

Investigation into the mobile genetic elements of
Clostridium difficile

Michael Sebastiaan Maria Brouwer

This thesis is submitted for the degree of PhD

Supervised by Dr. Adam P. Roberts, Dr. Elaine Allan and Prof. Peter Mullan

UCL Eastman Dental Institute, Department of Microbial Diseases

Declaration

I, Michael Sebastiaan Maria Brouwer, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Clostridium difficile is a pathogenic bacterium that can colonise both humans and various animals. Toxin production leads to clinical symptoms ranging from mild to severe diarrhoea and can result in potentially fatal pseudomembranous colitis. These symptoms are caused by the disruption of the cytoskeleton and tight junctions of gut epithelial cells by the toxins.

Genomic sequencing of *C. difficile* has indicated the chromosome carries a number of mobile genetic elements including conjugative transposons, which can encode antibiotic resistance genes. Analysing the sequence of a number of *C. difficile* strains indicated that each genome carries at least one and often multiple conjugative transposons. For many of the genes on these elements, functions were predicted using various bioinformatic tools.

The study of conjugative transposons in *C. difficile* has been limited by the lack of resistance genes encoded by the elements. Therefore, an antibiotic resistance gene was inserted into six of the elements in strains 630 and R20291 and filter-matings performed. Conjugative transfer was shown for all elements from strain 630 but not for Tn6103 from R20291.

The study of transconjugants of these matings showed the pathogenicity locus, encoding the two major toxins of *C. difficile*, to transfer at a low frequency into a non-toxicogenic recipient strain. Whole genome sequencing of transconjugants determined that the transfer is not limited to the pathogenicity locus but includes varying sizes of chromosomal DNA flanking the pathogenicity locus.

RNA-seq was used for the comparison of mutants for transcriptional regulators of conjugative transposons CTn2 and CTn4, however no significant differential expression was detected. Furthermore, strain 630 Δ *erm*, a commonly used laboratory strain for the generation of knockout mutants, was compared to the wildtype strain 630. A predicted oxidative stress operon was upregulated in 630 Δ *erm* which raises the question of the biological impact of these results on the knockout model.

Acknowledgements

I am very grateful to all three of my supervisors. Dr. Adam Roberts, Dr. Elaine Allan and Prof. Peter Mullany, thanks for all the help with the lab work, the writing and all the helpful discussions. I would like to thank Dr. Phil Warburton and Miss. Catherine Mackenzie for the work they have performed for this project. Also I would like to thank Dr. Haitham Hussain, Dr. Lena Ciriç, Dr. Rachel Williams and Dr. Rachael Whalan for their great technical assistance in the lab.

Furthermore I would like to thank Dr. Andre Dascal (McGill University) for providing *Clostridium difficile* strains. I would like to thank Dr. Jeroen Corver (Leids Universitair Medisch Centrum) for providing *C. difficile* M120 and for performing part of the bioinformatic analysis on Tn6164. I would like to thank Dr. Antonio Gomez (University of Barcelona) for his help using BYPASS. I would like to thank Dr. Tony Brooks and Dr. Warren Emmett (UCL Genomics) for performing genomic sequencing and RNAseq.

None of this work could have been performed without the support I have had from so many friends I have made at the Eastman; Dr. Zoie Aiken, Dr. Kat McCurrie, Dr. Sarah Tubby, Miss Linda Dekker and Dr. Farah Dalwai. Mr. Daniel Illes and Miss. Helen Garrahan who have been my best friends from the moment I arrived in London. I have also had a lot of love and support my family in the Netherlands. Papa en Mama, Jen en Ralph, lieve Donald, bedankt voor al jullie liefde en steun!

The work described in this thesis was performed as part of the HYPERDIFF project funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 223585.

Contents

Title page.....	1
Declaration.....	2
Abstract.....	3
Acknowledgements.....	5
Contents.....	6
List of Figures.....	12
List of Tables.....	15
List of abbreviations.....	16
Chapter 1 Introduction.....	19
1.1. <i>Clostridium difficile</i>	19
1.1.1. Isolation, pathogenesis and treatment.....	19
1.1.1.1. The isolation of <i>Clostridium difficile</i>	19
1.1.1.2. <i>C. difficile</i> disease and treatment.....	21
1.1.2. Virulence of <i>C. difficile</i>	23
1.1.2.1. <i>C. difficile</i> TcdA and TcdB.....	23
1.1.2.2. <i>C. difficile</i> binary toxin.....	27
1.1.2.3. Other <i>C. difficile</i> virulence factors.....	28
1.1.2.4. Sporulation.....	29
1.1.2.5. Antibiotic resistance in <i>C. difficile</i>	30
1.1.3. <i>C. difficile</i> epidemiology.....	31
1.1.3.1. Typing methods.....	31
1.1.3.2. Hypervirulence.....	32
1.2. Mobile genetic elements.....	35
1.2.1. Conjugation.....	35
1.2.1.1. Conjugative transposons.....	38

1.2.1.1.1.	Tn916.....	40
1.2.1.1.2.	Tn1549.....	42
1.2.1.1.3.	Excision and integration by site-specific recombination	42
1.2.1.2.	MGEs in <i>C. difficile</i>	45
1.2.1.2.1.	Tn5397.....	45
1.2.1.2.2.	CTNs in <i>C. difficile</i> 630.....	47
1.2.1.3.	Mobilisable elements in <i>C. difficile</i>	49
1.2.1.3.1.	Tn4453.....	49
1.2.1.3.2.	Tn5398.....	50
1.2.2.	Transduction	51
1.2.3.	Additional <i>C. difficile</i> MGEs.....	53
1.2.3.1.	<i>skin</i> ^{Cd}	53
1.2.3.2.	IStroms	54
1.2.4.	Transfer of DNA via nanotubes.....	55
1.3.	Aims of this work.....	57
Chapter 2	Materials and methods.....	59
2.1.	Bacterial strains and culture methods	59
2.2.	Molecular techniques.....	61
2.2.1.	DNA isolation.....	61
2.2.2.	PCR	62
2.2.3.	Purification of PCR products	62
2.2.4.	Transformation into competent <i>E. coli</i> CA-434	63
2.2.5.	DNA Sanger sequencing.....	64
2.3.	Filter-mating.....	65
Chapter 3	Bioinformatics search for novel elements and putative virulence factors	
	66	
3.1.	Introduction.....	66

3.2.	Aims of this chapter:	69
3.3.	Materials and methods	70
3.3.1.	Search for novel CTns	70
3.3.2.	Bioinformatics analysis to predict CDS functions	70
3.4.	Results and discussion	72
3.4.1.	CTn1-like elements	76
3.4.2.	CTn4-like elements	84
3.4.3.	CTn5-like elements	86
3.4.4.	Tn916	91
3.4.5.	Tn6164	93
3.4.6.	Non-conjugative genetic elements.....	96
3.4.7.	Cell surface proteins encoded on CTns	97
3.5.	Conclusions.....	99
Chapter 4	Mobility of putative mobile genetic elements.....	103
4.1.	Introduction.....	103
4.1.1.	CTns circularize upon excision	103
4.1.2.	CTn integration sites and copy number	104
4.1.3.	Mobilisation of non-conjugative DNA elements.....	105
4.1.4.	Co-transfer of CTns	106
4.1.5.	CloStron provides <i>C. difficile</i> targeted mutagenesis tool	107
4.1.6.	Using the CloStron system to create selectable markers.....	110
4.2.	Aims of this chapter:	111
4.3.	Materials and methods	112
4.3.1.	<i>C. difficile</i> growth curve	112
4.3.2.	Detection of CTn circularisation and excision from the chromosome ...	112
4.3.3.	CloStron mutagenesis.....	114
4.3.3.1.	CloStron mutagenesis with original protocol	115

4.3.3.2.	Clostron mutagenesis 2.0 with refined protocol.....	117
4.3.4.	Filter-matings.....	117
4.3.5.	Southern blotting.....	119
4.3.6.	PCR analysis for detection of co-transfer	120
4.3.7.	Nanotube transfer of antimicrobial resistance proteins.....	121
4.4.	Results	122
4.4.1.	Circularisation of CTns in different growth phases	122
4.4.2.	Circularisation of <i>C. difficile</i> CTns.....	124
4.4.3.	Clostron mutagenesis.....	128
4.4.4.	Conjugative transfer of CTns and analysis of transconjugants.....	132
4.4.4.1.	Conjugative transfer of CTn1 can occur with co-transfer of CTn5 and Tn5397	132
4.4.4.2.	Conjugative transfer of CTn2	137
4.4.4.3.	Conjugative transfer of CTn4 shows co-transfer of CTn5.....	138
4.4.4.4.	Conjugative transfer of CTn5 into <i>C. difficile</i> and <i>B. subtilis</i>	140
4.4.4.5.	Conjugative transfer of CTn7	142
4.4.4.6.	No conjugative transfer of Tn6103 was observed.....	144
4.4.4.7.	Conjugative transfer of the CTns of strain M120.....	144
4.4.5.	Nanotube transfer of antimicrobial resistance proteins	146
4.5.	Discussion	148
4.5.1.	Circularisation of <i>C. difficile</i> putative CTns	148
4.5.2.	Conjugative transfer of CTns.....	149
4.5.3.	Co-transfer of CTns from strain 630 Δ erm.....	151
4.5.4.	Number of integration sites in transconjugants	152
4.5.5.	<i>C. difficile</i> elements that were not shown to transfer	154
4.5.6.	Nanotube transfer of antimicrobial resistance proteins	156
4.5.7.	Conclusions	157

Chapter 5	Mobilisation of the PaLoc	158
5.1.	Introduction.....	158
5.2.	Aims of this chapter:	159
5.3.	Materials and methods	160
5.3.1.	Tissue culture and cytotoxicity assays.....	160
5.3.2.	Filter-matings.....	161
5.3.2.1.	PCR analysis.....	161
5.3.3.	Phage experiments	162
5.3.3.1.	Preparation of phage suspensions.....	162
5.3.3.2.	Phage infection assay.....	163
5.3.3.3.	Plaque assay.....	163
5.3.4.	DNA shotgun sequencing.....	163
5.4.	Results	165
5.4.1.	Transconjugant Paloc386 is cytotoxic.....	165
5.4.2.	PaLoc transfer is reproducible	166
5.4.3.	Phage assay to check for transfer of PaLoc	167
5.4.4.	Full genome sequencing	168
5.4.4.1.	CD37 genome shotgun sequence	168
5.4.4.2.	Sequencing of transconjugants.....	169
5.5.	Discussion	173
5.5.1.	Transfer of the PaLoc	173
5.5.2.	Conclusions	176
Chapter 6	The effects of CTNs on the transcriptome	177
6.1.	Introduction.....	177
6.2.	Aims of this chapter:	181
6.3.	Materials and methods	182
6.3.2.	RNA sequencing.....	182

6.4.	Results	186
6.4.1.	Transcription of CD0428 and CD1099	186
6.4.2.	Clostron mutants in R20291 alternative sigma factors	186
6.4.3.	RNA-seq results.....	189
6.4.3.1.	Biological variation between the samples	189
6.4.3.2.	Differential expression of the CTn2 mutant	191
6.4.3.3.	Differential expression of the CTn4 mutant	192
6.4.3.4.	Differential expression of 630 vs 630 Δ <i>erm</i>	193
6.5.	Discussion	195
6.5.1.	Design of the experimental methods	195
6.5.2.	Effects of CTn mutations on the transcriptome	196
6.5.3.	Comparison of the 630 and 630 Δ <i>erm</i> transcriptomes	197
6.5.4.	Conclusions	198
Chapter 7	General conclusions	200
Chapter 8	References.....	206
Appendix I	222
Appendix II	223
Appendix III	224

List of Figures

Figure 1.1 Gram staining of <i>C. difficile</i> strain 630.....	21
Figure 1.2 Effects of <i>C. difficile</i> infection.....	22
Figure 1.3 Schematic representation of the <i>C. difficile</i> PaLoc of strain 630.....	24
Figure 1.4 Schematic of <i>C. difficile</i> TcdA and TcdB.	24
Figure 1.5 Schematic representation of the <i>C. difficile</i> CDT locus of strain R20291.	28
Figure 1.6 Number of deaths attributed to <i>C. difficile</i>	34
Figure 1.7 Schematic representation of the transfer of a conjugative plasmid.	37
Figure 1.8 Schematic representation of the type IV secretion system of <i>A. tumefaciens</i>	38
Figure 1.9 Schematic of transposition and conjugative transfer.....	39
Figure 1.10 Schematic representation of Tn916.....	41
Figure 1.11 Schematic representation of Tn1549.....	42
Figure 1.12 Schematic representation of the insertion of Tn916.....	44
Figure 1.13 Schematic representation of the insertion of Tn5397.....	45
Figure 1.14 Schematic representation of Tn5397.....	46
Figure 1.15 Comparison of Tn916 to Tn5397, CTn1, CTn7 and CTn6.....	48
Figure 1.16 Comparison of Tn1549 to CTn4, CTn5 and CTn2.....	49
Figure 1.17 Schematic representation of Tn4453.....	50
Figure 1.18 Schematic representation of Tn5398.....	51
Figure 1.19 Schematic representation of <i>skin</i> ^{Cd} in strain 630.	54
Figure 3.1 Schematic representation of CTn1 and the CTn1-like elements in strains R20291, QCD-66C26, ATCC 43255, 23M63 and QCD-63Q42.....	78
Figure 3.2 Schematic representation of CTn1 and phage 1 in 630 and the CTn1-like element in QCD-63Q42.	81
Figure 3.3 Schematic representation of CTn1 and the CTn1-like elements in strains 2007855 and M68.	83

Figure 3.4 Schematic representation of CTn4 and the CTn4-like element in strain QCD-23M63.	85
Figure 3.5 Schematic representation of CTn5 and the CTn5-like elements in strains R20291, QCD-66C26, QCD-63Q42 and QCD-23M63.....	87
Figure 3.6 Schematic representation of CTn5 and the insertions present in Tn6103....	88
Figure 3.7 Schematic representation of CTn5 and the CTn5-like element in strain 2007855.....	90
Figure 3.8 Schematic representation of Tn916 compared the Tn916-like element in strain M68 and CTn1.....	92
Figure 3.9 Schematic representation of Tn6190 in M120 compared to Tn916 in <i>E. faecalis</i>	93
Figure 3.10 Schematic representation of Tn6164 and <i>E. faecalis</i> pEF418, <i>S. pneumoniae</i> Tn1806, <i>Thermoanaerobacter</i> sp. prophage and the <i>C. fetus</i> pathogenicity island.	95
Figure 3.11 Schematic representation of Tn5398 compared to the Tn5398-like element in ATCC-43255.	96
Figure 3.12 Schematic representation of the putative mobilisable element Tn6115 in QCD-63Q42.	97
Figure 3.13 Schematic representation of the cell surface proteins of <i>S. aureus</i> , <i>C. difficile</i> and <i>E. faecium</i>	98
Figure 4.1 Schematic representation of a conjugative transposon within the genome and circularized after excision.	104
Figure 4.2 Group II intron used for retrohoming.....	109
Figure 4.3 Method of gene disruption by the ClosTron.....	110
Figure 4.4 PCR screening for intron integration.	116
Figure 4.5 Excision of CTn1 in different growth phases.....	123
Figure 4.6 ORFs of CTn2.....	123
Figure 4.7 PCR products to show excision of CTn5 from the chromosome.	124
Figure 4.8 Sequences of transposon-chromosome junctions and direct repeat sequences.....	126
Figure 4.9 Excision of Tn6104.	127
Figure 4.10 ClosTron insertions in CTn1 and CTn7.	129

Figure 4.11 Clostron insertions in CTn2, CTn4 and CTn5.	130
Figure 4.12 Clostron insertion in Tn6103.	131
Figure 4.13 PCR results for the analysis of co-transfer in CD37 transconjugants.....	134
Figure 4.14 Southern blot of CTn1 transconjugants probed with the CTn1 recombinase CD0355.	136
Figure 4.15 Southern blot of CTn2 transconjugants probed with the CTn2 recombinase CD0436.	138
Figure 4.16 Southern blot for CTn4 transconjugants probed with the CTn4 recombinase CD1091.	140
Figure 4.17 Southern blot of CTn5 transconjugants probed with the CTn5 recombinase CD1878.	141
Figure 4.18 Southern blot of CTn7 transconjugants probed with the CTn7 recombinase CD3370.	143
Figure 4.19 Transconjugant analysis of M120 filter-matings	145
Figure 4.20 Nanotube transfer of antimicrobial resistance proteins.	147
Figure 5.1 Example of the effect of <i>C. difficile</i> culture supernatant on Hff-1 in the cytotoxicity assay.	165
Figure 5.2 Cytotoxicity assay.	166
Figure 5.3 PCR analysis for co-transfer of CTn1, CTn2, Tn5397, CTn5 and Tn5398 in the PaLoc transconjugants.	167
Figure 5.4 Circular representation of the regions used for SNP analysis on the 630 genome.	170
Figure 5.5 Fragments of 630 specific sequence in CD37 transconjugants.	172
Figure 6.1 RT-PCR for CD0428 and CD1099.	186
Figure 6.2 PCR for presence of retargeted Clostron plasmids in R20291.	187
Figure 6.3 Attempts to obtain Clostron insertions in Tn6104 and Tn6105.....	188
Figure 6.4 Size distribution of cDNA samples after fragmentation for RNA-seq.....	189
Figure 6.5 Plot for the variation between RNA-seq samples.....	191

List of Tables

Table 2.1 Bacterial strains used in this study.....	60
Table 3.1 Overview of CTns and non-conjugative elements in <i>C. difficile</i>	72
Table 4.1 Oligonucleotides for Tn excision	112
Table 4.2 Oligonucleotides for ClosTron mutagenesis	116
Table 4.3 Primers for the detection of CTns in transconjugant cells.....	118
Table 4.4 Primers for the identification of the transconjugants.	119
Table 4.5 Primers used for the Southern blot probe construction.....	120
Table 4.6 PCR primers for CTn detection.	121
Table 4.7 Transfer frequencies of CTn1.	132
Table 4.8 Summary of PCR analysis on transconjugants of CTn1 filter-matings.	135
Table 4.9 Summary of PCR analysis on transconjugants of CTn4 filter-matings.	139
Table 4.10 Transfer frequencies of CTn5.	141
Table 4.11 Summary of PCR analysis on putative transconjugants.	142
Table 5.1 PCR primers for CTn detection.	162
Table 5.2 Summary of the results of the SNP analysis in the transconjugants.	170
Table 5.3 Amount of 630 specific sequence for each of the Paloc transconjugants....	171
Table 6.1 Primers used for RT-PCR.....	182
Table 6.2 Differential expression of 630 Δ <i>erm</i> CD0428::ErmRAM vs 630 Δ <i>erm</i>	192
Table 6.3 Differential expression of 630 Δ <i>erm</i> CD1099::ErmRAM vs 630 Δ <i>erm</i>	193
Table 6.4 Differential expression of 630 vs 630 Δ <i>erm</i>	194

List of abbreviations

µg	microgram
µl	microlitre
µm	micrometre
°C	degrees Celsius
BHI	brain heart infusion
BHIB	brain heart infusion containing 5% defibrinated horse blood
bp	basepair
cDNA	complementary DNA
CDS	coding sequence
CDT	<i>C. difficile</i> binary toxin
CPE	cytopathic effect
CTn	Conjugative transposon
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide-triphosphate
EBS	exon binding site
g	gram
h	hours
Hfr	high frequency recombination
HGT	horizontal gene transfer
IBS	intron binding site
ICE	integrative and conjugative element
IPTG	Isopropyl-β-D-thio-galactoside
IS-element	insertion sequence element
kb	kilobase
l	litre

LCT	large clostridial toxin
LEO	Left end out
M	molar concentration
mg	milligram
MGE	mobile genetic element
min	minutes
ml	millilitre
mM	milli molar
mRNA	messenger RNA
ng	nanogram
nm	nanometre
OD ₆₀₀	optical density at 600 nm
ORF	open reading frame
<i>oriT</i>	origin of transfer
PaLoc	pathogenicity locus
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RAM	Retrotranspositional-activated marker
RIN	RNA integrity number
REO	Right end out
RNA	ribonucleic acid
RNA-seq	RNA sequencing
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	room temperature
RT-PCR	reverse transcriptase PCR
S-layer	surface layer

sec	seconds
SD	standard deviation
<i>skin</i>	<i>sigK</i> intervening sequence
SNP	single nucleotide polymorphism
TcdA	<i>C. difficile</i> toxin A
TcdB	<i>C. difficile</i> toxin B
TSF	Target site forward
TSR	Target site reverse
UTR	untranslated region
w/v	weight per volume
x g	times gravitational force

Chapter 1 Introduction

1.1. *Clostridium difficile*

1.1.1. Isolation, pathogenesis and treatment

1.1.1.1. The isolation of *Clostridium difficile*

Clostridium difficile was originally isolated from the faeces of healthy human newborns and designated *Bacillus difficilis* by Hall and O'Toole in 1935 [1]. This report, as well as that of Snyder [2] shortly after, characterised the organism as a Gram-positive rod-shaped bacterium (Figure 1.1). Furthermore, the bacterium is obligate anaerobic, forms endospores and produces a toxin that was shown to cause convulsions similar to tetanus toxin produced by *Clostridium tetani* when administered subcutaneously in to laboratory animals [2]. After the organism had been re-designated *C. difficile*, it was noted that the organism usually infects the human intestinal tract [3]. Later, the toxin of *C. difficile* was found in patients suffering from antibiotic associated pseudomembranous colitis [4] and *C. difficile* was recognised as the causative agent for most cases of clindamycin associated pseudomembranous colitis [5,6].

C. difficile strain 630 was isolated from a patient with severe pseudomembranous colitis in a hospital outbreak in Switzerland in 1982 [7]. Strain 630 is extensively studied and was therefore the first strain to be sequenced [8]. Epidemic outbreaks of *C. difficile* have become more frequent over the past decade and these isolates were described as being hypervirulent due to reports of higher toxin production and higher relapse rates in patients [9-11] (section 1.1.3.2). The largest outbreak of hypervirulent *C. difficile* in the UK was in the Stoke Mandeville hospital [12]. The isolate R20291 from

this outbreak is often used as a representative of hypervirulent strains and was shown to be particularly virulent [13]. As a representative strain of the hypervirulent ribotype, R20291 was the second strain to be sequenced [14].

The genome of *C. difficile* 630 consists of a 4.29 Mb chromosome and a 7.8 kb plasmid [8]. The average GC content of the chromosome is low (29%) however, a large number of mobile genetic elements (MGEs) were discovered which were higher in GC content (up to 47%) and comprised approximately 11% of the genome. Further genome sequencing of strains of *C. difficile* showed that the organism has very low genome conservation [15]. Other organisms described with low genome conservation include *Streptococcus pneumoniae* and *Campylobacter jejuni*, which respectively have 46.5% and 59.2% of the genome conserved between strains [16,17]. Based on the genome sequences of 15 strains, it was calculated that *C. difficile* has 23-26% genome conservation and the core genome consists of 947 to 1033 coding sequences (CDS) [15]. The pan genome is described as the collection of all genes present in all strains pooled together, a number that can be calculated from the core and whole genome size [18]. The pan genome of *Streptococcus agalactiae* is predicted to be 1806 CDS [18] and *S. pneumoniae* is 5100 CDS [16] whereas the pan genome of *C. difficile* is predicted to be 9640 CDS [15]. The reason *C. difficile* has the largest predicted pan genome published so far may be due to the large number of MGEs in this organism as these elements contribute greatly to the diversity within the species.



Figure 1.1 Gram staining of *C. difficile* strain 630.

C. difficile is Gram positive and rod shaped. The image was made using an Olympus BX51 microscope equipped with a QImaginig, MicroPublisher 5.0 RTV camera running the Simple PCI software.

1.1.1.2. *C. difficile* disease and treatment

Approximately 5-8% of adults are asymptomatic carriers of *C. difficile* whereas in children and neonates, prevalence ranges from 3-64% [19,20]. The normal colonic microflora is thought to protect from colonisation by *C. difficile*, a phenomenon referred to as colonisation resistance [19,21]. Disruption of the microflora by antibiotic treatment damages this protective barrier and can increase the chance of *C. difficile* colonisation and infection [22]. *C. difficile* can produce three toxins during infection which can all affect the gut epithelia and result in a mild to severe diarrhoea by disrupting tight junctions and the cytoskeleton of epithelial cells. In severe cases, the infection in combination with the influx of leukocytes generated by the immune response can cause pseudomembranous colitis, a condition where lesions of necrosis of the lining of the gut occur (Figure 1.2). This can lead to toxic megacolon, characterised by a dilated colon where the peristaltic movement is paralysed and for which currently the only treatment is colectomy [23,24].



copyright

all rights reserved

Figure 1.2 Effects of *C. difficile* infection.

a) Histology of pseudomembranous colitis in a *C. difficile* patient. The epithelial membrane of the mucosa has detached and is indicated by an arrow. b) Colonoscopy image of the colon of a patient infected with *C. difficile*, white pseudomembrane lesions are visible consisting of damaged epithelia and leukocytes from the immune system. Figure reproduced from Rupnik *et al.* [25]

Although *C. difficile* infection has been reported to occur after treatment with a range of antibiotics, there are a few antibiotics that are more commonly associated with *C. difficile* infection. Clindamycin, cephalosporins and fluoroquinolones have been shown to increase risk of *C. difficile* infection [6,23,25]. Initially, it was advised to prescribe vancomycin for the treatment of *C. difficile*, however, for fear of emerging vancomycin resistance, this was later changed to treatment with metronidazole in the UK and the US [23,26]. Recurrence of infection is a common complication seen in recovering *C. difficile* patients. Reinfection can occur with the strain that initially infected the patient although reinfection with a different strain can also occur [25]. As treatment with antibiotics is not always effective, alternative treatment includes the use of probiotics and faecal transplantation to provide the patient with a healthy microbiota and re-establish colonisation resistance [27,28]. Furthermore, studies are on-going for new treatment strategies such as a vaccine against the *C. difficile* TcdA for which Phase II

trials are currently underway [29]. Currently, prevention of spread is one of the most important strategies to battle *C. difficile* in the hospital setting by isolation of infected patients, disinfection of the patient environment and good hand hygiene for health care workers which has reduced the number of *C. difficile* outbreaks [30,31].

1.1.2. Virulence of *C. difficile*

1.1.2.1. *C. difficile* TcdA and TcdB

The main virulence factors of *C. difficile* are the cytotoxins TcdA and TcdB encoded on the pathogenicity locus (PaLoc) (Figure 1.3). The PaLoc is a genetic locus that is only present in toxigenic strains of *C. difficile* and encodes the toxin genes as well as their regulators, TcdD and TcdC. By comparing the sequence at the ends of the element to the sequence of non-toxigenic strains, it was concluded that the PaLoc has a single integration site in *C. difficile* and is inserted uni-directionally in all strains [32]. Analysis of five non-toxigenic strains identified a 115 bp linker sequence present at this site that is well conserved (> 90%) [32].

TcdA and TcdB [33] are both members of the family of large clostridial toxins (LCTs). This family also includes the lethal and haemorrhagic toxins (TcsL and TcsH) of *Clostridium sordelii*, α toxin (TcnA) of *Clostridium novyi* and TpeL of *Clostridium perfringens*, which are all structurally similar and have 30-70% amino acid sequence identity to TcdB [34,35].

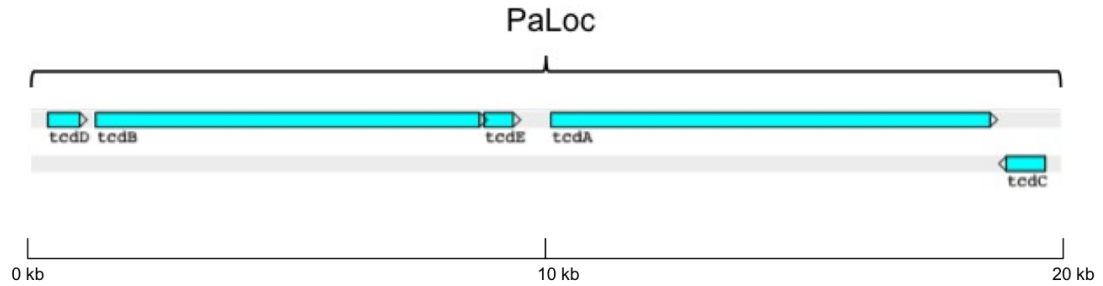


Figure 1.3 Schematic representation of the *C. difficile* PaLoc of strain 630.
ORFs are represented by blue arrows. The complete locus is 19.6 kb long.

Both TcdA and TcdB consist of functionally separate domains; a C-terminal binding domain for recognition of unknown receptors on the epithelia of the colon, a translocation domain for entering the target cell and an N-terminal catalytic domain for glycosylation of the target molecule [25,36] (Figure 1.4). After the toxin binds to a specific cell membrane receptor it is internalised via endocytosis. In the endosome, the cysteine protease within the translocation domain is responsible for proteolytic cleavage of the catalytic domain, which enters the cytoplasm [35]. Both toxins have been shown to inactivate small regulatory proteins of the Rho and Ras superfamilies of GTPases via glycosylation of these regulators [36,37]. This results in actin condensation and disruption of the cytoskeleton and tight junctions of the epithelia, leading to apoptosis [38,39].

© copyright
all rights reserved

Figure 1.4 Schematic of *C. difficile* TcdA and TcdB.

TcdA and TcdB have a similar structure consisting of three functional domains; a catalytic domain, a translocation domain and a binding domain. Figure reproduced from Rupnik *et al.* 2009 [25].

Inflammasome mediated recognition of the toxins by the innate immune system causes production of the cytokine IL-1 β [40,41]. This causes an influx of phagocytic cells such as macrophages resulting in further inflammation of the tissue [25]. The destruction of the epithelia and the inflammatory response cause the formation of the pseudomembranes, loss of the barrier function of the gut and fluid accumulation [25,36].

After the discovery that *C. difficile* produces two separate toxins [42], their relative roles in disease have been extensively studied [34,36,39,43-45] and are still subject to debate. Two recent studies from separate laboratories have independently made knockout mutants in both of the toxin genes and compared these mutants to the wildtype strain in an established hamster model [46]. The study by Lyras *et al.* [44] concluded TcdB is essential for virulence as the mortality rate decreased significantly for the animals challenged with the *tcdB* mutant whereas the mortality rate for animals challenged with the *tcdA* mutant was comparable to that of the wildtype. Kuehne *et al.* [45] used a similar methodology but found no significant differences in mortality rate for either of the mutants compared to the wildtype strain, although the time of death occurred later for the animals colonised by the *tcdB* mutant. A double knockout mutant for *tcdA* and *tcdB* was also used in this study and as expected none of the animals challenged by this mutant died. Both studies used molecular knockout systems based on interrupting the target gene by the insertion of an erythromycin resistance gene. Therefore, both studies used erythromycin sensitive derived isolates of strain 630 to construct their mutants. However, the erythromycin sensitive isolates used in these studies, JIR8094 or 630E [47] and 630 Δ *erm* [48] respectively, were

obtained in separate labs by repeated sub-culturing in the absence of erythromycin. Both isolates were reported to have lost the functional copy of *erm(B)*, which confers erythromycin resistance in 630. It is currently hypothesized that additional mutations have occurred during the generation of either of the erythromycin sensitive strains which affect virulence and have resulted in the opposing results of these two studies [46]. Because of the technical difficulties of the genetic manipulation of *C. difficile* and the relatively large size of the toxin genes, neither of the studies has been able to complement the mutations that were made.

In addition to *tcdA* and *tcdB*, there are three more ORFs present on the PaLoc; *tcdC*, *tcdD* and *tcdE* (Figure 1.3). TcdE is a putative phage-like holin, which was assumed for a long time to play a role in the release of the toxins into the extracellular environment. It was shown recently that TcdE is not essential for secretion of the toxins as no differences could be shown between the secretome of wildtype and *tcdE* mutant isolates and the precise function remains unknown [49].

During growth of *C. difficile* it was shown that *tcdC* is mainly transcribed during the lag and early exponential phase and decreases towards stationary growth. The transcription of the other PaLoc encoded genes is opposite and increases in the late exponential and stationary phase [50]. A model was proposed in which TcdD induces expression of the PaLoc genes, whereas TcdC reduces expression [50]. Transcriptional induction of *tcdA* and *tcdB* by TcdD was demonstrated [51], which was redesignated TxeR, and was subsequently shown to be an alternative sigma factor essential for toxin expression [52]. TcdC was shown to be an anti-sigma factor which negatively regulates the production of toxins and was predicted to interact with TcdD or the TcdD-RNA polymerase holoenzyme [53]. Furthermore, a study by Carter *et al.* [54] showed that

the introduction of a plasmid-encoded functional copy of *tcdC* into naturally occurring *tcdC*-mutant strains reduced toxin production *in vitro* and delayed disease symptoms in an *in vivo* mouse-model. However, these results are not supported by the results of the study by Cartman *et al.* [55], in which a similar *tcdC* mutation was restored by allelic exchange without any effect on toxin production. Although currently the opposing results of these studies cannot be explained, a third study by Bakker *et al.* (in press) has confirmed the lack of any effects on toxin production in a 630 Δ *erm tcdC* mutant (personal communication).

1.1.2.2. *C. difficile* binary toxin

In addition to TcdA and TcdB, some *C. difficile* strains also produce the binary toxin. This is a member of the family of clostridial binary toxins, which have ADP-ribosyltransferase activity, and include the *C. perfringens* iota toxin, *Clostridium spiroforme* toxin and the *Clostridium botulinum* C2 toxin C & D [56,57]. *C. difficile* binary toxin is encoded on the binary toxin locus or CDT locus, which consists of two toxins genes *cdtA* and *cdtB* as well as the positive transcriptional regulator *cdtR* [58], (Figure 1.5). The toxin consists of two peptides, CdtB which binds an unknown surface receptor and translocates CdtA, the catalytic domain, which ADP-ribosylates actin molecules and disrupts the cytoskeleton [56]. Natural occurring PaLoc negative but CDT positive isolates were shown to cause cytopathic effects (CPE) in *in vitro* cell line models and in an *ex vivo* rabbit ileal loop model. These isolates were also shown to colonise hamsters *in vivo* but did not cause disease [59]. CDT causes the formation of microtubule-based protrusions, which promote adherence of *C. difficile* to epithelial cells and may increase colonisation [60]. Although the presence of CDT alone does not cause disease in hamsters, it does affect the cytoskeleton of epithelial cells *in vitro* and

increases colonisation of hamsters. Therefore, the CDT is an interesting subject for further study and could prove to be a target for therapy against *C. difficile* infection.

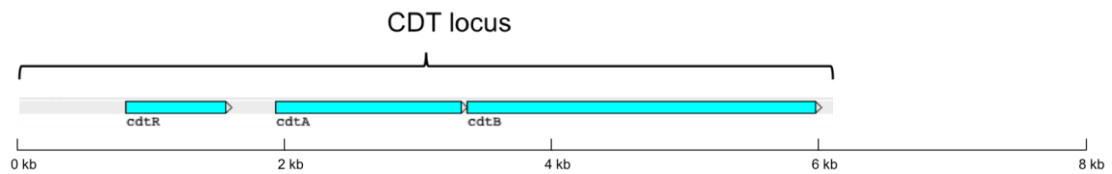


Figure 1.5 Schematic representation of the *C. difficile* CDT locus of strain R20291. ORFs are represented by blue arrows. The complete locus is 6.1 kb long.

1.1.2.3. Other *C. difficile* virulence factors

The *C. difficile* surface layer (S-layer) consists mainly of the S-layer protein Slp, a heterodimer consisting of the high molecular weight and low molecular weight Slps (HMW-SLP and LMW-SLP respectively) [61,62]. These are encoded by *slpA*, which is post-transcriptionally cleaved by cell wall protein 84 (Cwp84) into HMW-SLP and LMW-SLP [63]. The LMW-SLP is variable between strains, and is hypothesised to be an escape mechanism from the host immune system [62,64]. The HMW-SLP binds collagen I, thrombospondin and vitronectin, which are components of the extracellular matrix in the intestinal mucosa [61]. Other adhesins present in the *C. difficile* S-layer are the fibronectin binding protein Fbp68 [65] and Cwp66, which was shown to bind epithelial cells, but for which no precise substrate has been determined [66]. Adhesins are important virulence factors for many pathogenic bacteria to facilitate adhesion of the cells to specific host tissues [67].

A study of the antisera of several *C. difficile*-infected patients showed that both the HMW-SLP and LMW-SLP are recognised by nearly all antisera, whereas a number of

other cell wall proteins such as Cwp84 and Cwp2 are only recognised by a proportion of the tested antisera [68]. Further investigation into the putative adhesins and cell wall proteins of *C. difficile* is necessary as these proteins are expected to play a major role in virulence of the pathogen and they are potential therapeutic targets [67].

1.1.2.4. Sporulation

C. difficile can form endospores which are a metabolically dormant form of the bacterium that is excreted by infected patients. Spores are highly resilient to most commonly used disinfectants and can persist upon hospital surfaces, presumably for years [69]. New hosts are colonised through oral ingestion of these spores. The spores are resistant to gastric acid, which allows the spores to pass through the stomach after which bile salts in the gut signal the spores to germinate into vegetative cells [70]. The use of antibiotics was shown to increase the excretion of spores by infected patients which may account for the success of *C. difficile* as a nosocomial pathogen [31,71]. Although sporulation is considered to be part of the normal life cycle of *C. difficile* and is a feature non-pathogenic strains also exhibit, an increased rate of sporulation has been reported in a number of studies of hypervirulent strains [72,73](section 1.1.3.2). However, a recent paper criticizes the methods of these studies and uses a much larger number of isolates to show the capacity of sporulation is highly variable amongst isolates of specific ribotypes (section 1.1.3.1) of *C. difficile* [74]. Although the study has shown that the rate of sporulation is variable among related samples, it remains unclear if there is any correlation between the rate of sporulation and the severity of disease caused by these isolates.

1.1.2.5. Antibiotic resistance in *C. difficile*

The discovery of antibiotics led to a great advance in treatment of bacterial infections. Unfortunately, poor regulation and limited knowledge of the mechanisms of emerging antibiotic resistance, have quickly led to high occurrence of antimicrobial resistance in clinically relevant species due to over usage [75]. In 1962, early in the study of *C. difficile*, Smith and King reported that the bacterium was resistant to streptomycin, neomycin, colistin or oleandomycin [3]. Later, resistance to p-cresol, D-cycloserine and cefoxitin were described [76,77]. D-cycloserine and cefoxitin are now routinely used as selective supplements for the isolation of *C. difficile* from clinical samples [78]. Clindamycin and cephalosporins were specifically reported to induce pseudomembranous colitis [6,23] although only approximately 15% of *C. difficile* isolates are resistant to clindamycin [79].

A more recent study compared the resistance to 15 antibiotics of 237 isolates of *C. difficile* from a multi-institutional outbreak in Canada with 21 'historic strains' isolated between 1987 and 2001 [79]. This study showed all isolates were resistant to bacitracin and cefotaxime whereas high occurrence of resistance (> 85%) to ceftriaxone, clarithromycin, ciprofloxacin, levofloxacin, gatifloxacin and moxifloxacin was detected only for specific PFGE-type isolates (section 1.1.3.1).

Resistance to erythromycin, tetracycline and chloramphenicol are often encoded on mobile genetic elements (section 1.2). These elements can transfer from resistant isolates to susceptible recipients indicating the clinical importance of these elements [80-82].

1.1.3. *C. difficile* epidemiology

1.1.3.1. Typing methods

The most widely accepted typing methods for the study of *C. difficile* are currently restriction endonuclease analysis (REA), pulsed field gel electrophoresis (PFGE), PCR ribotyping, toxinotyping and multilocus sequence typing (MLST). REA and PFGE type isolates based on the DNA fingerprint after restriction digestion [83,84].

PCR ribotyping is based on the fact that the intergenic spacer region between the 16S and 23S rRNA differs between the multiple pairs of alleles in an individual cell. As most bacteria possess multiple copies of these genes, resulting in products of varying sizes in a PCR reaction using universal primers. Related isolates will have identical patterns of PCR products whereas more distantly related isolates will have variations which allows for grouping of isolates based on these variations [85].

Toxinotyping is a method that amplifies several fragments of the PaLoc in PCR reactions which are restricted with one to four different restriction enzymes in order to detect restriction fragment length polymorphisms of the PaLoc amplicons [86]. For MLST, six to seven housekeeping genes are sequenced to determine the genetic relationship between strains [87].

Recent studies of MLST data and whole genome sequence comparison have discovered that all strains can be divided into six major clades which were shown to have diverged millions of years ago [88]. However, sequence data on specific housekeeping genes and the PaLoc suggest that some recombination has occurred between members of these six clades [89,90]. Isolates from the epidemic outbreaks group together into two specific clades [90,91].

All of the different typing schemes have been developed in order to assign phenotypic differences, such as hypervirulence (section 1.1.3.2), to specific types. Overlap is seen between the results of the typing methods, for instance with the toxinotyping and ribotyping methods [92]. Over the past decade most literature has used ribotyping or a combination of methods to type isolates, therefore, in this thesis only the ribotypes of strains are indicated.

1.1.3.2. Hypervirulence

Since approximately 2000, reports of epidemic outbreaks of *C. difficile* in hospitals in the UK and other western countries increased [9,10]. The isolates from these outbreaks were reported to be more virulent, producing higher levels of toxin, were more often resistant to fluoroquinolones and were responsible for a higher relapse rate in patients [11]. The majority of these isolates were ribotype 027 and were reported to be hypervirulent [93].

As described above, TcdD is an alternative sigma factor that is required for the transcription of *tcdA* and *tcdB*, whereas TcdC is an anti-sigma factor that negatively regulates toxin production (Figure 1.3) [52,53]. It was shown that strains of ribotype 027 contain an 18-bp deletion in *tcdC* which results in a truncated protein being produced that is not capable of regulating toxin production [9]. In 2010 in the UK, ribotype 027 strains were reported to be isolated most frequently from patients [94], however, a report from the Netherlands showed the emergence of ribotype 078, of which the isolates contain a 39-bp deletion in *tcdC* as well as an early stop codon [95]. Recent studies of complementation of the *tcdC* mutation give inconclusive results leaving the role of *tcdC* in toxin regulation unclear [54,55](section 1.1.2.1) and the question remains if the *tcdC* genotype has any effects on hypervirulence.

The emergence of ribotype 002 has been reported in China [96], ribotype 001 in Germany [97] and ribotype 018 in Italy [98]. Hypervirulence in *C. difficile* is therefore no longer considered to be limited to ribotype 027. A study of samples from the past decade from the UK and US revealed that some isolates that were designated ribotype 027 were actually ribotypes 176, 198 and 244 [99]. These ribotypes are highly similar and are predicted to have evolved from ribotype 027 recently, however it is not reported if samples from these ribotypes contain the mutated *tcdC* allele.

Although the *in vitro* production of TcdA and TcdB was shown to be higher in ribotype 027 strains [9,72], concern over the biological relevance of toxin production in *in vitro* batch culture experiments has been raised [100]. Furthermore, a human gut model shows the production of toxins by ribotype 027 strains is not higher per unit of time but has a longer duration of toxin production, which may account for the increased severity of symptoms for patients infected with a ribotype 027 isolate [101].

It was also reported that ribotype 027 strains had a higher level of sporulation [72]. A recent study which included a large number of strains showed that the sporulation rate is very variable between strains within a ribotype, and high and low level sporulation is seen for both epidemic ribotypes and non-epidemic ribotypes, which suggests the amount of sporulation has no effect on severity of disease [74]. Although the epidemiology of *C. difficile* has certainly changed over the past two decades, the basis for the claim that certain ribotypes are hypervirulent is not solid and therefore these are now often referred to as epidemic isolates.

Due to the increase in *C. difficile* infections it became mandatory in the UK to report all cases of infection for epidemiological surveillance [31]. Between 2007 and 2010 the number of *C. difficile* related deaths in England and Wales has decreased from >8000

to <4000 per year (Figure 1.6). In this period, the number of *C. difficile* infections decreased from >55,000 to <22,000 in the UK [102]. The decrease in these numbers is most likely due to the improvements in hospital procedures. Although the prevalence of *C. difficile* is reducing, the occurrence of epidemic strains is still not understood and further study is necessary in order to improve treatment and prevention strategies.

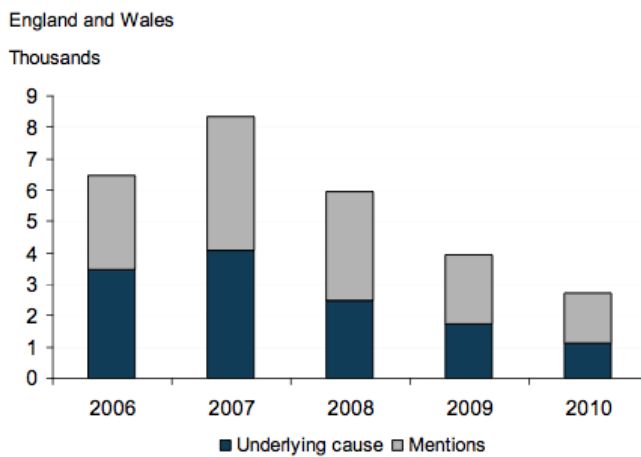


Figure 1.6 Number of deaths attributed to *C. difficile*.

The blue bar shows the number of patients where *C. difficile* is mentioned as the cause of death on the death certificate whereas the grey bar shows the number of patients where *C. difficile* is mentioned as a possible contributing factor. Figure reproduced from the Office for National Statistics website, last accessed 5th September 2012 [103].

1.2. Mobile genetic elements

Horizontal gene transfer (HGT) is the exchange of genetic material between organisms, a mechanism that is extensively used by some prokaryotes. Three molecular mechanisms have been identified for HGT: transformation, transduction and conjugation. Transformation is the uptake of naked DNA from the environment, mostly originating from damaged cells, and incorporation into the genome [104]. Bacteria that have this ability are designated competent cells. *C. difficile* has never been shown to be competent in laboratory settings [105-107] and transformation is therefore not further discussed here. Other forms of horizontal gene transfer are discussed below.

1.2.1. Conjugation

Conjugation is a form of genetic transfer between bacterial cells that occurs via cell to cell contact. The process was first discovered in *Escherichia coli* for the fertility plasmid (F-plasmid) [108]. Tatum and Lederberg [109] had shown that recombination of *E. coli* strains could occur when two auxotrophic mutant strains with specific nutritional requirements were grown in mixed cultures which resulted in prototrophic isolates. It was found that recombination depended on a fertility factor (F-factor) which would have to be present in at least one of the bacteria for a mating pair to be formed efficiently [110,111]. It was recognized that the F-factor was an autonomously replicating genetic element which can convert F- recipient strains into F+ donor strains and encodes the functions needed to promote conjugation [108,112].

The F-plasmid was also shown to integrate into the chromosome through a process of 'high frequency recombination' (Hfr). Transfer of the F-plasmid from Hfr cells leads to transfer of part or the whole of the donor chromosome which can recombine with the recipient chromosome after conjugation [112]. Experiments on the resulting

phenotypes of the recombinant cells showed that linear DNA was transferred unidirectional as specific genes were transferred more regularly and after a shorter time interval than other genes, depending on their distance on the chromosome from the origin of transfer (*oriT*) [113]. Hfr has been described in a number of other Gram negative organisms such as *Yersinia pseudotuberculosis* [114] and *Vibrio cholera* [115]. In the Gram positive organism *Lactococcus lactis*, a lactose fermentation plasmid is present which was shown to turn lactose fermentation negative strains into lactose fermentors [116]. Experiments on transfer from these strains also showed conjugational transfer of chromosomal DNA [117]. Conjugation was shown to be much more efficient in cells with a 'clumping' phenotype which was shown later to be facilitated by CluA, a chromosomally encoded membrane bound protein which is essential for conjugation [118].

CluA is encoded on the *L. lactis* sex factor, together with a number of other genes involved in conjugation. Recombination of these genes into the lactose fermentation plasmid increased the transfer frequency of the plasmid [119], however, the lactose fermentation plasmid was also shown to integrate into the chromosome within the sequence of the sex factor through recombination by the IS-element located within the plasmid [120]. The sex factor was shown to be responsible for the transfer of chromosomal DNA in a unidirectional manner, as previously shown for Hfr in *E. coli*, which was determined by the fact that the transfer frequency of genes located close to the sex factor in the clockwise orientation was higher than that of the genes lying in the anticlockwise direction [121]. Interestingly, the *L. lactis* sex factor is not present in any other lactid acid bacteria [122].

The process of conjugation starts with nicking at a single strand of the plasmid at the origin of transfer and unwinding of the strand by a relaxase. This strand is transferred

across the membranes to the recipient cell by machinery related to type IV secretion systems, where the DNA strand circularizes and the second strand is synthesized [104] (Figure 1.7).

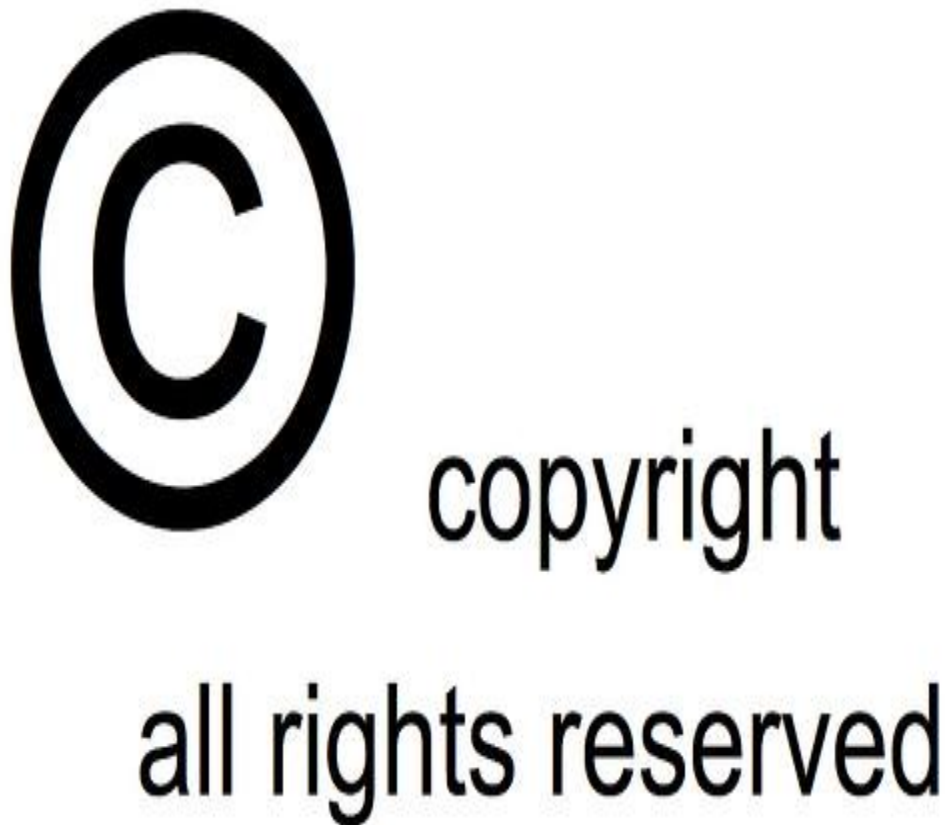


Figure 1.7 Schematic representation of the transfer of a conjugative plasmid.

a) The proteins of the conjugation pilus are encoded on the plasmid, shown in red and yellow. b) Cell to cell contact by the mating pair is made through the pilus, contraction of the pilus brings the cells together. c) The yellow strand of the plasmid is nicked, unwound and transferred to the recipient cell. d) The plasmid DNA circularizes in the recipient bacterium and a second strand is replicated in both cells. The recipient cell has become a potential donor cell of the plasmid. Figure was taken from Juhas *et al.* [123].

Although the conjugation machinery of Gram-negative and Gram-positive bacteria are both related to type IV secretion systems, the main difference is that Gram-negative type IV systems contain a pilus to enable cell to cell contact whereas Gram-positive

systems use an adhesin protein for this function [124]. Type IV secretion systems have three functional substructures; the coupling protein acts as a substrate receptor at the donor cytoplasmic side of the structure, a transport channel conducts the substrate across to the recipient cell and the pilus or adhesin maintains cell to cell contact during transfer [125]. One of the best characterized type IV secretion systems is that of *Agrobacterium tumefaciens* (Figure 1.8).



Figure 1.8 Schematic representation of the type IV secretion system of *A. tumefaciens*.

