difficile* was cultivated in 200 mL of LB broth, supplemented according to Table 2.5.1, at 37°C in the anaerobic workstation. The cultures were incubated for at least 10 days (mostly 14 days) to induce spore formation by starvation.

After the starvation, the samples were heat treated at 65°C for 20 min in the HB-1D hybridizing oven (Techne) to eliminate vegetative growth. The spores were left at room temperature to allow the mixture to cool down and split into 4 x 50 mL cultures. The cultures were centrifuged at 3000xg for 10 min at room temperature. The supernatant was discarded and the spore pellet resuspended in 10 mL dH_2O to which was added 10 mL 95 % EtOH and the mixture vortexed for 10 min. Centrifugation was repeated and the spore pellet was washed in 10 mL of dH20 and vortexed for 10 min. The washing step was repeated a further two times. Eventually, the spores were spun down, and the final pellet was resuspended in 44 µL of dH_20.

2.7.1 DNAse clean treatment of spores.

To the 44 µL of dH_2O spore mixture, 5 µL 10 X TURBO DNase Buffer and 1 µL TURBO DNase (Ambion) were added and incubated at 37°C for 30 min. The spores were spun at 3000xg for 10 min at room temperature and the supernatant was discarded from the pellet that was resuspended in 500 µL of 20 % HistoDenz (Merck).

2.7.3 Separation in Sucrose gradient.

The 500 µL of spore-20% HistoDenz mixture was gently transferred via pipetting into a 1 mL Eppendorf tube, containing 1 mL of 50 % HistoDenz. The solution was spun at 14.000 x g for 15 min at 4°C in a pre-cooled table centrifuge. The supernatant was discarded, and the spore pellet washed in 1 mL of dH2O and centrifuged at 5000 x g for 3 min at 4°C. The washing was repeated a further two times and the final pellet resuspended in 1 mL PBS for microscope quantification and final storage.

2.8 Transposon mutant library preparation.

2.8.1 Library preparation.

Genomic DNA samples were fragmented via sonication using a Covaris S-series sonolab at Deepseq (Nottingham). DNA fragments with an average size of around 500 bp were produced using the following settings: 20W, 200 cycles/burst, intensity 5, 10% dutycycle. A 1X bead purification was performed using NEB sample purification beads before fragments were end repaired and dA-tailed using NEBNext Ultra II DNA Library Prep reagents. A custom splinkerette adapter (

https://www.ncbi.nlm.nih.gov/pubmed/26794317) was ligated to the ends of fragments using the ligation reagents from the NEBNext Ultra II kit. Adapter-ligated samples were purified again with a 1X bead purification step before an I-*SceI* digest was performed to selectively cut between the library primer binding site and the splinkerette adapter on plasmid DNA only. Digested products were again purified using a 1X bead clean-up before a PCR was performed amplifying the transposon-genomic DNA junctions. The PCR employed the NEB Q5 high fidelity polymerase which amplified the region between the library primer "P7" and the adapter primer SplAP5.x where x represents the multiplex number. SplAP5 primers bind to the adapter and contain a unique region allowing multiplexing of samples. The library primer contains the P5 flow-cell binding region, while the adapter primer contains the P7 flow-cell binding region. PCRs were size-selected on a 2 % low-melt agarose gel and then gel extracted using a NEB monarch gel extraction kit. Gel extracted samples were quantified using an agilent 2100 bioanalyzer and a DNA 1000 chip to visualise the size distribution of the library. The concentration of DNA going onto the sequencing cartridge was assessed via Qubit and qPCR using the flow-cell binding regions to amplify.

2.8.2 MiSeq.

Samples were run at Durham University on a MiSeq using a custom sequencing primer "X" which was designed to read out of the transposon giving initially 12 bp of the transposon 'tag' and thereafter genomic DNA. A custom recipe MiSeq recipe was also used which employed dark cycles over the first 12 bp of reads hence preventing run failure from an inability to accurately detect clusters. The dark cycles resulted in accurate calling of cluster location since after 12 bp the bases are not homogenous throughout the whole sequencing lane.

Table 2.8.1: Primers utilized in this study (* Indicates Phosphorylation)

Primer Name	Sequence	Use
MTV10.1 F Aligns to pRF215	CGGCCGCTGTATTACCCTGTTATCCCTACCAG TGTGCTGGAATTCGCCCTTAG	For MTV10 alignment to pRF215
MTV10.1 R Aligns to pRF215	CGGGCGCGCCTTAAGACCCACTTTCACATTTA AGTTGTTTTTCTAATCCG	For MTV10 alignment to pRF215
MTV10.2 F Aligns to pRF215	TGGGTCTTAAGGCGCGCCCGCCCTTAAG	For MTV10 alignment to pMTL-84151
MTV10.2 R Aligns to pRF215	TCGCCCTTAGTAGGGATAACAGGGTAATACA GCGGCCGCGGTCATAG	For MTV10 alignment to pMTL-84151
ermB-R1	GCATCTAATTTAACTTCAATTCC	Inverse PCR of pRF215 and MTV10
catP_INV_R_1	TATTTGTGTGATATCCACTTTAACGGTCATGC TGTAGGTACAAGG	Inverse PCR of the CW-plasmids
catP_INV_F_1	TATTGTATAGCTTGGTATCATCTCATCATATA TCCCCAATTCACC	*catP* sequencing primer
catP_INV_R_2	GGCAAGTGTTCAAGAAGTTATTAAGTCGGGA GTGCAGTCGAAGTGG	
SplA5_top	G*AGATCGGTCTCGGCATTCCTGCTGAACCG CTCTTCCGATC*T	Primers for the Sprinklerette Adaptor
SplA5_bottom	G*ATCGGAAGAGCGGTTCAGCAGGTTTTTTT TTTCAAAAAAA*A	

SplAP5.7	C*AAGCAGAAGACGGCATACGAGATCAGATCTGGAGATCGGTCTCGGCATTC*C	Index Primer for the PCR during the library prep prior to MiSeq
SplAP5.8	C*AAGCAGAAGACGGCATACGAGATCATCAAGTGAGATCGGTCTCGGCATTC*C	
SplAP5.11	C*AAGCAGAAGACGGCATACGAGATCTGTAGCCGAGATCGGTCTCGGCATTC*C	
SplAP5.12	C*AAGCAGAAGACGGCATACGAGATAGTACAAGGAGATCGGTCTCGGCATTC*C	
SplAP5.13	C*AAGCAGAAGACGGCATACGAGATAACAACCAGAGATCGGTCTCGGCATTC*C	
SplAP5.14	C*AAGCAGAAGACGGCATACGAGATAACCGAGAGAGATCGGTCTCGGCATTC*C	
SplAP5.15	C*AAGCAGAAGACGGCATACGAGATAACGCTTAGAGATCGGTCTCGGCATTC*C	
SplAP5.16	C*AAGCAGAAGACGGCATACGAGATAAGACGGAGAGATCGGTCTCGGCATTC*C	
SplAP5.17	C*AAGCAGAAGACGGCATACGAGATAAGGTACAGAGATCGGTCTCGGCATTC*C	
SplAP5.18	C*AAGCAGAAGACGGCATACGAGATACACAGAAGAGATCGGTCTCGGCATTC*C	
ermB Tn_seq_ library_primer	AATGATACGGCGACCACCGAGATCTACACCTATCAACACACTCTTAAGTTTGCTTCTGTCAGA	Library Primer
ErmB_Tn_seq_sequencing primer	GCTTCTGTCAGACCGGGGACTTATCA	Sequencing Primer

2.8.3 Agarose gel electrophoresis.

Separation of PCR products, restriction fragments and plasmid DNA was performed via electrophoresis at 100 V for 60 min through 1.0 % agarose (Sigma) gels in TAE buffer containing 0.01 % (v/v) SYBR Safe DNA Gel Stain (Thermo Scientific). Separated DNA was subsequently viewed under blue light.

2.8.4 DNA extraction from agarose gels and reaction mixtures.

DNA fragments were visualized under blue light. Anticipated DNA fragments were excised from the gel using a scalpel. The DNA was extracted from the agarose using Monarch® PCR & DNA Clean-up Kit (NEB) based on manufacturer's protocol. DNA was eluted in 20 µL of distilled H_2O and stored in -20°C for later usage.

2.8.5 DNA sequencing and genome assembly.

DNA sequencing was performed by Eurofins Genomic (https://www.eurofinsgenomics.eu/en/ecom/cart/) without any deviations from protocol using the associated primers. Nanopore sequencing data were matched to the previously acquired Illumina sequencing data. Genome assembly was performed by Deepseq, using Vanu 2.0 (https://canu.readthedocs.io/en/latest/)

Chapter 3

Genetic Analysis of the Restriction Modification Barrier in *Clostridioides difficile* PCR RT 078

3.1 Introduction

3.1.1 Restriction modification systems in *C. difficile*

Restriction-Modification (R-M) systems are utilized in prokaryotic organisms to detect, cut, and modify specific DNA sequences. Generally, they protect the cell from invading DNA, such as that originating from bacteriophages. Defensive properties in bacteria were first observed in the 1950s, when *E. coli* B phage λ was found to replicate poorly in *E. coli* K-12. Restriction was observed for the non-methylated phage DNA, whilst the *E. coli* K12 genome remained intact due to site-specific methylation by the cognate methyltransferases (MTase) [105, 106]. Since R-M systems limit the entry of foreign DNA into the cell, they are generally considered to serve as an innate immune system for prokaryotes [107]. Initially, DNA modifications are carried out by MTase. These enzymes transfer methyl groups from the universal substrate, S-adenosyl-methionine, to their respective recognition site on the genome [108]. Bacteria then recruit restriction endonucleases (REases) which recognize non-methylated double stranded DNA. These REases clear the affected sites by causing DNA double strand cleavage. Restriction and methylation may be conducted by different enzymes or enzyme complexes, performing the individual steps utilizing different subunits. Based on

their activities, prokaryotic R-M systems can be differentiated into four classes, type I, type II, type III and type IV as depicted in Figure 3.1. Each of these systems perform restriction activity, while type I-III have additional methylation activity. Furthermore, R-M system I-III perform their activities on non-methylated DNA, while system IV exhibits nucleolytic activity against incorrectly modified DNA [109].

Type I restriction–modification systems. Type I R-M systems are encoded by three genes. The resulting proteins, transcribed form a pro tein complex. Its recognition sites on the genome consist of two allocated sequences, separated by a degenerate sequence. The R-M system I complex contains two individual target recognition domains (TRD), which detect and interact with the allocated DNA sequences. Additionally the MTase subunits interact with both strands of the DNA target. Many bacterial species, like *S. aureus* and *Mycoplasma spp.* contain several different TRD and methylation subunits. Responsible genes are in rotation of active and silent states, providing change in RM-system specificity [109].

Type II restriction–modification systems. Type II R-M systems are well investigated, as many endonucleases from this group are important enzymes in genetic engineering. Usually R-M system II encodes two genes, one for the restriction enzymes and another for the methyltransferase. Occasionally, these R-M systems also utilize other complexes, like transcription factors, nickases or DNA

repair systems to support their function. The Type II R-M system recognition site is often a 4-8 bp palindromic sequence. The utilized endonucleases are various and can be differentiated into 11 subfamilies. These enzymes bind to the unmethylated recognition site and cleave DNA at a defined location within or close to the recognition site.

Type III restriction–modification systems. R-M III-systems consist of two highly related subunits. These can modify and hydrolyze specific DNA sequences and recognize short nonpalindromic sequences. The methyltransferease consist of two subunits and modify one DNA strand after binding. The restriction enzymes cause DNA cleavage ~25-27 bp away from one of the recognition sites [110].

Type IV restriction–modification systems. In contrast to the previous systems, R-M system IV hydrolyzes only modified DNA. Thus, it is only made up of restriction enzyme components. Depending on the enzyme, these recognize modified cytosine residues or methylcytosine and methyladenine. The restriction components have a low specificity, which allows broad range protection against invading DNA with diverse methylation patterns. Like type II R-M systems, also type IV R-M systems are methyldirected and are encoded by one or two genes. There is no specific characteristic allowing to distinguish between Type IV and Type II R-M systems. Thus, some research suggests to reclassify specific Type II systems to Type IV [109].

Orphan methyltransferases. Orphan methyltransferases have no allocated restriction enzyme and act independently from the R-M systems. Most of these orphan enzymes are not well characterized. They can possess diverse cellular functions, including DNA replication/repair and even regulation of gene expression. Host protection from invading DNA in these cases is based on the underrepresentation of recognition sequence in the genome [111, 112].

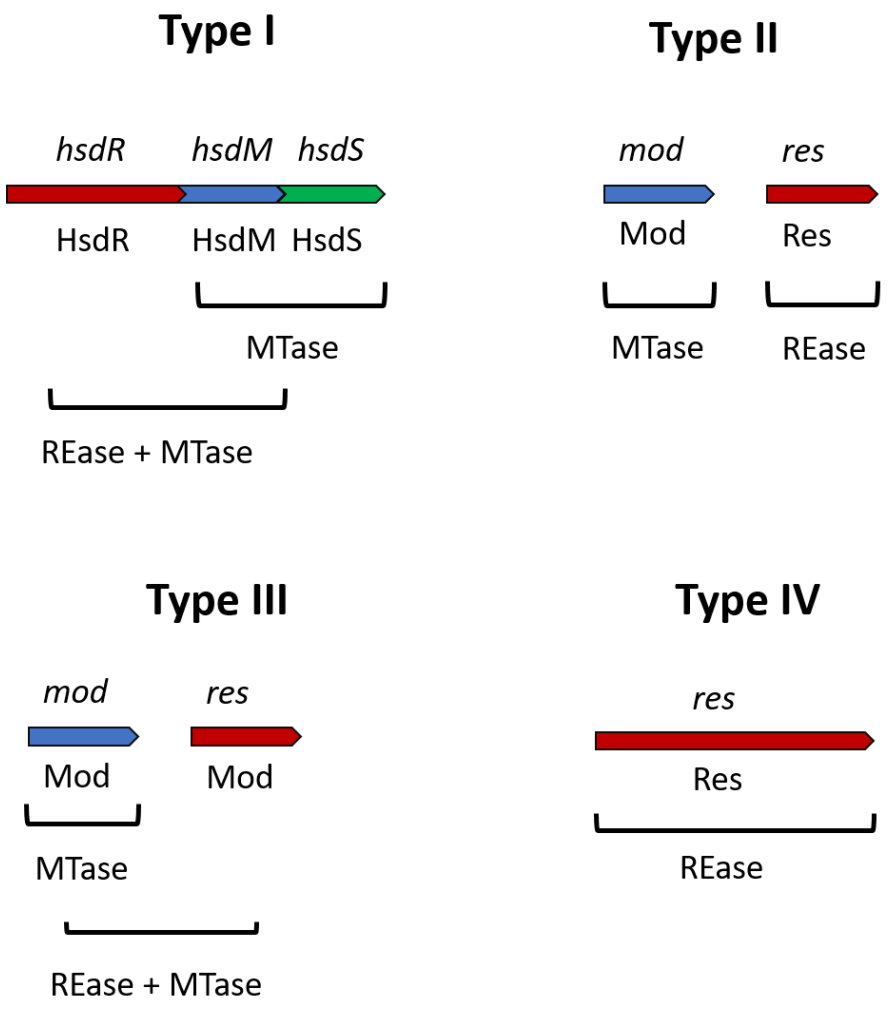


Figure 3.1: Schematic illustration of the R-M systems present in bacteria. Graphic was adapted from Andrew Dempster and Atack et al [113]. The R-M system type I contains 3 "host specific of DNA" or *hsd* genes. HsdR for restriction, *hsdM* for modification and *hsdS* for specificity. The *hsdM* and *hsdS* genes are transcribed from the same promotor and are important during methylation, while *hsdR* has its own promoter and is required for restriction. In type II systems the methylase and the endonuclease are two separate enzymes, Mod and Res. These R-M systems are the most prevalent in the bacterial kingdom. Like Type II systems, Type III R-M systems contain a Res and a Mod protein. In type III systems, these form a complex for modification and cleavage. Type IV systems do not contain a modification unit but are just comprised of an endonuclease, that detects and cuts modified DNA.

3.1.2 R-M Systems in clostridial genetics.

R-M systems in prokaryotic organisms are highly strain-specific. This has a direct impact on horizontal gene transfer between bacterial species and/or strains. Accordingly, transferring extrachromosomal DNA into *Clostridium spp*., by conjugation or transformation, may be impeded by the R-M system of the host. Efficient gene transfer into many strains of a number of *Clostridium* spp. is relatively straight forward, for example into *Clostridium beijerinckii* NCIMB 8052, *C. perfringens* 13, *C. difficile* strains CD37 and CD630 and *Clostridium botulinum* ATCC 3502 [67, 74, 114-116]. This is because although these strains possess at least one type II methylase gene, they lack the cognate restriction endonuclease gene [65]. In other cases, the presence of a complete R-M system (genes encoding both the restriction endonuclease and methylase enzymes) can prevent DNA transfer if the incoming DNA is not protected from cleavage by appropriate methylation of any targeted restriction enzyme recognition sequence. It follows that the successful transfer of DNA into a particular *Clostridium* recipient can require an understanding of number and specificity of any RM-system present. The action of these systems can then be avoided, either by *in vitro* or *in vivo* methylation of the donor DNA prior to its transfer or by negation of restriction activity by inactivating the encoding genes in the recipient [65]. RM-system specificity can be determined by the analysis of the fragmentation pattern obtained when plasmid DNA of known sequence is incubated with bacterial

lysates. By inspecting the resulting restriction pattern visible as DNA bands on an agarose gel, the targeted motives can be allocated. This can allow a gene encoding a methylase with the required specificity to be cloned and introduced into the *E. coli* donor strain. Most studies focused on evasion of RM-system type II, although the evasion of type I systems has also been reported. For example, countering the type I system of *Clostridium saccharobutylicum* NCP 262, resulted in an 8-10-fold increase in transfer frequencies [117]. In more recent times, the specificity of R-M systems has largely been determined through genome analysis, identifying specificity on the basis of homology to known systems [74, 115]. Genomics also allows the cloning of methylase genes in the recipient without needing to necessarily know the specificity of the R-M systems.

3.1.3 Rebase.

REBASE is a publicly accessible database, providing information about R-M systems [118, 119], which is reviewed and frequently updated. Single Molecule Real Time Sequencing (SMRT sequencing) improved the allocation of R-M systems, as DNA modifications can be directly detected during sequencing [120]. Restriction enzymes are always located in the immediate vicinity of their cognate methyltransferases. However, the identification of these complexes is difficult as

they undergo rapid evolutionary changes due to their protective function in host-parasite coevolution.

3.1.4 R-M's in *C. difficile*.

In some *Clostridium* species, successful introduction of foreign DNA plasmids is relatively easy to achieve. However, successful transformation of plasmid vectors in other strains, like *C. acetobutylicum* ATCC 824 [121] or *C. botulinum* type B [122] has only proven possible after circumvention of host restriction barriers. Until the early 2000s, gene transfer into *C. difficile* was based on conjugative transposons such as Tn*916*, which were transferred from *B. subtilis* donors using filter mating. These early protocols resulted in low frequencies of 10^{-8} per donor and the transposon insertion sites are limited to one single location into the genome [123]. In the early 2000s Purdy et al., characterized the R-M systems of two toxigenic strains of *C. difficile* by the conjugative transfer from *E. coli* of a number of autonomously replicating plasmids into *C. difficile* strain CD6 and CD3 [74]. The plasmids employed utilised the replication region of the *C. difficile* plasmid pCD6 and were transferred by *oriT*-mediated mobilisation. By analysing the fragmentation pattern obtained when DNA of a specific plasmid was incubated with the lysate of strain CD6 they concluded that it possessed two R-M systems: *Cdi*CD6I/M.*Cdi*CD6I and *Cdi*CD6II/M. *Cdi*CD6II, with equivalent

specificities to *Sau*96I/M. *Sau*96I (5'-GGN^{M}CC-3') and *Mbo*I/M. *Mbo*I (5'-G^{M}ATC-3'), respectively. Further analysis of the strain CD3, revealed that it possesses a type IIs restriction enzyme. The system, denoted *Cdi* I, cleaves the DNA sequence 5'-CATCG-3' between the fourth and fifth nucleotide resulting in a blunt-ended fragment. By measuring the frequency of plasmid transfer of various plasmids carrying different numbers of recognition sites, Purdy *et al*. [73], estimated that each site caused between a five- and 16-fold reduction in transfer efficiency representing a crucial factor in exogenous transfer into *C. difficile* [74]. In nontoxigenic *C. difficile* strain CD37 replication minus vectors were employed and designed to integrate into the genome by single crossover [76]. Successful integration of these plasmids was followed by the successful introduction of a mobilizable, replicative plasmid via conjugation with *E. coli* in CD37 [124]. Additionally, in 2003, Herbert *et al*, have shown that *oriT* based shuttle vectors may be transferred into *C. difficile* strain CD630 via conjugation at frequencies of 10^{-6} per donor cell. The transfer was not affected by the methylation status of the shuttle vector. Consistent with this observation, although the CD630 genome carried 5 different methylase genes but did not appear to carry any associated, cognate REase gene. This finding lead to the assumption, that the methylases play no role in the RM-system of CD630 but fulfil another function such as controlling of gene expression [115].

3.1.5 Improving Conjugation Efficiencies into *C. difficile.*

Conjugation may be the preferred way of DNA transfer into *C. difficile*. However, transfer frequencies of shuttle vectors remain low, especially into hypervirulent strains belonging to *C. difficile* PCR RT 027 [125]. R-M systems play a crucial role in DNA transfer and recognition thus impacting conjugation and successful plasmid transfer. R-M types I-III recognize non-modified DNA. R-M type IV, however, is conserved amongst most *Clostridium spp.* and recognises foreign modification patterns.

A variety of *E. coli* strains can function as conjugative donors in the transfer of plasmid vector DNA into clostridial recipients. Vectors are mobilised from the donor through the action of transfer (Tra) functions on a vector encoded origin of transfer (*oriT*) derived from plasmid pRK2. The Tra-functions may either be encoded on a co-resident large plasmid, such as the R-factor R702, or localised in the chromosome, as in the donor strains S17.1 and SM10 [65]. However, most genetic studies in *C. autoethanogenum*, *C. difficile* and *C. sporogenes* have utilized the *E. coli* strain CA434*,* where the *tra* genes are carried by R702 [74]. Recently, Woods *et al.,* created a novel conjugative donor to improve DNA transfer from *E. coli* to *C. autoethanogenum*, *C. sporogenes* and *C. difficile* R20291 [126]. Designated 'sExpress', as it is based on NEB Express (NEB Biolabs), it is a *dcm* mutant. As a consequence, the internal cytosine in the motif CCWGG is not

methylated. Therefore, plasmid DNA maintained in sExpress is not cleaved by cytosine-specific Type IV restriction systems. As most *C. difficile* strains carry a Type IV system, transfer of plasmids from sExpress occurs at 10-fold higher frequencies than from CA434. Similar improvements are evident to any clostridial donor that possesses a Type IV system [102].

3.2 Aims of the study

Many genetic protocols in bacteria heavily rely on preceding DNA transfer. However, establishing DNA transfer into *C. difficile* remains challenging and labour intensive. As the *C. difficile* RT078 has been poorly researched, we aim to find a suitable donor strain that yields a high rate of plasmid transfer. To answer this question we aim to evaluate which R-M systems can be identified in RT 078 by whole genome sequencing. Following these experimental set ups, we aim to establish, which rationale strategies result in improvements in DNA transfer frequency and subsequent mapping of transposon insertions following implementation of TraDIS.

3.3 Results

3.2.1 Genome sequencing

The initial step in the identification of possible R-M systems in RT 078 *C. difficile* strains was the selection of 10 different clinical isolates from the *Clostridia* Research Group culture collection (CRG, Nottingham). Cultures were grown overnight from which genomic DNA was prepared via phenol-chloroform extraction, as described in the Material and Method section 2.6.4 DNA samples were then sent for genome sequencing by both Illumina MiSeq and Oxford Nanopore technologies (DeepSeq, Nottingham). The reads generated by Illumina sequencing were aligned with the M120 draft genome annotation using CLC Genomics Workbench software. Variants were called according to the guidelines stated in Materials and Methods. Additional Oxford Nanopore analysis of the strains was performed to obtain maximum coverage of the respective genomes. Nanopore sequencing can also potentially provide information on those motifs that are methylated, which could allow the identification of R-M recognition sites. The 10 strains chosen are listed in Table 3.1. Strain CD1801, CD4401, CD9001, CD9301, CD9501 and CD9701 were originally sampled from patient material in

QMC, Nottingham. Strains EK24, EK26 and EK28 were isolated from across Europe by Ed Kuijper, while the M120 reference strain originated in London.

Table 3.1: Information about the C. *difficile* RT 078 strains.

C. difficile Strain	Working Name	Country	Isolated by	Source	GenBank Accession No.
9301	CD9301	UK	M. Lister	UoN	CP068561
9001	CD9001	UK	M. Lister	UoN	CP068560
9701	CD9701	UK	M. Lister	UoN	CP068559
9501	CD9501	UK	M. Lister	UoN	CP068558
4401	CD4401	UK	M. Lister	UoN	CP068667
1801	CD1801	UK	M. Lister	UoN	CP068556
M120	M120	UK	B. Wren	LSHTM, London	CP068555
EK24	CD2315	Hungary	Ed Kuijper	LUMC, Leiden	CP068554
EK26	2016	Ireland	Ed Kuijper	LUMC, Leiden	CP068553
EK28	7009825	NL	Ed Kuijper	LUMC, Leiden	CP068552

*UoN, University of Nottingham,

*LSHTM, London School of Hygiene & Tropical Medicine,

*LUMC, Leiden University Medical Centre

3.2.2 Phylogenetic analysis

To find taxonomic variation between the 10 strains, phylogenetic analysis was performed, via Whole Genome Alignment (CLC, Workbench 20). Short stretches of nucleotide sequences shared between multiple genomes are identified and similarities between the genomes are processed. The resulting Average Nucleotide Identity (ANI) algorithm qualifies the genetic distance in between the analysed genomes. The ANI can be used to further create a Phylogenetic tree of the screened organisms. The length of the tree branches illustrates the evolutionary time between the next nodes, displayed as substitutions/sequence site. Vertical lines represent evolutionary split to visualize connection. The CLC genomic alignment protocol applied on the sequenced RT 078 strains, resulted in clustering the candidates based on their pairwise nucleotide identity (Figure 3.2). The bioinformatic analysis identifies high levels of similarity of above 99% ANI between the genomes. The calculated ANI of control RT 027 strain R20291 and the RT 012 strain CD630 were estimated to be above 96%. This indicates a higher degree of genome similarity between the RT 078 genomes than with strains belonging to these other ribotypes. Some RT 078 strains, such as CD2315 or CD7009825, display taxonomic similarities of up to 100% ANI towards other RT 078 strains, in particular strain CD2016.

The Phylogenetic tree depicted in Figure 3.2B is based on ANI analysis. Strain CD9501 is clustered on a separate branch together with the other RT 078 strains (Figure 3.2B). As indicated in Figure 3.2A, all strain exhibits minimal differences in nucleotide analogy towards the other strains.

A

		CD630	CDR20291	1801	2016	4401	7009825	9001	9301	9501	9701	CD2315	M120
		1	2	3	4	5	6	7	8	9	10	11	12
CD630	**1**		98.96	96.48	96.50	96.49	96.50	96.50	96.49	96.60	96.50	96.50	96.50
CDR20291	**2**			96.50	96.50	96.51	96.50	96.50	96.48	96.52	96.50	96.50	96.50
1801	**3**				99.99	99.99	99.99	99.98	99.98	99.80	99.98	99.99	99.99
2016	**4**					99.99	100.00	99.98	99.98	99.81	99.98	100.00	99.99
4401	**5**						99.99	99.99	99.99	99.77	99.99	100.00	99.99
7009825	**6**							99.98	99.98	99.81	99.98	100.00	99.99
9001	**7**								99.99	99.80	99.99	99.98	99.99
9301	**8**									99.80	99.98	99.99	99.98
9501	**9**										99.81	99.81	99.81
9701	**10**											99.98	99.99
CD2315	**11**												99.99
M120	**12**												

B

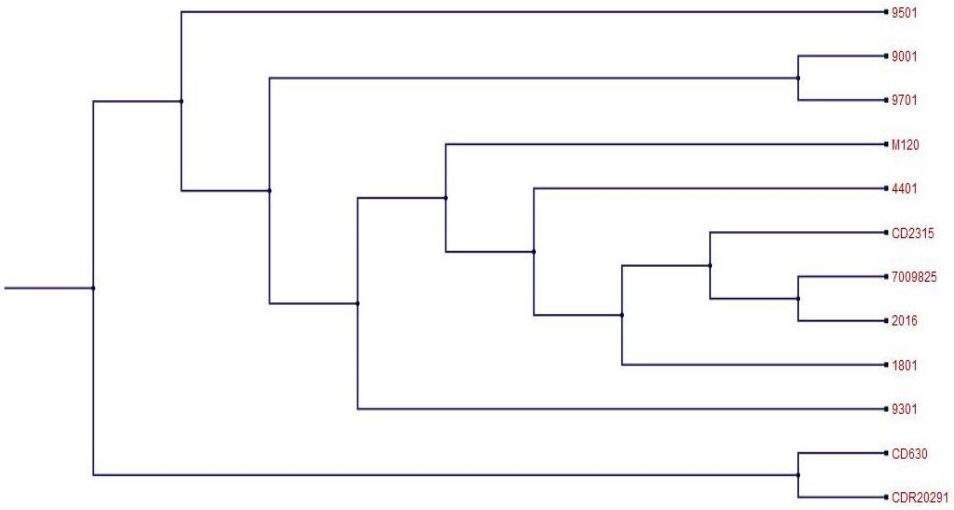


Figure 3.2: Phylogenetic tree and pairwise comparison of *C. difficile* RT 078 strains. (A)Pairwise comparison of genome sequences of the selected RT 078 strains. The comparison gradient indicates percentage identity between two genome sequences. Strain CDR20291 as well as strain CD630 have been included as reference sequences for other RTs. For the RT 078 strains, the GenBank identities have been used. (B)The Neighbour Joining (NJ) tree was constructed from the whole genome sequences of the selected C. difficile RT 078 strains. The branches are based on the comparative ANI, computed on the CLC workbench whole genome alignment software.

3.2.3 Finding a Conjugal Donor Strain for *C. difficile* RT 078.

To examine the efficiency of DNA transfer into the selected *C. difficile* RT 078 strains, shuttle vectors were transferred into each strain by conjugation and transfer frequencies calculated. As the transfer of shuttle vectors into hypervirulent *C. difficile* strains is observed at low efficiencies, the plasmid of choice for this assay was pMTL84151 [125], since this vector contains the *C. difficile* replicon CD6, which is stable and generally transferred at the highest frequency to this organism.

Conjugation is the preferred method of DNA transfer into *C. difficile*, thus a suitable conjugal donor must be selected. The HB101 *E. coli* strain CA434 carries the R factor R702 which contains the DNA transfer function, served as standard conjugal donor. The strain is *dam*$^+$ and *dcm*$^+$ which results in a Dam and Dcm

methylation pattern that may leave shuttle vectors susceptible to attack by Type IV restriction systems. As discussed in the introduction, *E. coli* strain 'sExpress,' is *dcm*- and as a consequence gives a 10-fold increase in DNA transfer efficiency into *C. difficile* R20291, which carries a Type IV R-M systems, relative to the widely used CA434 donor strain [102]. As such, sExpress could represent an optimal strain to transfer the shuttle vector pMTL84151 into the sequenced RT 078 strains, which were also found to carry type IV R-M systems (see Figure 3.5).

3.2.4 Comparison of conjugation efficiency between sExpress and CA434 for transfer into *C. difficile* RT 078.

To compare the relative efficiencies of CA434 and sExpress as conjugative donor strains each was transformed with a range of different *catP*–based, pMTL8x151 shuttle vectors carrying different Gram-positive replicons. As the RT 078 strains were shown to be highly genetically similar, CD9301 and M120 were chosen as representative recipient strains. Donor *E. coli* strains harbouring the different pMTL8X151 plasmids were mixed with overnight *C. difficile* according to the conjugation protocol described in section 2.5.9 and incubated together anaerobically on BHIS plates without selection for 24 hours. Subsequently, the plates were flushed with PBS and all growth harvested from the surface of the agar with a cell spreader. The cells were diluted to a factor of 10^{-7} and transferred to

solid selective media for selection of the desired transconjugants. After 48 hours the CFU were recorded, and the efficiency of plasmid transfer calculated. All conjugations were carried out in triplicate (Table 3.2)

Table 3.2: Conjugation efficiency into *C. difficile* CD9301 from *E. coli* donors CA434 and sExpress.

Test Plasmid	Gram-positive origin of replication	Conjugation efficiency (Transconjugants per recipient cell)	
		sExpress	CA434
pMTL81151	None	0	0
pMTL82151	pBP1	1.80×10^{-6}	6.34×10^{-8}
pMTL83151	pCB102	3.64×10^{-7}	5.57×10^{-8}
pMTL84151	pCD6	3.75×10^{-5}	2.93×10^{-7}
pMTL85151	pIM13	0	0
pMTL86151	pIP404	0	0

Of the tested replicons, transfer to strain CD9301 was only achieved when the replicons were those of pBP1, pCB102 and pCD6. As previously noted, [65] the highest frequency of transfer was obtained when the replicon was that of pCD6 (pMTL84151), whereas the lowest frequency of the three replication regions was obtained with the replicon of pCB102 (pMTL83151). The highest rates of transfer were obtained when the donor strain was sExpress, regardless of which replicon the vector employed.

Having established that sExpress was a superior donor relative to CA434 for transferring plasmids to the RT 078 strain CD9301, and that plasmid pMTL84151 was transferred at the highest frequencies, further conjugations were undertaken to see if any of the other clinical isolates were more effective recipients. Accordingly, conjugation experiments were repeated using the RT 078 strains CD1801, CD9501 and M120 as recipients. The results from these experiments are presented in Table 3.3 and Figure 3.3.

Table 3.3: Influence of Type IV R-M systems on the transfer of pMTL84151 into four different RT 078 strains. The total CFU refers to the colonies counted on BHIS medium supplemented with just CC to select against *E. coli* donor cells resulting in colonies solely formed by *C. difficile*. Positive transformants carry an erm^R are detected via their growth on BHIS supplemented with Em. The frequency of transfer has been developed by relating the total number of cells (Total CFU) and the number of transformant colonies on Em selection plates.

Recipient	sExpress			
	Total CFU	**Em^R CFU**	**Standard Deviation**	**Transfer Frequency**
1801	2.5×10^5	13.66	2.5	5.46×10^{-5}
9301	1.33×10^5	5.0	2.6	3.75×10^{-5}
9501	2.33×10^5	5.33	0.5	2.28×10^{-5}
M120	1.16×10^5	5.16	1.6	4.42×10^{-5}

Recipient	CA434			
	Total CFU	Em^R CFU	Transfer Frequency	Standard Deviation
1801	8.5×10^6	3.66	4.51×10^{-7}	1.6
9301	8.83×10^6	2.16	2.93×10^{-7}	1.5
9501	1.8×10^6	2.16	1.20×10^{-7}	0.7
M120	2.42×10^6	0	0	0

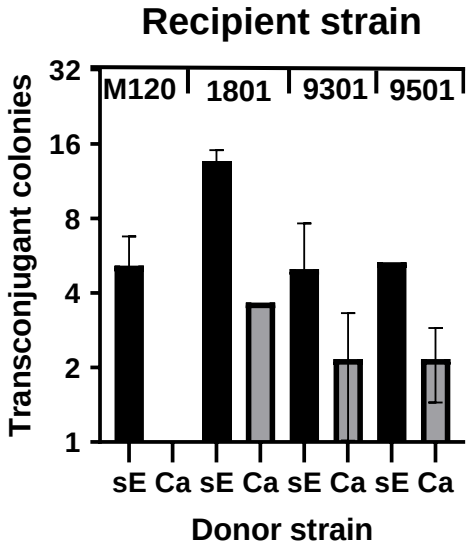
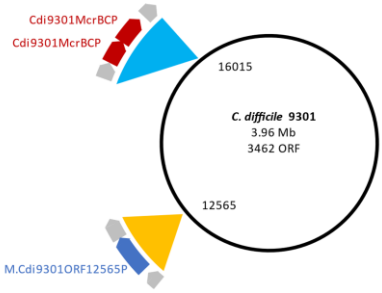


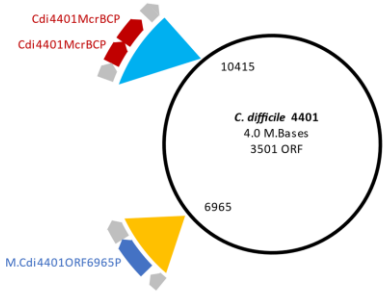
Figure 3.3: Analysis of the influence on conjugation of the Type IV R-M System. Graphic representation of the values presented in Table 3.3. Strains conjugates with the *E. coli* donor sExpress are depicted in black, whilst strains conjugated with Ca434 are presented in grey bars. This data shows the conjugation efficiency between the different conjugational donors in *C. difficile* strain 1801, 9301, 9501 and M120 utilizing the shuttle vector pMTL84151. After 72h, the CFU has been determined, the transformation efficiency has been calculated and plotted on the Graph. Experiments were performed in triplicate and the error bars represent the standard deviation depicted in Table 3.1. In all strains, CFU's of the strains conjugated with sExpress result in increased resistant colony growth on selective

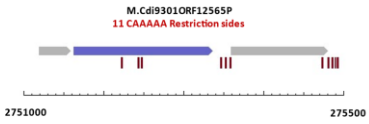
medium than strains conjugated with CA434. After 72h, M120 does not show any colony formation on BHIS supplementFed with Tm once CA434 was used as conjugational donor.

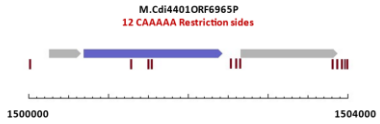
3.2.5 Evaluation of *C. difficile* RT 078 R-M Systems through Genome Sequencing.

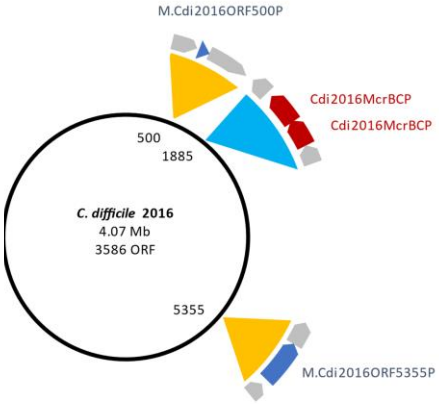
The determination of the entire nucleotide sequences and identification of possible components of R-M systems of the ten RT 078 strains revealed a high degree of similarity as well as some clear examples of diversity between strains with regard to R-M systems (see Fig 3.4).

