Development of Forward and Reverse Genetic Tools for *Eubacterium limosum*

Christopher James Millard

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

September 2017

Declaration

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

Christopher James Millard

September 2017

Abstract

Eubacterium limosum is an anaerobic, acetogenic Gram-positive bacterial species which is able to subsist on mixtures of CO, CO_2 and H_2. Additionally, it is able to grow on methanol. While growing autotrophically on these single-carbon feedstocks, *E. limosum* produces a range of fermentation products of industrial interest. However, no techniques for genetic manipulation of *E. limosum* have yet been published, and current studies concentrate on the optimisation of fermentation conditions. Here, techniques for both random insertional mutagenesis and directed genetic manipulation of *E. limosum* are described and exemplified. The transposon mutagenesis procedure demonstrated during this study should enable the generation of large insertional mutant libraries, which will in turn allow the study of gene essentiality during autotrophic growth on single-carbon gases and methanol. The directed genetic manipulations undertaken during this study include gene knockout via Allele-Coupled Exchange (ACE) and in-frame gene deletion via CRISPR-Cas9.

In addition to the establishment of genetic techniques in *E. limosum*, a novel Gram-positive replicon was isolated from a cryptic plasmid of *Clostridium carboxidivorans* during this study. This Gram-positive plasmid replicon is described and characterised. The replicon was modularised in accordance with the clostridial pMTL80000 plasmid system, and has proven suitable for use in *E. limosum* and in a range of other solventogenic and acetogenic clostridia. The copy number of the novel replicon was determined in *E. limosum* and *C. beijerinckii*, as was the copy number of the widely-used pCB102 replicon.

Acknowledgements

I would like to thank my supervisors Professor Nigel Minton and Doctor Ying Zhang for the support and guidance offered to me during the course of my PhD. I would like to thank Dr. Philippa Strong for her remarks on the first draft of this thesis. I would like to thank my friends in the Synthetic Biology Research Centre and elsewhere for their advice and for making B60 such an entertaining place to work. I would like to thank my family for their support over the last four years.

Finally, I would like to thank the BBSRC for funding my research and the University of Nottingham BBSRC DTP administrative staff for their organisational efforts.

Abbreviations

%	Percentage
µl	Microlitres
µg	Micrograms
µM	Micromolar
^{o}C	Degrees Celsius
ACE	Allele-Coupled Exchange
aTc	Anhydrotetracycline
AWM	*Acetobacterium woodii* Medium
BBSRC	Biotechnology and Biological Sciences Research Council
BCA	Bicinchoninic Acid
BLAST	
CAS	Chinese Academy of Sciences
CFU	Colony-Forming Unit
Cla	Clarithromycin
Cm	Chloramphenicol
CRISPR	Clustered Regularly Interspersed Palindromic Repeats
DMSO	Dimethylsulfoxide

DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
EPB	Electroporation Buffer
Erm	Erythromycin
g	Gram
G	Gravity
h	Hours
HAs	Homology Arms
Kb	Kilobase pair
kV	Kilovolt
LB	Luria-Bertani Medium
M	Molar
mg	Milligram
ml	Millilitres
mM	Millimolar
NTC	No-Template Control
OD	Optical Density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
qPCR	Quantitative Polymerase Chain Reaction
RCM	Reinforced Clostridial Medium

RPM	Revolutions Per Minute
SBRC	Synthetic Biology Research Centre
SEB	Sucrose Electroporation Buffer
SNP	Single-Nucleotide Polymorphism
SOEing	Splicing by Overlap Extension
Tm	Thiamphenicol
TraDIS	Transposon-Directed Insertion Site Sequencing
v/v	Volume per volume
w/v	Weight per volume
YTF	Yeast-Tryptone-Fructose Medium
YTG	Yeast-Tryptone-Glucose Medium
Ω	Resistance

Figures

Figure 1.1: The Wood-Ljungdahl pathway. 23

Figure 1.2: PhD project overview. 46

Figure 3.1: Illustration of *pyrE* knockout via Allele-Coupled Exchange (ACE). 89

Figure 3.2: Antibiotic resistance of *E. limosum*. 91

Figure 3.3: Effect of recovery time on transformant recovery in *E. limosum*. 93

Figure 3.4: Transformation efficiency of *E. limosum* at varying voltages. 94

Figure 3.5: Effect of input DNA amounts on transformation of *E. limosum*. 95

Figure 3.6: 5-FOA toxicity assay. 97

Figure 3.7: The ACE *pyrE* knockout vector employed in this study. 99

Figure 3.8: Colony PCR verification of *pyrE* deletion in putative mutants. 100

Figure 4.1: Map of *C. carboxidivorans* cryptic plasmid P19. 109

Figure 4.2: Structure of the P19 minimal replicon. 113

Figure 4.3: The pMTL80000 modular shuttle vector system. 114

Figure 4.4: The modular shuttle vector pMTL87151. 116

Figure 4.5: Diagnostic digest of pMTL87151. 117

Figure 4.7: Segregational stability assay for pMTL83151 and pMTL87151 in *E. limosum*. 121

Figure 4.8: 16S rRNA phylogenetic tree of Gram-positive species of interest, including acetogens and solventogenic clostridia. 126

Figure 4.9: Validation of primer sets used in qPCR experiments in this study. 128

Figure 4.10: Plasmid copy number (PCN) of pMTL87151 and pMTL83151 in *E. limosum* ATCC 8486. 129

Figure 4.11: Plasmid copy number (PCN) of pMTL87151 in *C. beijerinckii*. 130

Figure 4.12: Comparison of growth and plasmid copy number of *C. beijerinckii* harboring pMTL83151 and pMTL87151. 131

Figure 4.13: Plasmid copy number of cryptic plasmid P19 in *C. carboxidivorans*. 133

Figure 4.14: Functional dissection of the P19 replicon via truncation. 136

Figure 4.15: PCR across the Gram-positive replicon region of pMTL87151 and variants. . 137

Figure 4.16: Transformation efficiency of full-length and truncated P19 replicon variants into *E. limosum* ATCC 8486. 138

Figure 4.17: Comparison of bacterial growth and plasmid copy number of pMTL87151 (P19) and pMTL87151C (P19C) in *E. limosum*. 139

Figure 4.18: Comparison of segregational stability over time for plasmids pMTL87151 and pMTL87151C. 141

Figure 5.1: Plasmid map of pMTL-CW22. 158

Figure 5.2: Comparison of tetracycline-inducible systems. 160

Figure 5.3: pMTL-CW21 diagnostic digest. 164

Figure 5.4: pMTL-CW21 diagnostic digest with EcoRV. 165

Figure 5.5: Growth of *E. limosum* in medium supplemented with varying concentrations of anhydrotetracycline (aTc). 167

Figure 5.6: pMTL-JM102 diagnostic digest. 168

Figure 5.7: Preliminary CatP activity assay in *E. limosum*. 171

Figure 5.8: CatP time-course assay. 173

Figure 5.9: Transposon mutagenesis vectors employed in this study. 176

Figure 5.10: Transposase activity and plasmid loss efficiency in *E. limosum* transformant pools harboring pMTL-CW21 and pMTL-CW23. 179

Figure 5.11: Inverse PCR products from mutants grown on Tm + aTc (top gel) or Cla + aTc (bottom gel).....181

Figure 5.12: Growth of *E. limosum* harbouring transposon vectors during aTc-induced plasmid loss.....182

Figure 5.13: Plate test of the CatP reporter.....185

Figure 5.14: Transposon insertion sites on the *E. limosum* ATCC 8486 genome.....196

Figure 5.15: Cumulative total of transposon insertion sites plotted against total genome length of *E. limosum* ATCC 8486.....197

Figure 6.1: Overview of restriction-modification (R-M) systems in *E. limosum* ATCC 8486. 212

Figure 6.2: Diagram of *E. limosum* ATCC 8486 native CRISPR locus.....219

Figure 6.3: Comparative DNA transfer efficiencies of pMTL83151, 87151 and 87151C...222

Figure 6.4: The effect of removing putative restriction sites for R.Eli8486ORFD.....224