The type IV secretion system of the Gram negative *A. tumefaciens* consists of 11 VirB proteins VirB1–VirB11 and VirD4. The functional substructures are color coded. Yellow (including white): nucleoside triphosphatases that provide energy for the transfer and act as substrate receptor. Blue: components of the transmembrane channel. Red (including pink and orange): components of the pilus. Green: lytic transglycosylase responsible for the degradation of the murein (peptidoglycan) layer at the site of assembly. Figure was taken from Juhas *et al.* [123].

1.2.1.1. Conjugative transposons

CTNs, also known as integrative and conjugative elements (ICEs), are genetic elements that can transfer into a recipient strain through conjugation and integrate into the host genome through site-specific recombination [126,127] (section 1.2.1.1.3). For

transposition and conjugation, it was shown that CTNs are excised from the genomic target site and form a circular DNA molecule [128]. The circular intermediate can then insert into the previous target site, transpose to a different target site, transfer to a recipient cell via conjugation or can be lost from the host cell [129] (Figure 1.9).

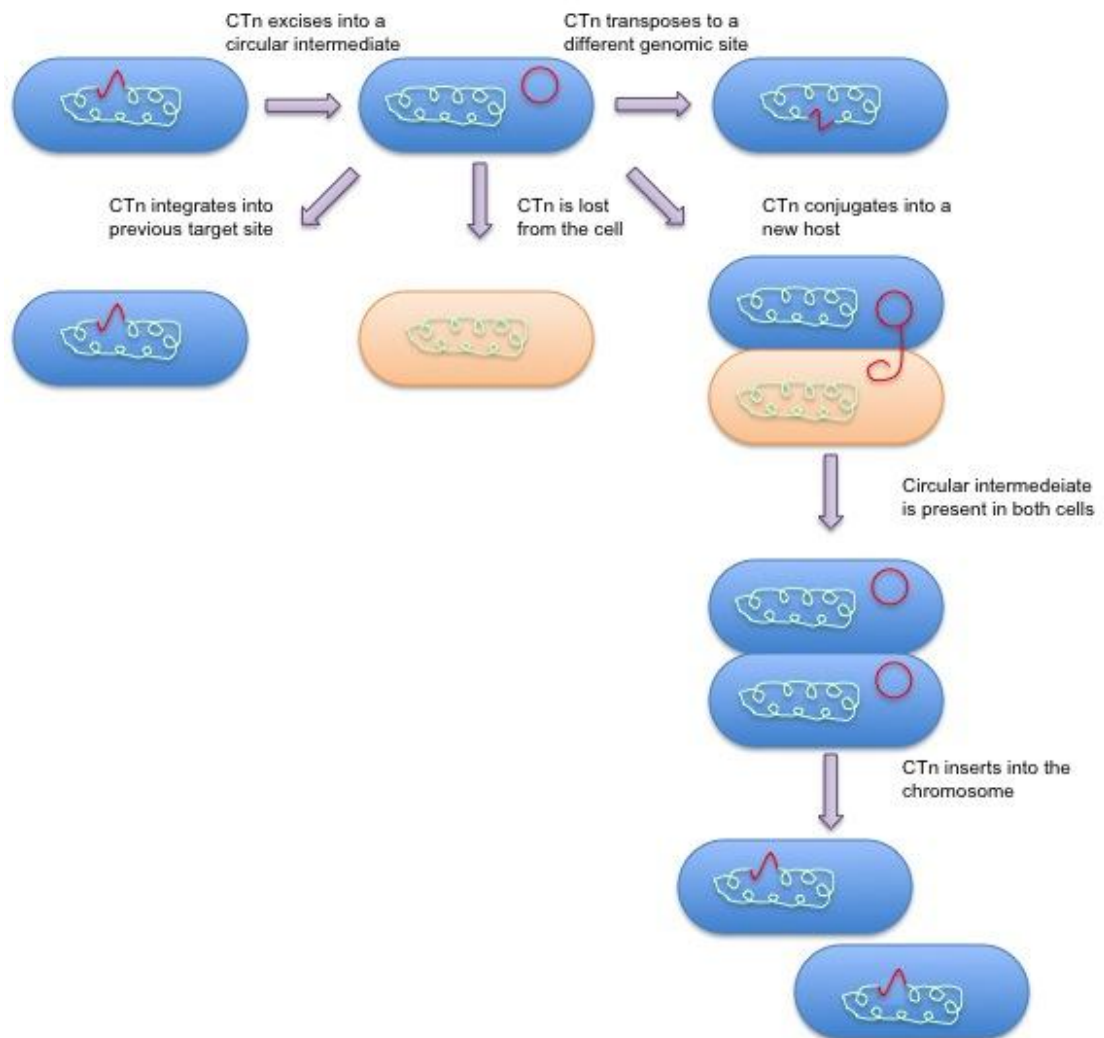


Figure 1.9 Schematic of transposition and conjugative transfer.

The chromosome is represented by a green circular molecule. The linear CTN integrated in the chromosome is represented by a red line. The circular intermediate of the CTN is represented by a red circle. A single strand of the CTN, nicked at the *oriT*, is transferred from this into the recipient cell. Blue cells represent a resistant phenotype, conferred by an antibiotic resistance gene on the CTN and orange cells represent a susceptible phenotype.

Because of their promiscuous nature and ability to integrate into genomes, CTNs have been used as shuttle vectors in order to introduce new DNA into recipients [130,131]. The fitness cost for maintaining CTNs and other mobile genetic elements is variable although generally not very costly to the bacterium [132]. A study into the fitness burden of the vancomycin resistance element Tn1549 indicated that the element in non-induced conditions did not have a high cost whereas constitutive expression of the vancomycin resistance operon resulted in a severe growth defect and modulated colonisation ability [133].

Similar to conjugative plasmids, it was shown that CTNs contain an origin of transfer (*oriT*) [134]. Although replication of CTNs was always considered to occur while integrated in the chromosome, a recent study by Lee *et al.* [135] shows that the circular intermediate of ICEBs1 can replicate starting at the *oriT* through a rolling circle replication mechanism, similar to the replication of plasmids. ICEBs1 is a CTN of the integrative and conjugative element (ICE) family, originally identified in *B. subtilis* [136]. It is unknown if this type of replication also occurs in other CTNs.

1.2.1.1.1. Tn916

Tn916 is the first designated CTN and was originally discovered in 1981 in *Enterococcus faecalis* DS-16 by Franke and Clewell who described it as a transposon that transferred into tetracycline susceptible strains from plasmid free donor strains and was therefore predicted to be self-transferable [137]. Since then many Tn916-like elements have been found in diverse bacterial species, which all have similar backbones [138]. A large number of these elements encode ribosomal protection proteins that confer resistance to tetracycline although other antibiotic resistance determinants are also

common [139]. Most of the research on this family of elements has been carried out on Tn916.

Comparison of Tn916-like elements and other conjugative transposons has revealed that the elements have a modular structure of which the accessory module is most diverse and can encode a number of functions such as resistance to tetracyclines, macrolides, lantibiotics, mercury and other heavy metals [139-141]. CTNs have also been reported to encode operons of genes for the production of and resistance to antimicrobial peptides [142,143], adhesin and virulence associated genes [144], chlorocatechol degradation [145] and metabolism of sucrose, biphenyl and salicylate [146,147]. The accessory module is often surrounded by genes for transcriptional regulation.

Another functional module of the CTNs is the integration module encoding the recombinase and excisionase proteins responsible for transposition of the element in the genome (section 1.2.1.1.3). The final module is the conjugation module which encodes all proteins necessary for conjugation [138] (Figure 1.10).

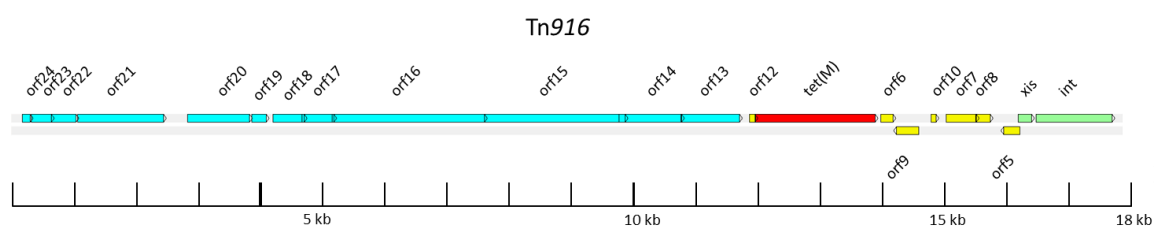


Figure 1.10 Schematic representation of Tn916.

Blue arrows represent the ORFs of the conjugation module, yellow arrows represent the ORFs involved in transcriptional regulation, the red arrow represents *tet(M)*, green arrows represent the ORFs of the transposition module.

1.2.1.1.2. Tn1549

Another family of CTNs are the Tn1549-like elements in which the conjugation module differs from Tn916 (Figure 1.11). Tn1549 was originally isolated in *E. faecalis* 268-10 where it is integrated in the plasmid pAD1. The element confers vancomycin resistance encoded by the *vanB* operon [148]. Transfer of a vancomycin resistance conferring Tn1549-like element from *Clostridium symbiosum* into *Enterococcus faecium* and *E. faecalis* was shown to occur *in vivo* in gnotobiotic mice [149]. This is a disturbing finding as vancomycin is still used as a last resort treatment in recurring cases of *C. difficile* infection.

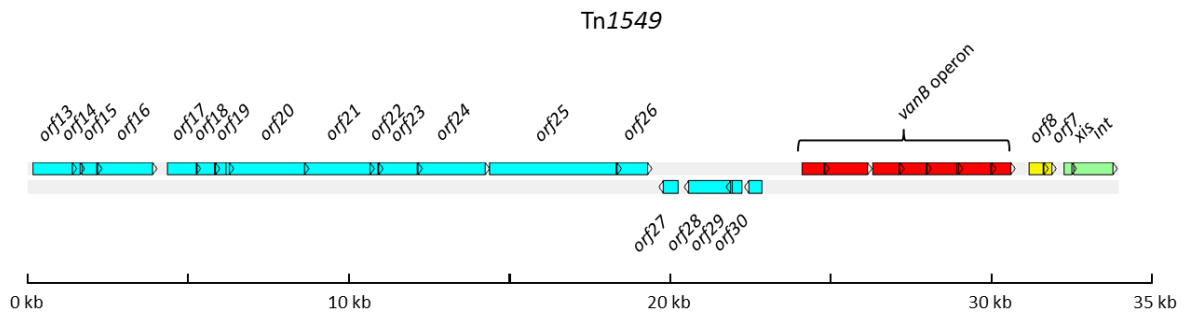


Figure 1.11 Schematic representation of Tn1549.

Blue arrows represent the ORFs of the conjugation module, red arrows represent the ORFs of the accessory module, yellow arrows represent the ORFs involved in transcriptional regulation, green arrows represent the ORFs of the transposition module.

1.2.1.1.3. Excision and integration by site-specific recombination

The excision and integration of CTNs is most often catalysed by site-specific recombinases of two distinct protein families, tyrosine recombinases and serine recombinases, named after the conserved residues that are present in the catalytic domains of these protein families [150]. Integrase are part of the family of tyrosine recombinases and carry out integration of genetic elements whereas the accessory

protein Xis promotes the excision function of the recombinase; this type of mechanism is encoded by Tn916 and Tn1549 as well as many other MGEs [148,151].

The tyrosine recombinase Int encoded on Tn916 excises Tn916 by forming staggered cuts at the coupling sequences at the ends of the element. These can form a heteroduplex at the joint of the circular molecule and form heteroduplexes at the site of integration on either side of the element [127,152,153]. These heteroduplexes are resolved after genomic replication (Figure 1.12). Recently the integration sites of Tn916 were studied in three *C. difficile* strains (ribotypes 009, 012 and 027). The consensus sequence that was determined was consistent with the sequence determined in other hosts [154] and was commonly found (up to 10^5 times) in all three *C. difficile* strains [155]. Although Tn916 was shown to insert in over 100 independent target sites in two of the strains, a preference appeared for intergenic regions whereas the third strain had a single highly preferred target site [155,156]. This has led to the hypothesis that the genomic sequence plays a role in target site selection but other factors are involved which have not yet been identified.

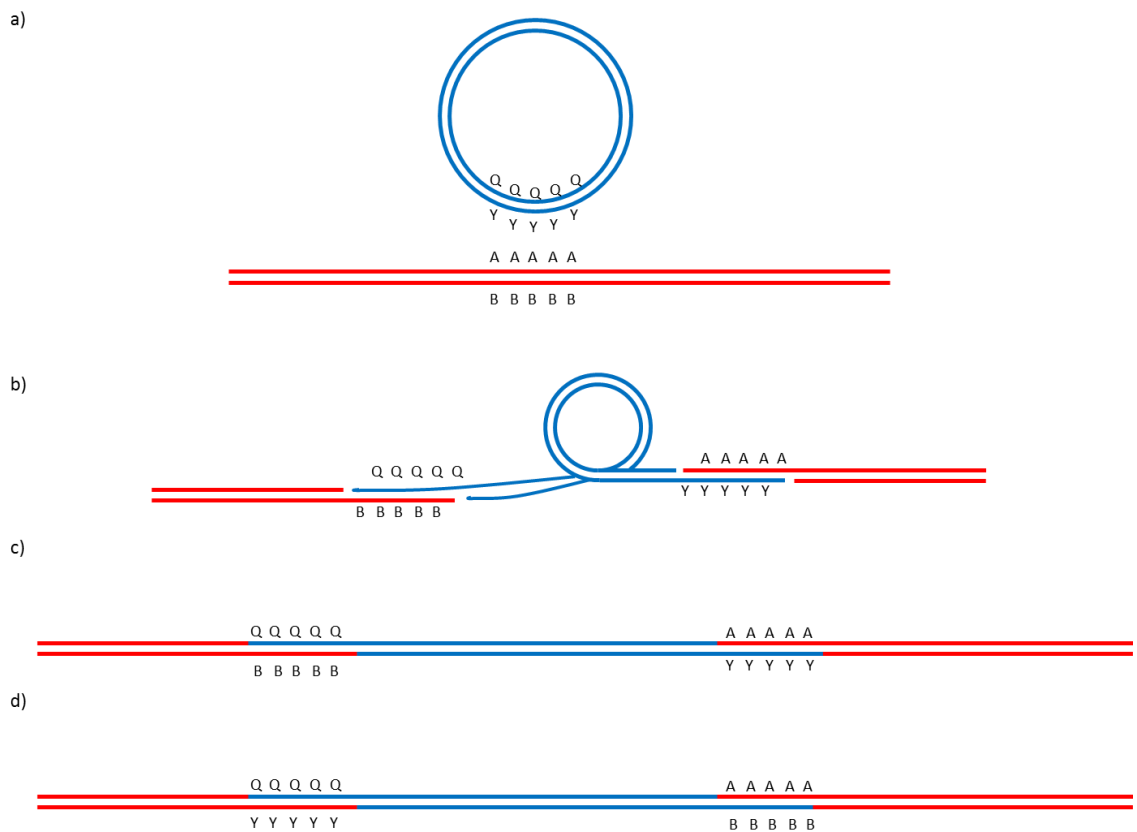


Figure 1.12 Schematic representation of the insertion of Tn916.

Tn916 is represented by the blue lines; the chromosomal DNA is represented by the red lines. The coupling sequences of the element and in the target site are represented by the primer pairs A-B and Q-Y. a) Tn916 is in the circular intermediate form and a suitable empty target site is present in the chromosome. b) Staggered cleavage of Tn916 and the chromosome by the tyrosine recombinase Int. c) Tn916 is ligated with the chromosomal DNA generating heteroduplexes. d) Replication of the chromosome containing Tn916 resolves the heteroduplexes. The figure was recreated based on the figure by Scott [157].

The insertion of CTNs and the excision of the element into a circular transfer intermediate can also be facilitated by serine recombinases. Tn5397, the tetracycline resistance conferring CTn which was originally discovered in *C. difficile* strain 630 (1.2.1.2.1), encodes the serine recombinase TndX [156]. Tn5397 is flanked on both sides by a GA dinucleotide when it is integrated in the genome (Figure 1.13). TndX can perform staggered cleavage at these sites leaving a dinucleotide overhang which is joined after strand exchange to form the circular molecule of Tn5397 and the original regenerated target site [158].

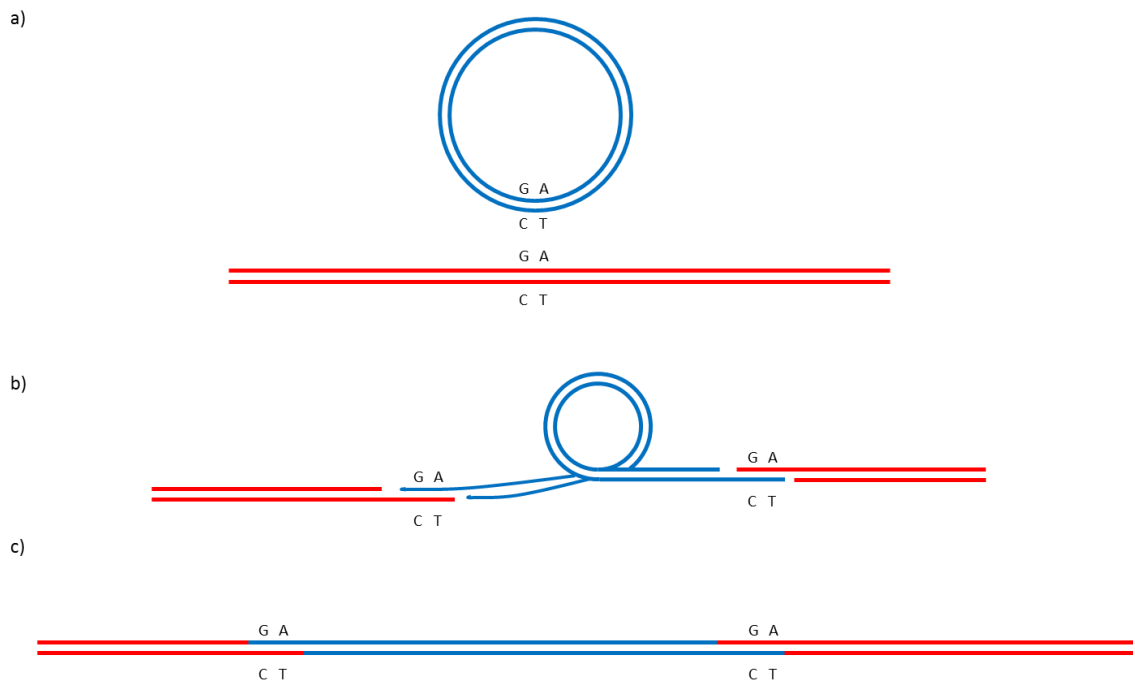


Figure 1.13 Schematic representation of the insertion of Tn5397.

The blue lines represent Tn5397 and the red lines represent the chromosomal DNA. The coupling sequences of the element and within the target site are represented by the nucleotide pairs G-C and A-T. a) Tn5397 is in the circular intermediate form and a suitable empty target site is present in the chromosome. b) Staggered cleavage of Tn5397 and the chromosome by the serine recombinase TndX. c) Tn5397 is ligated with the chromosomal DNA generating homoduplexes.

1.2.1.2. MGEs in *C. difficile*

1.2.1.2.1. Tn5397

The studies of Ionesco [159] and Smith *et al.* [106] were the first to demonstrate the transfer of tetracycline resistance between *C. difficile* strains. The groups could not isolate a plasmid responsible for the transfer of the resistance determinant although the transfer was shown to be cell to cell contact-dependant. As the first report of Tn916 had just been published [137], it was hypothesised that the resistance determinant could be encoded on a Tn916-like element. Transfer of the element from strain 630 at a low frequency (5×10^{-7} transconjugants per donor) was reported [160]

as well as homology between the tetracycline resistance element and Tn916 via DNA hybridisation [105]. Finally, the element was shown to be a conjugative transposon and designated Tn5397 [81].

Although the structure of Tn5397 and Tn916 are very similar, several differences were found. Tn916 encodes a tyrosine recombinase and an excisionase protein, which together are responsible for the excision of the element from the genome into a circular molecule and integration into the genome [161]. Tn5397 contains a single serine recombinase that is responsible for these functions [156]. Tn5397 also contains a group II intron integrated within *orf14* [81]. This class of mobile elements can transpose into specific genomic sites and are spliced from mRNA transcripts by the formation of a ribonucleo-protein by the mRNA and the protein encoded on the intron. *orf14* is located within the conjugation module and it is well conserved among Tn916-like elements and encodes a NlpC/P60 family protein [138]. The group II intron was shown to splice from the *orf14* transcript, however preventing this event by deletion of part of the intronic ORF did not have an effect on the conjugative transfer of Tn5397 [162]. It is hypothesised that either the function of Orf14 is substituted by host factors in these matings or that the non-spliced partial intron had no effect on the function of the protein as the element is integrated at the 3' end of *orf14* (Figure 1.14).

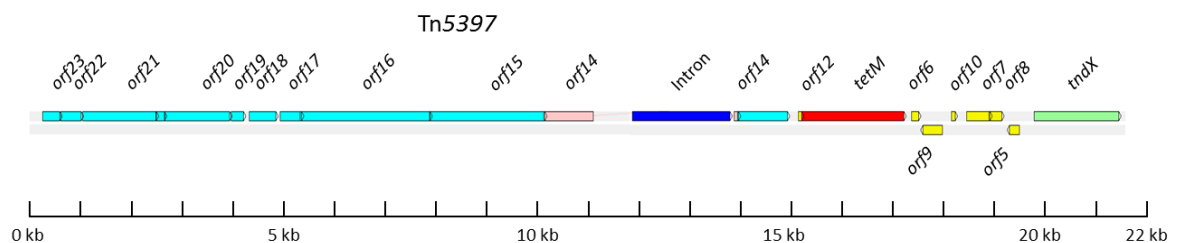


Figure 1.14 Schematic representation of Tn5397.

Blue arrows represent the ORFs of the conjugation module, the pink arrow represents the interrupted *orf14*, the dark blue arrow represents the intron encoded protein, the red arrow represents *tet(M)*,

yellow arrows represent the ORFs involved in transcriptional regulation, the green arrow represents the serine recombinase,

Tn5397 was shown to transfer by filter mating into *E. faecalis* [163], *S. aureus* [105] and *B. subtilis* [164]. *B. subtilis*::Tn5397 was subsequently used as donor of the element to show transfer of Tn5397 to a streptococcal recipient in a model for oral biofilms [165].

1.2.1.2.2. CTNs in *C. difficile* 630

Analysis of the genomic sequence of *C. difficile* strain 630 revealed that another six putative CTNs (in addition to Tn5397) are present in this strain [8]. Based on their location in the genome, the elements are named CTn1, CTn2, CTn4, CTn5, CTn6 and CTn7 and Tn5397 was named CTn3 although, mostly, the original designation is used in the literature. CTn1, CTn2, CTn4, CTn5, CTn6 and CTn7 were undetected for a long time due to the lack of any identifiable phenotypes encoded on these elements.

CTn1, CTn6 and CTn7 are related to Tn916, similarly to Tn5397 (Figure 1.15). CTn1 and CTn7 both encode predicted ABC transporters on their respective accessory modules although no substrates for these have been predicted. Furthermore, both elements encode a surface protein with homology to collagen binding proteins, further described in Chapter 3. CTn1 and CTn6 both encode a tyrosine recombinase and excisionases which are predicted to perform excision and insertion of their respective elements, whereas CTn7 encodes a serine recombinase. The accessory module of CTn6 encodes a number of hypothetical proteins for which no functions have been predicted.



copyright

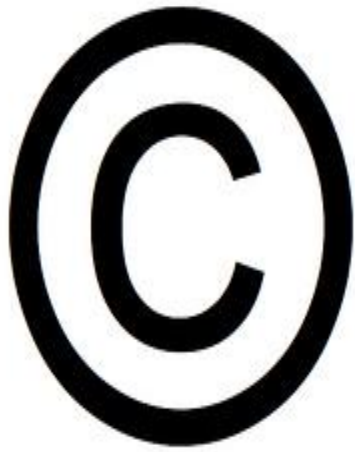
all rights reserved

Figure 1.15 Comparison of Tn916 to Tn5397, CTn1, CTn7 and CTn6.

The CDS on the elements are represented by coloured arrows. CDS are coloured by their function; pink: recombination, purple: accessory function, green: conjugation, light blue: transcriptional regulation. The dark blue region in Tn5397 is the group II intron. SR is serine recombinase, TR is tyrosine recombinase. Red lines between the elements represent protein similarity. Figure is taken from Sebahia *et al.* [8].

CTn2, CTn4 and CTn5 are all part of the Tn1549 family (Figure 1.16). CTn2 and CTn5 both encode serine recombinases which are predicted to insert and excise the respective elements from the chromosome, CTn4 encodes a tyrosine recombinase and excisionase for these functions. The accessory modules of CTn2, CTn4 and CTn5 all encode ABC transporters for which no substrate has been predicted.

Five of these six putative elements are predicted to encode ABC transporters. The substrates of these transporters cannot be predicted however, the transporter of CTn7 has homology to magnesium ABC transporters [8]. Currently, it is also unknown if these ABC transporters function in influx or efflux of specific substrates.



copyright

all rights reserved

Figure 1.16 Comparison of Tn1549 to CTn4, CTn5 and CTn2.

The CDS on the elements are represented by coloured arrows. CDS are coloured by their function; pink: recombination, purple: accessory function, green: conjugation, light blue: transcriptional regulation. SR is serine recombinase, TR is tyrosine recombinase. Red lines between the elements represent protein similarity. Figure is taken from Sebaihia *et al.*[8].

1.2.1.3. Mobilisable elements in *C. difficile*

1.2.1.3.1. Tn4453

Mobilisable transposons are elements that can be mobilised by Mob proteins to transfer via conjugation using the transfer machinery of another conjugative genetic element [166,167]. In *C. perfringens* chloramphenicol resistance is conferred by a chloramphenicol acetyl-transferase (CatP) protein encoded on the mobilisable transposon Tn4451 [166]. Analysis of *catP* in *C. difficile* strains found that this gene is located on a mobilisable transposon with an identical structure to Tn4451 designated Tn4453 [82] that has 89% sequence identity with Tn4451 [168,169] (Figure 1.17). The

element encodes *tnpX*, a serine recombinase that is responsible for excision of the element into a circular intermediate for transposition and conjugative transfer. Other genes encoded on the element are *tnpZ* which is involved in mobilisation of the element and *tnpV*, *tnpY* and *tnpW* for which the functions are unknown [169].

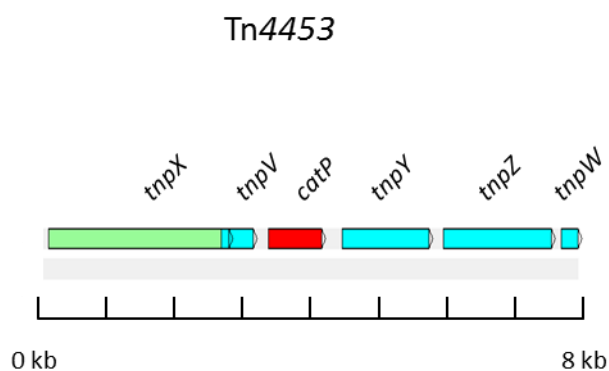


Figure 1.17 Schematic representation of Tn4453.

The green arrow represents *tnpX*, the red arrow represents *catP* other ORFs are represented by blue arrows.

1.2.1.3.2. Tn5398

In addition to the transfer of tetracycline resistance it was noted that erythromycin resistance could also be transferred from *C. difficile* 630 into *C. difficile* CD37 [160]. Subsequently the transfer of the resistance determinant was also shown from *C. difficile* 630 into *Staphylococcus aureus* [170] and *B. subtilis* [80]. Based on the fact that the element can enter the chromosome at specific sites in *C. difficile* but multiple sites in *B. subtilis* it is suggested that the element could be a conjugative transposon [80].

Subsequently it was shown that Tn5398 contains two rRNA adenine N-6-methyltransferase (*erm(B)*) genes [171] (Figure 1.18). *Erm(B)* confers erythromycin resistance through ribosomal protection but only *erm(B)2* on Tn5398 is functional [48]. Full characterisation of the element showed that Tn5398 is in fact not a conjugative or mobilisable transposon as it does not contain any predicted integration or excision

genes, and also no known mobilisation or conjugation related genes are present [172]. Apart from the two *erm(B)* genes, the element contains homologues to Tn916 *orf13*, *orf9* and an incomplete copy of *orf7*, which are predicted to function in regulation of the element. The protein encoded by *orf298* has similarity to the plasmid partitioning proteins ParA and Soj although there is no experimental evidence for this function [172]. Further CDS present on the element are *effR* and *effD*, which are predicted to function as an efflux protein and the associated regulator and two copies of *orf3* of which the function is unknown [169,172].

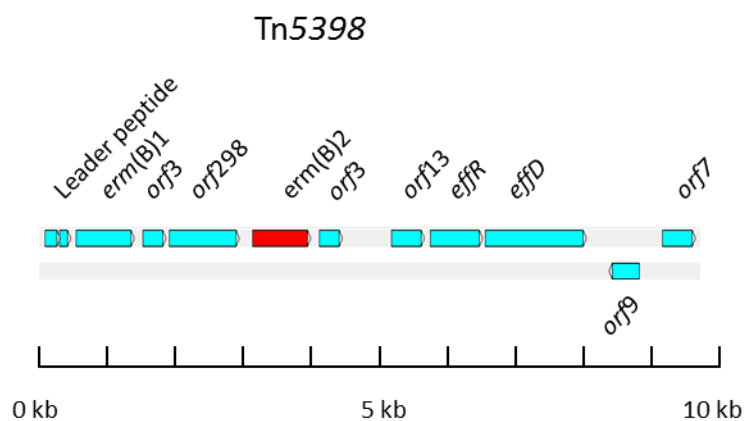


Figure 1.18 Schematic representation of Tn5398.

The red arrow represents *erm(B)2*, other ORFs are represented by blue arrows.

1.2.2. Transduction

Prophages are genetic elements that can replicate with the genome of the host cells during their lysogenic life cycle and can often integrate within the host chromosome. Upon induction, the phage can go into a lytic life cycle where it replicates in the host cell and its DNA gets packaged into phage particles. Upon lysis of the host cell these particles are released into the environment and can infect new host cells [104].

Transduction is a process of horizontal gene transfer in which host DNA is packaged into a phage head particle and subsequently inserted into the DNA of a recipient cell. Prophages are furthermore known to carry genes encoding virulence factors, for instance, nontoxigenic *C. botulinum* can be converted into toxin producing strains after transduction with toxin encoding prophages [173].

C. difficile strain 630 has been shown to carry two temperate prophages designated ϕ C630-1 and ϕ C630-2 [8] which are both *Myoviridae* phages [174,175]. Homology between *tcdE*, *tcdA* and *tcdB* and phage sequences has led to the hypothesis that the PaLoc was present on a phage that integrated within the genome of an ancestral *C. difficile* strain and conferred pathogenicity to the host cell [32,174,176]. Several studies have therefore focussed on the effect on virulence of the host cell upon integration of specific phages. Goh *et al.* [174] have shown that ϕ C2, ϕ C8 (*Myoviridae*) and ϕ C6 (*Siphoviridae*) all have different effects on TcdA and TcdB production by the host strain. Furthermore, the effects on the levels of *tcdA* and *tcdB* transcription were not comparable with the effects on toxin production, which is hypothesized to be caused by the acquisition of novel phage holin genes and effects on the host transcriptome by regulation systems present on the prophages. Govind *et al.* [176] found that infection with ϕ CD119 (*Myoviridae*) caused a significant reduction in both transcription and production of *tcdA* and *tcdB*. The Xre transcriptional regulator RepR was shown to cause this effect by binding the transcriptional regulator *tcdR* on the PaLoc. In contrast, Sekulovic *et al.* [177] have recently shown that ϕ CD38-2 (*Siphoviridae*) can induce the transcription and production of *tcdA* and *tcdB* in certain strains but will not do so in others. These studies all describe similar experiments with

very different outcomes, which may be attributed to the fact that all were using different phages from two distinct families.

1.2.3. Additional *C. difficile* MGEs

1.2.3.1. *skin*^{Cd}

In both *B. subtilis* and *C. difficile*, the RNA polymerase sigma factor encoding gene, *sigK* is interrupted by a prophage-like element [178,179]. The elements were designated *sigK* intervening sequence or *skin*^{Bs} and *skin*^{Cd}. Both elements were shown to excise into a circular molecule by site-specific recombination during late sporulation allowing transcription of *sigK*. In *B. subtilis* absence of the *skin*^{Bs} has no effect on sporulation [180] although overexpression of *skin*^{Bs} genes by repression of the *skin* repressor *sknR* resulted in a growth defect [181]. Absence of *skin*^{Cd} in *C. difficile* results in a growth defect and was only reported for the non-toxigenic strain CD37 and the toxigenic strain ATCC9689 [179]. All other *C. difficile* strains that have currently been sequenced contain a copy of *skin*^{Cd} [179,182]. The element in strain 630 encodes a serine recombinase hypothesized to be responsible for the excision of the element, a putative lipoprotein, a putative cell surface protein, two putative phage transcriptional regulators, seven phage proteins of unknown function and four hypothetical proteins (Figure 1.19). Furthermore, a *vanZ* teicoplanin resistance protein is present on the element but as noted by Sebaihia *et al.* [8], *C. difficile* strain 630 is sensitive to this antibiotic.

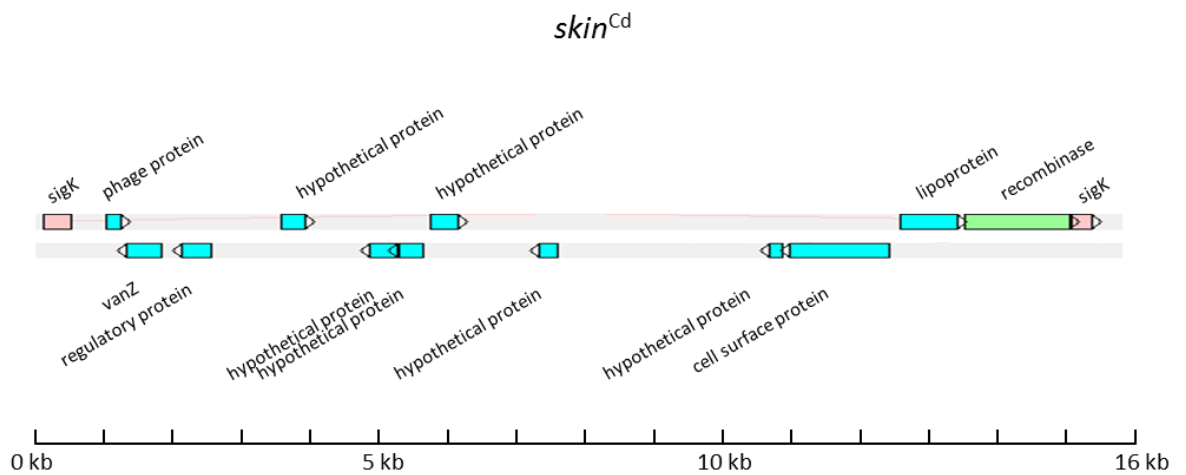


Figure 1.19 Schematic representation of *skin^{Cd}* in strain 630.

The green arrow represents the serine recombinase, other ORFs are represented by blue arrows. The interrupted *sigK* is represented by an interrupted pink arrow.

1.2.3.2. IStrons

Insertion sequences elements (IS-elements) are small genetic elements encoding generally one transposase that allows transposition of the element into different target sites [183]. Group I Introns on the other hand are genetic elements which encode ribozymes that catalyse their own splicing from mRNA, allowing exons to ligate to form mature mRNA molecules [104]. In *tcdA* in *C. difficile* C34, a chimeric genetic element was found with features of both of these elements that was designated an IStron; *CdIst1* [184,185]. *CdIst1* was shown to splice from the pre-mRNA and expression of TcdA was normal. The element consists of two CDS which encode a putative transposase and a ribozyme. Southern blot and PCR analysis indicated that *CdIst1* is commonly present in *C. difficile* strains [184], which was confirmed by the genome sequencing of strain 630 [8]. Variants of *CdIst1* are present due to (partial) deletion of the transposase genes and the length of the elements varies from

approximately 0.4-2.2 kb [178]. Alternative splice products detected in cells that contain *CdIst1* indicates that the element does not splice 100% accurately. This could mean that the element promotes an elevated mutation rate which could benefit *C. difficile* in certain circumstances although it could also decrease fitness [185]. A homolog with 63% sequence identity to *CdIst1* has been reported in *Bacillus cereus* after whole genome sequence analysis [186], however, IStrons have not been reported in any other organisms.

1.2.4. Transfer of DNA via nanotubes

The presence of a tubular network between *B. subtilis* cells was recently reported and named intercellular nanotubes [187]. Intercellular transfer of fluorescent proteins, antibiotic resistance proteins and non-conjugative plasmids was shown to occur through these tubes which were visualised using scanning electron microscopy. Based on their appearance and the fact they are sensitive to low concentrations of the detergent SDS, the nanotubes are proposed to consist of membrane components.

In contrast to conjugation, the transfer through intercellular nanotubes is bidirectional enabling proteins from either bacteria to cross the tubules to the intracellular milieu of the other cell. Furthermore, this type of transport is unique in that both hereditary and non-hereditary traits are transferred. It is suggested that metabolites and RNA may also be shared through the nanotubes, however, this was not tested for in this study.

Scanning electron microscopy of mixed cultures containing *B. subtilis*, *E. coli* and *S. aureus* cells showed the formation of interspecies nanotubes. It is hypothesized that

such multispecies communities can allow for the formation of new phenotypes in order for the communities to adept in changing environments.

A critical review article addressed concerns regarding the controls used in the study such as the potential of extracellular calcein carried over on the celmembranes in the fluorescence experiments [188]. Furthermore, it is stated that the nanotubes were only visualised for bacteria grown on solid growth media, raising the question if nanotubes can also be formed during other growth conditions.

As *C. difficile* is related to *B. subtilis*, there is a possibility that this organism is capable of forming intercellular nanotubes. Furthermore, transfer of the conjugative transposons described above could take place via this transfer system rather than conjugation. It is therefore important that the presence of a nanotube system in *C. difficile* is investigated to identify if such a system could be responsible for the previously reported transfer of conjugative elements.

1.3. Aims of this work

This work focuses on the study of the MGEs of *C. difficile* and the role these elements may play in virulence and other cellular processes. Chapter 3 describes a bioinformatics study of recently sequenced genomes of *C. difficile*. The aim of this chapter was to identify novel CTNs and to determine if the previously determined elements in strain 630 are present in other strains. Furthermore, the accessory modules of the novel elements, and those described in previous studies, were studied by bioinformatics in order to determine the likely function of the predicted proteins and predict the effect these elements may have on antibiotic resistance and virulence.

The aim of Chapter 4 was to determine if the CTNs of *C. difficile* are mobile. Experiments were set up to test if the CTNs in strains 630, R20291, QCD-32G25, QCD-23M63, QCD-63Q42 and M120 could excise from the genome. Subsequently, it was tested if the elements from strains 630, R20291 and M120 could transfer by performing filter-matings. The elements that did not contain a known selection marker were targeted by the ClosTron to introduce a selection marker. Transconjugants of the matings between *C. difficile* strains were studied to determine the number of integration sites of the CTNs that were transferred, using Southern blots. To determine if the CTNs of *C. difficile* only transfer under selection pressure, the transconjugants were examined for co-transfer of elements that do not contain selection markers.

A single CD37 transconjugant was discovered that contained the 630 PaLoc. The aim of Chapter 5 was to find if transfer of the PaLoc was reproducible and to analyse the transconjugants. Filter-matings were set up using a donor that contains a ClosTron insertion in the PaLoc to select for transfer of this genomic region. Whole genome

sequencing of the transconjugants was performed to determine the exact genomic regions that were transferred.

In Chapter 6, it was aimed to determine if the transcriptional regulators on CTns have effects on the transcriptome of the cell. Furthermore, it is determined if there are differences in the transcriptomes of strains 630 and 630 Δ *erm*. RNA-seq was used to compare the effects on the transcriptome of the mutations made in the transcriptional regulators encoded on CTn2 and CTn4 but also the effects of the mutations that have occurred between strains 630 and 630 Δ *erm*.

Chapter 2 Materials and methods

2.1. Bacterial strains and culture methods

Bacterial strains used in this study are listed in Table 2.1. *C. difficile* and *B. subtilis* strains were grown on Brain Heart Infusion (BHI) agar plates (Oxoid Ltd, Basingstoke, UK) supplemented with 5% defibrinated horse blood (BHIB) (E & O laboratories, Bonnybridge, UK) or in BHI broth. Broth and agar plates were pre-reduced before culturing *C. difficile*. For the cytotoxicity assay described in Chapter 5, *C. difficile* was grown in TY broth consisting of 3% Bacto Tryptose (BD Biosciences, Oxford, UK), 2% Yeast extract (Oxoid Ltd) and 0.1% Thioglycolate (Sigma-Aldrich Company Ltd, UK) at pH 7.4. All *C. difficile* strains were grown at 37 °C in anaerobic conditions (80% N₂, 10% H₂ and 10% CO₂).

E. coli was grown on LB agar plates or in LB broth (Sigma-Aldrich Company Ltd, Gillingham, UK). *E. coli* and *B. subtilis* strains were grown in aerobic conditions unless co-cultured with *C. difficile*.

All antibiotics were obtained from Sigma-Aldrich Company Ltd (Gillingham, UK).

Bacterium	Ribotype	Properties	Reference or source
<i>C. difficile</i> :			
630	012	Tc ^R Erm ^R Rif ^S	[7,160]
630 Δ erm	012	Tc ^R Erm ^S Rif ^S	[48]
R20291	027	Tc ^S Erm ^S Rif ^S	[13]
CD37	009	Tc ^S Erm ^S Rif ^R	[106]
		Recipient in filter-mating experiments	
CD843			Dr. S. Goh
QCD-66C26	027	Tc ^S Erm ^R	Dr. A. Dascal
QCD-32G58	027	Tc ^S Erm ^R	Dr. A. Dascal
QCD-23M63	078	Tc ^S Erm ^S	Dr. A. Dascal
M120	078	Tc ^R Erm ^S	Dr. J. Corver
CD37 pMTL960/GusA		Thi ^R	Dr. H. Hussain
R20291 pMTL9301		Linc ^R	Dr. H. Hussain
ClosTron mutants:			
630 Δ erm CD0364::ErmRAM	012	Tc ^R Erm ^R Linc ^R	Chapter 4
630 Δ erm CD0386::ErmRAM	012	Tc ^R Erm ^R Linc ^R	Chapter 4
630 Δ erm CD0428::ErmRAM	012	Tc ^R Erm ^R Linc ^R	Chapter 4
630 Δ erm CD1099::ErmRAM	012	Tc ^R Erm ^R Linc ^R	Chapter 4
630 Δ erm CD1873::ErmRAM	012	Tc ^R Erm ^R Linc ^R	Chapter 4
630 Δ erm CD1873::ErmRAM	012	Tc ^R Erm ^R Linc ^R	Chapter 4
R20291 1803::ErmRAM	027	Erm ^R Linc ^R	Chapter 4
630 Δ erm tcdB::ErmRAM	012	Tc ^R Erm ^R Linc ^R	[45]
Transconjugants:			
CD37::CTn1 CD0364::ErmRAM		4 transconjugants	Chapter 4
CD37::CTn1 CD0386::ErmRAM		9 transconjugants	Chapter 4
CD37::CTn2 CD0428::ErmRAM		10 transconjugants	Chapter 4
CD37::CTn4 CD1099::ErmRAM		6 transconjugants	Chapter 4
CD37::CTn5 CD1873::ErmRAM		10 transconjugants	Chapter 4
CD37::CTn7 CD3392::ErmRAM		6 transconjugants	Chapter 4
CD37 PaLoc tcdB::ErmRAM		7 transconjugants	Chapter 5
<i>B. subtilis</i> ::CTn5 CD1873::ErmRAM		10 transconjugants	Chapter 4
Other species:			
<i>B. subtilis</i> CU2189		Recipient in filter-mating experiments	[189]
<i>E. coli</i> CA434		Competent cells for transformations and conjugation into <i>C. difficile</i> .	[190]

Table 2.1 Bacterial strains used in this study

Tc = tetracycline, Erm = erythromycin, Linc = lincomycin, Thi = thiamphenicol, ^R = resistant, ^S = sensitive

2.2. Molecular techniques

2.2.1. DNA isolation

DNA was isolated using the Genra Puregene Yeast/Bact. Kit (Qiagen) according to the manufacturer's instructions; however, changes were made to the amount of the solutions that were used. Ten millilitres of overnight culture (16 h) was centrifuged for 5 min at 4,500 x g, and the cell pellet was resuspended in 1 ml PBS. The cell suspension was transferred to a microcentrifuge tube and centrifuged for 1 min at 16,000 x g. The cell pellet was resuspended in 600 µl cell suspension solution. Five microlitres of lytic enzyme solution was added and the tubes were inverted several times for mixing before they were incubated at 37 °C for 30 min. Tubes were centrifuged at 16,000 x g for 1 min and the supernatant was removed with a pipette. Cells were resuspended in 600 µl cell lysis solution and incubated at 80°C for 5 min. Next 5 µl of RNase A solution was added and the tubes were inverted 10 times to mix and incubated at 37 °C for 30 min. Samples were transferred to ice and 200 µl protein precipitation solution was added and vortexed for 20 sec after which the samples were incubated on ice for 30 min. The tubes were centrifuged for 3 min at 16,000 x g and the supernatant was poured into a microcentrifuge tube containing 600 µl isopropanol. The tubes were inverted approximately 50 times and centrifuged for 1 min at 16,000 x g. The supernatant was discarded and the tubes were drained on tissue paper. The DNA pellet was washed by adding 600 µl 70% ethanol and inverting several times. The tubes were centrifuged for 1 min at 16.000 x g and the supernatant was discarded and the tubes drained on tissue paper. The samples were air dried for 15 minutes and 100 µl sterile distilled H₂O was added. The DNA was incubated for 1 h at 65 °C and overnight

at RT. Samples were analysed using a Nanodrop 1000 for quality and quantity and stored at -20 °C.

2.2.2. PCR

Standard PCR amplification was carried out using the NEB Taq polymerase kit (New England Biolabs) according to the manufacturer's instructions. Reactions contained 10 mM dNTPs (NEB) and 0.4 µM primers (Sigma Aldrich, UK). PCR programs consisted of 4 min initial denaturation at 94 °C, followed by 25 to 35 cycles of 30 seconds denaturation at 94 °C, 1 min annealing at primer specific temperatures, extension at 72 °C for 1 min per 1-kb of expected product. All reactions finish with 10 min of final extension at 72 °C and then the sample was cooled to 4 °C. Annealing temperature is typically 5 °C below the melting temperature of the primer with the lowest melting temperature. Primers are listed in the materials and methods of each chapter.

PCR products were run on a 1% agarose gel at 80-120 mV for 30-60 min. Agarose gels contained Gelred (Biotium, Hayward, USA) at a one 10,000th (V/V) concentration. Gel imaging was carried out using an Alphaimager (AlphaInnotech, San Leandro, USA) using the Alphaview software package (version 1.0.1.10).

2.2.3. Purification of PCR products

PCR reactions for which only a single product was visible on an agarose gel were purified using the Spin column purification kit (NBS Biological Ltd, Cambridgeshire, UK) according to the manufacturer's instructions.

When multiple PCR products were present, the required product was agarose gel purified. The DNA was purified from the gel using a Spin column gel purification kit (NBS Biological Ltd) according to the manufacturer's instructions.

2.2.4. Transformation into competent *E. coli* CA-434

Competent cells were prepared by growing an overnight culture in LB broth and using 2 ml to inoculate a further 100 ml of LB broth. Cells were grown to $OD_{600} = 0.3$. The broth culture was centrifuged at $4,500 \times g$ for 5 min at $4^\circ C$. The cells were resuspended in 10 ml ice cold 100 mM $CaCl_2$ and incubated on ice for 20 min. The cells were centrifuged again, as above, and resuspended in 0.5 ml $CaCl_2$. After 15 min incubation on ice, 70 μ l DMSO was added, and the cells were incubated on ice for 15 min. Following this 70 μ l DMSO was added. Cells were aliquoted and snap frozen in dry ice with ethanol. The cells were stored at $-80^\circ C$. The protocol is based on the papers by Mandel and Higa [191] and Cohen *et al.* [192].

Transformation was performed as described for the p-GEM-T vector system (Promega, Southampton, UK). Competent cells were thawed on ice and 50 μ l of cells were added to a tube containing the target DNA. The cells were incubated on ice for 30 min after which the cells were heat-shocked at $42^\circ C$ for 45 sec after which the tube was placed on ice again for 2 min. 950 μ l of SOC medium (Sigma Aldrich Ltd, Gillingham, UK) was added to the tube and the cells were incubated at $37^\circ C$ for one hour with shaking at 200 rpm. After this, the cell suspension was transferred onto selective LB agar (Sigma Aldrich Ltd, Gillingham, UK.) plates and incubated O/N at $37^\circ C$.

2.2.5. DNA Sanger sequencing

DNA Sanger sequencing of PCR products was carried out by the Department of Biochemistry, University of Cambridge or the Wolfson Institute for Biomedical Research, University College London.

PCR products were purified with the Spin column purification kit (NBS Biological Ltd, Cambridgeshire, UK) as described above. DNA samples were 50-100 ng/ μ l. Both strands of the product were sequenced using the primers used for PCR.

2.3. Filter-mating

This method is based on filter-matings described by Mullany *et al.* [164]. Cultures of both donor and recipient strains were grown overnight for 16 h. These were used to start a 10 ml culture of the donor strain and a 50 ml culture of the recipient strain, both at an $OD_{600} = 0.1$. These were grown shaking at 50 rpm. After 4-6 h, when the OD_{600} was between 0.5 and 0.6, the cells were centrifuged for 10 min at 4,500 x g and the pellets were resuspended in 500 μ l BHI broth. The two cultures were mixed and 200 μ l of the mixture was spread onto each of four 0.45 μ m pore size cellulose nitrate filters (Sartorius, Epsom, UK), on BHIB plates. After 24 h the filters were placed into 25 ml tubes and 1 ml BHI broth was added. The tubes were vigorously vortexed and the resulting cell suspension was spread onto selective plates, selective for transconjugants. In filter-matings using *B. subtilis* as recipient, these plates are incubated aerobically in order to select against the *C. difficile* donor. Serial dilutions were spread on selective plates to count the number of donor and recipient cells harvested from the filters. After 72-96 h the putative transconjugants were counted and sub-cultured onto fresh selective plates.

Identity of the transconjugants was confirmed by PCR using a primer pair that is specific for the recipient strain and a primer pair that is specific for the transferring element; these are specified in the relevant chapters.

Chapter 3 Bioinformatics search for novel elements and putative virulence factors

3.1. Introduction

The sequencing of the first three *C. difficile* genomes, strains 630, R20291 and CD196, revealed a number of putative CTNs and other MGEs in each genome. The CTNs are all part of the Tn916-like or Tn1549-like families [8,14]. The genes in the accessory modules were often annotated as ABC transporters for unknown substrates or as hypothetical proteins. After these genomes were published, several other *C. difficile* genomes were sequenced and deposited in Genbank [88].

An essential feature for CTNs is the recombinase responsible for the integration and excision of the element [126]. The sequences of known functional recombinases were used here to search for recombinases and associated CTNs in the *C. difficile* genome sequences.

For *C. difficile* genomes most annotations were performed using gene finder programs that predict CDS such as Glimmer [193]. The function of the CDS is predicted based on DNA sequence homology alignment algorithms such as tBLASTn [194]. Unfortunately, in the 'omics era', Genbank and other databases are starting to fill up with sequences from genome sequencing studies where no biological data is available on most of the annotated CDS. As a result, when performing BLAST searches, a long list of hits is returned that have no supporting biological data and their actual functions may be unrelated to their current annotations.

To predict the biological function of a CDS more accurately another approach is to search with the amino acid sequence of the predicted protein using protein prediction tools. Hernandez *et al.* [195] have described a method for the in-depth analysis of protein sequences using a set of bioinformatics tools to predict the function of proteins annotated as 'hypothetical protein'.