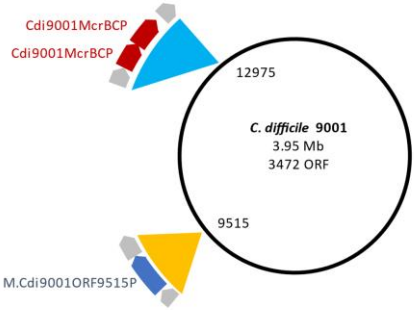

Cdi9301McrBCP
Cdi9301McrBCP
16015
C. difficile 9301
3.96 Mb
3462 ORF
12565
M.Cdi9301ORF12565P

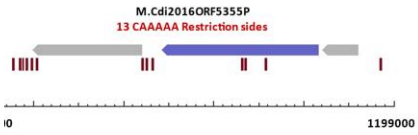

Cdi4401McrBCP
Cdi4401McrBCP
10415
C. difficile 4401
4.0 M.Bases
3501 ORF
6965
M.Cdi4401ORF6965P

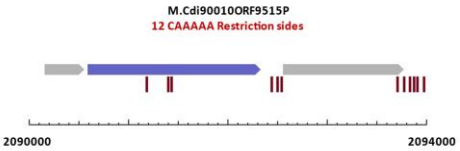

M.Cdi9301ORF12565P
11 CAAAAA Restriction sides
2751000
275500

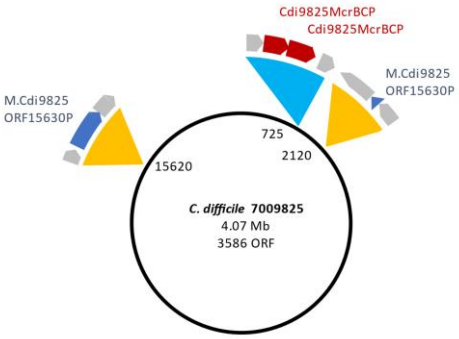

M.Cdi4401ORF6965P
12 CAAAAA Restriction sides
1500000
1504000

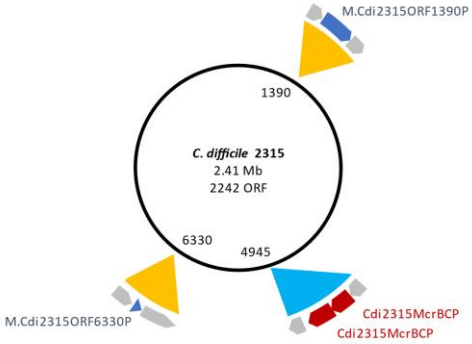

M.Cdi2016ORF500P
Cdi2016McrBCP
Cdi2016McrBCP
500
1885
C. difficile 2016
4.07 Mb
3586 ORF
5355
M.Cdi2016ORF5355P

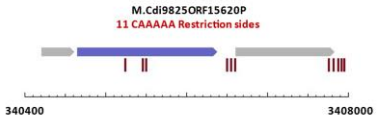

Cdi9001McrBCP
Cdi9001McrBCP
12975
C. difficile 9001
3.95 Mb
3472 ORF
9515
M.Cdi9001ORF9515P

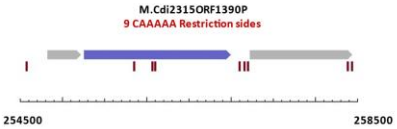

M.Cdi2016ORF5355P
13 CAAAAA Restriction sides
0
1199000

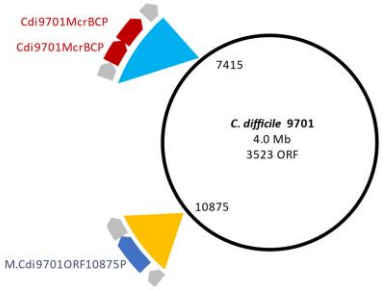

M.Cdi90010ORF9515P
12 CAAAAA Restriction sides
2090000
2094000

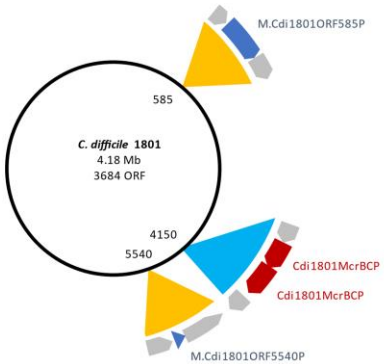

Cdi9825McrBCP
Cdi9825McrBCP
M.Cdi9825
ORF15630P
M.Cdi9825
ORF15630P
725
2120
15620
C. difficile 7009825
4.07 Mb
3586 ORF

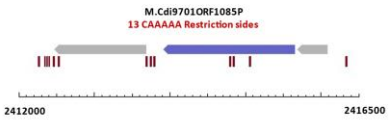

M.Cdi2315ORF1390P
1390
C. difficile 2315
2.41 Mb
2242 ORF
6330
4945
M.Cdi2315ORF6330P
Cdi2315McrBCP
Cdi2315McrBCP

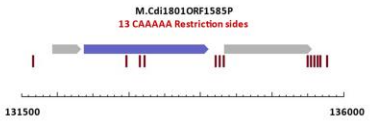

M.Cdi9825ORF15620P
11 CAAAAA Restriction sides
340400
3408000

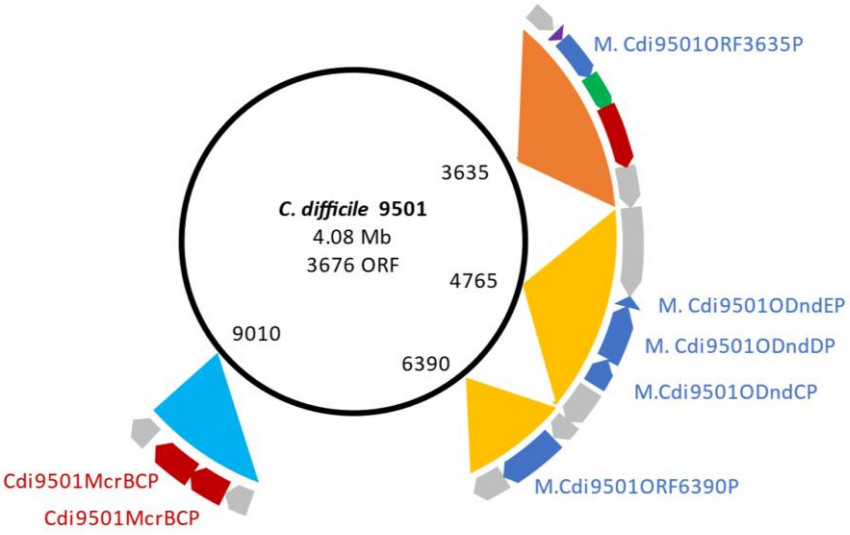

M.Cdi2315ORF1390P
9 CAAAAA Restriction sides
254500
258500

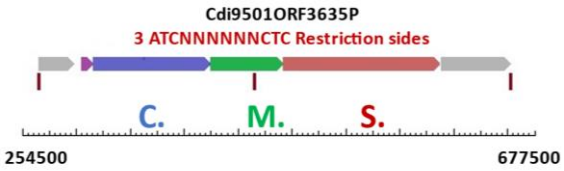

Cdi9701McrBCP
Cdi9701McrBCP
7415
C. difficile 9701
4.0 Mb
3523 ORF
10875
M.Cdi9701ORF10875P

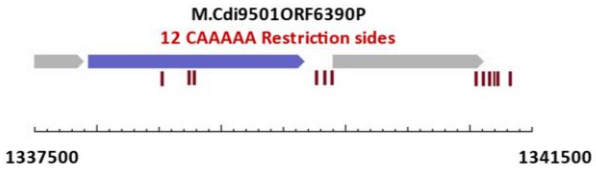

M.Cdi1801ORF585P
585
C. difficile 1801
4.18 Mb
3684 ORF
4150
5540
Cdi1801McrBCP
Cdi1801McrBCP
M.Cdi1801ORF5540P


M.Cdi9701ORF1085P
13 CAAAAA Restriction sides
2412000
2416500

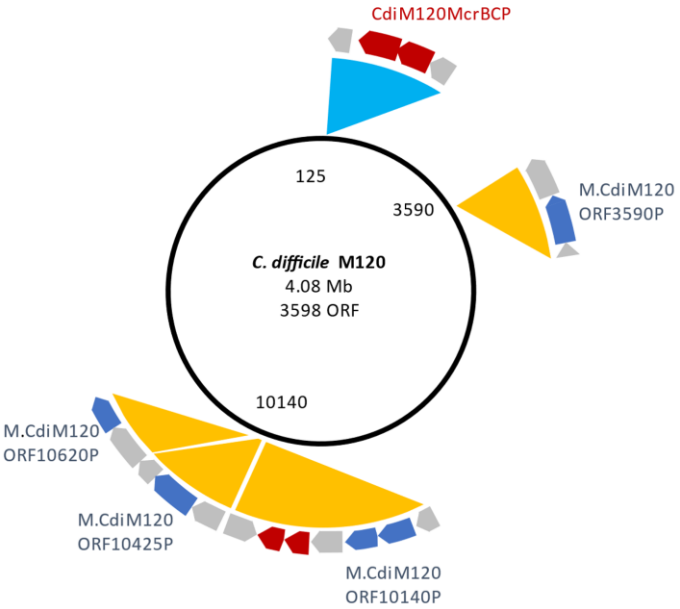

M.Cdi1801ORF1585P
13 CAAAAA Restriction sides
131500
136000

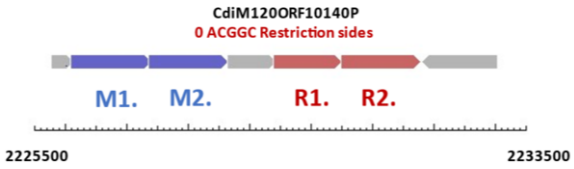

M. Cdi9501ORF3635P
3635
C. difficile 9501
4.08 Mb
3676 ORF
4765
M. Cdi9501ODndEP
M. Cdi9501ODndDP
9010
6390
M.Cdi9501ODndCP
Cdi9501McrBCP
Cdi9501McrBCP
M.Cdi9501ORF6390P

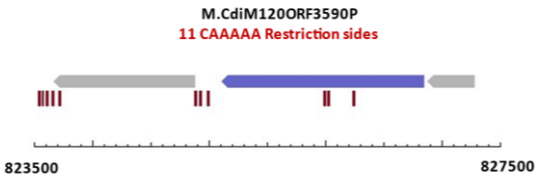

Cdi9501ORF3635P
3 ATCNNNNNNCTC Restriction sides
C.
M.
S.
254500
677500


M.Cdi9501ORF6390P
12 CAAAAA Restriction sides
1337500
1341500


ORF
Control Prot
Methylase
Restriction
Specificity
Type I
Type II
Type IV


CdiM120McrBCP
125
3590
M.CdiM120
ORF3590P
C. difficile M120
4.08 Mb
3598 ORF
10140
M.CdiM120
ORF10620P
M.CdiM120
ORF10425P
M.CdiM120
ORF10140P


CdiM120ORF10140P
0 ACGGC Restriction sides
M1.
M2.
R1.
R2.
2225500
2233500


M.CdiM120ORF3590P
11 CAAAAA Restriction sides
823500
827500


ORF
Control Prot
Methylase
Restriction
Specificity
Type I
Type II
Type IV

Figure 3.4: Graphic representation of the various R-M systems present in the PCR RT 078 strains, identified by REBASE. All strains contain a common Type IV R-M system, indicated by the blue guide comprising 2 adjacent REases. All strains contain at least one Type II orphan methylase (marked orange). Some strains (1801, 2315 and 2016) carry an additional gene coding for an orphan Type II methylase, while strain 9501 and M120 carry more than one. Accordingly, strain M120 carries a further Type II system comprising genes encoding two REases and two MTases, while strain 9501 is the only strain to carry a Type I system, marked brown.

3.2.6 Presence of a Common Type IV R-M System

Of particular note was the presence of a Type IV system comprised of two adjacent REases in all 10 strains. These are annotated as 'McrBCP' in all of the illustrated genomes (e.g., Cdi9301McrBCP and CdiM120McrBCP in strains CD9301 and M120, respectively) and share 100% identity between all ten strains. Their presence explains why plasmid DNA propagated in the *dcm*$^{+}$ *E. coli* donor strain CA434 is transferred at a lower frequency than from sExpress into strains, M120, 1801, 9301 and 9501. Further, it would be predicted that the use of the latter donor would give higher frequencies of transfer with the other sequenced RT 078 strains.

3.2.7 Presence of a Common Type II Orphan Methylase.

Similarly striking is the presence of a gene encoding a putative Type II N4-cytosine or N6-adenine DNA methyltransferase upstream of the above Type IV system probably recognizing CAAAAA. This gene is also conserved in all ten strains and apparently represents an orphan methylase having no cognate restriction enzyme gene in the near vicinity. This Type II methylase and the Type IV system comprising two REases, represent the only R-M systems carried by the strains CD4401, CD9001, CD9301 and CD9701. These methylases share 100% identity between strains, are encoded on the same DNA strand as the McrBCP REases and are positioned at a similar distance away, being between 3450bp and 3460bp. BLASTP searches demonstrated that a similar methylase was carried by a number of other clostridial species (Table 3.5).

Table 3.4: Homologous Systems to M2. CdiMORFEP in Other *Clostridium* spp. This can be calculated using the BlastN tool (NCBI). The maximum query coverage indicated the length of the sequences towards each other. The Identity refers to the extent to which two nucleotide sequences have the same residues at the same positions in an alignment.

Strains (Blast-P)	M2. CdiMORFEP (CAAAAA)	
	Max Query coverage (%)	Identity (%)
C. perfringens	59	26.65
C. botulinum	91	40.26
C. beijerinckii	57	29.21
C. sporogenes	88	39.96

3.2.8 Strain M120 carries two further Type II Systems.

The genome of strain M120 uniquely carries a cluster of 3 Type II R-M systems that are not present in any of the other nine sequenced strains. The larger of the three systems (CdiM120ORF10140P), comprises two sequentially located methylase genes (M1 and M2), followed almost immediately by two genes encoding a pair of distinct restriction endonucleases, R1 and R2. The two putative Type II N4-cytosine or N6-adenine DNA methyltransferase are annotated at REBASE as probably recognizing the sequence ACGGC. This same motif is assigned as the probable recognition sequence of the first (R1) of the two REases, a member of the LlaJI family. BLASTP searches of the methylases, demonstrated a high degree of

conservation of this system in other bacterial species (Table 3.6), including *Clostridium mangenotii LM2, Peptostreptococcaceae VA2* and *C. difficile* 630. The homologue in the latter species was previously designated CdiM (http://rebase.neb.com/cgi-bin/acroeget?CdiM+o5882).

Table 3.5: Homologous Systems in Other *Clostridium* spp.

Strains	M2. CdiMORFAP (ACGGC)	
	Max Query coverage (%)	Identity (%)
C. perfringens	**100**	**59.30**
C. botulinum	96	30.66
C. beijerinckii	**100**	**55.45**
C. sporogenes	89	26.28

Almost immediately downstream of the above, are two sequentially located Type II systems each composed of two orphan methyltransferases. The first, M.CdiM120ORF10245P, is a putative Type II N4-cytosine or N6-adenine DNA methyltransferase of unknown recognition sequence, while the second, M.CdiM120ORF10620P, is a putative Type II cytosine-5 DNA methyltransferase that also has an unknown recognition sequence. The absence of any restriction endonuclease genes in the immediate vicinity of these two systems suggest they do not play a role in restriction/modification.

3.2.9 A small Orphan MTase common to 4 Strains.

The Type IV system of four of the strains is characterised by the presence downstream of a Type II system comprising a small (186 amino acids) putative, orphan Type II N4-cytosine or N6-adenine DNA methyltransferase of unknown recognition sequence. The encoding gene is counter transcribed, relative to the McrBP genes, and in all four cases (CD1801 - M.Cdi1801ORF5540P, CD2016 - M.Cdi2016ORF500P, CD7009825 - M.Cdi9825ORF2120 and CD2315 - M.Cdi2315ORF6330P) is located some 300 kb away.

3.2.10 Only Strain CD9501 Carries a Type I R-M System.

Of the ten strains, only strain CD9501 appears to carry a Type I R-M system (Figure 3.5) comprising three genes encoding a HsdM methyltransferase (M.Cdi9501ORF3635P), a specificity function HsdS (S.Cdi9501ORF3635P) and a restriction endonuclease HsdR (R.Cdi9501ORF3635P), all with a recognition motif ATCNNNNNNCTC. Type I R-M systems are common in many types of bacteria and have been identified in several clostridial species. REBASE analysis previously predicted this type I system to be involved in *C. difficile* NCTC13748 and BLASTP reveals orthologs of this system in *C. botulinum, C. tetani, C. beijerinckii* and *C. sporogenes* (Table 3.10).

In addition to the Type I R-M system, the genome of strain CD9501 also carries two Type II systems that are devoid of any restriction endonuclease genes. The smaller of the two encodes a putative Type II N4-cytosine or N6-adenine DNA methyltransferase common to all ten strains (Table 3.8), probably recognizing CAAAAA. The larger system encompasses three putative Type II N4-cytosine or N6-adenine DNA methyltransferase of unknown recognition sequence. In descending order of size, M.Cdi9501DndDP (714 amino acids), M.Cdi9501DndCP (485 amino acids) and M.Cdi9501DndEP (127 amino acids) are not found in the other nine RT 078 strains.

3.2.11 Nanopore analysis of *C. difficile* methylation Motifs

To garner further insight on the R-M systems of the ten RT 078, their methylation patterns were analysed using Oxford Nanopore (Deepseq, Nottingham). The predicted N6-methyladenine (6ma) and 5-methylcytosine (5mc) methylation Motif plots are linked to the corresponding RT 078 strain and the GenBank assession No. and visualized in Table 3.8. The Nanopore sequencing data were matched to the previously acquired Illumina sequencing data. The coverage of the genome sequencing is represented in Table 3.7.

Table 3.6: Genome coverage of sequencing: This can be calculated by dividing the number of bases sequenced by the expected genome size and multiplying that by the percentage of bases that were placed in the final assembly. More simply it is the number of bases sequenced divided by the expected genome size.

Strain	Genome Coverage (x)
1801	200
4401	200
9001	183
9301	200
9501	200
9701	200
EK24	200
EK26	200
EK28	200
M120	200

The data derived from Nanopore yielded a bewildering array of putative methylation motifs. Some evidence was obtained to support the methylation of CAAAAA, as suggested by REBASE, but only in 9 of the 10 strains. Thus, no evidence for methylation of this sequence was obtained with strain CD9501. Data obtained with M120 DNA did not support the methylation of ACGGC nor the predicted recognition sequence (ATCNNNNNNCTC) of the CD9501 Type I system. The data was, therefore, considered inconclusive, and possible reflects the fact that the DeepSeq facility at Nottingham had no prior experience of undertaking this sort of analysis on bacterial genomes. Nanopore sequencing can detect

Table 3.7 : Methylation motifs identification by Nanopore sequencing

Strain name (Accession No.)	**Motif plot 5mc**	**Motif plot 6ma**
***C. difficile* 1801** (CP068556)	AGGAAC CTTC GTTCCT	AAATTA AGGAGC CCAA C**CAAAAA**H CWAATT
***C. difficile* 4401** (CP068557)	CTTC TCCT	AAAAAATAAC AAAT AGGAGC C**CAAAAA**H CCTATTAC TAAATTA TGAGTAA TTAAAG
***C. difficile* 9001** (CP068560)	CTTC TCCT	AAATTA AAGAGT AGGAGC **CAAAAA**H GAGDATA GTACTA TGAGTAA
***C. difficile* 9301** (CP068561)	CTTC GCCCCT GCCCT GTTCCT GTTTC	AAATTA ACCTAG AGGAGC CCAAAAAH TGAGTAA TTTAAGCAA
***C. difficile* 9501** CP068558	CTCT GTCC	CAAGT CAATC
***C. difficile* 9701** (CP068559(	CTTC GTTTC GTTTCC	AAATTA ACTTTAA AGGAGC C**CAAAAA**H
***C. difficile* EK-24** (CP068554)	CTTC CTTCGCGCTT GTTTC TGGCGC TTWCCT	AAATTA AAGAGT AGGAGC C**CAAAAA**H GAGCCAA TGAGTAA TTTAAGCA
***C. difficile* EK-26** (CP068553)	ACTTACATT CTTC CTTCGCGCTT GTTTC GVCGCGC TTACCT TTCCT	AAATTA AAGAGT AGGAGC CAAGGA CACTTTAA C**CAAAAA**H GACTAA
***C. difficile* EK-28** (CP068552)	CTTC GTTTC TGGCCC TTGCAT	AAATTA AGGAGC CCAAAAAH TGAGTAA

***C. difficile* M120** (CP068555)	CTDCT CTTCGCGCTT GCCGT GTTTC	AAATTA AGGAGC CAAATTTA C**CAAAAA**H GACACCTAG GGCCCAA

methylations of 5mc with an accuracy of up to 80% but the 6ma with an accuracy of lower than 70% [127, 128]. *PacBio* Single Molecule, Real-Time (SMRT) Sequencing has been described to most sensitively detects 4mC and 6mA but is outperformed by Nanopore in 5mc [127]. Nanopore based predictions, ascertained by subsequent SMRT analysis may result in an increasingly detailed and reliable estimation of methylation patterns involved in the RT 078 strains. Greater faith was therefore placed in the predictions of RM-associated genes identified by Rich Roberts (REBASE, NEB) using the gene prediction and homology searching tool, SEQWARE (Table 3.9). These genes, hare hypothesised to be involved in R-M systems. On the genome, these genes are typically encoded closely to each other and can be allocated towards a type of R-M system. Once the Type is predicted, candidate MTases are assigned to sequences predicted via Nanopore and known motif specificities in REBASE [129] .

Table 3.8: REBASE predicted motifs involved in *C. difficile* 078 methylation.

Organism	Enzyme	Recognition Sequence
Common, Type II methyltransferase		
***C. difficile* 7009825**	M.Cdi9825ORF15620P	CAAAAA
***C. difficile* 2016**	M.Cdi2016ORF5355P	CAAAAA
***C. difficile* CD2315**	M.Cdi2315ORF1390P	CAAAAA
***C. difficile* M120**	M.CdiM120ORF3590P	CAAAAA
***C. difficile* CD1801**	M.Cdi1801ORF585P	CAAAAA
***C. difficile* CD4401**	M.Cdi4401ORF6965P	CAAAAA
***C. difficile* CD9501**	M.Cdi9501ORF6390P	CAAAAA
***C. difficile* CD9701**	M.Cdi9701ORF10875P	CAAAAA
***C. difficile* CD9001**	M.Cdi9001ORF9515P	CAAAAA
***C. difficile* CD9301**	M.Cdi9301ORF12565P	CAAAAA
Unique Type II R-M System		
***C. difficile* M120**	M1.CdiM120ORF10140P	ACGGC
***C. difficile* M120**	M2.CdiM120ORF10140P	ACGGC
***C. difficile* M120**	R1.CdiM120ORF10140P	ACGGC
Unique Type I R-M System		
***C. difficile* CD9501**	Cdi9501ORF3635P	ATCNNNNNNCTC
***C. difficile* CD9501**	M.Cdi9501ORF3635P	ATCNNNNNNCTC
***C. difficile* CD9501**	S.Cdi9501ORF3635P	ATCNNNNNNCTC

Table 3.9: Homologues of the CD9501 Type I HsdM in other *Clostridia* species.

Strains	Type I system methyltransferase	
	Maximum Query coverage	Identity (%)
C. tetani	98	54.29
C. botulinum	98	52.76
C. beijerinckii	98	52.55
C. sporogenes	98	54.29

3.3 Discussion

As the overall goal of this thesis was to apply TraDIS to a RT 078 strain, initial steps focused on identifying the most 'competent' donor and recipient strain that would yield the highest frequency of transconjugants in conjugation experiments. As R-M systems can represent a formidable barrier to effective plasmid transfer, an initial goal was to use a combination of genomics and strain selection to identify and counter any problematical restriction endonucleases. At the outset, a total of ten strains were selected, including the RT 078 type strain, M120, and subjected to whole genome sequence. The strategy taken was to combine paired-end Illumina sequencing with Nanopore. The latter gave the long reads necessary to obtain a whole, contiguous genome, while the former reads were used to correct any errors made by Nanopore.

The create ANI Comparison tool (CLC, Workbench 20) was used to obtain a quantitative measure of the similarity between the 10 *C. difficile* RT 078 genomes. For comparative purposes the genomes of *C. difficile* RT 012 and RT 027 strains, CD630 and R20291, were included in the phylogenic analysis. From the plotted bifurcating tree of the RT 078 strains, it was evident that the phylogenetic similarities between strains was very high, with ANI values of above 99% for each pair-wise comparison of strains towards each other. Comparisons to the two strains belonging to either RT 027 (R20291) or RT 012 (CD630) each resulted in lower % values.

Previous genetic comparisons of various *C. difficile* RTs have described tight clustering of those strains of the PCR RT 078 [130]. Genetic similarities between other *C. difficile* strains such as R20291 (RT 027) are unlikely as they share fewer genes and usually represent an own lineage cluster [130]. Further, the data suggests ANI values of 100% for CD2315/2016, CD2315/4401, CD2315/7009825 and CD2016/7009825. This does not indicate that the genomes are 100% identical. In the consequent Phylogenetic tree, all strains are represented as individual branches with proximity. The 100% ANI similarity can be explained by mechanism of ANI computation. The comparative mechanism currently fulfils the gold standard as whole genome alignment method [131]. Alignment on decimal level

for strains with high genetic similarities are challenging to present and for simplicity rounded off to whole numbers.

Overall, the strong genetic similarity between the analysed ten RT 078 strains suggested that they are likely to behave similarly in terms of competency in the conjugative transfer of plasmids from *E. coli* donors. Most strikingly, it was evident that all strains carried an identical locus encoding two sequentially arranged Type IV restriction endonucleases (REase), which in all cases was accompanied by an upstream Type II systems composed of a single, putative orphan methyltransferase (MTase). The MTase gene resided at essentially the same distance upstream of the distal Type IV REase (some 1400 bp) and the encoded enzyme shared an identical amino acid sequence between all strains. It represented a Type II N4-cytosine or N6-adenine DNA methyltransferase that probably recognized a CAAAAA motif and was not accompanied by any associated gene encoding a REase. The latter suggests that this enzyme plays no role in restriction-modification. A similar conclusion was reached concerning the large number of genome-encoded orphan methyltransferases carried by *C. difficile* CD630 [115]. In bacteria, methylation is an essential process to generate self-distinction against potentially harmful genetic material of external sources such as phages or transposable elements. In some bacteria however, DNA methylation may have desirable side effects, such as increased temperature tolerance and antibioticO resistance. Solitary MTases have been implicated in the regulation of

expression of specific gene sets in bacterial including DNA and cell proliferation as well as infection. As DNA methylation status is replicated with the chromosome, they may be transmitted via gene transfer. Thus, epigenetic functions of bacterial strains can undergo horizontal transfer. Marinus and Morris confirmed, that MTases can be linked to transmittable gene expression and regulation [132, 133]. The transfer of mobile genetic elements is limited by the R-M systems. However, in some cases R-M systems may themselves be transposable elements in selective competition against each other. These so-called selfish genes may insert into the strains genome and rearrange or amplify existing structures. The system does not just defend, but directly influences the epigenetic order of a genome. A process that further drives prokaryotic evolution [134, 135].

Orphan MTases may be beneficial to the host and indicate strain evolution. Many of the orphan MTases have been linked to genome protection in the case of intra-host R-M system competition [136]. If two restriction enzymes recognize the same target sequence, methylation by the orphan methylases protect against genome fragmentation. This highlights the selective pressure on an organism to decrease its specificity against invading DNA via the participation of orphan methylases and its continuous adaptation to competitive environments [135]. In this study, the various strains contain a multitude of orphan methylases (Figure 3.5). Interestingly, it becomes evident when comparing these methylases with their phylogeny in Figure 3.2. Strains such as *C. difficile* strains CD7009825, CD2016,

CD2315 and 1801 express a highly similar methylation pattern, including a Type II R-M system and 2 orphan methylases. The genetically less related CD4401 and CD9301 on the other hand already lack an orphan methylase. Up to this point it can be concluded, that the presence or absence of specific orphan methylases may indicate close relationships between strains.

In contrast to the Type II systems, the Type IV system appears to play a universally important role in the effectiveness of gene transfer from *E. coli* donors to RT 078 *C. difficile* strains. The encoded enzymes represent the only REases that are always present in RT 078 strains, where they cleave any incoming DNA that is appropriately methylated. This was unequivocally demonstrated by showing that the use of a Dcm defective *E. coli* donor (strain sExpress) always led to an approximately 10-fold higher frequency of conjugative gene transfer compared to when a donor (strain CA434) with a functional Dcm system was employed. This improvement in DNA transfer was obtained with all four RT 078 strains tested, CD1801, CD9301, CD9501 and M120, and would be assumed to exhibit a similar effect in all RT 078 strains. Not investigated here was the relative importance of the two REase enzymes in terms of their effects on gene transfer. It would have been interesting, for instance to make CRISPR/Cas9-mediated single and double knock-outs of the two encoding genes and then to measure the effect on DNA transfer frequency. One advantage may have been to generate a new recipient strain that would not need to rely on the use of sExpress as the donor strain to

achieve the highest rate of DNA transfer but could use CA434. This is because one disadvantage of employing methylation mutants of *E. coli*, is their tendency to favour certain mutations and plasmid instabilities. The use of Dam/Dcm strains of *E. coli*, for instance, is not recommended by NEB, although this principally relates to Dam strains and mutations [137, 138]. Nonetheless, anecdotal evidence from other members of the SBRC has noted a tendency for plasmid rearrangements when using sExpress.

While the Type IV system represented a locus encoding REases in all RT 078, the two strains M120 and CD9501 carry other REase encoding genes. In the case of the type strain M120, one of its three additional Type II systems (CdiM120ORF10140P) comprises two sequentially located methylase genes (M1 and M2) followed by two REase genes (R1 and R2). The two MTases as well as the one REase (R1) are suggested by REBASE to probably recognize the sequence ACGGC. As no other RT 078 strains in this study carry this system, it has either been evolutionary lost in these strains or recently acquired by M120. To lose an already existing R-M system, however, is unlikely. Cells develop a dependency on existing R-M systems. If an R-M system is cured from the genome REase and MTase concentration are imbalanced. This results in cell death by REase inflicted genome fragmentation of non-methylated sites, termed post-segregational killing [134, 135].

The recognition sequence of CdiM120ORF10140P, ACGGC, is the same as the restriction enzyme BceAI. This particular REase is part of a Type II system of *Bacillus cereus* strain 1315 which, like CdiM120ORF10140P, is also composed of two MTases and the single REase, BceAI. Nucleotide blast of the DNA sequence of CdiM120ORF10140P against the *B. cereus* genome revealed a 92% query coverage. Considering that both, *C. difficile* and *B. cereus* are spore forming anaerobic pathogens and both known to colonize the gut of mammals, exchange of this R-M system could be possible through the involvement of mobile elements. Transposons are known to carry R-M associated genes in other clostridia [135, 139] and Tn*6164* has been shown to be present in several human RT 078 strains, including M120, through the use of PCR primers designed to detect its presence or the absence [139, 140]. When these primer sequences (Primer Absence - CCCACCTTTATAGCATCATATAG and Primer Presence - CTAACCTATCAACTCAACCCC) were blasted against the ten RT 078 genomes derived here, only M120 carried the Tn*6164* transposon. The absence of the transposon in the nine other RT 078 strains explains why their DNA was devoid of a ^{m4}C GCCGT/ACGGC methylation pattern.

Intriguingly, transfer of plasmid pMTL84151 into strain M120 could not be demonstrated from the *E. coli* donor strain CA434, only from sExpress (Table 3.3). The possibility exists that the CdiM120ORF10140P Type II system may have been the reason that transfer of pMTL84151 from CA434 could not be detected as

pMTL84151 has nine ACGGC sites. However, other plasmids that could be transferred from this donor have a similar number of ACGGC sites (nine in pMTL82151; ten in pMTL83151 & pMTL86151, and; eight in pMTL85151). The presence of this restriction activity may, therefore, not explain failure to transfer pMTL84151. The relevance of these enzymes could be explored by generating single and double mutants of the encoding genes using RiboCas and then testing the ability of the mutants created to act as donors for the transfer of pMTL84151, as well as assessing any general improvements in competence.

Strain CD9501 is the only RT 078 strain to carry what appears to be a complete Type I R-M system including a REase gene with a predicted restriction recognition site of ATCNNNNNNCTC. Despite its presence, the frequency with which plasmids were transferred to CD9501 from *E. coli* donors appeared little different to that observed with other *C. difficile* recipient strains lacking this type I system, such as CD9301 and CD1801. However, the recognition site is comparatively rare. For instance, the genomes of CD9501 and M120 carry a single and 9 sites, respectively, while all of the plasmids conjugated into these two RT 078 strains containing a single site. The presence of the latter clearly does not prevent plasmid transfer. Indeed, previous studies, albeit based on Type II R-M systems, have shown that the conjugal transfer of plasmids into *C. difficile* was still possible after the deliberate addition of one or even two restriction endonuclease target sites to the plasmid being transferred, but was eliminated if three sites were added [73].

BLASTN of the CD9501 Type I R-M system reveals that the same system occurs in a variety of *C. difficile* strains, but also in other Gram-positive bacteria such as *Streptococcus anginosus*. All these bacteria share the mammalian gut as a common habitat [141]. Horizontal gene transfer is hypothesised to considerably influences the distribution and evolution of R-M systems. Especially the exchange of R-M systems between various species protects against external threats [142]. Thus, genomic exchange of *C. difficile* strain CD9501 and other pathogens in its environment, such as *S. anginosus* harbouring the Type I R-M system is highly likely.

As a major aim of this initial work was to choose the most 'competent' *C. difficile* donor strain for TraDIS, the logical choice was to use the strain possessing the least R-M barriers, in which effective DNA transfer had been demonstrated. Of those strains in which DNA transfer was demonstrated, CD9301 fulfilled the required criteria as it represented one of 4 strains possessing only the universal Type IV systems and a Type II system lacking any REase. Strain CD1801 carried an additional Type II system, albeit lacking an REase, whereas strains M120 and CD9501 both possessed additional REases as part of a Type II and Type I system, respectively. Should the situation arise in the future where there was a need to use either M120 or CD9501, then it would be beneficial either to use RiboCas to inactivate the encoding genes or to endow the *E. coli* donor strain with cloned

copies of the appropriate methylase gene, thereby ensuring that any recognition site in the vector being transferred was appropriately protected from digestion through *in vivo* methylation in the *E. coli* donor. The *in vivo* methylation of plasmids in the donor is a common strategy for overcoming R-M barriers [68, 117]. In the case of Type I systems, it requires that both the HsdM- and HsdS-encoding genes are cloned and expressed in the donor [143]. Protecting the plasmid in the donor prior to transfer may be preferred over the KO of genes in the recipient, as the presence of restriction site in Cas9 may make it impossible to use RiboCas (for example, Cas9 carries 3 ATCNNNNNNCTC motifs), although the use of other systems such as ClosTron would be an option.

The study also identified the preferred *E. coli* donor as being the strain sExpress, as its use overcame the negative effects of the Type IV system in the selected *C. difficile* recipient strain CD9301, by ensuring the plasmid was not methylated by the inactivated donor, Dcm system.

Chapter 4

Developing TraDIS in *C. difficile* PCR RT 078

4.1 Introduction

4.1.1 Use of Transposons in *Clostridium.*

Transposon represents an important tool in clostridial forward genetics. They are used to generate mutant libraries composed of clones containing random transposon insertions. These libraries may be screened to identify desired phenotypes under specific environmental conditions and thereafter the genotype of selected mutants determined. In this way it is possible to discover genes that were previously unknown to be involved in a particular phenotype. Thus, it was the use of transposons that led to the discovery in 2009 by Videl and co-workers of the involvement of the quorum sensing genes *agrB* and *agrD* in the regulation of toxins in *C. perfringens* [144].

In the meantime several transposon systems have been developed for deployment in Gram-positive bacteria [145-149], including two further systems for *C. perfringens* [144, 150]. A promising approach in *C. perfringens* was based on a bacteriophage Mu-based transposon delivery system. Random mutant libraries with single transposon insertions were obtained via electroporation of the Mu-transposome complex into *C. perfringens.* While this phage derived Mu system

resulted in single insertion events, the transformation frequency obtained was inefficient resulting in strong insertion site bias that led to approximately 43% of total insertions in rRNA operons and 12% in intergenic regions [151]. The use of conjugative transposons based on Tn*916* and Tn*5397* has been investigated in *C. difficile*, but proved unsuitable for library production, as they exhibited a strong target site preference and were disadvantaged by the fact that clones generated frequently contained multiple insertions [103, 152]. Eventually the *Himar1 mariner* transposon system was successfully implemented for library production in *C. perfringens* where it was delivered on a self-replicating plasmid [153]. In many other *Clostridium* spp., such as in *C. difficile* [91], *C. acetobutylicum* and *C. sporogenes* [154], the *mariner* based system has been successfully implemented.

It is considered suitable as only the cognate *Himar1* transposase is required for transposition. Also, the transposable element of this system randomly inserts itself into an AT rich target side, minimizing insertion preferences into the AT-rich *Clostridium* genomes. In 2015, Dembek et al., performed the first Transposon Insertion Sequencing (TIS) experiment in *C. difficile* strains 630Δ*erm* and R20291. They identified 404 essential genes and 798 genes implicated in spore production [86].

4.1.2 Transposon Delivery Strategies

To obtain sufficiently large transposon mutant libraries, effective and efficient delivery of the transposon into the target organism's genome is crucial. Furthermore, the transposon-delivery should be a one-off event, taking place only once in each individual cell. Multiple insertions in the same clone make it impossible to draw conclusions as to the phenotypic changes responsible for any observed phenotype.

This makes transposon-transposase complex (transposome) or non-replicating plasmids a favoured tool in forward genetic studies. The advantages of non-replicating systems are that the transposome cannot be passed to descendant generations. The resulting lack of a plasmid encoded transposase gene prevents later movement of the inserted element, ensuring mutant stability. Further, the lack of plasmid backbone provided markers, indicates that all colonies observed on a transformation-selection plate can be identified as transposon mutants directly, without any further PCR analysis. Based on the colony count per plate, the total number of mutants in the library can be assessed. During the follow up library screening, the non-replicative system cannot be detected, as it cannot be retained by the organism, resulting in improved library interpretation.

4.1.3 Suicide vectors

A requirement for the use of suicide vector transposon-delivery is a high transformation rate. In practice this limits the technique to a relatively small subset of, typically more well characterised, species.

Transposome complexes have been integrated into a variety of Gram-positive and Gram-negative bacterial strains. Even though the Tn*5*-based transposon randomly inserts into the host's genome and creates knockouts, system delivery into the target host is challenging. Electroporation remains the technique of choice to transform the Transposome complex into various bacterial strains, allowing subsequent integration of transposon DNA. The efficiency of transposition varies between different techniques and species. Electroporation of Tn*4351* transposome into *Bacillus* spp., however, has been shown to be relatively inefficient and electroporation of Mu-based transposome complexes into *C. perfringens* needed to be further optimized [151, 155].

A major limitation in *C. difficile* research is that DNA transfer into the organism essentially relies on relatively inefficient conjugative plasmid transfer. Kirk et al., reported conjugation efficiencies could be improved by subjecting recipient *C. difficile* cells to a heat shock prior to mixing with donor *E. coli* cells, an outcome most likely due to heat inactivation of *C. difficile* R-M systems [77].

Further, Woods et al., developed a novel conjugative donor strain based largely on *E-coli* K-12, increasing the conjugation efficiency in *C. difficile* R20291 by avoiding the type-IV restriction barriers. Recently, a protocol for successful *C. difficile* electroporation has been reported. The transformation protocol achieves between 20 and 200 colonies per microgram of DNA amongst transformation of *C. difficile* strains R20291, CD630 and JSC10. The exact usability of this novel method, however, depends on the post-electroporation recovery time of the *C. difficile* recipients and has been suggested to be analysed in each lab individually [75].

Accordingly, transformation of DNA into important strains, such as 027 and 078 RTs of *C. difficile* remains challenging. Currently, low conjugation efficiencies hamper suicide vector mediated transposition, while suitable electroporation protocols have to be further researched.

4.1.4 Self-replicating plasmids

Low transposon vector transformation efficiencies are a major bottleneck during library production. Introducing self-replicative plasmid systems presents an alternative to suicide systems. These vectors are equipped with a suitable Gram-positive replicon and thus rapidly lead to large populations of plasmid harbouring cells. Suitable Gram-positive replicons for clostridia are obtained from plasmids

such as pBP1 (*C. botulinum*), pCB102 (*Clostridium butyricum*), pCD6 (*C. difficile*) and pIM13 (*Bacillus*). The utility of a replicon is dependent on its level of stability within the host. To determine the most suitable replicon, the vectors are transferred to the target clostridial strain and their segregational stability is assessed. This is achieved via direct measurement of bacterial growth on antibiotics or a full plasmid loss assay. The efficiency of the same replicon, however, may vary between different strains of the same species. Further, it remains challenging to obtain a transposon library using a replicon-based tool, as lingering plasmids may interfere with mutant stability and library analysis during downstream experiments.

To overcome this problem, Cartman and Minton successfully developed a pseudo-suicide transposon delivery system utilizing the segregationally unstable pBP1 replicon [91]. This method resulted in a transposition frequency of 4.5×10^4 per cell. However, the lack of effective control of transposition time and plasmid loss limited the library size. Furthermore, plasmids harbouring pBP1 may not be applicable in all *C. difficile* strains, as they have been described to be most stable in *C. difficile* strain 630 whilst being the most unstable plasmids in R20291 [65].

4.1.5 Inducible promotor systems

Although conditionally replicating plasmids can theoretically be used to control whether a particular gene or gene set is active within a cell, an alternative strategy would be to place the desired gene(s) under the control of an inducible promoter. This type of promoter remains inactive when in the uninduced state (tight repression), but upon induction is strongly expressed. This allows specific genes to be expressed only under defined conditions at the required time point. Broadly speaking, inducible promotor systems are of two types, based either on chemical or physiological inducers. The former are reliant on chemical compounds such as hormones, antibiotics or sugars while the latter react to external stimuli such as temperature, light, or osmotic stress. Choosing an optimal promoter is dependent on the utility of the inducer in the organism of choice. The toxicity of the inductive or repressive substance for the host should be taken into consideration. Furthermore, tight promoter activation is crucial, as leaky promotor systems may result in uncontrollable gene expression.