Figure 6.5: Illustration of gene editing via CRISPR-Cas9.....227

Figure 6.6: Schematic overview of the pMTL431511::ORFD knockout vector employed in this study.....228

Figure 6.7: Diagnostic colony PCR of *E. limosum* putative transformant colonies following transformation with ORFD knockout construct pMTL431511::ORFD.....231

Figure 6.8: PCR across both ends of the *R.Eli8486ORFD* locus confirming double-crossover and knockout.....232

Figure 6.9: Sequence of the *R.Eli8486ORFD-gmrSD* locus following deletion of *R.Eli8486ORFD* and *gmrSD*.....233

Figure 6.10: Comparison of transformation efficiency between wild-type *E. limosum* ATCC 8486 **(WT)** and *E. limosum* ATCC 8486 *ORFD* following transformation with pMTL-JM102. 235

Figure 8.1: Sequence of the P19 modularised replicon. 250

Figure 8.2: Sequence of the iteron region of the P19B modularised replicon. 251

Tables

Table 2.1: List of bacterial strains used in this study. 48

Table 2.2: A complete list of plasmids used in this study. 50

Table 2.3: A complete list of oligonucleotide primers used in this study. 52

Table 2.4: A complete list of non-primer oligonucleotides used in this study 54

Table 2.5: Chemical suppliers and associated chemicals used in this study. 55

Table 2.6: Media supplements used in this study 58

Table 2.7: A typical PCR programme as used in this study. 68

Table 2.8: Typical qPCR programme. 81

Table 3.1: *E. limosum* colony outgrowth under selection. 96

Table 4.1: Host range of the Gram-positive P19 replicon 124

Table 5.1: Insertion sites for initial inverse PCR sequencing of transposon mutants. 188

Table 5.2: Insertion site sequencing of 15 transposon mutants derived from three parent cells selected on clarithromycin. 191

Table 5.3: Insertion site sequencing of 30 transposon mutants, derived from three parent cells selected on clarithromycin and supplemented with 1 mM 5-FC. 193

Table 7.1: Methylation motifs associated with R-M systems of *E. limosum* ATCC 8486. 215

Table 7.2: Putative restriction-modification motifs within the pMTL80000 modular shuttle vector system. 217

Table 9.1: Transposon insertion sites identified during sequencing of individual transposon mutants via inverse PCR. 251

Table of Contents

Declaration 2
Abstract 3
Acknowledgements 4
Abbreviations 5
Figures 8
Tables 12
1 Introduction 20
1.1 Acetogenic Bacteria 20
1.1.1 The Industrial Potential of Acetogenic Bacteria 21
1.1.2 The Wood-Ljungdahl Pathway 22
1.1.3 Energy Conservation in Acetogens 25
1.1.4 Pathway Engineering in Acetogens 26
1.2 *Eubacterium limosum* 27
1.3 *Clostridium carboxidivorans* 32
1.4 Reverse Genetics in Clostridia 33
1.4.1 ClosTron 34
1.4.2 Allele-Coupled Exchange (ACE) 35
1.4.3 Negative or Counter-Selection Markers 36
1.4.4 CRISPR and CRISPR-Cas9 Gene Editing Technology 37
1.5 Forward Genetics in Clostridia 41
1.5.1 Transposon Mutagenesis 41
1.5.2 High-throughput Transposon Sequencing 42
1.6 Project Aims 43
2 Materials and Methods 48
2.1 Bacterial Strains and Plasmids 48
2.1.1 Bacterial Strains 48
2.1.2 Plasmids 50
2.2 Oligonucleotide Primers 52
2.3 Other Oligonucleotides 54

2.4 Chemicals and Suppliers 55
2.5 Bio-informatics Tools 55
2.5.1 DNA Sequence Analysis 55
2.5.2 Protein Sequence Analysis 56
2.5.3 Plasmid Design and Visualisation 56
2.5.4 Oligonucleotide Design 56
2.6 Data Analysis and Presentation 57
2.7 General Microbiological Techniques 57
2.7.1 Growth Media 57
2.7.2 Media Supplements 57
2.7.3 Anaerobic culture of microorganisms 58
2.7.4 Routine Culture of *Escherichia coli* 58
2.7.5 Routine Culture of *Eubacterium limosum* 59
2.7.6 Routine Culture of *Clostridium carboxidivorans* 61
2.7.7 Routine Culture of *Clostridium beijerinckii* 62
2.7.8 Routine Culture of *Bacillus subtilis* 62
2.7.9 Growth Curves 62
2.7.10 Strain Storage of *E. coli* 62
2.7.11 Strain Storage of Gram-positive organisms 63
2.7.12 Plasmid Segregational Stability Assay 63
2.8 General Molecular Techniques 64
2.8.1 Quantification of DNA 64
2.8.2 Preparation of Genomic DNA from Clostridia 64
2.8.3 Preparation of Plasmid DNA 66
2.8.4 Restriction Enzyme Digests 66
2.8.5 De-phosphorylation of Linear DNA 66
2.8.6 Ligation Reactions 66
2.8.7 Agarose Gel Electrophoresis 67
2.8.8 DNA Extraction from Agarose Gel 67
2.8.9 Polymerase Chain Reaction 67
2.8.10 Splicing by Overlap Extension (SOEing) Polymerase Chain Reaction 68

2.8.11 DNA Fragment Sequencing 68
2.9 Transformation Procedures 69
2.9.1 Preparation of *Eubacterium limosum* Electro-competent Cells 69
2.9.2 Electroporation of *Eubacterium limosum* 70
2.9.3 Transformation of *Bacillus subtilis* 70
2.9.4 Preparation of *Clostridium beijerinckii* Electro-competent Cells 71
2.9.5 Transformation of *Clostridium beijerinckii* 72
2.9.6 Transformation of *Clostridium pasteurianum* 73
2.10 Plasmid Vector Construction 74
2.10.1 Generation of the Clostridial Shuttle Vector pMTL87151 74
2.10.2 Generation of pMTL87151 Variants pMTL87151A, B, and C 74
2.10.3 Generation of the Clostridial Shuttle Vector pMTL87251C 75
2.10.4 Generation of *pyrE* Knockout Vector pMTL-JM101 75
2.10.5 Construction of CRISPR-Cas9 Knockout Vector pMTL431511::ORFD 76
2.11 Reporter Assays 78
2.11.1 CatP Activity Assay in *E. limosum* 78
2.12 Plasmid Copy Number Assay 80
2.12.1 Plasmid Copy Number Determination via Quantitative PCR (qPCR) 80
2.13 Generation of *E. limosum* Insertional Mutants via Transposon Mutagenesis 82
2.13.1 Inverse PCR of *E. limosum* Transposon Mutants 83
2.14 Generation of *E. limosum pyrE* Deletion Mutant via ACE 84
2.15 Generation of *E. limosum* Deletion Mutants via CRISPR-Cas9 84
3 Culture of *E. limosum* ATCC 8486 and notes on Electroporation Optimisation 87
3.1 Introduction 87
3.2 Results 90
3.2.1 Antibiotic Tolerance of *Eubacterium limosum* 90
3.2.2 Electroporation Optimisation 91
3.2.3 Replicon Performance in *E. limosum* 96

3.2.4 Extracellular Polysaccharides 96
3.2.5 5-FOA Toxicity in *E. limosum* 97
3.2.1 Construction of *pyrE* knockout vector 97
3.2.2 Knockout of *pyrE* in *E. limosum* 99
3.3 Discussion 101
3.4 Chapter Summary 102
4 Generation of a Novel Gram-positive Plasmid Replicon from *C. carboxidivorans* 105
4.1 Introduction 105
4.1.1 General Features of Plasmid Replicons 105
4.1.2 *C. carboxidivorans* and Plasmid P19 108
4.2 Results 108
4.2.1 Sequence Analysis of the P19 Cryptic Plasmid 108
4.2.2 Creation of the Modular P19 Replicon 114
4.2.3 Comparison of Transfer Frequencies Between pMTL83151 and pMTL87151 118
4.2.4 Segregational Stability Analysis 119
4.2.5 Segregational Stability of pMTL87151 and pMTL83151 in *E. limosum* 121
4.2.6 Host Range of pMTL87151 122
4.2.7 Plasmid Copy Number in *E. limosum* 126
4.2.8 Plasmid Copy Number in *C. beijerinckii* 130
4.2.9 P19 Copy Number in *C. carboxidivorans* 132
4.2.10 Functional Dissection of the P19 Replicon 135
4.2.11 Characterisation of the truncated replicon P19C 139
4.2.12 Employment of the P19 replicon in other studies 142
4.3 Discussion 142
4.4 Chapter Summary 149
5 Exemplification of an aTc-inducible *mariner* Transposon System in *Eubacterium limosum* 152
5.1 Introduction 152
5.1.1 Transposons 152
5.1.2 *Mariner* Transposons 153