PSI-BLAST is an alignment algorithm that performs BLAST searches on a protein database. The hits of the search can be combined to form a new consensus sequence, which is used to search the database again, known as the Position Specific Iteration. Several rounds of iterations will increase the number of hits, increasing the chance that a homologue with a similar predicted function as the query sequence is identified. However, sequences that are identified with the highest E-values do not necessarily have the most relevant predicted biological function. High E-values in BLAST searches may be caused by a common ancestry of the proteins or structural similarities rather than similar biological function [196].

The results of a PSI-BLAST search can be rearranged using the BYPASS program. This program makes use of a fuzzy logic algorithm that takes into account the amino acid content and length of the matched sequence stretch but also the hydrophobicity and flexibility profiles of the amino acid sequences [197]. These four parameters were all shown to improve the predictive power of the algorithm and therefore, to improve the prediction of the biological function of a query sequence.

Certain biological functions are localised to specific sites in cells. P-SORT uses the protein sorting signals encoded in the amino acid sequence in order to predict the location of the protein in the cell [198].

Both SMART [199] and PRO-DOM [200,201] are programs that can recognise conserved protein domains from various databases. The presence of conserved domains in an amino acid sequence can help to support the predicted the function of the protein.

These programs all predict different protein characteristics and were used on the protein sequences encoded on the *C. difficile* CTns to aid prediction of their biological function.

3.2.Aims of this chapter:

- Identifying novel CTNs in the genome sequences of *C. difficile* strains.
- Bioinformatics analyses of accessory CDS on previously identified and novel CTNs to predict the function of the proteins encoded on the CTNs.

3.3. Materials and methods

3.3.1. Search for novel CTNs

A library of sequences of tyrosine and serine recombinase genes (section 1.2.1.1.3) for which biological activity has been shown was assembled by Dr. Philip Warburton (UCL). The sequences were used to search the *C. difficile* genomes specifically using genomic BLAST. Positive genomes or contigs were compared to the 630 reference genome using DoubleACT [202] and comparisons were visualised using Artemis 12.0 [203] and the Artemis Comparison Tool (ACT) 9.0 [202,204] in order to determine if the recombinase genes were associated with a putative MGE. As a positive control, it was noted that the recombinases of the previously reported elements CTn1, CTn2, CTn4, CTn5, CTn6, CTn7 and Tn5397 all came up as hits when the genome of 630 [8] was analysed, as well as the two putative elements reported in R20291 [14].

3.3.2. Bioinformatics analysis to predict CDS functions

All CDS annotated on the CTNs in the sequence of *C. difficile* strains 630, R20291, CD196, QCD-23M63, QCD63Q42, ATCC 43255, QCD37X79, QCD-66C26, QCD-32G58, QCD76W55 and QCD-97B34 were studied using the set of bioinformatics programs described below.

For strains BI-1, BI-9, 2007855, M68 and CF5 only the CDS that have no homologues in previously studied elements were examined with the bioinformatics programs described below. For strain M120 part of the bioinformatics analysis was performed by Dr J. Corver (Leiden University Medical Centre, Leiden, The Netherlands), using only BLAST-P and SMART.

Putative CTNs in contigs and genome sequences that had not been annotated were analysed using the NCBI tool Open Reading Frame (ORF) Finder [205] and BLAST-P in order to predict CDS. All data was collected using Artemis.

Protein sequences were analysed using the following bioinformatics programs: PSI-BLAST [196], BYPASS [197], PRODOM [201], SMART [199] and P-SORT [198]. Programs that allow custom settings were used at the default settings. PSI-BLAST was performed in order to identify distantly related protein sequences. After the fifth iteration, or when no new hits were generated, the PSSM matrix was used in BYPASS analysis. PRODOM and SMART were used to predict functional domains in the protein sequences and P-SORT was used to predict the location of the protein in the cell.

3.4. Results and discussion

A bioinformatics search for novel putative CTNs in sequenced strains of *C. difficile* was performed using the nucleotide sequence of functional recombinase genes. During this study a total of 33 new elements were discovered in 15 *C. difficile* strains, as summarised in Table 3.1. New elements in strains ATCC 43255, QCD-23M63, QCD-32G58, QCD-37X79, QCD-63Q42, QCD-66C26, QCD-76W55, QCD-97B34 were described in Brouwer *et al.* [206] in which we reported a thorough bioinformatics analysis of the new elements, as well as the previously described elements in strains 630, R20291 and CD196. A follow-up study was performed after another five complete *C. difficile* genomes were made publicly available in Genbank (BI1, BI9, 2007855, M68, CF5) published in Brouwer *et al.* [207]. The elements in strain M120 were studied in collaboration with Dr. Jeroen Corver (LUMC, Leiden, The Netherlands) and were published recently [208].

A Tn number was assigned to new elements for which excision from the genome into a circular intermediate could be shown, in agreement with the guidelines of the Tn registry [209], as discussed in Chapter 4. For some elements where the exact integration site could be described, a Tn number was also assigned.

Table 3.1 Overview of CTNs and non-conjugative elements in *C. difficile*.

Data on strain CD37 is based on the genomic sequence that is presented in Chapter 6.

Strain	Ribotype	Accession number	Isolation details	Source	Family of element	Element	Accessory genes, properties
630	012	AM180355	Zurich, Switzerland, 1982	Human with severe pseudomembranous colitis	Tn916	CTn1	ABC transporter, putative cell surface protein.
					Tn916	CTn6	Hypothetical proteins

Strain	Ribotype	Accession number	Isolation details	Source	Family of element	Element	Accessory genes, properties
					Tn916	CTn7	ABC transporter, putative cell surface protein.
					Tn916	Tn5397	<i>tet</i> (M)
					Tn1549	CTn2	ABC transporter
					Tn1549	CTn4	ABC transporter
					Tn1549	CTn5	ABC transporter
					Unknown	Tn5398	<i>erm</i> (B)
R20291	027	FN545816	Stoke Mandeville, UK, 2006	Human with CDAD, associated with an outbreak	Tn916	CTn1-like	ABC transporter, putative cell surface protein.
					Tn1549	Tn6103	Accessory module of CTn5
					Mobilisable	Tn6104	Integrated in Tn6103, ABC transporter, toxin-antitoxin system
					Mobilisable	Tn6105	Integrated in Tn6103, radical SAM protein
					Mobilisable	Tn6106	Integrated in Tn6103, multidrug efflux pump
CD196	027	NC013315	Paris, France, 1985	Human	Tn916	CTn1-like	98-100% sequence ID to the R20291 CTn1-like element
ATCC-43255 (VPI10463)	087	CM000604		Human	Tn916	CTn1-like	ABC transporter, cation family efflux protein, <i>tet</i> (A), putative cell surface protein, integrated within <i>skin</i> ^{CD}
					Unknown	Tn5398-like	
QCD-23M63	078	CM000660	Montreal, Quebec, Canada		Tn916	Tn6073 CTn1-like	ABC-transporter, N-terminal hydrolase, putative cell surface protein
					Tn1549	CTn4-like	Accessory module of CTn4
					Tn1549	Tn6107 CTn5-like	ABC transporter
QCD-32G58	027	CM000287	Quebec, Canada, 2004		Tn916	CTn1-like	98-100% sequence ID to the R20291 CTn1-like element

Strain	Ribotype	Accession number	Isolation details	Source	Family of element	Element	Accessory genes, properties
					Tn1549	Tn6111 CTn5-like	99% sequence identity to Tn6103 in R20291, excluding Tn6105 and Tn6106
QCD-37X79	027	CM000658	London, Ontario, Canada, 2005		Tn916	CTn1-like	98-100% sequence ID to the R20291 CTn1-like element
					Tn1549	CTn5-like	99% sequence identity to Tn6103 in R20291, excluding Tn6105 and Tn6106
QCD-63Q42	001	CM000637	Quebec city, Quebec, Canada, 2005		Mobilisable	Tn6115	ABC transporter
					Tn916	CTn1-like	33 kb accessory module containing ABC transporter, antirestriction ardA, ABC transporter and 5 recombinases, putative cell surface protein
					Tn916	CTn1-like	99% identical to B19 element, lactoglutathione lyase, alpha/beta fold hydrolase ABC transporters, putative cell surface protein
					Tn916	CTn7-like	99% identical to CTn7, ABC transporter, cell surface protein
					Tn1549	CTn5-like	Restriction modification system, helicase.
QCD-66C26	027	CM000441	Quebec, Canada, 2004		Tn916	CTn1-like	98-100% sequence ID to the R20291 CTn1-like element
					Tn1549	Tn6110 CTn5-like	99% sequence identity to Tn6103 in R20291, excluding Tn6105 and Tn6106
QCD-76W55	027	CM000661	Minnesota, Minneapolis, USA, 1988		Tn916	CTn1-like	98-100% sequence ID to the R20291 CTn1-like element
QCD-97B34	027	CM000657	St John's, Newfoundland, Canada, 2004		Tn916	CTn1-like	98-100% sequence ID to the R20291 CTn1-like element
B11	027	FN668941	USA, 1988	Human	Tn916	CTn1-like	98-100% sequence ID to the R20291 CTn1-like element

Strain	Ribotype	Accession number	Isolation details	Source	Family of element	Element	Accessory genes, properties
B19	001	FN668944	USA, 2001	Human	Tn916	CTn1-like	99% identical to QCD-63Q42 CTn1-element
					Tn1549	CTn5-like	99% identical to QCD-63Q42 CTn5-element
					Tn916	CTn7-like	99% identical to CTn7, ABC transporter, cell surface protein part of conjugation module is not present due to a gap in the sequence
2007855	027	FN665654	USA, 2007	Bovine	Tn916	CTn1-like	98-100% sequence ID to the R20291 CTn1-like element
					Tn916	CTn1-like	Erythromycin resistance gene and toxin-antitoxin system, 99% identical to M68 CTn1-like element
					Tn1549	CTn5-like	ABC transporter, insertion of a putative mobilisable transposon containing aminoglycoside resistance
M68	017	FN668375	Ireland, 2006	Human, associated with an outbreak	Tn916	Tn916-like	Tetracycline and β -lactamase resistance, cell surface protein
					Tn916	CTn1-like	Erythromycin resistance and toxin-antitoxin system, 99% identical to 2007855 CTn1-like element
					Tn1549	CTn5-like	ABC transporter, identical to CF5 CTn5-like element
CF5	017	FN665652	Belgium, 1995	Human, asymptomatic patient	Tn1549	CTn5-like	ABC transporter, identical to M68 CTn5-like element
M120	078	FN665653	Ireland	Human diarrhoeal patient	Tn916	Tn6190	Accessory module of Tn916 including <i>tet(M)</i>
					Tn1549	Tn6164	Composite element, accessory genes include <i>tet(44)</i> , several aminoglycoside adenylyltransferases, and a restriction modification system

Strain	Ribotype	Accession number	Isolation details	Source	Family of element	Element	Accessory genes, properties
CD37	009	AHJJ01000000	United States, 1980	Clinical sample	Tn916	CTn1-like	ABC transporter
					Tn1549	CTn4-like	ABC transporter
					Tn916	CTn7-like	ABC transporter

3.4.1. CTn1-like elements

CTn1-like elements are CTNs of the Tn916 family with high sequence homology to CTn1. Strain R20291 contains two putative CTNs [14]. One of these elements is a CTn1-like element with approximately 82% sequence identity to CTn1, excluding the accessory region and an insertion in the conjugation module (Figure 3.1). One deletion at the right side of the element encompasses genes encoding homologues of two putative hypothetical proteins and the excisionase gene *xis*, predicted to play a role in the excision of CTn1 in strain 630. Experiments described in Chapter 4 could not detect the excision of the CTn1-like element. The accessory region of the element differs from CTn1 and encodes an ABC transporter system, a TetR family transcriptional regulator and a cation efflux family protein, amongst others. Information on the putative functions of the other CDS in this element are listed in Appendix I. Compared to CTn1 in 630, an insertion within the conjugation module contains further predicted accessory genes. The element is integrated in a novel target site compared to CTn1, between homologues of a hypothetical protein (CD3614) and a predicted transcriptional regulator (CD03615) from 630.

Strains QCD-66C26, QCD-32G58, QCD-37X79, QCD-76W55, QCD-97B34, BI1 and 2007855 all contain a CTn1-like element with 98-100% nucleotide sequence identity

with the element described in R20291. All the elements lack *xis* and are integrated in the same target site.

Strain ATCC 43255 contains a CTn1-like element that has integrated into the *skin*^{CD} element. The *skin*^{CD} is a genetic element involved in transcriptional regulation of *sigmaK* that is essential for efficient sporulation in *C. difficile* [179] (Chapter 1). The presence of the CTn1-like element within *skin*^{CD} does not prevent efficient sporulation [179]. The accessory region of the CTn1-like element differs from CTn1 in 630 and includes genes predicted to encode an ABC transporter system and a Tet(A) tetracycline resistance protein (Figure 3.1). The element contains an insertion within the conjugation module which is identical to the insertion seen in the CTn1-like element in R20291. Excluding these regions, the element has approximately 73% nucleotide sequence identity with CTn1 from 630.

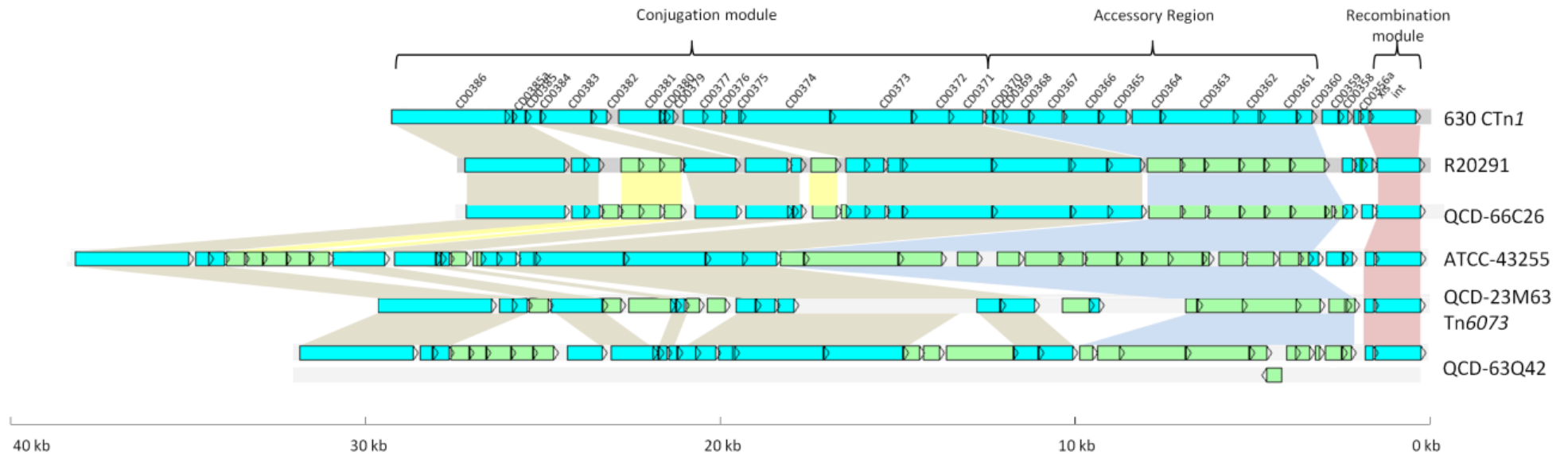


Figure 3.1 Schematic representation of CTn1 and the CTn1-like elements in strains R20291, QCD-66C26, ATCC 43255, 23M63 and QCD-63Q42.

CDS are represented by arrows. All blue CDS have homologues in CTn1 in strain 630. Green CDS are not present on CTn1. Red, brown and yellow boxes shows regions with sequence similarity. Red boxes are part of the recombination module, brown boxes are part of the conjugation module, and yellow boxes show sequence similarity between regions of the elements that are not present in CTn1 of 630. Blue boxes are part of the regulation and accessory module but do not show sequence similarity. The element shown for QCD-63Q42 also represents the element in strain BI9. The element shown for QCD-66C26 also represents the elements in strains QCD-32G58, QCD-37X79, QCD-76W55, QCD-97B34, BI1 and 2007855. A digital version of this image is included on the disc at the back of this thesis.

Strain QCD-23M63 contains the only CTn1-like element for which excision from the genome could be detected (Chapter 4) and was designated Tn6073. Tn6073 contains 3 small insertions in the conjugation module in addition to a different accessory module, compared to CTn1, (Figure 3.1). Excluding these insertions, the remainder of the sequence has 85% nucleotide sequence identity with CTn1. The accessory module of the element is predicted to encode an ABC transporter system, a two component regulatory system and a σ 70 sigma factor. The accessory module also encodes a predicted member of the N-terminal hydrolase family, a family which includes bile salt hydrolases involved in bile resistance, β -lactam acylases involved in resistance to β -lactams and N-acyl homoserine lactone acylases involved in quorum sensing. Tn6073 has integrated in a different target site compared to CTn1, intragenic between homologues of a predicted membrane protein (CD0651) and a transcriptional regulator (CD0652).

Strains QCD-63Q42 and BI9 contain nearly identical CTn1-like elements (99% sequence identity) which both have integrated between homologues of a predicted fructokinase (CD1806) and an NADPH-dependent FMN reductase (CD1807) (Figure 3.1). The accessory region of these CTn1-like elements differ from CTn1 in 630 and encode a putative ABC transporter system, a σ 70 sigma factor and three transcriptional regulators of the TetR, Xre and CheY families. Excluding the insertions in the conjugation region and the differences in the accessory region, the elements have 83% nucleotide sequence identity to CTn1.

Strain QCD-63Q42 contains a second CTn1-like element integrated between homologues of 630 hypothetical proteins (CD1564 and CD1565). An insertion of approximately 20 kb is present in the conjugation module, (Figure 3.2). This insertion

has 92% nucleotide sequence identity to sequences present in ϕ C630-1 of strain 630 but, as the sequence is currently incomplete, it is currently not possible to be sure if the element contains a complete phage [206]. The accessory region of most CTn1-like elements described here are approximately the same size (9.5 kb) as that of CTn1 in 630. However, the accessory region of the element in strain QCD-63Q42 is approximately 33 kb, and the length of the whole element is nearly 70 kb. The accessory module encodes 34 CDS including two predicted σ 70 sigma factors, five site specific recombinase proteins and two transcriptional regulators, one of the AraC and one of the MarR families. The accessory region also encodes a homologue of *ardA* also present on Tn916 and Tn6000 [210,211]. ArdA mimics the structure of DNA and binds endonucleases in order to evade the host restriction modification system. Excluding the phage-like insertion described above and the accessory region, the element has 73% nucleotide sequence identity to CTn1.

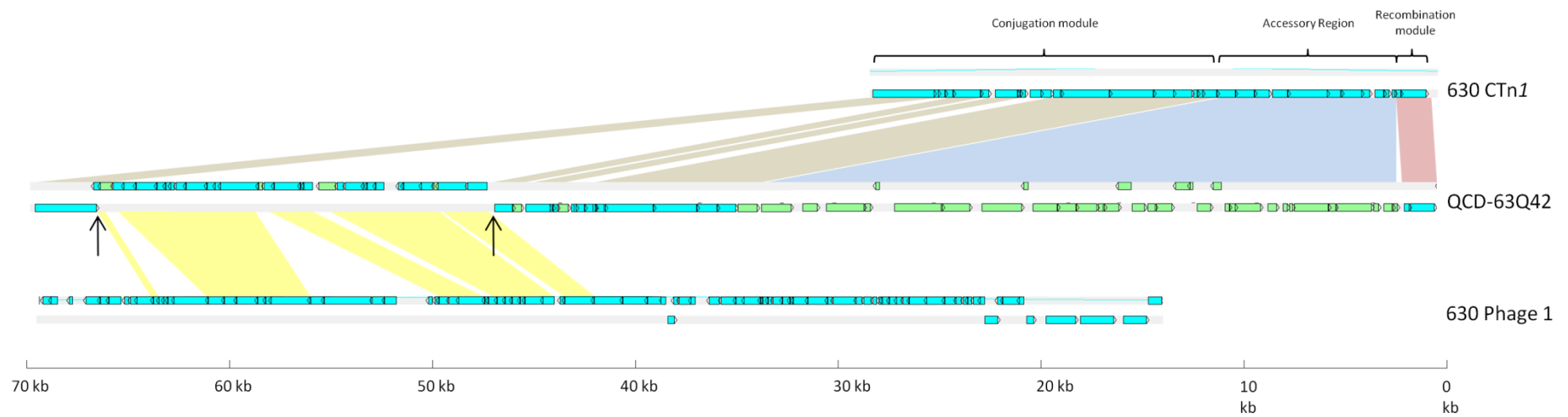


Figure 3.2 Schematic representation of CTn1 and phage 1 in 630 and the CTn1-like element in QCD-63Q42.

CDS are represented by arrows. All blue CDS have homologues in CTn1 in strain 630. Green CDS are not present in CTn1. Red brown and yellow boxes show regions with sequence similarity. Red boxes are part of the recombination module, blue boxes are part of the regulation and accessory module, brown boxes are part of the conjugation module, and yellow boxes are part of phage 1. Black arrows indicate gaps in the sequence of QCD-63Q42. Blue boxes are part of the regulation and accessory module but do not show sequence similarity. A digital version of this image is included on the disc at the back of this thesis.

Strains 2007855 and M68 contain CTn1-like elements which are 99% identical to each other. The conjugation region of these elements is 94% identical to CTn1 (Figure 3.3). Amino acid sequence comparison predicts the genes in the recombination module of the elements encode tyrosine recombinase and excisionase proteins, although these do not have any DNA sequence homology to any known recombinase and excisionase genes. Therefore, if functionality of the recombinase is shown in future study, it would be interesting to include in the recombinase library used here. The elements have integrated into different target sites, the element in 2007855 is intragenic in a gene encoding a hypothetical protein (homologue of 630 CD2831) and in M68 is intragenic in a putative oxidoreductase gene (CD3317). The accessory modules of the elements contain a region of approximately 2.2 kb that has 97% nucleotide sequence identity to a part of Tn5398 containing homologues of *erm(B)*, ORF3 and ORF298 [172]. Furthermore, the accessory modules of both elements encode homologues of the omega-epsilon-zeta operon, a toxin-antitoxin system which is involved in plasmid maintenance in Gram-positive bacteria [212]. In *Vibrio cholerae*, a similar system has been shown to stabilise the CTn SXT after circularisation [213,214]. A mutation in the zeta toxin gene results in an early stop codon in both elements at R52, halfway through the conserved toxin domain indicating that a toxin is unlikely to be produced. The fact that both elements encode the same mutation suggests the mutation was acquired before dissemination of the element to, or between these *C. difficile* strains. An alternative role for orphan antitoxin genes has previously been described as an anti-addiction system, protecting the cell from elements encoding similar toxin-antitoxin systems entering the cell [215], and the antitoxin and regulator from this operon may therefore still be functional.

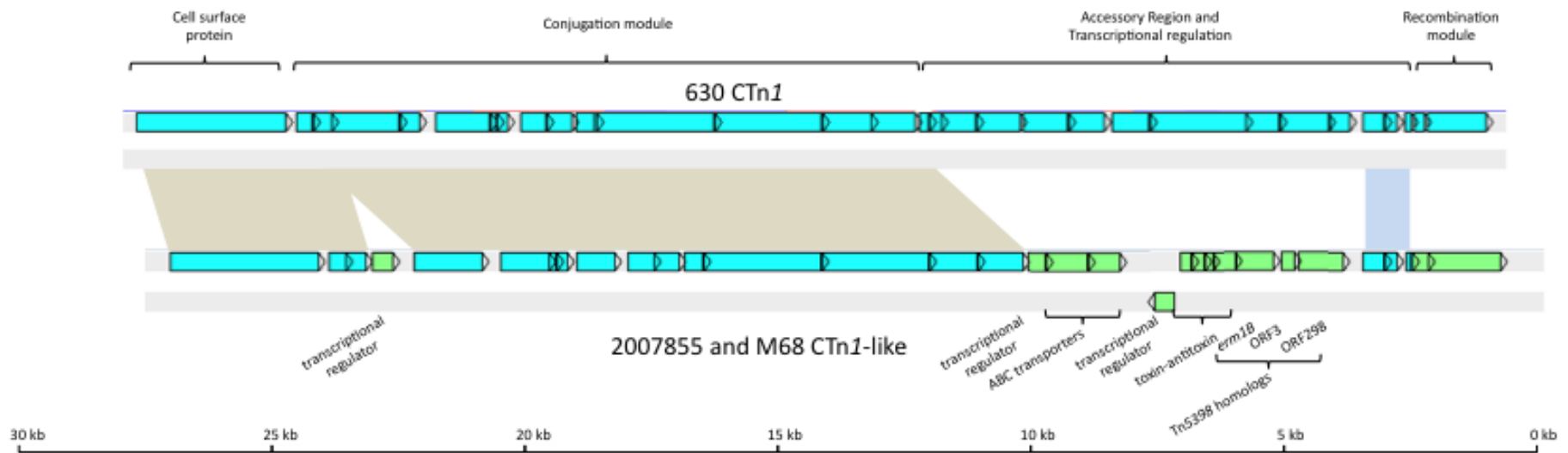


Figure 3.3 Schematic representation of CTn1 and the CTn1-like elements in strains 2007855 and M68.

CDS are represented by arrows. All blue CDS have homologues in CTn1 in strain 630. Green CDS are not present on CTn1. Coloured boxes show regions with sequence similarity. The blue box is part of the regulation and accessory module, brown boxes are part of the conjugation module. A digital version of this image is included on the disc at the back of this thesis.

3.4.2. CTn4-like elements

Strain QCD-23M63 contains a CTn4-like element (Figure 3.4) that has integrated in a different target site than CTn4 in 630, within a homologue a gene encoding a putative cell surface protein (CD1036). In this element, approximately 5 kb of the conjugation module of CTn4 is absent and therefore the element may not be capable of conjugative transfer.

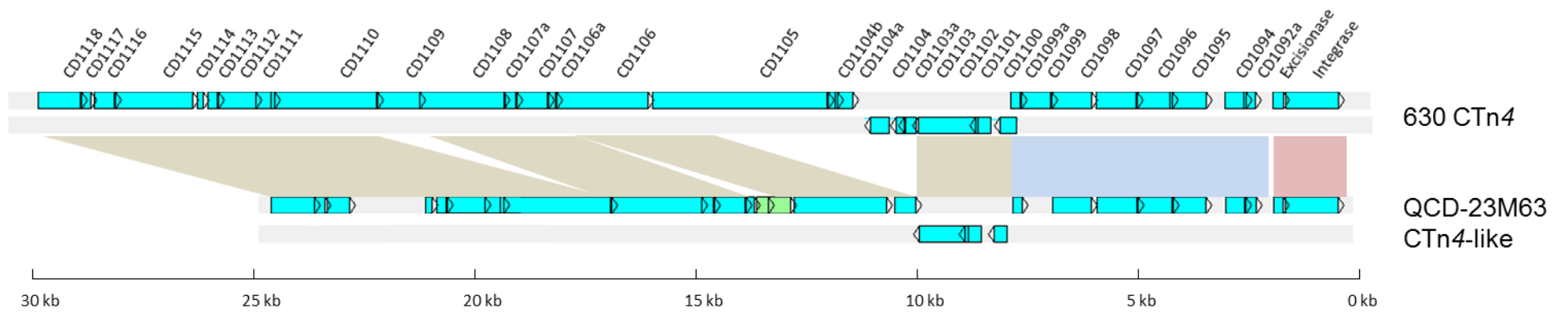


Figure 3.4 Schematic representation of CTn4 and the CTn4-like element in strain QCD-23M63.

CDS are represented by arrows. All blue CDS have homologues in CTn1 in strain 630. Green CDS are not present on CTn1. Coloured boxes show regions with sequence similarity. The red box is part of the recombination module, the blue box is part of the regulation and accessory module, brown boxes are part of the conjugation module. A digital version of this image is included on the disc at the back of this thesis.

3.4.3. CTn5-like elements

Strain R20291 contains an element that has previously been reported as unique to ribotype 027 strains, designated CTn027, and reported to contain a 20 kb phage island called Stoke Mandeville Phage Island (SMPI) [14]. Bioinformatics shows that only a part of the SMPI sequence is present in other ribotype 027 strains such as QCD-66C26 (Figure 3.5). Here, in-depth analysis of the integration site of the SMPI showed that the sequence actually consisted of three separate elements integrated into two homologues of CTn5 CDS (Figure 3.6). The complete element was renamed Tn6103 and the three insertions were designated Tn6104, Tn6105 and Tn6106 as they all resemble transposons in their own right.

The core element of Tn6103, excluding Tn6104, Tn6105 and Tn6106, has 85%-99% sequence identity with CTn5. Tn6103 is present in the same target site as CTn5 in 630 and nearly all CDS of CTn5 are present.

At the left end of Tn6104 (Figure 3.6), CDS 1744 and 1745 are homologues of the serine recombinase *tnpX* and *tnpV* of Tn4451 and Tn4453, respectively, mobilisable transposons of *C. perfringens* and *C. difficile* [166,216,217]. Although TnpV has some similarity to Xis from Tn916 (18%), only TnpX is essential in the excision of these elements from their respective chromosomes and the function of TnpV remains unknown [166,217]. Furthermore, Tn6104 also contains a *mobA/mobL* homologue (CDS 1758). In contrast to the single *catP* accessory gene present on Tn4451/Tn4453, Tn6104 encodes thirteen putative accessory proteins including a predicted transcriptional regulator of the Xre family (CDS 1747), a two-component regulatory system (CDS 1748-1749), an ABC transporter system (CDS 1750-1752), three σ_{24} sigma factors (CDS 1754-1756) and a toxin-antitoxin system (CDS 1759-1760).

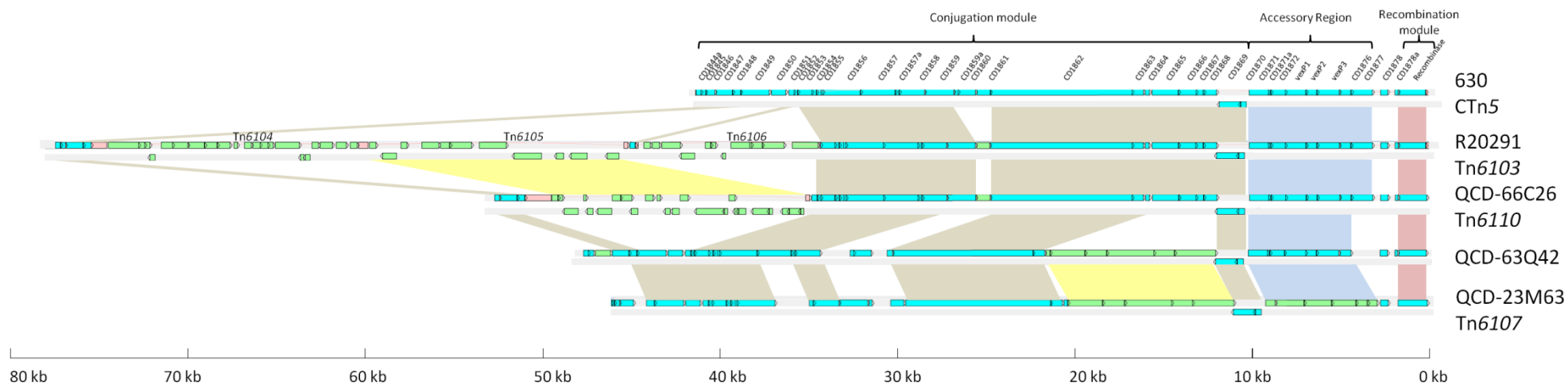


Figure 3.5 Schematic representation of CTn5 and the CTn5-like elements in strains R20291, QCD-66C26, QCD-63Q42 and QCD-23M63.

CDS are represented by arrows. All blue CDS have homologues in CTn5 in strain 630. Green CDS are not present in CTn5. Red, brown and yellow boxes show regions with sequence similarity. Red boxes are part of the recombination module, blue boxes are part of the regulation and accessory module, brown boxes are part of the conjugation module, yellow boxes show sequence similarity between regions of the elements that are not present in CTn5. Blue boxes are part of the regulation and accessory module but do not show sequence similarity. The element presented for Tn6110 in QCD-66C26 also represents Tn6111 in QCD-32G58 which is 99% identical. A digital version of this image is included on the disc accompanying the thesis.

Both Tn6105 and Tn6106 encode two predicted serine recombinases which could be involved in excision of these elements from the genome (respectively CDS1771-1772 and 1788) (Figure 3.6). Both elements also encode TnpV homologues (CDS 1765 and 1777) and putative mobilisation proteins (CDS 1768 and 1784). Tn6105 encodes eight putative accessory genes including a transcriptional regulator (CDS 1766) and a σ 70 sigma factor whereas Tn6106 encodes nine accessory proteins including a multidrug efflux pump protein (CDS 1779) and a transcriptional regulator of the Xre family.

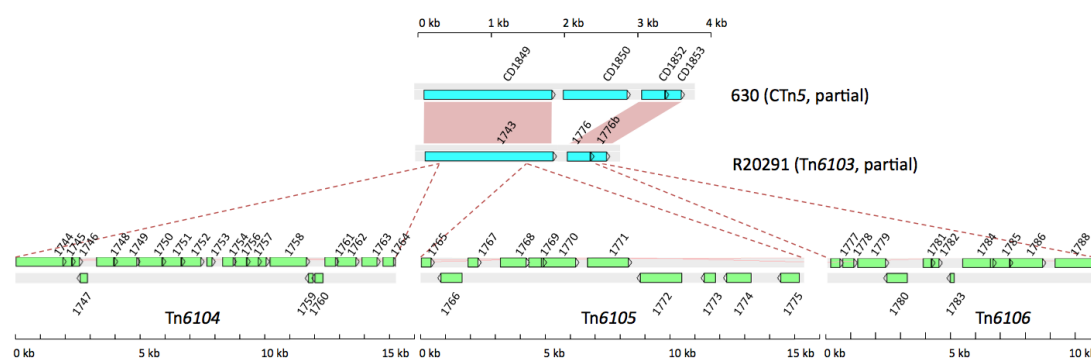


Figure 3.6 Schematic representation of CTn5 and the insertions present in Tn6103.

CTn5 and Tn6103, are compared to show the sites where Tn6104, Tn6105 and Tn6106 have inserted in the sequence. CDS are represented by arrows. All Blue CDS have homologues in CTn5 in strain 630, all green CDS are not present in CTn5, red boxes show sequence similarity. Insertion sites are indicated by red dotted arrows. A digital version of this image is included on the disc accompanying the thesis.

Strains QCD-66C26 and QCD-32G58 each contain a CTn5-like element which have been shown to excise from their respective chromosomes into circular intermediates (Chapter 4) and were designated Tn6110 and Tn6111 respectively. Strain QCD-37X79 has a CTn5-like element containing 99% sequence identity with Tn6110. All three elements have insertions identical to Tn6105 in Tn6103, however, none contains Tn6104 or Tn6106 (Figure 3.5). All three elements are present in similar target sites, between 630 homologues CD3369 and CD3393 which is the target site of CTn7 in 630.

Strains QCD-63Q42 and BI9 contain CTn5-like elements that are 99% identical to each other, containing similar accessory modules as CTn5 in 630 but containing a substitution in the conjugation module (Figure 3.5). Excluding this substitution, there is approximately 88% sequence identity between the elements in QCD-63Q42 and BI9 and CTn5 in 630. The substituted region in the new elements encodes a predicted restriction modification endonuclease and two methylase proteins. These could function as a form of molecular vaccination in which methylation protects the incoming element from host endonucleases and, following integration, will protect the host chromosome from endonucleases present on other mobile genetic elements when these enter the cell [218]. Due to frameshift mutations, there are no functional homologues of several CDS in the conjugation module and it is therefore conceivable that the element no longer transfers via conjugation. The CTn5-like elements in strains QCD-63Q42 and BI9 are both present in tandem with a homologue of CTn7 in the homologue of the CTn7 target site.

Strains QCD-23M63 (Figure 3.5) and 2007855 (Figure 3.7) both contain CTn5-like elements which have integrated in the CTn7 target site without homologues of CTn7 being present. The element in QCD-23M63 was shown to excise from the genome and was designated Tn6107 (Chapter 4). The element in strain 2007855 has an insertion in the homologue of a conjugation related gene which encodes a predicted serine recombinase and a MobA mobilisation protein and is therefore hypothesized to be a mobilisable transposon.

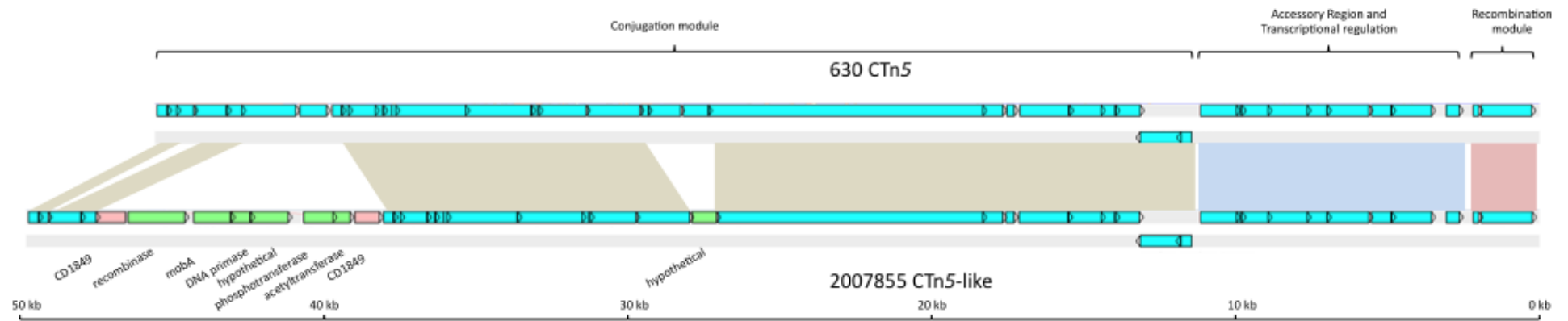


Figure 3.7 Schematic representation of CTn5 and the CTn5-like element in strain 2007855.

CDS are represented by arrows. All blue CDS have homologues in CTn5 in strain 630. Green CDS are not present in CTn5. Pink CDS are interrupted by an insertion. Coloured boxes show regions with sequence similarity. The red box is part of the recombination module, the blue box is part of the regulation and accessory module, brown boxes are part of the conjugation module. A digital version of this image is included on the disc at the back of this thesis.

3.4.4. Tn916

Strain M68 contains an element with 98% sequence identity to Tn916 excluding two insertions (Figure 3.8). The accessory module of Tn916 is present on the element, including *tet(M)*, which indicates the element may confer tetracycline resistance. The first of the two insertions is between homologues of Tn916 *orf21* and *orf22* and is predicted to encode a protein containing a metallo β -lactamase domain, which may play a role in resistance to β -lactams [219]. Although *C. difficile* isolates are generally resistant to a high number of β -lactams, not all isolates are resistant and β -lactamases may account for some of the resistance that has been recorded [220,221]. The second insertion replaces *orf24* with a homologue of the putative cell surface protein found on CTn1 (93% DNA sequence identity)(see 3.4.7). The element has integrated in a predicted exosporium glycoprotein gene (CD0332).

Strain M120 contains a Tn916-like element that has 97% sequence identity to Tn916 which was designated Tn6190 as conjugative transfer of the element has been shown (Chapter 4). The difference between Tn6190 and Tn916 is the absence of a homologue of *orf12* of on Tn6190 (Figure 3.9). ORF12 is involved in transcriptional regulation of Tn916 [222].

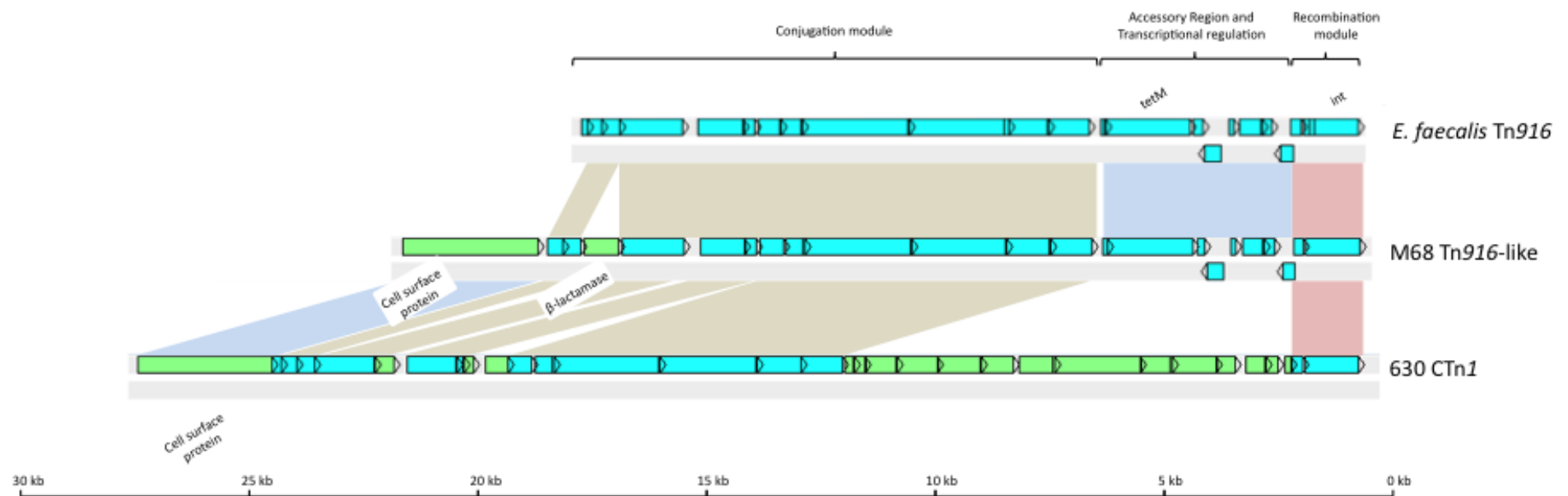


Figure 3.8 Schematic representation of Tn916 compared the Tn916-like element in strain M68 and CTn1.

CDS are represented by arrows. All blue CDS have homologues in Tn916 in *E. faecalis*. Green CDS are not present in Tn916. Coloured boxes show regions with sequence similarity. The red boxes are part of the recombination module, blue boxes are part of the regulation and accessory module, brown boxes are part of the conjugation module. A digital version of this image is included on the disc at the back of this thesis.

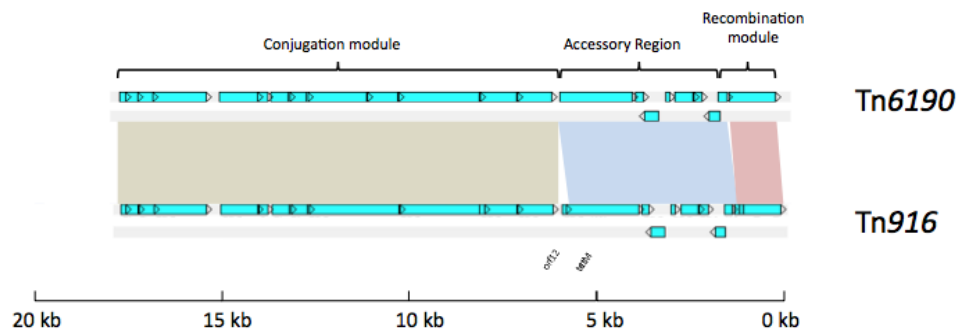


Figure 3.9 Schematic representation of Tn6190 in M120 compared to Tn916 in *E. faecalis*.

CDS are represented by arrows. Coloured boxes show regions with similarity. The red box is part of the recombination module. The blue box is part of the regulation and accessory module, the brown box is part of the conjugation module.

3.4.5. Tn6164

The sequence of a 107 kb element in strain M120 is derived from several different genetic elements (Figure 3.10). The element is present in the target site of CTn2. As circularization of the element has been shown, it was designated Tn6164 (Chapter 4). On the left end of the element is 7.3 kb of sequence that has no homology to known sequences. The fragment encodes two putative modification methylases and a LlaJI restriction endonuclease, which indicates a restriction modification system, as discussed above for the CTn5-like element of QCD-63Q42.

This sequence is followed by an element of approximately 39.5 kb, which shows 92% sequence identity to a *Thermoanaerobacter sp.* sequence. Both of these sequences appear to be complete prophages yet neither has previously been reported.

A 4.5 kb fragment of sequence is 99% identical to part of the *E. faecalis* plasmid pEF418 (Figure 3.10). The sequence encodes three putative nucleotidyl transferases which could play a role in resistance to aminoglycosides, one has most homology to

kanamycin resistance proteins, the second to streptomycin and the third to spectinomycin resistance proteins and it is known that pEF418 confers streptomycin and spectinomycin resistance [223].

The right end of the element is homologous to part of the conjugation and accessory modules of Tn1806 of *Streptococcus pneumoniae* [224]. An insertion of approximately 4.5 kb in the conjugation module has 90% sequence identity to a multidrug resistance cassette of the transferable pathogenicity island of *Campylobacter fetus subsp. fetus*. This sequence contains 4 CDS which encode three putative antibiotic resistance proteins and a padR transcriptional regulator. In *C. fetus subsp. fetus* the *tet(44)* and *ant(6)-Ib* genes confer tetracycline and streptomycin resistance respectively [225].

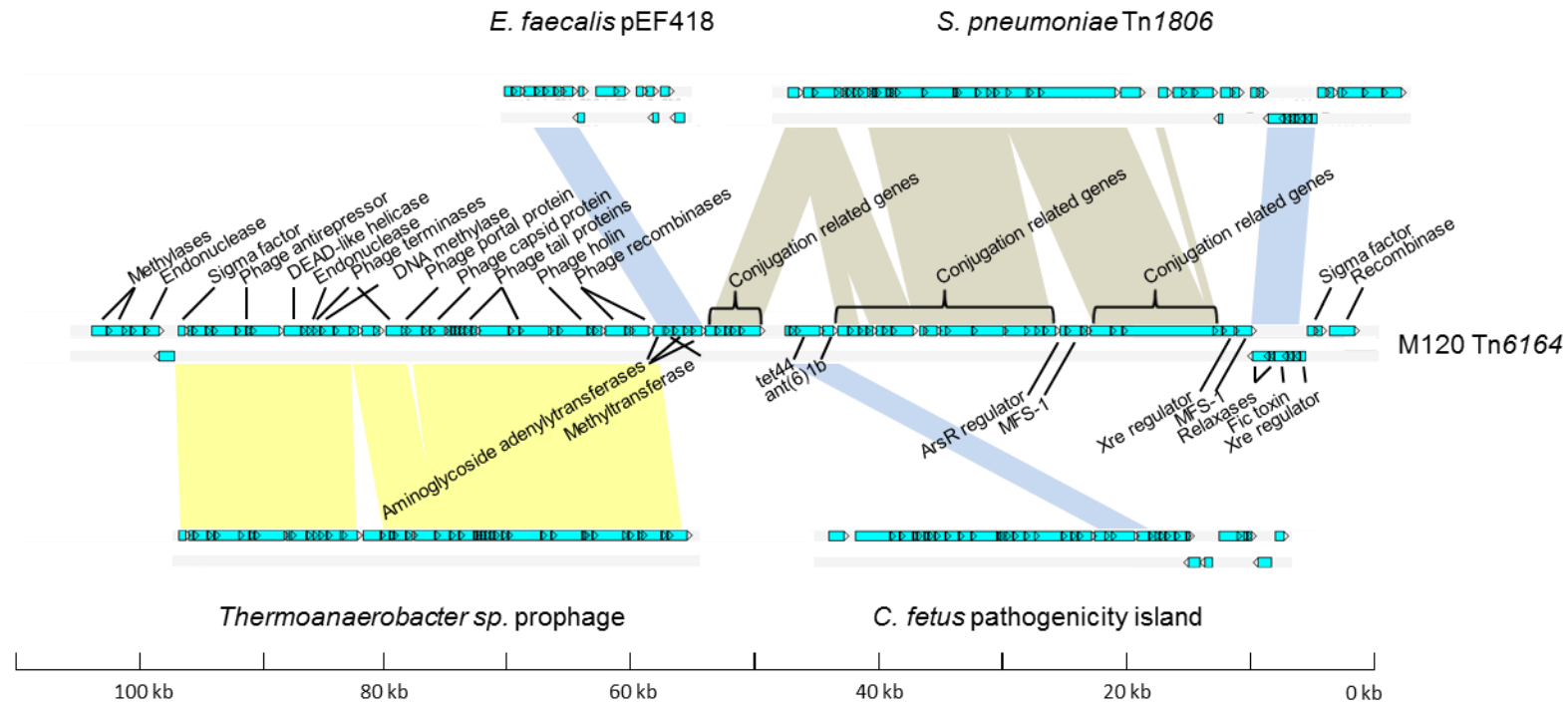


Figure 3.10 Schematic representation of Tn6164 and *E. faecalis* pEF418, *S. pneumoniae* Tn1806, *Thermoanaerobacter sp.* prophage and the *C. fetus* pathogenicity island. CDS are represented by blue arrows. Blue, brown and yellow boxes show regions with sequence similarity. Blue boxes are for accessory genes, brown boxes are part of the conjugation module, yellow boxes are phage related genes. A digital version of this image is included on the disc at the back of this thesis.

3.4.6. Non-conjugative genetic elements

A genetic element is present in strain ATCC 43255 that has 99% sequence identity to Tn5398 excluding the erythromycin resistance conferring *erm(B)* cassette (Figure 3.11). The sequence between the direct repeats of the *erm(B)* cassette is not present and a hypothetical gene is present which in Tn5398 has been interrupted by the *erm(B)* cassette. As it is currently unknown what the mechanism of transfer of Tn5398 is, it is impossible to predict if the element in strain ATCC-43255 can transfer in a similar fashion.

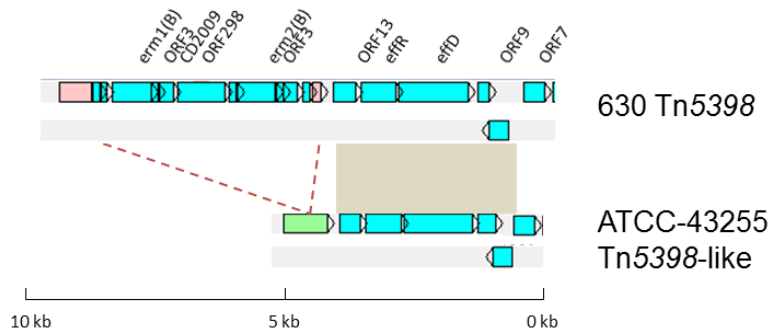


Figure 3.11 Schematic representation of Tn5398 compared to the Tn5398-like element in ATCC-43255. CDS are represented by arrows. The blue CDS are present in strain 630. The green CDS present in ATCC-43255 is present in 630, shown in pink, but was interrupted by the *erm(B)*-cassette. The CDS shown in pink has acquired a number of early stop codons and frame-shift mutations. The brown box shows CDS with similar predicted functions.

In strain QCD-63Q42 a novel putative transposon has integrated into a predicted transcriptional regulator (Figure 3.12). The element is flanked by a GG dinucleotide direct repeat and was designated Tn6115. The element consists of 12 CDS including a serine recombinase, an ABC transporter system and a two-component regulator system.

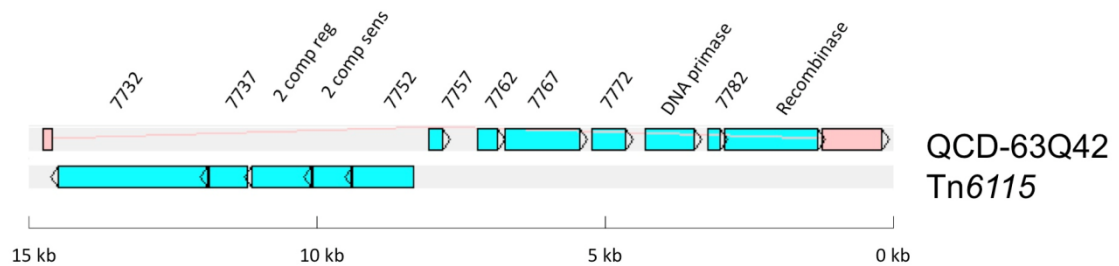


Figure 3.12 Schematic representation of the putative mobilisable element Tn6115 in QCD-63Q42. The blue arrows represent the CDS on the element, the pink arrow represents the interrupted CDS, homologue of CD1573, into which the element has integrated.