4.1.6 An Orthogonal Promotor System based on TcdR - P_{tcdB}

In addition, to optimally responding to conditions observed by the cell, inducible gene expression should be independent of the host's native regulatory control.

This is particularly important in systems that shuttle between hosts, as is the case with *E. coli*:*Clostridium* shuttle vectors, where it is desirable that the promoter system used has minimal activity in *E. coli* but becomes fully active in the clostridial target cell when supplied with the appropriate inducer. In *C. difficile*, such an orthogonal promotor system has been designed based on the promotors P_{tcdA} and P_{tcdB} of the two toxin genes, *tcdA* and *tcdB*. These promoters are specifically recognised and targeted by the alternate *C. difficile* sigma factor TcdR and are only poorly recognized by the sigma factors of the *E. coli* shuttle host [65].

This method has been developed successfully for transposon mutagenesis in *C. difficile* [91] in which P_{tcdB} is active in the presence of the alternative sigma factor TcdR [156]. Further, the *tcdR* gene was inserted in other *Clostridium* species such as *C. autoethanogenum* or *C. sporogenes* to improve orthologous expression of the transposase gene [154]. TcdR has no homolog in most clostridial species, making it highly strain specific and an optimal choice of application for transposon mutagenesis.

4.1.7 Tc inducible promotor

In *Clostridium*, several chemically inducible promotor systems have been developed to simultaneously control *Himar1C9* expression and plasmid loss. In *C. sporogenes,* an IPTG inducible promotor has been placed upstream of the pCB102 replicon. The plasmid (pMTL-YZ14) replicated normally in the absence of IPTG. Adding IPTG to the culture media, however, resulted in 100% plasmid loss. Further, pMTL-YZ14 achieved up to 80% plasmid loss in *C. acetobutylicum* [154]. This conditional vector has been shown to also be efficient in *C. beijerinckii*, *C. botulinum* and *C. autoethanogenum* [65].

Tests of the IPTG inducible system in *C. difficile*, however, remain challenging, as previous results from our group suggest that IPTG is not taken up by *C. difficile* (N. Minton, personal communication). Thus, a substitute in this particular clostridial host has to be found. As potential alternative to IPTG, Hartmann et al., described a lactose inducible system, successfully applied in *C. perfringens* [150].

In the early 1990s, Geissendorfer described a Tc inducible promotor in *E. coli* resulting in inducible, high-level expression in *B. subtilis* [93]. This system has been further developed and successfully implemented in a Tc-inducible expression vector for *S. aureus* [157]. Subsequently, in 2011, Fagan et al., introduced and optimized the Tc-inducible promoter (*Ptet*) for controlled protein expression in *C.*

difficile [94]. The novel transposon vector features Tc-dependent conditional replication into the *C. difficile* pCD6 replicon and tight control over the transposition of an *ermB*-based transposon. Two Tc inducible promotors are brought in line and thus drive expression of *tetR* and simultaneously express the codon optimized Himar1 transposase. Dembek et al., achieved nearly complete induced plasmid loss after 13 generation of growth in culture, while 40% non-induced bacteria retained the plasmid in *C. difficile* 630*Δerm* and strain R20291. Subsequently, transposon libraries were grown *in vitro* and during sporulation. Additional TraDIS analysis of these samples resulted in the identification of 404 and 798 essential genes for vegetative growth and sporulation, respectively [86].

In this study we aim to build a replicative approach in the clinical *C. difficile* isolate RT 078 strain CD9301 to establish which system may be the most optimal for library production.

4.2 Aims of the study

The overall aim of this Chapter is to evaluate a suitable system for the implementation of a TraDIS library in *C. difficile* RT 078. For this, various plasmid delivery systems are integrated in *C. difficile* strain CD9301 and evaluated based on their ability to achieve random transposon insertions into the genome and subsequent reliable plasmid clearing from the cell.

In short, we aimed to evaluate these three research questions:

- Which inducible systems can be used for transposase activation and plasmid loss in *C. difficile?*
- How do the Tc inducible plasmid delivery systems designed for *Clostridium* species developed in our lab perform in RT 078?
- How well does the Tc inducible transposon delivery system developed in Prof. Robert Fagan's laboratory [86] perform in RT 078?

4.3 Results and Discussion

4.3.1 Tc inducible plasmids pMTL-CW21 and pMTL-CW22

The aim of this chapter is to identify the most suitable transposon-delivery vector for the implementation of TraDIS in a *C. difficile* RT 078 strain. It is of importance that the vector of choice creates a mutant library of sufficient size and diversity and in which there is minimal plasmid retention. Central to this aim is the ability to force inducible plasmid loss, thereby selecting for true, genomic transposon mutants and against cells which harbour the transposon-delivery plasmid.

Fagan and colleagues introduced a divergent promotor system on their transposon plasmid that controlled transposase activity and plasmid loss in strain 630Δ*erm* and R20291 (Figure 4.1) [86]. To evaluate the efficiency of this system in other *Clostridium* species, a variety of transposon vectors utilizing the divergent promotor approach have been produced in our laboratory. The initial vectors utilized for production of a transposon mutant library in this study were created by Dr. Craig Woods. The two vectors, pMTL-CW21 and pMTL-CW22 have been tested in a variety of *Clostridium* species in our laboratory. Both plasmids were eventually used to produce large mutant libraries in *C. sporogenes* (pMTL-CW21) and *C. autoethanogenum* (pMTL-CW22). The two vectors are very similar. While

$P_{xylA/tetO}$ is driving the expression of the *mariner-Himar1C9* transposase gene, $P_{tetR/tetO}$ is driving the expression of *tetR* and reading into the respective Gram-positive pCB102 replicon (pMTL-CW21) (Figure 4.2) or that of pCD6 (pMTL-CW22) (Figure 4.3). These Nottingham-Tet constructs are derived from an expression vector, designed for *S. aureus [157].* The Tet-system has been mirrored to ensure divergent expression and was codon optimized by Dave Walker (Figure 4.1) (David Walker, PhD thesis).

Both plasmids carry an Em resistance gene (*ermB*) and contain the ColE1 Gram negative replicon. To prevent untimely transposition events in *E. coli* donor or cloning strains, the *catP* transposon is flanked by the *codA* gene on one side and the P_{thl} promoter on the other side. When the transposon is present in between these two elements, a *fdx* terminator that forms an integral part of the transposon prevents transcription from P_{thl} into the *codA* gene. In case of an undesired, early transposition event in *E. coli,* however, the transposon is excised leading to transcription of the *codA* and the production of the encoded cytosine deaminase. This enzyme is cytotoxic in the presence of 5-fluorocytosine (5-FC) leading to cell death. The mini-transposon is further flanked by ITRs. These ITRs are themselves flanked by I-*Sce*I recognition sequences. As the sites reside outside of the ITRs, the transposed DNA inserted into the *C. difficile* genome does not contain I-*Sce*I recognition sequences. This allows additional plasmid curing by I-*Sce*I targeted digestion of any remaining vector evading plasmid loss.

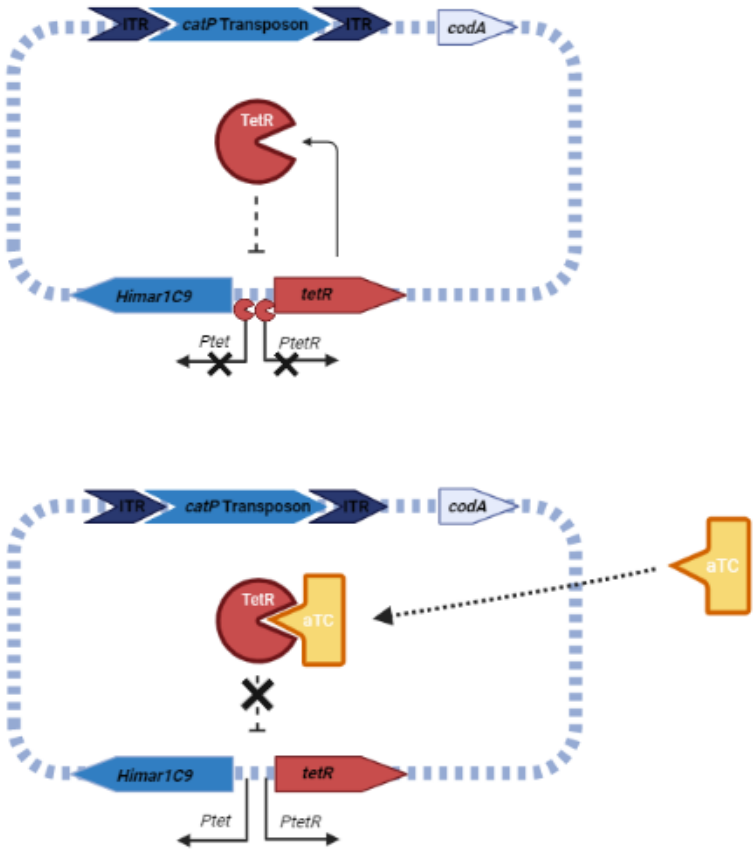


Figure 4.1: Tc inducible transposon-delivery plasmids.

A genetic schematic of the Tc-inducible transposon-delivery plasmids. The *himar1C9* transposase is controlled by the Tc-inducible promoter P_{tet} which is inhibited by the binding of TetR. TetR is expressed from the divergent promoter P_{tetR} which is also inhibited by TetR. In the presence of Tc or the analogue anhydrotetracycline (aTc) TetR is released from promoter binding and both P_{tet} and P_{tetR} are de-repressed. The addition of aTc therefore causes transposase expression and an increased expression of the TetR protein. Importantly, transcriptional read-through from P_{tetR} also causes plasmid instability via transcription into the Gram-positive origin of replication. The Gram-positive origin of replication is positioned downstream of *tetR* and is the only source of variation between pMTL-CW21 (pCB102), pMTL-CW22 (pCD6). The Pthl-Tn-*codA* module which confers plasmid stability in *E. coli* donor and storage strains is also present on the Tc-inducible transposon-delivery plasmids as well as I-*Sce*I recognition sites flanking the transposon.

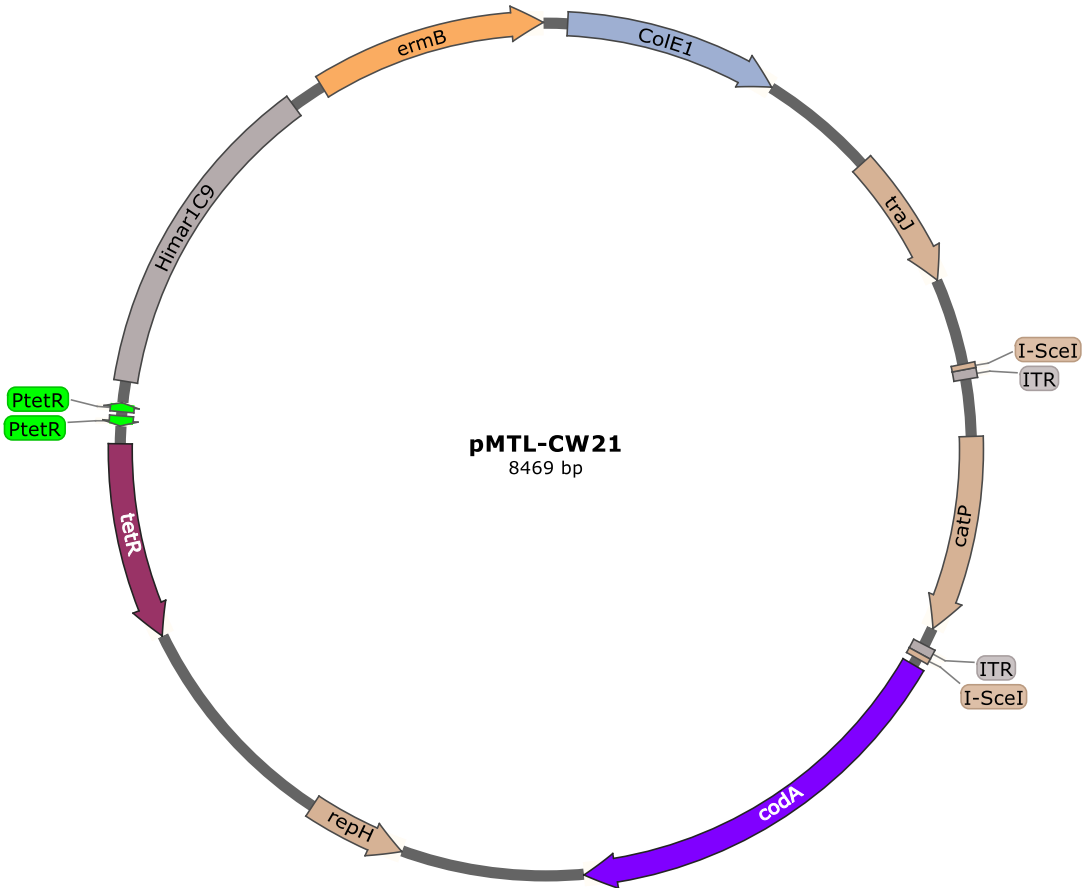


Figure 4.2: pMTL-CW21.

The divergent P_{tetR} (Figure 4.1) promoters transcribe the *Himar1C9* transposase gene and disrupt the function of the adjacent pCB102 replicon region (RepH). Cut and paste removal of the mini-transposon containing *catP*, through the action of the transposase on the ITR repeat sequences, leads to expression of *codA* (encodes cytosine deaminase) from an upstream P_{thl} promoter, which is lethal to the host in the presence of 5FC. The ITR sequences are flanked by I-*SceI* recognition sequences.

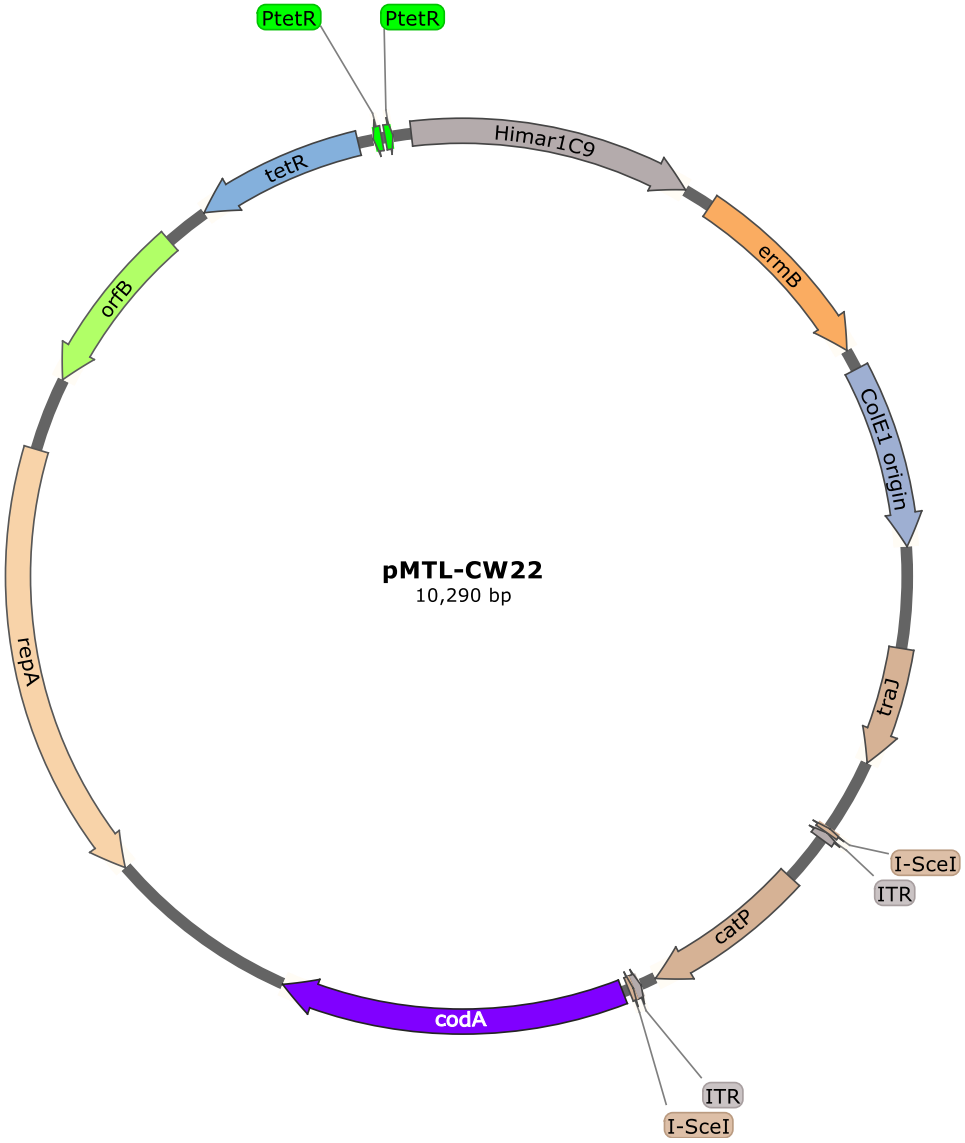


Figure 4.3: pMTL-CW22.

Components are identical to that of plasmid pMTL-CW21, except the the pCB102 replicon is replaced with that of pCD6, comprising *orfB* and *RepA*.

4.3.2 Assessment of pMTL-CW21 and pMTL-CW22

Transposon-delivery vectors were sourced from Dr. Craig Woods and can be found in the SBRC Nottingham culture collection. The plasmids, pMTL-CW21 (Figure 4.2), pMTL-CW22 (Figure 4.3), were transformed into *E. coli* conjugal donor strain sExpress and conjugated into *C. difficile* strain CD9301 with initial plasmid selection using cefoxitin, D-cycloserine (CC) and Em. Em selects for the marker on the plasmid backbone whilst cefoxitin and D-cycloserine select against a variety of Gram-positives and against the *E. coli* donor, respectively. All colonies from plasmid-selection plates were harvested and resuspended in PBS before being serially diluted and plated onto agar plates supplemented appropriately so as to quantify various aspects of the plasmid features. The features of interest are firstly plasmid loss efficiency, which determines to what extent adding the inducing compound reduces the number of plasmid-harbouring cells. This is calculated by comparing the CFU of plasmid backbone with and without the inducer compound. Plasmid loss efficiency is crucial to effective transposon-delivery plasmids.

$$Plasmid\ loss = \frac{\mathrm{CFU/mL(Em)} - \mathrm{CFU/mL\ (Em + aTc)}}{\mathrm{CFU/mL\ (Em)}} \times 100\%$$

A second feature of these plasmids is that of transposition frequency, which is an estimation of the activity of the transposase. This value will have repercussions for the ease of library creation. Transposition frequency is calculated by first estimating the number of transposon mutants in the sample. These mutants

should correspond to the CFU on plates selecting for the transposon-located antibiotic resistance marker. In case of pMTL-CW21 and CW22 the resistant marker is Tm. Further, these colonies help to indicate non-permissive condition for plasmid retention which is the presence of the inducer compound (aTc). Transposon mutants appear on these plates, but so too will any cells that have not lost the plasmid and so the plasmid-loss efficiency results must be taken into account here too. The transposition frequency is expressed as a proportion of plasmid-containing cells which is estimated by CFU counts of the plasmid backbone selective plates (containing Em).

$$Transposition\ frequency = \frac{\mathrm{CFU/mL(Tm + aTc)} - \mathrm{CFU/mL\ (Em + aTc)}}{\mathrm{CFU/mL\ (Em)}}$$

Table 4.1: Assessment of pMTL-CW21/22 *C. difficile* RT 078 strains CD9301.

aTc (100ng/mL) was used to induce transposition and plasmid-loss simultaneously. Both plasmids were transferred to *C. difficile* via conjugal transfer with *E. coli* sExpress donor strain. The transconjugant selection plate was flooded with PBS and the entire cell mass harvested. The transconjugant cell mixture was serially diluted down to 10^{-7} and plated to a) Em, b) Em, aTc, c) Tm aTc, d) Tm, and the respective CFU counts per mL were calculated. CFU counts were used to calculate transposition frequency ((Tm, aTc – Em, aTc) ÷ Tm; Plasmid loss efficiency (Em - Em, aTc) ÷ Em × 100; Data is the average of 3 biological replicates.

Plasmid	Biological Repeat	Backbone selection (Erm) CFU/ml	Non-permissive condition (aTc, Erm) CFU/ml	Transposon selection (aTc, Tm) CFU/ml	Plasmid Loss Efficiency (%)	Transposition Frequency
pMTL-CW21	1	1×10^{-7}	4×10^{-5}	1×10^{-6}	**96**	**0,06**
	2	15×10^{-6}	10×10^{-6}	1×10^{-6}	**33,3**	**0,6**
	3	2×10^{-7}	3×10^{-5}	1×10^{-5}	**98,5**	**-0,1**
pMTL-CW22	1	1×10^{-6}	6×10^{-6}	13×10^{-1}	**-500**	**0,16**
	2	6×10^{-5}	2×10^{-5}	1×10^{-5}	**66,66**	**3**
	3	17×10^{-6}	16×10^{-6}	19×10^{-6}	**5,88**	**1,06**

Table 4.2: Summary of Plasmid loss and transposition frequency achieved with transposon vector pMTLCW21 and pMTL-CW22 in *C. difficile* strain 9301. The standard deviation (SD) of the % plasmid loss and the transposon frequency achieved in strain 9301 using pMTL-CW21 and pMTL-CW22 was determined by calculating $\sigma=\sqrt{1N}\sum Ni=1(Xi-\mu)2$ $\sigma = 1 N \sum i = 1 N (X i - \mu) 2$. Where x_i is an individual value, μ is the mean/expected value and N is the total number of values.

Plasmid properties in strain 9301	Plasmid loss (%)	SD of Plasmid loss	Transposition frequency	SD of transposon frequency
pMTL-CW21	75.96	17.4	0.18	0.18
pMTL-CW22	-142.48	146.65	1.40	0.68

The data presented in Table 4.1 and summarized in Table 4.2 suggested that neither of the two vectors assessed were suitable for transposon delivery in *C. difficile*. Plasmid pMTL-CW21 (based on the pCB102 replicon) gave only low numbers of transposons in *C. difficile* strain CD9301 and exhibited a low frequency of plasmid loss following induction with aTc. Plasmid pMTL-CW22, in which the replicon had been changed to that of pCD6, gave even poorer rates of transposon mutant generation and plasmid loss. In fact, plating of cells carrying pMTL-CW22 on Em agar under non-permissive conditions (presence of aTc) resulted in a higher CFU than when the aTc inducer was absent. This indicates that the transformant colonies gained a fitness advantage under the influence of inducer. It is estimated that the pCD6 replicon enhances stability in *C. difficile* following induction with aTc, contrary to pCB102 in pMTL-CW21. However, even the latter plasmid exhibited sub-optimal plasmid curing and transposon frequencies. It was concluded that the use of the aTc-inducible P_{tet} promoter system may not be ideal for producing the required large transposon mutant library. This assumption is supported by the high deviations of σ, which indicate a suboptimal functioning in *C. difficle* strain 9301.

4.3.3 pMTL-CW17 A Lactose-inducible Transposon-delivery Plasmid

In 2014, Zhang et al., established a vector (pMTL-YZ14) in which transposase expression was under the control of an alternate sigma factor TcdR. Transcriptional readthrough into Gram-positive pCB102 replication region [154] was mediated by an upstream P_{fac} promoter which destabilised the replicon resulting in plasmid loss. The presence of a *lac* operator in P_{fac} meant that its transcription could be induced through the use of either the native inducer or its analog, isopropyl ß-D-1 thiogalactopyranoside (IPTG). Elsewhere [148,157], the native, lactose-inducible *bgaR*-P_{bgaL} system of *C. perfringens* has been exploited, where BgaR is a positive transcriptional regulator of *bgaL* (encodes ß-galactosidase).

As lactose-based inducible systems have recently been shown to be functional in *C. difficile* (Zuzana Grmanova, PhD thesis unpublished), a lactose inducible transposon vector pMTL-CW17 (Craig Woods) based on the use of *bgaR*-P_{bgaL} was investigated. It is similar to pMTL-CW21 and pMTL-CW22, except that the *himar1C9* transposase gene was placed under the control of *bgaR*-P_{bgaL}. Its plasmid-loss system remains reliant on conditional transcriptional read-through of the IPTG inducible Pfac promotor into pCB102 (Figure 4.4).

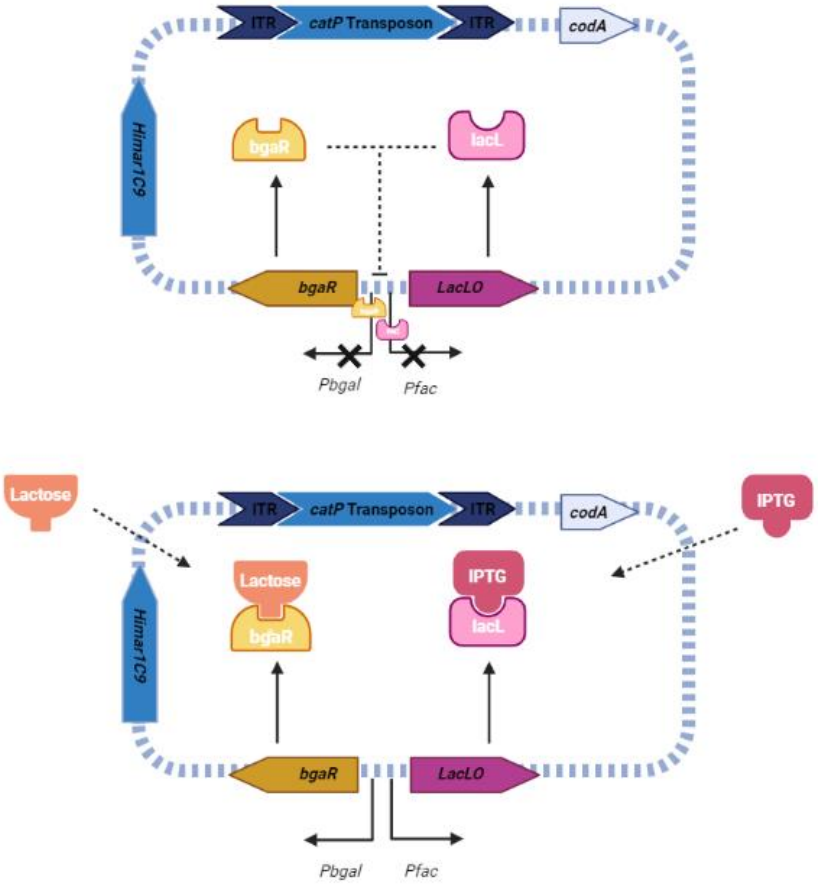


Figure 4.4: Illustration of the lactose and IPTG inducible transposon-delivery system. A genetic schematic of the lactose and IPTG -inducible transposon-delivery plasmids. The *himar1C9* transposase is controlled by the lactose-inducible *BgaR*-P_{bgal} promoter system. Lactose binds to BgalR and de-represses P_{bgal}. The addition of lactose therefore causes transposase expression. Further, transcriptional read-through from the P_{fac} promotor into the Gram-positive origin of replication allows plasmid loss. Transcription from P_{fac} is repressed by LacI which binds to the LacL-operator, blocking RNA Polymerase. The interaction of the substrate IPTG with LacI leads to causes conformational change of the protein. The occupied protein is now unable to bind the lac operator. This release allows transcriptional readthrough into the downstream pCB102 replicon. The P_{thl}-Tn-*codA* module which confers plasmid stability in *E. coli* donor and storage strains is also present on the lactose and IPTG transposon-delivery plasmids as well as I-*Sce*I recognition sites flanking the transposon.

Like pMTL-CW21 (Figure 4.3) and pMTL-CW22 (Figure 4.2), I-*Sce*I sites flank the vector *catP* gene. Hartman et al. have previously suggested that IPTG is not an inducer of the P_{bgaL} promoter in *C. perfringens* [150], however, an analysis of pMTL-CW17 (Figure 4.5) in *C. autoethanogenum* under various conditions revealed similar transposon efficiency to the orthologous P_{tcdB} promotor routinely used, suggesting IPTG inducible plasmid loss functioned in this *Clostridium* species (Nigel Minton, personally communicated). This study intends to analyse the potential of a lactose-IPTG inducible plasmid system *bgaR*-P_{bgaL}/P_{fac} in *C. difficile* 078 strains to produce sufficient ratios of unique transposon mutants for follow up library production, examining various lactose and IPTG concentrations.

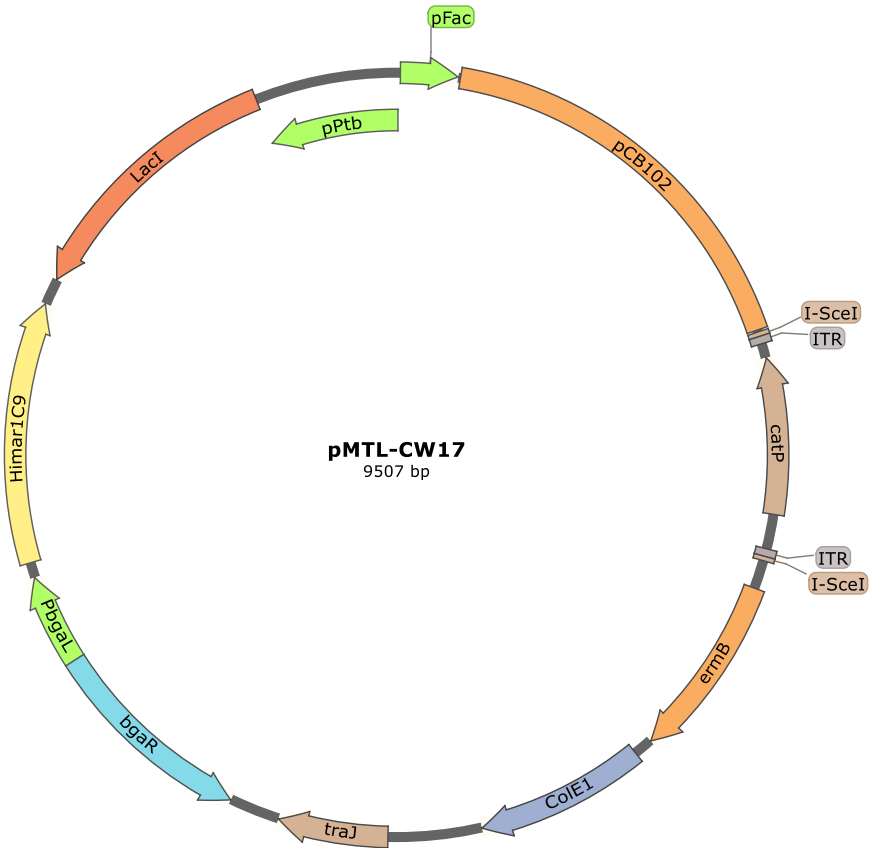


Figure 4.5: pMTL-CW17.

A plasmid map of pMTL-CW17. It exhibits lactose-induced expression of *himar1C9* transposase via *bgaR*-P_{bgaL}. Plasmid instability is triggered via transcriptional read through of the IPTG induced P_{fac} promotor into the Gram-positive replicon, pCB102. The P_{thl}-Transposon-*codA* module is present for increased plasmid stability in *E. coli* storage and donor strains, while the *catP* flanking I-*Sce*I sites facilitates removal of plasmid reads from the subsequent high throughput sequencing.

4.3.4 Assessment of pMTL-CW17.

Like pMTL-CW21 and 22, this vector was sourced from Dr. Craig Woods and can be found in the SBRC Nottingham culture collection. Utilizing sExpress as conjugational donor, pMTL-CW17 was conjugated into *C. difficile* strain 630Δ*erm* and CD9301. Three biological replicate colonies that grew on CC and Em were resuspended in PBS before being serially diluted and plated onto supplemented agar plates. The resulting numbers of colonies grown per plate visualize the efficiency of pMTL-CW17 to induce transposition and plasmid loss in *C. difficile* strain 630Δ*erm* and CD9301. The analysis of the vector pMTL-CW17 indicates that it is not a suitable transposon plasmid to introduce TraDIS in neither *C. difficile* strain 630Δ*erm* nor CD9301. Thus, there was little evidence that plasmid loss was occurring as was evident from there being little difference in the number of plasmid-containing cells, as estimated by CFU on plates supplemented with Erm, irrespective of whether IPTG was present or not. In the case of strain 630Δ*erm*, IPTG did cause some reduction in the CFU on Em plates compared to those cells not exposed to the inducer but the difference was marginal (Table 4.3). In contrast, the results with CD9301 were the opposite to what was expected, and CFUs on Em supplemented plates were higher with than without IPTG. As it was suggested by Zhang et al., [154] it appears that IPTG is an inducer and thus can not be taken up or cannot be metabolized by *C. difficile*.

Interestingly, during initial exploitation of the pMTL-YZ14 system in *C. autoethanogenum,* evidence was obtained that unwanted transposition events were taking place in *E. coli*, including insertions into the pCB102 replicon likely to affect the plasmid's replicative ability in the target host, *C. autoethanogenum* [154]. As pMTL-CW17 is based on pMTL-YZ14, transposition events in to pMTL-CW17 in *E. coli,* such as in the promotor units are considered possible. These plasmid alterations are a potential cause to influence plasmid loss or transposition frequency in the final *C. difficile* target cells.

In contrast to insufficient plasmid loss, pMTL-CW17 appeared to perform well in terms of mutant generation. The colony counts observed on the mutant selection plates exceed the counts of the non-lactose induced transposon control plates. This supports the view lactose-induction of the Himar1C9 transposase resulted in a stimulation of transposition and the generation of increased numbers of transposon mutants. However, the increased transposition frequency in neither 630Δ*erm* nor CD9301 indicate that the purity of the mutant library is insufficient and biased by resistant colonies due to remaining pMTL-CW17.

Overall, the low values in plasmid loss due to the problematic activation of IPTG driven plasmid loss via the P_{fac} promotor make pMTL-CW17 an unsuitable system for TraDIS in *C. difficile*.

Table 4.3: Assessment of pMTL-CW17. This table represents the effectivity of pMTL-CW17 in *C. difficile* strain 630Δ*erm* and 078 RT strain 9301. Isopropyl β-D-1-thiogalactopyranoside (IPTG) (10mM) was used to trigger plasmid loss via transcriptional read through the P_{fac} promotor into pCB102. Simultaneously, the *Himar1C9* transposase is activated by the (10 mM) lactose inducible *bgaR*-P_{bgaL}. The plasmid was introduced into both strains by the conjugational donor strain sExpress. After 8 hours of conjugation, the colonies were harvested in PBS and serially diluted down to 10^{-7} prior to plating. Respective CFU counts per mL were calculated. CFU counts were used to calculated transposition frequency ((Tm, lactose + IPTG – Em, IPTG) ÷ Tm; Plasmid loss efficiency (Em - Erm, IPTG) ÷Em × 100; Plasmid retention in presence of aTc (Em ÷ Em, IPTG) and Mutant selection plate purity which is the calculated percentage of colonies of mutant selection plates (Tm, lactose + IPTG) expected to represent transposon mutants (Tm, lactose + IPTG - Em, IPTG) ÷ Tm, lactose + IPTG.

C. difficile	Transposition frequency	Plasmid loss (%)	Plasmid retention with of IPTG
CD630*Δerm*			
Ø	10.8	33.33	1.66
Standard Deviation	5.8	31.50	0.21
CD9301			
Ø	3.96	10.0	1.64
Standard Deviation	15.66	24.33	0.34

4.3.5 An alternative approach - staggered transposon delivery systems

Since plasmid loss was extremely poor in all previously tested constructs, an alternative approach was attempted which does not rely on having a conditionally replicative plasmid vector, based on pMTL-CW21. This system separates the transposase and the transposon onto two separate plasmids. Initially a first plasmid (pMTL-pDIG1) is conjugated into *C. difficile*. This plasmid contains the Himar1C9 transposase under the control of a Tc inducible promoter. As it comprises the Gram-positive replicon pCB102, it will be maintained in *C. difficile*. Successful pMTL-pDIG01 transconjugants can be selected based on their plasmid derived Em resistance. These mutants are then transformed with the suicide plasmid pMTL-pDIG02 that carries a *catP* transposon. If both plasmids are simultaneously present *in C. difficile*, Tc induction of the transposase encoded in pMTL-pDIG01 triggers transposition of the *catP* transposon from pMTL-pDIG02 into the *C. difficile* genome. The key feature of this strategy is that plasmid pMTL-pDIG02, containing the transposon, does not have a functional Gram-positive replicon and therefore cannot persist through periods of growth. This means that when the transposon-located antibiotic resistance marker is selected for only chromosomally located transposon mutants will be able to proliferate.

4.3.6 Assessment of the staggered transposon-delivery system.

Transposon-delivery vectors were sourced from the lab intern culture collection (SBRC, Nottingham). Each can be found in the SBRC Nottingham culture collection using details from the materials and methods section. Both plasmids were transformed into separate *E. coli* conjugal donor strains. Initially, *E. coli* strains containing the transposase plasmid, pMTL-pDIG01 (Figure 4.6A), were conjugated into *C. difficile* 630 Δ*erm.* The medium was supplemented with antibiotics selecting against *E. coli* donor strains to select for *C. difficile* and against Gram-negative *E. coli.* Additional supplementation with Em selects for the marker on the plasmid backbone. Colonies were harvested and grown overnight in BHIS, supplemented with Em. Growing cultures were washed in sterile PBS twice and conjugated with *E. coli* donor strains, harbouring pMTL-pDIG02 (Figure 4.6B). Growth medium was supplemented with CC and Tm.

Figure 4.6: The staggered transposon delivery system pDG01 and pMTL-pDIG02. The transposon delivery system pMTL-pDIG01 (A) and pMTL-pDIG02(B) is derived from pMTL-CW21. The particular sequences, necessary for transposon delivery are split bewteen two plasmids. During a first conjugation, the transposase is introduced into the cell on a replicative plasmid (pMTL-pDIG01). Em on its backbone serves as selective marker for the construct. In a second conjugation the transposon plasmid, pMTL-pDIG02, is transferred into *C. difficile* strains containing pMTL-DIG01. This plasmid does not include a Gram-positive origin of replication. Transposon mutants are selected for on Tm.

While CC selects against the *E. coli*, Tm selects for transposon mutants. Since pMTL-pDIG02 is not equipped with a Gram-positive origin of replication, colonies that are Tm resistance are a consequence of successful transposon insertion. Colonies, which were growing on BHIS supplemented with Tm liquid culture, were serially diluted to a factor of 10^{-7} in PBS and plated on BHIS agar plates. These plates were supplemented with Tm and Tm including aTc to test for successful transposition. The cell dilutions were additionally plated on Em and Em containing aTc to evaluate plasmid retention of pMTL-pDIG01. Subsequently, colony formation on Tm with aTc indicates successful transposition. Since pMTL-pDIG02 does not contain a Gram-positive replicon Tm resistance is limited to the successful insertion of the *catP*-encoding transposon.

Table 4.4: Assessment of the staggered transposon system pMTL-pDIG01/ pMTL-pDIG02. *C. difficile* strain CD9301 and CD630Δ*erm* were transformed with pMTL-pDIG01 via conjugational transfer with sExpress. After 72 hours, Em positive colonies were selected and grown in supplemented BHIS medium over-night, prior to transformation with pMTL-pDIG02. The cells were harvested after 24 hours of conjugation in 1 mL of PBS. The cell mixture was diluted to 10^{-7} incubated on BHIS agar plates containing Em, Em + aTc, Tm, Tm + aTc and non-supplemented BHIS as control. The plasmids properties to form suitable transposon mutants were calculated as previous. Data is from three biological replicates.

C. difficile strain	Transposition frequency	Plasmid loss (%)	Resistance to plasmid loss	Mutant selection purity
CD630*Δerm*	-3.05	-212.50	0.31	-4618.74
CD9301	-160	0.38	-2.6	n.a.

From the data indicated in Table 4.4 it was immediately apparently pMTL-pDIG01 appeared to be entirely stable in both host strains, and loss of the plasmid, as measured by Em resistance CFU, did not occur in the presence of aTc. Rather, an increase in CFU was evident. Loss of transposase from the cell is crucial if any transposed transposons are to remain stably inserted in their new sites and not undergo re-transposition. Moreover, colony counts indicate that to the contrary, the presence of aTc causes higher ratios of plasmid retention in both samples, plated solely on Em. More troublesome was the lack of Tm resistant transformants that were observed following transfer of pMTL-pDIG02 into CD9301 cells carrying pMTL-pDIG01. The lack of colonies on Tm either with or without aTc shows that no *catP* induced antibiotic resistance was stably transferred into the

CD9301 cells harbouring pMTL-pDIG02. In contrast, in the absence of aTc, 39 Tm resistant transconjugant colonies were obtained with CD630Δ*erm* carrying pMTL-pDIG01 as the recipient and *E.coli* carrying pMTL-pDIG01 as the donor. Just under half this number (17.66), were obtained when aTc was included.