5.1.3 Transposon Mutagenesis in Clostridia 154
5.1.4 The pMTL-CW22 Transposon Vector Series 157
5.1.5 Anhydrotetracycline-inducible Gene Expression Systems 159
5.2 Results 163
5.2.1 Generation of Transposon Mutagenesis Vector pMTL-CW21 163
5.2.2 Generation of Transposon Mutagenesis Vector pMTL-CW23 166
5.2.3 aTc Growth Impact in *E. limosum* 166
5.2.4 Construction of Reporter Plasmid pMTL-JM102 167
5.2.5 CatP Reporter Assay in *E. limosum* 168
5.2.6 Transposon Mutagenesis Vectors pMTL-CW21 and –CW23. 174
5.2.7 Creation of Transposon Mutant Libraries in *E. limosum* 177
5.2.8 Efficiency of Transposase Induction and Plasmid Loss 178
5.2.9 Differentiating between Induced Plasmid Loss and Cell Death 182
5.2.10 Plate Testing of the CatP Reporter 183
5.2.11 Determining the Spontaneity of Transposition 185
5.2.12 Selecting against Spontaneous Transposition via *codA* Reassembly 191
5.2.13 Maximising Diversity in Transposon Mutant Libraries 195
5.2.14 Plotting of Transposon Insertion Sites. 195
5.2.15 Creation of Suicide Vector 198
5.3 Discussion 198
5.4 Chapter Summary 202
6 Knockout of a Restriction Endonuclease in *E. limosum* via CRISPR-Cas9 205
6.1 Introduction 205
6.2 Restriction-Modification Systems as a Barrier to Gene Transfer ... 208
6.3 Results 211
6.3.1 Sequence Analysis of R-M Systems in *E. limosum* ATCC 8486 211
6.3.2 The Endogenous CRISPR System of *E. limosum* ATCC 8486 218
6.3.3 Evaluating R-M System Efficiency 220

6.3.4 Construction of CRISPR KO Vectors....225
6.3.5 Knockout of *R.Eli8486ORFD*....229
6.3.6 SNP within the guide RNA of the knockout vector....236
6.4 Discussion....237
6.5 Chapter Summary....239
7 General Discussion....242
8 Appendix....249
8.1 P19 Modular Replicon Sequence....249
8.2 Transposon Insertion Sites Sequenced during this Study....251
9 References....255

CHAPTER ONE

INTRODUCTION

1 Introduction

Eubacterium limosum is a Gram-positive anaerobic microbe, which is of scientific interest both for its ecological niche as a commensal organism in ruminants, and for its acetogenic metabolism (Genthner et al., 1981). It is able to grow autotrophically on mixtures of CO, CO_2 and H_2 which may be generated by the gasification of biomass, or through the filtration of industrial off-gases. It is also capable of autotrophic growth on methanol, or heterotrophic growth on a variety of sugars. *E. limosum*, and other acetogenic bacteria, are of increasing interest because they are able to produce various acids and solvents of industrial interest. *E. limosum* has, to date, not been the target of published genetic modification, and tools for its manipulation have not been described.

1.1 Acetogenic Bacteria

Many bacteria are known to utilise CO or mixtures of CO_2-H_2 as their sole carbon and energy sources, producing acetyl-CoA as a result. Generally, acetate is produced as their primary fermentation product via the enzymes phosphotransacetylase and acetate kinase. These organisms are known as acetogens. The first acetogen to be isolated was *Clostridium aceticum*, though the culture was lost after World War 2. The first acetogen to be studied in great enzymological detail was *Clostridium thermoaceticum* (later *Moorella thermoaceticum*), whose acetyl-CoA pathway was elucidated by Harland Wood and Lars Ljungdahl (reviewed in Schiel-Bengelsdorf and Durre, 2012),

and later named after them. Other example species include *Acetobacterium woodii* (the first acetogen to be investigated primarily as an autotroph), *Clostridium autoethanogenum, Clostridium ljungdahlii, Clostridium carboxidivorans*, and *Eubacterium limosum*. *E. limosum* is the subject of this study. All acetogens identified thus far are facultative autotrophs, able to utilise not only CO and CO_2+H_2, but also a variety of sources of organic carbon (reviewed in Schuchmann and Muller, 2014).

Acetogens, and their reductive acetyl-CoA pathway in particular, have been identified as important contributors to the global cycling of carbon, playing an important ecological role (Drake et al., 2008). Microbial acetogenesis has been estimated to be responsible for the production of approximately 10 trillion kilograms of acetic acid per year (Ragsdale, 2008). Acetogens are ubiquitous in nature, and the acetate they produce can in turn be consumed by methanotrophic bacteria. The production of acetyl-CoA from CO and CO_2+H_2 by acetogens is possible due to the Wood-Ljungdahl Pathway (WLP), which is detailed below (see section 1.1.2).

1.1.1 The Industrial Potential of Acetogenic Bacteria

The industrial potential of acetogenic bacteria is often discussed in terms of their ability to grow on synthesis gas, or syngas (Liew et al, 2016). This is a mixture of CO, CO_2, and H_2. Syngas can be produced from the gasification of fossil fuels or biomass. It can also be captured from the exhaust of various industrial processes. The exact ratios of the constituent gases in syngas vary

depending on the source of the gases (Munasinghe and Khanal, 2010). It has been noted that syngas contains several contaminants which are inhibitory to microbial growth, and which must be removed prior to fermentation (Xu et al., 2011). These impurities include tar particulates, ammonia, and nitric oxide, as well as gases including methane, ethylene and acetylene.

Acetogenic bacteria have been recognised as a means by which pollutant gases can be remediated into valuable chemical commodities (Schiel-Bengelsdorf and Durre, 2012). As a result, understanding the metabolism of acetogens and developing tools to modify them is a subject of great interest from the viewpoint of developing fossil fuel-independent fuel and chemical production.

1.1.2 The Wood-Ljungdahl Pathway

The Wood-Ljungdahl pathway (WLP) which underlies acetogenic metabolism consists of two branches, the methyl and carbonyl branches (Ljungdahl, 1986). In the methyl branch of the WLP, CO_2 is reduced to formate and bound to tetrahydrofolate (THF). In the carbonyl branch, CO_2 is reduced to an enzyme-bound CO by CODH/ACS (see Figure 1.1). Because the redox potential of the CO_2/CO couple is so low, only reduced ferredoxin (Fd^{2-}) can supply the electrons necessary for this reaction (a variety of electron donors supply electrons to the reactions of the methyl branch). The bound methyl and carbonyl moieties are condensed into one molecule of acetyl-CoA. If the acetyl group is converted to acetate by phosphotransacetylase and acetate kinase, the CoA is recycled. The reduction of CO_2 to CO in the carbonyl

branch of the WLP has been identified as the greatest thermodynamic barrier that must be overcome during acetogenesis (Schuchmann and Muller, 2014).