3.4.7. Cell surface proteins encoded on CTNs

All CTn1-like elements, including CTn7 from 630, encode a putative cell surface protein which has been annotated as a putative collagen binding protein because of its amino acid sequence similarity to the *Staphylococcus aureus* collagen binding protein [226]. The *S. aureus* collagen binding protein consists of a CnaA domain, which can bind collagen, and seven CnaB domains, which can form a stalk-like structure projecting the collagen binding domain away from the cell wall [226] (Figure 3.13). The protein also contains an N-terminal signal peptide to transport it to the cell membrane and a C-terminal LPXTG domain, a signal for sortase mediated cell wall anchoring, which is found commonly in Gram-positive cell surface proteins [227]. A similar protein is predicted to be encoded on Tn5386, a Tn916-like element discovered in *E. faecium*, which contains eight CnaB domains [228]. These domains are predicted to act as a stalk for an intimin / invasion domain (Figure 3.13) a conserved domain that is involved in adherence in Gram-negative organisms [229].

The cell surface protein encoded on CTn1 contains two CnaB domains but does not contain any other known protein domains (Figure 3.13). This suggests the protein is

not involved in collagen binding [230]. The gene was not essential for conjugative transfer (Chapter 4), which indicates this is the only accessory gene conserved over all CTn1-like elements investigated here. With the exception of strains CF5 and M120, all *C. difficile* strains described here contain a CTn1-like element. The fact that a highly conserved accessory gene with no apparent function in conjugation is present in nearly all *C. difficile* strains studied here supports the idea that the cell surface protein would confer some advantage to its host cell.

The next step in the study of this protein should be to determine whether the protein is expressed and at what location, eg. on the cell surface or in the cytoplasm. If the protein is expressed on the cell surface it could be involved in interactions with the environment and the host, indicating the importance of further study of this protein.

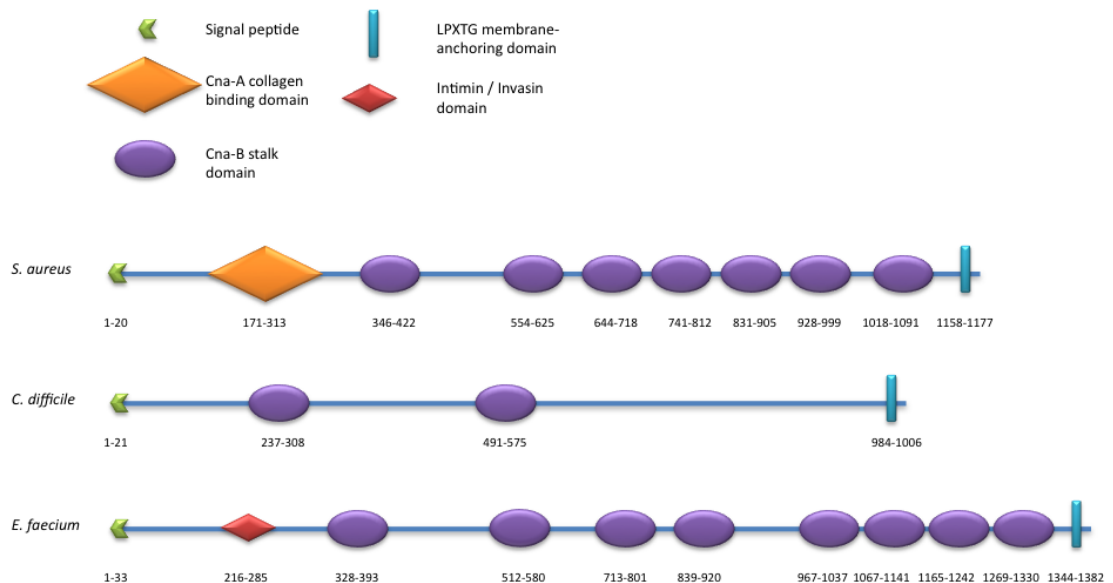


Figure 3.13 Schematic representation of the cell surface proteins of *S. aureus*, *C. difficile* and *E. faecium*.

Representation of the collagen binding protein of *S. aureus* and the putative cell surface proteins of *C. difficile* and *E. faecium*. Symbols for the protein domains are listed above the figure, the location of the domains on the peptide sequence is indicated by amino acid numbers below each domain. The N-terminus is depicted at amino acid 1.

3.5. Conclusions

A new approach was used during this study to identify MGEs by searching genome sequences for similarity with the conserved sequences of recombinase genes. A full overview of all *C. difficile* CTNs described so far is presented in Table 3.1. A total of 37 novel elements were identified by searching for known recombinases and also by searching with the CDS of CTNs of strain 630.

A re-annotation of all of the CDS of the *C. difficile* 630 genome was recently published [230]. This has led not only to an update to the putative functions of a large number of CDS but also to the modifications of predicted CDS start sites and the prediction of new CDS. However, no changes were made to the annotations of the mobile genetic elements in strain 630.

The analysis of the element previously reported as the SMPI on CTn027 [14] in strain R20291 showed the element is not a phage island. CTn027 was renamed Tn6103 [206] and the integrated SMPI was shown to consist of three separate insertions which were named Tn6104, Tn6105 and Tn6106. Some new predicted functions were assigned to the CDS on these three elements including four genes predicted to be sigma factors of the σ_{24} and σ_{70} families. It is interesting that these relatively small elements contain more than one sigma factor in combination with other transcriptional regulators and it is conceivable that these will not only affect the regulation of the genes on the element itself but may also affect the transcription of chromosomal genes.

During this bioinformatics analysis it became clear that many of the novel elements contain accessory genes which are homologues of accessory genes in elements in 630 and R20291. For those of which the accessory module was different, a number of ABC-

transporter systems were found. Prokaryotic ABC transporter systems consist of a minimum of two nucleotide binding domains (NBD) that bind and hydrolyse ATP and two transmembrane domains that provide the passageway through the membrane for the transported molecules encoded on one or several polypeptides [231]. In the results described above, when at least one NBD and one TMD was discovered in one or more CDS adjacent to each other, this was described as a predicted ABC transporter system, regardless of the number of NBD and TMD. In prokaryote ABC transporters the NBD and TMD are often separate polypeptide chains in importing transporters whereas for exporting transporters these domains are fused on a single polypeptide chain [231]. As all of the ABC transporter systems that were studied here consist of separate polypeptide chains for the NBD and TMD domains, these ABC transporters are likely to function as importers, although no prediction can be made for their substrates.

MGEs are known to encode antibiotic resistance proteins in their accessory domains. Predicted antibiotic resistance genes that were described here include tetracycline resistance genes *tet(A)* on a CTn1-like element in ATCC-43255, *tet(44)* on Tn6164 and *tet(M)* on Tn6190, both in M120. Furthermore, Tn6164 encodes a putative streptomycin resistance gene *ant(6)-Ib* as well as three predicted aminoglycoside resistance proteins which could confer resistance to kanamycin, streptomycin and spectinomycin [223]. Strain M120 is tetracycline, streptomycin and spectinomycin resistant but it is currently unknown if the proteins encoded on Tn6164 are responsible for this phenotype [208]. Further putative antibiotic resistance genes found in this study are *erm(B)* on the CTn1-like elements in strains M68 and 2007855 and the predicted aminoglycoside phosphotransferase and aminoglycoside

acetyltransferase on the CTn5-like element in 2007855 that may confer resistance to aminoglycosides.

Marsden *et al.* [232] have reported a comparative genomic hybridisation of 94 clinical isolates as well as a number of laboratory reference strains using probes of the entire 630 genome and part of the R20291 genome. The main aim of the study was to find a specific ribotype 027 or hypervirulence gene target which could be tested for in clinical isolates. With regards to the MGEs of strain 630 they conclude only CTn7 is present in a limited number of strains, in contrast with the results presented here where CTn1-like elements are found in nearly all strains. The observations by Marsden *et al.* were made from looking at the accessory modules of the elements (Marsden, personal communication). The work presented here shows the accessory module is not appropriate for analysing conservation of MGEs as the accessory module is the most variable.

The target sites of the elements described in this chapter are mostly new target sites in which no elements in *C. difficile* have previously been described. However, some of these target sites have been described before such as the homologue of the CTn2 target site in which Tn6164 has integrated in strain M120. Although both of these elements encode a serine recombinase, the amino acid sequence identity between the two proteins is only 66%. It is interesting that two recombinases with limited homology can integrate two different elements with little homology into identical target sites. Furthermore, it was shown here that 4 different CTn5-like elements in 5 different strains, some with up to 99% sequence identity with the CTn5 recombinase, have all integrated in the homologue of the target site of CTn7 in 630. Presumably, because the preferred target site in strain 630 is occupied by CTn7, CTn5 integrated in

a different target site. However, in strains QCD-63Q42 and BI9 CTn5-like and CTn7-like elements are present which have integrated in tandem in the homologue of the CTn7 target site. CTn5 in strain 630 may also be capable of integrating in tandem with CTn7.

In conclusion, I have shown here the presence of 37 novel elements in 16 sequenced strains of *C. difficile*. Most of these elements were shown to be related to CTn1 and CTn5 respectively of the Tn916 and Tn1549 families. All strains analysed contained at least one CTn of either of these families of elements although most strains were shown to contain multiple elements. The abundance of these elements in the genomes suggests an important role in the evolution of *C. difficile*. Although the accessory functions of most of these elements are still unknown, the bioinformatics analysis performed here has predicted functions for a number of genes including a number of sigma factors and other types of transcriptional regulators as well as antibiotic resistance proteins of several classes.

Chapter 4 Mobility of putative mobile genetic elements

4.1. Introduction

4.1.1. CTns circularize upon excision

During the evolution of genomes and the MGEs they contain, some MGEs become fixed in the genome as a result of mutations in genes associated with mobility. As the transfer of a CTn usually happens when the element is present as a circular intermediate, the excision of the element from the genome is the first step in conjugative transfer [127].

Excision from the chromosome and circularisation of the element can be shown by PCR [156]. A primer pair located on the chromosome, flanking the target site, can be used to show the regenerated target site is present in a sub-population of the cells. PCR using primers on the ends of the element, facing outwards, can be used to amplify the joint of the circular intermediate of the element (Figure 4.1). For CTn4 and CTn5 from strain 630 this method was previously used to determine the ends of these elements [8].

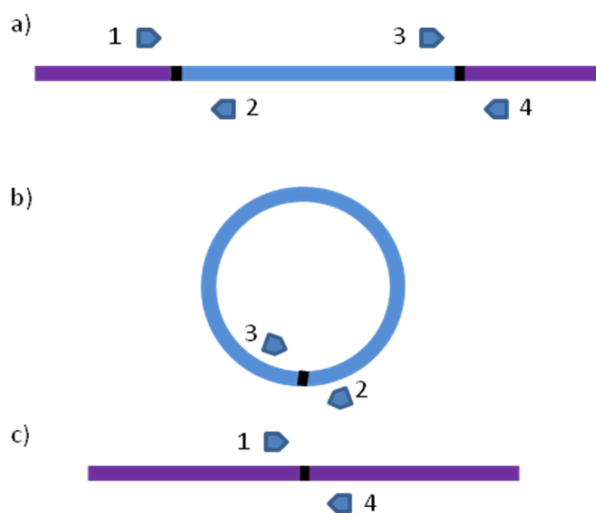


Figure 4.1 Schematic representation of a conjugative transposon within the genome and circularized after excision.

The genome is represented in purple lines, the conjugative transposon in blue. a) The conjugative transposon is inserted within the genome. b) The conjugative transposon is in its circular form. c) The genomic target site is regenerated. Primer pairs 1 & 2 and 3 & 4 are designed to amplify the genome-element junction, primer pair 1 & 4 to amplify the genomic target site and primer pair 2 & 3 to amplify the joint of the circular molecule.

4.1.2. CTn integration sites and copy number

The integration site of a CTn can be hard to predict in a recipient strain as elements often behave differently in different strains. For example, Tn916 has a highly preferred target site in CD37, however, in other strains it can insert in a large number of target sites. A recent study reported sequencing of 112 unique insertion sites in 630Δ*erm* and 102 unique insertion sites in R20291 for Tn916 with some transconjugants containing up to four copies of the element [155]. These sequences were used to generate a 15 bp degenerate consensus sequence which was similar for both strains and was present over 100,000 times in the genome of strain 630Δ*erm*. This consensus sequence is similar to the highly preferred target site in CD37 and is present abundantly in this genome, however, the element only integrates in one site in the genome of this strain, suggesting that other factors beside the target sequence play a role in target site selection [155,233].

Similarly, Tn5397 has a single preferred target site in *C. difficile* strain 630, which is present twice in strain R20291 and the element can insert in both [163]. This site is not present in *B. subtilis* and Tn5397 has been reported in five different sites [156], however, when the sequence from the 630 target site is cloned into *B. subtilis*, Tn5397 usually integrates at this site [234].

4.1.3. Mobilisation of non-conjugative DNA elements

The transfer of non-conjugative DNA elements has been reported for a number of Gram-positive organisms. Transconjugants from filter-matings between different *B. subtilis* strains have been shown to inherit not only the conjugative transposon present in the donor but also non-conjugative plasmids [235]. This type of *in trans* mobilisation only requires the plasmid to encode an origin of transfer (*oriT*) and a conjugative relaxase encoded by *mob* [127]. However, a recent study has shown that the replication relaxase encoded by *rep* and the origin of replication *oriV*, which are essential for plasmid replication, together can function to facilitate *in trans* mobilisation [236]. Furthermore, conjugative transposons can also integrate into non-conjugative plasmids, which can render the plasmid self-transmissible, a process known as mobilisation *in cis* [127,237].

Transfer of conjugative and mobilisable elements starts when a single strand of the DNA is nicked at the *oriT*. These specific regions usually consist of multiple inverted and direct repeats and have a somewhat higher AT ratio than the rest of the sequence of the element [134]. Nonetheless, the regions are not easily identified, a number of specific sequences has been described for conjugative plasmids but for CTns these have only been identified for the extensively studied Tn916 and ICEBs1 and the SXT/R391 family of ICEs [134,236,238].

Two types of putative mobilisable elements have been reported in *C. difficile*, Tn4453 and Tn5398. Mobilisable transposons can excise from the genome into a circular intermediate for transposition but while in this state, they can be mobilised in the same fashion as mobilisable plasmids [169]. In *C. difficile*, Tn4453 can be excised from the genome by a TnpX serine recombinase protein [82]. Furthermore, the element is mobilised by the TnpZ-RS_A mobilization system and carries a copy of *catP*, encoding a chloramphenicol resistance gene [82]. Tn5398 is an erythromycin resistance-conferring MGE for which the mechanism of transfer is still unknown as no obvious conjugation, mobilisation or recombination proteins are encoded by the element [169].

4.1.4. Co-transfer of CTNs

CTNs have the ability to enhance the transfer frequency of other CTNs. It was shown that the transfer frequency of Tn916 and Tn916 Δ E was higher when both elements were present than when only one of these element was present in the donor cell [239]. Additionally, over half of the transconjugants resulting from these matings contain both of the elements [239]. As Tn916 Δ E was derived from Tn916 and these elements have the same conjugation proteins, it is expected that the conjugation machinery of the elements is shared in cells that contain both elements. It is conceivable that other related CTNs can have similar effects on transfer frequency and co-transfer by sharing conjugation machinery. As the *C. difficile* 630 genome contains multiple Tn916 and Tn1549-like elements, transconjugants that were generated in this study were analysed to determine if co-transfer of other elements had occurred.

4.1.5. Clostron provides *C. difficile* targeted mutagenesis tool

The lack of genetic tools for clostridia has severely held back the study of this genus. Chemical transformation and electroporation of *C. difficile* is currently still not possible [190]. For a long time, DNA could only be introduced efficiently using conjugative transposons [131]. Finally, by using a plasmid replicon of the *C. difficile* plasmid pCD6 from strain CD6 (in the literature also referred to as R8375), and by evading the restriction modification systems by methylation of the vector or specifically taking out the restriction enzyme recognition site of the vector, the first autonomously replicating *C. difficile* shuttle vectors were produced [190].

Recently mutagenesis tools for clostridia have been developed. For targeted mutagenesis, there have been several reports of different systems, however, these were inefficient, mutation frequencies were not always reported and plasmids inserted into specific genes via single cross-over mutation looped out in a proportion of the population [47,107,190,240].

The Clostron system is the first efficient tool for the production of targeted gene knockouts in *C. difficile* and other clostridia [241]. The system is based on the TargeTron system, which is commercially available from Sigma-Aldrich and has been validated for the use in a number of organisms [242].

The TargeTron system is based on retargeting a group II intron into a specific target site. Group II introns splice from mRNA into an RNA lariat which is site-specifically integrated into a target site by the intron-encoded reverse transcriptase protein, a mechanism called retrohoming [243]. The specific site recognised within the chromosome for reverse splicing is designated the intron binding site (IBS) which is complementary to the exon binding site (EBS) encoded by the lariat RNA sequence.

The target site can therefore be changed by altering the sequence of the EBS, thus allowing the retargeting of the intron to a specific site (Figure 4.2). A computer algorithm was designed to identify suitable target sites [244].

In order to ensure the intron does not splice from its intended target site and integrate elsewhere, the reverse transcriptase gene is removed from the sequence and included on the shuttle vector that is used to introduce the intron into the recipient cell, allowing the reverse transcriptase to splice the intron *in trans* from the target gene only if the shuttle vector is present. However, the shuttle vector is an unstable plasmid and is lost from the cell after selection is removed [243]. A second selection marker is present at the original location of the reverse transcriptase on the group II intron, which is interrupted by a self-splicing group I intron [242]. The group I intron can only splice from the transcribed group II intron RNA lariat, allowing transcription of the resistance determinant only from the group II intron after it has integrated into the target gene and not from the shuttle vector, a tool called the retrotransposition-activated selectable marker (RAM). After integration the targeted cells should be resistant to the RAM encoded resistance and susceptible to the shuttle vector-encoded resistance. Integration of the group II intron can be tested for by PCR analysis.

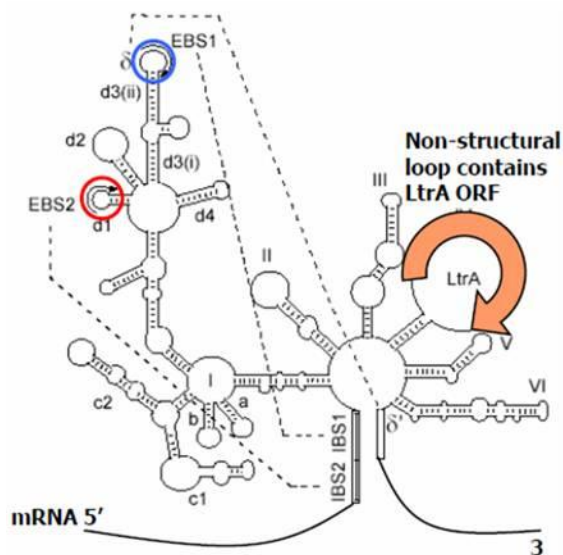


Figure 4.2 Group II intron used for retrohominy

Schematic representation of the secondary structure of the group II intron. EBS1 and 2 are complimentary to IBS1 and 2, which is recognised by the reverse transcriptase before it integrates the group II intron into the chromosomal target site. The non-structural loop LtrA ORF encodes the reverse transcriptase, which is replaced by the RAM in the TargeTron. Figure taken from the Protocol for clostridial Gene knockout using pMTL007 [245].

Heap *et al.* [241] adapted the TargeTron system for use in clostridia. The group II intron was cloned into a clostridial shuttle vector and the resistance genes for the RAM and shuttle vector were replaced for antibiotics clostridia are susceptible for (Figure 4.3). Furthermore, an IPTG inducible promotor was cloned in, although in a revised system, this promotor was exchanged for the constitutive *tcdB* promotor [246]. Further changes in the second version of the ClosTron system was the choice of different Gram+ replicons, which enabled the system to be used in *C. difficile* R20291, which was not efficiently targeted by the previous system [246].

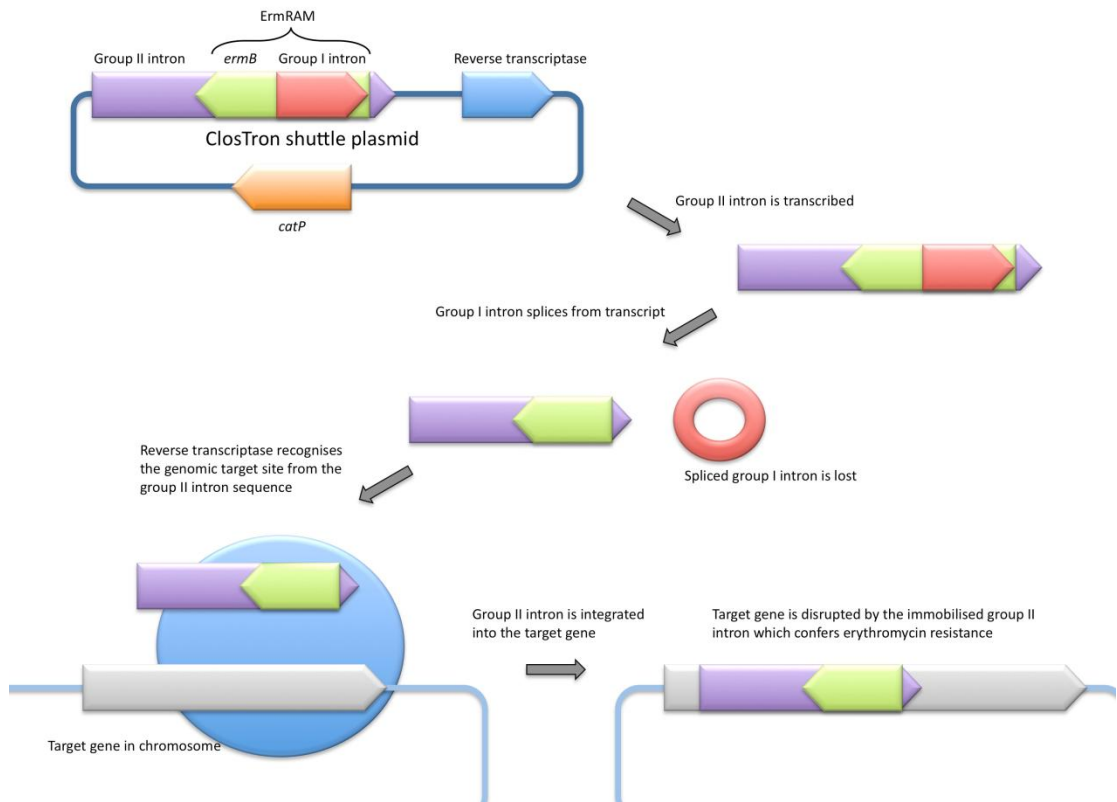


Figure 4.3 Method of gene disruption by the Clostron.

The Clostron shuttle plasmid is unstable and only maintained under the influence of chloramphenicol resistance conferred by *catP*. The group II intron is transcribed and the group I intron splices from the RNA and is lost. The EBS sites on the group II intron were changed to recognise regions in the target gene and the reverse transcriptase (blue circular protein) recognises this site and integrates the group II intron. After the construct is inserted in the target site it is reverse transcribed. Without the disruption of the group I intron, *erm(B)* can confer erythromycin resistance.

4.1.6. Using the Clostron system to create selectable markers

As the putative CTNs of strain 630 and R20291 do not contain any known resistance genes, Clostron mutagenesis was used to introduce *erm(B)*, encoding resistance to macrolides, lincosamides and streptogramin B antibiotics, into each of the elements for which excision from the genome and circularisation was shown. The target sites for the Clostron insertions were chosen in such a way that the chance of affecting the mobility functions of the elements was minimised. Target genes were all within the predicted accessory modules of the elements [8,138] and were therefore not expected to play a role in excision, insertion and / or conjugation.

4.2.Aims of this chapter:

- Determining if the previously described putative CTNs of strains 630 and R20291 and the novel elements described in Chapter 3 for strains M120, QCD-66C26, QCD-23M63 and QCD-32G58, are capable of excision from the genome and circularisation.
- Introducing a selectable marker into CTNs of strains 630 and R20291 which are capable of excision and circularisation.
- Determining if the CTNs of strains 630, R20291 and M120 can transfer via filter-matings.
- Determining the number of integration sites and the copy number for the CTNs in transconjugant cells.
- Determining if co-transfer of conjugative or non-conjugative elements has occurred.
- Determining if transfer is likely to occur by the recently described nanotubes in *B. subtilis*.

4.3. Materials and methods

4.3.1. *C. difficile* growth curve

Cultures of strains 630 and R20291 were grown for approximately 16 h in BHI broth and were used 1:10 (v/v) to set up a fresh culture of 20 ml at a starting OD₆₀₀ of 0.1. The OD₆₀₀ was measured every 60 min for 10 h until the cells were in stationary phase. When the OD₆₀₀ reached above 0.5, the sample was diluted 1:2 and the reading was multiplied accordingly.

4.3.2. Detection of CTn circularisation and excision from the chromosome

In order to amplify the joints between the genome-element junction, the circular forms and the empty target sites of CTns using PCR, four primers were designed for each element to act in 4 separate reactions as shown in Figure 4.1. Each product was sequenced in order to define the precise ends of each of the elements. The primers used for each of the elements are listed in Table 4.1.

Table 4.1 Oligonucleotides for Tn excision

Primer	Sequence 5' - 3'	Amplicon
CTn1 LEO	GGGCTTATCAAGCAATCCAA	0.7 kb
CTn1 REO	GAAATCTGGTCGGCACTCAT	
CTn1 TSF	GCAAAAAGTTGCAGGAATCAT	0.4 kb
CTn1 TSR	GAGAAAATGGCTCTGCTATGC	
CTn2 LEO	CGAAATTGATCCTGCAATGA	-
CTn2 LEO2	GTTCCAAGTCCGTTTTTCGCT	1.5 kb
CTn2 REO	GTTCCAAGTCCGTTTTTCGCT	
CTn2 TSF	ACCCTCCTAGACCAGGCATT	0.4 kb
CTn2 TSR	CAACATGAGAAACCAGACGGTA	
CTn3 LEO	CCACTTGATATGAAAAATCAAATGGCTC	0.3 kb
CTn3 REO	CCGTTCTATGCTGTGTGAGCTG	
CTn3 TSF	GGCGAAAAC T GCTTGGATT	0.4 kb
CTn3 TSR	CCTTTGTTTTCTACATCAAGTATAGCC	
CTn4 LEO	TGAACCCCAAAGCATTGCAG	0.7 kb

Primer	Sequence 5'- 3'	Amplicon
CTn4 REO	TGGTGGTTGCTCATAACCG	
CTn4 TSF	GGATAATGCAATAGAAGCATGTG	0.5 kb
CTn4 TSR	CATATCACCTTCCTATCCAATTCTT	
CTn5 LEO	CATCCCACCTGATAGGCAAG	1.3 kb
CTn5 REO	GATGCCGTTGCTGAGGTAAT	
CTn5 TSF	GCGCCATTGGAATGACTAT	0.6 kb
CTn5 TSR	CCAGTCCATGGTTGAAAAGTT	
CTn6 LEO	CCAAAGACGCAAGCCAGTAT	0.6 kb
CTn6 REO	CGGCAGGGAGGATAACTACA	
CTn6 TSF	CGCTCTTGCCCATATTTTCAG	0.8 kb
CTn6 TSR	CGCATCAGGAACATTGACTA	
CTn7 LEO	GACCCTGCACAGTTGACCTT	0.5 kb
CTn7 REO	CTTGTGAGCTTCGCCATACT	
CTn7 TSF	CCGAAATCGCTTTACAGTGG	0.3 kb
CTn7 TSR	CCCACCTAGAAAGGGATGTG	
R20291 CTn1 LEO	GTGAGGGATTGCCAAAGGTA	0.6 kb
R20291 CTn1 REO	CGGCACTCATAACGGTACTT	
R20291 CTn1 TSF	GATTTCGTTAAATGCCTCTTTATTG	0.6 kb
R20291 CTn1 TSR	GAGTGAAGAAGACTGGTTGACAGA	
R20291 Tn6103 LEO	CCAACCATTTGGTATTCCGTA	1.3 kb
R20291 Tn6103 REO	GATGCCGTTGCTGAGGTAAT	
R20291 Tn6103 TSF	GCGCCATTGGAATGACTAT	0.6 kb
R20291 Tn6103 TSR	CCAGTCCATGGTTGAAAAGTT	
R20291 Tn6104 LEO	GCTCATGTCCCTTGATGACGA	0.6 kb
R20291 Tn6104 REO	GCTCTGGTCGGACAAGGTAG	
R20291 Tn6104 TSF	GCTTGAAGATAACGGATATGAGATAA	0.7 kb
R20291 Tn6104 TSR	CTCTCTTGAACGGATTGTTGC	
R20291 Tn6105 LEO	GCTTTTCAGTTCCTCGGTGAC	0.6 kb
R20291 Tn6105 REO	GATGTTCCATACCGAGCAGG	
R20291 Tn6105 TSF	GTATAGGCGGACGATTACC	0.4 kb
R20291 Tn6105 TSR	CATTTTCCACCGTCCATTAC	
R20291 Tn6106 LEO	CTCACAATTCAGCGTTTCGC	0.5 kb
R20291 Tn6106 REO	CCATTCAGGAATACGCCAAT	
R20291 Tn6106 TSF	CTTTGAAAAAGGCGATTTGA	1.0 kb
R20291 Tn6106 TSR	GGTTGAAATAGCAACTACAGTTC	
23M63 CTn1 LEO	GAAGCCCGTACACAATCGTT	0.9 kb
23M63 CTn1 REO	GGGCTTATCAAGCAATCCAA	
23M63 CTn1 TSF	GGCTTCTTTAATTTGCTCTTTA	1.0 kb
23M63 CTn1 TSR	GGTTGTGTTTGCAGCTGTTA	
23M63 CTn5 LEO	CGCCTTTGGTAACCATTTGT	1.0 kb
23M63 CTn5 REO	GCATTGAAGAAAAGGCTTGAG	
23M63 CTn5 TSF	GGGAAGGCAGAGGAAATGT	0.8 kb
23M63 CTn5 TSR	GCTAAAGGAATCTTGGAAATTTGG	
66C26 CTn1 LEO	GTGAGGGATTGCCAAAGGTA	0.6 kb
66C26 CTn1 REO	CGGCACTCATAACGGTACTT	
66C26 CTn1 TSF	GATTTCGTTAAATGCCTCTTTATTG	0.6 kb
66C26 CTn1 TSR	GAGTGAAGAAGACTGGTTGACAGA	
66C26 CTn5 LEO	CGCCTTTGGTAACCATTTGT	1.3 kb
66C26 CTn5 REO	GCATTGAAGAAAAGGCTTGAG	
66C26 CTn5 TSF	GGGAAGGCAGAGGAAATGT	0.6 kb

Primer	Sequence 5'- 3'	Amplicon
66C26 CTn5 TSR	GCTAAAGGAATCTTGGAATTTGG	
32G58 CTn1 LEO	GTGAGGGATTGCCAAAGGTA	0.6 kb
32G58 CTn1 REO	CGGCACTCATACGGGTACTT	
32G58 CTn1 TSF	GATTCGTAAATGCCTCTTTATTG	0.6 kb
32G58 CTn1 TSR	GAGTGAAGAAGACTGGTTGACAGA	
32G58 CTn5 LEO	CCAACCATTGGTATTCCGTA	1.3 kb
32G58 CTn5 REO	GATGCCGTTGCTGAGGTAAT	
32G58 CTn5 TSF	GTCTGGAAATCACCTCCTCATT	0.6 kb
32G58 CTn5 TSR	TGGAAAGGCAGAGGAAATTG	
43255 CTn1 LEO	GTGAGCTTGCGTCTGTTCTG	1.2 kb
43255 CTn1 REO	CGTGAGCTTCCATTGACAGA	
43255 CTn1 TSF	GGAGAAGATGGGAAGAGTTCAA	1.0 kb
43255 CTn1 TSR	GAAACCAACAATAAGGGCG	
M120 Tn6164 LEO	TCTATCCTGCCTTCTCAACC	0.5 kb
M120 Tn6164 REO	GAGGAATTTAACAGAACAGTATTT	
M120 Tn6164 TSF	GTTTATTGCTGGGGATGTGG	0.5 kb
M120 Tn6164 TSR	CATTGATGAAGTACCACCTATTGC	
M120 Phage REO	GGAAATGCCAAGGGACAATA	0.5 kb
M120 Phage TSR	TATGGAGATGTTGGCAGTGG	
M120 Phage REO2	CGAATCGCTGAAATGACTGA	0.5 kb
M120 Phage TSR2	GGCAGTAACAATTCCGTGCT	

4.3.3. ClosTron mutagenesis

In this study, several mutants were generated using the ClosTron system. The original system consisted of designing a retargeting sequence and cloning this sequence into the shuttle vector pMTL007 [241]. After the plasmid has been transferred into the target *C. difficile* strain, expression of the gene encoded on the group II intron is induced by IPTG.

For the refined ClosTron protocol, the plasmids can be ordered commercially. Furthermore, the group II intron is expressed constitutively, making the IPTG induction unnecessary [246].

4.3.3.1. Clostron mutagenesis with original protocol

Clostron mutagenesis was performed as described by Heap *et al.* [241]. Suitable Clostron target sites were identified using the TargetTron algorithm at the TargetTron design site for which access was gained through the TargetTron Gene Knockout System kit (Sigma-Aldrich Ltd). Primers were designed by the algorithm and synthesized by Sigma-Aldrich Ltd, primers IBS, EBS1d and EBS2 are target site specific, primer EBS universal is common to all PCRs for intron retargeting (Table 4.2). The Clostron plasmids were constructed as described by Heap *et al.* [241] and sequenced using primers 5402F-F1 and pMTL007-R1 to confirm the correct sequence for intron retargeting. The plasmids were conjugated into *C. difficile* 630 Δ *erm* and presence of the plasmid was checked by performing a colony PCR on the sub-cultured plates using the ErmRAM-F and ErmRAM-R primers (Table 4.2). Colonies that were shown to contain the non-spliced ErmRAM product, a 1.3 kb PCR product, were used for the IPTG induction. After this, mutants were selected on BHIB plates containing 40 μ g/ml lincomycin and *C. difficile* selective supplement instead of 25 μ g/ml erythromycin. Colonies were sub-cultured onto fresh selective plates and checked by PCR with the ErmRAM-F and ErmRAM-R primers to contain the integrated intron with an expected size of 0.9 kb. After this, PCR was performed for the specific target site using ErmRAM-F and Target-R as well as Target-F and ErmRAM-R (Table 4.2) (Figure 4.4). PCR products for these PCR reactions were all sequenced to confirm the intron is integrated in the correct target site.

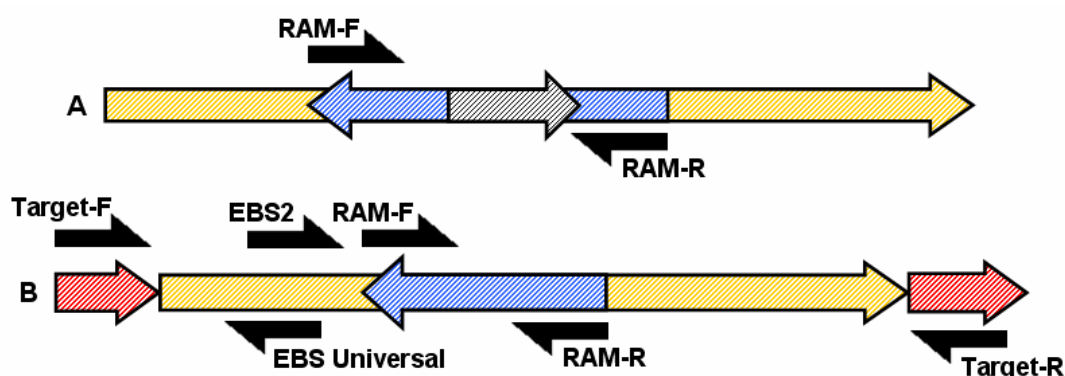


Figure 4.4 PCR screening for intron integration.

The intron on the plasmid is represented by the upper yellow arrow. The intron integrated in the target site is shown by the lower yellow arrow. The red arrows show the interrupted target site. The blue arrows represent the ErmRAM, which in the upper schematic is interrupted by the group I intron, shown as a grey arrow. PCR primers are represented by black half-arrows. Reproduced from the protocol for clostridial Gene Knockout using pMTL007 [245].

Table 4.2 Oligonucleotides for Clostron mutagenesis

As numerous primer combinations were used, the expected amplicon sizes are not shown.

Primer	Sequence 5'-3'
EBS universal	CGAAATTAGAACTTGCCTTCAGTAAAC
IBS CD0364	AAAAAAGCTTATAATTATCCTTATGACACAGAAA TGTGCGCCAGATAGGGTG
EBS1d CD0364	CAGATTGTACAAATGTGGTGATAACAGATAAGTC AGAAATGTTAACTTACCTTTCTTTGT
IBS CD0386/CD3392	AAAAAAGCTTATAATTATCCTTAGTCATCCATTA CGTGCGCCAGATAGGGTG
EBS1d CD0386/CD3392	CAGATTGTACAAATGTGGTGATAACAGATAAGTC CATTACTTTAACTTACCTTTCTTTGT
EBS2 CD0386/CD3392	TGAACGCAAGTTTCTAATTTTCGATTATGACTCGA TAGAGGAAAGTGTCT
IBS CD1873	AAAAAAGCTTATAATTATCCTTATAAACATTTA CGTGCGCCAGATAGGGTG
EBS1d CD1873	CAGATTGTACAAATGTGGTGATAACAGATAAGTC ATTTACCATAACTTACCTTTCTTTGT
EBS2 CD1873	TGAACGCAAGTTTCTAATTTTCGTTTTTTTATCGA TAGAGGAAAGTGTCT
pMTL007-R1	AGGGTATCCCCAGTTAGTGTAAAGTCTTGG
5402F-F1	TTAAGGAGGTGTATTTTCATATGACCATGATTACG
CD0364 TSF	GATTAAAGCCCTCGCACAAG
CD0364 TSR	CGTAACCAGCGGACACATTA
CD0386/CD3392 TSF	CCCAAATGAAAGGAGCATT
CD0386/CD3392 TSR	GCGTCCGCTAATTTTGTGAT
CD1873 TSF	CAATAATGAAATCAAGCAATGAAA
CD1873 TSR	GCTTCGAGTGAAACAACATTCT
CD0428 TSF	CATGGAATCTTCGCATCAGA
CD0428 TSR	GCGAATGTCCTCACTAACCG
CD1099 TSF	GTCCATACCGGCATCATAA
CD1099 TSR	CAGAAAGGACGGGTGTTGTT

Primer	Sequence 5'-3'
ErmRAM-F	ACGCGTTATATTGATAAAAATAATAATAGTGGG
ErmRAM-R	ACGCGTGCGACTCATAGAATTATTTCCCTCCCG

4.3.3.2. ClosTron mutagenesis 2.0 with refined protocol

The intron target site and the intron redirection sequence were determined using the Intron targeting and design tool as described by Heap *et al.* [246]. The intron redirection sequence was ordered to be cloned into vector pMTL007C-E5 (DNA2.0, Menlo Park, USA).

ClosTron mutants were constructed as described by Heap *et al.* [246] with the only modification that mutants were selected on BHIB agar plates containing 40 µg/ml lincomycin. PCR was used to determine if the ErmRAM had spliced and to determine if the intron had inserted in the desired target site as described above.

4.3.4. Filter-matings

Filter-matings were performed as described in Chapter 2. ClosTron mutants were used as donor strains for CTn1 (630Δ*erm* CD0364::ErmRAM and 630Δ*erm* CD0386::ErmRAM), CTn2 (630Δ*erm* CD0428::ErmRAM), CTn4 (630Δ*erm* CD1099::ErmRAM), CTn5 (630Δ*erm* CD1873::ErmRAM), CTn7 (630Δ*erm* CD3392::ErmRAM) and Tn6103 (R20291 1803::ErmRAM). CD37 and *B. subtilis* CU2189 were used as recipient strains. The filter-matings with R20291 1803::ErmRAM as donor were performed using *C. difficile* 630Δ*erm*, CD37 and *B. subtilis* CU2189 as recipients. Strain M120 was used as donor for the elements Tn6164 and Tn6190 where CD37 was used as recipient strain.

Serial dilutions of the mixture of donor, recipient and transconjugant cells from the filters were set up to determine the cell counts for donor and recipient cells. Donor cells of the elements containing the Clostron insertion were selected on lincomycin 40 µg/ml plates. M120 donor cells were selected on tetracycline 10 µg/ml plates. CD37 recipient cells were selected on rifampicin 25 µg/ml plates. *B. subtilis* CU2189 recipient cells were selected on antibiotic free plates incubated in aerobic conditions. 630Δ*erm* recipient cells were selected on tetracycline 10 µg/ml plates.

CD37 transconjugant cells were selected on lincomycin 40 µg/ml and rifampicin 25 µg/ml plates or tetracycline 10 µg/ml and rifampicin 25 µg/ml plates. *B. subtilis* CU2189 recipient cells were selected on lincomycin 40 µg/ml plates incubated in aerobic conditions. 630Δ*erm* recipient cells were selected on lincomycin 40 µg/ml and tetracycline 10 µg/ml plates.

Identity of the transconjugants was confirmed by two sets of PCRs, one using a primer pair that is specific for the recipient strain and using a primer pair that is specific for the CTn that was transferred (Table 4.3, Table 4.4). For elements that contain *erm*(B) introduced by the Clostron system, the ErmRAM primers were used. The 16S PCR product of *B. subtilis* was sequenced and checked using BLASTn.

Table 4.3 Primers for the detection of CTNs in transconjugant cells.

Element	Primer	Sequence 5'-3'	Amplicon
Clostron	ErmRAM-F	ACGCGTTATATTGATAAAAAATAATAATA GTGGG	0.9 kb
	ErmRAM-R	ACGCGTGCGACTCATAGAATTATTTTCCT CCCG	
Tn6164	Tn6164 Acc reg-F	CAGCTGCAGTTTTTCCATGA	0.5 kb
	Tn6164 Acc reg-R	GCAGCTAACGGTGATGACAA	
Tn6190	Tn916Int-F	GACGGAAGATACTTATACA	1 kb
	Tn916Int-R	GCCTTTGGATTTCATTCCTGC	

Table 4.4 Primers for the identification of the transconjugants.

CD37 is negative for the Binary toxin PCR but positive for the Lok PCR that amplifies the region in which the PaLoc has integrated in toxigenic strains (Braun *et al.* 1996). The 16S primers were previously described in (Weisburg *et al.* 1991). The CD0242 primer pair partially amplifies CD0242 in strains 630 and 630 Δ *erm*, a CDS not present in R20291. Y=(C/T)

Strain	Primer	Sequence 5'-3'	Amplicon
CD37	Binary toxin-F	YAATACTACTTACAAGGCTCC	0.4 kb
	Binary toxin-R	TTTCGTTTTGATTTTTYTGTTC	
	Lok1	AAAATATACTGCACATCTGTATAC	0.6 kb
Lok3	TTTACCAGAAAAAGTAGCTTTAA		
630 Δ <i>erm</i>	CD0242-F	GTAAACTGTGCAGGTAGAGGAAG	0.6 kb
	CD0242-R	CCTCAATTCAACATGAGCTAT	
<i>B. subtilis</i>	16S F	AGAGTTTGATCMTGGCTCAG	1.4 kb
	16S R	TACGGYTACCTTGTTACGACTT	

4.3.5. Southern blotting

Southern blots were carried out using the Amersham ECL Direct Nucleic Acid Labelling and Detection System (GE Healthcare, Bucks, UK) according to the manufacturer's instructions.

DNA was digested overnight using HindIII at 37 °C. The DNA sample was loaded onto a 0.9% agarose gel and run overnight at low voltage (20V). The DNA was blotted onto a Hybond N+ nylon membrane (Amersham, UK) by capillary blotting. The DNA was UV-fixed to the nylon membrane using a UV Stratalinker 1800 (Stratagene, Cheshire, UK). The membrane was pre-hybridized in hybridization buffer for 1-3 h before the labeled probe was added.

Probes were prepared by standard PCR amplification and gel extraction as described in Chapter 2. Probes were constructed using primers listed in Table 4.5. The probe and DNA marker were labeled in separate reactions and both were mixed into the hybridization buffer for hybridization.

After hybridization, the membrane was washed and 10 ml detection reagent was spread onto the membrane. After 1 min incubation the reagent was drained on a piece of filter paper and wrapped in plastic foil in a film cassette. In a dark room, a sheet of Hyperfilm ECL (Amersham, UK) was exposed in the film cassette for initially 1 minute. Depending on the signal that was visible on this film a second sheet of film was exposed for 1 to 16 h. Films were developed using an AGFA CURIX 260 (AGFA, UK).

Table 4.5 Primers used for the Southern blot probe construction.

Primer	Sequence 5'-3'	Amplicon
CD0355 958F	GGTATGGTGCAGGACTGAT	0.2 kb
CD0355 1184R	GCCAGTCTTTCCAGTTCTGC	
CD0436 48F	AACTGCATCAGGCTGTGTTG	0.4 kb
CD0436 443R	GTCAAAGCAATTCGCCATC	
CD0511 1215F	CGAGCGGAAGTCAGAAGATT	0.2 kb
CD0511 1399R	GCTTTGCACTACCTCTGCTTG	
CD1091 691F	GGATTGACGGAAGCTGACAT	0.2 kb
CD1091 928R	CGTTCCGGTTAAGGAACAGA	
CD1878 904F	GGCAAGAACATGAGGACACA	0.2 kb
CD1878 1122R	CTCAGCAACGGCATCATCTA	
CD3370 447F	TTCTGTTGCACAGCAGGAAG	0.2 kb
CD3370 647R	CCAGCACCATCAACATAACG	

4.3.6. PCR analysis for detection of co-transfer

To perform the analysis of co-transfer and to confirm the identity of each of the transconjugants, PCR analysis was performed on each sample with a number of primer pairs, all listed in Table 4.6.

Table 4.6 PCR primers for CTn detection.

Primer pair CD0364 F+R is for the detection of the presence of CTn1. Primer pair CD0428 F+R is for the detection of the presence of CTn2. Primer pair CD1873 F+R is for the detection of the presence of CTn5. Primer pair CD0511 F+R is for the detection of the presence of Tn5397. Primer pair 9371 F + 9387 R is for the detection of the presence of Tn5398. Primer pair *tcdB* 3120 F + 3521 R is for the detection of the presence of *tcdB* encoded on the PaLoc. Primer pair Lok1 + Lok3 is for the detection of the empty target site of the PaLoc. Primer pair *cdtA* F+R is for the detection of the presence of *cdtA* encoded on the binary toxin locus.

Primer	Sequence 5'- 3'	Amplicon
CTn1 CD0364 F	GATTAAAGCCCTCGCACAAG	0.3 kb
CTn1 CD0364 R	GCGTCCGCTAATTTTGTGAT	
CTn2 CD0428 F	CATGGAATCTTTCGCATCAGA	0.4 kb
CTn2 CD0428 R	GCGAATGTCCTCACTAACCG	
CTn5 CD1873 F	CAATAATGAAATCAAGCAATGAAA	0.2 kb
CTn5 CD1873 R	GCTTCGAGTGAAACAACATTCT	
Tn5397 CD0511 F	CGAGCGGAAGTCAGAAGATT	0.2 kb
Tn5397 CD0511 R	GCTTTGCACTACCTCTGCTTG	
Tn5398 9371F	GGATAGAAATACTCGTCAACAG	0.9 kb
Tn5398 9387 R	ATTTTTTATTTTTTAGGAGTCATA	
<i>tcdB</i> 3120 F	AGGTGCAGCAATCAAAGAGC	0.4 kb
<i>tcdB</i> 3521 R	CCACCTTCCATTCTCCAGAT	
Lok1	AAAATATACTGCACATCTGTATAC	0.6 kb
Lok3	TTTACCAGAAAAAGTAGCTTTAA	
<i>cdtA</i> F	YAATACTACTTACAAGGCTCC	0.3 kb
<i>cdtA</i> R	TTTCGTTTTGATTTTYTGTTTC	

4.3.7. Nanotube transfer of antimicrobial resistance proteins

CD37, R20291, CD37 pMTL960/GusA and R20291 pMTL9301 were grown for 16 h (section 2.1). CD37 pMTL960/GusA and R20291 pMTL9301 were grown in broth supplemented with 30 µg/ml thiamphenicol 40 µg/ml lincomycin, respectively. 30 µl of each culture was spotted on an antibiotic free BHIB agar plate, as well as a mixture of equal volumes of CD37 pMTL960/GusA and R20291 pMTL9301. After 4 h of incubation, the cells were replica plated onto BHIB agar plates supplemented with 30 µg/ml thiamphenicol, 40 µg/ml lincomycin, 30 µg/ml thiamphenicol and 40 µg/ml lincomycin or antibiotic free. Plates were incubated for 24 h at 37°C before growth was assessed.

4.4. Results

4.4.1. Circularisation of CTNs in different growth phases

A growth curve was performed to determine the growth phases of *C. difficile* (Figure 4.5.a). Genomic DNA samples of strains 630 and R20291 were isolated from cells in lag, exponential and stationary phase, respectively after 2 h, 5 h or 10 h of growth. These samples were used as template in PCR reactions to determine when the CTNs excise. CTn1, CTn2, CTn4, CTn5, CTn7 and Tn5397 excised in all of these growth phases whereas excision of CTn6 could not be detected. An example is shown for CTn1 (Figure 4.5.b + c). The template DNA of strain R20291 was used as a control; the homologue of the CTn1 target site in R20291 is empty and was expected to generate a similar product as seen for the target site of CTn1 in 630. Primers for the product of the joint of the circular intermediate are the LEO + REO primers, the joint of the empty target site is amplified by the TSF + TSR primers (Table 4.1).

Sebahia *et al.* [8] have predicted the size and the ends of the putative CTNs based on the 630 genomic sequence. The primers used here were designed using these predictions. For CTn2, the PCR with the CTn2 LEO + REO amplifying the circular intermediate did not produce any product. PCR for the regenerated target site with CTn2 TSF + TSR was larger than expected. Sequencing of this product showed that the region CD0404 - CD0406 was present in the regenerated target site. Primer CTn2 LEO2 was designed further into the left side of CTn2 that resulted in a product of the circular intermediate of CTn2. This confirms that the region CD0404 - CD0406 is not part of the element (Figure 4.6).

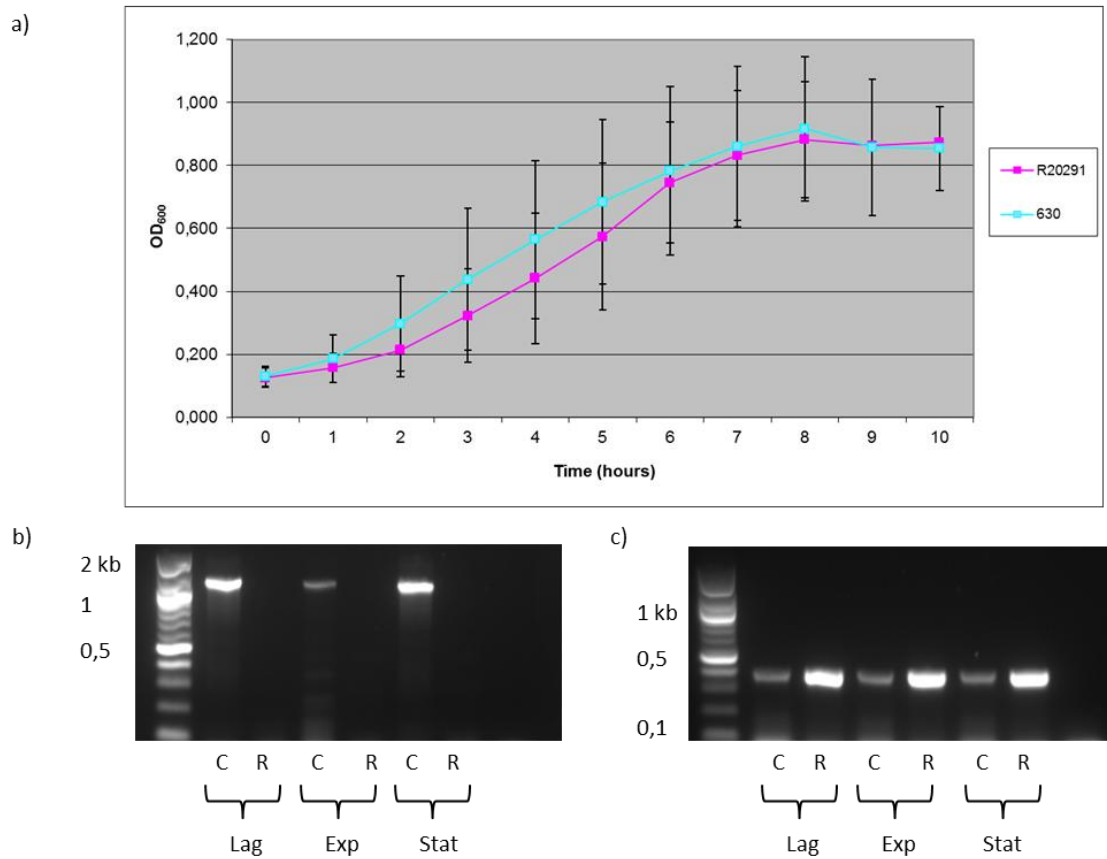


Figure 4.5 Excision of CTn1 in different growth phases.