These findings lead to the conclusion that the staggered pMTL-pDIG01 and pMTL-pDIG02 system is not suitable for implementation of transposon mutagenesis in neither strain 630Δ*erm* nor CD9301. Further increased CfUs are observed on Em plates, induced with aTc than on Em without inducer. In this system, plasmid loss is induced via increased transcription from *tetR* into the pCB102 replicon.

Table 4.5: Plasmid loss analysis of pMTL-pDIG01 in 630Δerm. *C. difficile* strain CD630Δ*erm* were transformed with pMTL-pDIG01 Em positive colonies were selected and grown Em in supplemented BHIS medium over-night. The plasmids properties to cause plasmid loss, was calculated as previous. Data is from three biological replicates.

C. difficile* strain CD630*Δerm	**Colonies growing on Erm**	**Colonies growing on Em and aTC**	**Plasmid loss (%)**
Replicate 1	4×10^{-5}	1×10^{-5}	75
Replicate2	10×10^{-6}	8×10^{-6}	20
Replicate 3	3×10^{-5}	6×10^{-5}	-100
Ø			-1.66
Standard Deviation			73.06

Evaluating pMTL-pDIG1 in co-culture with pMTL-pDIG2, our data suggests that the activation of the aTc system triggers stimulated growth instead of plasmid elimination (Table 4.5). Colony formation of Em of strain 630Δ*erm* reveals that pMTL-pDIG01 is retained throughout the assay. A possible explanation is that during the long 24h conjugation time, the *tetR* on pMTL-pDIG01 becomes leaky. Pre-induced transposition in a fraction of *C. difficile* strains possibly results in Tm resistance, indicating pMTL-pDIG02 retention. However, the lack of a Gram-positive replicon in this plasmid makes the retention of pMTL-DIG02 unlikely. Thus, possible plasmid integration events of the suicide vector into the *C. difficile* genome may be hypothesized. Overall, the increased number of pMTL-pDIG01 and thus transposase retention makes a broad scale transposon mutant selection impossible in 630Δ*erm*. The lack of Tm positive colonies in strain 9301 indicates that no transposon mutants have been established.

These findings lead to the conclusion that the staggered pMTL-pDIG1 and pMTL-pDIG2 system is not suitable for implementation of transposon mutagenesis in neither strain 630Δ*erm* nor 9301.

4.3.7 A tetracycline inducible suicide vector pMTL-CW20ΔR

Suicide vector delivery systems are preferred over a replicative system, as it ensures that mutant pools are not 'contaminated' with cells carry plasmid-born transposons. Suicide, however, can only be used if high efficient transfer methods are available. It is possible that the recent improvements in DNA transfer rates into *Clostridium* spp., based on the novel conjugational donor sExpress, may make the application of suicide vectors in transposon delivery *in C. difficile* [77, 102, 158]. To investigate this, the utility of a suicide vector was investigated. Derived from pMTL-CW21 through deletion of its Gram-positive replicon (via an *Asc*I – *Fse*I digest), its Himar1C9 transposase is under the control of the Tc inducible P_{tet} promotor, lacking P_{tcdB} (Figure 4.7).

4.3.10 Evaluating pMTL-CW20ΔR.

The ability of pMTL-CW20ΔR (Figure 4.7) to deliver the transposon from *E. coli* to *C. difficile* CD9301 was evaluated including separate conjugations into *C. difficile* 630Δ*erm* and CD9301. In essence high CFUs were obtained on Em compared to Tm, and the former resistance was apparently not lost.

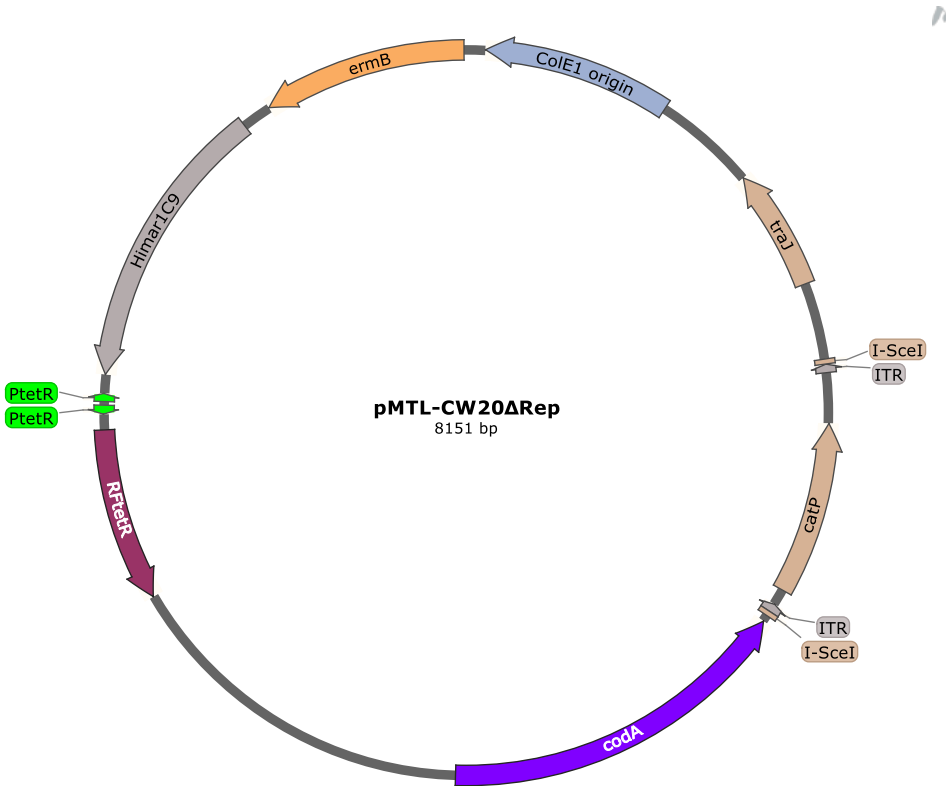


Figure 4.7: pMTL-CW20ΔR.

This plasmid is derived from pMTL-CW22/CW21. The divergent Tc promoters control the transposase and the *tetR* gene. I-*Sce*I recognition sequences flank the transposon as in pMTL-CW22/CW21 and the conformation P_{thl}-Transposon-*codA* is present for increased plasmid stability in *E. coli* storage and donor strains. As the Gram-positive replicon has been removed, this plasmid will not replicate in the *C. difficile* host.

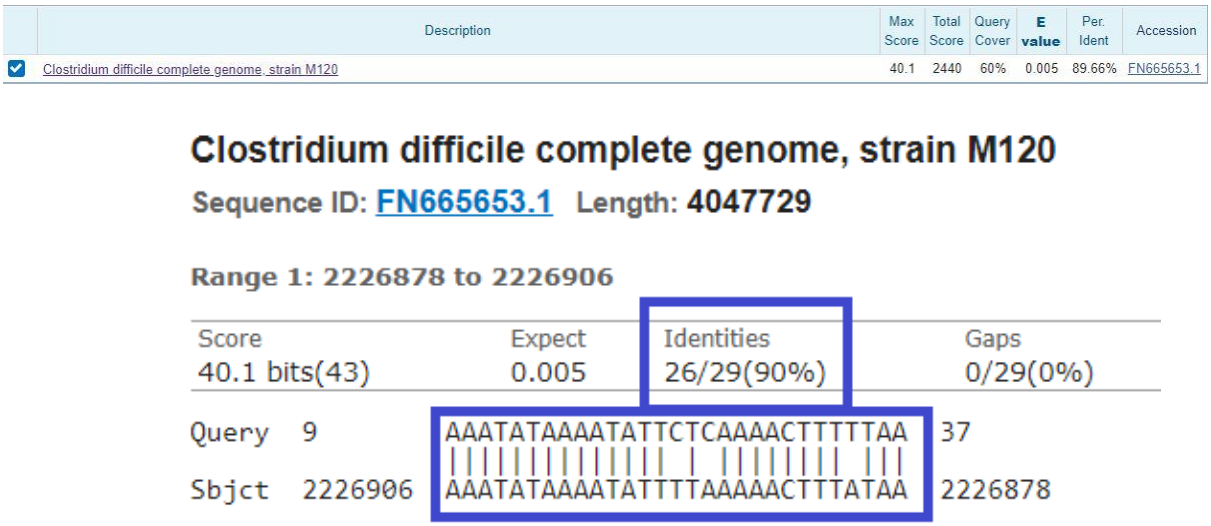
If integration of the plasmid is occurring, it is not clear what the homologous regions shared by the chromosome and plasmid might be responsible. All plasmids used incorporate the *ermB* gene but this or similar gene(s) are not carried by RT 078 strains, which are all sensitive to this antibiotic. However, this transposon vector also demonstrated an increased frequency of plasmid retention as indicated by colony formation on media supplemented with Em (Figure 4.7).

Table 4.6: Conjugal transfer of pMTL-CW20ΔR from *E. coli* to *C. difficile* CD9301 and CD630Δerm. After 8 and 24 hours of conjugation time, transconjugant selection plates were flooded with PBS and the cells harvested. The transconjugant cell mixture was serially diluted down to 10^{-7} and plated onto BHIS supplemented with CC and either Em +aTc, or Tm +aTc and the respective CFU counts per mL were determined. Plasmid performance properties were calculated according to section 4.3.2.

	***C. difficile* Recipient**			
	CD630Δ*erm*		**CD9301**	
	Conjugation time			
	8 hours	24 hours	8 hours	24 hours
Plasmid Loss Efficiency	-200.0	85.72	-83.3	-20.05
Resistance to Plasmid Loss	0.33	7.0	0.54	0.83
Transposon Frequency	-0.5	-0.14	0.0	1.0
Mutant Selection Purity	-20.0	No growth	0.0	45.44

The erythromycin antibiotic resistance gene *ermB* was always employed as the marker gene on all plasmids employed in these experiments. All strains of *C. difficile* utilized in this study, however, are known to be Em sensitive. A search of the genome of the RT 078 strain M120 with *ermB* using BLASTN indicated that small regions of around 30 bp of homology between *ermB* and sections of the

M120 genome of up to 90% were present (Fig. 4.10). It is not clear whether these small regions could mediate integration of the plasmid. In *C. cellulolyticum*, the minimum level of sequence homology required for demonstrable homologous recombination during CRISPR-assays was 100 bp [159], while the *Streptomyces* phage ϕBT1 requires a minimal homology of 36 bp and 48 bp to successfully integrate DNA through the action of site-specific integrases [160]. Thus, the overall base pair homology of the plasmid derived *ermB* and the M120 reference genome are of insufficient length even in the presence of integrases.



Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Clostridium difficile complete genome, strain M120	40.1	2440	60%	0.005	89.66%	FN665653.1

Clostridium difficile complete genome, strain M120

Sequence ID: FN665653.1 Length: 4047729

Range 1: 2226878 to 2226906

Score	Expect	Identities	Gaps
40.1 bits(43)	0.005	26/29(90%)	0/29(0%)

```
Query  9        AAATATAAAATATTCTCAAAACTTTTTAA  37
               ||||||||||||||| | ||||||||  |||
Sbjct  2226906  AAATATAAAATATTTTAAAAACTTTATAA  2226878
```

Figure 4.8: BlastN of *ermB* against the *C. difficile* strain M120 genome.

Blasting the *ermB* sequence utilized in the pMTL-8000 series against the genome of the *C. difficile* RT 078 model strain M120. The program used for this sequence comparison is available on NCBI, BlastN (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch). For an optimal analysis, the option: “Somewhat similar sequence (blastn)”, has been selected.

While it is not inconceivable that the above regions do mediate integration of the plasmid, further evidence would need to be acquired to support this, ultimately whole genome sequencing. It appears that the integration of the entire plasmid into the *C. difficile* genome, evasion of plasmid curing and eventual gain of Em resistance may be caused by an alternative mechanism. *C. difficile* has a long-standing history of acquiring, accumulating and maintaining antibiotic resistance genes mediating, for instance, resistance to metronidazole, vancomycin and fidaxomicin [161-165].

4.3.11 Employing the transposon vector pRF215.

None of the transposon vectors examined to this point possessed the properties required for deployment of TraDIS. Attention was turned to the plasmid pRF215 of Fagan et al [84], presented in Figure 4.9. A pivotal difference between this plasmid and those described above was that the antibiotics used as the backbone and transposon marker had been reversed. Thus, the *catP* gene in the mini transposon had been exchanged with the *ermB* marker. Utilizing this plasmid Dembek and colleagues identified 404 essential genes for surivival in 630Δ*erm* and 798 in R20291 [86]. The transposon vector pRF215 (Figure 4.9) features divergent aTc-inducible promoters with overlapping *tetO* sequences in each. Similar to the pMTL-CW systems, P_{tetR} regulates the expression of *tetR*, encoding the

transcriptional regulator for *tetO*, repressing both promoters. In the presence of aTc both P_{tet} promoters are transcriptionally active. Even though, the aTc inducible system of pMTL-CW2X/ pMTL-pDIG-01 and pRF215 are very similar, the system in the pMTL vectors contain an additional RBS between the P_{tet} promotor and the *Himar1C9* transposase, not present in pRF215.

In pRF215 inducible plasmid loss is dependent on transcription into the *C. difficile* pCD6 replicon. Almost complete plasmid loss was reported within 13 generations of growth following induction [86]. Further, the divergent P_{tet} promotor tightly controls the Himar1C9 transposase and thus transposition of an *ermB* transposon. Dembek *et al*. detected Em-resistant mutants with a frequency of 1.18×10^{-4} featuring wide range of colony morphologies, suggesting successful random transfer of the transposon. For library production approximately 85,000 colonies of R20291 have been pooled and TraDIS identified 70,000 unique insertion sites of the transposon, with average of one insertion every 97 bp. Subsequently 404 essential genes were characterized.

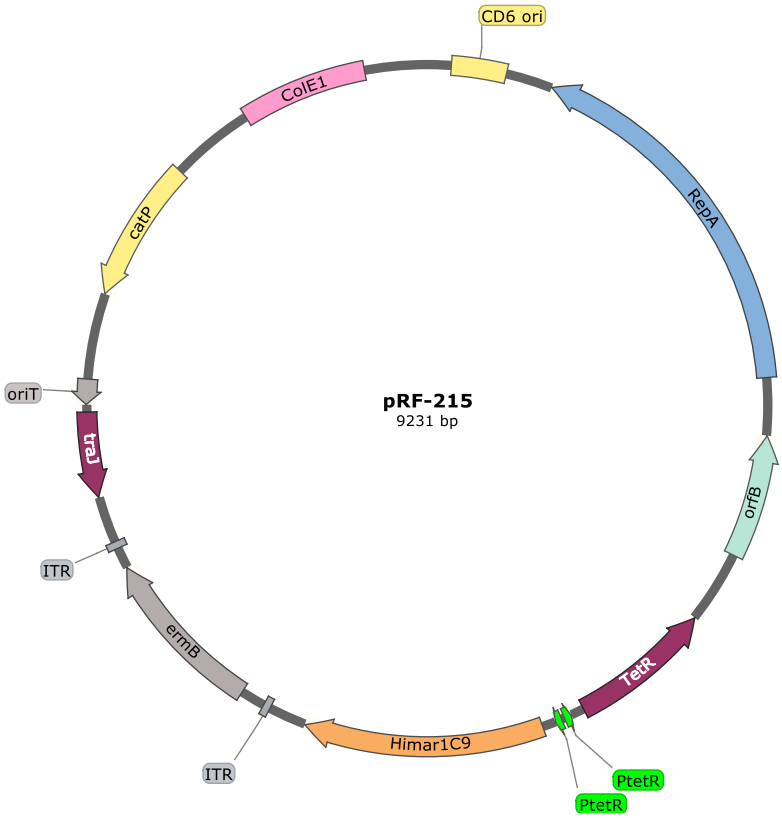


Figure 4.9: Transposon delivery vector pRF215.

The *transposon* delivery vector developed by Dembek et al. [84], for *C. difficile*. It employs the conditional replicon causing plasmid loss via transcription into pCD6. Plasmid loss and the Himar1C9 transposase-*ermB* transposon complex are controlled via inducible P_{tet} promoter. In contrast to the previously employed plasmid systems, pRF215 contains *ermB* as transposon marker and the *catP* acquired Tm resistance on the plasmid backbone. Successful replication in *E. coli* is achieved via the Gram-negative replicon *ColE1.*

Generally, the density of insertion did not appear sufficient to saturate the genome and allow the precise identification of nonessential genes. However, pRF215 serves as an example of a transposon-delivery system suitable to produce large transposon libraries in *C. difficile* in 630Δ*erm* and R20291. Its application in RT 078 *C. difficile* strains was therefore examined.

4.3.12 Evaluation of pRF215 in *C. difficile* CD9301

The transposons vector pRF215 was assessed in the same manner as applied to pMTL-CW22/21 in terms of measuring plasmid loss and transposon frequency following aTc (100 ng/mL) induction.

Table 4.7: Evaluation of pRF215 in *C. difficile* CD9301. Transposition frequency and plasmid loss efficiency of pRF215 and pMTL-CW22 in *C. difficile* RT 078 clinical isolate CD9301. Mutant colonies were selected after 48 hours growth on selection plates. Emerging colonies were counted and the quality of pRF215 in strain CD9301 was assessed. Three biological replicates were undertaken. The extensive statistical analysis is outlined in Table 4.9 and Figure 4.13.

Plasmid	Plasmid Loss Efficiency (%)	Resistance to plasmid loss	Transposition frequency
pMTL-CW22	53.02	3.48	0.23
pRF-215	96.23	193.7	0.018

Despite the two plasmids utilising essentially the same aTc-inducible system and pCD6 replicon, it was clear that pRF215 exhibited greater plasmid loss (96%) than pMTL-CW22 (53.02%) after induction (Table 4.5). Overall, high levels of transposition were obtained with pRF215, while just 1 out of 193.7 cells were resistant to plasmid loss. The data obtained with pRF215 additionally showed a high level of consistency over the three biological replicates. On this basis, pRF215 appeared to represent a better option for undertaking TraDIS in RT 078. However, unlike pMTL-CW22, the mini-transposon of pRF215 does not carry flanking I-*Sce*I recognition sequences. The presence of these restriction sites represents a potential improvement to the process of plasmid deletion. During downstream molecular protocols, these sites allow a significant reduction in the presence of plasmid reads during the subsequent library preparation step.

4.3.13 Inverse PCR analysis.

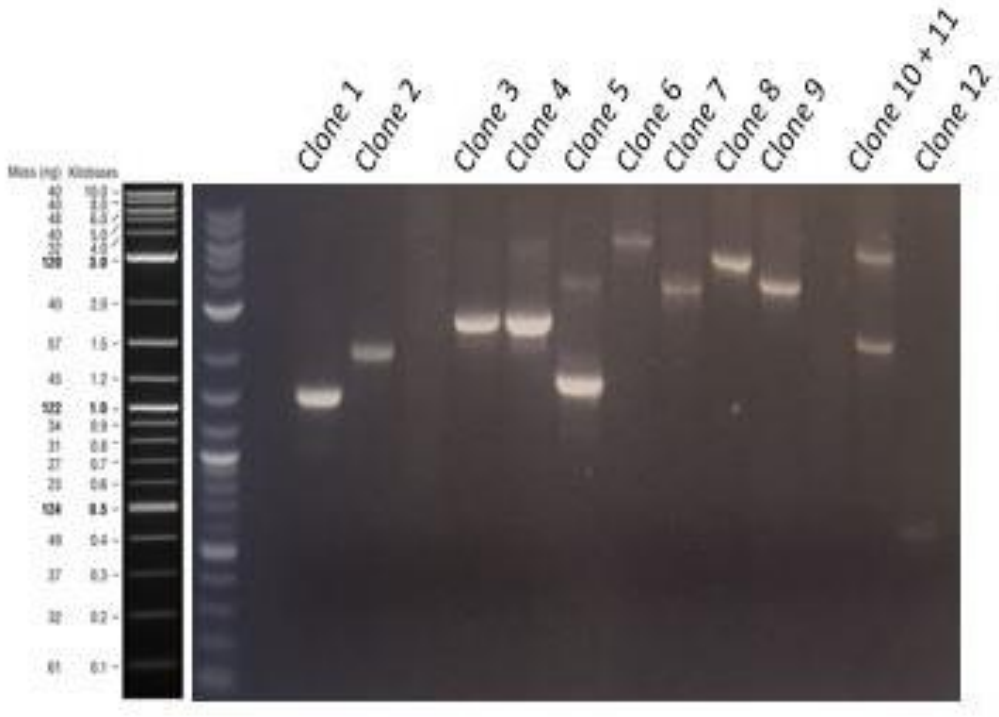


Figure 4.10: Insertion sites for randomly chosen transposon mutants.

Agarose gel electrophoresis of the putative amplified DNA fragments following inverse PCR on 12 randomly selected transposon mutants generated in *C. difficile* strain CD9301 using pRF215. Putative mutants were inoculated into liquid culture from which genomic DNA was isolated after 48 hours growth. Lane 1 contains the molecular marker 1kb Plus ladder (Thermo Scientific). Lane 2-14 contains randomly chosen mutant colonies that underwent iPCR treatment.

To evaluate if the cells of colonies observed on the Tm-supplemented transposon selection plates, containing aTc, contained unique insertions into the *C. difficile* genome, a screening method based on inverse PCR was conducted. Extracted gDNA was harvested from all 14 colonies, followed by a *Hin*dIII digest and a ligation

with T4 ligase. The processed gDNA was used as template for inverse PCR, utilizing transposon specific primers. Amplified DNA could only be visualised in 12 of the 14 samples on an agarose gel (Figure 4.10). These bands were isolated and sent for Sanger sequencing, utilizing the primer "ermB-F1". All of the sequences obtained matched the RT 078 genome. Ten represented unique sequences, whereas the remaining two were duplicates

Table 4.8: Insertion sites for randomly chosen transposon mutants. Further analysis of the DNA fragments produced by inverse PCR, visualized in Figure 4.10. Transposon mutants of *C. difficile* strain CD9301 obtained using pRF215 were picked and isolated from the gel. Sanger sequencing was performed using the sequencing primer ermB-R1 by Eurofins Genomics. The resultant sequences were compared to the RT 078 M120 genome using BLASTn. The Tn was inserted in the indicated genes.

Sample No.	Gene Name	CDS
1	Hypothetical protein	AYD18829.1
2	S-layer protein precursor (*slp*A) gene, complete *cds*; and SecA (*sec*A) gene, partial *cds*	AAZ05982.1 AAZ05981.1
3	Drug/sodium antiporter	AXU66631.1
4	Endopeptidase, cation transporter	AXU69693.1
5	Signalling protein	AXU68700.1
6	Methionine aminopeptidase	AXU67932.1
7	Heat shock protein 90	AXU88930.1
8	Sensory transduction histidine kinase	AXU88711.1
9 and 10	O-sialoglycoprotein endopeptidase	AXU89701.1
11	Transcriptional regulator	AXU67057.1
12	Hypothetical protein	AXU67742.1

4.3.14 Analysis of pRF215 TraDIS sequencing data in *C. difficile.*

The experiments conducted using pRF215 in strain CD9301 demonstrated the plasmid consistently underwent a reasonable degree of plasmid loss and produced single *ermB* transposon insertions, both essential to the generation of a saturated transposon library. To gain an insight into whether pRF215 also represents an optimal plasmid for broad range library production and eventually TraDIS analysis, the overall ratios of plasmid curing during TraDIS assays in *C. difficile* strains 630Δ*erm* and R20291, the raw data files of the TraDIS assay conducted by Dembek *et al*. were analysed [86]. The evaluation of the TraDIS data in the two strains of *C. difficile* shows the pRF215 transposon vector maps to an average of 6.39% total reads in 630 Δ*erm* (Table 4.7). In R20291, however, the plasmid is associated with 22.01% of the transposon insertion reads (Table 4.7). The transposon reads which corresponded to insertions into the *C. difficile* genome were 89.67% in 630Δ*erm* and 71.55% in R20291. In contrast to this, 80% and 100% of chromosomal reads were obtained for *C. acetobutylicum* and *C. sporogenes* respectively [154]. Furthermore, the transposon plasmid pMTL-YZ14 resulted in 89-100% of genome insertions *C. autoethanogenum* (Craig Woods, PhD thesis).

Table 4.7: Distribution of mini-transposon reads during TraDIS in *C. difficile* strain 630Δ*erm* and R20291. The percentage of TraDIS reads mapping to the chromosome and plasmid in a transposon library made in two *C. difficile* using the plasmid pRF215. Respective statistics are evaluated and depicted in in Table 4.9.

READS (%)			
630Δ*erm*[a]			R20291[b]
Chromosome	Plasmid	Chromosome	Plasmid
89.67	6.40	71.56	22.02

4.3.15 Analysing transposition events into pRF215.

Analysis of the TraDIS sequencing data, revealed that many reads were assigned to plasmid contamination due to remaining pRF215 in the library. During library sequencing, plasmid contaminations may not interfere extensively with the purity of the chromosome library. Interestingly, larger number of cells that retain the plasmid (as measured by Erm^R) occur with R20291 than *630*Δ*erm*. Increased avoidance of plasmid loss, through stabilisation of the plasmid, may be linked to transposition events into the pRF215 plasmid itself. To obtain evidence for this, the transposon reads from the Dembek et al study [86] were examined using the Artemis Gene browser tool (Sanger Institute). Figure 4.11 shows that a dominant fraction of the transposition events have occurred into the aTc-inducible promotor region (RFtetR) and *orfB* of pRF215. The disruption of this region most likely interferes with transcription into the pCD6 replicon causing inactivation of the inducible plasmid loss mechanism. Similar events are likely to occur during the use

of comparable plasmids, like pMTL-CW21 and CW22. These events complicate the use of conditional systems developed to eliminate replicative forms of the plasmid from transposon libraries.

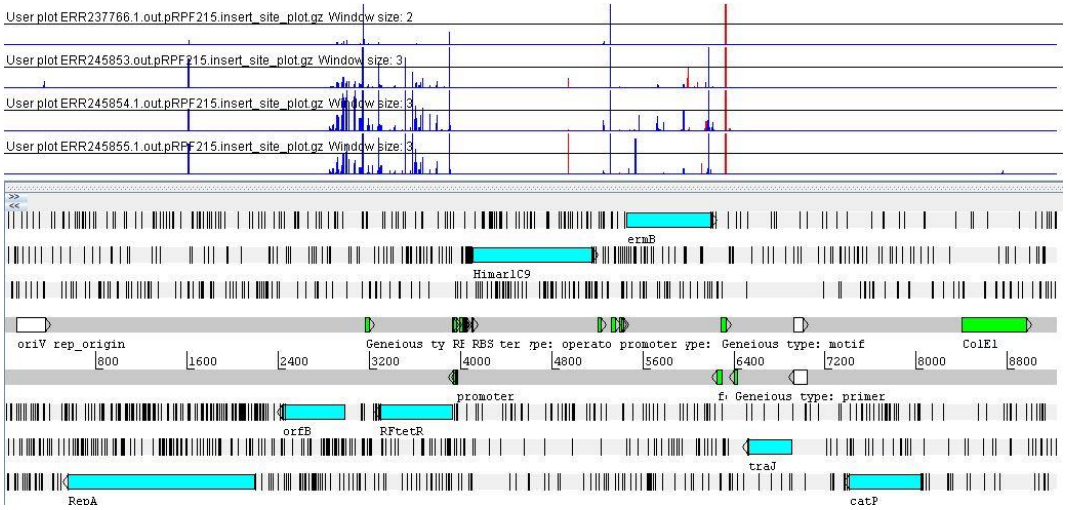


Figure 4.11: Transposon integrations into pRF215.

Produced using the Artemis free genome browser and annotation tool (Sanger Institute). The remaining transposon plasmid reads during the TraDIS assay of Dembek *et al*. were analysed [86]. The integration sides of the *ermB* transposon into the remaining pRF215 plasmid have been determined. The data is based on the four-library sequencing runs of the pRF215 transposon vector into *C. difficile* strain 630Δ*erm*.

4.3.16 A novel transposon delivery system in *C. difficile*, pMTL-MtV10

When evaluating transposon vectors in *C. difficile*, plasmid pRF215 generally preforms better than pMTL-CW21/22 in terms of plasmid loss from the cells, however, the latter plasmids allow plasmid reads to be eliminated during library preparation through the use of the I-*Sce*I sites. A new vector was therefore made, pMTL-MtV10, which combines the features of the two plasmid series and is depicted in Figure 4.12. The new plasmid was made by combining the DNA fragment from pRF215 comprising the inducible *tetO*/P_{tet} based divergent promoter system, the Himar1C9 transposase gene and *ermB* transposon of pRF215 with a fragment of plasmid pMTL84151 encompassing the selectable marker *catP* and the pCD6 Gram-positive. The two fragments were PCR amplified and joined by and HiFi assembly using the templates and primers as shown in Table 4.8. Importantly, during HiFi assembly, I-*SceI* recognition sites were cloned into the vectors flanking the *ermB,* mini-transposon. The final plasmid also incorporated the modular flanking *Asc*I and *Fse*I sites around the pCD6 replicon which facilitates the future exchange of the plasmid replication region if required. Transformants were selected on LB agar supplemented with chloramphenicol and Em and inoculated to overnight cultures with the same supplementation. Plasmid DNA was prepared from overnight cultures and verified with diagnostic digests and Sanger sequencing.

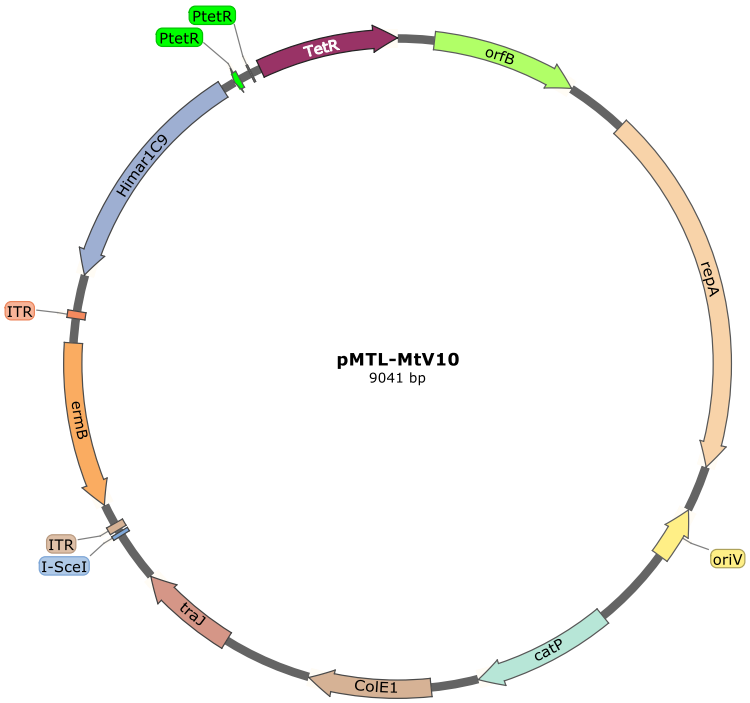


Figure 4.12: pMTL-MtV10.

A plasmid map of pMTL-MtV10. The divergent Tc promoters originating from pRF215 control the transposase and the TetR protein, causing relocation of the *ermB* transposon. Inducible transcriptional read-through into pCD6 also causing plasmid instability. I-*SceI* recognition sites flank the transposon.

Table 4.8 Primers and templates used for construction of pMTL-MtV10. pMTL-MtV10 was constructed using HiFi assembly. This table presents the primers names and relevant templates used to generate the DNA fragments with overlapping regions allowing the HiFi reaction to form a completed molecule. The respective primer sequences are depicted in Table 2.5.3.

Primer Name	Template	Description
MTV10.1 F	pRF215	Containing the I-*Sce*I spacer, tet-Promotor system, Himar1C9 transposase, *ermB* Transposon
MTV10.1 R	pRF215	
MTV10.2 F	pMTL84151	Containing the modular pCD6 Gram-positive replicon, Tm resistance
MTV10.2 R	pMTL84151	

4.3.18 Assessment of pMTL-MtV10 in *C. difficile* CD9301.

The properties of the novel transposon-delivery vector pMTL-MtV10 were assessed as previously described, inducing plasmid loss and transposition with aTc (100 ng/mL). The performance of the vector is summarised in Table 4.9. The novel plasmid exhibits a high degree of instability following addition of aTc with 97.07% of the CD9301 Em resistant transconjugants exhibiting plasmid loss, compared to 96.23% with pJF215 and 53.02% with pMTL-CW22. Comparing the numbers of the aTc-induced mutant selection plates vs the non-induced controls revealed the efficiency of pMTL-MtV10 to form transposon libraries. Overall, it is evident that pMTL-MtV10 performs well as transposon delivery vector. Despite its operational similarity to pMTL-CW22, the novel vectors quality as transposon delivery system outcompeted the progenitor in every function necessary for successful library production.

Further, pMTL-MtV10 performs like pRF215 in the tested RT 078 strain CD9301. As both plasmids mediate plasmid loss through the same Tc reactive system into the pCD6 replicon, plasmid loss values are of comparatively good quality with pMTL-MtV10 at 97.07% and pRF215 at 96.23%. Interestingly, the resistance to plasmid loss and the transposon frequency in pMTL-MtV10 slightly exceed that of pRF-215. As both plasmids contain the same reactive centre, responsible for transposon frequency and plasmid loss, similarly good quality was achieved. Determining the statistical significance between these data sets revealed, that the p-Value of plasmid loss and transposon frequency between pRF215 to pMTL-MtV10 was no underlying statistical difference.

Table 4.9: Assessment of pMTL-MtV10 in *C. difficile* CD9301. Transconjugants carrying one of the three transposon-delivery vectors were cultivated in the presence of Tm, overnight before being serially diluted on non-selective media as wells as selective media (Em or Tm) with or without aTc (100 ng/mL). Following counting of the CFU, estimates were made of: Transposition frequency ((Em, aTc – Tm, aTc) ÷ Em; Plasmid loss efficiency (Em – Em, aTc) ÷ Em × 100; Plasmid retention in presence of aTc (Em ÷ Em, aTc). The p-Value was determined in GraphPad Prism 8.4.3 using the Two-stage linear step-up procedure of Benjamin, Krieger and Yekutieli (with Q=1%). Statistical significance between the datasets of pRF215 to pMTL-Mt10 regarding plasmid loss is 0.958 and of transposition frequency 0.664, stating no statistically significant differences between the datasets.

Plasmid	Plasmid loss (%)	SD of Plasmid loss	p-Value of Plasmid loss	Resistance to plasmid loss	Transposon frequency
pMTL-CW22	53.02	31.72		3.48	0.23

pRF-215	96.23	3.65	0.958	193.7	0.02
pMTL-MtV10	97.07	3.45		345	0.0067

Figure 4.13: Plasmid loss efficiency of three TraDIS vectors.

Comparison of the 3 TraDIS vectors pMTL-MtV10, pRF215 and pMTL-CW22. All vectors contain a Tc inducible divergent promotor unit. Transcriptional read-through into the Himar1C9 transposase causes transposition of the mariner Transposon and translocation of the transposon into the *C. difficile* CD9301 genome. Transcription into the pCD6 replicon causes plasmid instability. Plasmid loss efficiency of pMTL-MtV10 was superior to the other two vectors with 96.07% of the cells in the population losing the transposon plasmid after induction with aTc (SD 3.45). Cultures created with pRF215 expressed 96.23% of plasmid loss (SD 3.65), while CD9301 cells treated with pMTL-CW22 lost the plasmid with 53.02% accuracy (SD 31.27). Significance has been addressed by calculating the p-Value using the two-stage linear step-up procedure of Benjamin, Krieger and Yekutieli (with Q=1%) and resulted in 0.958 between pRF215 and pMTL-MtV10, stating no significant difference between the datasets. pMTL-CW22 has not been statistically analysed further as it is considered not suitable for further TraDIS studies based on its general poor performance.

4.4 Discussion

Transposon directed mutagenesis represents a promising tool to elucidate genetic processes under numerous conditions in *C. difficile*. However, implementation of TraDIS remain difficult. Major impediments are the poor frequencies with which DNA may be introduced into *C. difficile* as well as finding hypothetical constructs to create a transposon mutant library suitable for TraDIS analysis [166]. Recently, advances in DNA transformation into *C. difficile*, such as electroporation or evading the R-M systems with novel conjugational donor strains have been described [126, 158]. This shifts the focus of attention towards finding the optimal tool for transposon delivery and additional plasmid elimination. Successful transposon mutagenesis essentially relies on single transposon insertion into the target genome by a transposase. Further, additional loss of the transposase is crucial to prepare a stable mutant. Simultaneous and controlled activation of plasmid loss and transposition has been well described in *Clostridium*, using inducible promotor systems [91, 154].

To select a suitable inducer system for implementation of TraDIS in *C. difficile*, certain characteristics must be considered: wide dynamic range of activity, tight repression, optimal bioavailability of the inducer, quick induction, and the ability to work in multiple *Clostridium* spp. This chapter summarizes the implementation

and analysis of several different transposon delivery vectors in *C. difficile* strains CD9301 and 630Δ*erm*.

In this study, the lactose and IPTG inducible pMTL-CW17 has been tested in *C. difficile*. The Himar1C9 transposase activation is coupled to the lactose inducible *bgaR*-P_{bgaL}, while plasmid loss is triggered by the IPTG induced P_{fac} promoter. The plasmid pMTL-CW17 has been conjugated and induced in the target *C. difficile* strains. However, the achieved values of plasmid loss as well as transposition frequency were not sufficient for TraDIS. This may be explained by the rapid volatilization of lactose, which increasingly hampers the transposase activity over the experimental set up [150]. In *C. perfringens* a lactose inducible system has been developed. The lactose however was metabolized after 75 min, resulting in weakening induction [150]. Further it is hypothesized that *C. difficile* cannot metabolize or take up the gratuitous IPTG, rendering induced plasmid loss in pMTL-CW17 impossible [154].

As an alternative to the lactose inducible systems, a Tet system, developed by Dr. Craig Woods was selected, as it met all the characteristics required for the envisaged TraDIS vector. Comprising a divergent promotor system, P_{tet} controls the expression of the target gene, while P_{tetR} transcribes *tetR* which encodes the Tc repressor. (Anhydrous)-tetracycline binds the TetR repressor, and the system becomes activated. The Himar1C9 transposase delivers the *catP* transposon into

the target genome while transcription read through into the Gram-positive replicon promotes plasmid loss.

The aTc inducible plasmids pMTL-CW21 and pMTL-CW22 are identical except for the Gram-positive replicon employed. However, neither transcription into pCB102 (pMTL-CW21) or pCD6 (pMTL-CW22) resulted in sufficient plasmid loss. Specifically, fluctuating values of plasmid loss efficiency between biological replicates were observed. While these plasmids have been utilized to generate transposon mutant libraries suitable for TraDIS analysis in other *Clostridium* species (PhD theses of Raquel Rodrigues and Craig Woods) they were found to be not suitable for *C. difficile*.

Trying to prevent the issue of low plasmid loss, a suicide vector pMTL-CW20ΔR was created by removing the Gram-positive replicon from pMTL-CW2X. Suicide vectors lack the Gram-positive replicon and are therefore not maintained and passed onto daughter cells, ensuring there loss from the cell population. The Himar1C9 transposase is under the control of the Tc inducible promotor, P_{tet}. Interestingly, several *C. difficile* transformants expressed resistance against the Em, suggesting that the plasmid, or at least the *ermB* antibiotic resistance marker it carried, remained in the cells.

Similar results have been observed, when dividing the P_{tet} transposase and the transposon on two different plasmids in the staggered system of pMTL-pDIF01/02. Analysis of this system suggested that the number of cells carry the plasmid, as measured by the CFU on media supplemented with Em, increased after induction of the P_{tet} promoter with aTc

This makes these plasmid systems unsuitable for the deployment of TraDIS, as the presence of the transposase causes unstable mutations. The persistence of the plasmid encoded Em^R phenotype may be a consequence of integration of the plasmid, partially or in total, into the *C. difficile* genome, due to sequence homology with the *ermB* gene on the plasmid [161, 162]. Furthermore, the spread of antibiotic resistances in *C. difficile* has been shown to persist and spread amongst an entire culture, even in absence of antibiotic selective pressure [163-165].

So far, the plasmid systems analysed in this study have demonstrated them not to be suitable for the deployment of TraDIS in *C. difficile* RT 078 strain CD9301. At this point it remains questionable if the aTc systems employed were malfunctioning or if the *ermB* backbone resistance causes increased plasmid stability in the organism. However, the predominant problem is insufficient plasmid loss.