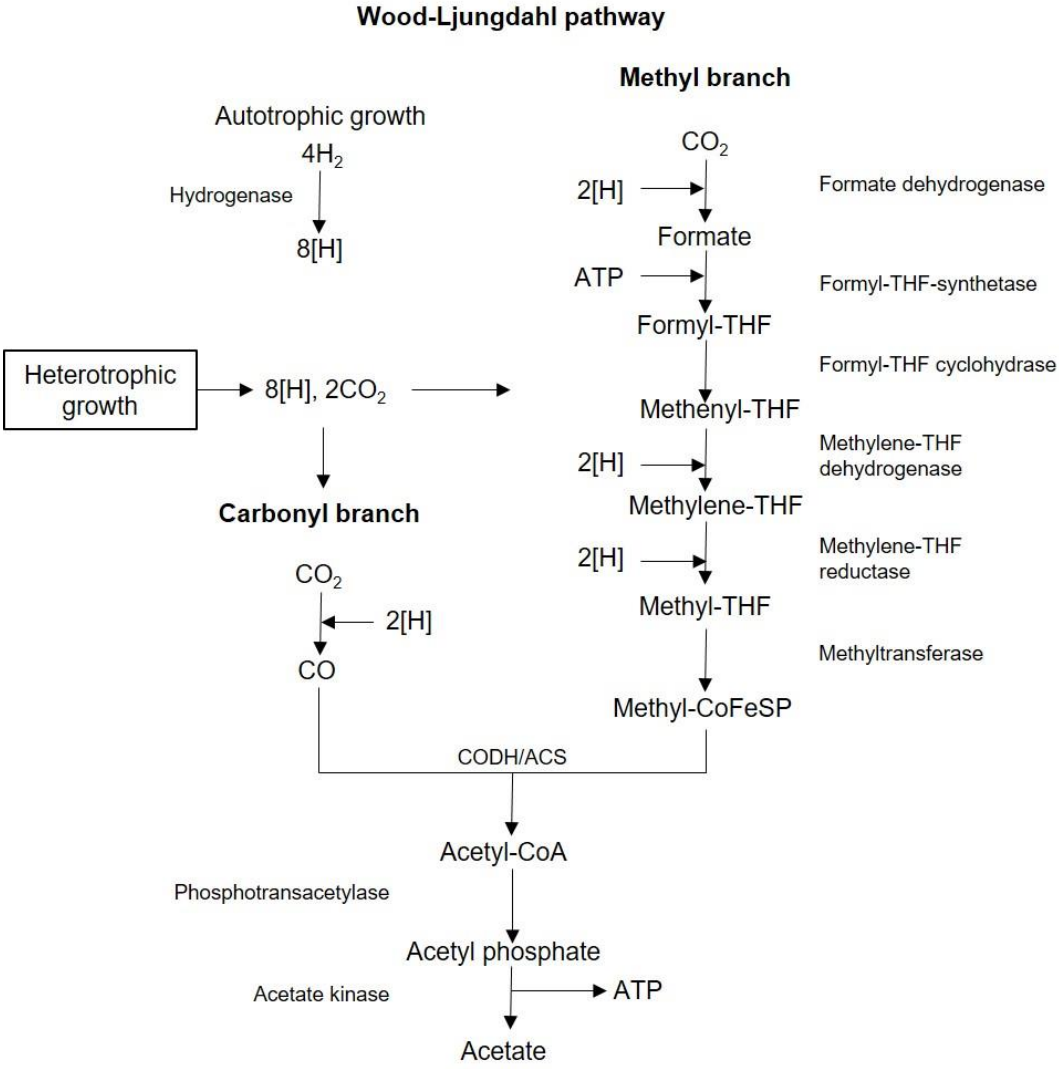


Figure 1.1: The Wood-Ljungdahl pathway. In the methyl branch, a CO_2 molecule is successively reduced to a methyl group while being bound to tetrahydrofolate (THF). The bound methyl group is transferred to a corrinoid iron-sulfur protein (CoFeSP) and then to CO dehydrogenase/Acetyl-CoA synthase (CODH/ACS). CODH/ACS joins the methyl group to the CO from the carbonyl branch, generating acetyl-CoA. [H] represents e- + H^+. Figure adapted from Schuchmann and Müller, 2014.

During heterotrophic acetogenesis (e.g. during growth on fructose), acetyl-CoA and CO_2 are produced by the oxidation of pyruvate to acetyl-CoA via

pyruvate ferredoxin oxidoreductase (PFOR) (reviewed in Ragsdale, 2008). This CO_2 is then reduced to formate and CO in the methyl and carbonyl branches of the WLP. PFOR accordingly acts as a linkage between heterotrophic growth and the Wood-Ljungdahl pathway.

The ultimate product of the WLP is acetyl-CoA, which in turn is used to produce various fermentation products, including acetate, butyrate, and ethanol. The specific range of fermentation products varies depending on the acetogen in question, and several organisms are of particular interest due to their product range. *C. carboxidivorans*, for example, is noteworthy for its production of butanol (Bruant et al., 2010).

Production of alcohols during acetogenic fermentation generally results in the production of ATP when CO is the electron donor, in contrast to production of organic acids such as butyrate. The production of acetate is distinct from the production of other fermentation products in acetogens because it generates ATP, rather than requiring it. Acetogens can therefore resort to side-production of acetate when the energy balance of another fermentation product is negative (Bertsch and Muller, 2015). Production of acetate has different energetic qualities depending on whether it is produced from H_2+CO_2 or CO alone. The $\Delta G^{0'}$ of the reaction is -95 kJ/mol in the former case, and -175 kJ/mol in the latter.

1.1.3 Energy Conservation in Acetogens

It has been noted that while the enzymatic reactions involved in the flow of carbon through the WLP are highly conserved among acetogens, the reactions involving the supply of electrons to this pathway vary widely between organisms (Schuchmann and Muller, 2014). This has been attributed to the fact that for several reductive steps in the WLP, there are several energetically-viable electron donors, such as NADPH, ferredoxin, and H_2. The specific reducing equivalents used by particular organisms have been elucidated in some cases. For the reduction of formate, for example, *Moorella thermoacetica* uses NADPH and *A. woodii* uses H_2 (Wang et al., 2013).

A. woodii is the acetogen in which energy conservation is best-understood, with the reductants known for every redox step in its WLP. The energy conservation system in *A. woodii* can act as a model for other acetogens, and consists of four basic components. The first is the WLP itself. The second is a cytoplasmic hydrogenase, which oxidises molecular hydrogen and produces NADH and reduced ferredoxin (Fd^{2-}). These reducing equivalents are used for the reductive steps of the WLP. The third component is the membrane-bound Rnf complex, which pumps Na+ out of the cell by oxidising Fd^{2-}. Na+ returns through the fourth component, the ATP synthase (reviewed in Bertsch and Muller, 2015). When *A. woodii* is growing on CO alone, reducing equivalents for the WLP are supplied by carbon monoxide dehydrogenase-acetyl CoA synthase (CODH/ACS), which oxidises CO to CO_2, producing Fd^{2-} in the process.

It has been proposed that acetogens should be divided based on the electron transport complex they use to establish the ion gradient that enables ATP production (either the Rnf or Ech complexes), and by the ion that is transported by this complex (Na^+ or H^+). Hence *A. woodii* would be classified as an "Rnf-containing, Na^+-dependent" acetogen (Schuchmann and Muller, 2014). By this scheme, *E. limosum* (the subject of this study) can likewise be classified as an Rnf-containing, Na^+-dependent acetogen, according to recently-obtained experimental evidence (Jeong et al., 2015).

1.1.4 Pathway Engineering in Acetogens

Many acetogens are members of the genus *Clostridium*, which has been noted for both its metabolic diversity and lack of amenability to genetic manipulation (Tracy et al., 2012). This has been attributed to various factors, including the presence of strong restriction-modification systems, as well as to a lack of molecular tools. The present study is concerned with the acetogen *E. limosum.* The genetic tools and strategies reviewed below were established for use in pathogenic clostridia as well as those of biotechnological interest, and were applied to *E. limosum* in this study.