Excision of CTn1 from the genome into a circular intermediate was shown in lag, exponential and stationary growth phases. a) Growth curve for *C. difficile* strains to determine the growth phases. DNA samples were extracted for stationary phase after 1 h growth, for exponential phase after 4 h growth and for stationary phase after 8 h growth. b) PCR product of the joint of the circular intermediate generated with primers CTn1 LEO + REO. c) PCR product of the regenerated target site generated with primers CTn1 TSF + TSR. The template for each reaction is shown below the lanes. C is strain 630, R is strain R20291. Lag is lag phase, Exp is exponential phase, Stat is stationary phase. The left lane in b) and c) is the molecular weight marker.

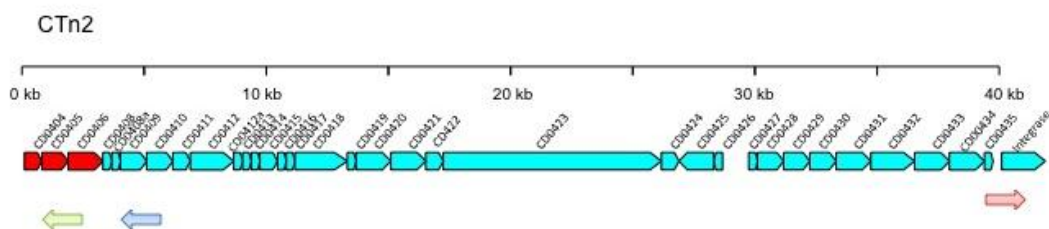


Figure 4.6 ORFs of CTn2

ORFs shown in blue are part of CTn2 and are present in the circular form of the element. ORFs shown in red were previously predicted to be part of the element but were shown to remain in the regenerated target site in the genome after the element excised. The green arrow represents primer CTn2 LEO that could not produce a PCR product of the circular molecule with primer CTn2 REO shown in red. The blue arrow represents primer CTn2 LEO2 that amplified the circular intermediate of CTn2 with CTn2 REO.

4.4.2. Circularisation of *C. difficile* CTns

PCR was carried out using primers to amplify the junctions of the circular intermediates of the CTns from strains 630, R20291, M120, QCD-23M63, QCD-66C26, QCD-32G58 and ATCC-43255. These PCRs were performed with the Left End Out and Right End Out (LEO and REO) primers (Table 4.1). In addition, the regenerated target site in the chromosome was amplified with the Target Site Forward and Target Site Reverse (TSF and TSR) primers (Table 4.1). The element-chromosome junctions were amplified with primer pairs LEO + TSF and REO + TSR. Products of the expected size were obtained for CTn1, CTn2, CTn4, CTn5 and CTn7 from strain 630, Tn6103 from R20291, Tn6073 and Tn6107 from QCD-23M63, Tn6115 from QCD-63Q42, Tn6110 from QCD-66C26, Tn6111 from QCD-32G58 and Tn6164 from M120. An example for this is shown for CTn5 in Figure 4.7. These PCR products were sequenced to determine the direct repeats on either side of the element (Figure 4.8).

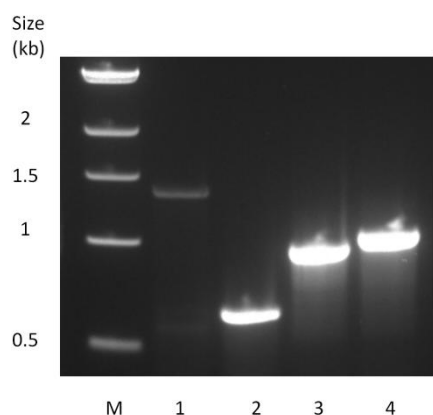


Figure 4.7 PCR products to show excision of CTn5 from the chromosome.

PCRs were carried out for the joints of the circular molecule, the regenerated target site and the ends of the transposon in the chromosome. Lane 1 is product of the joint of the circular intermediate of CTn5 using primers CTn5 LEO+REO. Lane 2 is product of the regenerated target site within the chromosome using primers CTn5 TSF+TSR. Lane 3 and 4, respectively, are products of the left and right junctions of the element and the chromosome using primer pairs CTn5 TSF+LEO and TSR+REO. Template in all reactions was DNA of 630 isolated from the exponential growth phase. Lane M is a molecular weight marker. Size of the fragments is shown left of the figure.

CTn1, CTn4 and Tn6073 all encode predicted tyrosine recombinases and an excisionase which are predicted to be responsible for the excision and integration of these elements. CTn2, CTn5, CTn7, Tn5397, Tn6103, Tn6104, Tn6107, Tn6110, Tn6111 and Tn6164 all encode predicted serine recombinases which are probably responsible for the excision and integration of these elements.

- a) CTn1 630 Joint of circular form
 CTATTAGTCTAAAAAATAACACAAAAAGAAA
 GATAATCAGATTTTTTATTGTGTTTTCTTT
 Left end in chromosome
 GCATCCAAATTTAAATAACACAAAAAGAAA
 CGTAGGTTTAAATTTATTGTGTTTTCTTT
 Right end in chromosome
 CTATTAGTCTAAAAAATAAATAGAAAAATCG
 GATAATCAGATTTTTTATTTATCTTTTAGC
 Regenerated target site
 GCATCCAAATTTAAATAAATAGAAAAATCG
 CGTAGGTTTAAATTTATTATCTTTTAGC
- b) CTn2 630 Joint of circular form
 TCCTATTGACAACACATGTGGAGTGATAGCG
 AGGATAACTGTTGTGTACACCTCACGTATCGC
 Left end in chromosome
 TCCCAAAATACGCCACATGTGGAGTGATAGCG
 AGGGTTTTATGCGGTGTACACCTCACGTATCGC
 Right end in chromosome
 TCCTATTGACAACACATGTGTGAGACTGTAGTTA
 AGGATAACTGTTGTGTAACACCTCTGACATCAAT
 Regenerated target site
 TCCCAAAATACGCCACATGTGTGAGACTGTAGTTA
 AGGGTTTTATGCGGTGTAACACCTCTGACATCAAT
- c) CTn7 630 Joint of circular form
 ATAGTAGTAACGACACATGTGGAGACTGTGTGTTGCTA
 TATCATCATGCTGTGTACACCTCTGACACACACGAT
 Left end in chromosome
 CCCTCATCTTACACACATGTGGAAACTGTGTGTTGCTA
 GGGAGTAAGATGTGTGTACACCTTTGACACACACGAT
 Right end in chromosome
 ATAGTAGTAACGACACATGTGGAGACTGTAGCTAAGATA
 TATCATCATGCTGTGTACACCTCTGACATCGATTCTAT
 Regenerated target site
 CCCTCATCTTACACACATGTGGAAACTGTAGCTAAGATA
 GGGAGTAAGATGTGTGTACACCTTTGACATCGATTCTAT
- d) Tn6073 QCD-23M63 Joint of circular form
 AGTTTTTATAGTTAAAAACACAATAAAAAAT
 TCAAAAATATCAATTTTTTGTGTATTTTTA
 Left end in chromosome
 TAGATTTTAAATATAAAAAACACAATAAAAAAT
 ATCTAAAAATATATTTTTTGTGTATTTTTA
 Right end in chromosome
 AGTTTTTATAGTTAAAAAGTAAAATTTAAAT
 TCAAAAATATCAATTTTTTCATTTAAATTTAA
 Regenerated target site
 TAGATTTTAAATATAAAAAAGTAAAATTTAAAT
 ATCTAAAAATATATTTTTTCATTTAAATTTAA
- e) Tn6103 R20291 Joint of circular form
 TTACTACTCGTTCAACTCTGCCATCGCAA
 AATGATGAGCAAGTTGAGACGGTAGCGTT
 Left end in chromosome
 CCTTGACATTGCCAACTCTGCCATCGCAA
 GGAACGTAAACGGTTGAGACGGTAGCGTT
 Right end in chromosome
 TTACTACTCGTTCAACTACCTCATAGACAT
 AATGATGAGCAAGTTGATGGAGTATCTGTA
 Regenerated target site
 CCTTGACATTGCCAACTACCTCATAGACAT
 GGAACGTAAACGGTTGATGGAGTATCTGTA
- f) Tn6104 R20291 Joint of circular form
 CAATTTGCTCTGGTGATATATGCTC
 GTTAAACAGAGACCCTATATACGAG
 Left end in chromosome
 AAATTTGGCTGCGGGTGATATATGCTC
 TTTAACCGACGCCACTATATACGAG
 Right end in chromosome
 CAATTTGCTCTGGAAAACTGCCAA
 GTTAAACAGAGACCCTTTTGGACGGTT
 Regenerated target site
 AAATTTGGCTGCGGGAAAACTGCCAA
 TTTAACCGACGCCCTTTTGGACGGTT
- g) Tn6107 QCD-23M63 Joint of circular form
 TCCTATGGACAAACACATGTGAGACCATAGCG
 AGGATACCTGTTGTGTACAGCTCTGTATCGC
 Left end in chromosome
 TCCTCATCTACTCATGTGAGACCATAGCG
 AGGGAGTAAGATGAGTACAGCTCTGTATCGC
 Right end in chromosome
 TCCTATGGACAAACACATGTGAAACTGTAGCT
 AGGATACCTGTTGTGTACACTTTGACATCGA
 Regenerated target site
 TCCTCATCTACTCATGTGAAACTGTAGCT
 AGGGAGTAAGATGAGTACACTTTGACATCGA
- h) Tn6110 QCD-66C26 Joint of circular form
 AATGATGAGCAAGTTGAGACGGTAGCGTT
 TTACTACTCGTTCAACTCTGCCATCGCAA
 Left end in chromosome
 TCATTCTACACATGTTGAGACGGTAGCGTT
 AGTAAGATGTGTACAACTCTGCCATCGCAA
 Right end in chromosome
 AATGATGAGCAAGTTGACGAGTAGTAA
 TTACTACTCGTTCAACTTGCTCATCATT
 Regenerated target site
 TCATTCTACACATGTTGAAACGAGTAGTAA
 AGTAAGATGTGTACAACTTGCTCATCATT
- i) Tn6111 QCD-32G58 Joint of circular form
 AATGATGAGCAAGTTGAGACGGTAGCGTT
 TTACTACTCGTTCAACTCTGCCATCGCAA
 Left end in chromosome
 CATTCTACACATGTGGAAGACGGTAGCGTT
 GTAAGATGTGTACACCTCTGCCATCGCAA
 Right end in chromosome
 AATGATGAGCAAGTTGAAACTGTAGCTAA
 TTACTACTCGTTCAACTTTGACATCGATT
 Regenerated target site
 CATTCTACACATGTTGAAACTGTAGCTAA
 GTAAGATGTGTACAACTTTGACATCGATT
- j) Tn6164 M120 Joint of circular intermediate
 GTGTTCTGTGCGTACACCTCACGCA
 CACAAGACAACGCATGTGGAGTGCGT
 Left end in the chromosome
 GGCTTATGCGGAGTACACCTCACGCA
 CCGAATACGCCTCATGTGGAGTGCGT
 Right end in the chromosome
 GTTTTCTGTTGCGTAACGCTCTGATA
 CAAAAGACAACGCATGCGAGACTAT
 Regenerated target site
 GGCTTATGCGGAGTAACGCTCTGATA
 CCGAATACGCCTCATGCGAGACTAT

Figure 4.8 Sequences of transposon-chromosome junctions and direct repeat sequences

Sequences of the PCR products of the joint of the circular molecule, the joint of the chromosome-CTn junction and the regenerated target site of each of the CTns which were shown to excise from the chromosome. Bases in red are part of the CTn, bases in blue are part of the chromosome, and bases in black and underlined are the direct repeat sequences at the ends of the elements.

For the putative mobilisable transposon Tn6104, located on Tn6103 in R20291, only the product for the circular intermediate of the element was amplified (Figure 4.9). Amplification of the empty target site could not be shown. For the two other putative mobilisable transposons within Tn6103, Tn6105 and Tn6106, only products for the joints of the elements in the chromosome could be amplified. This suggests that these elements did not excise from the chromosome in the bacterial population tested.

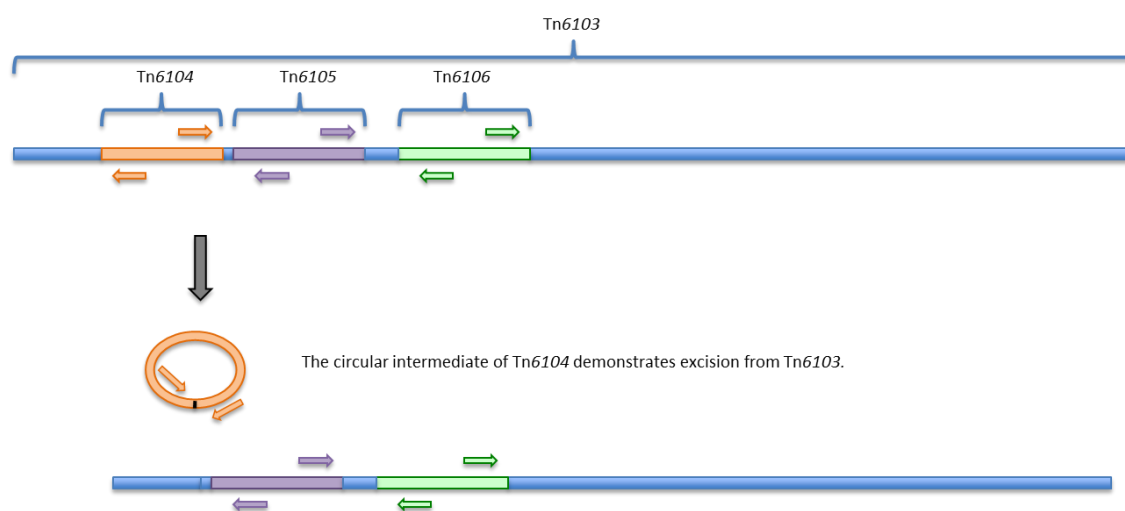


Figure 4.9 Excision of Tn6104.

PCR amplification of the joint of the circular intermediate of the putative mobilisable transposon Tn6104 demonstrates the element excises from Tn6103. Tn6103 is shown in blue, Tn6104 is shown in orange, Tn6105 is shown in purple, Tn6106 is shown in green. The elements are not drawn to scale.

The joint of the empty target site or the circular intermediate could not be amplified for the CTn1-like elements in strains R20291, QCD-66C26 and QCD-32G58, indicating these elements did not excise from the chromosome in the tested bacterial populations. Strain ATCC-43255 contains a CTn1-like element which has integrated in *skin*^{Cd}, see Chapter 3. Although circularization of the CTn1-like element could not be shown here, circularization of the *skin*^{Cd}, which includes the CTn1-like element was previously shown in this strain [179].

4.4.3. ClosTron mutagenesis

Using the original ClosTron protocol [241], the intron was retargeted to two locations in CTn1 in separate isolates of 630 Δ erm. The first location was in CD0364, an ABC transporter ATP binding protein, the second in CD0386, a putative cell surface protein. CTn7 has a homologue of this cell surface protein, CD3392, which could also be targeted by the same ClosTron construct (Figure 4.10). Sequencing of the intron-target site junctions of six mutants demonstrated that in half of the mutants, the intron was inserted in CD0386 and in half of the mutants the intron was inserted in CD3392.

The ClosTron 2.0 protocol [246] was used with 630 Δ erm to retarget the intron to CTn2 CD0428, a predicted AraC transcriptional regulator. For CTn4, the intron was retargeted to CD1099, a predicted two component response regulator. For CTn5, the intron was retargeted to CD1873, a predicted ABC transporter permease protein (Figure 4.11).

Strain R20291 contains Tn6103, which is 85% identical to CTn5. The intron was retargeted in this strain using the same construct used for CD1873 on CTn5. Sequencing of the intron-genome junction confirmed the ClosTron had inserted in ORF 1803, the homologue of CD1873 on CTn5 (Figure 4.12).

PCR was used to confirm that the introduction of the intron did not prevent excision of the elements from the genome.

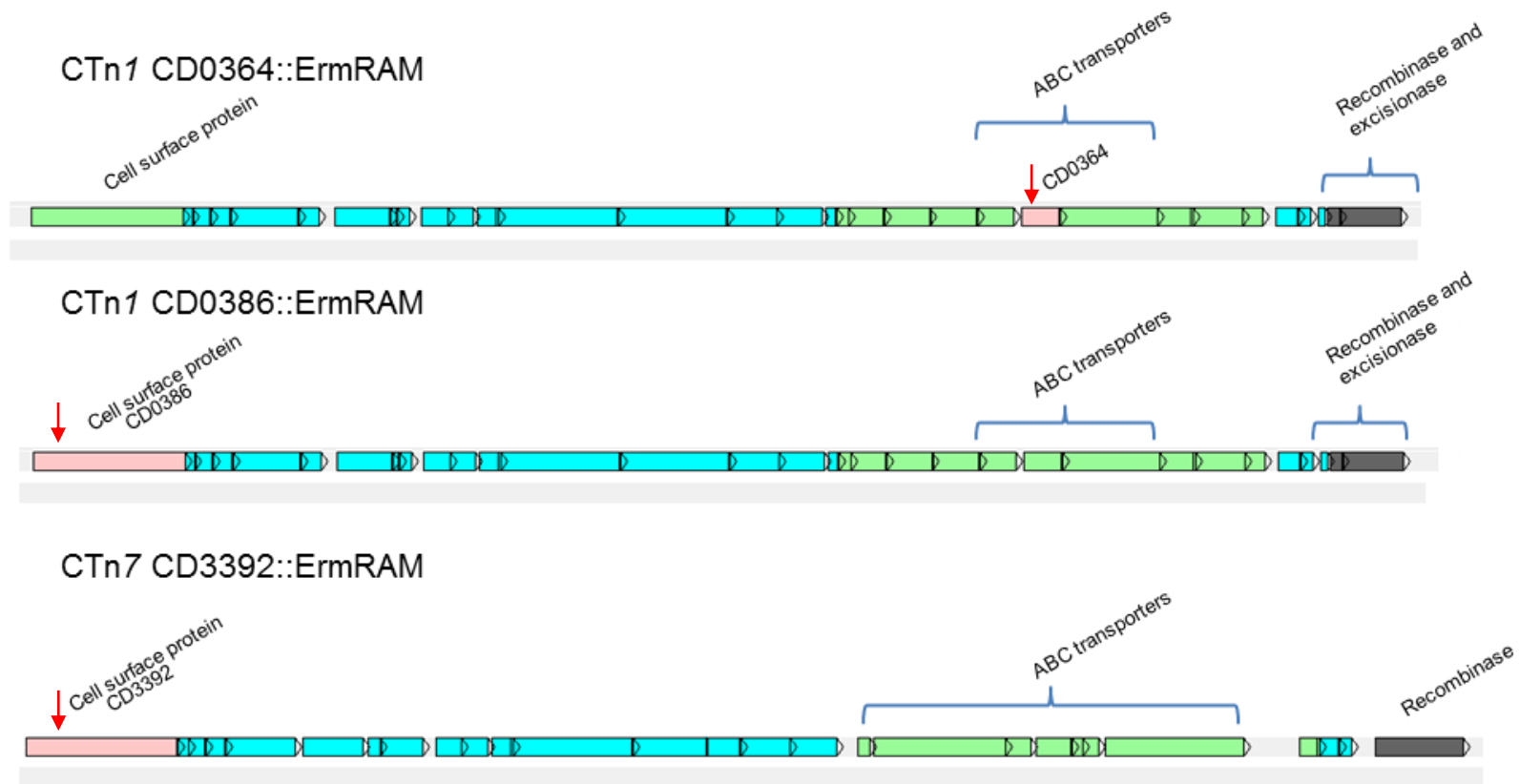


Figure 4.10 Clostron insertions in CTn1 and CTn7.

ORFs that were shown to contain the Clostron insertion are shown in pink, the integration sites of the Clostron constructs are indicated by red arrows. ORFs in the recombination module are shown in grey. ORFs on the conjugation module are shown in blue. ORFs on the accessory module are shown in green.

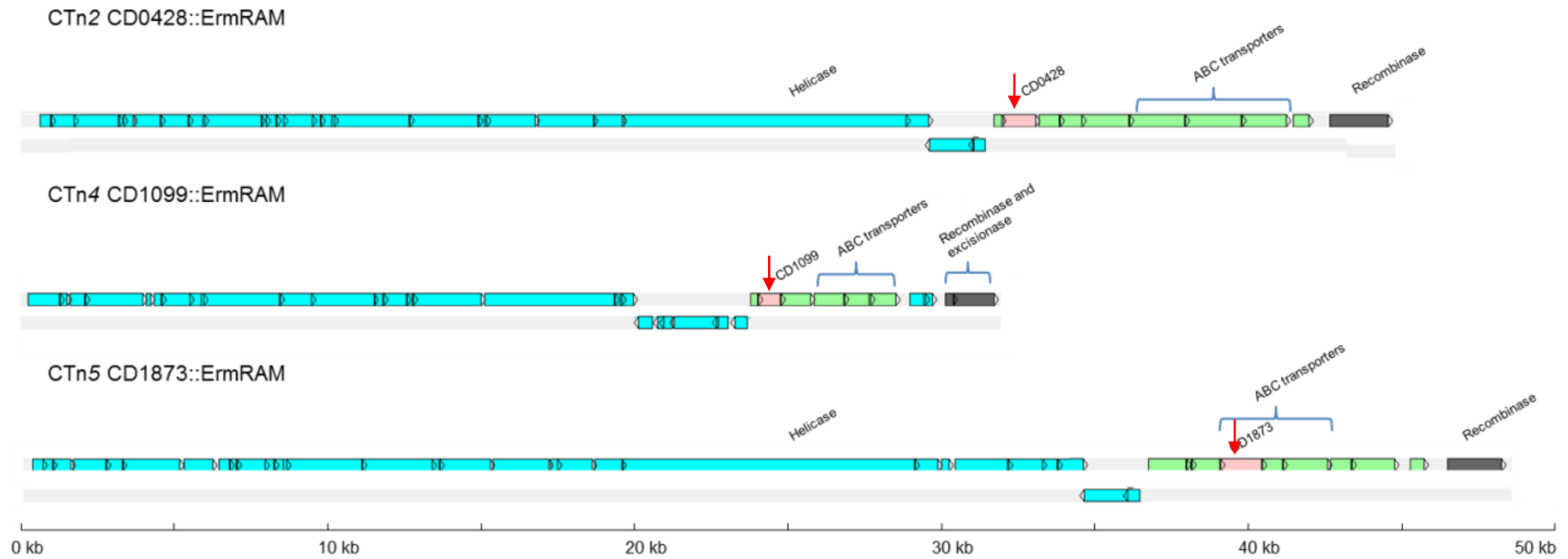


Figure 4.11 Clostron insertions in CTn2, CTn4 and CTn5.

ORFs that were shown to contain the Clostron insertion are shown in pink, the integration sites of the Clostron constructs are indicated by red arrows. ORFs in the recombination module are shown in grey. ORFs in the conjugation module are shown in blue. ORFs in the accessory module are shown in green.

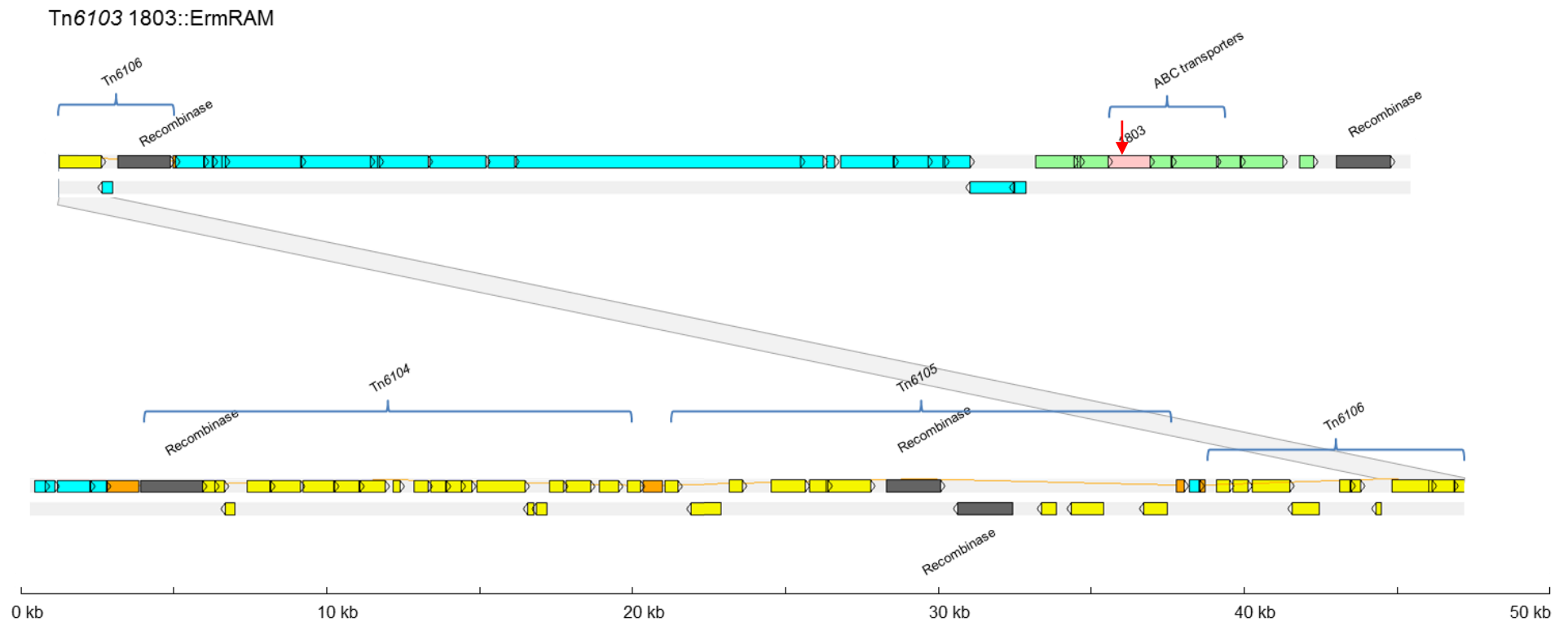


Figure 4.12 Clostron insertion in Tn6103.

The schematic of Tn6103 is shown as two parts which should be joined at the grey diagonal line. ORF 1803 that was shown to contain the Clostron insertion is shown in pink, the integration site of the Clostron construct is indicated by a red arrow. ORFs on the recombination module of Tn6103 and the recombinases of Tn6104, Tn6105 and Tn6106 are shown in grey. ORFs on the conjugation module of Tn6103 are shown in blue. ORFs on the accessory module are shown in green. ORFs that were interrupted by an insertion of a DNA fragment, compared to CTn5, are shown in orange. ORFs of the putative mobilisable elements Tn6104, Tn6105 and Tn6106 are shown in yellow.

4.4.4. Conjugative transfer of CTns and analysis of transconjugants

4.4.4.1. Conjugative transfer of CTn1 can occur with co-transfer of CTn5 and Tn5397

Conjugative transfer of CTn1 was carried out using the donor strains 630 Δ erm containing *erm*(B) in the ABC-transporter gene (CD0364) or in the cell surface gene (CD0386) (Figure 4.10) using *C. difficile* CD37 and *B. subtilis* CU2189 as recipients. Transconjugants were only detected where CD37 was recipient. Donor, recipient and transconjugant cell numbers were determined and the transfer frequency was calculated as the number of transconjugant cells per donor cell (Table 4.7). Putative transconjugants were sub-cultured onto fresh selective plates and screened by PCR to confirm the CTn is present by amplifying the ErmRAM.

Table 4.7 Transfer frequencies of CTn1.

The transfer frequency was calculated per filter as the number of transconjugants that resulted from a filter-mating, divided by the number of donor cells present on the filter. Each filter-mating experiment consisted of four technical replicates and each experiment was repeated at least three times. The mean of the transfer frequency and standard deviation were calculated from the frequencies of each independent filter. The detection limit for the filter-matings is approximately 10^{-9} transconjugants per donor.

Original element (strain)	Gene containing the Clostron insert	Recipient strain	Frequency of transconjugants per donor (σ)
CTn1 (630 Δ erm)	CD0364	CD37	7.0×10^{-9} (9.0×10^{-9})
		<i>B. subtilis</i> CU2189	$< 10^{-9}$
CTn1 (630 Δ erm)	CD0386	CD37	2.0×10^{-8} (1.7×10^{-8})
		<i>B. subtilis</i> CU2189	$< 10^{-9}$

PCR analysis was carried out on all transconjugants to detect co-transfer of elements that were not selected for in the filter-matings. Examples of the PCR results are shown in Figure 4.13 including results for the transconjugants described below in sections 4.4.4.1 - 4.4.4.5.

The control primer pairs *tcdB* and PaLoc TS are expected to be mutually exclusive as these generate a product from *tcdB* within the PaLoc and the empty PaLoc target site in the genome, respectively. As expected, all transconjugants generated a product with only one of these primer pairs. As a number of isolates were positive for *tcdB*, a second control PCR was performed with *cdtA* primers, amplifying the gene encoding one of the peptide chains of the binary toxin on the CDT locus, which is absent in CD37 [58]. PCR analysis using this primer pair was carried out by Dr. P. Warburton. The *tcdB* and *cdtA* primers generated products of the expected size in the transconjugant samples CD0364-4, CD0386-7 + 8, CD1099-2 + 4 and CD3392-6. All of the CTns tested for were also present in these samples and these were concluded to be donor isolates that became rifampicin resistant by spontaneous mutation. These isolates will be indicated in the relevant sections.

PCR results for the transconjugants of CTn1 are summarised in Table 4.8. Isolates CD0386-5, 6 and 11 contain copies of Tn5397. Sub-culturing these isolates on plates containing tetracycline confirmed that these isolates had become tetracycline resistant, as expected. Furthermore, CTn5 was also detected in CD0386-6. The transfer of these elements had not been selected for in the filter-matings. Additionally, CD0386-6 also contains a copy of the PaLoc but is negative for other 630 specific CTns and is also negative for *cdtA* encoded on the binary toxin locus. This transconjugant is further analysed in Chapter 5.

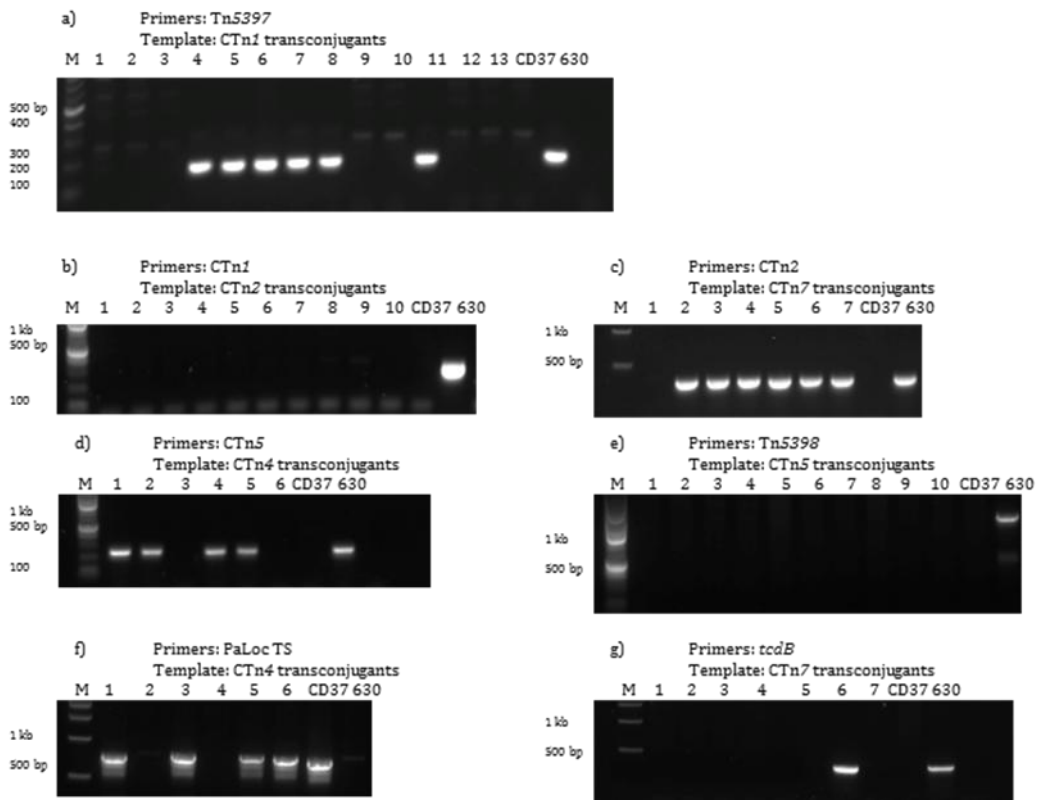


Figure 4.13 PCR results for the analysis of co-transfer in CD37 transconjugants.

Examples of each of the PCR reactions are shown. The order of the templates for each reaction is similar to the samples of the previously performed Southern blots. a) Template CTn1 transconjugants, primers Tn5397 CD0511 F+R. b) Template CTn2 transconjugants, primers CTn1 CD0364 F+R. c) Template CTn7 transconjugants, primers CTn2 CD0428 F+R. d) Template CTn4 transconjugants, primers CTn5 CD1873 F+R. e) Template CTn5 transconjugants, primers Tn5398 9371F + 9387R. f) Template CTn4 transconjugants, primers PaLoc TS Lok1+Lok3. g) Template CTn7 transconjugants, primers *tcdB* 3120F + 3521R.

Table 4.8 Summary of PCR analysis on transconjugants of CTn1 filter-matings.

PCR reactions that contain a product of the expected size are shown in green, reactions that do not contain the expected product are red. Template 1-13 are putative transconjugant DNA, CD37 and 630 are positive and negative controls. Primers: CTn2 CD0428 F+R, CTn5 CD1873 F+R, Tn5397 CD0511 F+R, Tn5398 9371 F-9387 R, *tcdB* ToxB 3120F + 3521R, PaLoc TS Lok1 + Lok3, *cdtA* BiTox-1F + 1R.

	CD0364				CD0386										CD37	630
	1	2	3	4	5	6	7	8	9	10	11	12	13			
CTn2	Red	Red	Red	Green	Red	Red	Green	Green	Red	Red	Red	Red	Red	Red	Green	
CTn5	Red	Red	Red	Green	Red	Green	Green	Green	Red	Red	Red	Red	Red	Red	Green	
Tn5397	Red	Red	Red	Green	Green	Green	Green	Red	Red	Red	Green	Red	Red	Red	Green	
Tn5398	Red	Red	Red	Green	Red	Red	Green	Green	Red	Red	Red	Red	Red	Red	Green	
<i>tcdB</i>	Red	Red	Red	Green	Red	Green	Green	Green	Red	Red	Red	Red	Red	Red	Green	
PaLoc TS	Green	Green	Green	Red	Green	Red	Red	Red	Green	Green	Green	Green	Green	Green	Red	
<i>cdtA</i>	Red	Red	Red	Green	Red	Red	Green	Green	Red	Red	Red	Red	Red	Red	Green	

Southern blots were performed in order to detect if CTn1 integrated into the recipient genome at multiple sites and if multiple copies of the element were present in the transconjugants. The probe binds to the recombinase which is present on a fragment of at least 3 kb, but possibly larger depending on the HindIII sites surrounding the target site of the element. The transconjugants for CTn1, both for the element marked in the ABC-transporter and the element marked in the cell surface protein, were all probed on a single blot. The probe was specific for CD0355, the tyrosine recombinase gene in CTn1 (Figure 4.14). This blot was performed to identify the number of target sites that CTn1 has in CD37. The control of CD37 DNA indicates that the probe hybridises to a fragment in the CD37 genome >10 kb, all transconjugants have hybridisation to a similar fragment. This result was expected as CD37 was shown to have a CTn1-like element [182].

For the DNA of the CD0364 transconjugants 1, 2 and 3, the probe hybridised to fragments of different sizes and to multiple fragments in each sample. This indicates that multiple copies of CTn1 have integrated into different sites in the chromosome. Putative transconjugants CD0364-4, CD0386-7 and 8 are expected to be rifampicin resistant donors, which explains hybridisation to a similar size fragment as seen for the 630 control DNA. As expected based on the 630 sequence, the minimal fragment size is >3 kb.

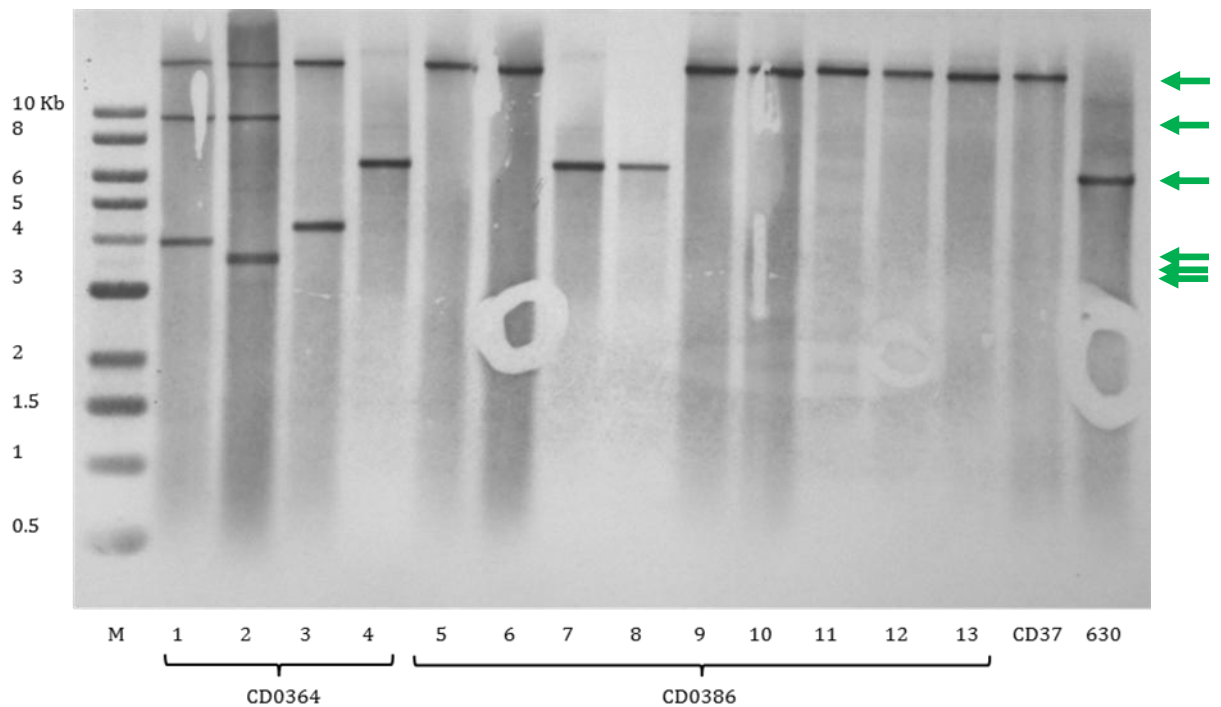


Figure 4.14 Southern blot of CTn1 transconjugants probed with the CTn1 recombinase CD0355.

All DNA was HindIII digested. Lane 1-4 is DNA of CD37 transconjugants containing CTn1 marked in CD0364. Lane 5-13 is DNA of CD37 transconjugants containing CTn1 marked in CD0386. Lane CD37 is DNA of strain CD37, lane 630 is DNA of strain 630, and lane M is a molecular marker. Green arrows indicate DNA fragments that are expected to be visualised through specific binding of the probe, assessment was made based on the size of the DNA fragments.

4.4.4.2. Conjugative transfer of CTn2

Filter-matings were carried out for CTn2 using the 630 Δ *erm* ClosTron mutant containing *erm*(B) in the transcriptional regulator CD0428 (Figure 4.11) and *C. difficile* CD37 and *B. subtilis* CU2189 as recipient strains. The transfer frequency into recipient CD37 was 2.1×10^{-4} transconjugants per donor (standard deviation is 5.5×10^{-4}). No transfer was observed into the *B. subtilis* recipient and the transfer frequency is considered below the detection limit of 10^{-9} transconjugants per donor. PCR analysis could not detect co-transfer of elements that were not selected in the filter-matings.

The Southern blot for the transconjugants of CTn2 was probed with the CTn2 serine recombinase gene CD0436 (Figure 4.15) to detect the number of integration sites and copy number of CTn2 in transconjugants. A HindIII restriction site is present in CTn2 and the recombinase is expected to hybridize to a fragment > 10 kb. The probe hybridised to a single 6 kb fragment in all samples, including the positive and negative controls which is therefore predicted to be non-specific binding. In addition, the probe hybridised to one or more greater than 10 kb fragments of the DNA from the transconjugants and 630 Δ *erm*. This suggests the element can integrate into multiple target sites and multiple copies of the element are present in three of the transconjugants.

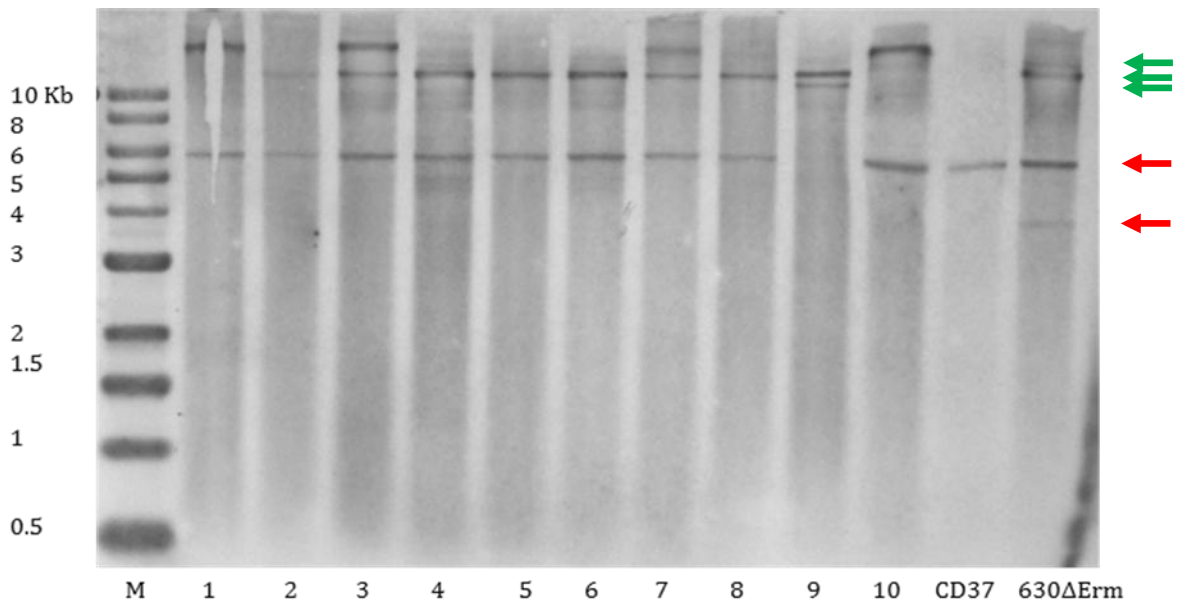


Figure 4.15 Southern blot of CTn2 transconjugants probed with the CTn2 recombinase CD0436.

All DNA was HindIII digested. Lane 1-10 is DNA of CD37 transconjugants containing CTn2 marked in CD0428. Lane CD37 is DNA of strain CD37, lane 630Δ*erm* is DNA of strain 630Δ*erm*, and lane M is a molecular marker. Coloured arrows indicate DNA fragments of varying sizes. Green arrows indicate DNA fragments that are expected to be visualised through specific binding of the probe, red arrows indicate DNA fragments that are expected to be visualised through non-specific binding of the probe, assessment was made based on the size of the DNA fragments.

4.4.4.3. Conjugative transfer of CTn4 shows co-transfer of CTn5

Filter-matings for CTn4, carried out using the 630Δ*erm* ClosTron mutant containing *erm*(B) in the transcriptional regulator CD1099 (Figure 4.11) as donor and *C. difficile* CD37 and *B. subtilis* CU2189 as recipients. Transfer into CD37 was observed at a frequency of 2.3×10^{-6} transconjugants per donor (standard deviation is 4.2×10^{-6}), whereas no transfer in *B. subtilis* was observed and the transfer frequency is considered to be below the detection limit of 10^{-9} transconjugants per donor.

PCR analysis for the co-transfer of elements that were not selected showed co-transfer of CTn5 in transconjugants 1 and 5 (Table 4.9). The PCRs also suggest that isolates 2

and 4 are not transconjugants but spontaneous mutants of the donor isolate which became rifampicin resistant, as described in section 4.4.4.1.

Table 4.9 Summary of PCR analysis on transconjugants of CTn4 filter-matings.

PCR reactions that contain a product of the expected size are shown in green, reactions that do not contain the expected product are red. Template 1-6 are putative transconjugant DNA, CD37 and 630 are positive and negative controls. Primers: CTn1 CD0364 F+R, CTn2 CD0428 F+R, CTn5 CD1873 F+R, Tn5397 CD0511 F+R, Tn5398 9371 F-9387 R, *tcdB* ToxB 3120F + 3521R, PaLoc TS Lok1 + Lok3, *cdtA* BiTox-1F + 1R.

	1	2	3	4	5	6	CD37	630
CTn1	Red	Green	Red	Green	Red	Red	Red	Green
CTn2	Red	Green	Red	Green	Red	Red	Red	Green
CTn5	Green	Green	Red	Green	Green	Red	Red	Green
Tn5397	Red	Green	Red	Green	Red	Red	Red	Green
Tn5398	Red	Green	Red	Green	Red	Red	Red	Green
<i>tcdB</i>	Red	Green	Red	Green	Red	Red	Red	Green
PaLoc TS	Green	Red	Green	Red	Green	Green	Green	Red
<i>cdtA</i>	Red	Green	Red	Green	Red	Red	Red	Green

Southern blot for the transconjugants of CTn4 was performed with a probe for the CTn4 recombinase. No known HindIII sites are present in the sequence of CTn4. The genomic DNA from *C. difficile* CD37 hybridises to a > 10 kb fragment as expected since a CTn4-like element is known to be present [182]. The probe hybridises in each of the transconjugants to a fragment of approximately the same size. This could indicate that CTn4 only integrates in the transconjugants at the same site as the CTn4-like element that is native to CD37, or a fragment of similar size.

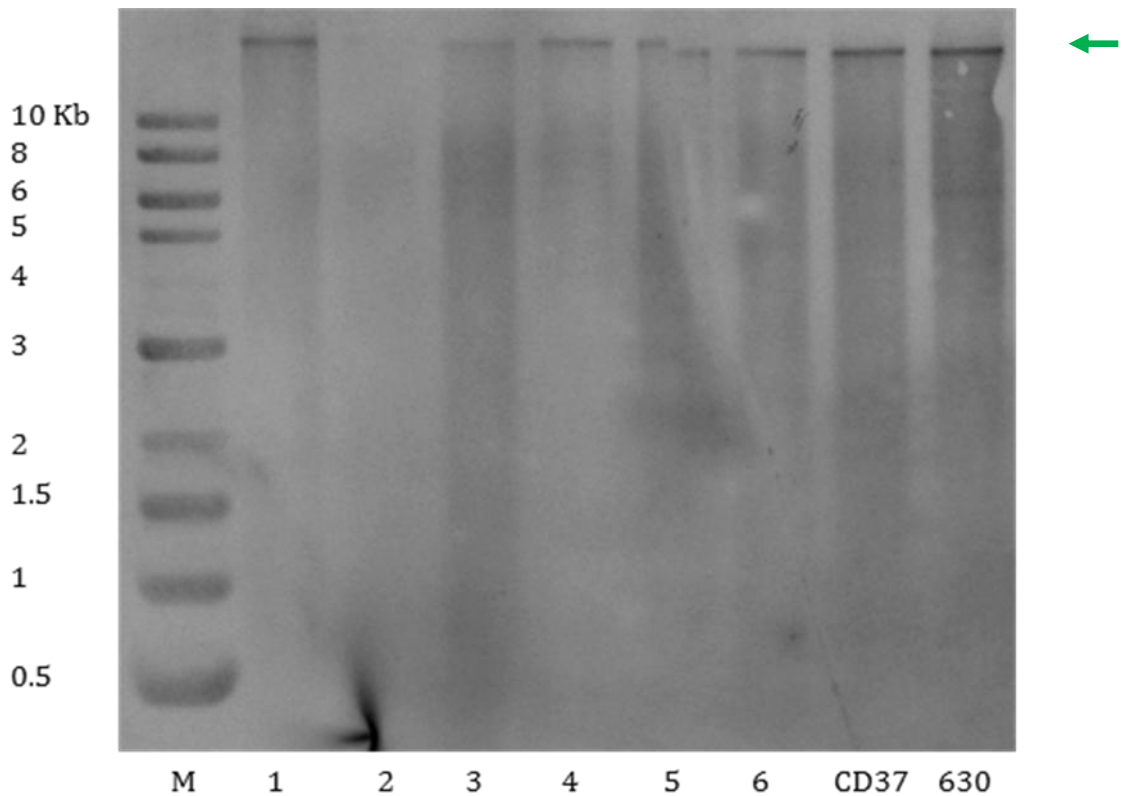


Figure 4.16 Southern blot for CTn4 transconjugants probed with the CTn4 recombinase CD1091.

All DNA was HindIII digested. Lane 1-6 is DNA of CD37 transconjugants containing CTn4 marked in CD1099. Lane CD37 is DNA of strain CD37, lane 630 is DNA of strain 630, and lane M is a molecular marker. A green arrow indicates the DNA fragment that is expected to be visualised through specific binding of the probe, assessment was made based on the size of the DNA fragment.

4.4.4.4. Conjugative transfer of CTn5 into *C. difficile* and *B. subtilis*

Filter-matings were carried out using the ClosTron mutant in 630 Δ *erm* for the ABC-transporter on CTn5, CD1873 (Figure 4.11). Transconjugants were detected using *C. difficile* CD37 or *B. subtilis* CU2189 as recipient strains. The frequency of transfer was determined for both recipients (Table 4.10). The frequency of transfer for CTn5 into *B. subtilis* was determined by Miss C. MacKenzie, UCL.

PCR analysis of the transconjugants of CD37 indicated that no elements had co-transferred into the CTn5 transconjugants.

Table 4.10 Transfer frequencies of CTn5.

The transfer frequency was calculated per filter as the number of transconjugants that resulted from a filter-mating, divided by the number of donor cells present on the filter. Each filter-mating experiment consisted of four technical replicates and each experiment was repeated at least three times. The mean of the transfer frequency and standard deviation were calculated from the frequencies of each independent filter. The detection limit for the filter-matings is approximately 10^{-9} transconjugants per donor.

Original element (strain)	Gene containing the ClosTron insert	Recipient strain	Frequency of transconjugants per donor (σ)
CTn5 (630 Δ erm)	CD1873	CD37	2.8×10^{-5} (2.6×10^{-5})
		<i>B. subtilis</i> CU2189	4.0×10^{-6} (3.4×10^{-6})

Southern blots were performed for the CTn5 transconjugants using the CTn5 recombinase as probe. The CTn5 recombinase is present on an internal HindIII fragment of approximately 4 kb. As expected, the probe hybridised to a 4 kb fragment for the 630 control DNA and all transconjugants (Figure 4.17).