To circumvent the problem, the transposon delivery vector pRF215 was analysed. It was previously used to identify essential genes in *C. difficile* strains 630Δ*erm* and R20291 and employs a different inducible Tet system and a backbone conferring resistance to Tm [86]. The Tet inducible system of pMTL-CW2X/ pMTL-pDIG-01 and pRF215 are similar. However, the Nottingham-Tet constructs are derived from an expression vector, designed for *Staphylococcus aureus* [157]. *This* system has been mirrored to ensure divergent expression and was codon optimized by Dave Walker (Dave Walker, PhD thesis). Further, pMTL-Tet systems contain an additional RBS between the P_{tet} promotor and the *Himar1C9* transposase, not present in pRF215. When conjugated into strain CD9301, pRF215 consistently outperformed the pMTL vectors in terms of high rates of plasmid loss, transposition frequency and maintenance of the conditional nature of its replicon, hampering the clearance of the plasmid from the cell. Inverse PCR analysis of random transposon mutants indicated transposon insertions into various parts of the CD9301 genome. However, the library sequencing data of two *C. difficile* strains, conducted by Dembek et al., revealed that up to 22.01% of total reads map to pRF215 [86]. The majority of remaining plasmid has a transposon insertion into the Tc inducible promotor region (RFtetR) and the *orfB*, disrupting the plasmid loss mechanism.

Based on these findings we aimed to combine the clearly functional divergent Tc system, the Himar1C9 transposase as well as the *ermB* transposon of pRF215 with

the additional I-*SceI* restriction site, present in the pMTL-CW2X constructs. This site can be targeted during library preparation and allows additional removal of plasmid backbone during the sequencing stage. The vector was termed pMTL-MtV10 and in terms of plasmid loss, transposition frequency and resistance to plasmid loss, it has a comparable performance to pRF215 in strain CD9301. These results are as expected, considering both plasmids share the same transposase, Tc inducible promoter and replicon. Overall, pMTL-MtV10 outperforms the previous transposon vectors employed in this study and its specific improvements make it an optimal vector for mutant library production for TraDIS analysis.

Chapter 5

Establishing TraDIS in *C. difficile* Ribotype 078

5.1 Introduction

5.1.1 Transposon mutagenesis in *C. difficile.*

As discussed in Chapter 1 and Chapter 4, transposable elements are a promising tool for performing forward genetics and have been used for mutant production in several clostridia species [65, 86]. Large random transposon insertions can be used to evaluate essentiality of genes and the fitness of a given phenotype under various environmental conditions [65, 86]95]. These extensive transposon libraries are analysed by high throughput sequencing technologies, such as Transposon Insertion Sequencing (TIS), Insertion sequencing (INSeq), high-throughput insertion track by deep *sequencing* (HITS) and Transposon sequencing (Tn-seq) which upon minor variations, generally follow a similar workflow [78]. As described by Barquiste et al., [78] the *in vitro* transposition of a transposable element into a target organism is followed by mutant pooling and subsequent DNA extraction. During the library preparation, the DNA is fragmented, and transposon insertions are tagged, and PCR enriched. Subsequent sequencing and *in silico* mapping of the transposon insertion sides reveals its location into the hosts

genome [98]. TIS is a high-throughput functional genomics approach as it allows to compare various conditions such as bacterial stress responses or assay essential genes quickly.

Identification of the transposon insertion within a library is achieved by high-throughput sequencing. Simultaneous mapping of all insertion sites within the library against the original genome allows the identification of essential genes, represented as genes that lack a transposon insertion. Continuous growth of the library under various environmental conditions with subsequent sequencing allows the simultaneous evaluation of the fitness of mutants. Using TIS, Langridge et al introduced the Transposon Directed Insertion-site Sequencing (TraDIS) technique to sufficiently screen mutant related fitness within a population of single transposon mutants [87]. The workflow of TraDIS is visualized in Figure 5.1 and requires a fragmentation and PCR based preparation of a transposon library. The employment of a variety of adaptable primers, makes TraDIS adjustable for multiple systems such as Tn*5*-, Tn*917*- and *Himar1*- based transposons, discussed in Chapter 1 [98]. Since TraDIS has been described, many library preparation and sequencing improvements have been performed [78, 98].

A crucial step during sequencing of a TraDIS library is the PCR amplification (presented in Figure 5.1) of the transposon prior to the Illumina sequencing. Transposon sequence amplification is necessary to reinforce the number of

transposons reads towards the *C. difficile* genome. A TraDIS library preparation starts with the fragmentation of the 4.2 million base pair *C. difficile* genome into approximately 10,500 400bp fragments per genome, of which most sections of the genome contain a transposon insertion.

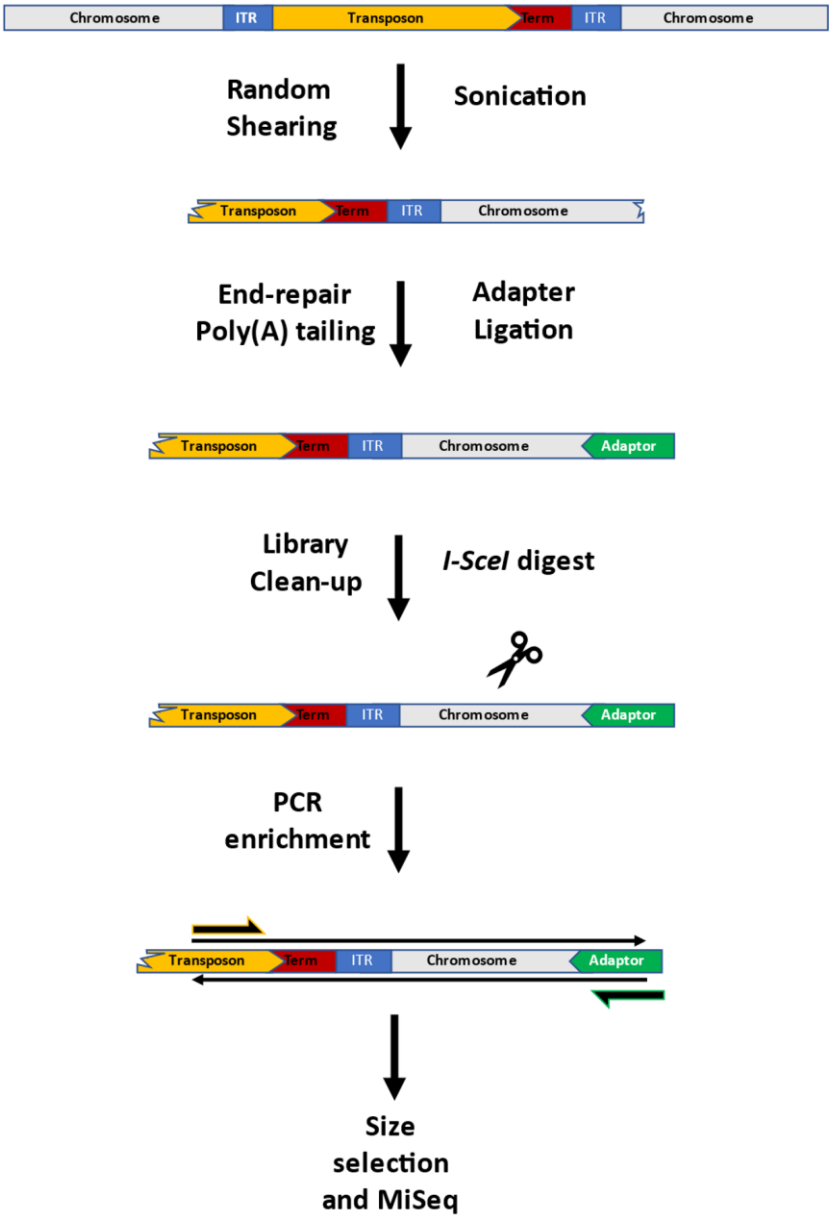


Figure 5.1.: DNA treatment during TraDIS library preparation. The genomic DNA is randomly sheared via sonication to an approximate length of 400bp. Fragments are end repaired and NEBNext UltraTM II DNA Library Prep Kit adapters were ligated to them. Remaining fragments were removed via a clean-up step, using Ampure XP magnetic beads. A PCR step, with primers, targeting the transposon and sequence and adaptor sequence amplified the library. The library was than digested with I-*SceI* overnight at 37°C. A second PCR step was conducted to enrich the transposon junctions and run on an agarose gel, to extract DNA -/= to 250bp. Extracted DNA quality was validated via a MiSeq platform.

During the TraDIS protocol, it is important to specifically target the transposon insertion site and leave the remaining gDNA without adapter, to specifically target transposon insertions into the genome. During this protocol, the same P7 primer is used, while the P5 primer is customized and referred to as the library primer. As depicted in Figure 5.1, the library primer is attached to the transposon and thus ensures exponential amplification of only transposon DNA and the insertion associate genomic side.

5.1.3 Essential genes

Essential genes are described as genes, that carry out biological functions, necessary for the survival of an organism. Identification if these genes during library screening is challenging as death is the only phenotype in the absence of

such an essential gene from the library. The knowledge about essential genes in hospital acquired diseases such as *C. difficile* PCR Ribotype 078 may be of pharmaceutical value and is specifically important for the implementation of further genetic approaches such as phage derived killing or as target sites for novel antibiotics [167]. Furthermore, knowledge of essential genes can be import during selection of target sites during genetic engineering of *C. difficile* strains. Knowing the essentiality of genes during basic cellular functions prevent the frustrating possibility of unintentionally targeting genomic sides, necessary for the organism survival.

A TraDIS study on *C. difficile* already revealed the essential genes in PCR ribotype 027 [86]. For our study it is of great interest to compare the similarity between the set of genes involved in these highly virulent *C. difficile* strains survival.

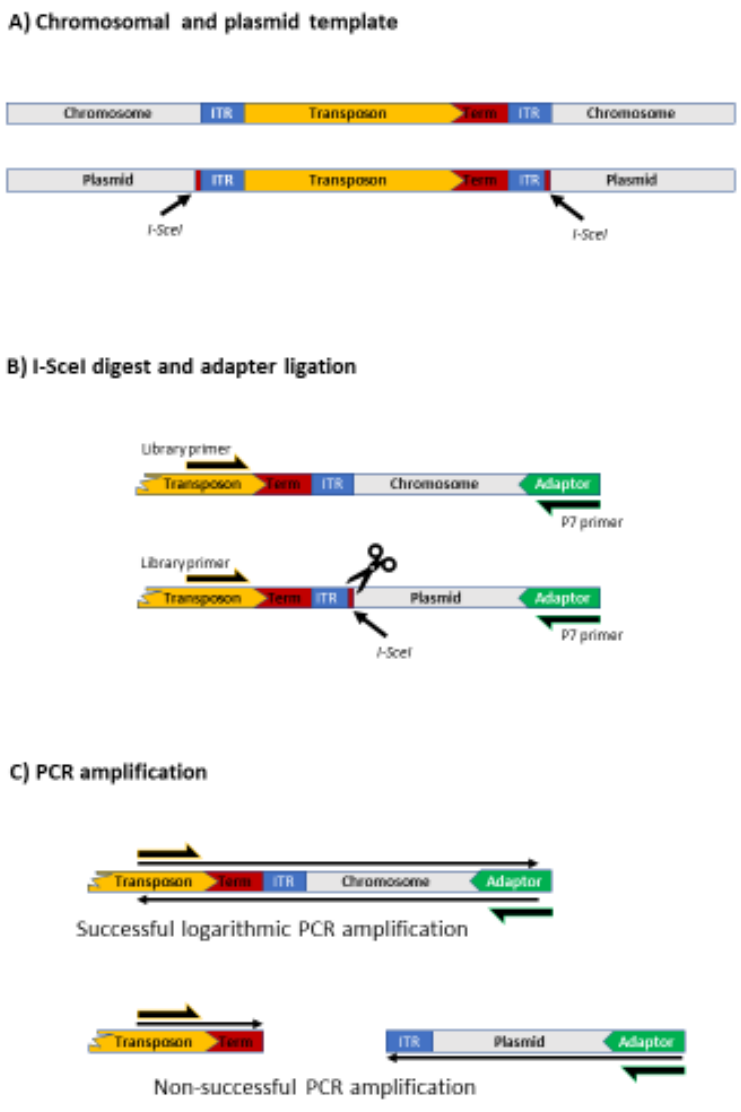


Figure 5.2: I-*SceI* digestion during the TraDIS library preparation.

The additional I-*SceI* side within plasmids such as pMTL-CW2X and pMTL-MtV10 allows additional plasmid curing during the library preparation. A) The difference between Chromosomal insertions of the transposon and the plasmid, where I-*SceI* restriction sides flank the transposon. B) Adaptor ligation and *I-SceI* digest of the remaining plasmid. C) Just the transposon sequences in the genome can be PCR amplified during the library preparation, as both, the P5 and P7 primers are able to bind to the template. Graphic was adapted from Syed et al., [168] the PhD thesis of Craig woods [169].

5.1.4 Sporulation and germination marker.

The genetic mechanism underlying clostridial spore biology have usually been predicted based on comparisons to the model organism, *B. subtilis.* Until today, the environmental factors, important during *C. difficile* sporulation and germination are largely unknown.

C. difficle sporulation is a major threat to clinical settings. Transmission of *C. difficile* to patients often occurs via health care personal carrying spores. Spores are highly resistant to most clinical disinfectants and thus are hard to treat with commercial methods and can survive for month in hospital environments. Even tough, *C. difficile* specific hygiene has improved during the last years, it has been suggested, that environmental contaminations with spores still are likely to contribute to 10% of new CDI cases [170]. Besides the how ratio of infections and transmissions of spores in hospitals, not much is known about spore interactions with the patient's gastrointestinal system, how it causes over colonialization of the patients gut and how it eventually results in CDI [171]. To enhance the understanding of the *C. difficile* spore biology and resulting patient complaints in hospital settings a better understanding of its genetic mode of actions is necessary.

Using TraDIS in *C. difficile* strain 630Δ*erm* and R20291, Dembek et al., [86] discovered 798 genes, associated with sporulation, of which only a few genes have been previously studied. Apart of the RNA sigma factors and the master regulator Spo0A they predicted a several genes that have no previously characterized role during *C. difficile* spore biology. Even though they confirmed the validity of their transposon library and consider the approach scalable, they did not consider the size and density of the mutant pool as sufficient to perform a statistical analysis of gene essentiality.

It is of great interest, to compare the sporulation and germination relevant features of the strains used by Dembek and colleagues [86] with the RT 078 strain CD9301. The addition of an I-*SceI* mediated plasmid curing step during the transposon library preparation of the pMTL-MtV10 TraDIS vector increases the library density by eliminating plasmid backbone, which potentially empathises genes with less frequent transposon insertion sides (Figure 5.2).

This study aims to produce and sequence a transposon mutant library in *C. difficile* RT 078 using TraDIS. To evaluate if an additional I-*SceI* digest improves the quality of the library, pRF215 is compared with pMTL-MTV10.

5.2 Aims of the study

The main objective of this research was to determine the essential genes needed for growth on rich media and essential genes needed for sporulation. A secondary objective was to compare the newly created MTV-10 transposon-delivery plasmid with a previously published plasmid called pRF-215. Due to the inclusion of I-*SceI*, MTV-10 was designed to be superior to pRF-215 [169]. Finally, the efficacy of the I-*SceI* plasmid-read elimination strategy was tested.

In short, we aimed to answer these three questions:

- Which genes are needed for growth on rich media BHIS?
- Which genes are needed for sporulation?
- How effective is the I-*SceI* plasmid-read elimination strategy?

5.3 Results

5.3.1 Library creation

The objective of this experiment was to determine the essential genes needed for sporulation in *C. difficile* PCR ribotype 078. To evaluate these genes, a transposon library was created. In the first chapter it was established that CD9301 is an optimal representative strain to perform TraDIS in *C. difficile* RT 078. In the follow up chapter the aTc inducible transposon vector pMTL-MtV10 was constructed as an alternative to the established pRF215. Using the conjugation *E-coli* donor strain sExpress pMTL-MtV10 and pRF215 were transformed into CD9301. Transconjugant colonies were selected on BHIS supplemented with CC, Em and Tm with concentrations according to Table 2.4. After 48 hours, 3 emerging colonies per plasmid were harvested and incubated in 10 mL BHIS broth supplemented with CC, Em and Tm and incubated at 37°C in the anaerobic cabinet overnight. Subsequently, the transconjugant cultures were incubated in fresh 10 mL BHIS supplemented with CC, Em and Tm and left to grow until an OD of 0.05 was reached. Afterwards, colonies were serial diluted and incubated on BHIS agar, specifically supplemented to assess the libraries qualities. Subsequently, 100 µl of the remaining transconjugant cultures were transferred to mutant selection plates

containing CC, Em, and aTc. Dembek et al, harvested the growth on mutant selection plates after 18 hours [86]. During the experimental set up performed in this chapter, the growth was harvested after 48 hours, as it is anticipated that slow growing mutants delay colony formation. Growth was harvested by applying 2 mL of BHIS broth supplemented with CC, Em and aTc per plate and detaching colonies with a scraper. The resulting cell suspension was harvested in 3 separate 50 mL BHIS broth supplemented with CC, Em and aTc. Freezer stocks were created in 10% DMSO and stored at -80°C to serve as inoculum in future outgrowth experiments. Genomic DNA of 2 mL bacterial culture was obtained via phenol-chloroform extraction, detailed in materials and methods.

The OD of the cell suspension was determined and freshly inoculated to an OD of 0.1 into 50 mL of BHIS, supplemented with Em and aTc. After 24 hours, 2 mL of inoculate was harvested per 50 mL falcon tube containing the *C. difficile* TraDIS library of essential genes, produced via pMTL-MtV10 or pRF215. The optical density of the remaining pMTL-MtV10 and pRF215 libraries was determined and inoculated to an OD600 of 0.1 into fresh BHIS medium, containing Em and aTc to initiate the sporulation library. Sporulation of the cultures was induced by nutrient starvation of these cultures for 14 days in the anaerobic workbench. After 14 days, the samples were harvested, and spores were isolated. Genomic DNA was extracted via ultrasonication of the spores, followed by phenol-chloroform extraction, detailed in materials and methods.

5.3.2 Comparison between pRPF-215 and pMTL-MTV10.

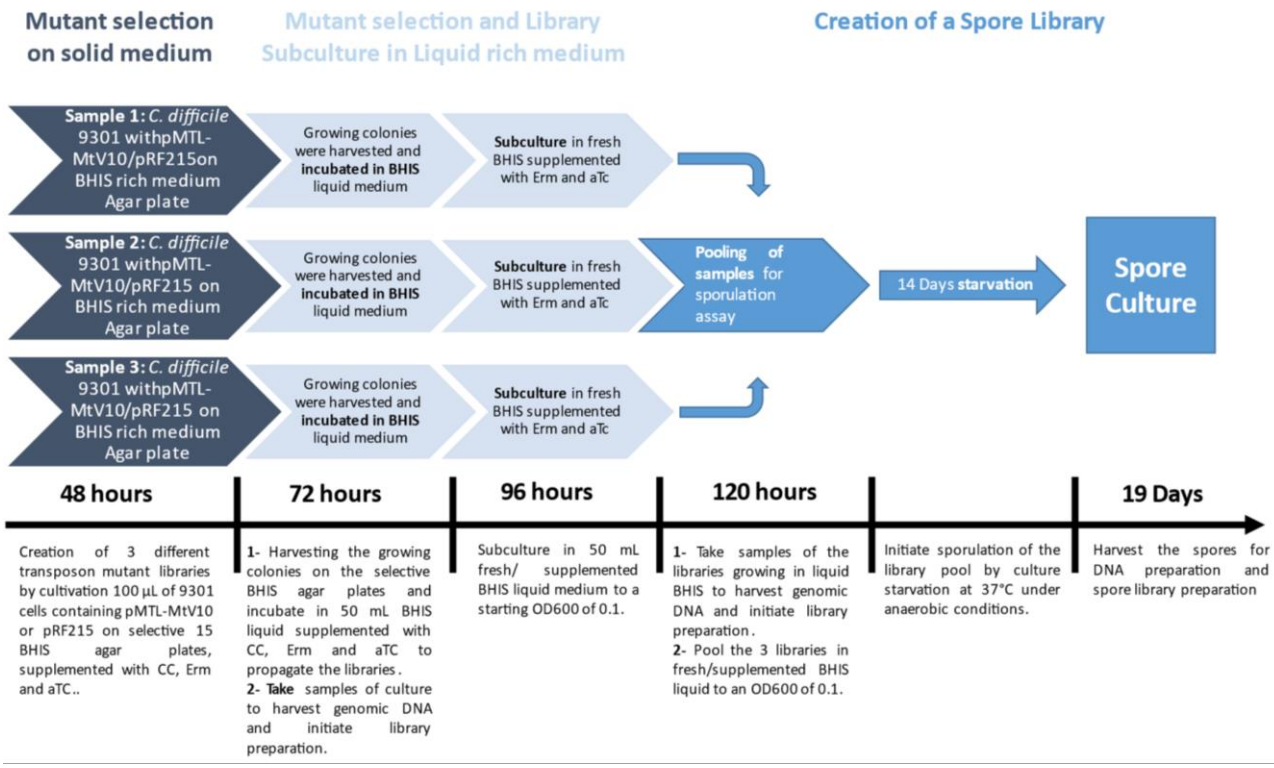


Figure 5.3: Construction of a TraDIS library in *C. difficile*, using either pMTL-MtV10 or pRF215. Initially *C. difficile* cells are conjugated with *E. coli* donor cells, containing rather pRF215 or pMTL-MtV10. Cells indicating successful conjugation were harvested and proliferated in liquid culture. Transposon mutagenesis was induced via cultivation on BHIS plates, supplemented with Em and aTc. After 48 hours, growth was harvested using a cell scraper and further inoculated in 50 mL of fresh BHIS containing Em and aTc to select for transposon mutants. The TraDIS library for essential genes was generated by inoculating the mutant culture in fresh liquid rich BHIS medium, supplemented with Em and aTc. After 20 hours of incubation at 37°C in the anaerobic chamber, samples were harvested for gDNA isolation and subsequent TraDIS library preparations. Further the remaining cultures were pooled and transferred to fresh BHIS containing Em and aTc. The cultures were maintained at under anaerobic conditions for 14 days to induce sporulation of the cells by starvation. After 14 days, samples were harvested for DNA isolation and

library preparation of the *C. difficile* RT 078 library of essential genes involved during sporulation.

Table 5.1 presents the mapping statistics for the transposon libraries created with the two different transposon-delivery plasmids. The incorporation of the I-*SceI* site and subsequent application of I-*SceI* into the sequencing library preparation was intended to give a reduced percentage of reads mapping to the plasmid. This is the case in the rich media samples, but in the spore library there was a very large percentage of reads which mapped to the delivery plasmid. There was clearly an issue with the library preparation either during the microbiology stage or at the sequencing library stage because for the pRPF-215 library there was a very low percentage of reads mapping to the genome too (24.4%). This could be an effect of the difficulties of preparation genomic DNA from spores, or because of inaccuracies in the sequencing library preparation.

Table 5.1: Mapping statistics of the TraDIS libraries created with pMTL-MtV10 and pRF215. The Transposon library samples that have been created according to the protocol described in Figure 5.2. Using the TraDIS toolkit analysis pipeline [98] the samples have been mapped to the CD9301 genome.

Sample	Genome (%)	Delivery Plasmid (%)	Other (%)
pMTL-MtV10 samples in BHIS combined libraries	90.1	1.1	8.7
pRPF215 samples BHIS combined libraries	91.1	5.4	3.4
pMTL-MtV10 spore library	6.1	75.9	18.0
pRPF215 spore library	24.4	5.2	70.4

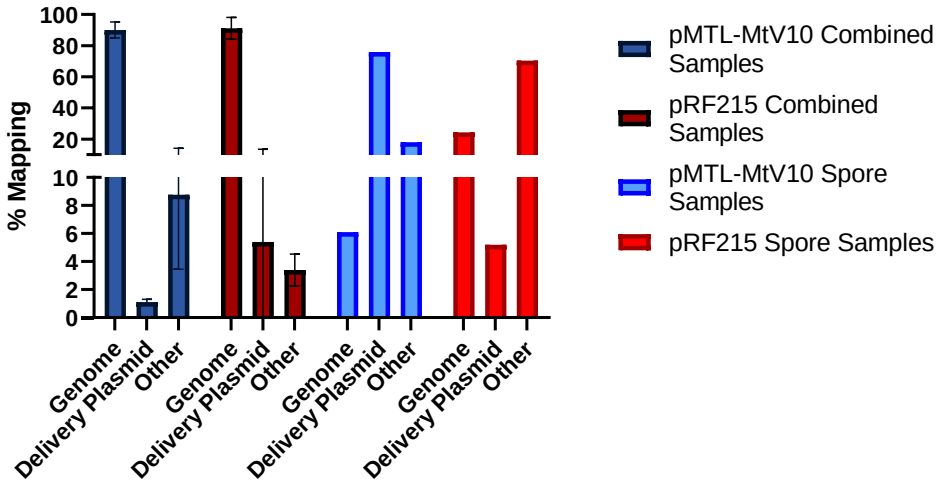


Figure 5.4: Mapping statistics of the TraDIS libraries. Graphical visualization of the values determined in Table 5.1. The standard deviation of the individual values has been evaluated using GraphPad Prism version 8.4. The SD for the individual samples is determined as: For pMTL-MtV10 samples in BHIS mapping to the genome (5.12), delivery plasmid (0.2), other (5.32). For pRPF215 samples BHIS combined libraries mapping to the genome (6.95), delivery plasmid (8.05), other (1.13). To obtain the spore libraries, the three individual essential gene Libraries have been pooled into one sample, according to the protocol shown in Figure 5.2. Thus, no standard deviation can be determined.

Table 5.2: Mapping statistics of the TraDIS library achieved by Dembek et al, in *C. difficile* 630*Δerm* and R20291 [86]. Extended table of the TraDIS library mapping statistics towards plasmid read contaminations, already presented in Table 4.7.

	Total Reads	Reads Mapped	%Mapped
C. difficile* strain 630*Δerm			
Chromosome	32650314	29279004	89.67
Plasmid		2088244	6.39
***C. difficile* strain R20291**			
Chromosome	171969196	123057426	71.55
Plasmid		37866472	22.02

It is worth noting that in our experiment the plasmid reads for pRPF215 were significantly lower than in the previously published work. The reason for this is not

clear, but it could be due to the difference in biological transposon library creation technique. In the previously published work [86] a single transconjugant colony was taken and grown in liquid culture to amplify the plasmid-containing pool. This is different to the approach we have taken whereby all transconjugant colonies were pooled. It is also entirely possible that this difference has occurred by chance. The stochastic nature of the results is evident with the plasmid-read percentage diverging wildly between 2.9% and 22% without any obvious reason for it.

It is difficult to come to any firm conclusions as to the usefulness of pMTL-MtV10 over pRPF215 from the data described here. If pRPF215 had not been used alongside pMTL-MtV10 in the experiments described here, then the observation of a low plasmid-read value of 1.1% for pMTL-MtV10-created libraries would appear to be an outstanding result compared to the previously published values of 6-22%. However, a better comparison is the value of the pRPF215 results that were obtained alongside our pMTL-MtV10 results which were insignificantly different. So far, the combined data for the rich media samples has been discussed, but it is worth looking at the individual biological samples as shown in Table 5.3 where it can be seen that in one of the pRPF215 libraries there is a much higher domination of plasmid reads (14%). Perhaps the domination of plasmid-containing cells in biological libraries is an effect which occurs only semi-regularly and when it does the I-*SceI* site strategy could prevent those reads taking up valuable sequencing.

Table 5.3: Individual mapping statistics of the TraDIS libraries, created using pRF215 and pMTL-MtV10 in *C. difficile* strain 9301. The mapping statistics from the individual essential gene libraries set ups in BHIS, as described in Figure 5.2 prior to pooling (Table 5.1). Statical analysis using the two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli in GraphPad Prism Version 8.4 resulted with no significant statistical difference between the plasmids: Genome (p-Value: 0.84) Delivery Plasmid (p-Value: 0.40) Other (p-Value: 0.16).

Sample	Genome (%)	Plasmid (%)	Other (%)
pMTL-MtV10-e1	84.1	0.9	14.9
pMTL-MtV10-e2	93.2	1.3	5.4
pMTL-MtV10-e3 + I*SceI*	92.9	1.1	6.0
pRF215-e1	95.1	1.0	3.9
pRF215-e2	95.2	0.5	4.2
pRF215-e3 + I*SceI*	83.1	14.7	2.1

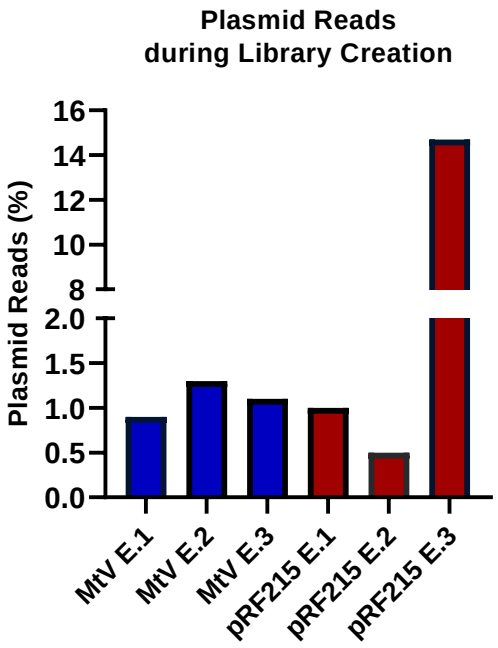


Figure 5.5: Graphic visualization of the Transposon to plasmid mapping statistics of the essential gene libraries. The TraDIS libraries were created using pRF215 and pMTL-MtV10 in *C. difficile* strain 9301 according to the protocol described in Figure 5.2. Sample MtV-E.3 and pRF215-E.3 were additionally digested with I-*SceI* during the library preparation (Figure 5.1).

A final point to note on the plasmid-read mapping is that the I-*SceI* strategy of MtV-10 did not appear to work for the spore sample where 75.9% of the reads mapped to the genome. It is not clear why this occurred since the I-*SceI* digest should have cleaved the DNA rendering the subsequent PCR unable to amplify products from plasmids as detailed in Chapter 4. Further, no difference was observed in the DNA samples obtained from the essential gene Library in BHIS processed with or without I-*SceI* (Table 5.3).

5.3.3 Essential genes for rich media.

The data from six biological repeats was combined from both the pMTL-MTV-10 and pRPF-215 transposon pools. This file was used to determine the essential gene set for growth on rich media. The TraDIS toolkit analysis pipeline was used to analyse the sequencing data. Reads were mapped to the *C. difficile* M120 genome (NC_017174.1). The mapping statistics are shown in Table 5.4. These results appeared promising, especially the 126,079 unique insertion sites.

Table 5.4: Mapping statistics of the essential gene set for growth on rich BHIS media.

Sample	Total Reads	Mapped to Genome (%)	Unique Insertion Sites in the Genome	Average base pairs between insertion sites
Combined rich media	156,922,014	91.8	126,079	32.1

Gene-level analysis was then performed. This involved calculating an insertion index for each gene which is the number of insertions found in that gene divided by the gene length. The insertion index allows the prevalence of transposon hits between genes to be compared since it normalises for gene length. The insertion indexes for all the genes in the genome are then compared and from that comparison a set of essential genes can be predicted. All these functionalities can be achieved using the TraDIS toolkit.

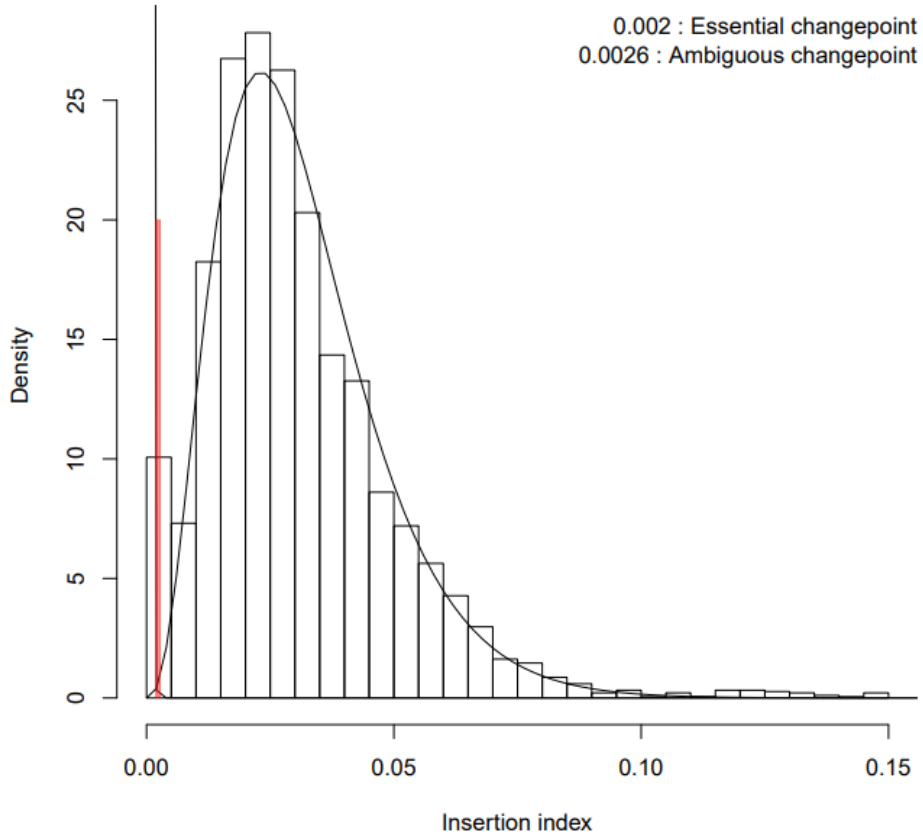


Figure 5.6: Histogram of the distribution of Transposon insertions obtained in this study. The non-essential gene normal distribution is shifted left compared to the example data in Figure 5.7.

Despite the relatively high number of unique insertion sites, the TraDIS_essentiality.R script predicted only 172 essential genes which is far lower than what was expected and much less than what is biologically likely. The essentiality prediction script can also deem a gene to be ambiguous in its essentiality status, but only one gene was called as ambiguous. Figure 5.6 shows a histogram of insertion index with the essential and ambiguous cut-offs marked. On this figure, the essential genes are expected to be represented on the left-hand side as a peak with a low insertion index, while one would expect non-essential

genes to be clustered in a distribution to the right of that with a smaller number of genes in the ambiguous zone between. This figure explains why the TraDIS toolkit script predicted the unexpectedly low number of essential genes. The non-essential gene distribution largely merges with the left-hand distribution which should represent essential genes. The script clearly does not deal with this situation and this has resulted in the unrealistic number of predicted essential genes.

This is best understood by comparison with a histogram of some example data provided by the tutorial document in the TraDIS toolkit GitHub in Figure 5.7. In this figure there is a very clear gap between the right-hand distribution (non-essential genes) and the left-hand distribution.

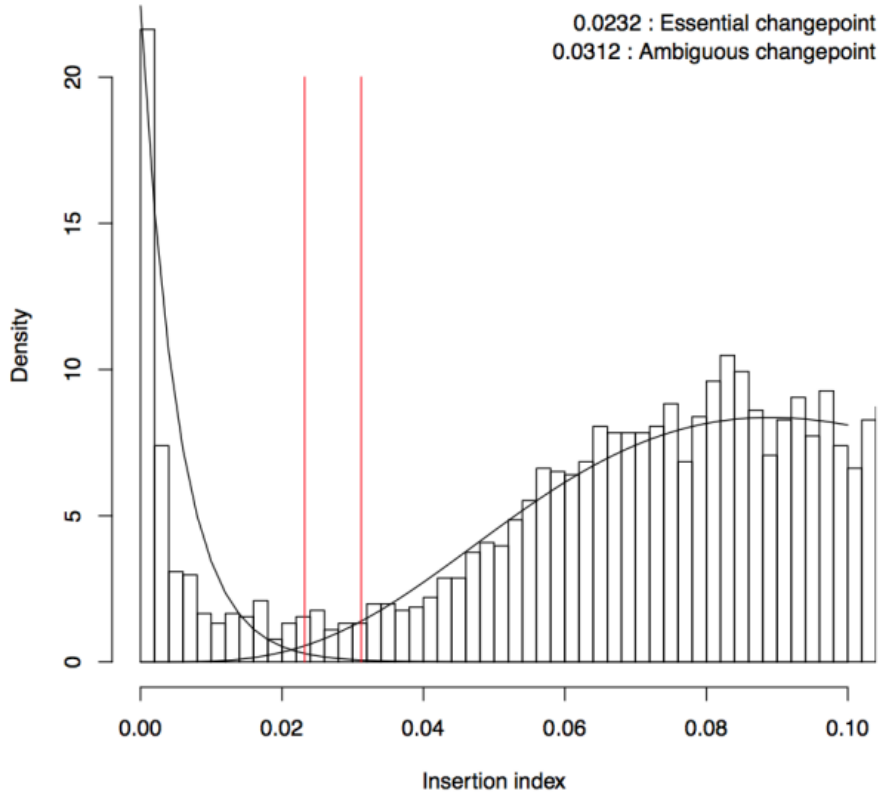


Figure 5.7: Histogram representing the expected normal distribution of a saturated Transposon library. As referred to in the data from the TraDIS toolkit paper(https://academic.oup.com/bioinformatics/article/32/7/1109/1744349?login=true). Gene essentiality is predicted *in silico* using the tradis_essentiality.R. This script fits distributions towards the insertion indexes (insertion sites divided by gene length) with the insertion density of genes. This produces tables of putative essential and ambiguous genes, along with plots that can be used to evaluate the predictions. This plot is showing the automated fit, along with thresholds for essentiality and ambiguity.

The TraDIS toolkit paper contains a supplementary information document which describes a situation analogous to what was observed in this study. It states that if a library is not sufficiently saturated then the two distributions collapse to the

left and all that can be done is to call only those genes which completely lack insertions as essential. This is what has occurred here, as can be seen in Figure 5.6 which is another histogram of the data obtained in this study, with a much smaller bin size. It is clear that the only genes being attributed as essential by the script are those with zero insertions.

"Note that in cases where a library is not sufficiently saturated, the two distributions will collapse to the left, preventing a reasonable prediction of essential genes. An example of this can be seen in the analysis of a Clostridium difficile 630 library in Dembek et al., mBio 2015. In this case, the best that can be done is to report genes completely lacking insertion sites, with the understanding that some of these will certainly be "false positive" [98].

However, the explanation that the library created in the present work is insufficiently saturated may not be the best one, especially considering the rather high number of unique insertions (126,079). An alternative explanation is that transposition events were continuing to occur in the library due to the continued presence of the transposon delivery plasmid. In this scenario essential genes which previously lacked insertions could have acquired a transposon sufficiently late in the experiment to render them unable to be counter-selected. The histogram with two distributions only occurs if the essential genes are unable to compete in the

culture. An essential gene may move along the histogram by a late acquisition of a transposon.

Before giving up on this rich media data entirely an attempt was made to define an essentiality cutoff manually. From the histogram Figure 5.8 and given that one can expect the number of essential genes to be similar to that predicted in [86], a manual cut-off was used at 0.0125 as indicated with the dashed red line on the figure. This cutoff produced 448 predicted essential genes, which is a more realistic number. This list of essential genes can be found in table 7.1 in the appendix. This approach will inevitably produce an abundance of false positives and false negatives and so should only be used tentatively.