Pathway engineering in acetogens has included both plasmid-based strategies as well as those based on direct modification of the chromosome. The acetogen *C. ljungdahlii* formed the basis of several early studies (Kopke et al., 2010, Leang et al., 2013, Banerjee et al., 2014). In one example, a plasmid-borne operon consisting of the genes responsible for butanol synthesis in *C.*

acetobutylicum was transformed into *C. ljungdahlii* under the control of a single *Clostridium acetobutylicum* native promoter and terminator (Kopke et al., 2010). Butanol was produced at a maximum level of 2 mM in the exponential phase of transformant *C. ljungdahlii* cultures, though this was consumed in the course of growth and converted to butyrate by stationary phase, demonstrating that *C. ljungdahlii* is able to metabolise butanol. Chromosomal modification of *C. ljungdahlii* was demonstrated in 2013, enabling knockout of genes and facilitating pathway engineering (Leang et al., 2013). This same study also demonstrated the function of several plasmid replicative origins in *C. ljungdahlii*, an important tool for heterologous expression of genes. Another study from the same group demonstrated the functionality of a lactose-induced gene expression system, increasing ethanol production 1.5-fold by expression of a plasmid-borne aldehyde/alcohol dehydrogenase gene, *adhE1* (Banerjee et al., 2014). That study also demonstrated the redirecting of a significant proportion of *C. ljungdahlii*'s carbon flux towards the production of acetone via induced expression of an operon of four genes. These studies demonstrate the utility of pathway engineering in acetogens.

1.2 *Eubacterium limosum*

E. limosum is a rod-shaped, Gram-positive, obligately anaerobic firmicute bacterium from the class Clostridiale. It was noted decades ago as being

associated with methanogenic bacteria in an acetate-fed culture enriched from anaerobic digester sludge (Ward et al., 1978).

Decades prior to this study, attention was paid to *E. limosum* both for its role as a ruminant commensal microbe, and for its ability to assimilate H_2-CO_2 and methanol (Genthner et al., 1981). It is worth noting that the study of Genthner and coworkers made use of *E. limosum* ATCC 8486, the type strain which was isolated from human faeces (Eggerth, 1935) and which is the primary subject of this study.

Interestingly, attention has been paid to *E. limosum* as a methane mitigation tool in the context of ruminant digestion. This is because it may compete for H_2 with commensal hydrogenotrophic methanogens in sheep and cattle (Kelly et al., 2016), and has been noted to be involved in inter-species transfer of hydrogen and metabolic products (Rode et al., 1981). *E. limosum* has also been investigated for its ability to synthesise Vitamin B12 (Lamm et al., 1980).

Early studies on the metabolic versatility of *E. limosum* noted its ability to grow on methanol and CO_2-H_2, with subsequent studies confirming its ability to grow with carbon monoxide as its sole carbon and energy source (Genthner and Bryant, 1982). This early study utilised *E. limosum* RF, a strain isolated from the rumen of molasses-fed sheep. The study by Genthner and Bryant was the first to note acetate production during growth on CO as the sole energy source. It was further apparent that CO was utilised preferentially to the CO_2-H_2 mixture, with most CO being consumed before noticeable amounts of H_2

were utilised. *E. limosum* was also observed to possess a notably high tolerance for CO, being able to grow without inhibition when CO made up 50% of the gas phase of the culture. This was in contrast to the CO-oxidising clostridia known at the time, which were inhibited at CO concentrations above 5%. *E. limosum* was shown to be able to produce acetate, butyrate, and caproate during growth on methanol (Genthner et al., 1981). Subsequent studies have likewise identified acetate and butyrate as the primary fermentation products of *E. limosum* (Kelly et al., 2016).

Its ability to produce butyrate during growth on CO as a sole carbon and energy source places *E. limosum* in a minority among acetogens. Butyrate production in *E. limosum* begins at mid-exponential phase when growing on CO (reviewed in Jeong et al., 2015). Other acetogens capable of this include *Clostridium drakei*, *C. carboxidivorans*, and *Butyribacterium methylotrophicum* (reviewed in Schiel-Bengelsdorf and Durre, 2012). The enzymology of butyrate production in *E. limosum* during fermentation of CO has been examined (Jeong et al., 2015), and found to be reliant on a gene cluster that is well-conserved among other butyrate forming bacteria, including *C. acetobutylicum*, *Clostridium beijerinckii*, and *C. carboxidivorans*, of which only the latter is an acetogen. This underlies the previously-established point that acetogens are not a phylogenetic grouping, but rather that acetogenesis is a widely-distributed strategy (Drake et al., 2008).

Increasing interest is being paid to *E. limosum* as a potential vehicle for industrial biotechnology. This is due to its ability to ferment carbon monoxide

via the Wood-Ljungdahl pathway, and for its ability to assimilate methanol as a carbon source. Both of these carbon sources can be generated by industrial processes, and *E. limosum* has therefore been noted as a biocatalyst for the production of organic acids such as acetate and butyrate (Liew et al., 2016).

Much of the existing work on *E. limosum* has been conducted in the strain *E. limosum* KIST612, which was originally isolated in 1997 (reviewed briefly in Jang et al., 2017). A complete genome is available for this strain, the first published for *E. limosum* (Roh et al., 2011). Experiments and sequence analysis have been conducted to elucidate the energy conservation systems of *E. limosum*, with recent evidence suggesting that its energy conservation is based on an Rnf complex establishing a gradient of Na+ ions (Jeong et al., 2015), similar to that found in *A. woodii*.

The performance of *E. limosum* KIST612 during continuous fermentation of CO has been assessed, in order to better characterise it as a potential industrial biocatalyst (Chang et al., 2001). The study by Chang and coworkers observed that cell yields obtained (5 g L^{-1}) were very high, the highest reported for acetogenic bacteria under the given reactor conditions.

Growth of *E. limosum* KIST612 in batch culture with CO has recently been subject to kinetic modelling (Jang et al., 2017) based on sequence analysis and the data generated by Chang *et al*. These recent efforts demonstrate the increasing interest in *E. limosum* as a platform for sustainable chemical production.

Use of *E. limosum* as a biocatalyst would be facilitated by the establishment of robust methods for its genetic manipulation, whether by plasmid-based strategies or via direct chromosomal modification. Extending the fermentation profile of *E. limosum* requires an understanding of its metabolism, which in turn requires tools for the knockout and perturbation of specific target genes.

Two complete genome sequences have recently been published for *E. limosum* in the wider literature. These are *E. limosum* KIST612 (Roh et al., 2011) as mentioned above, and *E. limosum* SA11 (Kelly et al., 2016). A draft genome for *E limosum* ATCC 8486 has recently been published (Song and Cho, 2015). A closed genome sequence for *E. limosum* ATCC 8486 was completed concomitantly with this study, and is the third complete genome for an *E. limosum* strain (R. Mansfield, personal communication). *E. limosum* KIST612 was isolated from an anaerobic digester, SA11 from a sheep rumen, and ATCC 8486 from human faeces (Eggerth, 1935). The genomes for these organisms were measured at 4.3, 4.15, and 4.4 Mb, respectively.

There are several interesting differences between the published genomes. *E. limosum* SA11 was noted as possessing no restriction-modification systems (Kelly et al., 2016), whereas ATCC 8486 appears to possess at least three functional systems, which are examined in the current study. SA11 and KIST612 possess 1 and 3 integrated prophages respectively, all distinct from each other.

At the time of writing, no plasmid-based or chromosomal modification of *E. limosum* has been published. There has been no transformation protocol

published, nor have viable plasmid origins of replication been published for this strain. This study presents and exemplifies a novel plasmid origin of replication for use in *E. limosum*, and exemplifies both forward genetic approaches (insertional mutagenesis via *mariner* transposon) and reverse genetic approaches via selective, in-frame gene deletion via CRISPR-Cas9. Additionally, work was undertaken to optimise an electroporation protocol recently established in *E. limosum*. During this study, work involving the acetogen *C. carboxidivorans* was also undertaken (see below).

1.3 *Clostridium carboxidivorans*

C. carboxidivorans was first isolated from an agricultural settling lagoon (Liou et al., 2005). Like *E. limosum*, *C. carboxidivorans* is an acetogen, able to grow on CO or H_2+CO_2 as sole carbon and energy sources. During autotrophic growth on CO, *C. carboxidivorans* produces acetate, butyrate, ethanol and butanol (reviewed in Bruant et al., 2010).

As an acetogen that produces butanol, *C. carboxidivorans* is an organism of particular industrial interest. Butanol is employed as an industrial solvent and is of interest as a biofuel because it is non-hygroscopic and therefore less corrosive than ethanol (reviewed in Bertsch and Muller, 2015).