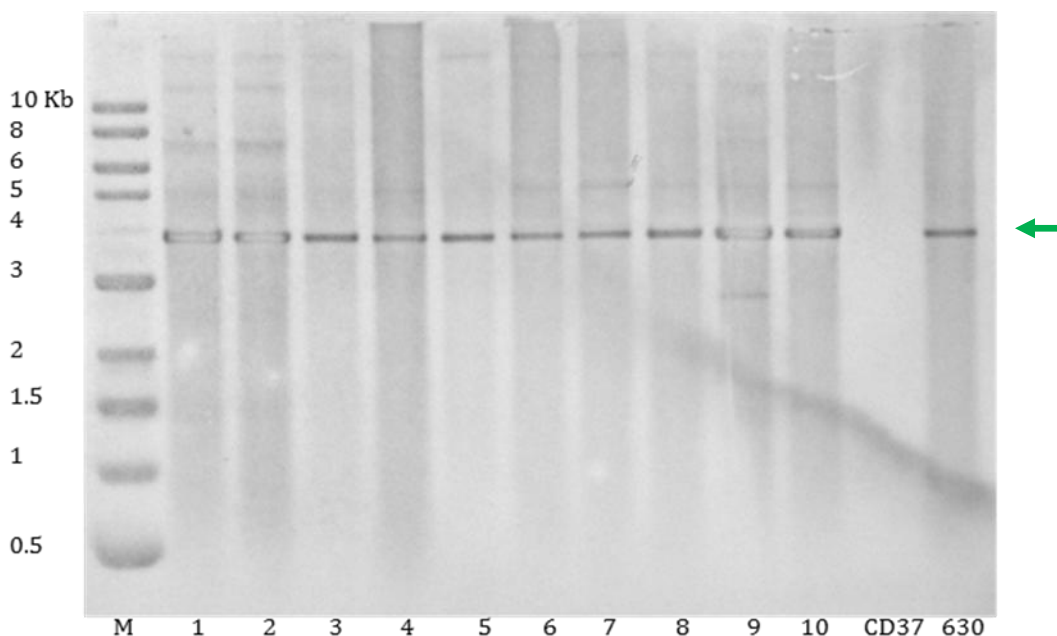


Figure 4.17 Southern blot of CTn5 transconjugants probed with the CTn5 recombinase CD1873.

All DNA was HindIII digested. Lane 1-10 is DNA of CD37 transconjugants containing CTn5 marked in CD1873. Lane CD37 is DNA of strain CD37, lane 630 is DNA of strain 630, and lane M is a molecular marker. A green arrow indicates the DNA fragment that is expected to be visualised through specific binding of the probe, assessment was made based on the size of the DNA fragment.

4.4.4.5. Conjugative transfer of CTn7

Filter-matings were carried out for CTn7 using 630 Δ *erm* containing a Clostron mutation in the cell surface protein CD3392 (Figure 4.10) as donor strain and *C. difficile* CD37 and *B. subtilis* CU2189 as recipients. The transfer frequency into recipient CD37 was 9.6×10^{-9} transconjugants per donor (standard deviation is 4.4×10^{-9}). No transfer was observed into the *B. subtilis* recipient and the transfer frequency is considered below the detection limit of 10^{-9} transconjugants per donor.

PCR analysis was performed to detect co-transfer of elements other than CTn7 whose transfer was not selected during the filter-mating experiments. Co-transfer of CTn2 was seen in five out of six transconjugants (Table 4.11). Sample 6 was shown not be a transconjugant but a spontaneous mutant of the donor isolate for rifampicin resistance, as described in section 4.4.4.1.

Table 4.11 Summary of PCR analysis on putative transconjugants.

PCR reactions that contain a product of the expected size are shown in green, reactions that do not contain the expected product are red. Template 1-7 are putative transconjugant DNA, CD37 and 630 are positive and negative controls. Primers: CTn1 CD0364 F+R, CTn2 CD0428 F+R, CTn5 CD1873 F+R, Tn5397 CD0511 F+R, Tn5398 9371 F-9387 R, *tcdB* ToxB 3120F + 3521R, PaLoc TS Lok1 + Lok3, *cdtA* BiTox-1F + 1R.

	1	2	3	4	5	6	7	CD37	630
CTn1	Red	Red	Red	Red	Red	Green	Red	Red	Green
CTn2	Red	Green	Green	Green	Green	Green	Green	Red	Green
CTn5	Red	Red	Red	Red	Red	Green	Red	Red	Green
Tn5397	Red	Red	Red	Red	Red	Green	Red	Red	Green
Tn5398	Red	Red	Red	Red	Red	Green	Red	Red	Green
<i>tcdB</i>	Red	Red	Red	Red	Red	Green	Red	Red	Green
PaLoc TS	Green	Green	Green	Green	Green	Red	Green	Green	Red
<i>cdtA</i>	Red	Red	Red	Red	Red	Green	Red	Red	Green

The Southern blot of the CTn7 transconjugants was probed with the CTn7 serine recombinase CD3370 to identify the number of target sites and the copy number of the element in CD37 transconjugants (Figure 4.18). The probe is expected to hybridize to a fragment >2.5 kb, based on the 630 sequence. The probe hybridised with DNA fragments of 6 kb in 6 out of 7 transconjugants and in controls CD37 and 630 Δ erm. CD37 contains a CTn7-like element [182] but transconjugants 1, 3 and 4 only hybridise to a single fragment, and it is hypothesised that the native CTn7-like element may have been lost in these transconjugants.

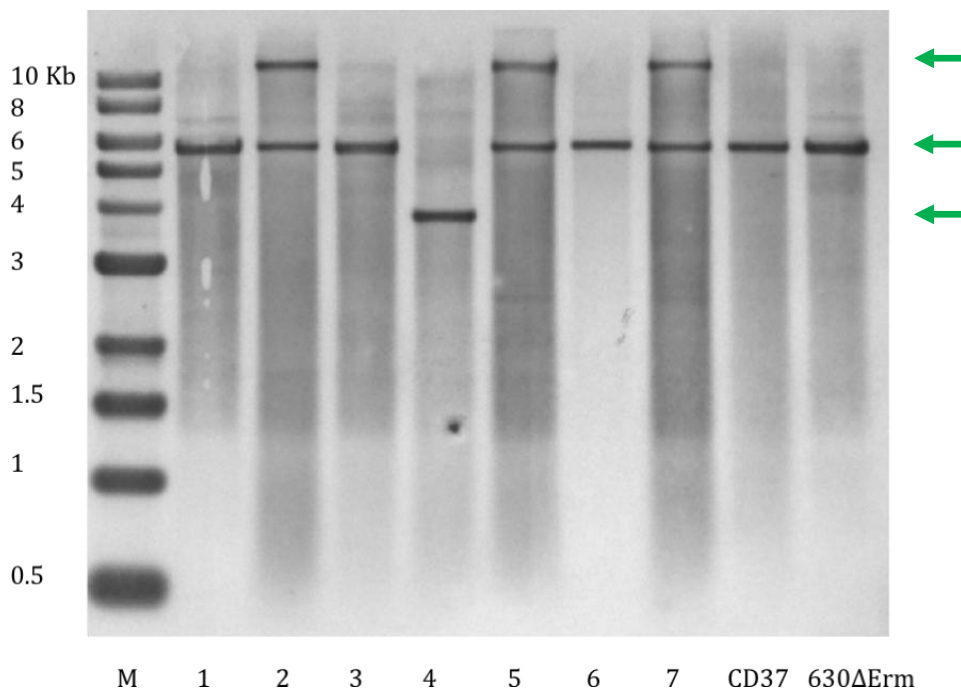


Figure 4.18 Southern blot of CTn7 transconjugants probed with the CTn7 recombinase CD3370.

All DNA was HindIII digested. Lane 1-7 is DNA of CD37 transconjugants containing CTn7 marked in CD3392. Lane CD37 is DNA of strain CD37, lane 630 Δ erm is DNA of strain 630 Δ erm, and lane M is a molecular marker. Green arrows indicate DNA fragments that are expected to be visualised through specific binding of the probe, assessment was made based on the size of the DNA fragments.

4.4.4.6. No conjugative transfer of Tn6103 was observed

No transconjugants were observed in the matings using R20291 as the donor of the element Tn6103 with any of the recipient strains CD37, 630 Δ *erm* or *B. subtilis* CU2189. The transfer frequency for this element into any of these recipients is therefore considered to be below the detection limit of 10^{-9} transconjugants per donor cell.

4.4.4.7. Conjugative transfer of the CTns of strain M120

Strain M120 was shown to contain 2 putative CTns, one 107 kb element designated Tn6164 and a Tn916-like element designated Tn6190 [88] (Chapter 3). Filter matings were carried out using M120 as a donor and CD37 as a recipient. As both Tn6164 and Tn6190 encode a predicted tetracycline resistance gene, filter-matings selecting for rifampicin and tetracycline resistance would select for CD37 transconjugants containing either of these elements. A large number of putative transconjugants were obtained and 70 putative transconjugants from 2 separate experiments were checked by PCR for identity of the strain with the Lok1+Lok3 primer pair. This primer pair spans the PaLoc integration site and can only produce an amplicon if the PaLoc is not present, confirming the isolate is the non-toxigenic recipient strain CD37 [32]. Furthermore, the presence of both Tn6164 and Tn6190 were tested with the primer pairs Tn6164 acc region F+R and Tn916 integrase F+R (Table 4.3), respectively. All transconjugants were shown to be CD37 containing the Tn916-like element Tn6190 (Figure 4.19). Conjugative transfer of Tn6164 was not detected.

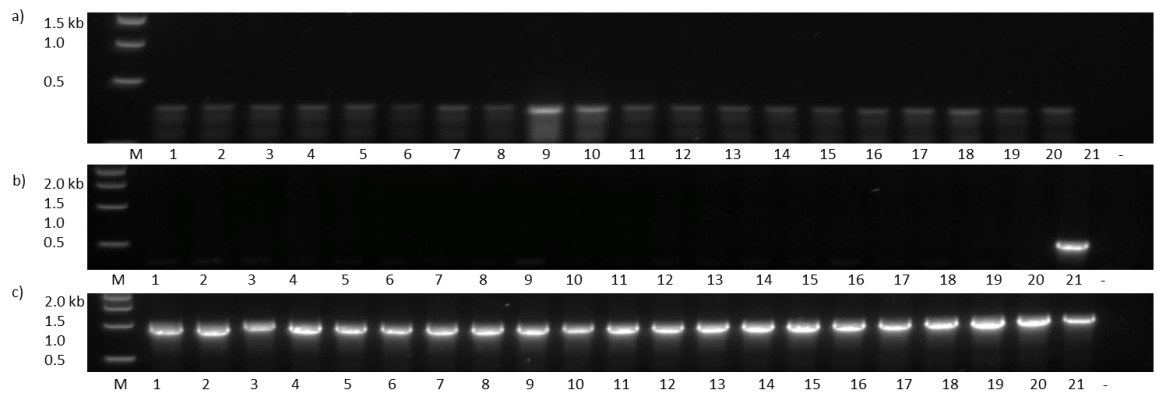


Figure 4.19 Transconjugant analysis of M120 filter-matings

Panel A shows the product of the PCR to confirm the putative transconjugants are CD37, the primer pair is lok1 + lok3 and the expected product size is 0.8 kb. Panel B shows the PCR analysis for the presence of Tn6164 with primer pair Tn6164 acc region F+R, expected product size is 0.5 kb. Panel C shows the PCR analysis for the presence of the Tn6190 with primer pair Tn916 integrase F+R, expected product size is 1.0 kb. The template in lanes 1-20 are transconjugants 1-20, lanes 21 are M120 and lanes – are H₂O. Lanes M are a molecular weight marker, the size of the fragments is indicated left of the figures.

4.4.5. Nanotube transfer of antimicrobial resistance proteins

Nanotubes were shown to allow for the transfer of antibiotic resistance proteins between bacterial cells without the genetic element encoding the resistance gene transferring [187]. Nanotubes were also shown to be capable of transferring non-conjugative DNA elements between *B. subtilis* strains. To study if nanotubes are formed between *C. difficile* cells and to see if antibiotic resistance proteins are transferred, the experiments described by Dubey and Ben-Yehuda [187] were repeated with *C. difficile*.

Cultures of single *C. difficile* strains and mixed cultures of strains with different resistance phenotypes were co cultured on an antibiotic free agar plate for 24 hours after which these were replica plated onto plates containing different combinations of antibiotics and to an antibiotic free plate. These plates were checked after 24 and 48 h incubation and the experiment was performed three times with similar results. On thiamphenicol plates, CD37 pMTL960/GusA (thiamphenicol^R) and the mixed culture of CD37 pMTL960/GusA and R20291 pMTL 9301 grew (Figure 4.20). On lincomycin plates, R20291 pMTL 9301 (lincomycin^R) and the mixed culture grew. On the thiamphenicol + lincomycin plates none of the cultures grew. This indicates that antibiotic resistance proteins and non-conjugative plasmids were not transferred when two cultures with different resistance phenotypes were co-cultured. On the antibiotic free plates all cultures grew.



Figure 4.20 Nanotube transfer of antimicrobial resistance proteins.

The occurrence of transfer of antimicrobial resistance proteins through nanotubes was studied by co-culturing *C. difficile* strains which are resistant for different antibiotics. a) Position of strains on agar plates; 1 CD37 pMTL960/GusA (thiamphenicol^R, lincomycin^S); 2 wildtype CD37 (thiamphenicol^S, lincomycin^S); 3 mixed culture of CD37 pMTL960/GusA and R20291 pMTL9301; 4 wildtype R20291 (thiamphenicol^S, lincomycin^S); 5 R20291 pMTL9301(thiamphenicol^S, lincomycin^R). b) Agar plates of replica plated cells after 24 h incubation. Plates were also checked after 48 h incubation. The supplemented antibiotics present in the agar are indicated above the plates; Thi 30 is 30 µg/ml thiamphenicol, Linc 40 is 40 µg/ml lincomycin, Abf is antibiotic free.

4.5. Discussion

4.5.1. Circularisation of *C. difficile* putative CTns

The first step in conjugative transfer of a CTn is considered to be excision from the genome into a circular intermediate [128]. Therefore, the presence of these circular intermediates was investigated for each element before further study. The joint between the ends of the elements present in a circular molecule was PCR amplified for CTn1, CTn2 and CTn7 present in strain 630 and the excision of CTn4 and CTn5 as previously described was confirmed [8]. In five other strains the excision of another seven CTns was shown which demonstrates that a large number of *C. difficile* predicted CTns are active elements.

Sequencing of the circular intermediate, empty target site and the junctions between the CTn and the chromosome has identified the direct repeats at the ends of the elements. Most of the elements that were analysed here encode serine recombinases and all of these were shown to create the original regenerated target site upon excision. Two elements encoding serine recombinases in *C. difficile* that were previously studied have 2 bp direct repeats at the ends of the element as well as in the target site and joint of the circular intermediate [156,168], although a 5 bp direct repeat was described for CTn5 [8]. Only two out of eight serine recombinase-encoding elements analysed here had a 2 bp direct repeat, whereas the direct repeats of the other elements ranged between 8-15 bp. Further study could determine if the sequence of these direct repeats are required to be conserved in order for integration and excision of these elements to take place.

Excision of two elements encoding tyrosine recombinases described here resulted in regeneration of the target site without any heteroduplexes. These elements have 6-7 bp direct repeats at either end of the element. Although tyrosine recombinases often create heteroduplexes upon integration and excision, as described for CTn4 in *C. difficile* [8,157], if the direct repeats of the element are identical to the target site and joint of the circular intermediate, no heteroduplexes are formed [247].

In strains R20291, QCD-66C26, and QCD-32G58, CTn1-like elements are present that share 98-100% sequence identity and all lack a homologue to the predicted *xis* of CTn1 (Chapter 3). Excision could not be shown for any of these elements under our growth conditions, which is presumably because of the absence of a *xis* homologue. To demonstrate if *xis* is essential for the excision of these elements, CTn1 could be introduced into these strains to provide *xis in trans* which may complement the deletion in these strains.

Most of the elements that were analysed here were able to excise from the genome into a circular intermediate, indicating these are active mobile elements that play a role in the plasticity of the *C. difficile* genome. These elements may be capable of conjugative transfer and so contribute to genetic exchange between *C. difficile* isolates and between *C. difficile* and other species, which warrants further study of these elements.

4.5.2. Conjugative transfer of CTns

Six CTns that were shown to excise from the genomes in strains 630 and R20291 were marked by introducing *erm(B)* into the element using the CloStron. In order to

minimise the effect of this insertional mutation on the conjugative properties of the element, the mutations were made in predicted accessory genes.

Filter-mating experiments showed that all of the elements to which the Clostron was retargeted in strain 630 were capable of conjugation, confirming that none of these genes are essential for conjugation. The conjugation frequencies of the Tn916-like elements CTn1 and CTn7 were similar and ranged between 10^{-8} - 10^{-9} transconjugants per donor, similar to the frequencies reported for the transfer of Tn916 and Tn5397 between *C. difficile* strains [164,248].

The transfer frequencies of the Tn1549-like elements CTn2, CTn4 and CTn5 ranged from 10^{-4} - 10^{-6} , which is comparable to the transfer frequency of Tn1549 between *E. faecalis* strains [148]. Although these frequencies are somewhat higher than those of the Tn916-like elements studied here, it is unknown if this is a statistically significant difference and further study is needed to determine if Tn1549-like elements actually have a higher transfer frequency.

Culturing a donor strain of Tn916 in the presence of tetracycline prior or during matings enhances the transfer frequency up to 15-fold [249]. This has led to a model for the regulatory mechanism of Tn916 in which the presence of tetracycline results in the destruction or silencing of the transcriptional terminator *orf12*. This allows expression of the downstream genes including the transposition and conjugation genes [138,250]. As none of the elements in *C. difficile* strain 630 encode *orf12* homologues, it is unlikely that these elements are affected by tetracycline. It is possible though that the regulation of these elements is also affected by environmental factors and future study determining these factors is important for fundamental research of these elements.

4.5.3. Co-transfer of CTns from strain 630 Δ *erm*

The filter-matings performed here have all selected for the transfer of a single element carrying *erm*(B). Besides the elements that were selected for, co-transfer of several elements was demonstrated by Southern blot and PCR. When selecting for the transfer of CTn1, the transfer of Tn5397 was observed multiple times, once in combination with CTn5 and the PaLoc. CTn5 was also shown to transfer when the transfer of CTn4 was selected for. However, both of these elements have also transferred independently, indicating transfer of neither element is reliant on the other element. In contrast, the transfer of CTn7 was observed in combination with the transfer of CTn2 in five out of six transconjugants. Although it appears that CTn7 is a complete conjugative transposon, it is possible that the element is dependent on CTn2 for its transfer. The CTn7 transconjugant in which CTn2 is not present may have lost the element during transposition. To test if CTn7 relies on CTn2 for transfer, more transconjugants could be analysed for co-transfer. If this is inconclusive, then knockout mutants could be made for any genes on CTn2 that are essential for transfer and determining the effect of this on the transfer of CTn7.

Although the term co-transfer has been used here, it is uncertain that the transconjugants containing more than one element have always arisen from a single transfer event. Possibly, when two cells are in close range, after the first element has been transferred, a second transfer event between these two cells could occur, or one recipient cell may also receive elements from more than one donor cell. This could help to explain the observation of co-transfer of CTn2 when only the transfer of CTn7 is selected for because the transfer frequency of CTn2 is so much higher than that of the other elements. Furthermore, the co-transfer of CTn5 was observed several times,

which is the element with the highest transfer frequency after CTn2. However, this does not explain why the co-transfer is not seen in the CTn1 and CTn4 transconjugants.

During the selection for the transfer of CTn1, the co-transfer of Tn5397 was detected several times. After selection on erythromycin only, these cells were also plated out onto tetracycline, showing that tetracycline resistance was acquired during these matings without any selection for this event. The transfer of antibiotic resistance genes without selection pressure demonstrates the clinical importance of the research into horizontal gene transfer in order to understand which factors may induce these events.

4.5.4. Number of integration sites in transconjugants

The results of the PCRs that were performed to detect co-transfer led to the conclusion that samples CD0364-4, CD0386-7 and 8, CD1099-2 and 4 and CD3392-6 are rifampicin resistant donor strain mutants. All of these samples contain all of the 630 CTns as well as the PaLoc and CDT-locus which means no CD37 specific sequence could be determined. Southern blots were performed on these samples before these conclusions had been reached.

To probe the Southern blots, the recombinases of the transposons were used. Sequence comparison demonstrated diversity between the genes which suggested these probes would not hybridize to the sequence of other CTns. However, during the experiments, it became apparent that CD37 has a number of related CTns with similar recombinases. During the PCR analysis of the transconjugants to look for co-transfer of other unmarked CTns, genes on the accessory regions of the conjugative transposons were used after it was determined these are specific for the 630 CTns. Co-transfer of CTn4 and CTn7 could not be tested for here as CD37 was positive for the accessory

genes of these elements. After these experiments were performed, the shotgun genome sequence of CD37 (Chapter 6) showed that CTn4 and CTn7-like elements, with 96% and 98% sequence identity respectively, are present in CD37.

The recombinases of the CTns were also used as probes for the Southern blots as these genes are all at one side of the element which limits the chance of HindIII sites being present between the probe and the end of the element in the target site. This enables to determine if the element can integrate in multiple sites in the recipient genome. However, examination of the results of the CTn5 transconjugants showed the probe used hybridized to an internal HindIII fragment and could not be used to determine the number of integration sites. This experiment could be repeated using a restriction enzyme that does not have restriction sites between the CTn5 recombinase and the end of the element.

In the Southern blots of the CTn1 transconjugants the CTn1 recombinase probe hybridised to two or three fragments of different sizes for the DNA of transconjugants CD0364-1, 2 and 3. This indicates that the element has integrated into at least 5 different target sites in the transconjugants and that multiple copies of the element are present. The fact that the probe only hybridises to a single fragment for all CD0386 transconjugants may indicate that the marked CTn1 element has integrated in the same fragment as the native CTn1 recombinase or that the marked element has replaced the native element carrying the CTn1-like recombinase.

Southern blot of the CTn2 transconjugants shows hybridisation with the CTn2 recombinase probe to >10 kb fragments in all transconjugants; at least 3 separate size fragments can be distinguished indicating multiple integration sites and multiple copies of the element in at least three of the transconjugants.

The Southern blot of the transconjugants of CTn4 was hybridised with the CTn4 recombinase probe which hybridised at a >10 kb fragment in all transconjugant samples as well as CD37 and 630. There was no indication of two copies of the element, which could indicate the native CTn4-like element was lost in the transconjugants or the elements have recombined, however, CTn4 and the CTn4-like element from CD37 may have been present on DNA fragment of similar sizes and were too big to be separated on the gel. To determine if both elements are present, the Southern blot could be repeated with another restriction enzyme that has a restriction site within CTn4 or the gel for the Southern blot could be performed pulsed field gel electrophoresis in order to get a better separation of the large DNA fragments.

CD37 contains a CTn7-like element integrated in the same target site as in 630. In three out of six transconjugants, the CTn7 recombinase probe hybridised with two fragments, which could mean both the native CD37 CTn7-like element is present as well as the donor copy and the element can integrate in a different target site when the preferred target site is occupied. Only a single copy of the element is present in the other three transconjugants which suggests that the native CTn7-like element was lost or the elements have recombined. Whole genome sequencing of the transconjugants could determine what exactly happened with these elements.

4.5.5. *C. difficile* elements that were not shown to transfer

In Tn6103 from strain R20291, the Clostron was retargeted to an ABC-transporter, a homologue of the ABC-transporter in 630 where the Clostron was retargeted in CTn5. No transfer was observed for Tn6103 from R20291 into CD37, 630 Δ erm or *B. subtilis* CU2189. Although CTn5 can transfer from 630 Δ erm into CD37 and *B. subtilis* CU2189,

this was not unexpected as three predicted mobilisable transposons are integrated in predicted conjugal transfer genes in Tn6103 (Chapter 3).

Strain M120 contains two predicted conjugative transposons that encode tetracycline resistance proteins, Tn6164 and Tn6190 [208]. Tn6190 is 97% identical to Tn916, whereas Tn6164 is a composite element that consists of fragments of several MGEs (Chapter 3). It was therefore expected that filter-matings using M120 as donor and selecting for tetracycline resistant transconjugants, would result in isolates containing Tn6190 and possibly in isolates containing Tn6164. All of the transconjugants contained Tn6190. Although no conjugation of Tn6164 was shown in the experiments performed here, the element was shown to excise from the genome into a circular molecule, indicating that the element is an active MGE. No final conclusion can be drawn on whether Tn6164 may be able to transfer via conjugation as our findings could be the result of Tn6190 having a much higher transfer frequency than Tn6164. Studying a larger number of transconjugants could have resulted in detection of a Tn6164 transconjugant. Furthermore, Tn6164 may transfer under different conditions than those we have used in our experiments or strain CD37 may not be a suitable recipient strain for Tn6164. Although Tn6164 is considerably larger than any of the elements in strain 630 for which transfer has been shown here, transfer of large CTns have been shown in other species such as CTnscr94 from *Salmonella senftenberg* (100 kb) and the R7a symbiosis island from *Mesorhizobium loti* (500 kb) [146,251].

Further study of the transfer of Tn6164 could be performed by constructing a mutant strain in which the *tet(M)* of Tn6190 is knocked out. This would allow for filter-mating experiments to be conducted in which the selection on tetracycline would select only for Tn6164 transconjugants. Alternatively, *erm(B)* could be introduced into Tn6164

using the Clostron system, as was done here for the CTns from strains 630 and R20291.

4.5.6. Nanotube transfer of antimicrobial resistance proteins

Dubey and Ben-Yehuda have shown that co-culture of two organisms can result in a mixed culture where both cell types show a resistant phenotype although only one of the organisms contains the specific antibiotic resistance gene necessary to survive in a specific environment [187]. The network of nanotubes connects cells to each other, even from different species, and it was shown that antibiotic resistance proteins as well as non-conjugative plasmids could be transferred through them. Although the study shows that the transfer of the plasmid does not occur through transformation, it is not clear if the donor strain used in these experiments contains other MGEs that could mobilise the non-conjugative plasmid.

It was attempted here, to show if this transfer of resistance proteins could also be shown to occur between *C. difficile* strains and to determine if transfer of CTns could occur through nanotubes. The experiments as described by Dubey and Ben-Yehuda [187] were repeated using *C. difficile* but no evidence was found for transfer of resistance proteins. Possibly, a nanotube transfer system does not exist for *C. difficile*, although the related species *B. subtilis* was used in the Dubey study. Although great care was taken to reproduce the conditions of the experiments described by Dubey and Ben-Yehuda, the conditions in which the experiments were undertaken may affect the frequency at which nanotubes are formed [188], and variations, such as the fact that the experiments here were performed under anaerobic conditions, are inevitable. To exclude this possibility, these experiments could be repeated using *B. subtilis* under aerobic and anaerobic conditions. At the moment, because of the lack of evidence for

the existence of a nanotube system for *C. difficile*, the transfer of DNA seen in this chapter is considered to occur through conjugation.

4.5.7. Conclusions

The mobility of 15 *C. difficile* CTns was tested here and 10 were shown to be able to excise from their respective genomes indicating that most of the *C. difficile* CTns are active elements that can excise from the genome into a circular transfer intermediate. The conjugative transfer between *C. difficile* strains was shown for six out of eight tested elements and one element was shown to transfer into *B. subtilis*, showing it is possible these elements contribute to genetic exchange with different strains and species.

Southern blots of transconjugants were performed to test if the elements could integrate into multiple target sites and to see if multiple copies are present in the transconjugant genome. CTn1, CTn2 and CTn7 were shown to be able to insert in three to five different sites each in CD37 and CTn1 and CTn2 were shown to have multiple copies in transconjugant cells.

Ten out of forty transconjugants that were studied here were shown to contain CTns that co-transferred when the transfer of another element was selected. It is unknown if these elements transferred in a single transfer event or in multiple events. The transfer of Tn5397, conferring tetracycline resistance to the transconjugants when lincomycin resistance encoded on a different element was selected for, indicates these transfer events can have clinical implications.

One CTn1 transconjugant was demonstrated to contain CTn5, Tn5397 and the PaLoc and this transconjugant is further studied in Chapter 5.

Chapter 5 Mobilisation of the PaLoc

5.1. Introduction

The PaLoc is a 19.6 kb genetic element encoding 5 predicted proteins including the main virulence factors TcdA and TcdB (discussed in Chapter 1). The PaLoc was shown to be a separate genetic element that has integrated uni-directionally in the same target site in all toxigenic *C. difficile* strains, whereas a 115 bp sequence is present at this locus in all nontoxigenic strains [32]. As the PaLoc encodes a putative phage holin and other members of the family of large clostridial toxins, *C. botulinum* neurotoxins C1 and D, are encoded on the neurotoxin-converting phages [252] the PaLoc has been hypothesized to be of phage origin [174]. It was therefore hypothesized that the presence of prophages in the genome may have an effect on toxin production through the presence of phage transcriptional regulators and several studies have looked into this [174,176,177]. Although these studies show that phages can affect toxin production, these effects are strain and phage specific and no common effect has been described.

5.2.Aims of this chapter:

- Reproducing the transfer of the PaLoc from the donor *630Δerm* into the non-toxicogenic recipient strain CD37.
- Determining if the transconjugant Paloc386 from Chapter 4, which contains the PaLoc, produces the PaLoc toxins.
- Studying the CD37 transconjugants containing the PaLoc by genome sequencing in order to predict what transfer mechanism is likely to be responsible for PaLoc transfer.

5.3. Materials and methods

5.3.1. Tissue culture and cytotoxicity assays

Human foreskin fibroblast (HFF)-1 cells (ATCC SCRC-1041) were grown in Dulbecco's Modified Eagle's Medium (Gibco-Invitrogen, UK) supplemented with 10% foetal calf serum (Hyclone, USA), 0.1% glutamine (Cambrex, Belgium), penicillin/streptomycin mix (100 U ml⁻¹) (Lonza, Belgium). Cells were grown at 37 °C at 5% CO₂ in vented flasks. Depending on confluence of growth, the cells were split 1-2 times per week by washing the cells twice with sterile PBS (Sigma Aldrich Ltd.) using 1 ml trypsin (Cambrex, Belgium) and incubating approximately 15 min at 37 °C. When most cells were not adhering, pre-warmed medium was added to the cells and these were split over 2 or 3 flasks.

Bacterial strains were grown in TY broth for 48 h and spun down for 10 min at 4500 x g. The supernatant was filter sterilised (filter pore size 0.45 µm) and stored at 4 °C. The cytotoxicity assay was performed using the *C. difficile* toxB test kit (Techlab Inc, Blacksburg, VA, USA) according to the manufacturer's instructions. HFF-1 cells were trypsinized and seeded into a 96-well plate at 1x10⁴ cells per well. After 24 h it was checked that the growth of the cells was confluent in the wells. Filter sterilised bacterial supernatant was 4 fold serial diluted and neat to 1:4096 were all added to the fibroblast monolayer in duplicate. Polyclonal anti *C. difficile* TcdB antibodies were added to one series of dilutions. Purified *C. difficile* TcdB was added to one well as a positive control. Cells were incubated at 37 °C for 24 h before they were assessed under an inverted microscope for the characteristic cell rounding also referred to as the cytopathic effect (CPE) [44]. The toxin end titre point has been defined as the first

dilution point where no difference is seen between the sample and the negative control that has been treated with anti TcdB antibodies [45]. The toxin end titre point was determined for the dilution series of each of the samples.

5.3.2. Filter-matings

Filter-matings were performed as described in Chapter 2. The Clostron mutant 630 Δ *erm tcdB::ErmRAM* was used as the donor strain as this strain has an *erm(B)* insertion in *tcdB*. CD37 was used as the recipient strain. Serial dilutions of the mixture of donor, recipient and transconjugant cells from the filters were set up to determine the cell counts for donor and recipient cells. Donor cells were selected on lincomycin (40 μ g/ml) containing plates and recipient cells were selected for on rifampicin (25 μ g/ml) containing plates. Transconjugant cells were selected for on lincomycin (40 μ g/ml) and rifampicin (25 μ g/ml) containing plates.

5.3.2.1. PCR analysis

PCR was used to identify transconjugants and to look for co-transfer of elements that were not selected for in the filter-mating. PCR analysis was performed as described in Chapter 2. PCR was performed with the primers listed in Table 5.1.

Table 5.1 PCR primers for CTn detection.

Primer pair CD0364 F+R is for the detection of the presence of CTn1. Primer pair CD0428 F+R is for the detection of the presence of CTn2. Primer pair CD1873 F+R is for the detection of the presence of CTn5. Primer pair CD0511 F+R is for the detection of the presence of Tn5397. Primer pair 9371 F + 9387 R is for the detection of the presence of Tn5398. Primer pair ToxA F + R is for the detection of the presence of *tcdA* encoded on the PaLoc. Primer pair Lok1 + Lok3 is for the detection of the empty target site of the PaLoc. Primer pair *cdtA* F+R is for the detection of the presence of *cdtA* encoded on the binary toxin locus.

Primer	Sequence 5'- 3'	Amplicon
CTn1 CD0364 F	GATTAAAGCCCTCGCACAAG	0.3 kb
CTn1 CD0364 R	GCGTCCGCTAATTTTGTGAT	
CTn2 CD0428 F	CATGGAATCTTTCGCATCAGA	0.4 kb
CTn2CD0428 R	GCGAATGTCCTCACTAACCG	
CTn5 CD1873 F	CAATAATGAAATCAAGCAATGAAA	0.2 kb
CTn5CD1873 R	GCTTCGAGTGAAACAACATTCT	
Tn5397 CD0511 F	CGAGCGGAAGTCAGAAGATT	0.2 kb
Tn5397 CD0511 R	GCTTTGCACTACCTCTGCTTG	
Tn5398 9371F	GGATAGAAATACTCGTCAACAG	0.9 kb
Tn5398 9387 R	ATTTTTTATTTTTAGGAGTCATA	
ToxA F	CATCTATAAGTTCTCATATTCCTTC	1.2 kb
ToxA R	CCAGAACTTATATTGTCC	
Lok1	AAAATATACTGCACATCTGTATAC	0.6 kb
Lok3	TTTACCAGAAAAAGTAGCTTTAA	
<i>cdtA</i> F	YAATACTACTTACAAGGCTCC	0.3 kb
<i>cdtA</i> R	TTTCGTTTTGATTTTYTGTTCC	

5.3.3. Phage experiments

5.3.3.1. Preparation of phage suspensions

A culture of 630 Δ *erm tcdB::ErmRAM* was grown for 16 h and was used to inoculate a 50 ml culture 1:10 (v/v), which was grown for approximately 20 h. Mytomycin C was added to a final concentration of 3 μ g/ml and the cells were incubated for 4 h anaerobically. The cells were centrifuged for at 7800 x g for 10 min, at 4 °C. 1 M NaCl and 20% (w/v) PEG 6000 were added to the supernatant [253]. The phages were left to precipitate on ice at 4 °C overnight. The phages were harvested by centrifugation at 7800 x g for 20 min, at 4 °C. The supernatant was discarded, the phages were

resuspended in 2 ml PBS and filter sterilised. The phage suspensions were stored at 4 °C.

5.3.3.2. Phage infection assay

A CD37 culture was grown in 10 ml broth for 16 h after which it was incubated with 1 ml of the phage suspension. The cells were incubated for 2 h at 37 °C and the cells were centrifuged 10 min at 4500 x g. The pellet was resuspended into 1 ml of pre-reduced broth and 200 µl spread onto plates containing rifampicin (25 µg/ml) and lincomycin (40 µg/ml).

5.3.3.3. Plaque assay

CD37 and CD843 cultures were grown for 16 h. CD843 was previously described as a propagating strain for both ϕ C630-1 and ϕ C630-2 of strain 630 [175]. 100 µl of culture and 100 µl of phage suspension was added to 4 ml of 50 °C top agar, BHI 0.6% agar, in sterile polystyrene tubes (Sterilin Ltd, Newport, UK). The tube was inverted several times for mixing and spread out onto a BHI plate. Plaque plates were evaluated after 24 h incubation for the presence of clear plaques in the bacterial lawn that had grown.

5.3.4. DNA shotgun sequencing

Full genome shotgun sequencing was carried out by UCL Genomics, University College London. Dr. Tony Brooks prepared the samples as part of a 72 bp paired end multiplex reaction, and sequencing was performed on an Illumina GAII-X.

Dr. Warren Emmett performed the assembly and annotation of the genome sequences. Initial genome assembly was done using the Velvet software suite [254]. The assembled genome was compared and mapped against the annotated CD630

reference genome (Genbank AM180355) using xBase [255]. Gene prediction was performed using Glimmer [193], tRNA genes were searched for with tRNAScan-SE [256], Ribosomal RNA genes were searched for with RNAmmer [257], Protein BLAST [258] was run using the translated coding sequences as a query against the CD630 reference sequence. The best result for each BLAST search was imported as the gene annotation. A comparison file with concatenated reference and sequence files was produced using MUMMER [259]. Analysis of the genome was performed using ACT [204].

5.4. Results

5.4.1. Transconjugant Paloc386 is cytotoxic

To determine if the transconjugant isolate Paloc386 described in Chapter 4 that contains the PaLoc expresses the toxin genes, a cytotoxicity assay was performed. Filter-sterilised cell free supernatant was serial diluted and added to HFF-1 cells. Supernatant of Paloc386 was compared to that of the recipient strain CD37 and donor strain 630. The CPE was assessed visually for rounding of the cells using an inverted microscope (Figure 5.1). The toxin end titre point was determined for each of the samples (Figure 5.2). Similar toxin levels were present in the culture supernatant of donor 630 Δ erm and transconjugant Paloc386, whereas no CPE was seen for the culture supernatant of recipient CD37. No difference could be distinguished between the negative control containing only tissue culture medium and the wells containing the neat culture supernatant that was neutralized with the TcdB antibodies.

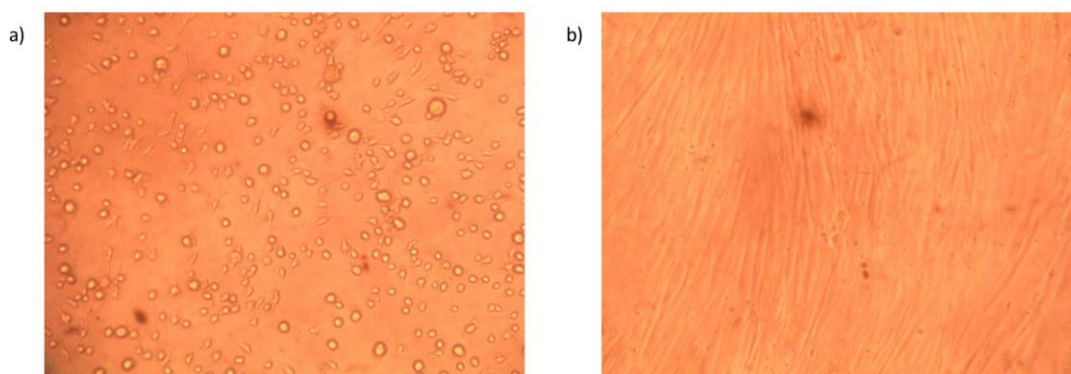


Figure 5.1 Example of the effect of *C. difficile* culture supernatant on Hff-1 in the cytotoxicity assay. HFF-1 cells were incubated with cell free supernatants for 24 hours. a) Supernatant of transconjugant Paloc386 causes CPE and therefore rounding of the cells can be observed, similar to strain 630 Δ erm. b) Supernatant of CD37 causes no CPE.

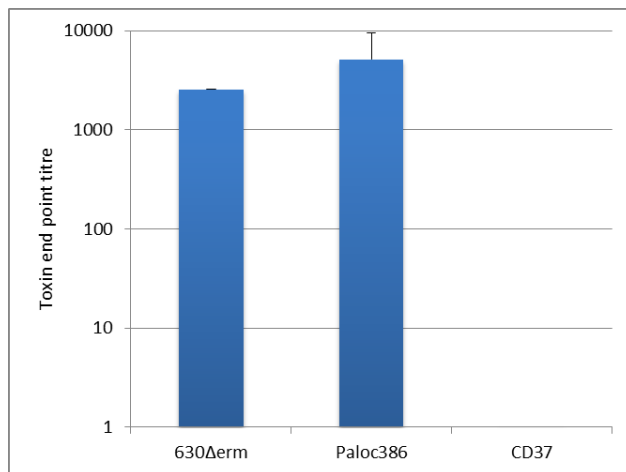


Figure 5.2 Cytotoxicity assay.

HFF-1 cells were incubated with cell free supernatant for 24 h. End toxin titre point was assessed by inverted microscope. The mean is shown for 3 independent experiments, error bars indicate the standard deviation.

5.4.2. PaLoc transfer is reproducible

In order to determine if the PaLoc transfer from 630Δerm into CD37 was reproducible, a *tcdB* Clostron mutant was used as donor in filter-matings using CD37 as the recipient. Rifampicin and lincomycin resistant colonies were screened for the presence of the *tcdA* gene, present on the PaLoc, as well as for the presence of the binary toxin. Ten putative transconjugants that were *tcdA* positive and *cdtA* negative were screened by PCR for the co-transfer of other mobile elements. PCR amplification was performed on the same genes to study co-transfer as described in Chapter 4. An example of transconjugant Paloc 3.5 is shown (Figure 5.3). Co-transfer of CTn5 has occurred in this transconjugant. Co-transfer of Tn5397 was also observed in transconjugant Paloc 3.7. This transconjugant was tested for tetracycline resistance, as Tn5397 contains *tet(M)*, and was tetracycline resistant.

The transfer frequency for the PaLoc into CD37 was calculated at 7.5×10^{-9} transconjugants per donor cell (standard deviation is 4.1×10^{-9}), based on 3 biological replicates each containing at least 3 technical replicates.

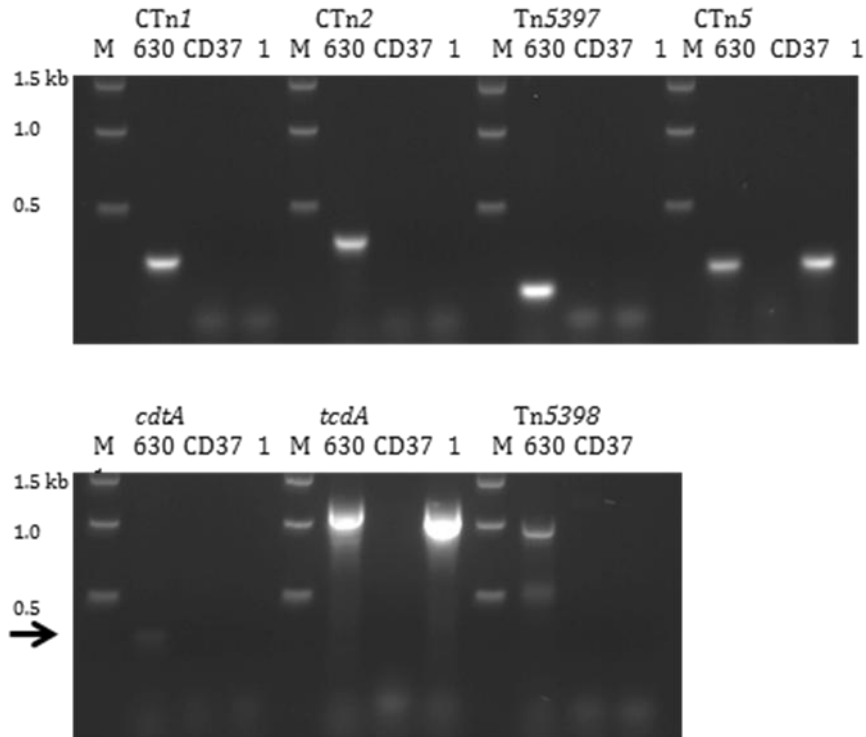


Figure 5.3 PCR analysis for co-transfer of CTn1, CTn2, Tn5397, CTn5 and Tn5398 in the PaLoc transconjugants.

Examples of each of the PCR reactions are shown. Template in lanes 630 is 630 Δ *erm* DNA, lanes CD37 is CD37 DNA, lanes 1 is transconjugant 1 DNA. Lanes M are molecular marker. Primers for the reactions are: CTn1 CD0364 F+R, CTn2 CD0428 F+R, Tn5397 CD0511 F+R, CTn5 CD1873 F+R, *cdtA* BiTox-1F + 1R, *tcdA* F+R, Tn5398 9371F + 9387R. The positive control 630 with primers for *cdtA* resulted in a faint band which is indicated by a black arrow.

5.4.3. Phage assay to check for transfer of PaLoc

As the PaLoc has been proposed to be of phage origin [175], the PaLoc may be transferred via generalised transduction by either of the prophages that are present in strain 630 [8,104]. Concentrated phage suspensions were made by mitomycin C induction of 630 Δ *erm tcdB::ErmRAM*. These phage suspensions were incubated with CD37 and then grown on selective plates to determine if transfer of the PaLoc occurred. No lincomycin resistant transductants were obtained in this phage assay.

To check if CD37 is sensitive for the ϕ C630-1 and ϕ C630-2 phages of 630, a plaque assay was performed with CD37 and CD843, a control strain which is sensitive to both ϕ C630-1 and ϕ C630-2 [175]. The plaque assays resulted in 100-200 plaques per CD843 plate but none on the CD37 plates; this experiment was performed in triplicate. These results indicate that CD37 is not susceptible to ϕ C630-1 and ϕ C630-2 and that the mechanism of transfer of the PaLoc is not through phage transduction by ϕ C630-1 and ϕ C630-2.

5.4.4. Full genome sequencing

In order to be able to predict the mechanism of transfer of the PaLoc and to see if further DNA was transferred, the genomes of the CD37 recipient strain, the Paloc386 transconjugant, containing the whole PaLoc, and 6 other Paloc transconjugants containing the *tcdB* ClosTron mutation were sequenced.

5.4.4.1. CD37 genome shotgun sequence

The sequencing of strain CD37 has resulted in a total of 7,515,781 72bp paired end reads, which is approximately 130-fold coverage of the genome. The sequence of the 245 contigs has been deposited in Genbank as a shotgun sequence under accession number AHJJ01000000 [182]. These contigs were mapped to the 630 reference sequence and annotated. The sequence was analysed using ACT.

As expected, the PaLoc is not present in CD37 and a 115 bp sequence is present at the PaLoc integration site. This sequence is identical to the sequence in strain 7322 [32]. Furthermore, the sequence confirms the previous report that the CDT-locus is not present in CD37 [58].

Analysis of the CD37 sequence demonstrates that the genome contains CTn1-like, CTn4-like and CTn7-like elements as well as a putative prophage [182](see Chapter 3).

5.4.4.2. Sequencing of transconjugants

Whole genome shotgun sequencing was performed on the transconjugant Paloc386 containing the wildtype PaLoc as well as for six of the transconjugants containing the PaLoc with a ClosTron insertion. PCR analysis had previously shown that transconjugants Paloc35 and Paloc37 respectively contain a copy of CTn5 and Tn5397, whereas Paloc386 contains a copy of both of these elements, (see Chapter 4). This was confirmed by the genome sequence of these isolates. Furthermore, the genetic elements described for CD37 were present in each of the transconjugants.

To determine what part of the genome of the transconjugants is CD37 specific and what part is 630 specific, 30 regions of 10 kb were chosen, evenly spaced along the 630 genome (Figure 5.4). Regions were chosen not to include CTns, Tn5398 the *skin*^{Cd} or either of the prophages. These regions were aligned with CD37 and for each region there were 7-123 SNPs between the CD37 and 630 sequences. The sequence of these regions is included in Appendix II. All of the seven PaLoc transconjugants were then aligned with these regions of CD37 and 630 in order to determine the origin of the DNA. All of the sequence was CD37 or 630 specific although some gaps in the sequences were observed. Each of the transconjugants contained 630 specific SNPs in region 6 or 7 and Paloc386 contained 630 specific SNPs in regions 5 and 6. The PaLoc is situated between region 6 and 7 (Figure 5.4). Regions 1-4 and 8-30 were CD37 specific for each of the transconjugants (Table 5.2).

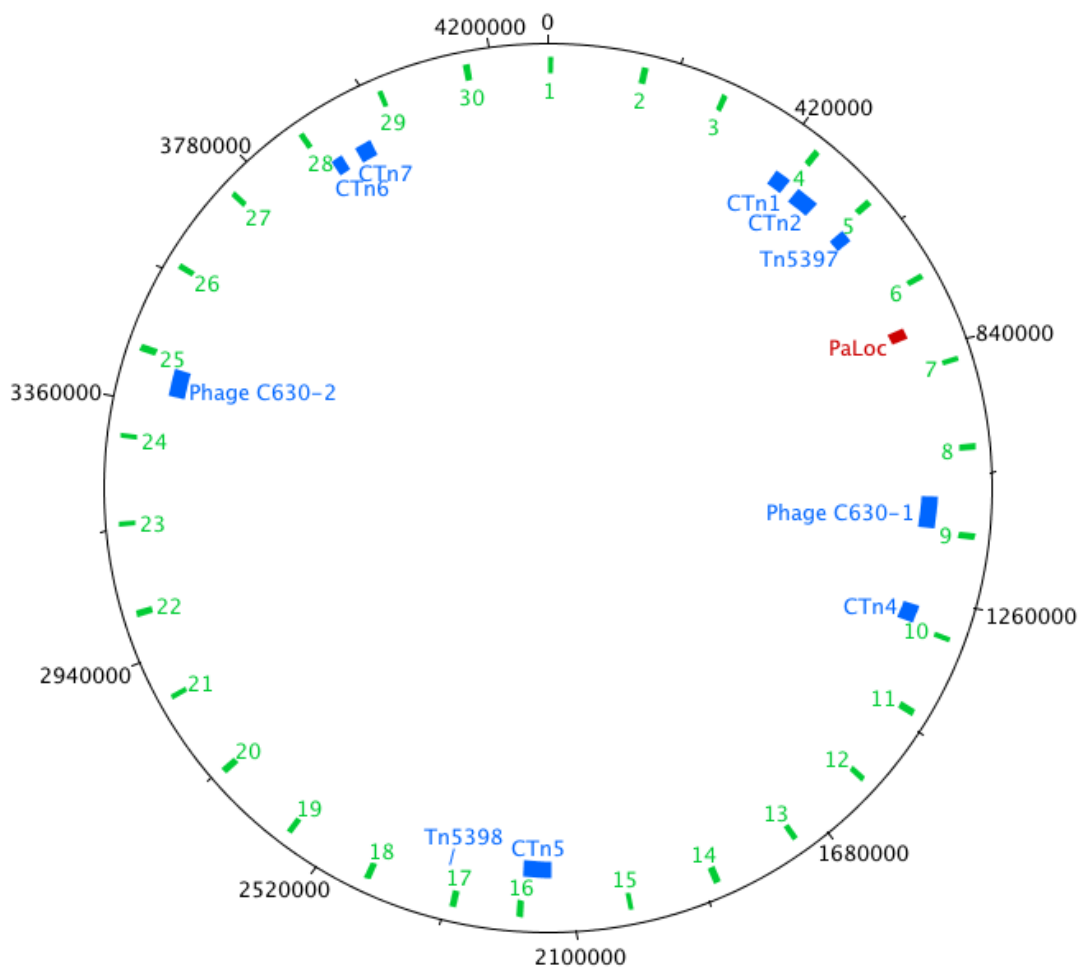


Figure 5.4 Circular representation of the regions used for SNP analysis on the 630 genome.
Regions used for SNP analysis are shown in green. Positions of the MGEs are shown in blue. The PaLoc is shown in red.

Table 5.2 Summary of the results of the SNP analysis in the transconjugants.

Region 5, 6 or 7 contained 630 specific SNPs in each of the transconjugants. The PaLoc is situated between region 6 and region 7. Regions 1-4 and 8-30 were CD37 specific for each of the transconjugants.

Region	Paloc386	Paloc12	Paloc22	Paloc26	Paloc35	Paloc37	Paloc46
4	CD37	CD37	CD37	CD37	CD37	CD37	CD37
5	630	CD37	CD37	CD37	CD37	CD37	CD37
6	630	CD37	630	CD37	630	CD37	630
7	CD37	630	CD37	630	CD37	630	CD37
8	CD37	CD37	CD37	CD37	CD37	CD37	CD37

The sequence of the transconjugants between regions 4 and 8 were aligned with the 630 and CD37 sequences in order to determine the precise size of each of the 630 specific fragments that had transferred. The distance between the first and last 630 specific fragment ranges between 66 kb and 271 kb (Table 5.3). The precise site of the 630 specific fragments in the transconjugant genomes are shown in Figure 5.5. The sequence at the ends of the transferred DNA in each of the transconjugants is included in Appendix II.

Table 5.3 Amount of 630 specific sequence for each of the Paloc transconjugants.