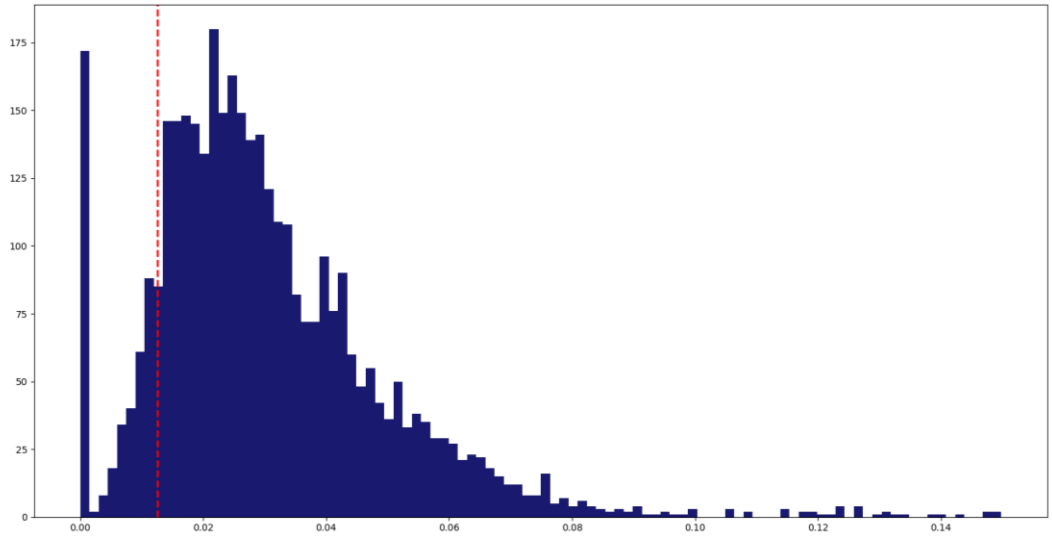


Figure 5.8: Adapted Histogram Transposon insertions obtained in this study. Selecting a manual cut-off at 0.0125

5.3.4 Genes required for sporulation

In order to predict the essential gene set for sporulation all the data from gDNA from spores was combined and analysed. As previously discussed, a much lower percentage of reads from these samples successfully mapped to the genome. The depth of sequencing, however, was still sufficient to identify 62,423 unique insertion sites. This coverage is unlikely to give accurate essentiality predictions with an average distance between insertion sites of 65 base pairs. The same TraDIS toolkit pipeline was followed for this data regardless and unlike the rich media data did succeed in producing a realistic number of predicted essential genes – 398 genes out of a total of 3,694 genes in the genome (11%). This number is believable, since it is in line with previous results in *C. difficile* [86]. However, it is likely to have missed many essential genes too as this number is lower than efforts into a minimal genome have managed to produce [172]. The list of predicted essential genes for sporulation is provided in Table 7.2 in the appendix.

Table 5.5: Mapping statistics of the essential gene set for *C. difficile* RT 078 Sporulation

Sample	Total Reads	Mapped to Genome (%)	Unique Insertion Sites in the Genome	Average base pairs between insertion sites
Combined spore samples	60,854,152	38.3	62,423	64.8

5.4 Conclusions

Evidence that the incorporation of I-*SceI* sites into pMtV-10 has improved its suitability for use in TraDIS was not obtained in this study. Reasonable transposon libraries were generated which had a large number of unique insertions in a population grown on rich media. This number of unique insertions would usually be sufficient to make good essentiality calls. However, the non-essential gene right-hand distribution generated in the TraDIS toolkit pipeline appears to be left-shifted which makes essentiality predictions difficult. The reason for this shift is either because of insufficient time for cells with insertions in essential genes to be outcompeted (and therefore out represented), or because of an active transposase moving transposons from non-essential regions into essential regions as growth occurs. This latter point is an unfortunate consequence of having a replicative vector containing the transposase still in a significant number of cells. Nevertheless, this theory is very unlikely as although movement of the transposon from the plasmid to the genome is highly efficient, subsequent movement from the chromosome is much less efficient. This can be explained as the nature of the mini-Tn in the plasmid (which is a small highly supercoiled CCDNA molecule) supports cut and paste better than the chromosomal DNA.

The data generated in this chapter provides some preliminary insights, but it is also clear that much better data should be generated with much the same approach and that data would warrant the large amount of manual curation and thorough bioinformatic investigations that would be needed. It may be the case that merely using a HiSeq instrument as the previous *C. difficile* transposon sequencing publication did rather than the relatively low-throughput MiSeq that was used here would mask any problems with plasmid reads and would generate sufficient data for firm essentiality predictions. Any possibility of taking this approach was ruled out by the constraint on working conditions imposed by the pandemic.

Chapter 6

Final Discussion

6.1.1 *C. difficile* RT 078, an emerging threat to healthcare systems.

CDI is the leading cause of antibiotic-associated diarrhoea, especially in immunocompromised patients. *C. difficile* can hyper-colonize the guts and cause major infections. CDI is a significant burden for healthcare providers around the world due to patients needing to stay for extended periods within a clinical setting, which in turn also increases treatment costs [12]. Presently it is reported that CDI has overtaken cases of *MRSA* infections in healthcare environments, such as hospitals and nursing homes [17]. The lack of gold-standard diagnosis techniques for CDI, as well as the rapid evolution of the organism's epidemiology, makes appropriate treatment difficult.

Outbreaks of CDI are mainly attributed to RT 001, 017, 027 and 078. RT 027 is common in hospital settings [8, 24-26], but in recent years, hyper-virulent RT 078 strains have become more prevalent [32].

6.1.2 Improving DNA transfer into *C. difficile* ribotype 078

Genetic procedures such as CRISPR or TraDIS are considered an effective means of obtaining insights into the emerging *C. difficile* PCR ribotype 078 [65]. One challenge in establishing these protocols is that the mechanism of DNA transfer in *C. difficile* remains poorly understood [74, 115]. In this study, we analysed the

genomes of 10 patient-derived PCR RT 078 strains using Illumina and Nanopore sequencing and subsequent genome assembly. It has been established that these strains have a high genetic similarity of up to 99.99%, as explained in Chapter 3. Based on these phylogenetic similarities, it is expected that all strains utilize related restriction modification systems and thus have a comparative potential to uptake DNA via conjugation. Methylome analysis reveals that although strains of RT 078 are genetically similar, they express a variety of restriction barriers. In the literature, *C. difficile* type M120 is considered representative of PCR RT 078, so most of the research has focused on this strain. Table 3.3 shows that M120 deploys several R-M barriers with diverse recognition sequences. This complicates further implementation of genetic protocols such as CRISPR, TraDIS and ACE. To enhance our understanding of the organism and thus the entire ribotype, it is of crucial importance to improve DNA transfer into *M120*.

The remaining RT 078 strains analysed in this study express the Type-IV R-M system, which is circumvented by the conjugational donor strain of *E. coli* known as 'sExpress' [126]. This improvement in DNA transformability makes them more suitable for the introduction of plasmid-borne tools such as CRISPR or TraDIS. In Chapter 3, we established that type 93-01 represents an optimal strain for further analysis of the PCR RT 078 strains. The experimental set-up showed improved DNA transformability. Increased DNA transfer into *C. difficile* facilitates successful downstream genetic manipulations in this organism.

In 2019, Woods et al. [102] demonstrated that the circumvention of R-M barriers in *Clostridium* species is a promising approach to enhance DNA transfer. By creating a conjugational donor strain, sExpress, they bypassed the Dcm sites (CCWGG), and thus cytosine-specific Type IV methylation in diverse *Clostridium spp*. Eventually this technique improved DNA transfer into *C. autoethanogenum*, *C. sporogenes* and *C. difficile* R20291, a RT 027 strain. The experiments conducted in Chapter 3 of this thesis showed that sExpress did indeed increase conjugation in *C. difficile* PCR RT 078, including the M120 strain. However, the methylation systems deployed in M120 (Table 3.9) represent an additional methylation barrier during DNA transformation into the strain. The M120 orphan methylase M.CdiM120ORF3590P recognizes CAAAAA and has been shown to lack a restriction component and thus does not represent any R-M barrier [173]. However, this work introduces M1.CdiM120ORF10140P and M2.CdiM120ORF10140P, which recognize ACGGC. By creating conjugational donor cells that circumvent these R-M type II systems, DNA transfer into M120 is potentially increased.

Only R-M type IV systems are known to recognize foreign modified DNA, whilst systems I – III are expected to target non-methylated DNA [107]. Thus, methyltransferases targeting ACGGC (M1.CdiM120ORF10140P) need to be expressed in the sExpress donor. Constructing a plasmid-derived vector expressing

this enzyme may result in the optimal DNA modifications needed to circumvent not just R-M Type IV but also Type II systems, resulting in a potentially super-conjugant *E. coli* donor cell for PCR RT 078.

An alternative to using a super-conjugant donor may be DNA treatment with a methyltransferase. This mimics the *C. difficile* methylation pattern *in vitro* prior to DNA integration. However, performing such *in vitro* DNA methylation requires knowledge of the R-M barriers present within the respective target strains. In this case, the availability of such commercially available methylases may be restricted to well-understood organisms and not be disposable for less-studied bacterial strains.

A possible solution to this problem may be to use an intermediate host bacterium, which *in vivo* methylates the target DNA according to the pattern required in *C. difficile*. Re-extracting the methylated DNA from the intermediate host enables it to be transformed into *C. difficile* via electroporation. This method has been successfully established in *C. acetobutylicum*, by integrating the *B. subtilis* phi 3T methyltransferase into an *E. coli* intermediate host. *In vivo* methylation of the target plasmids in *E. coli*, followed by DNA extraction and subsequent electroporation into *C. acetobutylicum*, result in increased frequencies of transformation [68, 121].

Thus, both *in vivo* and *in vitro* methylation, and subsequent electroporation or chemical integration of modified DNA, represent a potential alternative to using a super-conjugant donor strain.

Despite these advances, implementing effective DNA transformation using chemical or physical stressors within *C. difficile* remains difficult. In 2020, Bhattacharjee et al. successfully introduced an electroporation protocol in the *C. difficile* strains R20291, CD630 and JSC10 [158]. However, neither transfer of DNA into *C. difficile* via electroporation, nor *in vitro* methylation of DNA have yet been fully understood and established [74, 115]. The conjugation method developed by Purdy et al [74] and improved by Kirk et al., still represents the most reliable method of DNA transfer into *C. difficile* [76, 77]. As this method allows simultaneous target-specific modifications of the methylation sites by the donor strain, it remains overall the gold standard for DNA transformation into *C. difficile*. Another option to increase DNA transfer into M120 may be to genetically modify the conjugational recipient. The activity of RM-Type II may be amended by using target-specific genetic approaches, such as CRISPR/Cas9. By engineering M.CdiM120ORF3590P or M2.CdiM120ORF10140P promotor- or operator-deficient M120 cell lines, their potential to detect foreign DNA may be reduced. Woods et al. have created a *C. autoethanogenum* knockout strain using CRISPR/Cas9-mediated genome editing. They targeted CLAU_0514, encoding for a fused Type II R-M enzyme, eventually creating a strain which is 10 times more

amenable to DNA transfer than the wild type. Although Wilkowska et al. successfully modulated the activity of an R-M Type II-specific promotor in *E. coli*, the group discovered that the bacteria R-M defence mechanism against invading phage DNA is reduced. This indicates that altering the R-M type II specific response may conflict with the underlying bacterial methylation balance [174].

In multiple *C. difficile* isolates of various origins, the Type II 6mA DNA MTase, termed '*C. difficile* adenine methyltransferase A (camA)', is constantly expressed, targeting the CAAAAA methylation motif. Inspection of the TraDIS library data of Dembeck et al., demonstrates that the CamA homologue of R202091 is reduces spore purification efficiency [86]. Inactivation of *camA* resulted in a sporulation defect in the reference strain CD630. Oliverei et al. [173] suggest that, based on its highly conserved nature, additional phenotypes may be regulated by CamA beyond sporulation. They observe that, especially in genes encoding for sporulation, membrane transport or mortality, CAAAAA sites are overrepresented. Furthermore, they discovered that the inactivation of *camA* causes alterations in the organism's transcriptional signatures, *in vivo* colonization and biofilm formation compared to the wild type. This suggests that the functional impact of R-M patterns goes beyond DNA transfer protection against invading species but that these patterns are involved in regulation of signalling cascades [173, 174].

It is hypothesised that an enhancing R-M system of Type II-specific operons potentially reduces DNA methylation and thus stabilizes DNA transfer in strain M120. However, targeted stable modifications of the R-M machinery by ACE or CRISPR in host strains may not be without its consequences. The intracellular complexity of DNA methylation is diverse and such modifications could have unpredictable unwanted phenotypic effects [173].

6.1.3 Establishing a TraDIS library for *C. difficile* PCR RT 078.

The work conducted in Chapter 3 suggests that, currently, conjugation with sExpress is the most suitable method to transfer DNA into *C. difficile* PCR RT 078. This resolves the most crucial barrier to implementing TraDIS in these strains. The next steps to successfully implement TraDIS involve finding a suitable transposon delivery tool and establishing a protocol resulting in optimal library generation and preparation. Rapid transposon liquid enrichment sequencing (TnLE-seq) is the method of choice for organisms with a limited genetic toolkit. Many of these transposon insertion sequencing (TIS) methods rely on library propagation using non-replicative suicide plasmids. After conjugation-mating, positive mutants are selected in a liquid selection stage, selecting for transposon mutants and counter-selecting against the *E-coli* donor. This allows for fast library proliferation by avoiding complications that can arise from agar-based manual selection methods.

Further, the selection of antibiotics selecting for transposons but not plasmid backbone allows plasmid loss and clearance of the active transposase from the target cell after transposition.

The failure of the TnLE-seq protocol utilizing the suicide vector system pDIG01/pDIG02 points instead towards using a plasmid-derived system. However, elimination of the transposase after transposition is still required to produce a stable mutant. With non-replicative transposon-delivery, the process of attempting to obtaining a mutant colony can be broken down into two basic phases, DNA transfer and transposase activity in the cell.

Transposon delivery via a replicative and inducible vector is more complex than using suicide plasmids; however, it does not produce high frequency DNA transfer. A single host cell, carrying the transposon delivery system, can be selected on an agar medium and proliferated via an additional liquid step.

In 2013, Fels et al. successfully conducted a liquid-based TIS experiment in which the transposon-delivery plasmid was unable to replicate in the conjugational *E. coli* donor, a diaminopimelic acid (DAP) auxotroph. The *E. coli* cells were then counter-selected by the absence of DAP [175]. The advantage of this experimental set-up was that the combination enrichment/competition phase was carried out entirely in liquid culture, minimizing the manual labour required. Even so, the

researchers were able to demonstrate TnLE-seq within *Desulfovibrio vulgaris*. The organism genome size of 3.773.159 bp is comparatively small and few transconjugants were needed to allow a genome-saturating library to be generated with no replicative plasmid using a single mating plate [175].

In recent years, several successful innovations in DNA sequencing and library preparation have pioneered the use of replicative plasmid systems for less transformable organisms [86, 176]. Le Breton et al. created a TIS library for *Streptococcus pyogenes*, using the temperature-sensitive replicon pW01. Just 6% of their overall reads could be assigned to the bacterial chromosome; instead, most reads mapped to the delivery plasmid [177]. Only after continuous passaging of the library in non-permissive conditions were the plasmid reads mitigated. Yet this still has the potential to form the biological bias of the library by promoting mutants that gained a fitness advantage. Thus, Le Breton et al. concluded that an effective conditional replicon is essential to eliminate replicative transposon delivery systems from an organism [177].

Several plasmid delivery systems utilizing transcriptional read-through into a conditional replicon have been developed for *Clostridium spp.* [86, 154]. Zhang et al. [154] produced a conditional replicon-controlling plasmid maintenance system via transcriptional read-through, induced by an IPTG sensitive promotor. This system has been exemplified in *C. acetobutylicum* and *C. sporogenes* and is

potentially capable of being implemented in any *Clostridium* species. Our group demonstrated that a replicative plasmid delivery vector consisting of a conditional replicon coupled with an inducible promotor system is an optimal tool to achieve randomized transposition and subsequent plasmid loss in *C. difficile.* The plasmid employs the P_{tcdB} promoter system, activated via alternative sigma factor TcdR, present on the host genome. Importantly, P_{tcdB} is not active in the *E. coli* donor strain but in a multitude of clostridia [91, 154].

Chapter 4 describes how various inducible promotor systems were analysed in the *C. difficile* PCR RT 078 strain CD9301, based on their ability to induce optimal transposition frequencies and plasmid loss ratios. The lactose-inducible *bgaR*-P_{bgaL}, developed in *C. perfringens* and the IPTG-induced P_{fac} promoters, did not express qualities sufficient for TraDIS, possibly due to the dissipitation of lactose from the system [150]. The Tc-inducible systems, using the divergent P_{tet}-P_{tetR} in pMTL-CW21 and pMTL-CW22, were shown to be functional in *C. difficile* but neither plasmid loss efficiency nor transposition frequency was optimal to establish a saturated TraDIS library. The persistent plasmid retention is expected to be linked to sequence homology of *ermB* on the *C. difficile* genome with the plasmid-derived *ermB* and potential plasmid integration [161, 162]. Further PCR analysis, targeting plasmid specific DNA sequences, may support this theory. Dembek et al. [86] have successfully created a TraDIS library, using the replicative vector pRF215 in *C. difficile* strains 630Δ*erm* and R20291. In pRF215, plasmid loss

via transcriptional read-through and Himar1C9 activation are directed by a deviant Tet-inducible system to the Nottingham constructs [86].

Transforming pRF215 into strain CD9301 resulted in increased plasmid loss efficiency, transposition frequency and randomized insertion into the host's genome. However, after library sequencing of the two *C. difficile* strains, up to 22.01% of the total reads could be assorted to pRF215, with the majority of plasmid loss evaders revealing transposon insertion into RFtetR [86].

In 2013, Fels et al. introduced the inclusion of an additional BgIII digestion step during the library preparation, prior to sequencing to delete plasmid backbone reads [175]. More recently, Charbonneau et al. (2017) have conducted a TIS*,* similar to the experimental set-up in this thesis. A transposon plasmid, containing a conditional replicon, was transformed in *Streptococcus equi* cells. Plasmid harbouring cells were plated on non-permissive conditions and selecting for the presence of the transposon. During the library preparation stage, the adaptor-ligated DNA fragments were exposed to SmaI, to remove plasmid reads similar to the I-*SceI* step deployed in Chapter 4 [176]. Based on this improved method of plasmid clearing, the novel conjugational donor pMTL-MtV10 was constructed. This plasmid utilizes the divergent promotor system, transposase and *ermB* transposon from pRF215. However, it is additionally equipped with an I-*SceI*

digestion sites, to achieve optimal plasmid clearance based on Fels et al. [175, 176].

The initial attempts to create a TraDIS library using pMTL-MtV10 and pRF215 resulted in insufficient identification of essential genes for RT 078 survival in rich medium and for genes involved in sporulation. Just the adaptation of a manual cutoff, as described in the TraDIS toolkit [98] caused an approximate calling of a number of essential genes. The severe impact of the COVID19 pandemic and the subsequent downshift of the feasible laboratory routines caused irretrievable retardation of the final TraDIS library production. The work outlined in this thesis contains the initial efforts and assay methods to evaluate the TraDIS library potential of pMTL-MtV10 and pRF215. Unfortunately, COVID inflicted lack of contingency forced us to repurpose these initial tests to create the full-scale transposon library, presented in Chapter 5.

For further tests of the plasmid constructs in RT 078, the full workforce of ours and the collaborators laboratory facilities needs to be restored. Once the daily routines necessary to conduct each step of the TraDIS pipeline, are back in place, the true potential of pMTL-MtV10 and pRF215 to create a thorough transposon library can be assessed rigorously.

6.1.4 Future prospects

Once the transposon library for *C. difficile* strain CD9301 has been successfully generated, a multitude of applications may be created with it. A major problem in combating CDI is the current reliance on antibiotic therapies. Over time, varieties of *C. difficile* have accumulated antimicrobial resistance, such as against fluoroquinolones and clindamycin. Sholeh et al. have reported recently that Tc resistance is present in one fifth of human clinical *C. difficile* isolates.

*C. difficile´*s increasing accumulation of antimicrobial mechanisms makes the task of developing CDI treatments more challenging [178]. Plating a TraDIS library on BHIS agar plates, supplemented with clinically interesting antibiotics, leads to anomalous colony formation. These colonies carry a transposition event, allowing them to evade antimicrobial killing. Analysing the insertions of such mutants allows researchers to infer the involvement of specific genes during antimicrobial resistance. This makes them interesting targets to develop novel therapeutics against multi-resistant CDIs.

C. difficile is an important human pathogen. However, little is known about the bacterium's adaptive abilities against stress factors. Strain CD630 has been proven to withstand clinically relevant heat stress. Jain et al. have revealed that several proteins are significantly modulated under mild heat stress [179]. The upregulation of several hydrogenases and oxidoreductases suggests a modified

electron flow between the enzymes. In a similar vein, Giordano et al. cultivated the anaerobic *C. difficile* strain CD630 in a modified hypoxic chamber, exposed to various concentrations of oxygen. They observed significant growth of the bacterium in up to 2% oxygen. Twenty-five genes were identified as those most induced during growth in 2% oxygen, many of which are hypothetical proteins [180].

Using a TraDIS library may be a promising approach to identify genes involved in both *C. difficile* heat and oxygen stress responses. It is especially interesting if the specific gene knockouts result in similar fitness advantages during exposure to various stress factors. This data may provide an insight into *C. difficile* adaptation of physiology and metabolism while reacting to the altering host environment, for example during pyrexia.

6.2 Final Conclusion

This thesis elucidates the different stages of trying to implement transposon mutagenesis in a *C. difficile* RT 078 strain, starting from enhancing DNA transfer into the organism, to choosing the correct transposon delivery system and then implementing library preparation and sequencing optimization.

In the course of this research, a deeper understanding of DNA transformation into *C. difficile* PCR RT 078 was achieved. The genetic diversity and complexity of different strains of the same ribotype have been illuminated. Based on the complexity of its employed R-M systems, the RT 078 reference strain M120 is considered a challenging strain for the experimentalist. In this thesis, the clinical isolate-derived RT 078 strain CD9301 was selected as a suitable target to establish further genetic procedures.

Using this novel strain for implementation of TraDIS revealed that the current gold standard of employed TraDIS vectors in *C. difficile* suffers from disruptive plasmid retention. By introducing an additional I-Sce1 restriction side into the vector backbone, this work has provided a potential improvement to the performance of TraDIS in *C. difficile*. Our current data does not result in the identification of essential genes under different conditions. However, the suggested improvements promise to yield an optimal protocol to establish superior transposon libraries in *C. difficile* PCR RT 078. Overall, TraDIS represents a quicker and cheaper tool to elucidate unknown mechanisms of *C. difficile* adaptation, resistances and pathogenesis compared with the labour-intensive classical *C. difficile* reverse genetic approaches.

Altogether, this work has provided a platform for making progress in understanding the genetic modes of actions in the emerging yet inadequately

understood *C. difficile* PCR RT 078 strains. The developments made can be transferred to other clinically relevant *C. difficile* strains such as RT 027 or 014. Eventually, these improvements will help to further the understanding of the pathogen and contribute to increased patient wellbeing while reducing healthcare burdens.

7. Appendix

Table 7.1: List of essential genes in *C. difficile* PCR RT 078 in rich media. This data has been produced by manually defining the essentiality cutoff to 0.0125. Using this method 448 essential genes are predicted, including false positives and negatives.

locus_tag	gene_name	read_count	Fcn
CDM120_RS00005	*dnaA*	0	chromosomal replication initiator protein DnaA
CDM120_RS00015	*yaaA*	6	S4 domain-containing protein YaaA
CDM120_RS00125	CDM120_RS00125	3	
CDM120_RS00180	CDM120_RS00180	0	
CDM120_RS00205	CDM120_RS00205	0	
CDM120_RS00210	CDM120_RS00210	0	
CDM120_RS00305	CDM120_RS00305	0	
CDM120_RS00310	CDM120_RS00310	0	
CDM120_RS00505	CDM120_RS00505	724	PTS sugar transporter subunit IIA
CDM120_RS00540	CDM120_RS00540	7	2-C-methyl-D-erythritol 2
CDM120_RS00600	*rpmG*	0	50S ribosomal protein L33
CDM120_RS00605	*secE*	1	preprotein translocase subunit SecE
CDM120_RS00615	*rplK*	10	50S ribosomal protein L11
CDM120_RS00655	*rpsG*	7	30S ribosomal protein S7
CDM120_RS00670	*rpsJ*	9	30S ribosomal protein S10
CDM120_RS00685	*rplW*	4	50S ribosomal protein L23
CDM120_RS00695	*rpsS*	14	30S ribosomal protein S19
CDM120_RS00715	*rpmC*	4	50S ribosomal protein L29
CDM120_RS00745	*rpsH*	12	30S ribosomal protein S8
CDM120_RS00775	*secY*	26	preprotein translocase subunit SecY
CDM120_RS00785	*map*	10	type I methionyl aminopeptidase
CDM120_RS00790	CDM120_RS00790	11	KOW domain-containing RNA-binding protein
CDM120_RS00795	*infA*	2	translation initiation factor IF-1
CDM120_RS00810	*rpsK*	8	30S ribosomal protein S11
CDM120_RS00870	CDM120_RS00870	0	
CDM120_RS00895	CDM120_RS00895	0	
CDM120_RS00910	CDM120_RS00910	0	
CDM120_RS00915	CDM120_RS00915	0	
CDM120_RS00940	CDM120_RS00940	0	

CDM120_RS00960	CDM120_RS00960	0	
CDM120_RS00970	CDM120_RS00970	0	
CDM120_RS00980	CDM120_RS00980	0	
CDM120_RS01065	CDM120_RS01065	0	
CDM120_RS01070	CDM120_RS01070	0	
CDM120_RS01205	CDM120_RS01205	555	PTS lactose/cellobiose transporter subunit IIA
CDM120_RS01810	CDM120_RS01810	398	hypothetical protein
CDM120_RS01865	CDM120_RS01865	64	helix-turn-helix domain-containing protein
CDM120_RS02230	*fba*	31	class II fructose-1
CDM120_RS02250	CDM120_RS02250	0	helix-turn-helix transcriptional regulator
CDM120_RS02255	CDM120_RS02255	0	DNA adenine methylase
CDM120_RS02260	CDM120_RS02260	0	restriction endonuclease subunit M
CDM120_RS02265	CDM120_RS02265	0	hypothetical protein
CDM120_RS02270	CDM120_RS02270	0	AAA family ATPase
CDM120_RS02275	CDM120_RS02275	0	LlaJI family restriction endonuclease
CDM120_RS02280	CDM120_RS02280	0	CGNR zinc finger domain-containing protein
CDM120_RS02285	CDM120_RS02285	0	sigma-70 family RNA polymerase sigma factor
CDM120_RS02290	CDM120_RS02290	0	hypothetical protein
CDM120_RS02295	CDM120_RS02295	0	hypothetical protein
CDM120_RS02300	CDM120_RS02300	0	DUF2800 domain-containing protein
CDM120_RS02305	CDM120_RS02305	0	DUF2815 family protein
CDM120_RS02310	CDM120_RS02310	0	DNA polymerase
CDM120_RS02315	CDM120_RS02315	0	phage antirepressor
CDM120_RS02320	CDM120_RS02320	0	hypothetical protein
CDM120_RS02325	CDM120_RS02325	0	virulence-associated E family protein
CDM120_RS02330	CDM120_RS02330	0	VRR-NUC domain-containing protein
CDM120_RS02335	CDM120_RS02335	0	DEAD/DEAH box helicase
CDM120_RS02340	CDM120_RS02340	0	phage-associated protein
CDM120_RS02345	CDM120_RS02345	0	HNH endonuclease
CDM120_RS02350	CDM120_RS02350	0	P27 family phage terminase small subunit
CDM120_RS02355	CDM120_RS02355	0	methionine adenosyltransferase
CDM120_RS02360	CDM120_RS02360	0	site-specific DNA-methyltransferase
CDM120_RS02365	CDM120_RS02365	0	virulence protein
CDM120_RS02370	CDM120_RS02370	0	DUF4314 domain-containing protein
CDM120_RS02375	CDM120_RS02375	0	amidoligase family protein
CDM120_RS02380	CDM120_RS02380	0	gamma-glutamylcyclotransferase
CDM120_RS02385	CDM120_RS02385	0	DUF5049 domain-containing protein
CDM120_RS02390	CDM120_RS02390	0	terminase large subunit
CDM120_RS02395	CDM120_RS02395	0	hypothetical protein
CDM120_RS02400	CDM120_RS02400	0	phage portal protein
CDM120_RS02405	CDM120_RS02405	0	Clp protease ClpP

CDM120_RS02410	CDM120_RS02410	0	phage major capsid protein
CDM120_RS02415	CDM120_RS02415	0	hypothetical protein
CDM120_RS02420	CDM120_RS02420	0	head-tail connector protein
CDM120_RS02425	CDM120_RS02425	0	phage head closure protein
CDM120_RS02430	CDM120_RS02430	0	HK97 gp10 family phage protein
CDM120_RS02435	CDM120_RS02435	0	hypothetical protein
CDM120_RS02440	CDM120_RS02440	0	phage major tail
CDM120_RS02445	CDM120_RS02445	0	hypothetical protein
CDM120_RS02450	CDM120_RS02450	0	tape measure protein
CDM120_RS02455	CDM120_RS02455	0	phage tail family protein
CDM120_RS02460	CDM120_RS02460	0	phage tail protein
CDM120_RS02465	CDM120_RS02465	0	hypothetical protein
CDM120_RS02470	CDM120_RS02470	0	hypothetical protein
CDM120_RS02475	CDM120_RS02475	0	glycosyl hydrolase
CDM120_RS02480	CDM120_RS02480	0	phage holin family protein
CDM120_RS02485	CDM120_RS02485	0	N-acetylmuramoyl-L-alanine amidase
CDM120_RS02490	CDM120_RS02490	0	hypothetical protein
CDM120_RS02495	CDM120_RS02495	0	recombinase family protein
CDM120_RS02500	CDM120_RS02500	0	recombinase family protein
CDM120_RS02505	CDM120_RS02505	0	recombinase family protein
CDM120_RS02510	CDM120_RS02510	0	single-stranded DNA-binding protein
CDM120_RS02515	CDM120_RS02515	0	nucleotidyltransferase domain-containing protein
CDM120_RS02520	CDM120_RS02520	0	class I SAM-dependent methyltransferase
CDM120_RS02525	CDM120_RS02525	0	aminoglycoside nucleotidyltransferase ANT(6)-Ia
CDM120_RS02530	CDM120_RS02530	0	adenine phosphoribosyltransferase
CDM120_RS02535	CDM120_RS02535	0	ANT(9) family aminoglycoside nucleotidyltransferase Spw
CDM120_RS02545	CDM120_RS02545	0	replication initiator protein A
CDM120_RS02550	CDM120_RS02550	0	ATP-binding protein
CDM120_RS02555	CDM120_RS02555	0	PcfB family protein
CDM120_RS02560	CDM120_RS02560	0	phage antirepressor KilAC domain-containing protein
CDM120_RS02565	CDM120_RS02565	0	pseudogene
CDM120_RS02570	CDM120_RS02570	0	metalloregulator ArsR/SmtB family transcription factor
CDM120_RS02575	CDM120_RS02575	0	permease
CDM120_RS02580	CDM120_RS02580	0	thioredoxin family protein
CDM120_RS02585	CDM120_RS02585	0	VOC family protein
CDM120_RS02590	CDM120_RS02590	0	PadR family transcriptional regulator
CDM120_RS02595	tet44	0	Tc resistance ribosomal protection protein Tet(44)
CDM120_RS02600	CDM120_RS02600	0	aminoglycoside nucleotidyltransferase ANT(6)-Ib
CDM120_RS02605	CDM120_RS02605	0	replication initiator protein A
CDM120_RS02610	CDM120_RS02610	0	ATP-binding protein

CDM120_RS19220	CDM120_RS19220	0	pseudogene
CDM120_RS19740	CDM120_RS19740	0	pseudogene
CDM120_RS02620	CDM120_RS02620	0	hypothetical protein
CDM120_RS19545	CDM120_RS19545	0	hypothetical protein
CDM120_RS02625	CDM120_RS02625	0	PcfB family protein
CDM120_RS02630	CDM120_RS02630	0	phage antirepressor KilAC domain-containing protein
CDM120_RS02635	CDM120_RS02635	0	type IV secretory system conjugative DNA transfer family protein
CDM120_RS02640	CDM120_RS02640	0	hypothetical protein
CDM120_RS02645	CDM120_RS02645	0	hypothetical protein
CDM120_RS02650	CDM120_RS02650	0	Maff2 family protein
CDM120_RS02655	CDM120_RS02655	0	hypothetical protein
CDM120_RS02660	CDM120_RS02660	0	hypothetical protein
CDM120_RS02665	CDM120_RS02665	0	PrgI family protein
CDM120_RS02670	CDM120_RS02670	0	ATP-binding protein
CDM120_RS02675	CDM120_RS02675	0	C40 family peptidase
CDM120_RS02680	CDM120_RS02680	0	hypothetical protein
CDM120_RS02685	CDM120_RS02685	0	copper amine oxidase
CDM120_RS02690	CDM120_RS02690	0	DUF4366 domain-containing protein
CDM120_RS02695	CDM120_RS02695	0	DUF6088 family protein
CDM120_RS02700	CDM120_RS02700	0	nucleotidyl transferase AbiEii/AbiGii toxin family protein
CDM120_RS19000	CDM120_RS19000	0	pseudogene
CDM120_RS02705	CDM120_RS02705	0	helix-turn-helix domain-containing protein
CDM120_RS02710	CDM120_RS02710	0	MFS transporter
CDM120_RS02715	CDM120_RS02715	0	GNAT family N-acetyltransferase
CDM120_RS02720	CDM120_RS02720	0	DNA topoisomerase 3
CDM120_RS02725	dcm	0	DNA (cytosine-5-)-methyltransferase
CDM120_RS02730	CDM120_RS02730	0	DEAD/DEAH box helicase family protein
CDM120_RS02735	CDM120_RS02735	0	hypothetical protein
CDM120_RS18570	CDM120_RS18570	0	helix-turn-helix domain-containing protein
CDM120_RS02745	CDM120_RS02745	0	MFS transporter
CDM120_RS02750	CDM120_RS02750	0	relaxase/mobilization nuclease domain-containing protein
CDM120_RS02755	*mobC*	0	plasmid mobilization relaxosome protein MobC
CDM120_RS02760	CDM120_RS02760	0	virulence RhuM family protein
CDM120_RS02765	CDM120_RS02765	0	hypothetical protein
CDM120_RS02770	CDM120_RS02770	0	ImmA/IrrE family metallo-endopeptidase
CDM120_RS02775	CDM120_RS02775	0	helix-turn-helix domain-containing protein
CDM120_RS02780	CDM120_RS02780	0	hypothetical protein
CDM120_RS02785	CDM120_RS02785	0	sigma-70 family RNA polymerase sigma factor
CDM120_RS19550	CDM120_RS19550	0	hypothetical protein

CDM120_RS02790	CDM120_RS02790	0	recombinase family protein
CDM120_RS02835	CDM120_RS02835	5	hypothetical protein
CDM120_RS02855	CDM120_RS02855	8	DUF3784 domain-containing protein
CDM120_RS03010	CDM120_RS03010	7	MBL fold metallo-hydrolase
CDM120_RS03110	CDM120_RS03110	352	sigma-70 family RNA polymerase sigma factor
CDM120_RS03120	CDM120_RS03120	2	hypothetical protein
CDM120_RS03130	CDM120_RS03130	2	hypothetical protein
CDM120_RS03135	CDM120_RS03135	19	sigma-70 family RNA polymerase sigma factor
CDM120_RS03145	CDM120_RS03145	4	hypothetical protein
CDM120_RS03165	CDM120_RS03165	5	hypothetical protein
CDM120_RS03170	CDM120_RS03170	6	hypothetical protein
CDM120_RS03245	CDM120_RS03245	23	ABC transporter ATP-binding protein
CDM120_RS03470	CDM120_RS03470	9	BlaI/MecI/CopY family transcriptional regulator
CDM120_RS03690	CDM120_RS03690	9	hypothetical protein
CDM120_RS03735	CDM120_RS03735	0	spore coat protein CotJB
CDM120_RS03820	CDM120_RS03820	0	YvrJ family protein
CDM120_RS03825	CDM120_RS03825	2	DUF2922 domain-containing protein
CDM120_RS19035	CDM120_RS19035	2	hypothetical protein
CDM120_RS04075	CDM120_RS04075	44	methyltransferase domain-containing protein
CDM120_RS04105	*infC*	9	translation initiation factor IF-3
CDM120_RS04115	*rplT*	11	50S ribosomal protein L20
CDM120_RS04180	CDM120_RS04180	7	TrkA family potassium uptake protein
CDM120_RS04310	CDM120_RS04310	14	cyclodeaminase/cyclohydrolase family protein
CDM120_RS04385	CDM120_RS04385	4	CooT family nickel-binding protein
CDM120_RS04410	CDM120_RS04410	6	hypothetical protein
CDM120_RS04435	CDM120_RS04435	0	
CDM120_RS04670	CDM120_RS04670	0	sulfite exporter TauE/SafE family protein
CDM120_RS04690	CDM120_RS04690	24	enoyl-CoA hydratase/isomerase family protein
CDM120_RS04810	CDM120_RS04810	49	rubrerythrin family protein
CDM120_RS04920	CDM120_RS04920	4	GIY-YIG nuclease family protein
CDM120_RS04950	CDM120_RS04950	47	pseudogene
CDM120_RS05015	CDM120_RS05015	5	PTS sugar transporter subunit IIB
CDM120_RS05170	CDM120_RS05170	1	cold-shock protein
CDM120_RS05190	CDM120_RS05190	8	hypothetical protein
CDM120_RS05335	CDM120_RS05335	16	ABC transporter ATP-binding protein
CDM120_RS05360	CDM120_RS05360	6	GNAT family N-acetyltransferase
CDM120_RS19480	CDM120_RS19480	3	hypothetical protein
CDM120_RS05510	CDM120_RS05510	805	cellulose biosynthesis cyclic di-GMP-binding regulatory protein BcsB
CDM120_RS05690	CDM120_RS05690	2	hypothetical protein
CDM120_RS05695	CDM120_RS05695	6	hypothetical protein

CDM120_RS05710	CDM120_RS05710	2	hypothetical protein
CDM120_RS18600	*ssrS*	0	
CDM120_RS05755	CDM120_RS05755	11	PTS sugar transporter subunit IIA
CDM120_RS05910	CDM120_RS05910	0	
CDM120_RS05920	CDM120_RS05920	6	dephospho-CoA kinase
CDM120_RS05990	*maf*	7	Maf family protein
CDM120_RS06000	CDM120_RS06000	17	rod shape-determining protein
CDM120_RS06025	*minD*	18	septum site-determining protein MinD
CDM120_RS06080	*rplU*	12	50S ribosomal protein L21
CDM120_RS06090	*rpmA*	5	50S ribosomal protein L27
CDM120_RS06165	*fapR*	28	transcription factor FapR
CDM120_RS06220	CDM120_RS06220	1552	gamma-glutamyl-gamma-aminobutyrate hydrolase family protein
CDM120_RS06240	*spoIIIAA*	17	stage III sporulation protein AA
CDM120_RS06250	*spoIIIAC*	5	stage III sporulation protein AC
CDM120_RS06280	CDM120_RS06280	6	Asp23/Gls24 family envelope stress response protein
CDM120_RS06300	*xseB*	239	exodeoxyribonuclease VII small subunit
CDM120_RS06310	CDM120_RS06310	7	divergent PAP2 family protein
CDM120_RS06360	CDM120_RS06360	34	copper transporter
CDM120_RS06430	CDM120_RS06430	117	sigma-70 family RNA polymerase sigma factor
CDM120_RS06495	CDM120_RS06495	4	hypothetical protein
CDM120_RS06525	*efp*	16	elongation factor P
CDM120_RS06550	*ftsY*	36	signal recognition particle-docking protein FtsY
CDM120_RS06555	CDM120_RS06555	6	putative DNA-binding protein
CDM120_RS06560	*ffh*	52	pseudogene
CDM120_RS06565	*rpsP*	2	30S ribosomal protein S16
CDM120_RS06585	*rplS*	5	50S ribosomal protein L19
CDM120_RS06655	CDM120_RS06655	3	YraN family protein
CDM120_RS06690	CDM120_RS06690	11	Rrf2 family transcriptional regulator
CDM120_RS06695	*nifS*	28	cysteine desulfurase NifS
CDM120_RS06705	*mnmA*	23	tRNA 2-thiouridine(34) synthase MnmA
CDM120_RS06715	CDM120_RS06715	0	IreB family regulatory phosphoprotein
CDM120_RS06750	CDM120_RS06750	0	spore protein
CDM120_RS06785	CDM120_RS06785	55	DUF2953 domain-containing protein
CDM120_RS19580	CDM120_RS19580	0	hypothetical protein
CDM120_RS06835	CDM120_RS06835	11	ribosome maturation factor RimP
CDM120_RS06845	CDM120_RS06845	4	YlxR family protein
CDM120_RS06850	CDM120_RS06850	5	ribosomal L7Ae/L30e/S12e/Gadd45 family protein
CDM120_RS06855	*infB*	56	translation initiation factor IF-2
CDM120_RS06875	CDM120_RS06875	19	bifunctional riboflavin kinase/FAD synthetase
CDM120_RS06910	CDM120_RS06910	2	YlmC/YmxH family sporulation protein