In the course of this study, attempts were made to establish gene transfer in *C. carboxidivorans*. These attempts were unsuccessful, probably due to the extensive restriction systems present in that organism. However, in the course of that work a novel Gram-positive plasmid origin of replication was isolated.

This novel replicon, designated P19 after the name of the cryptic plasmid it is derived from, was characterised in terms of its host range, segregational stability, and copy number. The replicon was modularised to render it compatible with the pMTL80000 modular shuttle vector system, which is widely used for gene transfer to clostridia (Heap et al., 2009). In the form of pMTL87151, the new replicon was found to be functional in a range of acetogenic and solventogenic clostridia, including *C. autoethanogenum* and *C. beijerinckii*, as well as within *Bacillus subtilis*, a Gram-positive species outside of the clostridia (see Chapter 4).

1.4 Reverse Genetics in Clostridia

Several techniques exist for the targeted disruption of genes in clostridia. Following the disruption of specific genes, the phenotypes of the resulting mutants can be assessed. While early reverse genetic methods involved the insertional disruption of targeted genes, more recent techniques allow in-frame deletion of genes. This avoids the creation of polar effects which may complicate the analysis of mutant phenotypes. The most recent genome editing techniques are based on counter-selection by the programmable CRISPR-Cas9 system.

1.4.1 **ClosTron**

Genetic engineering of clostridia was initially reliant upon the transfer of replicative plasmids, or upon generation of single-crossover mutants via the integration of non-replicating plasmids. Plasmid-based approaches were limited by the need for selection in favour of the plasmid (for example, with antibiotic supplementation of media). Single-crossover mutants were susceptible to reversion, resulting in instability of mutant phenotypes. However, techniques were developed for the generation of stable mutants. One such method is the ClosTron (Heap et al., 2007). This method employs a targetable intron that inactivates the target gene upon insertion. The intron includes a retro-transposition activated marker (RAM) based on the *ermB* gene, conferring resistance to erythromycin/clarithromycin only after integration has occurred. This allows positive selection in favour of stable integrant mutants. Such intron-based insertional mutagens had previously been employed in *E. coli*, using the *Lactococcus lactis* L1.LtrB intron (Karberg et al., 2001), but the ClosTron marked the introduction of this method into the clostridia. This method was initially used to rapidly generate 5 mutants in *C. difficile* and 6 in the solventogenic clostridium *C. acetobutylicum* (Heap et al., 2007). The latter species is of long-standing interest among the clostridia, representing the most studied solventogenic clostridial species and the process organism of commercial Acetone-Butanol-Ethanol (ABE) fermentation (reviewed in Papoutsakis, 2008).

1.4.2 **Allele-Coupled Exchange (ACE)**

Subsequent developments have provided the means to replace intron-based insertional mutagens with recombination-based methods that allow the substitution of wild-type alleles with mutant variants. A foundational tool for in-frame deletion and integration of heterologous DNA into clostridia is Allele-Coupled Exchange (ACE) technology, which was first exemplified in *C. acetobutylicum* (Heap et al., 2012). This strategy requires the integration of a plasmid into the target chromosome in such a way that a selectable phenotype is created upon the occurrence of a double crossover event. In the original ACE study, and in many subsequent implementations, the genomic target selected is the orotate phosphoribosyltransferase gene *pyrE,* which is involved in pyrimidine biosynthesis. This gene was selected because it can function as both a selective and counter-selective marker. A functional *pyrE* confers uracil prototrophy, while a non-functional *pyrE* confers resistance to 5-fluoroorotic acid (5-FOA) toxicity. Use of analogous genes has been long-established in yeast strains (Sakai et al., 1991). In order to knock out *pyrE* within *C. acetobutylicum,* an integrative plasmid construct (pMTL-JH12) was utilised. Transformants were selected based on a *catP* selective marker encoded by the pMTL-JH12 backbone, whilst double integrants were selected for by growth on media supplemented with 5-FOA. The order of recombination events is dictated by the varying homology arm lengths in the ACE vector. The initial recombination results in plasmid integration (and consequently faster growth, as cells replicate at normal speed rather than being

limited by the replication-deficient plasmid). The second recombination event results in truncation of *pyrE*, with the phenotype described above.

ACE was subsequently used to repair *pyrE* and demonstrate the reversibility of the strategy, with double crossover events being selected for by growth on media lacking uracil, which does not support growth of *pyrE* knockouts unable to synthesise it by themselves. Thus, by selection of an appropriate target gene which can be selected both positively and negatively, reversible modifications can be made to the chromosome of an organism of interest. By alternating the method of selection, iterative modifications can be made to the target organism; in the study by Heap and coworkers, the entire genome of bacteriophage lambda was incorporated into the genome of *C. acetobutylicum* (Heap et al., 2012). During the same study, ACE was used to insert an NADPH-dependent alcohol dehydrogenase (*adh*) from *C. beijerinckii* into *C. acetobutylicum*. A decrease in acetone production was observed, alongside an increase in isopropanol production.

1.4.3 **Negative or Counter-Selection Markers**

The use of heterologous *pyrE* genes located on ACE vectors as negative selection markers is critical to the generation of knockout mutants via ACE. The use of alternative negative selection markers has also been explored. Examples include the use of alternative counterselection targets such as the cytosine deaminase (*codA*) in *Clostridium difficile* (Cartman et al., 2012) and the mRNA interferase *mazF* in *C. acetobutylicum* (Al-Hinai et al., 2012).

Cytosine deaminase catalyses the conversion of 5-fluorocytosine (5-FC) to highly toxic 5-fluorouracil (5-FU), and is employed as a counter-selective marker during the transposon mutagenesis work conducted during this study (see Chapter 5).

1.4.4 CRISPR and CRISPR-Cas9 Gene Editing Technology

Restriction endonucleases are not the only bacterial barrier to invasive nucleic acids. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) form an acquired immune system in bacteria and archaea, targeting foreign DNA on the basis of particular recognition sequences (Horvath and Barrangou, 2010). CRISPR was first noted in *E. coli* in 1987, and subsequently identified in numerous species of bacteria and archaea. CRISPR loci consist of several direct repeat sequences separated by spacers which are generally derived from invasive plasmid or viral DNA sequences (the sequence that a spacer is derived from is known as a proto-spacer). They are generally located adjacent to CRISPR-associated (*cas*) genes.

CRISPR-associated gene clusters appear to contain a variety of genes, including helicases, nucleases and DNA polymerases (Haft et al., 2005). Cas1, for example, is thought to be a dsDNA endonuclease (Wiedenheft et al., 2009). Four of these genes (*cas1* to *cas4*) are always associated with CRISPR loci. Additional *cas* genes are often found at these loci, as well as other ancillary genes such as those encoding "repeat-associated mysterious proteins" (RAMP).

By definition, CRISPR loci contain a repeat region adjacent to their *cas* genes. This region consists of a number of palindromic direct repeats with unique spacer regions between them. The recognition that these spacer regions were derived from bacteriophage and plasmid sequences was key in identifying CRISPR loci as bacterial defensive systems (Bolotin et al., 2005). CRISPR-mediated acquired immunity depends upon recognition of invasive DNA by a complex of Cas proteins, after which invasive DNA is processed and integrated into the repeat region of the CRISPR locus, forming a new spacer sequence. To enable an immune response against the sequences integrated into the repeat region, it is transcribed into a pre-crRNA which is processed in turn to generate a number of crRNAs (CRISPR RNAs) which target the Cas effector protein(s) against foreign DNA. The foreign DNA sequences (proto-spacers) selected for integration as spacers within the CRISPR locus are selected by the CRISPR system because they are adjacent to specific proto-spacer-associated motifs, or PAMs (Mojica et al., 2009). It has been speculated that the absence of the PAM within the CRISPR locus itself prevents self-directed ("autoimmune") activity by the CRISPR system (Horvath and Barrangou, 2010).