Transconjugant	Size of 630 specific fragment
Paloc386	271992 bp
Paloc12	151692 bp
Paloc22	157688 bp
Paloc26	66034 bp
Paloc35	224973 bp
Paloc37	250027 bp
Paloc46	141658 bp

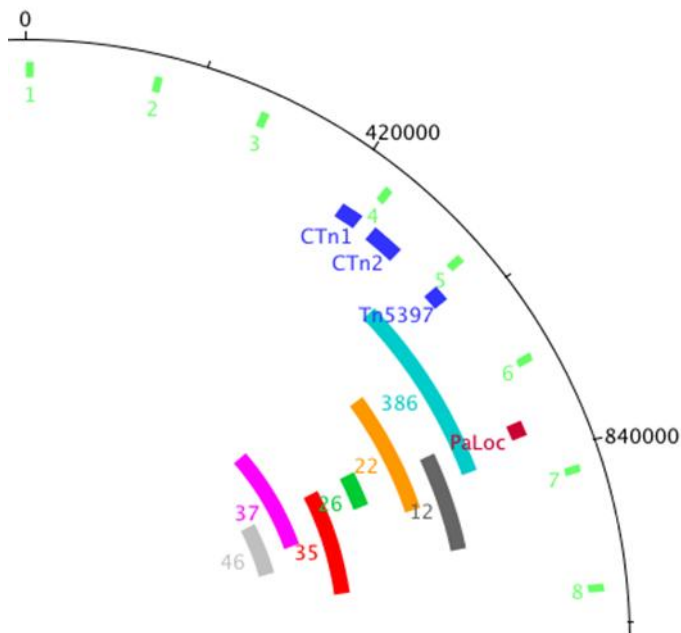


Figure 5.5 Fragments of 630 specific sequence in CD37 transconjugants.

A quarter of the circular representation of the 630 chromosome as shown in Figure 5.4 was magnified to show the locations of the 630 specific sequences present in each of the PaLoc transconjugants. All genetic fragments are labelled. Regions used for SNP analysis are shown in green. Positions of the MGEs are shown in blue. The PaLoc is shown in red. The 630 specific DNA fragment present in each of the transconjugants is shown with the transconjugant number.

5.5. Discussion

5.5.1. Transfer of the PaLoc

Reproducible transfer of the PaLoc has been demonstrated here for the first time. Although the PaLoc has been hypothesised to be a mobile genetic element or a phage remnant [32], the transfer events described here are unlikely to be PaLoc specific as variable amounts of 630 DNA have transferred. Analysing the transconjugants containing the PaLoc could not indicate a specific MGE with which the PaLoc co-transfers, although co-transfer of the PaLoc with CTn1, CTn5 and Tn5397 was demonstrated. Although Tn5397 had transferred on the same transferred fragment as the PaLoc, CTn1 and CTn5 never transferred on the PaLoc fragment and transfer of these elements could be a separate event. Because of the presence of CTn4 and CTn7-like elements in CD37, co-transfer of these two elements cannot be excluded.

The transfer mechanism of the PaLoc was not determined, although a number of known mechanisms are considered to be highly unlikely. Transformation of free DNA has never been observed in *C. difficile* despite multiple attempts in our laboratory and others [105,107]. However, to rule out this possibility, filter-matings will be repeated in our lab on agar plates containing DNase.

The PaLoc could have been transferred via transduction however, the size of the DNA fragments that were shown to transfer makes this unlikely as the transferred DNA is over three times the size of the phage genomes of ϕ C630-1 and ϕ C630-2 of strain 630. Furthermore, CD37 was shown not to be susceptible to the two prophages of 630 and transduction of the PaLoc from phage suspensions could not be shown.

The transfer of the PaLoc could be facilitated by a high frequency recombination (Hfr)-type mechanism. During Hfr, a conjugative element mobilises the chromosome *in cis* by integrating into the chromosome or *in trans* by mobilising another element within the chromosome after which (part of) the chromosome is transferred into a recipient cell [114,115,260]. The transferred fragment of DNA will recombine with the chromosome of the recipient cell through a double-crossover event replacing part of the recipient genome, a process which depends on the recombination protein RecA [104,114,115]. The mechanism has been described in a number of Gram negative bacteria but in Gram positive bacteria it has only been described in *L. lactis* [261]. The *L. lactis* sex factor is chromosomally located and can mobilise the chromosome but it can also transpose through site-specific recombination into non-conjugative plasmids in order to mobilise these [120,121] (Chapter 1).

To determine if the transfer event is Hfr, a number of experiments need to be performed. As Hfr is a RecA dependent event in the recipient strain, a CD37 *recA* mutant could be constructed and used as recipient in filter-matings. Hfr in *C. difficile* 630 would most likely be mobilisation *in cis* by one of the CTns. *In cis* mobilisation would involve using the *oriT* of the CTn that facilitated the transfer which means the element rarely comes across completely as part of the element would be the last sequence to be transferred to the transconjugant. Part of the element that does come across may be lost during the double crossover event that integrates the transferred DNA fragment into the chromosome. Furthermore, if the element responsible for transfer is unique to 630 then recombination would not occur with this sequence and the transferred fragment of the responsible element would be lost. As the whole sequence of Tn5397 was present on the DNA fragment that transferred into CD37 in

two of the transconjugants, it is unlikely that the *oriT* of this element is used as the whole of the chromosome would have to have been transferred into the recipient cell. However, CTn1 and CTn2 are located close to this element and were not present in any of the transconjugants on the integrated DNA fragment (Figure 5.5). Therefore, one of these two elements may facilitate the Hfr transfer. In order to test this hypothesis, mutants could be made for genes that are essential for conjugative transfer of these elements. If genes which are essential for one of the elements are also essential for Hfr, then the element responsible for Hfr is identified. However, more than one of the CTns may be capable of mobilising the genome in this manner.

The PaLoc contains the two major virulence factors that have been identified in *C. difficile*, making transfer of this locus clinically important. *C. difficile* has been shown here, and in previous studies, to exchange DNA with a number of species including *B. subtilis* [164], *E. faecalis* [163] and *S. aureus* [170]. Transfer of the PaLoc into any of these species is therefore theoretically possible and when the transfer mechanism of the PaLoc transfer is determined, experiments should be carried out in order to determine the host range. If the transfer mechanism is shown to be Hfr, transfer into unrelated species would be less likely.

The cytotoxicity assay showed that the CD37 transconjugant has the same level of cytotoxicity as the donor strain. Although horizontally acquired DNA can be silenced after transfer, the PaLoc encodes two hypothetical regulators, *tcdC* and *tcdD*, which may also regulate toxin expression in transconjugants [52,53]. Furthermore, a recent paper by Dingle *et al.* [89] described MLST data which suggests that a toxigenic sequence type-11 strain diverged millions of years ago but had only recently acquired

the PaLoc. This data suggests that the PaLoc transfer that was described here may have also occurred *in vivo*.

Protective colonization with non-toxigenic *C. difficile* strains is a new strategy to prevent colonization by toxigenic *C. difficile*. Protective colonization in hamsters was shown to lead to colonization resistance of toxigenic strains and to prevent disease symptoms [262]. Currently, human phase II trials are being conducted [263] however, if *in vivo* transfer of the PaLoc can occur, this treatment strategy may be less effective than expected and the effect on *C. difficile* epidemiology cannot be predicted.

5.5.2. Conclusions

The data in this chapter describes the transfer of the PaLoc from a toxigenic donor strain to a non-toxigenic recipient strain. The recipient strain becomes cytotoxic and produces similar levels of toxin as the wildtype donor strain.

The transfer mechanism of the PaLoc transfer is still unknown although a number of transfer methods have been ruled out. Transfer via transformation is unlikely due to the natural inability of *C. difficile* to take up DNA under laboratory conditions. Furthermore, the size of the transferred DNA fragment in some of the transconjugants is more than three times larger than the phage genomes of ϕ C630-1 and ϕ C630-2 of strain 630, making transduction of the DNA fragment highly unlikely. As the size of the DNA fragment that was incorporated in the recipient chromosome varies between the transconjugants, the donor DNA could be incorporated by non-specific recombination. This leads to the hypothesis that a Hfr-like transfer mechanism could be responsible for the transfer of the PaLoc into non-toxigenic strains.

Chapter 6 The effects of CTNs on the transcriptome

6.1. Introduction

6.1.1. The effect of MGE acquisition on the transcriptome

Many CTNs contain predicted transcriptional regulators, two-component regulatory systems and alternative $\sigma 70$ sigma factors, as discussed in Chapter 3. These regulators are hypothesised to regulate the genes encoded on the element [138]. However, some elements encode a disproportionately large number of putative transcriptional regulators such as Tn6103 in R20291, which encodes a two-component regulator system and also has the putative mobilisable elements Tn6104, Tn6105 and Tn6106 inserted into it, which encode another five putative transcriptional regulators and five predicted alternative sigma factors (Chapter 3).

The transcriptional regulation system of CTNs has been studied primarily in Tn916 but this is still not completely understood. A mechanism of transcriptional attenuation by *orf12* on Tn916 regulates *tet(M)* transcription, which is de-repressed by the presence of tetracycline [222]. Read-through from this gene into the downstream recombination and conjugation modules is hypothesised to be responsible for the increase in transfer frequency of Tn916 in the presence of tetracycline [127,138].

The ORFs involved in transcriptional regulation of Tn916 are conserved in Tn5397 and these elements induce the loss of the other element upon entering the cell [156]. This loss of a residing copy of Tn916 is not seen when a second copy of Tn916, marked with *erm(B)*, is introduced. Similarly, Tn5397 can stably reside in strain 630 with 5 other CTNs. It is expected that the transcriptional regulators of Tn916 and Tn5397 can act *in trans*, however, differences in the regulation of the excision and integration modules

of these elements cause the loss of the residing element upon introduction of the new element [156].

Another example of MGEs affecting the expression of other MGEs is the *Salmonella* Pathogenicity Island 1 (SPI-1) which is essential for *Salmonella enterica* to invade epithelial cells in the intestine. Recently, the *std* fimbrial cluster, another horizontally acquired genetic element, was shown to regulate the expression of SPI-1 genes and overexpression of the *std* genes resulted in an invasion defect [264].

Besides the effects that transcriptional regulators encoded on MGEs may have on the transcription of genes encoded on other MGEs, MGEs can also affect the general transcriptome of their host cell. In three *Pseudomonas* spp. a comparison of the transcriptome was made using a microarray between plasmid-free strains and strains containing the incompatibility group P7 archetype plasmid pCAR1. Between 73-125 chromosomal genes per host were shown to be differentially expressed when pCAR1 was present. Although most of these genes overlap between the hosts, the effects that the element has on the transcriptome differs between strains [265].

These examples demonstrate that MGEs can have effects *in trans* on the genes of other MGEs and genes of their host chromosome. The transcriptional regulators encoded on the CTns in *C. difficile* may also affect the general transcriptome. As described in Chapter 4, mutations were made for the AraC transcriptional regulator CD0428 encoded on CTn2 and for the two-component transcriptional regulator CD1099 on CTn4. The transcriptomes of these mutants were compared with the transcriptome of the parental strain, 630 Δ *erm*. 630 Δ *erm* is an erythromycin sensitive derivative strain from 630 containing a deletion in the erythromycin resistance conferring element Tn5398 [48]. The strain was derived through repeated sub-

culturing and, apart from this deletion, acquired 21 SNP mutations of which 5 are located in CDS (Dr. S. Kuehne, personal communication). It is unknown if these SNPs have any influence on the phenotype of *630Δerm*. As *630Δerm* is often used for mutational analysis, the transcriptome of this strain was compared to the transcriptome of the wildtype strain 630 to determine the biological relevance of these SNPs. Furthermore, it was attempted to target the Clostron to alternative sigma factors encoded on Tn6104 and Tn6105 in R20291 to compare the transcriptomes of these mutants to wildtype R20291.

6.1.2. RNA-seq

RNA-seq is a novel technique using high-throughput sequencing of cDNA in order to determine the sequences of all transcripts in a cell. Generally, the sequencing data can be used for two purposes. First, the start of the transcripts will allow for the detection of start sites, untranslated regions (UTRs) and small regulatory RNA molecules. Furthermore, the number of reads for each gene can be used to determine the level of transcription for each gene in the transcriptome [266,267].

So far, RNA-seq has been used in a number of prokaryote studies. One of these is the study of the effect of ethanol on the human pathogen *Acinetobacter baumannii* studied to explain a previously reported increase in virulence in the presence of ethanol [268]. Comparison of the cultures grown in the presence and absence of ethanol showed expression of 49 genes was induced by ethanol whereas 21 were repressed. Most of these genes are involved in metabolism, stress response and permeases of efflux pumps [268]. Furthermore, the upregulation of phospholipase C was shown, which is a

known virulence factor in a number of bacteria due to its hydrolase activity of phospholipids [269]. A phospholipase C knockout mutant in *A. baumannii* was made which was shown to have reduced virulence [268].

RNA-seq was used here to study the effects on the transcriptome of specific mutations of transcriptional regulators encoded on CTNs and to analyse differences between the transcriptomes of 630 and 630 Δ *erm*.

6.2.Aims of this chapter:

- Retargeting the Clostron to alternative sigma factors encoded on Tn6104 and Tn6105 in R20291.
- Determining if transcriptional regulators CD0428 and CD1099 encoded on CTn2 and CTn4, respectively, are transcriptionally active.
- Determining the effect on the transcriptome of insertional mutations of CD0428 and CD1099.
- Comparing the transcriptome of strain 630 and 630 Δ *erm*.

6.3. Materials and methods

6.3.1. RT-PCR

RT-PCR was performed using cDNA as template, generated as described in section 6.3.2.4. The RNA that was used for cDNA generation was used as a negative control. PCR was performed as described in Chapter 2. Primers are listed in Table 6.1.

Table 6.1 Primers used for RT-PCR.

Primer	Sequence 5'- 3'	Amplicon
CD0428 F	CATGGAATCTTCGCATCAGA	0.4 kb
CD0428 R	GCGAATGTCCTCACTAACCG	
CD1099 F	GTCCATACCGGGCATCATAA	0.2 kb
CD1099 R	CAGAAAGGACGGGTGTTGTT	

6.3.2. RNA sequencing

Up to the stage of the generation of cDNA, the protocols described here were all performed in an environment that was thoroughly cleaned with RNase Zap (Invitrogen, Paisley, UK). Nuclease free H₂O and nuclease-free barrier tips and plastic ware were used throughout.

6.3.2.1. RNA extraction

Cells were grown for 16 h BHI broth. 2 ml was added to 18 ml of fresh broth and grown for approximately 4 h, to OD₆₀₀ of approximately 0.5 (exponential phase). 10 ml of cells were then added to 20 ml of RNA protect (Qiagen, UK) in an anaerobic incubator. Cells were vortexed, incubated at RT for 5 min and recovered by centrifugation for 10 min at 4500 x g. The supernatant was discarded and the tubes were kept upside down for 1

min on a piece of absorbent paper. Pellets were stored at -20 °C for a maximum of 2 weeks.

RNA was extracted using the Fast RNA pro blue extraction (MP Biomedicals, USA) according to the manufacturer's instructions. RNA pellets were resuspended in 100 µl H₂O. After 5 min of incubation at RT, a 2 µl aliquot was taken for assessment on a Bioanalyzer and the remaining sample was stored at -70 °C.

The RNA was analysed on an Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Nano Kit (Agilent technologies) for quantification and quality assessment according to the manufacturer's instructions. The Agilent 2100 expert software calculated the RNA integrity number (RIN) and the RNA concentration. Where RIN could not be calculated, the electropherogram was used to assess the quality of the sample.

6.3.2.2. DNase treatment

RNA samples were treated with the Ambion DNA-free kit (Invitrogen, Paisley, UK) according to the manufacturer's instructions. Ten microgram RNA was used per reaction, H₂O was added to a total volume of 44 µl. Five microliters of DNase buffer and 1 µl of DNase I (2 units per µl) were added after which samples were incubated at 37 °C for 20 min. Five and a half microliters DNase inactivation reagent was added to the reaction. Tubes were incubated at RT for 2 min after which the samples were centrifuged at 4 °C for 90 sec at 10,000 x g and the RNA was transferred to a fresh tube. EDTA was added to the solution to a final concentration of 1 mM.

6.3.2.3. mRNA isolation

mRNA was isolated using a MICROBexpress kit (Invitrogen, Paisley, UK) according to the manufacturer's instructions. 10 µl of DNA-free treated RNA was used as input containing 2 to 5 µg RNA. Samples were stored at -80 °C and analysed by a Bioanalyzer before these were used for downstream protocols.

6.3.2.4. Generation of cDNA

cDNA was generated using the Ovation RNA-seq System (NuGEN, Bemmell, The Netherlands) according to the manufacturer's instructions. To prepare the samples, H₂O was added to an aliquot of 30 ng of mRNA up to a final volume of 5 µl. cDNA was purified using the Qiaquick PCR purification kit (Qiagen) with minor modifications to the manufacturer's instructions, which are included in the protocol of the NuGEN kit.

6.3.2.5. RNA library preparation

cDNA was fragmented and adaptor molecules added using the Encore NGS Multiplex system I kit (NuGEN, Bemmell, The Netherlands), according to the manufacturer's instructions.

An aliquot of 5 µg cDNA was diluted to a total volume of 120 µl in TE buffer and fragmented in a Covaris S-series sonication system. System settings as recommended by the NuGEN kit: Duty cycle 10%, Intensity level 5, 100 cycles, 5 min. Samples were analysed on the Bioanalyzer and stored at -20 °C.

Fragmented cDNA samples were concentrated using the Qiaquick PCR purification kit as explained above. 50 µl of H₂O heated to 50 °C was loaded to the middle of the column and incubated 5 min at RT to elute the cDNA. After 1 min centrifugation at

16,000 x g the sample was loaded back onto the column and centrifuged again. The samples were quantified on a Nanodrop 1000.

Aliquots of 200 ng fragmented cDNA were diluted to a total volume of 7 μ l in H₂O to prepare the cDNA libraries. The samples were analysed on a Bioanalyzer before sequencing.

6.3.2.6. RNA sequencing

Dr. Tony Brooks performed sequencing of the cDNA libraries at UCL genomics, the samples were sequenced as a 72 bp paired end multiplex reaction on an Illumina GAIIX.

Dr. Warren Emmett assembled the RNAseq data. Custom de-multiplexing was performed to sort the reads to the specific samples. The reads were filtered to identify exact duplicates after which they were aligned against the reference genome 630 (AM18035) using the BWA software [270]. HTseq software [271] was used to generate counts of the reads that aligned specifically to 630 genes. The read counts were processed using the DEseq package and comparisons were made between the transcription levels of genes in strains 630, 630 Δ *erm*, 630 Δ *erm* CD0428::*ErmRAM* and 630 Δ *erm* CD0428::*ErmRAM* [272].

6.4. Results

6.4.1. Transcription of CD0428 and CD1099

Before RNA-seq was performed on strains containing the mutated transcriptional regulators CD0428 on CTn2 and CD1099 on CTn4, the expression of these genes was studied. cDNA was produced from RNA samples of 630 in exponential growth phase. This cDNA was used as template for RT-PCR for these two genes using primer pairs CD0428 F+R and CD1099 F+R. mRNA was used as a negative control to determine the presence of DNA. Both genes are transcribed in 630 during exponential growth (Figure 6.1) and the mutant versions of these genes were selected for the RNA-seq study in which the transcriptome of the mutants was compared with that of the parental strain.

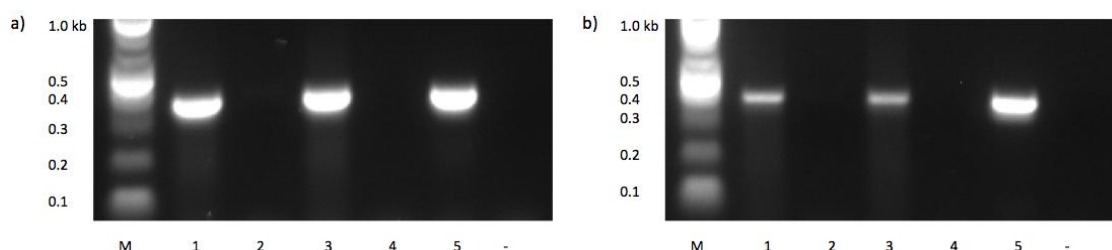


Figure 6.1 RT-PCR for CD0428 and CD1099.

RT-PCR was performed on cDNA produced from RNA of strain 630 cells in exponential growth. a) RT-PCR for CD0428 using primer pair CD0428 TSF+TSR, expected product size 400 bp. b) RT-PCR for CD1099 using primer pair CD1099 TSF+TSR, expected product size 400 bp. Template in lanes 1 & 3 is 630 cDNA, lanes 2 & 4 is 630 mRNA, lane 5 is 630 genomic DNA, lane – is H₂O, lane M is a molecular marker. Size of the fragments is shown left of the figure.

6.4.2. ClosTron mutants in R20291 alternative sigma factors

R20291 Tn6103 contains insertions of the putative mobilisable elements Tn6104, Tn6105 and Tn6106, which encode five predicted alternative sigma factors. The ClosTron was retargeted to three of these genes; Tn6104 CDS1748 and CDS1754 and

to Tn6105 CDS1770 (Figure 6.3). In parallel with these, the Clostron was also retargeted to CDS1803 on Tn6103, as described in Chapter 4. The Clostron plasmids were conjugated from *E. coli* into R20291 which resulted in thiamphenicol resistant colonies. These were shown to contain the unspliced ErmRAM as expected (Figure 6.2). Although the gel bands appear to have shifted a little in lanes 2 and 3 for an unknown reason, all of these products are expected to be Clostron plasmid products. Culturing of these isolates on lincomycin only resulted in lincomycin resistant colonies for CDS1803 (Chapter 4) and no Clostron mutants were obtained for CDS 1748, CDS1754 or CDS 1770 encoded on Tn6104 and Tn6105.

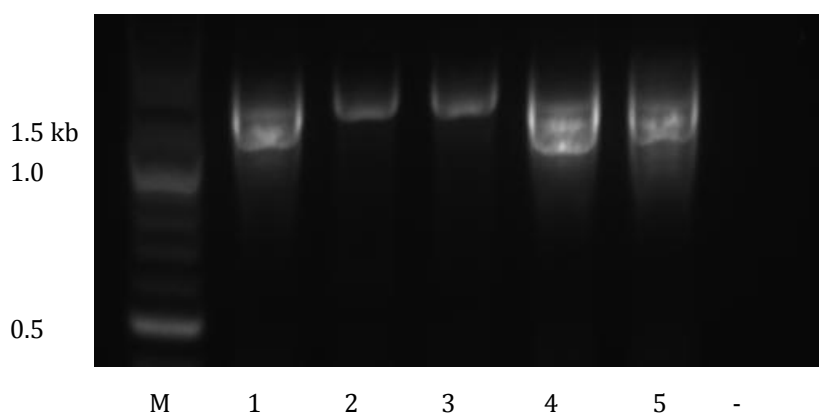


Figure 6.2 PCR for presence of retargeted Clostron plasmids in R20291.

Colony PCR was performed on thiamphenicol resistant R20291 isolates to confirm presence of the plasmid containing the retargeted Clostron. The primer pair in all reactions was ErmRAM F+R and the expected product size was 1.2 kb. The template in lane 1 was R20291 containing the Clostron plasmid retargeted to CDS1754, lanes 2, 3 and 4 was R20291 containing the Clostron plasmid retargeted to CDS1748 and lane 5 was R20291 containing the Clostron plasmid retargeted to CDS1803. The template in lane – is H₂O, lane M is a molecular marker. Size of the fragments is shown left of the figure.

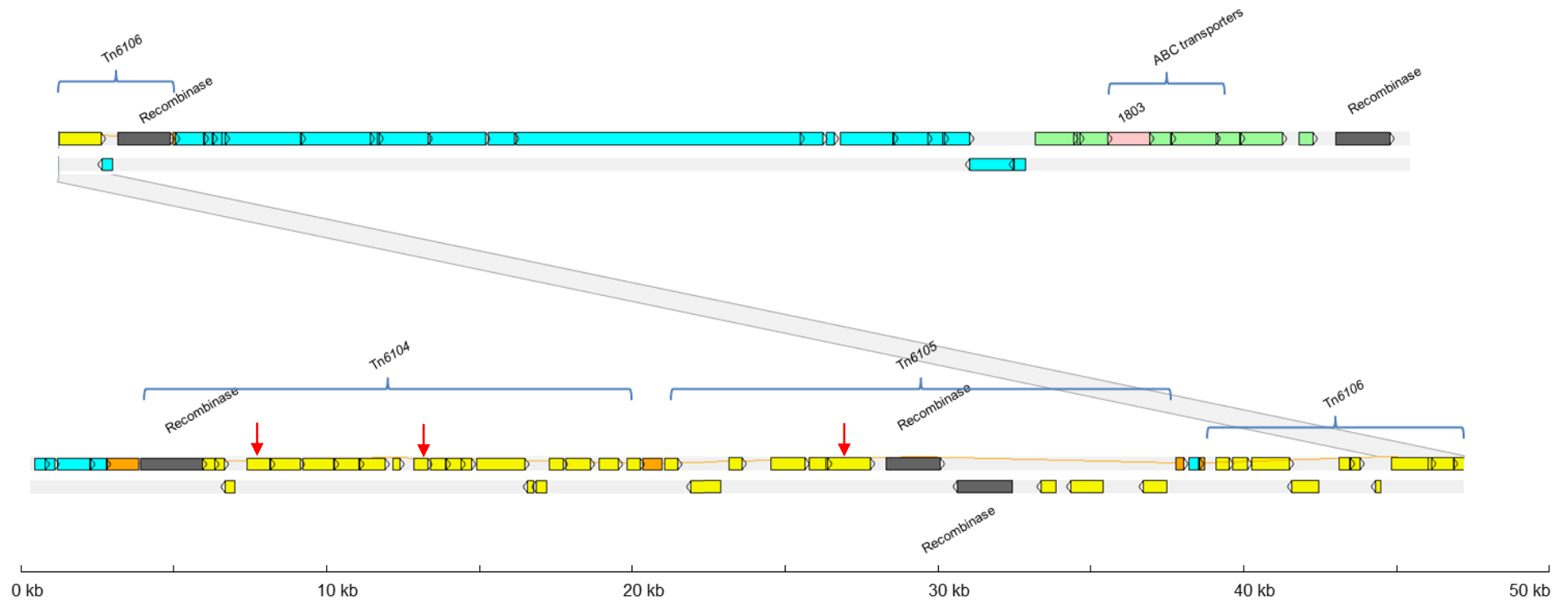


Figure 6.3 Attempts to obtain Clostron insertions in Tn6104 and Tn6105.

The schematic of Tn6103 is shown as two parts which should be joined at the grey diagonal line. The targets for Clostron insertions on Tn6104 and Tn6105 are indicated by red arrows. ORF 1803 that was shown to contain the Clostron insertion is shown in pink. ORFs of Tn6104, Tn6105 and Tn6106 are shown in yellow. ORFs on the recombination module of Tn6103 and the recombinases of Tn6104, Tn6105 and Tn6106 are shown in grey. ORFs on the conjugation module of Tn6103 are shown in blue. ORFs on the accessory module are shown in green. ORFs that were interrupted by an insertion of a DNA fragment, compared to CTn5, are shown in orange.

6.4.3. RNA-seq results

6.4.3.1. Biological variation between the samples

RNA from strains 630, 630 Δ erm, 630 Δ erm CD0428::ErmRAM (CTn2 mutant) and 630 Δ erm CD1099::ErmRAM (CTn4 mutant) was isolated and reverse transcribed into cDNA. The cDNA was fragmented into 150 to 400 bp fragments as measured by the Bioanalyzer (Figure 6.4). Adaptor ligation was carried out after which sequencing was performed by Dr. Tony Brooks at UCL Genomics.

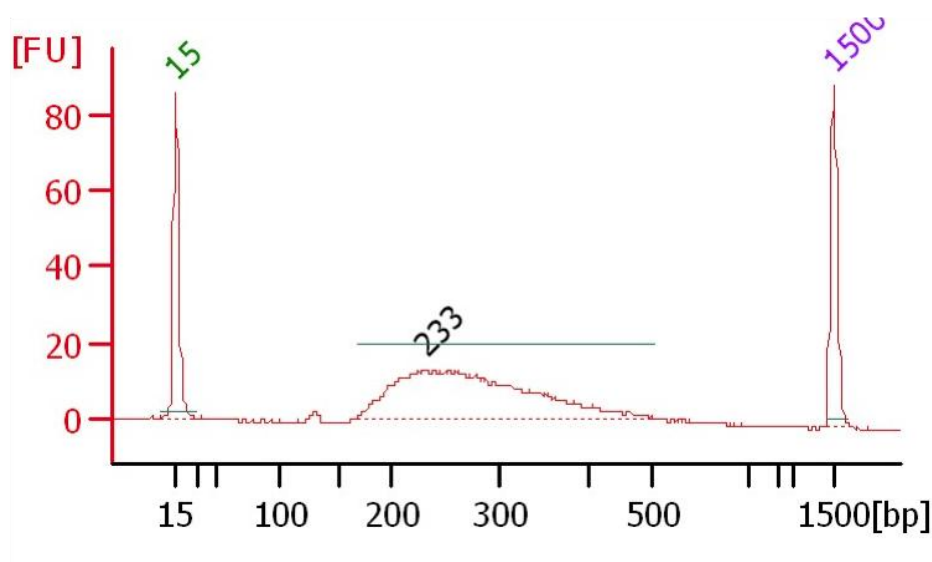


Figure 6.4 Size distribution of cDNA samples after fragmentation for RNA-seq.

The Bioanalyzer measures the intensity of fluorescence over time to detect the size of each DNA fragment in the sample. This is plotted as fluorescent units on the y-axis and number of basepairs on the x-axis. The peaks at 15 and 1500 bp show the calibration controls in the sample. The horizontal line above the sample peak shows the region of the distribution over which the sample concentration was calculated. The sample concentration is shown at the top of the distribution in ng/ μ l.

Initial analysis of the RNA-seq data was performed by Dr. Warren Emmet using a number of bioinformatics packages. The DEseq package was used to process read counts and perform statistical analysis [272]. The expected variation between technical replicates and the actual biological variation between the RNA samples are shown in

Figure 6.5. These are calculated following a data-driven algorithm that looks at the relationship between the mean and standard deviation of the transcription of each gene [272]. The variation is expected to be high for low transcription counts but is expected to decrease for higher counts as seen for sample 630G (purple graph in Figure 6.5), which is a sample that was prepared in parallel for another study. All other samples show high variation between the replicates for the high transcription counts. The plot demonstrates that significant biological variation between the replicates of each sample is present. Adjusted p-values for the biological variation are calculated for each gene by the package. These adjusted p-values will be high for the genes where variation between the replicates is high which may result in differential expression of genes being false-negative. This could be corrected for by including a higher number of biological replicates, however, these could not be included in this study.

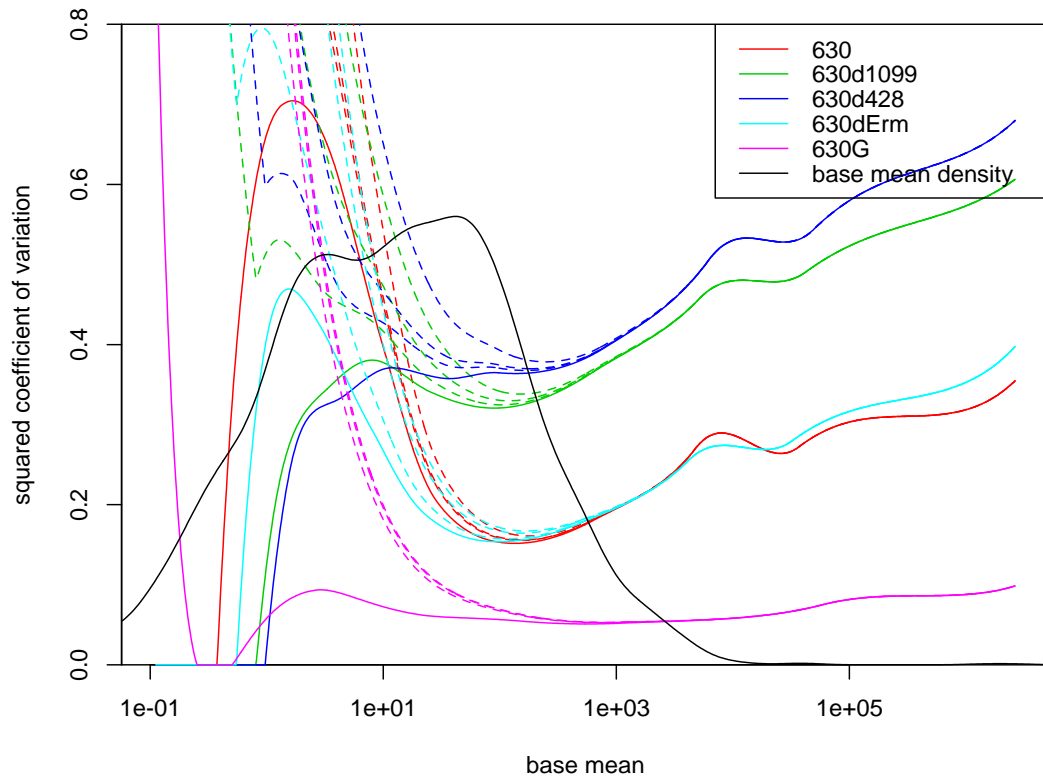


Figure 6.5 Plot for the variation between RNA-seq samples.

The squared coefficient of variation is plotted against the mean of transcription levels of the RNA samples. The y-axis show the expected technical variation based on the samples, the x-axis shows the read count of the transcripts. The dotted lines depict the expected variation between technical replicates whereas the solid lines show the biological variation between the samples. The legend in the figure shows which colour belongs to each sample. 630G are RNA samples for another study, which are included to show the expected variation between biological samples.

6.4.3.2. Differential expression of the CTn2 mutant

The transcriptome of the CTn2 AraC transcriptional regulator mutant, 630 Δ *erm* CD0428::*ErmRAM*, was compared to the transcriptome of 630 Δ *erm* (Table 6.2). The transcription reads of the *erm(B)* gene on the *ErmRAM* in the CTn2 mutant were aligned to the two *erm(B)* copies on Tn5398 in 630 by the bioinformatics software. Although only the non-functional *erm(B1)* is present on Tn5398 in this strain, the differences observed in transcription between both *erm(B)* genes can be ignored as these are false positive results generated by the *Clostron* insertion.

In the CTn2 mutant, a predicted proline iminopeptidase, CD1997, was upregulated compared to 630Δ*erm*. Bioinformatics analysis of the predicted protein indicates domains with peptidase activity. Furthermore, the orphan response regulator CD2047 was upregulated in the CTn2 mutant. Both of these genes are not located on or near CTn2 in the chromosome. As the precise functions of these genes are unknown, no predictions can be made towards the effects of the differential expression of these genes between the mutant and 630Δ*erm*.

Table 6.2 Differential expression of 630Δ*erm* CD0428::ErmRAM vs 630Δ*erm*.

Fold change was calculated as expression in 630Δ*erm* CD0428::ErmRAM divided by expression in 630Δ*erm*. p-adj is the adjusted p-value.

Gene	Function	p-adj	Fold change
CD1997	Proline iminopeptidase	3.52×10^{-2}	11.15
CD2007	Tn5398 <i>erm</i> (B2)	1.81×10^{-8}	31.64
CD2010	Tn5398 <i>erm</i> (B1)	8.47×10^{-8}	23.14
CD2047	2-component response regulator	6.97×10^{-6}	18.55

6.4.3.3. Differential expression of the CTn4 mutant

The transcriptome of the CTn4 two-component transcriptional regulator mutant, 630Δ*erm* CD1099::ErmRAM, was compared to the transcriptome of 630Δ*erm* (Table 6.3). Transcription of the *erm*(B) gene on the ErmRAM in the CTn4 mutant was aligned to the two *erm*(B) copies on Tn5398, although only the non-functional *erm*(B1) is present on Tn5398 in this strain. The orphan response regulator CD2047 was upregulated in the CTn4 mutant compared to 630Δ*erm*, similar to the CTn2 mutant. As

the function of the gene is unknown, no predictions can be made towards the effects of the differential expression of CD2047 between the mutant and 630 Δ *erm*.

Table 6.3 Differential expression of 630 Δ *erm* CD1099::ErmRAM vs 630 Δ *erm*.

Fold change was calculated as expression in 630 Δ *erm* CD0428::ErmRAM divided by expression in 630 Δ *erm*. p-adj is the adjusted p-value.

Gene	Function	p-adj	Fold change
CD2007	Tn5398 <i>erm</i> (B2)	1.59×10^{-8}	30.76
CD2010	Tn5398 <i>erm</i> (B1)	1.59×10^{-8}	25.07
CD2047	2-component response regulator	1.62×10^{-5}	13.68

6.4.3.4. Differential expression of 630 vs 630 Δ *erm*

The differential mRNA expression of strain 630 Δ *erm* compared to 630 is summarised in Table 6.4. 630 Δ *erm* is a derivative of strain 630 which is erythromycin sensitive due to a deletion in Tn5398. This deletion has also affected the transcription of the adjacent ORFs *erm*(B1), *orf3a* and *orf13* (Table 6.4). Transcription of the other ORFs encoded on Tn5398 has not changed: minimal transcription is measured for *effD*, *orf7* and *orf9* while the predicted transcriptional regulator *effR* is highly transcribed in both strains (data not shown).

Other genes that were differentially expressed between 630 and 630 Δ *erm* are all four genes of a predicted oxidative stress operon (CD0825-CD0828) which were upregulated in 630 Δ *erm*. Rubrerythrin is an iron binding protein first identified in the anaerobic *Desulfovibrio vulgaris* of which no particular function has been described yet [273]. The ferric uptake regulator (Fur) has been shown to regulate iron uptake of the cell in the facultative anaerobe *Pseudomonas aeruginosa* where it also regulates a

number of other cellular functions [274]. Desulfoferrodoxin is another iron binding protein from *D. vulgaris* which has oxidoreductase activity [275]. CD0828 contains two iron-binding domains with predicted oxidoreductase activity, which has been annotated as an oxidative stress protein [8], although it seems there is no experimental proof of this function.

The orphan response regulator CD2047 is downregulated in *630Δerm*, however, it is unknown what the function of the regulator is. Furthermore, the trehalose-6-phosphate hydrolase *treA* is upregulated in *630Δerm*, which may have a function in sugar metabolism.

Table 6.4 Differential expression of 630 vs 630Δerm.

Fold change was calculated as expression in *630Δerm* divided by expression in 630. 0 means no expression was detected in *630Δerm*. p-adj is the adjusted p-value.

Gene	Function	p-adj	Fold change
CD0825	Rubryerythrin	7.70×10^{-8}	25.15
CD0826	Fur transcriptional regulator	4.61×10^{-7}	17.34
CD0827	Desulfoferrodoxin	4.67×10^{-6}	18.93
CD0828	Oxidative stress protein	1.89×10^{-7}	21.91
CD2005	Tn5398 ORF13	4.08×10^{-7}	0.06
CD2006	Tn5398 ORF3b	4.30×10^{-5}	0.03
CD2007	Tn5398 <i>erm</i> (B)2	4.15×10^{-13}	0.02
CD2008	Tn5398 ORF298	1.48×10^{-16}	0.01
CD2009	Tn5398 ORF3a	3.07×10^{-5}	0
CD2010	Tn5398 <i>erm</i> (B)1	9.26×10^{-13}	0.03
CD2047	2-component response regulator	2.34×10^{-5}	0.10
CD3091	Trehalose-6-phosphate hydrolase	8.62×10^{-5}	7.21

6.5. Discussion

6.5.1. Design of the experimental methods

Although great care was taken to treat the samples in identical conditions and the cultures were treated with RNA Protect to prevent transcriptomic effects by oxygen shock or RNA degradation during centrifugation of the culture or other stages of RNA extraction, the RNA-seq results described here suffered from high biological variation between the triple replicates of each sample,

Adjusted p-values were calculated for each differentially expressed gene and only those genes with an adjusted p-value ≤ 0.05 were described here. Most RNA-seq studies that were described in the literature so far have used different methods for the comparison of samples. For instance, Camarena *et al.* [268] used RNA from two replicates but combined the reads before analysis. Effectively, this means there are no controls for variation between the samples, which eliminates the possibility to adjust the p-value for the biological variation between the replicates and p-values of these results may therefore be less significant than as they are presented. Dotsch *et al.* [276] used a similar approach as used in the study described here, using the DEseq package for statistical analysis of the biological variation of the samples, however, this study only uses duplicate samples for each of the conditions that was studied. Although the data from these studies is still valid, it is less robust than studies that use a higher number of samples.

Although using less biological replicates would have resulted in lower biological variation in the study described here, the results would have been less significant. To reduce the biological variation between samples a higher number of replicates could

be used. Furthermore, our samples were sequenced in multiplex reactions with 4 samples per sequencing lane. Using the same number of replicates but fewer samples per lane would increase the cost of the study but could reduce the biological variation that was observed in this study.

6.5.2. Effects of CTn mutations on the transcriptome

No significant differences were observed between the transcriptomes of *630Δerm* and the CTn2 and CTn4 transcriptional regulator mutants. Both mutants appeared to upregulate CD2047 compared to *630Δerm*, however, the comparison of *630Δerm* to 630 shows a downregulation in *630Δerm*. Comparison of transcription of CD2047 between 630 and both mutants shows no significant difference (data not shown) which could indicate the effects on CD2047 could be an artefact in the transcriptomic data for *630Δerm*.

The lack of differential expression in the transcriptome may have been caused by the variation that was measured between the biological replicates. However, other factors could have contributed to this such as the type of media that was used to culture the strains and the growth phase in which the RNA was isolated. Although these regulators appear to have little effect on the transcriptome, further study of other transcriptional regulators could identify CTns that affect the transcription of genes in the chromosome.

In order to study the effects of predicted sigma factors encoded on Tn6104 and Tn6105 in R20291, it was attempted to make Clostron mutants with the Clostron 2.0 system. Although several attempts were made, none of the mutants were generated. Although these predicted sigma factors are encoded on MGEs, it is possible that these have become essential, or very important, genes for the survival of the strain.

However, other research groups have confirmed difficulty of generating mutants in R20291 (personal communication with Dennis Bakker, LUMC, and Melissa Martin, LSHTM) and choosing different target sites within these genes may result in the mutations.

6.5.3. Comparison of the 630 and 630 Δ *erm* transcriptomes

The erythromycin susceptible strain 630 Δ *erm*, derived from 630, has a deletion in the genetic element Tn5398 which includes *orf298*, *orf3b* and *erm(B2)* [48]. This deletion has led to the loss of erythromycin resistance but no other phenotypical differences have been described between the strains and 630 Δ *erm* is now a commonly used laboratory strain as it was the first strain in which targeted mutagenesis using the Clostron system was demonstrated [241]. RNA-seq was used to compare the transcriptomes of 630 and 630 Δ *erm* to determine if secondary mutations in 630 Δ *erm* may have caused transcriptomic differences between these strains. In contrast with 630, the transcription of the ORFs adjacent to the deletion in Tn5398, *erm(B1)*, *orf3a* and *orf13*, was minimal. Transcription of the other ORFs encoded on Tn5398 has not changed.

Differential gene expression was also observed in 630 Δ *erm* for some genes outside of Tn5398. A predicted oxidative stress operon was upregulated in 630 Δ *erm* spanning CD0825-CD0828 which encodes a putative ferric uptake regulator (*fur*) and three iron binding proteins with predicted oxido-reductase activity. The sequence of 630 Δ *erm* shows there is a threonine to alanine mutation in codon 41 in the sequence of Fur which may be responsible for the differential expression of the operon. This mutation was not found in the sequence of any of the genomes of natural isolates that have been sequenced.

Iron is an essential nutrient for pathogenic bacteria and iron acquisition in Gram-negative pathogens has been studied thoroughly [277]. Although less is known of iron uptake by Gram-positive bacteria, the process is regulated by Fur which also plays a role in regulation of further cellular processes [277,278]. Although the upregulation of *fur* would be expected to lead to great differences in the transcriptome, only differential expression of a small number of genes was observed here. Perhaps further differences in expression were not measured because of the biological variation in the samples that was discussed above.

The trehalose-6-phosphate hydrolase *treA* (CD3091) is upregulated in *630Δerm*, compared to 630. *treA* can hydrolyse trehalose-6-phosphate into glucose and glucose-6-phosphate which can be used by the cell as an energy source [279]. Changes in the sugar metabolism of the cell could prove to be advantageous adaptations for growth in laboratory conditions.

6.5.4. Conclusions

Although great care was taken to culture all cells for RNA isolation in identical conditions, significant biological variation was observed in the results described here. In future experiments, this variation will have to be eliminated in order to obtain higher quality results.

The only significant difference between the mutants and the parental strain *630Δerm* was the upregulation of the proline iminopeptidase for the mutant of the regulator on CTn2. However, some effects may have been masked by the biological variation between replicate samples and no definite conclusion can be drawn yet.

The results presented in this chapter demonstrate the importance of comparing results for Clostron generated mutants in *630Δerm* to a *630Δerm* control instead of to *630* wildtype. Further study of the differentially expressed genes between *630* and *630Δerm* will have to establish what precise effects these changes in the transcriptome may have on the physiology of the cell.

Chapter 7 General conclusions

C. difficile is a nosocomial pathogen with very low genome conservation which is hypothesized to be due to the large number of MGEs that all sequenced strains contain [15,88]. The bioinformatics study here of 15 sequenced genomes of *C. difficile* resulted in the discovery of 34 novel putative CTNs and mobilisable transposons (Chapter3). It was expected that these elements would encode for some antimicrobial resistance genes, however, putative resistance genes were only found on 7 of these elements. Further analysis showed that accessory regions of the elements described here mainly encode ABC-transporters. Furthermore, most of these elements encode a number of transcriptional regulators, sigma factors and two-component regulatory systems. It was hypothesised that a number of these may have effects on genes outside of the CTn on which they are encoded, and acquisition of these CTNs could lead to phenotypical changes including up or down regulation of virulence factors.

This was tested by comparing knockout mutants for transcriptional regulators encoded on CTn2 and CTn4 generated in strain 630 Δ *erm* to the wildtype 630 Δ *erm* using RNA-seq (Chapter 6). No significant differences were measured between the CTn4 mutant and only upregulation of a predicted proline iminopeptidase was observed for the CTn2 mutant. Additionally, 630 Δ *erm* was also compared to the wildtype 630 and showed significant upregulation of all four genes from an oxidative stress operon including the Fur transcriptional regulator which is involved in diverse cellular functions in other species [274]. As Tn5398 is located away from this operon in the genome, it is unlikely that the deletion of *erm*(B) from this element would cause these effects. The difference in transcription of this operon between 630 and 630 Δ *erm* could

have effects on virulence and needs to be considered by researchers that currently use the strain to produce mutants for virulence and other phenotypic studies.

A previous study described a putative collagen binding protein at the ends of CTn1 and CTn7 [8]. Bioinformatics analysis has shown this gene to be present on CTn1-like elements in 16 out of 18 strains of *C. difficile* that were studied here (Chapter 3). Additionally, a recent study by He *et al.* [280] showed that CTn1-like elements were present in all 151 ribotype 027 strains sequenced for a global epidemiological study of this ribotype. The putative cell surface gene has significant homology to a known collagen binding protein from *S. aureus* [226] but does not contain the collagen binding domains and the function of the gene in *C. difficile* is unknown. Clostron insertion in these genes in CTn1 and CTn7 did not stop conjugative transfer and the protein is assumed to have an accessory function (Chapter 4). The fact that the gene has been conserved on so many related CTns indicates it may confer some advantages to the host, however, the function remains currently unknown.

Excision of a CTn from the genome into a circular intermediate is hypothesized to be the first step before conjugative transfer [127]. Circular molecules were detected for 11 elements in six different strains (Chapter 4). Using the Clostron system, an *erm(B)* gene was introduced into the elements of strains 630 and R20291 that excised. This allowed for the selection of erythromycin resistance during filter-mating assays for the transfer of these elements. Transfer was shown from strain 630 Δ *erm* into *C. difficile* CD37 for all five elements tested; CTn1, CTn2, CTn4, CTn5 and CTn7. Additionally, CTn5 also transferred to *B. subtilis*. Tn6103 from R20291 did not transfer in the studies

reported here, possibly due to three insertions that are present within genes of the conjugation module.

Transconjugants of these filter-matings were analysed to determine if the elements can integrate in multiple target sites in CD37 and to see if multiple copies of the elements are present in the transconjugants. CTn1, CTn2 and CTn5 were present in transconjugants at several different integration sites in the genome. CD37 contains a CTn4 and a CTn7-like element [182], which complicates the analysis of the transconjugants of these elements. CTn7 at least does have multiple target sites in the genome in CD37 transconjugants. Further analysis could be performed if the genomes of the transconjugants were sequenced. Further study of these transconjugants looked into co-transfer of the CTns. As only one element per donor contained *erm(B)*, only this element is selected during filter-matings. However, CTn2, CTn5 and Tn5397 were shown to co-transfer without any selection. The Tn5397 transconjugants were tetracycline resistant, as expected, which demonstrates the clinical importance of research into this type of (co-)transfer to better understand which factors may induce such events.

As CTns are very promiscuous elements, they can spread to other species and disperse antibiotic resistance genes and other genes which may be beneficial to pathogens. If factors that induce the transfer of CTns are identified then spread may be prevented or minimised. Specifically, this could be beneficial when CTns are identified in clinical samples. If the transfer inducing substances are present in particular foodstuffs or medicine, changing the patient's diet to exclude these substances could prevent dissemination of CTns in the gut.

In one of the transconjugants, transfer of the PaLoc into the recipient had occurred. The transconjugant was shown to produce similar levels of toxins as the donor. PaLoc transfer was shown to be reproducible by selecting for transfer using a mutant containing *erm(B)* in the PaLoc.

Genomic sequencing of a set of the transconjugants showed that large fragments of donor-specific sequence is present on either side of the PaLoc which indicates the transfer is not specific for the PaLoc. The donor-specific fragment of DNA varies in size in the transconjugants leading to the hypothesis that the DNA is integrated by RecA-dependent homologous recombination rather than a site-specific recombination mechanism. Hfr is a RecA-dependent transfer mechanism described in *E. coli* where chromosomal DNA is mobilised by the integration of a conjugative plasmid after which the transferred DNA is recombined with the recipient chromosome [104]. To confirm if the PaLoc transfer occurs through a Hfr-like transfer mechanism, a RecA knockout recipient strain will have to be generated. As CTn1 and CTn2 are located close to the transferred sequence in the chromosome, either of these elements could be responsible for transfer. Generating mutants of these elements that are incapable of conjugative transfer could confirm this. However, involvement of the other CTNs cannot be excluded. Performing filter-matings with donor strains that contain selection markers at random locations throughout the chromosome would determine if the transfer is specific for the region of the chromosome that contains the PaLoc.

Further work is currently being carried out to determine if this transfer is specific for these donor and recipient strains or if transfer can also occur between other strains. Transfer from the PaLoc from donor strain R20291 was observed by Dr. Haitham

Hussain, UCL-EDI , as well as transfer into the non-toxigenic recipient strain OX-904 (personal communication).

Transfer of the PaLoc and transformation of non-toxigenic strains into pathogenic, toxin producing strains could have great clinical implications. Little research is carried out on non-toxigenic strains and it is therefore unknown if these strains have additional virulence factors, such as adhesions, that have allowed these strains not to be outcompeted by toxigenic strains.

Furthermore, the PaLoc of hypervirulent strains contains mutations in the negative regulator TcdC which results in more expression of the toxins which may be responsible for the increased virulence of these strains. Transfer of the PaLoc from these strains into non-hypervirulent toxigenic strains could result in replacement of the native PaLoc by the hypervirulent PaLoc and could potentially increase the strain's virulence. Finally, transfer of the PaLoc *in vivo* to other species is potentially also possible. If transfer is Hfr mediated than the species would have to be related for homologous recombination to occur making it an unlikely scenario, however, the possibility cannot be excluded and could result in a highly dangerous pathogen.

Although PaLoc transfer has currently only been reported under laboratory conditions, it was previously reported that the PaLoc of strains from ribotype 078 are highly similar to the PaLoc from divergent ribotypes [89]. It was suggested that the PaLoc in these strains was introduced after they diverged from a common ancestor which is in line with our findings here. This further stresses the clinical importance of the findings as it eliminates the question if PaLoc transfer could occur *in vivo*.