CDM120_RS06965	CDM120_RS06965	202	UvrD-helicase domain-containing protein
CDM120_RS07045	CDM120_RS07045	14	lantibiotic protection ABC transporter ATP-binding protein
CDM120_RS07065	CDM120_RS07065	25	pyridoxamine kinase
CDM120_RS07075	CDM120_RS07075	2	cold-shock protein
CDM120_RS07105	CDM120_RS07105	118	helix-turn-helix transcriptional regulator
CDM120_RS07155	CDM120_RS07155	2	DUF2577 domain-containing protein
CDM120_RS07330	CDM120_RS07330	248	glycerophosphodiester phosphodiesterase
CDM120_RS07370	*ddlR*	41	transcriptional regulator DdlR
CDM120_RS07475	*pdaA*	1604	delta-lactam-biosynthetic de-N-acetylase
CDM120_RS07575	*folP*	7	dihydropteroate synthase
CDM120_RS07595	CDM120_RS07595	46	DNA primase
CDM120_RS07600	*rpoD*	28	RNA polymerase sigma factor RpoD
CDM120_RS07605	CDM120_RS07605	16	class I SAM-dependent methyltransferase
CDM120_RS07755	CDM120_RS07755	15	hypothetical protein
CDM120_RS07805	CDM120_RS07805	47	DUF1002 domain-containing protein
CDM120_RS07835	CDM120_RS07835	32	metal ABC transporter permease
CDM120_RS07965	CDM120_RS07965	23	tyrosine--tRNA ligase
CDM120_RS08045	*gltA*	41	NADPH-dependent glutamate synthase
CDM120_RS08060	CDM120_RS08060	47	ABC transporter permease
CDM120_RS08180	CDM120_RS08180	6	ACT domain-containing protein
CDM120_RS08195	CDM120_RS08195	313	manganese catalase family protein
CDM120_RS08275	CDM120_RS08275	21	alpha/beta hydrolase
CDM120_RS08310	CDM120_RS08310	18	K(+)-transporting ATPase subunit C
CDM120_RS08370	CDM120_RS08370	22	ABC transporter ATP-binding protein
CDM120_RS08385	CDM120_RS08385	18	ABC transporter ATP-binding protein
CDM120_RS08480	CDM120_RS08480	77	superoxide dismutase
CDM120_RS08595	CDM120_RS08595	5957	lipoate--protein ligase
CDM120_RS08645	CDM120_RS08645	2	hypothetical protein
CDM120_RS08670	CDM120_RS08670	77	hypothetical protein
CDM120_RS08730	CDM120_RS08730	45	flavodoxin domain-containing protein
CDM120_RS08860	*thiF*	25	thiamine biosynthesis protein ThiF
CDM120_RS08875	CDM120_RS08875	26	thiamine phosphate synthase
CDM120_RS08920	CDM120_RS08920	7	MOSC domain-containing protein
CDM120_RS08945	CDM120_RS08945	2	hypothetical protein
CDM120_RS08965	CDM120_RS08965	1	hypothetical protein
CDM120_RS08995	CDM120_RS08995	21	nitrogenase iron protein NifH
CDM120_RS09175	CDM120_RS09175	29	ABC transporter ATP-binding protein
CDM120_RS09185	CDM120_RS09185	320	hypothetical protein
CDM120_RS09195	*def*	17	peptide deformylase
CDM120_RS09210	CDM120_RS09210	12	class I SAM-dependent methyltransferase

CDM120_RS09245	CDM120_RS09245	8	DUF4624 family lipoprotein
CDM120_RS09250	*gap*	20	type I glyceraldehyde-3-phosphate dehydrogenase
CDM120_RS09320	CDM120_RS09320	20	hypothetical protein
CDM120_RS09325	CDM120_RS09325	11	response regulator transcription factor
CDM120_RS09330	CDM120_RS09330	46	HAMP domain-containing histidine kinase
CDM120_RS09350	CDM120_RS09350	1522	MerR family transcriptional regulator
CDM120_RS09355	CDM120_RS09355	12	DUF4363 family protein
CDM120_RS09380	CDM120_RS09380	15	C-GCAxxG-C-C family protein
CDM120_RS09395	CDM120_RS09395	63	sulfite reductase subunit C
CDM120_RS09480	CDM120_RS09480	8	LytTR family transcriptional regulator
CDM120_RS09490	CDM120_RS09490	17	histidine phosphatase family protein
CDM120_RS09505	CDM120_RS09505	8	1-acyl-sn-glycerol-3-phosphate acyltransferase
CDM120_RS09510	CDM120_RS09510	31	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
CDM120_RS09565	CDM120_RS09565	125	response regulator transcription factor
CDM120_RS09695	CDM120_RS09695	18	helix-turn-helix domain-containing protein
CDM120_RS09710	CDM120_RS09710	725	aminoglycoside 6-adenylyltransferase
CDM120_RS19605	CDM120_RS19605	2	hypothetical protein
CDM120_RS09775	*eutA*	264	ethanolamine ammonia-lyase reactivating factor EutA
CDM120_RS09785	*eutC*	1608	ethanolamine ammonia-lyase subunit EutC
CDM120_RS09825	CDM120_RS09825	0	EutN/CcmL family microcompartment protein
CDM120_RS09905	CDM120_RS09905	19	acetyl-CoA carboxylase carboxyltransferase subunit alpha
CDM120_RS09910	CDM120_RS09910	20	acetyl-CoA carboxylase
CDM120_RS09915	*accC*	29	acetyl-CoA carboxylase biotin carboxylase subunit
CDM120_RS09920	*accB*	1	acetyl-CoA carboxylase biotin carboxyl carrier protein
CDM120_RS09925	CDM120_RS09925	8	hypothetical protein
CDM120_RS09935	CDM120_RS09935	50	hypothetical protein
CDM120_RS09985	CDM120_RS09985	1175	FtsX-like permease family protein
CDM120_RS09990	CDM120_RS09990	345	ABC transporter ATP-binding protein
CDM120_RS10095	*miaA*	11	tRNA (adenosine(37)-N6)-dimethylallyltransferase MiaA
CDM120_RS10130	CDM120_RS10130	0	hypothetical protein
CDM120_RS19495	CDM120_RS19495	2	hypothetical protein
CDM120_RS10190	CDM120_RS10190	0	helix-turn-helix domain-containing protein
CDM120_RS10200	CDM120_RS10200	223	hypothetical protein
CDM120_RS10205	CDM120_RS10205	19	hypothetical protein
CDM120_RS10210	CDM120_RS10210	10	hypothetical protein
CDM120_RS10215	CDM120_RS10215	2172	hypothetical protein
CDM120_RS10250	CDM120_RS10250	0	hypothetical protein
CDM120_RS10255	CDM120_RS10255	4	helix-turn-helix transcriptional regulator
CDM120_RS19615	CDM120_RS19615	4	hypothetical protein
CDM120_RS10325	CDM120_RS10325	35	2

CDM120_RS10345	CDM120_RS10345	16	TetR/AcrR family transcriptional regulator
CDM120_RS10350	CDM120_RS10350	11	pseudogene
CDM120_RS10355	CDM120_RS10355	0	tyrosine-type recombinase/integrase
CDM120_RS10360	CDM120_RS10360	0	excisionase
CDM120_RS10365	CDM120_RS10365	0	helix-turn-helix domain-containing protein
CDM120_RS10370	CDM120_RS10370	0	sigma-70 family RNA polymerase sigma factor
CDM120_RS10375	CDM120_RS10375	0	helix-turn-helix transcriptional regulator
CDM120_RS10380	CDM120_RS10380	0	cysteine-rich KTR domain-containing protein
CDM120_RS10385	*tetM*	0	Tc resistance ribosomal protection protein Tet(M)
CDM120_RS10390	CDM120_RS10390	0	conjugal transfer protein
CDM120_RS10395	CDM120_RS10395	0	bifunctional lysozyme/C40 family peptidase
CDM120_RS10400	CDM120_RS10400	0	YtxH domain-containing protein
CDM120_RS10405	CDM120_RS10405	0	pseudogene
CDM120_RS10410	CDM120_RS10410	0	conjugal transfer protein
CDM120_RS10415	CDM120_RS10415	0	antirestriction protein ArdA
CDM120_RS10420	CDM120_RS10420	0	hypothetical protein
CDM120_RS10425	CDM120_RS10425	0	replication initiation factor domain-containing protein
CDM120_RS10430	CDM120_RS10430	0	hypothetical protein
CDM120_RS10435	CDM120_RS10435	0	DUF87 domain-containing protein
CDM120_RS10440	CDM120_RS10440	0	YdcP family protein
CDM120_RS10445	CDM120_RS10445	0	YdcP family protein
CDM120_RS10460	CDM120_RS10460	10	VanZ family protein
CDM120_RS10465	CDM120_RS10465	0	hypothetical protein
CDM120_RS10540	*argF*	644	ornithine carbamoyltransferase
CDM120_RS10555	*argJ*	35	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ
CDM120_RS10600	CDM120_RS10600	1	hypothetical protein
CDM120_RS10655	CDM120_RS10655	27	hypothetical protein
CDM120_RS10700	CDM120_RS10700	3	hypothetical protein
CDM120_RS10755	CDM120_RS10755	2369	purine/pyrimidine permease
CDM120_RS10770	CDM120_RS10770	58	amidohydrolase family protein
CDM120_RS10785	CDM120_RS10785	11	2Fe-2S iron-sulfur cluster binding domain-containing protein
CDM120_RS10865	CDM120_RS10865	4	QueT transporter family protein
CDM120_RS10895	CDM120_RS10895	0	hypothetical protein
CDM120_RS11005	CDM120_RS11005	29	magnesium transporter CorA family protein
CDM120_RS11055	CDM120_RS11055	98	helix-turn-helix domain-containing protein
CDM120_RS11115	CDM120_RS11115	0	diguanylate cyclase
CDM120_RS11125	CDM120_RS11125	16	phosphatidate cytidylyltransferase
CDM120_RS11130	CDM120_RS11130	7	isoprenyl transferase
CDM120_RS11140	*frr*	12	ribosome recycling factor
CDM120_RS11145	CDM120_RS11145	25	UMP kinase

CDM120_RS11160	CDM120_RS11160	43	D-alanyl-D-alanine carboxypeptidase
CDM120_RS11170	CDM120_RS11170	23	helix-turn-helix domain-containing protein
CDM120_RS11200	CDM120_RS11200	0	endonuclease/exonuclease/phosphatase family protein
CDM120_RS19650	CDM120_RS19650	2	hypothetical protein
CDM120_RS11250	CDM120_RS11250	22	putative ABC transporter permease
CDM120_RS11255	CDM120_RS11255	16	VanZ family protein
CDM120_RS11300	CDM120_RS11300	21	4Fe-4S binding protein
CDM120_RS11385	CDM120_RS11385	146	DUF3794 domain-containing protein
CDM120_RS11435	CDM120_RS11435	10	ferritin
CDM120_RS11595	*asrA*	44	anaerobic sulfite reductase subunit AsrA
CDM120_RS11620	CDM120_RS11620	34	YhcH/YjgK/YiaL family protein
CDM120_RS11630	CDM120_RS11630	24	N-acetylneuraminate lyase
CDM120_RS11705	CDM120_RS11705	2	pseudogene
CDM120_RS11835	CDM120_RS11835	15	PTS sugar transporter subunit IIA
CDM120_RS11905	CDM120_RS11905	6	hypothetical protein
CDM120_RS11925	CDM120_RS11925	12	hypothetical protein
CDM120_RS11945	CDM120_RS11945	2	cold-shock protein
CDM120_RS12070	CDM120_RS12070	33	IMP dehydrogenase
CDM120_RS12095	CDM120_RS12095	0	hypothetical protein
CDM120_RS12155	CDM120_RS12155	9	glycine/betaine reductase A
CDM120_RS12225	CDM120_RS12225	39	hypothetical protein
CDM120_RS12425	CDM120_RS12425	5	spore coat protein CotJB
CDM120_RS12440	*dut*	12	dUTP diphosphatase
CDM120_RS12550	CDM120_RS12550	32	PLP-dependent aminotransferase family protein
CDM120_RS12570	CDM120_RS12570	55	oxidoreductase
CDM120_RS12580	CDM120_RS12580	0	4Fe-4S binding protein
CDM120_RS12630	CDM120_RS12630	15	cytidine deaminase
CDM120_RS12655	CDM120_RS12655	90	YabP/YqfC family sporulation protein
CDM120_RS12685	*mtaB*	40	tRNA (N(6)-L-threonylcarbamoyladenosine(37)-C(2))-methylthiotransferase MtaB
CDM120_RS12700	CDM120_RS12700	61	CRISPR-associated helicase/endonuclease Cas3
CDM120_RS12705	*cas5*	11	CRISPR-associated protein Cas5
CDM120_RS12710	CDM120_RS12710	27	DevR family CRISPR-associated autoregulator
CDM120_RS12715	CDM120_RS12715	25	hypothetical protein
CDM120_RS12720	*cas6*	17	CRISPR-associated endoribonuclease Cas6
CDM120_RS19385	CDM120_RS19385	0	hypothetical protein
CDM120_RS12810	*rpsT*	0	30S ribosomal protein S20
CDM120_RS19390	CDM120_RS19390	1	
CDM120_RS12870	CDM120_RS12870	19	helix-turn-helix transcriptional regulator
CDM120_RS12890	CDM120_RS12890	5	PTS fructose transporter subunit IIB
CDM120_RS12910	*manA*	22	mannose-6-phosphate isomerase

CDM120_RS13035	CDM120_RS13035	12	hypothetical protein
CDM120_RS13055	CDM120_RS13055	60	leucine--tRNA ligase
CDM120_RS13095	CDM120_RS13095	9	antibiotic biosynthesis monooxygenase
CDM120_RS13225	*coaD*	9	pantetheine-phosphate adenylyltransferase
CDM120_RS13260	CDM120_RS13260	4	Rpn family recombination-promoting nuclease/putative transposase
CDM120_RS13315	CDM120_RS13315	31	thiamine diphosphokinase
CDM120_RS13365	CDM120_RS13365	35	methionyl-tRNA formyltransferase
CDM120_RS13390	*gmk*	5	guanylate kinase
CDM120_RS19115	CDM120_RS19115	24	YicC family protein
CDM120_RS19120	CDM120_RS19120	18	diaminopimelate epimerase
CDM120_RS13485	*trpS*	21	tryptophan--tRNA ligase
CDM120_RS13555	CDM120_RS13555	700	hypothetical protein
CDM120_RS13595	CDM120_RS13595	32	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
CDM120_RS13615	CDM120_RS13615	7	hypothetical protein
CDM120_RS13635	CDM120_RS13635	19	response regulator transcription factor
CDM120_RS13675	*ftsZ*	70	cell division protein FtsZ
CDM120_RS13680	CDM120_RS13680	7	small basic family protein
CDM120_RS13770	CDM120_RS13770	18	AraC family transcriptional regulator
CDM120_RS13820	CDM120_RS13820	802	acetyl-CoA C-acetyltransferase
CDM120_RS13835	CDM120_RS13835	20	3-hydroxybutyrate dehydrogenase
CDM120_RS13860	CDM120_RS13860	10	flavodoxin family protein
CDM120_RS13875	CDM120_RS13875	8	hypothetical protein
CDM120_RS19670	CDM120_RS19670	0	hypothetical protein
CDM120_RS13885	CDM120_RS13885	3	alpha/beta-type small acid-soluble spore protein
CDM120_RS13975	CDM120_RS13975	25	pseudogene
CDM120_RS14230	CDM120_RS14230	74	bifunctional (p)ppGpp synthetase/guanosine-3'
CDM120_RS14285	CDM120_RS14285	1	HPr family phosphocarrier protein
CDM120_RS14370	CDM120_RS14370	24	glycosyltransferase family 2 protein
CDM120_RS14385	CDM120_RS14385	22	glycosyltransferase family 2 protein
CDM120_RS14410	*murJ*	45	murein biosynthesis integral membrane protein MurJ
CDM120_RS14465	*secA*	86	preprotein translocase subunit SecA
CDM120_RS14565	CDM120_RS14565	7	pseudogene
CDM120_RS14590	CDM120_RS14590	0	hypothetical protein
CDM120_RS14670	CDM120_RS14670	561	serine hydroxymethyltransferase
CDM120_RS14990	CDM120_RS14990	582	V-type ATP synthase subunit E
CDM120_RS19675	CDM120_RS19675	2	hypothetical protein
CDM120_RS15125	CDM120_RS15125	23	ABC transporter ATP-binding protein
CDM120_RS15295	CDM120_RS15295	7	C-GCAxxG-C-C family protein
CDM120_RS15740	CDM120_RS15740	17	6-phosphofructokinase
CDM120_RS15840	CDM120_RS15840	3	metalloregulator ArsR/SmtB family transcription factor

CDM120_RS18615	*ssrA*	16	
CDM120_RS15985	*smpB*	11	SsrA-binding protein SmpB
CDM120_RS16010	CDM120_RS16010	9	YfbM family protein
CDM120_RS16370	CDM120_RS16370	34	4-hydroxy-tetrahydrodipicolinate reductase
CDM120_RS16795	CDM120_RS16795	97	DUF3783 domain-containing protein
CDM120_RS19700	CDM120_RS19700	0	hypothetical protein
CDM120_RS17330	CDM120_RS17330	2	PTS lactose/cellobiose transporter subunit IIA
CDM120_RS17460	*atpF*	11	F0F1 ATP synthase subunit B
CDM120_RS17510	CDM120_RS17510	21	threonylcarbamoyl-AMP synthase
CDM120_RS17535	*rpmE*	2	50S ribosomal protein L31
CDM120_RS17855	*greA*	12	transcription elongation factor GreA
CDM120_RS17920	CDM120_RS17920	14	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase
CDM120_RS18230	CDM120_RS18230	30	putative bacteriocin export ABC transporter
CDM120_RS18280	CDM120_RS18280	10	PTS sugar transporter subunit IIB
CDM120_RS18290	CDM120_RS18290	201	cysteine hydrolase
CDM120_RS18435	CDM120_RS18435	12	50S ribosomal protein L9
CDM120_RS18450	CDM120_RS18450	9	MazG-like family protein
CDM120_RS18455	CDM120_RS18455	3	30S ribosomal protein S18
CDM120_RS18480	CDM120_RS18480	3	DUF3343 domain-containing protein

Table 7.2: List of predicted essential genes for sporulation

locus_tag	gene_name	read_count	fcn
CDM120_RS00005	*dnaA*	0	chromosomal replication initiator protein DnaA
CDM120_RS00015	*yaaA*	0	S4 domain-containing protein YaaA
CDM120_RS00075	CDM120_RS00075	0	
CDM120_RS00090	CDM120_RS00090	0	
CDM120_RS00100	*rrf*	0	

CDM120_RS00125	CDM120_RS00125	0	
CDM120_RS18595	*ffs*	0	
CDM120_RS00165	*rrf*	0	
CDM120_RS00170	CDM120_RS00170	0	
CDM120_RS00175	CDM120_RS00175	0	
CDM120_RS00180	CDM120_RS00180	0	
CDM120_RS00185	CDM120_RS00185	0	
CDM120_RS00190	CDM120_RS00190	0	
CDM120_RS00195	CDM120_RS00195	0	
CDM120_RS00200	CDM120_RS00200	0	
CDM120_RS00210	CDM120_RS00210	0	
CDM120_RS00215	CDM120_RS00215	0	
CDM120_RS00220	CDM120_RS00220	0	
CDM120_RS00225	CDM120_RS00225	0	
CDM120_RS00230	CDM120_RS00230	0	
CDM120_RS00235	CDM120_RS00235	0	
CDM120_RS00245	CDM120_RS00245	0	
CDM120_RS00250	CDM120_RS00250	0	
CDM120_RS00255	CDM120_RS00255	0	
CDM120_RS00260	CDM120_RS00260	0	
CDM120_RS00265	CDM120_RS00265	0	
CDM120_RS00270	CDM120_RS00270	0	
CDM120_RS00275	CDM120_RS00275	0	
CDM120_RS00280	CDM120_RS00280	0	
CDM120_RS00285	CDM120_RS00285	0	
CDM120_RS00290	CDM120_RS00290	0	
CDM120_RS00295	CDM120_RS00295	0	
CDM120_RS00300	CDM120_RS00300	0	
CDM120_RS00305	CDM120_RS00305	0	
CDM120_RS00310	CDM120_RS00310	0	
CDM120_RS00315	CDM120_RS00315	0	
CDM120_RS00320	CDM120_RS00320	0	
CDM120_RS00325	CDM120_RS00325	0	
CDM120_RS00330	CDM120_RS00330	0	
CDM120_RS00335	CDM120_RS00335	0	
CDM120_RS00340	CDM120_RS00340	0	
CDM120_RS00345	CDM120_RS00345	0	
CDM120_RS00350	CDM120_RS00350	0	
CDM120_RS00360	CDM120_RS00360	0	
CDM120_RS00365	CDM120_RS00365	0	
CDM120_RS00370	CDM120_RS00370	0	

CDM120_RS00380	CDM120_RS00380	0	
CDM120_RS00565	CDM120_RS00565	1	cysteine--tRNA ligase
CDM120_RS00600	*rpmG*	0	50S ribosomal protein L33
CDM120_RS00625	CDM120_RS00625	0	50S ribosomal protein L10
CDM120_RS00650	*rpsL*	0	30S ribosomal protein S12
CDM120_RS00685	*rplW*	0	50S ribosomal protein L23
CDM120_RS00690	*rplB*	0	50S ribosomal protein L2
CDM120_RS00705	*rpsC*	0	30S ribosomal protein S3
CDM120_RS00710	*rplP*	0	50S ribosomal protein L16
CDM120_RS00715	*rpmC*	0	50S ribosomal protein L29
CDM120_RS00745	*rpsH*	0	30S ribosomal protein S8
CDM120_RS00760	*rpsE*	0	30S ribosomal protein S5
CDM120_RS00765	*rpmD*	0	50S ribosomal protein L30
CDM120_RS00770	*rplO*	0	50S ribosomal protein L15
CDM120_RS00775	*secY*	0	preprotein translocase subunit SecY
CDM120_RS00790	CDM120_RS00790	0	KOW domain-containing RNA-binding protein
CDM120_RS00800	*rpmJ*	0	50S ribosomal protein L36
CDM120_RS00805	*rpsM*	0	30S ribosomal protein S13
CDM120_RS00810	*rpsK*	0	30S ribosomal protein S11
CDM120_RS00850	*rplM*	0	50S ribosomal protein L13
CDM120_RS00855	*rpsI*	0	30S ribosomal protein S9
CDM120_RS00870	CDM120_RS00870	0	
CDM120_RS00880	*rrf*	0	
CDM120_RS00885	CDM120_RS00885	0	
CDM120_RS00890	CDM120_RS00890	0	
CDM120_RS00895	CDM120_RS00895	0	
CDM120_RS00900	CDM120_RS00900	0	
CDM120_RS00910	CDM120_RS00910	0	
CDM120_RS00915	CDM120_RS00915	0	
CDM120_RS00920	CDM120_RS00920	0	
CDM120_RS00925	CDM120_RS00925	0	
CDM120_RS00930	CDM120_RS00930	0	
CDM120_RS00935	CDM120_RS00935	0	
CDM120_RS00945	CDM120_RS00945	0	
CDM120_RS00950	CDM120_RS00950	0	
CDM120_RS00955	CDM120_RS00955	0	
CDM120_RS00960	CDM120_RS00960	0	
CDM120_RS00965	CDM120_RS00965	0	
CDM120_RS00970	CDM120_RS00970	0	
CDM120_RS00980	CDM120_RS00980	0	
CDM120_RS00990	*rrf*	0	

CDM120_RS01015	CDM120_RS01015	0	
CDM120_RS01030	CDM120_RS01030	0	
CDM120_RS01035	CDM120_RS01035	0	
CDM120_RS01040	CDM120_RS01040	0	
CDM120_RS01050	CDM120_RS01050	0	
CDM120_RS01055	CDM120_RS01055	0	
CDM120_RS01060	CDM120_RS01060	0	
CDM120_RS01065	CDM120_RS01065	0	
CDM120_RS01070	CDM120_RS01070	0	
CDM120_RS01165	CDM120_RS01165	0	rod shape-determining protein
CDM120_RS01170	*fabZ*	0	3-hydroxyacyl-ACP dehydratase FabZ
CDM120_RS01285	*rimI*	0	ribosomal protein S18-alanine N-acetyltransferase
CDM120_RS01300	*hpdC*	0	4-hydroxyphenylacetate decarboxylase small subunit
CDM120_RS01310	CDM120_RS01310	0	hypothetical protein
CDM120_RS01365	CDM120_RS01365	0	4Fe-4S binding protein
CDM120_RS01470	CDM120_RS01470	0	co-chaperone GroES
CDM120_RS01595	CDM120_RS01595	0	hypothetical protein
CDM120_RS18635	CDM120_RS18635	0	pseudogene
CDM120_RS01890	CDM120_RS01890	0	hypothetical protein
CDM120_RS01990	*rrf*	0	
CDM120_RS19215	CDM120_RS19215	0	pseudogene
CDM120_RS02110	CDM120_RS02110	0	metallophosphoesterase
CDM120_RS02250	CDM120_RS02250	0	helix-turn-helix transcriptional regulator
CDM120_RS02255	CDM120_RS02255	0	DNA adenine methylase
CDM120_RS02260	CDM120_RS02260	0	restriction endonuclease subunit M
CDM120_RS02265	CDM120_RS02265	0	hypothetical protein
CDM120_RS02270	CDM120_RS02270	0	AAA family ATPase
CDM120_RS02275	CDM120_RS02275	0	LlaJI family restriction endonuclease
CDM120_RS02280	CDM120_RS02280	0	CGNR zinc finger domain-containing protein
CDM120_RS02285	CDM120_RS02285	0	sigma-70 family RNA polymerase sigma factor
CDM120_RS02290	CDM120_RS02290	0	hypothetical protein
CDM120_RS02295	CDM120_RS02295	0	hypothetical protein
CDM120_RS02300	CDM120_RS02300	0	DUF2800 domain-containing protein
CDM120_RS02305	CDM120_RS02305	0	DUF2815 family protein
CDM120_RS02310	CDM120_RS02310	0	DNA polymerase
CDM120_RS02315	CDM120_RS02315	0	phage antirepressor
CDM120_RS02320	CDM120_RS02320	0	hypothetical protein
CDM120_RS02325	CDM120_RS02325	0	virulence-associated E family protein
CDM120_RS02330	CDM120_RS02330	0	VRR-NUC domain-containing protein
CDM120_RS02335	CDM120_RS02335	0	DEAD/DEAH box helicase

CDM120_RS02340	CDM120_RS02340	0	phage-associated protein
CDM120_RS02345	CDM120_RS02345	0	HNH endonuclease
CDM120_RS02350	CDM120_RS02350	0	P27 family phage terminase small subunit
CDM120_RS02355	CDM120_RS02355	0	methionine adenosyltransferase
CDM120_RS02360	CDM120_RS02360	0	site-specific DNA-methyltransferase
CDM120_RS02365	CDM120_RS02365	0	virulence protein
CDM120_RS02370	CDM120_RS02370	0	DUF4314 domain-containing protein
CDM120_RS02375	CDM120_RS02375	0	amidoligase family protein
CDM120_RS02380	CDM120_RS02380	0	gamma-glutamylcyclotransferase
CDM120_RS02385	CDM120_RS02385	0	DUF5049 domain-containing protein
CDM120_RS02390	CDM120_RS02390	0	terminase large subunit
CDM120_RS02395	CDM120_RS02395	0	hypothetical protein
CDM120_RS02400	CDM120_RS02400	0	phage portal protein
CDM120_RS02405	CDM120_RS02405	0	Clp protease ClpP
CDM120_RS02410	CDM120_RS02410	0	phage major capsid protein
CDM120_RS02415	CDM120_RS02415	0	hypothetical protein
CDM120_RS02420	CDM120_RS02420	0	head-tail connector protein
CDM120_RS02425	CDM120_RS02425	0	phage head closure protein
CDM120_RS02430	CDM120_RS02430	0	HK97 gp10 family phage protein
CDM120_RS02435	CDM120_RS02435	0	hypothetical protein
CDM120_RS02440	CDM120_RS02440	0	phage major tail
CDM120_RS02445	CDM120_RS02445	0	hypothetical protein
CDM120_RS02450	CDM120_RS02450	0	tape measure protein
CDM120_RS02455	CDM120_RS02455	0	phage tail family protein
CDM120_RS02460	CDM120_RS02460	0	phage tail protein
CDM120_RS02465	CDM120_RS02465	0	hypothetical protein
CDM120_RS02470	CDM120_RS02470	0	hypothetical protein
CDM120_RS02475	CDM120_RS02475	0	glycosyl hydrolase
CDM120_RS02480	CDM120_RS02480	0	phage holin family protein
CDM120_RS02485	CDM120_RS02485	0	N-acetylmuramoyl-L-alanine amidase
CDM120_RS02490	CDM120_RS02490	0	hypothetical protein
CDM120_RS02495	CDM120_RS02495	0	recombinase family protein
CDM120_RS02500	CDM120_RS02500	0	recombinase family protein
CDM120_RS02505	CDM120_RS02505	0	recombinase family protein
CDM120_RS02510	CDM120_RS02510	0	single-stranded DNA-binding protein
CDM120_RS02515	CDM120_RS02515	0	nucleotidyltransferase domain-containing protein
CDM120_RS02520	CDM120_RS02520	0	class I SAM-dependent methyltransferase
CDM120_RS02525	CDM120_RS02525	0	aminoglycoside nucleotidyltransferase ANT(6)-Ia
CDM120_RS02530	CDM120_RS02530	0	adenine phosphoribosyltransferase
CDM120_RS02535	CDM120_RS02535	0	ANT(9) family aminoglycoside nucleotidyltransferase Spw

CDM120_RS02545	CDM120_RS02545	0	replication initiator protein A
CDM120_RS02550	CDM120_RS02550	0	ATP-binding protein
CDM120_RS02555	CDM120_RS02555	0	PcfB family protein
CDM120_RS02560	CDM120_RS02560	0	phage antirepressor KilAC domain-containing protein
CDM120_RS02565	CDM120_RS02565	0	pseudogene
CDM120_RS02570	CDM120_RS02570	0	metalloregulator ArsR/SmtB family transcription factor
CDM120_RS02575	CDM120_RS02575	0	permease
CDM120_RS02580	CDM120_RS02580	0	thioredoxin family protein
CDM120_RS02585	CDM120_RS02585	0	VOC family protein
CDM120_RS02590	CDM120_RS02590	0	PadR family transcriptional regulator
CDM120_RS02595	*tet44*	0	Tc resistance ribosomal protection protein Tet(44)
CDM120_RS02600	CDM120_RS02600	0	aminoglycoside nucleotidyltransferase ANT(6)-Ib
CDM120_RS02605	CDM120_RS02605	0	replication initiator protein A
CDM120_RS02610	CDM120_RS02610	0	ATP-binding protein
CDM120_RS19220	CDM120_RS19220	0	pseudogene
CDM120_RS19740	CDM120_RS19740	0	pseudogene
CDM120_RS02620	CDM120_RS02620	0	hypothetical protein
CDM120_RS19545	CDM120_RS19545	0	hypothetical protein
CDM120_RS02625	CDM120_RS02625	0	PcfB family protein
CDM120_RS02630	CDM120_RS02630	0	phage antirepressor KilAC domain-containing protein
CDM120_RS02635	CDM120_RS02635	0	type IV secretory system conjugative DNA transfer family protein
CDM120_RS02640	CDM120_RS02640	0	hypothetical protein
CDM120_RS02645	CDM120_RS02645	0	hypothetical protein
CDM120_RS02650	CDM120_RS02650	0	Maff2 family protein
CDM120_RS02655	CDM120_RS02655	0	hypothetical protein
CDM120_RS02660	CDM120_RS02660	0	hypothetical protein
CDM120_RS02665	CDM120_RS02665	0	PrgI family protein
CDM120_RS02670	CDM120_RS02670	0	ATP-binding protein
CDM120_RS02675	CDM120_RS02675	0	C40 family peptidase
CDM120_RS02680	CDM120_RS02680	0	hypothetical protein
CDM120_RS02685	CDM120_RS02685	0	copper amine oxidase
CDM120_RS02690	CDM120_RS02690	0	DUF4366 domain-containing protein
CDM120_RS02695	CDM120_RS02695	0	DUF6088 family protein
CDM120_RS02700	CDM120_RS02700	0	nucleotidyl transferase AbiEii/AbiGii toxin family protein
CDM120_RS19000	CDM120_RS19000	0	pseudogene
CDM120_RS02705	CDM120_RS02705	0	helix-turn-helix domain-containing protein
CDM120_RS02710	CDM120_RS02710	0	MFS transporter
CDM120_RS02715	CDM120_RS02715	0	GNAT family N-acetyltransferase

CDM120_RS02720	CDM120_RS02720	0	DNA topoisomerase 3
CDM120_RS02725	*dcm*	0	DNA (cytosine-5-)-methyltransferase
CDM120_RS02730	CDM120_RS02730	0	DEAD/DEAH box helicase family protein
CDM120_RS18570	CDM120_RS18570	0	helix-turn-helix domain-containing protein
CDM120_RS02745	CDM120_RS02745	0	MFS transporter
CDM120_RS02750	CDM120_RS02750	0	relaxase/mobilization nuclease domain-containing protein
CDM120_RS02755	*mobC*	0	plasmid mobilization relaxosome protein MobC
CDM120_RS02760	CDM120_RS02760	0	virulence RhuM family protein
CDM120_RS02765	CDM120_RS02765	0	hypothetical protein
CDM120_RS02770	CDM120_RS02770	0	ImmA/IrrE family metallo-endopeptidase
CDM120_RS02775	CDM120_RS02775	0	helix-turn-helix domain-containing protein
CDM120_RS02780	CDM120_RS02780	0	hypothetical protein
CDM120_RS02785	CDM120_RS02785	0	sigma-70 family RNA polymerase sigma factor
CDM120_RS19550	CDM120_RS19550	0	hypothetical protein
CDM120_RS02790	CDM120_RS02790	0	recombinase family protein
CDM120_RS02930	CDM120_RS02930	0	ornithine aminomutase subunit alpha
CDM120_RS03050	CDM120_RS03050	0	hypothetical protein
CDM120_RS03120	CDM120_RS03120	0	hypothetical protein
CDM120_RS03145	CDM120_RS03145	0	hypothetical protein
CDM120_RS03150	CDM120_RS03150	0	hypothetical protein
CDM120_RS03170	CDM120_RS03170	0	hypothetical protein
CDM120_RS03685	CDM120_RS03685	0	hypothetical protein
CDM120_RS03825	CDM120_RS03825	0	DUF2922 domain-containing protein
CDM120_RS19570	CDM120_RS19570	0	hypothetical protein
CDM120_RS04015	CDM120_RS04015	0	hypothetical protein
CDM120_RS04105	*infC*	0	translation initiation factor IF-3
CDM120_RS04110	*rpmI*	0	50S ribosomal protein L35
CDM120_RS19250	CDM120_RS19250	0	pseudogene
CDM120_RS04190	*pheS*	1	phenylalanine--tRNA ligase subunit alpha
CDM120_RS04200	*zapA*	0	cell division protein ZapA
CDM120_RS04305	CDM120_RS04305	3	formate--tetrahydrofolate ligase
CDM120_RS04410	CDM120_RS04410	0	hypothetical protein
CDM120_RS04435	CDM120_RS04435	0	
CDM120_RS04680	*nadE*	0	NAD(+) synthase
CDM120_RS04820	CDM120_RS04820	0	superoxide reductase
CDM120_RS05170	CDM120_RS05170	0	cold-shock protein
CDM120_RS05540	CDM120_RS05540	0	
CDM120_RS05545	*rrf*	0	
CDM120_RS05690	CDM120_RS05690	0	hypothetical protein
CDM120_RS05695	CDM120_RS05695	0	hypothetical protein

CDM120_RS18600	*ssrS*	0	
CDM120_RS05740	CDM120_RS05740	0	DUF896 domain-containing protein
CDM120_RS05885	CDM120_RS05885	0	helix-turn-helix transcriptional regulator
CDM120_RS05910	CDM120_RS05910	0	
CDM120_RS05920	CDM120_RS05920	0	dephospho-CoA kinase
CDM120_RS05995	*radC*	0	DNA repair protein RadC
CDM120_RS06000	CDM120_RS06000	1	rod shape-determining protein
CDM120_RS06035	*rodA*	1	rod shape-determining protein RodA
CDM120_RS06080	*rplU*	0	50S ribosomal protein L21
CDM120_RS06095	*obgE*	1	GTPase ObgE
CDM120_RS06170	*plsX*	0	phosphate acyltransferase PlsX
CDM120_RS06190	*fabG*	0	3-oxoacyl-[acyl-carrier-protein] reductase
CDM120_RS06195	*acpP*	0	acyl carrier protein
CDM120_RS06215	CDM120_RS06215	0	TIGR03905 family TSCPD domain-containing protein
CDM120_RS06280	CDM120_RS06280	0	Asp23/Gls24 family envelope stress response protein
CDM120_RS06310	CDM120_RS06310	0	divergent PAP2 family protein
CDM120_RS06355	CDM120_RS06355	2	hypothetical protein
CDM120_RS06370	CDM120_RS06370	1	glycosyl transferase
CDM120_RS06525	*efp*	0	elongation factor P
CDM120_RS06560	*ffh*	0	pseudogene
CDM120_RS06565	*rpsP*	0	30S ribosomal protein S16
CDM120_RS06585	*rplS*	0	50S ribosomal protein L19
CDM120_RS06690	CDM120_RS06690	0	Rrf2 family transcriptional regulator
CDM120_RS06705	*mnmA*	1	tRNA 2-thiouridine(34) synthase MnmA
CDM120_RS06715	CDM120_RS06715	0	IreB family regulatory phosphoprotein
CDM120_RS06720	CDM120_RS06720	6	hypothetical protein
CDM120_RS06840	*nusA*	0	transcription termination factor NusA
CDM120_RS06845	CDM120_RS06845	0	YlxR family protein
CDM120_RS06885	*rpsO*	0	30S ribosomal protein S15
CDM120_RS06910	CDM120_RS06910	0	YlmC/YmxH family sporulation protein
CDM120_RS18735	CDM120_RS18735	0	transposase
CDM120_RS07075	CDM120_RS07075	0	cold-shock protein
CDM120_RS07100	CDM120_RS07100	0	XRE family transcriptional regulator
CDM120_RS07155	CDM120_RS07155	0	DUF2577 domain-containing protein
CDM120_RS07190	CDM120_RS07190	0	hypothetical protein
CDM120_RS07280	CDM120_RS07280	0	RidA family protein
CDM120_RS07370	*ddlR*	1	transcriptional regulator DdlR
CDM120_RS07455	CDM120_RS07455	0	hypothetical protein
CDM120_RS07580	*folB*	0	dihydroneopterin aldolase
CDM120_RS07605	CDM120_RS07605	0	class I SAM-dependent methyltransferase

CDM120_RS07725	CDM120_RS07725	0	FeoB-associated Cys-rich membrane protein
CDM120_RS08040	CDM120_RS08040	1	sulfide/dihydroorotate dehydrogenase-like FAD/NAD-binding protein
CDM120_RS08275	CDM120_RS08275	0	alpha/beta hydrolase
CDM120_RS08420	CDM120_RS08420	28	ABC transporter ATP-binding protein
CDM120_RS08605	CDM120_RS08605	0	
CDM120_RS08625	CDM120_RS08625	0	pyrimidine/purine nucleoside phosphorylase
CDM120_RS08770	CDM120_RS08770	1	hypothetical protein
CDM120_RS08945	CDM120_RS08945	0	hypothetical protein
CDM120_RS08965	CDM120_RS08965	0	hypothetical protein
CDM120_RS09505	CDM120_RS09505	0	1-acyl-sn-glycerol-3-phosphate acyltransferase
CDM120_RS09640	CDM120_RS09640	0	cupin domain-containing protein
CDM120_RS09825	CDM120_RS09825	0	EutN/CcmL family microcompartment protein
CDM120_RS09915	*accC*	1	acetyl-CoA carboxylase biotin carboxylase subunit
CDM120_RS09920	*accB*	0	acetyl-CoA carboxylase biotin carboxyl carrier protein
CDM120_RS10205	CDM120_RS10205	0	hypothetical protein
CDM120_RS10355	CDM120_RS10355	0	tyrosine-type recombinase/integrase
CDM120_RS10360	CDM120_RS10360	0	excisionase
CDM120_RS10365	CDM120_RS10365	0	helix-turn-helix domain-containing protein
CDM120_RS10370	CDM120_RS10370	0	sigma-70 family RNA polymerase sigma factor
CDM120_RS10375	CDM120_RS10375	0	helix-turn-helix transcriptional regulator
CDM120_RS10380	CDM120_RS10380	0	cysteine-rich KTR domain-containing protein
CDM120_RS10385	*tetM*	0	Tc resistance ribosomal protection protein Tet(M)
CDM120_RS10390	CDM120_RS10390	0	conjugal transfer protein
CDM120_RS10395	CDM120_RS10395	0	bifunctional lysozyme/C40 family peptidase
CDM120_RS10400	CDM120_RS10400	0	YtxH domain-containing protein
CDM120_RS10405	CDM120_RS10405	0	pseudogene
CDM120_RS10410	CDM120_RS10410	0	conjugal transfer protein
CDM120_RS10415	CDM120_RS10415	0	antirestriction protein ArdA
CDM120_RS10420	CDM120_RS10420	0	hypothetical protein
CDM120_RS10425	CDM120_RS10425	0	replication initiation factor domain-containing protein
CDM120_RS10430	CDM120_RS10430	0	hypothetical protein
CDM120_RS10435	CDM120_RS10435	0	DUF87 domain-containing protein
CDM120_RS10440	CDM120_RS10440	0	YdcP family protein
CDM120_RS10445	CDM120_RS10445	0	YdcP family protein
CDM120_RS10470	CDM120_RS10470	0	helix-turn-helix domain-containing protein
CDM120_RS10600	CDM120_RS10600	0	hypothetical protein
CDM120_RS10700	CDM120_RS10700	0	hypothetical protein
CDM120_RS10865	CDM120_RS10865	0	QueT transporter family protein
CDM120_RS10895	CDM120_RS10895	0	hypothetical protein