CRISPR systems can be classified based on their effector proteins. These may consist of the 'Cascade' (Cas Complex for Antiviral Defence) proteins (Cas5-8), the Cas9 endonuclease, or the CSM/CMR complexes. These are known as Type I, II or IIIA/IIIB CRISPR systems (reviewed in Cass et al., 2015). CRISPR loci consist of the core *cas* genes (*cas1-4*, as described above), the effector protein genes described here, and the repeat regions containing the

spacer sequences. The CRISPR locus of *E. limosum* ATCC 8486 can be seen in Figure 6.2. While the *E. limosum* CRISPR system characterised *in silico* in this study is a Type I system, the Type II (with Cas9 as the effector protein) systems are of the greatest biotechnological interest.

CRISPR-based defence against invasive DNA was originally thought to rely on an RNA interference mechanism (Makarova et al., 2006). Subsequent studies resulted in the characterisation of the Cas9 endonuclease as an effector of the CRISPR system. This prompted the development of a gene editing technology which exploits the particular qualities of Cas9 (Charpentier and Marraffini, 2014, Jinek et al., 2012). This technology is referred to as CRISPR-Cas9.

The basis of CRISPR-Cas9 is that Cas9 can be targeted against DNA sequenced specified by an artificial single-guide RNA or sgRNA (in the wild-type context, Cas9 is guided by spacer-derived crRNAs complexed to a small tracrRNA which is transcribed from a locus adjacent to *cas9*). The only stipulation for the targeting of CRISPR-Cas9 is that the target sequence must have an associated PAM (proto-spacer associated motif, as described above). The development of Cas9 from bacterial defence mechanism to gene editing tool is reminiscent of the development of restriction endonucleases as biomolecular tools.

Cas9 targeted against a bacterial gene of interest provides a powerful counter-selection mechanism. In combination with a recombination cassette which removes the targeted region, CRISPR-Cas9 can be used to efficiently generate

deletion mutants (Jiang et al., 2013). CRISPR-Cas9 systems based on this principle have been exemplified in a variety of bacterial species, including *E. coli* (Jiang et al., 2015, Pyne et al., 2015) and various *Streptomyces* species (Cobb et al., 2015). All of these examples are relatively recent. The study by Pyne *et al* (2015) is notable for combining CRISPR-Cas9 genome editing and lambda Red recombineering.

CRISPR-Cas9 genome editing has also been employed in clostridia, being used recently to repress expression of the sporulation protein *spo0A* in *C. beijerinckii and C. acetobutylicum* (Li et al., 2016). The single-plasmid system employed by Li and coworkers is similar to the one employed in this study. They (Li et al., 2016) employed both standard Cas9, a mutant Cas9 nickase which produces a single-stranded breakage rather than a double-stranded one, and dCas9 (dead Cas9), which does not possess endonuclease activity but which does bind to the target sequence and interfere with target gene expression. As in this study, the CRISPR vector employed by Li and coworkers placed the Cas9 under the control of a constitutive promoter (in that study, the P_{ptb} promoter of the *C. acetobutylicum* phosphotransbutyrylase).

Interestingly, the *C. acetobutylicum* strain employed in the study by Li et al was a deletion mutant lacking a Type II restriction endonuclease, Cac824I. This deletion was accomplished prior to their CRISPR work.

In the current study, CRISPR-Cas9 was employed to generate a gene deletion mutant in *E. limosum*. The restriction endonuclease subunit of a Type I

restriction-modification system was deleted and a significant increase in DNA transfer efficiency was observed.

1.5 Forward Genetics in Clostridia

Forward genetic approaches are based on the generation of random mutants of the desired phenotype, with subsequent determination of the genotype responsible. Of particular interest in the study of clostridia is random insertional mutagenesis via the use of transposons. In this study, a clostridial transposon mutagenesis vector developed for use in *C. autoethanogenum* was used to mutagenise *E. limosum*, and the merits of this strategy evaluated.

1.5.1 Transposon Mutagenesis

The origins and development of transposon technology are detailed in Chapter 5. The transposon mutagenesis technique employed in this study is a development of previous methods employed in clostridia. Previous targets of transposon mutagenesis in clostridia include *C. difficile* (Cartman and Minton, 2010), and more recently *Clostridium perfringens* (Liu et al., 2013). Both of these studies were based on the *mariner* transposon. In both of these studies, the transposable element contained *catP* (chloramphenicol acetyltransferase), a selective marker conferring resistance to chloramphenicol and thiamphenicol. Transposon mutagenesis of *C. perfringens* had previously been demonstrated using transposases which were bound to transposons *in vitro* and

then electroporated into *C. perfringens*. This is in contrast to the strategies above, which were based on the use of replicative mutagenesis vectors containing the transposon, and from which the transposase was expressed. This was not practicable in *C. difficile*, as no electroporation method exists for that species, with all gene transfer accomplished via conjugation (Cartman and Minton, 2010). Accordingly, a novel transposon mutagenesis vector was developed and employed. Mapping of insertion sites demonstrated that the transposon inserted randomly, with single-insert mutants constituting the majority (98.3%) of the mutants assayed. This was in contrast to previous systems utilised in *C. difficile*, which were based on conjugative transposons and which had been found to possess strong insertion-site biases and a tendency towards producing multiple insertions per transposon mutant (Hussain et al., 2005, Wang et al., 2006). Both of these features are undesirable in a mutant library, the former because it increases the proportion of the target genome which is uncovered by insertions, and the latter because it complicates genotypic analysis of phenotypically-selected mutants. Following the example of Cartman and Minton, several subsequent transposon mutagenesis studies in clostridia have utilised the plasmid-based *mariner* transposon system (e.g. Zhang et al., 2015), as does this study.

1.5.2 High-throughput Transposon Sequencing

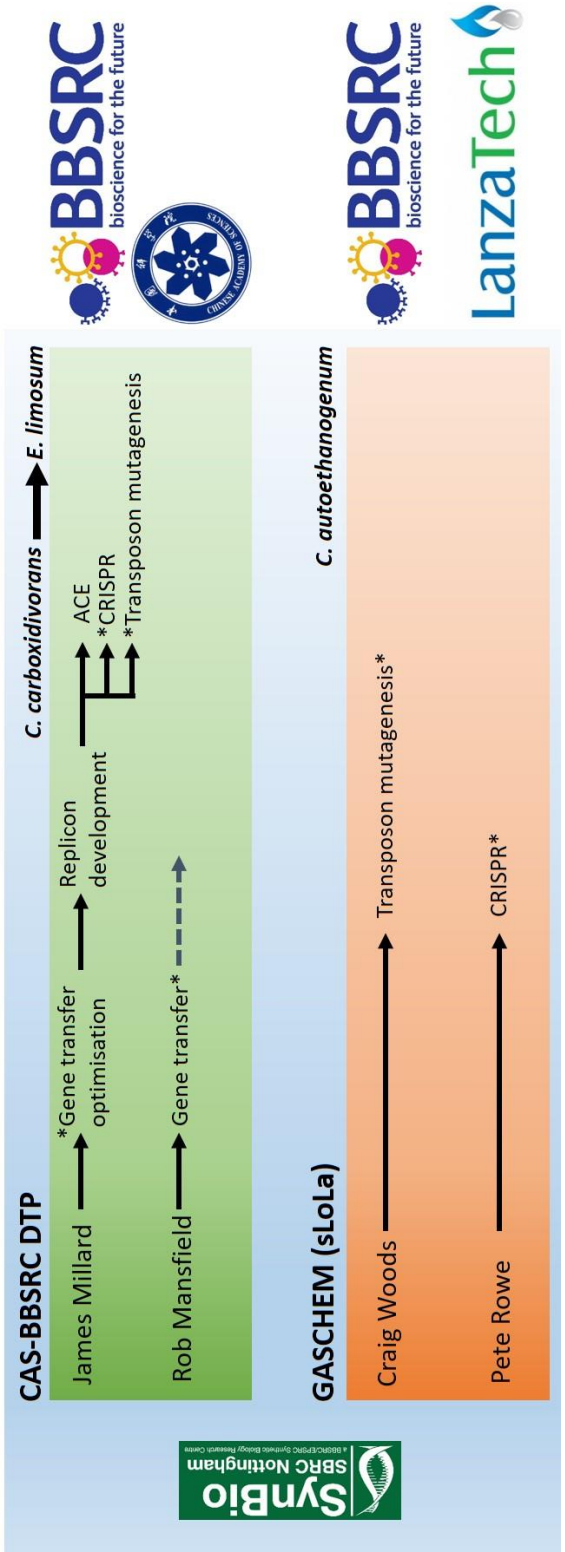
Several methods exist in which insertional mutagenesis via transposons is combined with high-throughput sequencing technologies. Examples include

Insertion Sequencing (INSeq) (Goodman et al., 2009), Transposon Sequencing (Tn-seq) (van Opijnen et al., 2009), High-Throughput Insertion Tracking by Deep Sequencing (HITS) (Gawronski et al., 2009) and Transposon-Directed Insertion Site Sequencing (TraDIS) (Langridge et al., 2009) (reviewed in van Opijnen and Camilli, 2013). Both Tn-seq and INSeq are based on the *mariner* transposon, with both methods making a single base change in the transposon inverted repeat region in order to introduce a restriction enzyme cut site for downstream sample processing purposes. These high-throughput methods are valuable because they allow the determination of gene essentiality through the sequencing of large mutant pools.