This study has demonstrated that CTNs are present in all strains of *C. difficile* that were sequenced so far and 34 novel putative CTNs and mobilisable transposons were described here. CTNs were shown to transfer when selection pressure was present but co-transfer of elements that were not selected also took place, for instance generating tetracycline resistant transconjugants in the absence of tetracycline. Furthermore, transfer of the PaLoc encoding the major virulence factors of the organism through an undetermined mechanism was shown to transform non-toxigenic into toxigenic strains, possibly revealing an important role for mobile genetic elements in *C. difficile* virulence.

Chapter 8 References

1. Hall IC, O'Toole E (1935) Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, *Bacillus difficilis*. . Am J Dis Child 49: 390-402.
2. Snyder ML (1937) Further Studies on *Bacillus difficilis* (Hall and O'Toole). The Journal of Infectious Diseases 60: 223-231.
3. Smith LD, King EO (1962) Occurrence of *Clostridium difficile* in infections of man. J Bacteriol 84: 65-67.
4. Viscidi R, Willey S, Bartlett JG (1981) Isolation rates and toxigenic potential of *Clostridium difficile* isolates from various patient populations. Gastroenterology 81: 5-9.
5. Bartlett JG, Moon N, Chang TW, Taylor N, Onderdonk AB (1978) Role of *Clostridium difficile* in antibiotic-associated pseudomembranous colitis. Gastroenterology 75: 778-782.
6. Bartlett JG, Onderdonk AB, Cisneros RL, Kasper DL (1977) Clindamycin-associated colitis due to a toxin-producing species of *Clostridium* in hamsters. J Infect Dis 136: 701-705.
7. Wust J, Sullivan NM, Hardegger U, Wilkins TD (1982) Investigation of an outbreak of antibiotic-associated colitis by various typing methods. J Clin Microbiol 16: 1096-1101.
8. Sebahia M, Wren BW, Mullany P, Fairweather NF, Minton N, et al. (2006) The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet 38: 779-786.
9. Warny M, Pepin J, Fang A, Killgore G, Thompson A, et al. (2005) Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. Lancet 366: 1079-1084.
10. McDonald LC, Killgore GE, Thompson A, Owens RC, Jr., Kazakova SV, et al. (2005) An epidemic, toxin gene-variant strain of *Clostridium difficile*. N Engl J Med 353: 2433-2441.
11. Kuijper EJ, Coignard B, Tull P (2006) Emergence of *Clostridium difficile*-associated disease in North America and Europe. Clin Microbiol Infect 12 Suppl 6: 2-18.
12. Commision TH (2006).
13. Stabler RA, Gerding DN, Songer JG, Drudy D, Brazier JS, et al. (2006) Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. J Bacteriol 188: 7297-7305.
14. Stabler RA, He M, Dawson L, Martin M, Valiente E, et al. (2009) Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol 10: R102.
15. Scaria J, Ponnala L, Janvilisri T, Yan W, Mueller LA, et al. (2010) Analysis of ultra low genome conservation in *Clostridium difficile*. PLoS One 5: e15147.
16. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, et al. (2007) Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. J Bacteriol 189: 8186-8195.
17. Champion OL, Gaunt MW, Gundogdu O, Elmi A, Witney AA, et al. (2005) Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source. Proc Natl Acad Sci U S A 102: 16043-16048.

18. Tettelin H, Massignani V, Cieslewicz MJ, Donati C, Medini D, et al. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* 102: 13950-13955.
19. McFarland LV, Brandmarker SA, Guandalini S (2000) Pediatric *Clostridium difficile*: a phantom menace or clinical reality? *J Pediatr Gastroenterol Nutr* 31: 220-231.
20. Ferraris L, Butel MJ, Campeotto F, Vodovar M, Roze JC, et al. (2012) Clostridia in premature neonates' gut: incidence, antibiotic susceptibility, and perinatal determinants influencing colonization. *PLoS One* 7: e30594.
21. Spencer RC (1998) The role of antimicrobial agents in the aetiology of *Clostridium difficile*-associated disease. *J Antimicrob Chemother* 41 Suppl C: 21-27.
22. Coia JE (2009) What is the role of antimicrobial resistance in the new epidemic of *Clostridium difficile*? *Int J Antimicrob Agents* 33 Suppl 1: S9-12.
23. Bartlett JG (2008) Historical perspectives on studies of *Clostridium difficile* and *C. difficile* infection. *Clin Infect Dis* 46 Suppl 1: S4-11.
24. Gan SI, Beck PL (2003) A new look at toxic megacolon: an update and review of incidence, etiology, pathogenesis, and management. *Am J Gastroenterol* 98: 2363-2371.
25. Rupnik M, Wilcox MH, Gerding DN (2009) *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol* 7: 526-536.
26. Burdon DW, Brown JD, Youngs DJ, Arabi Y, Shinagawa N, et al. (1979) Antibiotic susceptibility of *Clostridium difficile*. *J Antimicrob Chemother* 5: 307-310.
27. Bakken JS, Borody T, Brandt LJ, Brill JV, Demarco DC, et al. (2011) Treating *Clostridium difficile* infection with fecal microbiota transplantation. *Clin Gastroenterol Hepatol* 9: 1044-1049.
28. Rohlke F, Surawicz CM, Stollman N (2010) Fecal flora reconstitution for recurrent *Clostridium difficile* infection: results and methodology. *J Clin Gastroenterol* 44: 567-570.
29. Foglia G, Shah S, Luxemburger C, Pietrobon PJ (2012) *Clostridium difficile*: Development of a novel candidate vaccine. *Vaccine* 30: 4307-4309.
30. Gerding DN, Muto CA, Owens RC, Jr. (2008) Measures to control and prevent *Clostridium difficile* infection. *Clin Infect Dis* 46 Suppl 1: S43-49.
31. Berrington A, Borriello SP, Brazier JS, Duckworth G, Foster K, et al. (2004) National *Clostridium difficile* Standards Group: Report to the Department of Health. *J Hosp Infect* 56 Suppl 1: 1-38.
32. Braun V, Hundsberger T, Leukel P, Sauerborn M, von Eichel-Streiber C (1996) Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene* 181: 29-38.
33. Rupnik M, Dupuy B, Fairweather NF, Gerding DN, Johnson S, et al. (2005) Revised nomenclature of *Clostridium difficile* toxins and associated genes. *J Med Microbiol* 54: 113-117.
34. Carter GP, Rood JI, Lyras D (2012) The role of toxin A and toxin B in the virulence of *Clostridium difficile*. *Trends Microbiol* 20: 21-29.
35. von Eichel-Streiber C, Boquet P, Sauerborn M, Thelestam M (1996) Large clostridial cytotoxins--a family of glycosyltransferases modifying small GTP-binding proteins. *Trends Microbiol* 4: 375-382.
36. Voth DE, Ballard JD (2005) *Clostridium difficile* toxins: mechanism of action and role in disease. *Clin Microbiol Rev* 18: 247-263.

37. Just I, Selzer J, Wilm M, von Eichel-Streiber C, Mann M, et al. (1995) Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* 375: 500-503.
38. Aktories K, Just I (1995) Monoglucosylation of low-molecular-mass GTP-binding Rho proteins by clostridial cytotoxins. *Trends Cell Biol* 5: 441-443.
39. Chaves-Olarte E, Weidmann M, Eichel-Streiber C, Thelestam M (1997) Toxins A and B from *Clostridium difficile* differ with respect to enzymatic potencies, cellular substrate specificities, and surface binding to cultured cells. *J Clin Invest* 100: 1734-1741.
40. Petrilli V, Papin S, Tschopp J (2005) The inflammasome. *Curr Biol* 15: R581.
41. Ng J, Hirota SA, Gross O, Li Y, Ulke-Lemee A, et al. (2010) *Clostridium difficile* toxin-induced inflammation and intestinal injury are mediated by the inflammasome. *Gastroenterology*.
42. Taylor NS, Thorne GM, Bartlett JG (1981) Comparison of two toxins produced by *Clostridium difficile*. *Infect Immun* 34: 1036-1043.
43. Lima AA, Lyerly DM, Wilkins TD, Innes DJ, Guerrant RL (1988) Effects of *Clostridium difficile* toxins A and B in rabbit small and large intestine in vivo and on cultured cells in vitro. *Infect Immun* 56: 582-588.
44. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, et al. (2009) Toxin B is essential for virulence of *Clostridium difficile*. *Nature* 458: 1176-1179.
45. Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, et al. (2010) The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* 467: 711-713.
46. Kuehne SA, Cartman ST, Minton NP (2011) Both, toxin A and toxin B, are important in *Clostridium difficile* infection. *Gut Microbes* 2: 252-255.
47. O'Connor JR, Lyras D, Farrow KA, Adams V, Powell DR, et al. (2006) Construction and analysis of chromosomal *Clostridium difficile* mutants. *Mol Microbiol* 61: 1335-1351.
48. Hussain HA, Roberts AP, Mullany P (2005) 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* 54: 137-141.
49. Olling A, Seehase S, Minton NP, Tatge H, Schroter S, et al. (2011) Release of TcdA and TcdB from *Clostridium difficile* cdi 630 is not affected by functional inactivation of the *tcdE* gene. *Microb Pathog*.
50. Hundsberger T, Braun V, Weidmann M, Leukel P, Sauerborn M, et al. (1997) Transcription analysis of the genes *tcdA-E* of the pathogenicity locus of *Clostridium difficile*. *Eur J Biochem* 244: 735-742.
51. Moncrief JS, Barroso LA, Wilkins TD (1997) Positive regulation of *Clostridium difficile* toxins. *Infect Immun* 65: 1105-1108.
52. Mani N, Dupuy B (2001) Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proc Natl Acad Sci U S A* 98: 5844-5849.
53. Matamouros S, England P, Dupuy B (2007) *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Mol Microbiol* 64: 1274-1288.
54. Carter GP, Douce GR, Govind R, Howarth PM, Mackin KE, et al. (2011) The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of *Clostridium difficile*. *PLoS Pathog* 7: e1002317.
55. Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP (2012) Precise Manipulation of the *Clostridium difficile* Chromosome Reveals a Lack of Association between the *tcdC* Genotype and Toxin Production. *Appl Environ Microbiol* 78: 4683-4690.

56. Popoff MR, Rubin EJ, Gill DM, Boquet P (1988) Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infect Immun* 56: 2299-2306.
57. Perelle S, Gibert M, Bourlioux P, Corthier G, Popoff MR (1997) Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infect Immun* 65: 1402-1407.
58. Carter GP, Lyras D, Allen DL, Mackin KE, Howarth PM, et al. (2007) Binary toxin production in *Clostridium difficile* is regulated by CdtR, a LytTR family response regulator. *J Bacteriol* 189: 7290-7301.
59. Geric B, Carman RJ, Rupnik M, Genheimer CW, Sambol SP, et al. (2006) Binary toxin-producing, large clostridial toxin-negative *Clostridium difficile* strains are enterotoxic but do not cause disease in hamsters. *J Infect Dis* 193: 1143-1150.
60. Schwan C, Stecher B, Tzivelekidis T, van Ham M, Rohde M, et al. (2009) *Clostridium difficile* toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. *PLoS Pathog* 5: e1000626.
61. Calabi E, Calabi F, Phillips AD, Fairweather NF (2002) Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. *Infect Immun* 70: 5770-5778.
62. Fagan RP, Albesa-Jove D, Qazi O, Svergun DI, Brown KA, et al. (2009) Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Molecular Microbiology* 71: 1308-1322.
63. Kirby JM, Ahern H, Roberts AK, Kumar V, Freeman Z, et al. (2009) Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile*. *J Biol Chem* 284: 34666-34673.
64. Cerquetti M, Molinari A, Sebastianelli A, Diociaiuti M, Petruzzelli R, et al. (2000) Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. *Microb Pathog* 28: 363-372.
65. Hennequin C, Janoir C, Barc MC, Collignon A, Karjalainen T (2003) Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiology* 149: 2779-2787.
66. Waligora AJ, Hennequin C, Mullany P, Bourlioux P, Collignon A, et al. (2001) Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. *Infection and Immunity* 69: 2144-2153.
67. Foster TJ, McDevitt D (1994) Surface-associated proteins of *Staphylococcus aureus*: their possible roles in virulence. *FEMS Microbiol Lett* 118: 199-205.
68. Wright A, Drudy D, Kyne L, Brown K, Fairweather NF (2008) Immunoreactive cell wall proteins of *Clostridium difficile* identified by human sera. *J Med Microbiol* 57: 750-756.
69. Lawley TD, Clare S, Deakin LJ, Goulding D, Yen JL, et al. (2010) Use of purified *Clostridium difficile* spores to facilitate evaluation of health care disinfection regimens. *Appl Environ Microbiol* 76: 6895-6900.
70. Heeg D, Burns DA, Cartman ST, Minton NP (2012) Spores of *Clostridium difficile* clinical isolates display a diverse germination response to bile salts. *PLoS One* 7: e32381.
71. Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, et al. (2009) Antibiotic treatment of *clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. *Infect Immun* 77: 3661-3669.
72. Merrigan M, Venugopal A, Mallozzi M, Roxas B, Viswanathan VK, et al. (2010) Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as robust toxin production. *J Bacteriol* 192: 4904-4911.

73. Akerlund T, Persson I, Unemo M, Noren T, Svenungsson B, et al. (2008) Increased sporulation rate of epidemic *Clostridium difficile* Type 027/NAP1. J Clin Microbiol 46: 1530-1533.
74. Burns DA, Heeg D, Cartman ST, Minton NP (2011) Reconsidering the sporulation characteristics of hypervirulent *Clostridium difficile* BI/NAP1/027. PLoS One 6: e24894.
75. Levy SB, Marshall B (2004) Antibacterial resistance worldwide: causes, challenges and responses. Nat Med 10: S122-129.
76. George WL, Sutter VL, Citron D, Finegold SM (1979) Selective and differential medium for isolation of *Clostridium difficile*. J Clin Microbiol 9: 214-219.
77. Hafiz S, Oakley CL (1976) *Clostridium difficile*: isolation and characteristics. J Med Microbiol 9: 129-136.
78. Oxoid Ltd *Clostridium difficile* Selective Supplement. Oxoid Ltd.
- .
79. Bourgault AM, Lamothe F, Loo VG, Poirier L (2006) In vitro susceptibility of *Clostridium difficile* clinical isolates from a multi-institutional outbreak in Southern Quebec, Canada. Antimicrob Agents Chemother 50: 3473-3475.
80. Mullany P, Wilks M, Tabaqchali S (1995) Transfer of macrolide-lincosamide-streptogramin B (MLS) resistance in *Clostridium difficile* is linked to a gene homologous with toxin A and is mediated by a conjugative transposon, Tn5398. J Antimicrob Chemother 35: 305-315.
81. Mullany P, Pallen M, Wilks M, Stephen JR, Tabaqchali S (1996) A group II intron in a conjugative transposon from the gram-positive bacterium, *Clostridium difficile*. Gene 174: 145-150.
82. Lyras D, Storie C, Huggins AS, Crellin PK, Bannam TL, et al. (1998) Chloramphenicol resistance in *Clostridium difficile* is encoded on Tn4453 transposons that are closely related to Tn4451 from *Clostridium perfringens*. Antimicrob Agents Chemother 42: 1563-1567.
83. Clabots CR, Johnson S, Bettin KM, Mathie PA, Mulligan ME, et al. (1993) Development of a rapid and efficient restriction endonuclease analysis typing system for *Clostridium difficile* and correlation with other typing systems. J Clin Microbiol 31: 1870-1875.
84. Gal M, Northey G, Brazier JS (2005) A modified pulsed-field gel electrophoresis (PFGE) protocol for subtyping previously non-PFGE typeable isolates of *Clostridium difficile* polymerase chain reaction ribotype 001. J Hosp Infect 61: 231-236.
85. Cartwright CP, Stock F, Beekmann SE, Williams EC, Gill VJ (1995) PCR amplification of rRNA intergenic spacer regions as a method for epidemiologic typing of *Clostridium difficile*. J Clin Microbiol 33: 184-187.
86. Rupnik M, Avesani V, Janc M, von Eichel-Streiber C, Delmee M (1998) A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. J Clin Microbiol 36: 2240-2247.
87. Lemee L, Dhalluin A, Pestel-Caron M, Lemeland JF, Pons JL (2004) Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. J Clin Microbiol 42: 2609-2617.
88. He M, Sebahia M, Lawley TD, Stabler RA, Dawson LF, et al. (2010) Evolutionary dynamics of *Clostridium difficile* over short and long time scales. Proc Natl Acad Sci U S A 107: 7527-7532.

89. Dingle KE, Griffiths D, Didelot X, Evans J, Vaughan A, et al. (2011) Clinical *Clostridium difficile*: clonality and pathogenicity locus diversity. PLoS One 6: e19993.
90. Stabler RA, Dawson LF, Valiente E, Cairns MD, Martin MJ, et al. (2012) Macro and micro diversity of *Clostridium difficile* isolates from diverse sources and geographical locations. PLoS One 7: e31559.
91. Knetsch CW, Terveer EM, Lauber C, Gorbalenya AE, Harmanus C, et al. (2012) Comparative analysis of an expanded *Clostridium difficile* reference strain collection reveals genetic diversity and evolution through six lineages. Infect Genet Evol.
92. Rupnik M, Brazier JS, Duerden BI, Grabnar M, Stubbs SL (2001) Comparison of toxinotyping and PCR ribotyping of *Clostridium difficile* strains and description of novel toxinotypes. Microbiology 147: 439-447.
93. Razavi B, Apisarnthanarak A, Mundy LM (2007) *Clostridium difficile*: emergence of hypervirulence and fluoroquinolone resistance. Infection 35: 300-307.
94. Cartman ST, Heap JT, Kuehne SA, Cockayne A, Minton NP (2010) The emergence of 'hypervirulence' in *Clostridium difficile*. Int J Med Microbiol.
95. Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, et al. (2008) Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clin Infect Dis 47: 1162-1170.
96. Cheng VC, Yam WC, Lam OT, Tsang JL, Tse EY, et al. (2011) *Clostridium difficile* isolates with increased sporulation: emergence of PCR ribotype 002 in Hong Kong. Eur J Clin Microbiol Infect Dis 30: 1371-1381.
97. Borgmann S, Kist M, Jakobiak T, Reil M, Scholz E, et al. (2008) Increased number of *Clostridium difficile* infections and prevalence of *Clostridium difficile* PCR ribotype 001 in southern Germany. Euro Surveill 13.
98. Spigaglia P, Barbanti F, Dionisi AM, Mastrantonio P (2010) *Clostridium difficile* isolates resistant to fluoroquinolones in Italy: emergence of PCR ribotype 018. J Clin Microbiol 48: 2892-2896.
99. Valiente E, Dawson LF, Cairns MD, Stabler RA, Wren BW (2012) Emergence of new PCR ribotypes from the hypervirulent *Clostridium difficile* 027 lineage. J Med Microbiol 61: 49-56.
100. Freeman J, Fawley W, Baines S, Wilcox M (2006) Measurement of toxin production by *Clostridium difficile*. Lancet 367: 982-983; author reply 983-984.
101. Freeman J, Baines SD, Saxton K, Wilcox MH (2007) Effect of metronidazole on growth and toxin production by epidemic *Clostridium difficile* PCR ribotypes 001 and 027 in a human gut model. J Antimicrob Chemother 60: 83-91.
102. Health Protection Agency (2011) Annual counts and rates of *C. difficile* infection (April 2007 - March 2011) by PCO for patients aged 2 years and over.
103. Office for National Statistics (2011) Deaths involving *Clostridium difficile*: England and Wales, 2006 to 2010.
104. Madigan MT, Martinko JM, Parker J (2003) Brock Biology of microorganisms. Upper Saddle River, N.J.: London : Prentice Hall ; Pearson Education. xxv, 1019, [1058] p. p.
105. Hachler H, Kayser FH, Berger-Bachi B (1987) Homology of a transferable tetracycline resistance determinant of *Clostridium difficile* with *Streptococcus (Enterococcus) faecalis* transposon Tn916. Antimicrob Agents Chemother 31: 1033-1038.

106. Smith CJ, Markowitz SM, Macrina FL (1981) Transferable tetracycline resistance in *Clostridium difficile*. *Antimicrob Agents Chemother* 19: 997-1003.
107. Minton N, Carter G, Herbert M, O'Keeffe T, Purdy D, et al. (2004) The development of *Clostridium difficile* genetic systems. *Anaerobe* 10: 75-84.
108. Hayes W (1953) Observations on a transmissible agent determining sexual differentiation in *Bacterium coli*. *J Gen Microbiol* 8: 72-88.
109. Tatum EL, Lederberg J (1947) Gene Recombination in the Bacterium *Escherichia coli*. *J Bacteriol* 53: 673-684.
110. Hayes W (1953) The mechanism of genetic recombination in *Escherichia coli*. *Cold Spring Harb Symp Quant Biol* 18: 75-93.
111. Lederberg J, Cavalli LL, Lederberg EM (1952) Sex Compatibility in *Escherichia Coli*. *Genetics* 37: 720-730.
112. Adelberg EA, Pittard J (1965) Chromosome Transfer in Bacterial Conjugation. *Bacteriol Rev* 29: 161-172.
113. Bouck N, Adelberg EA (1963) The relationship between DNA synthesis and conjugation in *Escherichia coli*. *Biochem Biophys Res Commun* 11: 24-27.
114. Lesic B, Zouine M, Ducos-Galand M, Huon C, Rosso ML, et al. (2012) A Natural System of Chromosome Transfer in *Yersinia pseudotuberculosis*. *PLoS Genet* 8: e1002529.
115. Hochhut B, Marrero J, Waldor MK (2000) Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a constin found in *Vibrio cholerae* O139. *J Bacteriol* 182: 2043-2047.
116. Gasson MJ, Davies FL (1980) High-frequency conjugation associated with *Streptococcus lactis* donor cell aggregation. *J Bacteriol* 143: 1260-1264.
117. McKay LL, Baldwin KA, Walsh PM (1980) Conjugal transfer of genetic information in group N streptococci. *Appl Environ Microbiol* 40: 84-89.
118. Stentz R, Jury K, Eaton T, Parker M, Narbad A, et al. (2004) Controlled expression of CluA in *Lactococcus lactis* and its role in conjugation. *Microbiology* 150: 2503-2512.
119. Anderson DG, McKay LL (1984) Genetic and physical characterization of recombinant plasmids associated with cell aggregation and high-frequency conjugal transfer in *Streptococcus lactis* ML3. *J Bacteriol* 158: 954-962.
120. Gasson MJ, Swindell S, Maeda S, Dodd HM (1992) Molecular rearrangement of lactose plasmid DNA associated with high-frequency transfer and cell aggregation in *Lactococcus lactis* 712. *Mol Microbiol* 6: 3213-3223.
121. Gasson MJ, Godon JJ, Pillidge CJ, Eaton TJ, Jury K, et al. (1995) Characterization and exploitation of conjugation in *Lactococcus lactis*. *Int Dairy Journal* 5: 757-762.
122. Davidson BE, Kordias N, Dobos M, Hillier AJ (1996) Genomic organization of lactic acid bacteria. *Antonie Van Leeuwenhoek* 70: 161-183.
123. Juhas M, Crook DW, Hood DW (2008) Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell Microbiol* 10: 2377-2386.
124. Alvarez-Martinez CE, Christie PJ (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* 73: 775-808.
125. Zechner EL, Lang S, Schildbach JF (2012) Assembly and mechanisms of bacterial type IV secretion machines. *Philos Trans R Soc Lond B Biol Sci* 367: 1073-1087.
126. Burrus V, Pavlovic G, Decaris B, Guedon G (2002) Conjugative transposons: the tip of the iceberg. *Mol Microbiol* 46: 601-610.

127. Salyers AA, Shoemaker NB, Stevens AM, Li LY (1995) Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* 59: 579-590.
128. Scott JR, Kirchman PA, Caparon MG (1988) An intermediate in transposition of the conjugative transposon Tn916. *Proc Natl Acad Sci U S A* 85: 4809-4813.
129. Clewell DB, Flannagan SE, Jaworski DD (1995) Unconstrained Bacterial Promiscuity - the Tn916-Tn1545 Family of Conjugative Transposons. *Trends in Microbiology* 3: 229-236.
130. Scott JR, Churchward GG (1995) Conjugative transposition. *Annu Rev Microbiol* 49: 367-397.
131. Roberts AP, Hennequin C, Elmore M, Collignon A, Karjalainen T, et al. (2003) Development of an integrative vector for the expression of antisense RNA in *Clostridium difficile*. *J Microbiol Methods* 55: 617-624.
132. Enne VI, Delsol AA, Davis GR, Hayward SL, Roe JM, et al. (2005) Assessment of the fitness impacts on *Escherichia coli* of acquisition of antibiotic resistance genes encoded by different types of genetic element. *J Antimicrob Chemother* 56: 544-551.
133. Foucault ML, Depardieu F, Courvalin P, Grillot-Courvalin C (2010) Inducible expression eliminates the fitness cost of vancomycin resistance in enterococci. *Proc Natl Acad Sci U S A* 107: 16964-16969.
134. Jaworski DD, Clewell DB (1995) A Functional Origin of Transfer (Orit) on the Conjugative Transposon Tn916. *Journal of Bacteriology* 177: 6644-6651.
135. Lee CA, Babic A, Grossman AD (2010) Autonomous plasmid-like replication of a conjugative transposon. *Mol Microbiol* 75: 268-279.
136. Auchtung JM, Lee CA, Monson RE, Lehman AP, Grossman AD (2005) Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc Natl Acad Sci U S A* 102: 12554-12559.
137. Franke AE, Clewell DB (1981) Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of "conjugal" transfer in the absence of a conjugative plasmid. *J Bacteriol* 145: 494-502.
138. Roberts AP, Mullany P (2009) A modular master on the move: the Tn916 family of mobile genetic elements. *Trends Microbiol* 17: 251-258.
139. Roberts AP, Mullany P (2011) Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol Rev* 35: 856-871.
140. Van Houdt R, Monchy S, Leys N, Mergeay M (2009) New mobile genetic elements in *Cupriavidus metallidurans* CH34, their possible roles and occurrence in other bacteria. *Antonie Van Leeuwenhoek* 96: 205-226.
141. Rice LB, Carias LL, Marshall S, Rudin SD, Hutton-Thomas R (2005) Tn5386, a novel Tn916-like mobile element in *Enterococcus faecium* D344R that interacts with Tn916 to yield a large genomic deletion. *J Bacteriol* 187: 6668-6677.
142. Rauch PJ, De Vos WM (1992) Characterization of the novel nisin-sucrose conjugative transposon Tn5276 and its insertion in *Lactococcus lactis*. *J Bacteriol* 174: 1280-1287.
143. Rice LB, Carias LL, Marshall SH, Hutton-Thomas R, Rudin S (2007) Characterization of Tn5386, a Tn916-related mobile element. *Plasmid* 58: 61-67.
144. Seth-Smith HM, Fookes MC, Okoro CK, Baker S, Harris SR, et al. (2012) Structure, Diversity, and Mobility of the *Salmonella* Pathogenicity Island 7 Family of

- Integrative and Conjugative Elements within *Enterobacteriaceae*. J Bacteriol 194: 1494-1504.
145. Ravatn R, Studer S, Zehnder AJ, van der Meer JR (1998) Int-B13, an unusual site-specific recombinase of the bacteriophage P4 integrase family, is responsible for chromosomal insertion of the 105-kilobase *clc* element of *Pseudomonas* sp. Strain B13. J Bacteriol 180: 5505-5514.
 146. Hochhut B, Jahreis K, Lengeler JW, Schmid K (1997) CTnscr94, a conjugative transposon found in enterobacteria. J Bacteriol 179: 2097-2102.
 147. Nishi A, Tominaga K, Furukawa K (2000) A 90-kilobase conjugative chromosomal element coding for biphenyl and salicylate catabolism in *Pseudomonas putida* KF715. J Bacteriol 182: 1949-1955.
 148. Garnier F, Taourit S, Glaser P, Courvalin P, Galimand M (2000) Characterization of transposon Tn1549, conferring VanB-type resistance in *Enterococcus* spp. Microbiology 146 (Pt 6): 1481-1489.
 149. Launay A, Ballard SA, Johnson PD, Grayson ML, Lambert T (2006) Transfer of vancomycin resistance transposon Tn1549 from *Clostridium symbiosum* to *Enterococcus* spp. in the gut of gnotobiotic mice. Antimicrob Agents Chemother 50: 1054-1062.
 150. Grindley ND, Whiteson KL, Rice PA (2006) Mechanisms of site-specific recombination. Annu Rev Biochem 75: 567-605.
 151. Taylor KL, Churchward G (1997) Specific DNA cleavage mediated by the integrase of conjugative transposon Tn916. J Bacteriol 179: 1117-1125.
 152. Caparon MG, Scott JR (1989) Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. Cell 59: 1027-1034.
 153. Rice LB, Carias LL (1994) Studies on excision of conjugative transposons in enterococci: evidence for joint sequences composed of strands with unequal numbers of nucleotides. Plasmid 31: 312-316.
 154. Scott JR, Bringel F, Marra D, Van Alstine G, Rudy CK (1994) Conjugative transposition of Tn916: preferred targets and evidence for conjugative transfer of a single strand and for a double-stranded circular intermediate. Mol Microbiol 11: 1099-1108.
 155. Mullany P, Williams R, Langridge GC, Turner DJ, Whalan R, et al. (2012) Behavior and target site selection of conjugative transposon Tn916 in two different strains of toxigenic *Clostridium difficile*. Appl Environ Microbiol 78: 2147-2153.
 156. Wang H, Roberts AP, Lyras D, Rood JI, Wilks M, et al. (2000) Characterization of the ends and target sites of the novel conjugative transposon Tn5397 from *Clostridium difficile*: excision and circularization is mediated by the large resolvase, TndX. J Bacteriol 182: 3775-3783.
 157. Scott JR (1992) Sex and the single circle: conjugative transposition. J Bacteriol 174: 6005-6010.
 158. Wang H, Mullany P (2000) The large resolvase TndX is required and sufficient for integration and excision of derivatives of the novel conjugative transposon Tn5397. J Bacteriol 182: 6577-6583.
 159. Ionesco H (1980) Transferable tetracycline resistance in "*Clostridium difficile*" (author's transl)]. Ann Microbiol (Paris) 131A: 171-179.
 160. Wust J, Hardegger U (1983) Transferable resistance to clindamycin, erythromycin, and tetracycline in *Clostridium difficile*. Antimicrob Agents Chemother 23: 784-786.

161. Su YA, Clewell DB (1993) Characterization of the left 4 kb of conjugative transposon Tn916: determinants involved in excision. *Plasmid* 30: 234-250.
162. Roberts AP, Braun V, von Eichel-Streiber C, Mullany P (2001) Demonstration that the group II intron from the Clostridial Conjugative transposon Tn5397 undergoes splicing In vivo. *J Bacteriol* 183: 1296-1299.
163. Jasni AS, Mullany P, Hussain H, Roberts AP (2010) Demonstration of conjugative transposon (Tn5397)-mediated horizontal gene transfer between *Clostridium difficile* and *Enterococcus faecalis*. *Antimicrob Agents Chemother* 54: 4924-4926.
164. Mullany P, Wilks M, Lamb I, Clayton C, Wren B, et al. (1990) Genetic analysis of a tetracycline resistance element from *Clostridium difficile* and its conjugal transfer to and from *Bacillus subtilis*. *J Gen Microbiol* 136: 1343-1349.
165. Roberts AP, Pratten J, Wilson M, Mullany P (1999) Transfer of a conjugative transposon, Tn5397 in a model oral biofilm. *FEMS Microbiol Lett* 177: 63-66.
166. Bannam TL, Crellin PK, Rood JI (1995) Molecular genetics of the chloramphenicol-resistance transposon Tn4451 from *Clostridium perfringens*: the TnpX site-specific recombinase excises a circular transposon molecule. *Mol Microbiol* 16: 535-551.
167. Crellin PK, Rood JI (1998) Tn4451 from *Clostridium perfringens* is a mobilizable transposon that encodes the functional Mob protein, TnpZ. *Mol Microbiol* 27: 631-642.
168. Lyras D, Rood JI (2000) Transposition of Tn4451 and Tn4453 involves a circular intermediate that forms a promoter for the large resolvase, TnpX. *Mol Microbiol* 38: 588-601.
169. Adams V, Lyras D, Farrow KA, Rood JI (2002) The clostridial mobilisable transposons. *Cell Mol Life Sci* 59: 2033-2043.
170. Hachler H, Berger-Bachi B, Kayser FH (1987) Genetic characterization of a *Clostridium difficile* erythromycin-clindamycin resistance determinant that is transferable to *Staphylococcus aureus*. *Antimicrob Agents Chemother* 31: 1039-1045.
171. Farrow KA, Lyras D, Rood JI (2000) The macrolide-lincosamide-streptogramin B resistance determinant from *Clostridium difficile* 630 contains two *erm*(B) genes. *Antimicrob Agents Chemother* 44: 411-413.
172. Farrow KA, Lyras D, Rood JI (2001) Genomic analysis of the erythromycin resistance element Tn5398 from *Clostridium difficile*. *Microbiology* 147: 2717-2728.
173. Fujii N, Oguma K, Yokosawa N, Kimura K, Tsuzuki K (1988) Characterization of bacteriophage nucleic acids obtained from *Clostridium botulinum* types C and D. *Appl Environ Microbiol* 54: 69-73.
174. Goh S, Chang BJ, Riley TV (2005) Effect of phage infection on toxin production by *Clostridium difficile*. *J Med Microbiol* 54: 129-135.
175. Goh S, Ong PF, Song KP, Riley TV, Chang BJ (2007) The complete genome sequence of *Clostridium difficile* phage phiC2 and comparisons to phiCD119 and inducible prophages of CD630. *Microbiology* 153: 676-685.
176. Govind R, Vedyappan G, Rolfe RD, Dupuy B, Fralick JA (2009) Bacteriophage-mediated toxin gene regulation in *Clostridium difficile*. *J Virol* 83: 12037-12045.
177. Sekulovic O, Meessen-Pinard M, Fortier LC (2011) Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. *J Bacteriol* 193: 2726-2734.

178. Hasselmayer O, Braun V, Nitsche C, Moos M, Rupnik M, et al. (2004) *Clostridium difficile* IStron CdlSt1: discovery of a variant encoding two complete transposase-like proteins. *J Bacteriol* 186: 2508-2510.
179. Haraldsen JD, Sonenshein AL (2003) Efficient sporulation in *Clostridium difficile* requires disruption of the sigmaK gene. *Mol Microbiol* 48: 811-821.
180. Kunkel B, Losick R, Stragier P (1990) The *Bacillus subtilis* gene for the development transcription factor sigma K is generated by excision of a dispensable DNA element containing a sporulation recombinase gene. *Genes Dev* 4: 525-535.
181. Kimura T, Amaya Y, Kobayashi K, Ogasawara N, Sato T (2010) Repression of *sigK* intervening (skin) element gene expression by the CI-like protein SknR and effect of SknR depletion on growth of *Bacillus subtilis* cells. *J Bacteriol* 192: 6209-6216.
182. Brouwer MS, Allan E, Mullany P, Roberts AP (2012) Draft Genome Sequence of the Nontoxigenic *Clostridium difficile* Strain CD37. *J Bacteriol* 194: 2125-2126.
183. Mahillon J, Chandler M (1998) Insertion sequences. *Microbiol Mol Biol Rev* 62: 725-774.
184. Braun V, Mehlig M, Moos M, Rupnik M, Kalt B, et al. (2000) A chimeric ribozyme in *Clostridium difficile* combines features of group I introns and insertion elements. *Mol Microbiol* 36: 1447-1459.
185. Hasselmayer O, Nitsche C, Braun V, von Eichel-Streiber C (2004) The IStron CdlSt1 of *Clostridium difficile*: molecular symbiosis of a group I intron and an insertion element. *Anaerobe* 10: 85-92.
186. Tourasse NJ, Helgason E, Okstad OA, Hegna IK, Kolsto AB (2006) The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. *J Appl Microbiol* 101: 579-593.
187. Dubey GP, Ben-Yehuda S (2011) Intercellular nanotubes mediate bacterial communication. *Cell* 144: 590-600.
188. Ficht TA (2011) Bacterial Exchange via Nanotubes: Lessons Learned from the History of Molecular Biology. *Front Microbiol* 2: 179.
189. Christie PJ, Korman RZ, Zahler SA, Adsit JC, Dunny GM (1987) Two conjugation systems associated with *Streptococcus faecalis* plasmid pCF10: identification of a conjugative transposon that transfers between *S. faecalis* and *Bacillus subtilis*. *J Bacteriol* 169: 2529-2536.
190. Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, et al. (2002) Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol Microbiol* 46: 439-452.
191. Mandel M, Higa A (1970) Calcium-dependent bacteriophage DNA infection. *J Mol Biol* 53: 159-162.
192. Cohen SN, Chang AC, Hsu L (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Natl Acad Sci U S A* 69: 2110-2114.
193. Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23: 673-679.
194. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
195. Hernandez S, Gomez A, Cedano J, Querol E (2009) Bioinformatics annotation of the hypothetical proteins found by omics techniques can help to disclose additional virulence factors. *Curr Microbiol* 59: 451-456.

196. Tian W, Skolnick J (2003) How well is enzyme function conserved as a function of pairwise sequence identity? *J Mol Biol* 333: 863-882.
197. Gomez A, Cedano J, Espadaler J, Hermoso A, Pinol J, et al. (2008) Prediction of protein function improving sequence remote alignment search by a fuzzy logic algorithm. *Protein J* 27: 130-139.
198. Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* 14: 897-911.
199. Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95: 5857-5864.
200. Wootton JC, Federhen S (1996) Analysis of compositionally biased regions in sequence databases. *Methods Enzymol* 266: 554-571.
201. Corpet F, Servant F, Gouzy J, Kahn D (2000) ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res* 28: 267-269.
202. Underwood A, Green J DoubleAct v2 Available at http://www.hpa-bioinfotools.org.uk/pise/double_act.html.
203. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, et al. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944-945.
204. Carver T, Berriman M, Tivey A, Patel C, Bohme U, et al. (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24: 2672-2676.
205. Tatusov T, Tatusov R Open Reading Frame Finder.
206. Brouwer MS, Warburton PJ, Roberts AP, Mullany P, Allan E (2011) Genetic Organisation, Mobility and Predicted Functions of Genes on Integrated, Mobile Genetic Elements in Sequenced Strains of *Clostridium difficile*. *PLoS One* 6: e23014.
207. Brouwer MS, Roberts AP, Mullany P, Allan E (2012) In silico analysis of sequenced strains of *Clostridium difficile* reveals a related set of conjugative transposons carrying a variety of accessory genes. *Mobile Genetic Elements* 2: 1-5.
208. Corver J, Bakker D, Brouwer MS, Harmanus C, Hensgens MP, et al. (2012) Analysis of a *Clostridium difficile* PCR ribotype 078 100 kilobase island reveals the presence of a novel transposon, Tn6164. *BMC Microbiol* 12: 130.
209. Roberts AP, Chandler M, Courvalin P, Guedon G, Mullany P, et al. (2008) Revised nomenclature for transposable genetic elements. *Plasmid* 60: 167-173.
210. Brouwer MS, Mullany P, Roberts AP (2010) Characterization of the conjugative transposon Tn6000 from *Enterococcus casseliflavus* 664.1H1 (formerly *Enterococcus faecium* 664.1H1). *FEMS Microbiol Lett* 309: 71-76.
211. Serfiotis-Mitsa D, Roberts GA, Cooper LP, White JH, Nutley M, et al. (2008) The Orf18 gene product from conjugative transposon Tn916 is an ArdA antirestriction protein that inhibits type I DNA restriction-modification systems. *J Mol Biol* 383: 970-981.
212. Zielenkiewicz U, Ceglowski P (2005) The toxin-antitoxin system of the streptococcal plasmid pSM19035. *J Bacteriol* 187: 6094-6105.
213. Wozniak RA, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 8: 552-563.
214. Wozniak RA, Waldor MK (2009) A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet* 5: e1000439.

215. Van Melderen L, Saavedra De Bast M (2009) Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet* 5: e1000437.
216. Abraham LJ, Rood JI (1987) Identification of Tn4451 and Tn4452, chloramphenicol resistance transposons from *Clostridium perfringens*. *J Bacteriol* 169: 1579-1584.
217. Lyras D, Adams V, Lucet I, Rood JI (2004) The large resolvase TnpX is the only transposon-encoded protein required for transposition of the Tn4451/3 family of integrative mobilizable elements. *Mol Microbiol* 51: 1787-1800.
218. Kobayashi I (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res* 29: 3742-3756.
219. Carfi A, Pares S, Duee E, Galleni M, Duez C, et al. (1995) The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO J* 14: 4914-4921.
220. Muytjens HL, van der Ros-van de Repe J (1982) Comparative activities of 13 beta-lactam antibiotics. *Antimicrob Agents Chemother* 21: 925-934.
221. Chow AW, Cheng N, Bartlett KH (1985) In vitro susceptibility of *Clostridium difficile* to new beta-lactam and quinolone antibiotics. *Antimicrob Agents Chemother* 28: 842-844.
222. Su YA, He P, Clewell DB (1992) Characterization of the tet(M) determinant of Tn916: evidence for regulation by transcription attenuation. *Antimicrob Agents Chemother* 36: 769-778.
223. van Schaik W, Top J, Riley DR, Boekhorst J, Vrijenhoek JE, et al. (2010) Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. *BMC Genomics* 11: 239.
224. Camilli R, Del Grosso M, Iannelli F, Pantosti A (2008) New genetic element carrying the erythromycin resistance determinant erm(TR) in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* 52: 619-625.
225. Abril C, Brodard I, Perreten V (2010) Two novel antibiotic resistance genes, tet(44) and ant(6)-Ib, are located within a transferable pathogenicity island in *Campylobacter fetus subsp. fetus*. *Antimicrob Agents Chemother* 54: 3052-3055.
226. Rich RL, Demeler B, Ashby K, Deivanayagam CC, Petrich JW, et al. (1998) Domain structure of the *Staphylococcus aureus* collagen adhesin. *Biochemistry* 37: 15423-15433.
227. Navarre WW, Schneewind O (1999) Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* 63: 174-229.
228. Rice LB, Carias LL, Hutton-Thomas R, Rudin S (2007) Interaction of related Tn916-like transposons: analysis of excision events promoted by Tn916 and Tn5386 integrases. *J Bacteriol* 189: 3909-3917.
229. Tsai JC, Yen MR, Castillo R, Leyton DL, Henderson IR, et al. (2010) The bacterial intimins and invasins: a large and novel family of secreted proteins. *PLoS One* 5: e14403.
230. Monot M, Boursaux-Eude C, Thibonnier M, Vallenet D, Moszer I, et al. (2011) Reannotation of the genome sequence of *Clostridium difficile* strain 630. *J Med Microbiol* 60: 1193-1199.

231. Hollenstein K, Dawson RJ, Locher KP (2007) Structure and mechanism of ABC transporter proteins. *Curr Opin Struct Biol* 17: 412-418.
232. Marsden GL, Davis IJ, Wright VJ, Sebahia M, Kuijper EJ, et al. (2010) Array comparative hybridisation reveals a high degree of similarity between UK and European clinical isolates of hypervirulent *Clostridium difficile*. *BMC Genomics* 11: 389.
233. Wang H, Roberts AP, Mullany P (2000) DNA sequence of the insertional hot spot of Tn916 in the *Clostridium difficile* genome and discovery of a Tn916-like element in an environmental isolate integrated in the same hot spot. *FEMS Microbiol Lett* 192: 15-20.
234. Wang H, Smith MC, Mullany P (2006) The conjugative transposon Tn5397 has a strong preference for integration into its *Clostridium difficile* target site. *J Bacteriol* 188: 4871-4878.
235. Naglich JG, Andrews RE, Jr. (1988) Tn916-dependent conjugal transfer of PC194 and PUB110 from *Bacillus subtilis* into *Bacillus thuringiensis subsp. israelensis*. *Plasmid* 20: 113-126.
236. Lee CA, Thomas J, Grossman AD (2012) The *Bacillus subtilis* conjugative transposon ICEBs1 mobilizes plasmids lacking dedicated mobilization functions. *J Bacteriol* 194: 3165-3172.
237. Rice LB (1998) Tn916 family conjugative transposons and dissemination of antimicrobial resistance determinants. *Antimicrob Agents Chemother* 42: 1871-1877.
238. Ceccarelli D, Daccord A, Rene M, Burrus V (2008) Identification of the origin of transfer (oriT) and a new gene required for mobilization of the SXT/R391 family of integrating conjugative elements. *J Bacteriol* 190: 5328-5338.
239. Flannagan SE, Clewell DB (1991) Conjugative transfer of Tn916 in *Enterococcus faecalis*: trans activation of homologous transposons. *J Bacteriol* 173: 7136-7141.
240. Liyanage H, Kashket S, Young M, Kashket ER (2001) *Clostridium beijerinckii* and *Clostridium difficile* detoxify methylglyoxal by a novel mechanism involving glycerol dehydrogenase. *Appl Environ Microbiol* 67: 2004-2010.
241. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP (2007) The Clostron: a universal gene knock-out system for the genus *Clostridium*. *J Microbiol Methods* 70: 452-464.
242. Zhong J, Karberg M, Lambowitz AM (2003) Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res* 31: 1656-1664.
243. Karberg M, Guo H, Zhong J, Coon R, Perutka J, et al. (2001) Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat Biotechnol* 19: 1162-1167.
244. Perutka J, Wang W, Goerlitz D, Lambowitz AM (2004) Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J Mol Biol* 336: 421-439.
245. The Minton Laboratory UoN (2006) Protocol for Clostridial Gene Knockout using pMTL007. In: Actions MC, editor.
246. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, et al. (2009) The Clostron: Mutagenesis in *Clostridium* refined and streamlined. *J Microbiol Methods*.

247. Roberts AP, Davis IJ, Seville L, Villedieu A, Mullany P (2006) Characterization of the ends and target site of a novel tetracycline resistance-encoding conjugative transposon from *Enterococcus faecium* 664.1H1. *J Bacteriol* 188: 4356-4361.
248. Mullany P, Wilks M, Tabaqchali S (1991) Transfer of Tn916 and Tn916 delta E into *Clostridium difficile*: demonstration of a hot-spot for these elements in the *C. difficile* genome. *FEMS Microbiol Lett* 63: 191-194.
249. Showsh SA, Andrews RE, Jr. (1992) Tetracycline enhances Tn916-mediated conjugal transfer. *Plasmid* 28: 213-224.
250. Celli J, Trieu-Cuot P (1998) Circularization of Tn916 is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. *Mol Microbiol* 28: 103-117.
251. Sullivan JT, Trzebiatowski JR, Cruickshank RW, Gouzy J, Brown SD, et al. (2002) Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J Bacteriol* 184: 3086-3095.
252. Sakaguchi Y, Hayashi T, Kurokawa K, Nakayama K, Oshima K, et al. (2005) The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. *Proc Natl Acad Sci U S A* 102: 17472-17477.
253. Mullen LM, Nair SP, Ward JM, Rycroft AN, Williams RJ, et al. (2007) Comparative functional genomic analysis of *Pasteurellaceae* adhesins using phage display. *Vet Microbiol* 122: 123-134.
254. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821-829.
255. Chaudhuri RR, Loman NJ, Snyder LA, Bailey CM, Stekel DJ, et al. (2008) xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Res* 36: D543-546.
256. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955-964.
257. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, et al. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35: 3100-3108.
258. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
259. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5: R12.
260. Griffiths AJF, National Center for Biotechnology I, National Institutes of Health . PubMed C (2000) An introduction to genetic analysis. New York: Freeman. xvii, 860 p. p.
261. Wegmann U, O'Connell-Motherway M, Zomer A, Buist G, Shearman C, et al. (2007) Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis subsp. cremoris* MG1363. *J Bacteriol* 189: 3256-3270.
262. Merrigan MM, Sambol SP, Johnson S, Gerding DN (2009) New approach to the management of *Clostridium difficile* infection: colonisation with non-toxigenic *C. difficile* during daily ampicillin or ceftriaxone administration. *Int J Antimicrob Agents* 33 Suppl 1: S46-50.
263. Gerding DN (2012) *Clostridium difficile* infection prevention: biotherapeutics, immunologics, and vaccines. *Discov Med* 13: 75-83.

264. Lopez-Garrido J, Casadesus J (2012) Crosstalk between Virulence Loci: Regulation of *Salmonella enterica* Pathogenicity Island 1 (SPI-1) by Products of the *std* Fimbrial Operon. *PLoS One* 7: e30499.
265. Shintani M, Takahashi Y, Tokumaru H, Kadota K, Hara H, et al. (2010) Response of the *Pseudomonas* host chromosomal transcriptome to carriage of the IncP-7 plasmid pCAR1. *Environ Microbiol* 12: 1413-1426.
266. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18: 1509-1517.
267. Fu X, Fu N, Guo S, Yan Z, Xu Y, et al. (2009) Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics* 10: 161.
268. Camarena L, Bruno V, Euskirchen G, Poggio S, Snyder M (2010) Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA-sequencing. *PLoS Pathog* 6: e1000834.
269. Songer JG (1997) Bacterial phospholipases and their role in virulence. *Trends Microbiol* 5: 156-161.
270. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754-1760.
271. Anders S HTseq: Analysing high-throughput sequencing data with Python.
272. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106.
273. Moura I, Tavares P, Ravi N (1994) Characterization of three proteins containing multiple iron sites: rubrerythrin, desulfoferrodoxin, and a protein containing a six-iron cluster. *Methods Enzymol* 243: 216-240.
274. Pohl E, Haller JC, Mijovilovich A, Meyer-Klaucke W, Garman E, et al. (2003) Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator. *Mol Microbiol* 47: 903-915.
275. Moura I, Tavares P, Moura JJ, Ravi N, Huynh BH, et al. (1990) Purification and characterization of desulfoferrodoxin. A novel protein from *Desulfovibrio desulfuricans* (ATCC 27774) and from *Desulfovibrio vulgaris* (strain Hildenborough) that contains a distorted rubredoxin center and a mononuclear ferrous center. *J Biol Chem* 265: 21596-21602.
276. Dotsch A, Eckweiler D, Schniederjans M, Zimmermann A, Jensen V, et al. (2012) The *Pseudomonas aeruginosa* Transcriptome in Planktonic Cultures and Static Biofilms Using RNA Sequencing. *PLoS One* 7: e31092.
277. Tong Y, Guo M (2009) Bacterial heme-transport proteins and their heme-coordination modes. *Arch Biochem Biophys* 481: 1-15.
278. Vasileva D, Janssen H, Honicke D, Ehrenreich A, Bahl H (2012) Effect of iron limitation and *fur* gene inactivation on the transcriptional profile of the strict anaerobe *Clostridium acetobutylicum*. *Microbiology* 158: 1918-1929.
279. Rimmele M, Boos W (1994) Trehalose-6-phosphate hydrolase of *Escherichia coli*. *J Bacteriol* 176: 5654-5664.
280. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, et al. (2012) Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat Genet* 45: 109-113.

Appendix I

Appendix I is on the disk included with the thesis.

A summary of all of the bioinformatics results as described in Chapter 3 are present in:
Bioinformatics Table.xsl.

Figures from Chapter 3 which may not be clear in print are included as figure
number.png.

Appendix II

Appendix II is on the disk included with the thesis.

Comparisons of the 30 regions of 630 and CD37 that were used for the SNP analysis of the PaLoc transconjugants are included as fasta files.

Sequences of the ends of the genomic fragments that transferred from 630 into CD37 are included as fasta files.

Appendix III

Publications resulting from this study:

Brouwer MS, Warburton PJ, Roberts AP, Mullany P, Allan E (2011) Genetic Organisation, Mobility and Predicted Functions of Genes on Integrated, Mobile Genetic Elements in Sequenced Strains of *Clostridium difficile*. PLoS One 6: e23014.

Brouwer MS, Roberts AP, Mullany P, Allan E (2012) In silico analysis of sequenced strains of *Clostridium difficile* reveals a related set of conjugative transposons carrying a variety of accessory genes. Mobile Genetic Elements 2: 1-5.

Brouwer MS, Allan E, Mullany P, Roberts AP (2012) Draft Genome Sequence of the Nontoxigenic *Clostridium difficile* Strain CD37. J Bacteriol 194: 2125-2126.

Corver J, Bakker D, Brouwer MS, Harmanus C, Hensgens MP, Roberts AP, Kuijper EJ, Van Leeuwen HC (2012) Analysis of a *Clostridium difficile* PCR Ribotype 078 100 Kilobase Island reveals the presence of a novel transposons; Tn6164. BMC Microbiol 12: 130

These publications are included at the back of this thesis.