CDM120_RS19645	CDM120_RS19645	0	hypothetical protein
CDM120_RS19370	CDM120_RS19370	0	diguanylate cyclase
CDM120_RS11125	CDM120_RS11125	0	phosphatidate cytidylyltransferase
CDM120_RS11135	CDM120_RS11135	0	hypothetical protein
CDM120_RS19650	CDM120_RS19650	0	hypothetical protein
CDM120_RS11300	CDM120_RS11300	0	4Fe-4S binding protein
CDM120_RS11435	CDM120_RS11435	0	ferritin
CDM120_RS11660	CDM120_RS11660	0	DUF3787 domain-containing protein
CDM120_RS11895	CDM120_RS11895	0	helix-turn-helix transcriptional regulator
CDM120_RS11945	CDM120_RS11945	0	cold-shock protein
CDM120_RS12005	CDM120_RS12005	13	transketolase
CDM120_RS12155	CDM120_RS12155	0	glycine/betaine reductase A
CDM120_RS12580	CDM120_RS12580	0	4Fe-4S binding protein
CDM120_RS12630	CDM120_RS12630	0	cytidine deaminase
CDM120_RS12670	CDM120_RS12670	0	GatB/YqeY domain-containing protein
CDM120_RS12675	CDM120_RS12675	0	30S ribosomal protein S21
CDM120_RS12760	*grpE*	0	nucleotide exchange factor GrpE
CDM120_RS12780	CDM120_RS12780	0	CCA tRNA nucleotidyltransferase
CDM120_RS12810	*rpsT*	0	30S ribosomal protein S20
CDM120_RS19390	CDM120_RS19390	0	
CDM120_RS12890	CDM120_RS12890	0	PTS fructose transporter subunit IIB
CDM120_RS13010	CDM120_RS13010	0	PTS glucose transporter subunit IIA
CDM120_RS19395	CDM120_RS19395	0	helix-turn-helix domain-containing protein
CDM120_RS13255	CDM120_RS13255	0	hypothetical protein
CDM120_RS13325	*rsgA*	0	ribosome small subunit-dependent GTPase A
CDM120_RS13370	*def*	0	peptide deformylase
CDM120_RS13375	*priA*	215	primosomal protein N'
CDM120_RS13545	CDM120_RS13545	0	cell division protein SepF
CDM120_RS13595	CDM120_RS13595	0	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
CDM120_RS13635	CDM120_RS13635	0	response regulator transcription factor
CDM120_RS13675	*ftsZ*	0	cell division protein FtsZ
CDM120_RS13770	CDM120_RS13770	2	AraC family transcriptional regulator
CDM120_RS13860	CDM120_RS13860	0	flavodoxin family protein
CDM120_RS19670	CDM120_RS19670	0	hypothetical protein
CDM120_RS13885	CDM120_RS13885	0	alpha/beta-type small acid-soluble spore protein
CDM120_RS18590	*rnpB*	0	
CDM120_RS14225	CDM120_RS14225	0	D-tyrosyl-tRNA(Tyr) deacylase
CDM120_RS14235	CDM120_RS14235	0	adenine phosphoribosyltransferase
CDM120_RS14335	CDM120_RS14335	1	LCP family protein
CDM120_RS14350	CDM120_RS14350	2	polysaccharide biosynthesis protein

CDM120_RS14370	CDM120_RS14370	2	glycosyltransferase family 2 protein
CDM120_RS14415	CDM120_RS14415	3	cell wall-binding protein Cwp7
CDM120_RS14495	*scfA*	0	six-cysteine ranthipeptide SCIFF
CDM120_RS14995	CDM120_RS14995	0	V-type ATP synthase subunit K
CDM120_RS15005	CDM120_RS15005	0	hypothetical protein
CDM120_RS16045	*secG*	0	preprotein translocase subunit SecG
CDM120_RS16070	CDM120_RS16070	0	triose-phosphate isomerase
CDM120_RS16085	CDM120_RS16085	0	central glycolytic genes regulator
CDM120_RS16340	CDM120_RS16340	0	small
CDM120_RS16360	CDM120_RS16360	0	aspartate-semialdehyde dehydrogenase
CDM120_RS16825	CDM120_RS16825	0	helix-turn-helix transcriptional regulator
CDM120_RS17030	CDM120_RS17030	0	RNHCP domain-containing protein
CDM120_RS19700	CDM120_RS19700	0	hypothetical protein
CDM120_RS17170	*rrf*	0	
CDM120_RS17430	CDM120_RS17430	0	holo-ACP synthase
CDM120_RS17445	*atpG*	0	ATP synthase F1 subunit gamma
CDM120_RS17465	*atpE*	0	ATP synthase F0 subunit C
CDM120_RS17475	CDM120_RS17475	0	ATP synthase subunit I
CDM120_RS17480	CDM120_RS17480	0	AtpZ/AtpI family protein
CDM120_RS17520	*prfA*	1	peptide chain release factor 1
CDM120_RS17535	*rpmE*	0	50S ribosomal protein L31
CDM120_RS17675	CDM120_RS17675	5	ribose-phosphate pyrophosphokinase
CDM120_RS17850	*lysS*	1	lysine--tRNA ligase
CDM120_RS17855	*greA*	0	transcription elongation factor GreA
CDM120_RS17920	CDM120_RS17920	1	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase
CDM120_RS17990	CDM120_RS17990	0	single-stranded DNA-binding protein
CDM120_RS18390	CDM120_RS18390	1	DnaD domain protein
CDM120_RS18405	CDM120_RS18405	0	
CDM120_RS18410	CDM120_RS18410	0	
CDM120_RS18415	CDM120_RS18415	0	
CDM120_RS18455	CDM120_RS18455	0	30S ribosomal protein S18
CDM120_RS18465	CDM120_RS18465	0	30S ribosomal protein S6
CDM120_RS18475	CDM120_RS18475	0	DUF951 domain-containing protein
CDM120_RS18550	*rnpA*	0	ribonuclease P protein component

8. References

1. Hall I, O.T.E., *Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, Bacillus difficilis.* Am J Dis Child, 1935. **49**(390).
2. Lawson, P.A., et al., *Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prevot 1938.* Anaerobe, 2016. **40**: p. 95-9.
3. Smits, W.K., et al., *Clostridium difficile infection.* Nature reviews. Disease primers, 2016. **2**: p. 16020-16020.
4. Bartlett, J.G., *Clostridium difficile: history of its role as an enteric pathogen and the current state of knowledge about the organism.* Clin Infect Dis, 1994. **18 Suppl 4**: p. S265-72.
5. George, W.L., et al., *Aetiology of antimicrobial-agent-associated colitis.* Lancet, 1978. **1**(8068): p. 802-3.
6. Rupnik, M., M.H. Wilcox, and D.N. Gerding, *Clostridium difficile infection: new developments in epidemiology and pathogenesis.* Nat Rev Microbiol, 2009. **7**(7): p. 526-36.
7. Kelly, C.P. and J.T. LaMont, *Clostridium difficile--more difficult than ever.* N Engl J Med, 2008. **359**(18): p. 1932-40.
8. Goorhuis, A., et al., *Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078.* Clin Infect Dis, 2008. **47**(9): p. 1162-70.
9. Cole, S.A. and T.J. Stahl, *Persistent and Recurrent Clostridium difficile Colitis.* Clin Colon Rectal Surg, 2015. **28**(2): p. 65-9.
10. England, P.H., *Voluntary surveillance of Clostridium difficile, England, Wales and Northern*

Ireland: 2014. 2015. **9**(21): p. 5-12.

11. Heinlen, L. and J.D. Ballard, *Clostridium difficile Infection.* Am J Med Sci, 2010. **340**(3): p. 247-52.
12. Dubberke, E.R. and M.A. Olsen, *Burden of Clostridium difficile on the healthcare system.* Clin Infect Dis, 2012. **55 Suppl 2**: p. S88-92.
13. He, M., et al., *Evolutionary dynamics of Clostridium difficile over short and long time scales.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(16): p. 7527-7532.
14. Paredes-Sabja, D., A. Shen, and J.A. Sorg, *Clostridium difficile spore biology: sporulation, germination, and spore structural proteins.* Trends in microbiology, 2014. **22**(7): p. 406-416.
15. Hernandez-Rocha, C., et al., *[Clostridium difficile associated infections: an updated view].* Rev Chilena Infectol, 2012. **29**(4): p. 434-45.
16. Aktories, K., C. Schwan, and T. Jank, *Clostridium difficile Toxin Biology.* Annu Rev Microbiol, 2017. **71**: p. 281-307.
17. Miller, B.A., et al., *Comparison of the burdens of hospital-onset, healthcare facility-associated Clostridium difficile Infection and of healthcare-associated infection due to methicillin-resistant Staphylococcus aureus in community hospitals.* Infect Control Hosp Epidemiol, 2011. **32**(4): p. 387-90.
18. Tenover, F.C., et al., *Laboratory Diagnosis of Clostridium difficile Infection: Can Molecular Amplification Methods Move Us Out of Uncertainty?* The Journal of Molecular Diagnostics : JMD, 2011. **13**(6): p. 573-582.

19. Gürtler, V., *Typing of Clostridium difficile strains by PCR-amplification of variable length 16S-23S rDNA spacer regions.* J Gen Microbiol, 1993. **139**(12): p. 3089-97.
20. O'Neill, G.L., et al., *Modification of a PCR Ribotyping Method for Application as a Routine Typing Scheme forClostridium difficile.* Anaerobe, 1996. **2**(4): p. 205-209.
21. Janezic, S., *Direct PCR-Ribotyping of Clostridium difficile.* Methods Mol Biol, 2016. **1476**: p. 15-21.
22. Bartlett, J.G., *Narrative review: the new epidemic of Clostridium difficile-associated enteric disease.* Ann Intern Med, 2006. **145**(10): p. 758-64.
23. Cheknis, A.K., et al., *Distribution of Clostridium difficile strains from a North American, European and Australian trial of treatment for C. difficile infections: 2005-2007.* Anaerobe, 2009. **15**(6): p. 230-3.
24. Baldan, R., et al., *Clostridium difficile PCR Ribotype 018, a Successful Epidemic Genotype.* J Clin Microbiol, 2015. **53**(8): p. 2575-80.
25. Eyre, D.W., et al., *Emergence and spread of predominantly community-onset Clostridium difficile PCR ribotype 244 infection in Australia, 2010 to 2012.* Euro Surveill, 2015. **20**(10): p. 21059.
26. Nyc, O., et al., *Clostridium difficile PCR ribotype 176 in the Czech Republic and Poland.* Lancet, 2011. **377**(9775): p. 1407.
27. Wilcox, M.H., et al., *Changing epidemiology of Clostridium difficile infection following the introduction of a national ribotyping-based surveillance scheme in England.* Clin Infect Dis, 2012. **55**(8): p. 1056-63.
28. Keel, K., et al., *Prevalence of PCR Ribotypes among Clostridium difficile Isolates from Pigs, Calves, and Other Species.* Journal of Clinical Microbiology, 2007. **45**(6): p. 1963-1964.
29. Kim, H.Y., et al., *High prevalence of Clostridium difficile PCR ribotype 078 in pigs in Korea.* Anaerobe, 2018. **51**: p. 42-46.
30. Stabler, R.A., et al., *Comparative Phylogenomics of Clostridium difficile Reveals Clade Specificity and Microevolution of Hypervirulent Strains.* Journal of Bacteriology, 2006. **188**(20): p. 7297-7305.
31. Gerding, D.N., et al., *Clostridium difficile binary toxin CDT.* Gut Microbes, 2014. **5**(1): p. 15-27.
32. van Dorp, S.M., et al., *Ribotype 078 Clostridium difficile infection incidence in Dutch hospitals is not associated with provincial pig farming: Results from a national sentinel surveillance, 2009-2015.* PLoS One, 2017. **12**(12): p. e0189183.
33. Henriques, A.O. and C.P. Moran, Jr., *Structure, assembly, and function of the spore surface layers.* Annu Rev Microbiol, 2007. **61**: p. 555-88.
34. Edwards, A.N., et al., *Chemical and Stress Resistances of Clostridium difficile Spores and Vegetative Cells.* Frontiers in Microbiology, 2016. **7**: p. 1698.
35. Fawley, W.N., et al., *Efficacy of hospital cleaning agents and germicides against epidemic Clostridium difficile strains.* Infect Control Hosp Epidemiol, 2007. **28**(8): p. 920-5.
36. Setlow, P., *Germination of Spores of Bacillus Species: What We Know and Do Not Know.* Journal of Bacteriology, 2014. **196**(7): p. 1297-1305.
37. D. Higgins, J.D., *Recent progress in Bacillus subtilis sporulation.* FEMS Microbiol Rev. **36 (2012)**: p. 131-148.
38. McKenney, P.T., A. Driks, and P. Eichenberger, *The Bacillus subtilis endospore: assembly and functions of the multilayered coat.* Nat Rev Microbiol, 2013. **11**(1): p. 33-44.
39. Permpoonpattana, P., et al., *Surface layers of Clostridium difficile endospores.* J Bacteriol, 2011. **193**(23): p. 6461-70.
40. Diaz-Gonzalez, F., et al., *Protein composition of the outermost exosporium-like layer of Clostridium difficile 630 spores.* J Proteomics, 2015. **123**: p. 1-13.
41. Talukdar, P.K., et al., *Updates on the sporulation process in Clostridium species.* Res Microbiol, 2015. **166**(4): p. 225-35.

42. Durre, P., *Ancestral sporulation initiation.* Mol Microbiol, 2011. **80**(3): p. 584-7.
43. Steiner, E., et al., *Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in Clostridium acetobutylicum.* Mol Microbiol, 2011. **80**(3): p. 641-54.
44. Stephenson, K. and R.J. Lewis, *Molecular insights into the initiation of sporulation in Gram-positive bacteria: new technologies for an old phenomenon.* FEMS Microbiol Rev, 2005. **29**(2): p. 281-301.
45. Underwood, S., et al., *Characterization of the Sporulation Initiation Pathway of Clostridium difficile and Its Role in Toxin Production.* Journal of Bacteriology, 2009. **191**(23): p. 7296-7305.
46. Al-Hinai, M.A., S.W. Jones, and E.T. Papoutsakis, *The Clostridium sporulation programs: diversity and preservation of endospore differentiation.* Microbiol Mol Biol Rev, 2015. **79**(1): p. 19-37.
47. Pettit, L.J., et al., *Functional genomics reveals that Clostridium difficile Spo0A coordinates sporulation, virulence and metabolism.* BMC Genomics, 2014. **15**: p. 160.
48. Dawson, L.F., et al., *Characterisation of Clostridium difficile biofilm formation, a role for Spo0A.* PLoS One, 2012. **7**(12): p. e50527.
49. Mackin, K.E., et al., *Spo0A differentially regulates toxin production in evolutionarily diverse strains of Clostridium difficile.* PLoS One, 2013. **8**(11): p. e79666.
50. Kroos, L., et al., *Control of sigma factor activity during Bacillus subtilis sporulation.* Mol Microbiol, 1999. **31**(5): p. 1285-94.
51. Fimlaid, K.A., et al., *Global Analysis of the Sporulation Pathway of Clostridium difficile.* PLoS Genetics, 2013. **9**(8): p. e1003660.
52. de Hoon, M.J.L., P. Eichenberger, and D. Vitkup, *Hierarchical evolution of the bacterial sporulation network.* Current biology : CB, 2010. **20**(17): p. R735-R745.
53. Pereira, F.C., et al., *The spore differentiation pathway in the enteric pathogen Clostridium difficile.* PLoS Genet, 2013. **9**(10): p. e1003782.
54. Al-Hinai, M.A., S.W. Jones, and E.T. Papoutsakis, *sigmaK of Clostridium acetobutylicum is the first known sporulation-specific sigma factor with two developmentally separated roles, one early and one late in sporulation.* J Bacteriol, 2014. **196**(2): p. 287-99.
55. Harry, K.H., et al., *Sporulation and enterotoxin (CPE) synthesis are controlled by the sporulation-specific sigma factors SigE and SigK in Clostridium perfringens.* J Bacteriol, 2009. **191**(8): p. 2728-42.
56. Yutin, N. and M.Y. Galperin, *A genomic update on clostridial phylogeny: Gram-negative spore-formers and other misplaced clostridia.* Environ Microbiol, 2013. **15**(10): p. 2631-41.
57. Howerton, A., N. Ramirez, and E. Abel-Santos, *Mapping Interactions between Germinants and Clostridium difficile Spores.* J Bacteriol, 2011. **193**(1): p. 274-82.
58. Lawley, T.D., et al., *Proteomic and genomic characterization of highly infectious Clostridium difficile 630 spores.* J Bacteriol, 2009. **191**(17): p. 5377-86.
59. Sorg, J.A. and A.L. Sonenshein, *Bile salts and glycine as cogerminants for Clostridium difficile spores.* J Bacteriol, 2008. **190**(7): p. 2505-12.
60. Theriot, C.M., A.A. Bowman, and V.B. Young, *Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for Clostridium difficile Spore Germination and Outgrowth in the Large Intestine.* mSphere, 2016. **1**(1).
61. Sorg, J.A. and A.L. Sonenshein, *Chenodeoxycholate is an inhibitor of Clostridium difficile spore germination.* J Bacteriol, 2009. **191**(3): p. 1115-7.
62. Lewis, B.B., R.A. Carter, and E.G. Pamer, *Bile acid sensitivity and in vivo virulence of clinical Clostridium difficile isolates.* Anaerobe, 2016. **41**: p. 32-36.
63. Francis, M.B., et al., *Bile acid recognition by the Clostridium difficile germinant receptor, CspC, is important for establishing infection.* PLoS Pathog, 2013. **9**(5): p. e1003356.

64. Gutelius, D., et al., *Functional analysis of SleC from Clostridium difficile: an essential lytic transglycosylase involved in spore germination.* Microbiology, 2014. **160**(Pt 1): p. 209-16.
65. Minton, N.P., et al., *A roadmap for gene system development in Clostridium.* Anaerobe, 2016. **41**: p. 104-112.
66. Schwarz, K.M., et al., *Towards improved butanol production through targeted genetic modification of Clostridium pasteurianum.* Metabolic Engineering, 2017. **40**: p. 124-137.
67. Scott, P.T. and J.I. Rood, *Electroporation-mediated transformation of lysostaphin-treated Clostridium perfringens.* Gene, 1989. **82**(2): p. 327-33.
68. Mermelstein, L.D. and E.T. Papoutsakis, *In vivo methylation in Escherichia coli by the Bacillus subtilis phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of Clostridium acetobutylicum ATCC 824.* Appl Environ Microbiol, 1993. **59**(4): p. 1077-81.
69. Jennert, K.C., et al., *Gene transfer to Clostridium cellulolyticum ATCC 35319.* Microbiology, 2000. **146 Pt 12**: p. 3071-80.
70. Minton, J.D.O.M.L.T.-J.S.J.K.B.D.E.T.N.P., *Introduction of plasmids into whole cells of Clostridium acetobutylicum by electroporation.* FEMS Microbiology Letters, 01 November 1988. **Volume 56**(Issue 1): p. Pages 83–88,

.
71. Pyne, M.E., et al., *Improved Draft Genome Sequence of Clostridium pasteurianum Strain ATCC 6013 (DSM 525) Using a Hybrid Next-Generation Sequencing Approach.* Genome Announc, 2014. **2**(4).
72. Jiraskova, A., et al., *Rapid protocol for electroporation of Clostridium perfringens.* J Microbiol Methods, 2005. **62**(1): p. 125-7.
73. Ackermann, G., et al., *Electroporation of DNA sequences from the pathogenicity locus (PaLoc) of toxigenic Clostridium difficile into a non-toxigenic strain.* Mol Cell Probes, 2001. **15**(5): p. 301-6.
74. Purdy, D., et al., *Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier.* Mol Microbiol, 2002. **46**(2): p. 439-52.
75. !!! INVALID CITATION !!! {}.
76. Liyanage, H., et al., *Clostridium beijerinckii and Clostridium difficile detoxify methylglyoxal by a novel mechanism involving glycerol dehydrogenase.* Appl Environ Microbiol, 2001. **67**(5): p. 2004-10.
77. Kirk, J.A. and R.P. Fagan, *Heat shock increases conjugation efficiency in Clostridium difficile.* Anaerobe, 2016. **42**: p. 1-5.
78. Barquist, L., C.J. Boinett, and A.K. Cain, *Approaches to querying bacterial genomes with transposon-insertion sequencing.* RNA Biol, 2013. **10**(7): p. 1161-9.
79. Muñoz-López, M. and J.L. García-Pérez, *DNA Transposons: Nature and Applications in Genomics.* Current Genomics, 2010. **11**(2): p. 115-128.
80. Casacuberta, E. and J. Gonzalez, *The impact of transposable elements in environmental adaptation.* Mol Ecol, 2013. **22**(6): p. 1503-17.
81. Reilly, M.T., et al., *The Role of Transposable Elements in Health and Diseases of the Central Nervous System.* J Neurosci, 2013. **33**(45): p. 17577-86.
82. Levin, H.L. and J.V. Moran, *Dynamic interactions between transposable elements and their hosts.* Nat Rev Genet, 2011. **12**(9): p. 615-27.
83. Ravindran, S., *Barbara McClintock and the discovery of jumping genes.* Proc Natl Acad Sci U S A, 2012. **109**(50): p. 20198-9.
84. Mullany, P., E. Allan, and A.P. Roberts, *Mobile genetic elements in Clostridium difficile and their role in genome function.* Res Microbiol, 2015. **166**(4): p. 361-7.

85. Berg, D.E., et al., *Transposition of R factor genes to bacteriophage lambda.* Proceedings of the National Academy of Sciences of the United States of America, 1975. **72**(9): p. 3628-3632.
86. Dembek, M., et al., *High-throughput analysis of gene essentiality and sporulation in Clostridium difficile.* MBio, 2015. **6**(2): p. e02383.
87. Langridge, G.C., et al., *Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants.* Genome Res, 2009. **19**(12): p. 2308-16.
88. Mullany, P., et al., *Behavior and target site selection of conjugative transposon Tn916 in two different strains of toxigenic Clostridium difficile.* Appl Environ Microbiol, 2012. **78**(7): p. 2147-53.
89. Robertson, H.M. and D.J. Lampe, *Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera.* Mol Biol Evol, 1995. **12**(5): p. 850-62.
90. Lampe, D.J., et al., *Hyperactive transposase mutants of the Himar1 mariner transposon.* Proc Natl Acad Sci U S A, 1999. **96**(20): p. 11428-33.
91. Cartman, S.T. and N.P. Minton, *A mariner-based transposon system for in vivo random mutagenesis of Clostridium difficile.* Appl Environ Microbiol, 2010. **76**(4): p. 1103-9.
92. Moncrief, J.S., L.A. Barroso, and T.D. Wilkins, *Positive regulation of Clostridium difficile toxins.* Infect Immun, 1997. **65**(3): p. 1105-8.
93. Geissendorfer, M. and W. Hillen, *Regulated expression of heterologous genes in Bacillus subtilis using the Tn10 encoded tet regulatory elements.* Appl Microbiol Biotechnol, 1990. **33**(6): p. 657-63.
94. Fagan, R.P. and N.F. Fairweather, *Clostridium difficile has two parallel and essential Sec secretion systems.* J Biol Chem, 2011. **286**(31): p. 27483-93.
95. Hutchison, C.A., et al., *Global transposon mutagenesis and a minimal Mycoplasma genome.* Science, 1999. **286**(5447): p. 2165-9.
96. Gallagher, L.A., J. Shendure, and C. Manoil, *Genome-scale identification of resistance functions in Pseudomonas aeruginosa using Tn-seq.* MBio, 2011. **2**(1): p. e00315-10.
97. Goodman, A.L., et al., *Identifying genetic determinants needed to establish a human gut symbiont in its habitat.* Cell Host Microbe, 2009. **6**(3): p. 279-89.
98. Barquist, L., et al., *The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries.* Bioinformatics, 2016. **32**(7): p. 1109-11.
99. Eckert, S.E., et al., *Retrospective application of transposon-directed insertion site sequencing to a library of signature-tagged mini-Tn5Km2 mutants of Escherichia coli O157:H7 screened in cattle.* J Bacteriol, 2011. **193**(7): p. 1771-6.
100. Dziva, F., et al., *Sequencing and functional annotation of avian pathogenic Escherichia coli serogroup O78 strains reveal the evolution of E. coli lineages pathogenic for poultry via distinct mechanisms.* Infect Immun, 2013. **81**(3): p. 838-49.
101. Weerdenburg, E.M., et al., *Genome-wide transposon mutagenesis indicates that Mycobacterium marinum customizes its virulence mechanisms for survival and replication in different hosts.* Infect Immun, 2015. **83**(5): p. 1778-88.
102. Woods, C., et al., *A novel conjugal donor strain for improved DNA transfer into Clostridium spp.* Anaerobe, 2019. **59**: p. 184-191.
103. Hussain, H.A., A.P. Roberts, and P. Mullany, *Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites.* J Med Microbiol, 2005. **54**(Pt 2): p. 137-41.
104. Heeg, D., et al., *Spores of Clostridium difficile clinical isolates display a diverse germination response to bile salts.* PLoS One, 2012. **7**(2): p. e32381.
105. Bertani, G. and J.J. Weigle, *Host controlled variation in bacterial viruses.* J Bacteriol, 1953. **65**(2): p. 113-21.

106. Luria, S.E. and M.L. Human, *A nonhereditary, host-induced variation of bacterial viruses.* J Bacteriol, 1952. **64**(4): p. 557-69.
107. Vasu, K. and V. Nagaraja, *Diverse functions of restriction-modification systems in addition to cellular defense.* Microbiol Mol Biol Rev, 2013. **77**(1): p. 53-72.
108. Hagemann, M., et al., *Identification of the DNA methyltransferases establishing the methylome of the cyanobacterium Synechocystis sp. PCC 6803.* DNA Res, 2018. **25**(4): p. 343-352.
109. Ershova, A.S., et al., *Role of Restriction-Modification Systems in Prokaryotic Evolution and Ecology.* Biochemistry (Mosc), 2015. **80**(10): p. 1373-86.
110. Rao, D.N., D.T. Dryden, and S. Bheemanaik, *Type III restriction-modification enzymes: a historical perspective.* Nucleic Acids Res, 2014. **42**(1): p. 45-55.
111. Adhikari, S. and P.D. Curtis, *DNA methyltransferases and epigenetic regulation in bacteria.* FEMS Microbiol Rev, 2016. **40**(5): p. 575-91.
112. Rodic, A., et al., *Understanding key features of bacterial restriction-modification systems through quantitative modeling.* BMC Syst Biol, 2017. **11**(Suppl 1): p. 377.
113. Atack, J.M., et al., *Phasevarions of Bacterial Pathogens: Methylomics Sheds New Light on Old Enemies.* Trends Microbiol, 2018. **26**(8): p. 715-726.
114. Oultram, J.D., et al., *Introduction of plasmids into whole cells of Clostridium acetobutylicum by electroporation.* FEMS Microbiology Letters, 1988. **56**(1): p. 83-88.
115. Herbert, M., et al., *Gene transfer into Clostridium difficile CD630 and characterisation of its methylase genes.* FEMS Microbiol Lett, 2003. **229**(1): p. 103-10.
116. Meaney, C.A., et al., *The role of small acid-soluble proteins (SASPs) in protection of spores of Clostridium botulinum against nitrous acid.* Int J Food Microbiol, 2016. **216**: p. 25-30.
117. Lesiak, J.M., W. Liebl, and A. Ehrenreich, *Development of an in vivo methylation system for the solventogen Clostridium saccharobutylicum NCP 262 and analysis of two endonuclease mutants.* J Biotechnol, 2014. **188**: p. 97-9.
118. Roberts, R.J., et al., *REBASE--a database for DNA restriction and modification: enzymes, genes and genomes.* Nucleic Acids Res, 2015. **43**(Database issue): p. D298-9.
119. Roberts, R.J., et al., *REBASE--enzymes and genes for DNA restriction and modification.* Nucleic Acids Res, 2007. **35**(Database issue): p. D269-70.
120. Clark, T.A., et al., *Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing.* Nucleic Acids Res, 2012. **40**(4): p. e29.
121. Mermelstein, L.D., et al., *Expression of cloned homologous fermentative genes in Clostridium acetobutylicum ATCC 824.* Biotechnology (N Y), 1992. **10**(2): p. 190-5.
122. Davis, T.O., et al., *Development of a transformation and gene reporter system for group II, non-proteolytic Clostridium botulinum type B strains.* J Mol Microbiol Biotechnol, 2000. **2**(1): p. 59-69.
123. Mullany, P., et al., *Gene cloning in Clostridium difficile using Tn916 as a shuttle conjugative transposon.* Plasmid, 1994. **31**(3): p. 320-3.
124. Mani, N., et al., *Environmental response and autoregulation of Clostridium difficile TxeR, a sigma factor for toxin gene expression.* J Bacteriol, 2002. **184**(21): p. 5971-8.
125. Cartman, S.T., et al., *The emergence of 'hypervirulence' in Clostridium difficile.* International Journal of Medical Microbiology, 2010. **300**(6): p. 387-395.
126. Woods, C., et al., *A novel conjugal donor strain for improved DNA transfer into Clostridium spp.* Anaerobe, 2019. **59**: p. 184-191.
127. McIntyre, A.B.R., et al., *Single-molecule sequencing detection of N6-methyladenine in microbial reference materials.* Nat Commun, 2019. **10**(1): p. 579.
128. Gouil, Q. and A. Keniry, *Latest techniques to study DNA methylation.* Essays Biochem, 2019. **63**(6): p. 639-648.
129. Beaulaurier, J., E.E. Schadt, and G. Fang, *Deciphering bacterial epigenomes using modern sequencing technologies.* Nat Rev Genet, 2019. **20**(3): p. 157-172.

130. Kurka, H., et al., *Sequence similarity of Clostridium difficile strains by analysis of conserved genes and genome content is reflected by their ribotype affiliation.* PloS one, 2014. **9**(1): p. e86535-e86535.
131. Richter, M. and R. Rosselló-Móra, *Shifting the genomic gold standard for the prokaryotic species definition.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(45): p. 19126-19131.
132. Harris, A.J. and A.D. Goldman, *The complex phylogenetic relationships of a 4mC/6mA DNA methyltransferase in prokaryotes.* Mol Phylogenet Evol, 2020. **149**: p. 106837.
133. Marinus, M.G. and N.R. Morris, *Biological function for 6-methyladenine residues in the DNA of Escherichia coli K12.* Journal of Molecular Biology, 1974. **85**(2): p. 309-322.
134. Furuta, Y. and I. Kobayashi, *Restriction-Modification Systems as Mobile Epigenetic Elements*. 2013.
135. Kobayashi, I., *Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution.* Nucleic Acids Res, 2001. **29**(18): p. 3742-56.
136. Chinen, A., et al., *Evolution of sequence recognition by restriction-modification enzymes: selective pressure for specificity decrease.* Mol Biol Evol, 2000. **17**(11): p. 1610-9.
137. Glickman, B.W. and M. Radman, *Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction.* Proc Natl Acad Sci U S A, 1980. **77**(2): p. 1063-7.
138. Marinus, M.G., et al., *Insertion mutations in the dam gene of Escherichia coli K-12.* Mol Gen Genet, 1983. **192**(1-2): p. 288-9.
139. Hargreaves, K.R., et al., *As Clear as Mud? Determining the Diversity and Prevalence of Prophages in the Draft Genomes of Estuarine Isolates of Clostridium difficile.* Genome Biol Evol, 2015. **7**(7): p. 1842-55.
140. Corver, J., et al., *Analysis of a Clostridium difficile PCR ribotype 078 100 kilobase island reveals the presence of a novel transposon, Tn6164.* BMC Microbiol, 2012. **12**: p. 130.
141. Ruoff, K.L., *Streptococcus anginosus ("Streptococcus milleri"): the unrecognized pathogen.* Clin Microbiol Rev, 1988. **1**(1): p. 102-8.
142. Jeltsch, A. and A. Pingoud, *Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems.* J Mol Evol, 1996. **42**(2): p. 91-6.
143. Obarska-Kosinska, A., et al., *HsdR subunit of the type I restriction-modification enzyme EcoR124I: biophysical characterisation and structural modelling.* Journal of molecular biology, 2008. **376**(2): p. 438-452.
144. Vidal, J.E., et al., *Use of an EZ-Tn5-based random mutagenesis system to identify a novel toxin regulatory locus in Clostridium perfringens strain 13.* PLoS One, 2009. **4**(7): p. e6232.
145. Gao, L.Y., et al., *Transposon mutagenesis of Mycobacterium marinum identifies a locus linking pigmentation and intracellular survival.* Infect Immun, 2003. **71**(2): p. 922-9.
146. Le Breton, Y., N.P. Mohapatra, and W.G. Haldenwang, *In vivo random mutagenesis of Bacillus subtilis by use of TnYLB-1, a mariner-based transposon.* Appl Environ Microbiol, 2006. **72**(1): p. 327-33.
147. Maier, T.M., et al., *In vivo Himar1-based transposon mutagenesis of Francisella tularensis.* Appl Environ Microbiol, 2006. **72**(3): p. 1878-85.
148. Cao, M., A.P. Bitar, and H. Marquis, *A mariner-based transposition system for Listeria monocytogenes.* Appl Environ Microbiol, 2007. **73**(8): p. 2758-61.
149. Wilson, A.C., M. Perego, and J.A. Hoch, *New transposon delivery plasmids for insertional mutagenesis in Bacillus anthracis.* J Microbiol Methods, 2007. **71**(3): p. 332-5.
150. Hartman, A.H., H. Liu, and S.B. Melville, *Construction and characterization of a lactose-inducible promoter system for controlled gene expression in Clostridium perfringens*. Appl Environ Microbiol, 2011. **77**(2): p. 471-8.

151. Lanckriet, A., et al., *Generation of single-copy transposon insertions in Clostridium perfringens by electroporation of phage mu DNA transposition complexes.* Appl Environ Microbiol, 2009. **75**(9): p. 2638-42.
152. Jasni, A.S., et al., *Demonstration of Conjugative Transposon (Tn5397)-Mediated Horizontal Gene Transfer between Clostridium difficile and Enterococcus faecalis.* Antimicrob Agents Chemother, 2010. **54**(11): p. 4924-6.
153. Liu, H., et al., *Use of a mariner-based transposon mutagenesis system to isolate Clostridium perfringens mutants deficient in gliding motility.* Journal of bacteriology, 2013. **195**(3): p. 629-636.
154. Zhang, Y., A. Grosse-Honebrink, and N.P. Minton, *A universal mariner transposon system for forward genetic studies in the genus Clostridium.* PLoS One, 2015. **10**(4): p. e0122411.
155. Shoemaker, N.B., et al., *Tn4351 transposes in Bacteroides spp. and mediates the integration of plasmid R751 into the Bacteroides chromosome.* Journal of bacteriology, 1986. **165**(3): p. 929-936.
156. Mani, N. and B. Dupuy, *Regulation of toxin synthesis in Clostridium difficile by an alternative RNA polymerase sigma factor.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(10): p. 5844-5849.
157. Corrigan, R.M. and T.J. Foster, *An improved tetracycline-inducible expression vector for Staphylococcus aureus.* Plasmid, 2009. **61**(2): p. 126-9.
158. Bhattacharjee, D. and J.A. Sorg, *Factors and Conditions That Impact Electroporation of Clostridioides difficile Strains.* mSphere, 2020. **5**(2).
159. Xu, T., et al., *Efficient Genome Editing in Clostridium cellulolyticum via CRISPR-Cas9 Nickase.* Appl Environ Microbiol, 2015. **81**(13): p. 4423-31.
160. Zhang, L., et al., *Highly efficient in vitro site-specific recombination system based on streptomyces phage phiBT1 integrase.* J Bacteriol, 2008. **190**(19): p. 6392-7.
161. Spigaglia, P., F. Barbanti, and P. Mastrantonio, *Detection of a genetic linkage between genes coding for resistance to tetracycline and erythromycin in Clostridium difficile.* Microb Drug Resist, 2007. **13**(2): p. 90-5.
162. Peng, Z., et al., *Update on Antimicrobial Resistance in Clostridium difficile: Resistance Mechanisms and Antimicrobial Susceptibility Testing.* Journal of clinical microbiology, 2017. **55**(7): p. 1998-2008.
163. Spigaglia, P., *Recent advances in the understanding of antibiotic resistance in Clostridium difficile infection.* Ther Adv Infect Dis, 2016. **3**(1): p. 23-42.
164. Spigaglia, P., F. Barbanti, and P. Mastrantonio, *Horizontal transfer of erythromycin resistance from Clostridium difficile to Butyrivibrio fibrisolvens.* Antimicrob Agents Chemother, 2005. **49**(12): p. 5142-5.
165. Spigaglia, P., P. Mastrantonio, and F. Barbanti, *Antibiotic Resistances of Clostridium difficile.* Adv Exp Med Biol, 2018. **1050**: p. 137-159.
166. Cartman, S.T., et al., *Precise manipulation of the Clostridium difficile chromosome reveals a lack of association between the tcdC genotype and toxin production.* Appl Environ Microbiol, 2012. **78**(13): p. 4683-90.
167. Lin, D.M., B. Koskella, and H.C. Lin, *Phage therapy: An alternative to antibiotics in the age of multi-drug resistance.* World J Gastrointest Pharmacol Ther, 2017. **8**(3): p. 162-173.
168. Syed, F., H. Grunenwald, and N. Caruccio, *Next-generation sequencing library preparation: simultaneous fragmentation and tagging using in vitro transposition.* Nature Methods, 2009. **6**(11): p. i-ii.
169. Woods, C., *Forward Genetics in Clostridial Acetogens as a Route to Process Improvements.* Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, 2017.

170. Starr, J.M., et al., *Spatio-temporal stochastic modelling of Clostridium difficile.* J Hosp Infect, 2009. **71**(1): p. 49-56.
171. McDonald, L.C., et al., *Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA).* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2018. **66**(7): p. e1-e48.
172. Hutchison, C.A., 3rd, et al., *Design and synthesis of a minimal bacterial genome.* Science, 2016. **351**(6280): p. aad6253.
173. Oliveira, P.H., et al., *Epigenomic characterization of Clostridioides difficile finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis.* Nat Microbiol, 2020. **5**(1): p. 166-180.
174. Wilkowska, K., et al., *Low-level expression of the Type II restriction–modification system confers potent bacteriophage resistance in Escherichia coli.* DNA Research, 2020. **27**(1).
175. Fels, S.R., et al., *Rapid transposon liquid enrichment sequencing (TnLE-seq) for gene fitness evaluation in underdeveloped bacterial systems.* Appl Environ Microbiol, 2013. **79**(23): p. 7510-7.
176. Charbonneau, A.R.L., et al., *Defining the ABC of gene essentiality in streptococci.* BMC Genomics, 2017. **18**(1): p. 426.
177. Le Breton, Y., et al., *Essential Genes in the Core Genome of the Human Pathogen Streptococcus pyogenes.* Sci Rep, 2015. **5**: p. 9838.
178. Sholeh, M., et al., *Antimicrobial resistance in Clostridioides (Clostridium) difficile derived from humans: a systematic review and meta-analysis.* Antimicrob Resist Infect Control, 2020. **9**(1): p. 158.
179. Jain, S., et al., *Quantitative Proteomic Analysis of the Heat Stress Response in Clostridium difficile Strain 630.* Journal of Proteome Research, 2011. **10**(9): p. 3880-3890.
180. Giordano, N., J.L. Hastie, and P.E. Carlson, *Transcriptomic profiling of Clostridium difficile grown under microaerophillic conditions.* Pathogens and Disease, 2018. **76**(2).