1.6 Project Aims

Acetogenic bacteria are of interest due to their ability to produce a variety of acids and solvents using single-carbon gases as a sole carbon and energy source. It has been noted that the main obstacles to utilisation of acetogens for industrial processes are a lack of understanding of their energy metabolism and a lack of genetic tools for the modification of these organisms (Bertsch and Muller, 2015). This study began as an attempt to develop the acetogenic species *Clostridium carboxidivorans* as a potential industrial process strain. This project was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and the Chinese Academy of Science (CAS), and

conducted via a PhD project within a BBSRC Doctoral Training Programme



(DTP) (refer to Figure **1.2**, below).

However, this species proved to be genetically intractable. Following unsuccessful attempts to establish gene transfer in *C. carboxidivorans*, the focus of the project was switched to the acetogen *Eubacterium limosum*. The present study aims to extend and exemplify the range of tools available for the genetic modification of *E. limosum* ATCC 8486, facilitating future development of this strain into a potential industrial process organism.

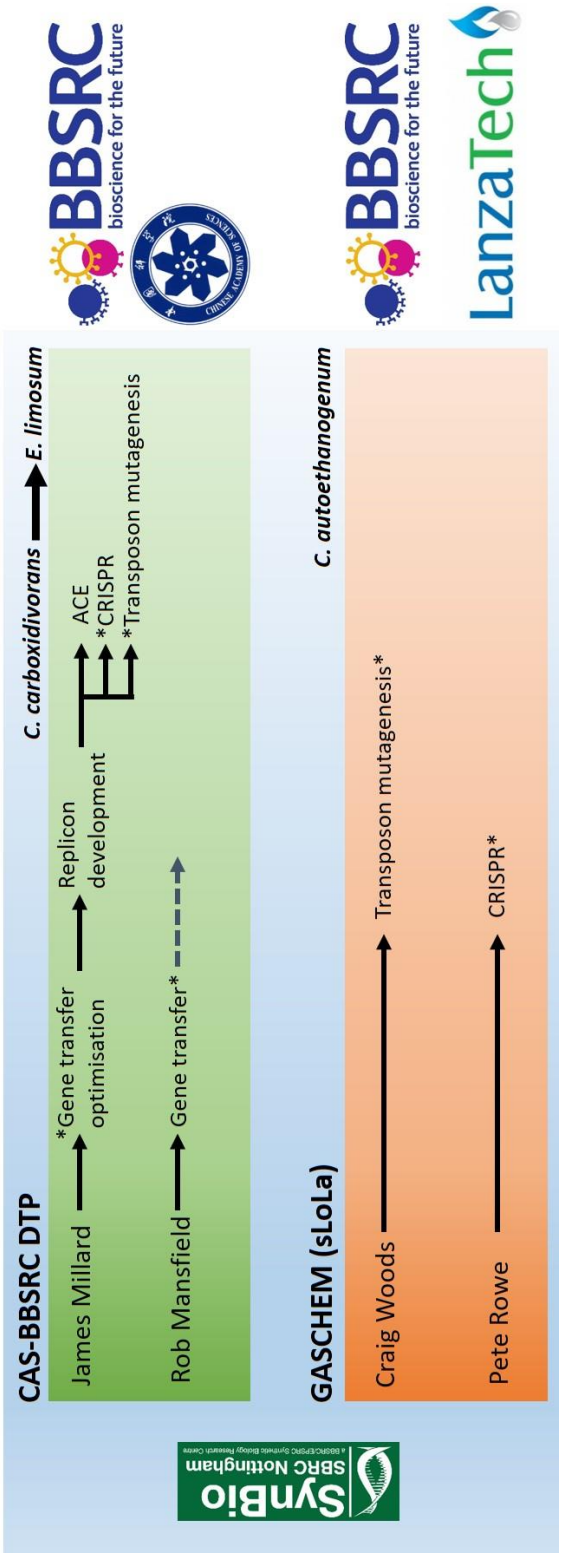


Figure 1.2: PhD project overview. All work during this PhD was conducted within SBRC Nottingham. This project employed and expanded upon molecular tools developed during the parallel GASCHEM project.

CHAPTER TWO

MATERIALS AND METHODS

2 Materials and Methods

2.1 Bacterial Strains and Plasmids

2.1.1 Bacterial Strains

Table 2.1: List of bacterial strains used in this study. DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH.

Strain	**Description/Genotype**	**Source/Reference**
A. woodii DSMZ 1030	Acetogenic species; type strain.	DSMZ; W. E. Balch, 1977
B. subtilis 168	Gram-positive aerobe.	SBRC Culture Collection (Nottingham, UK)
C. autoethanogenum DSMZ 10061	Acetogenic clostridial species.	DSMZ; J. Abrini 1994
Clostridium carboxidivorans P7 (a.k.a. DSMZ-15243)	Solventogenic and acetogenic clostridial species; type strain.	DSMZ; Liou et al 2005.
Clostridium beijerinckii 59B	Solventogenic clostridial species. Isolated by Dr. G. Little.	Dr. Gareth Little (SBRC, Nottingham, UK).
C. ljungdahlii DSMZ 13528	Acetogenic clostridial species; type strain.	DSMZ; R. Tanner 1993
Eubacterium limosum ATCC 8486 (a.k.a. DSMZ-20543)	Acetogenic clostridial species; type strain.	DSMZ; A. H. Eggerth, 1935
Eubacterium limosum DSMZ-2593	Acetogenic clostridial species.	DSMZ; A. H. Eggerth, 1935
Eubacterium limosum DSMZ-2594	Acetogenic clostridial species.	DSMZ; A. H. Eggerth, 1935
Clostridium pasteurianum	Solventogenic clostridial species.	Dr. Alexander Grosse-Honebrink; Angel Pech-Canul (SBRC, Nottingham, UK)
E. coli DH5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96*	New England Biolabs

	recA1 relA1 endA1 thi-1 hsdR17	
Geobacillus thermoglucosidasius DSMZ 2542	Thermophilic, solventogenic Gram-positive species.	DSMZ; Y. Suzuki 1984

2.1.2 **Plasmids**

Table 2.2: A complete list of plasmids used in this study.

Plasmid	**Description**	**Source/Reference**
pMTL82151	pMTL80000-series plasmid, having the pBP1 Gram-positive replicon, *catP* selective marker, ColE1 Gram-negative replicon, *traJ* conjugation locus, and a multiple cloning site (MCS).	(Heap et al., 2009)
pMTL83151	As above, with the pCB102 Gram-positive replicon.	(Heap et al., 2009)
pMTL84151	As above, with the pCD6 Gram-positive replicon.	(Heap et al., 2009)
pMTL85151	As above, with the pIM13 Gram-positive replicon.	(Heap et al., 2009)
pMTL87151	As above, with the P19 Gram-positive replicon.	This Study
pMTL87151A	As pMTL87151, with the putative promoter sequence of the P19 *rep* gene removed.	This Study
pMTL87151B	As pMTL87151, with both of the iterons (binding sites) of the P19 Rep protein removed.	This Study
pMTL87151C	As pMTL87151, with one of the iterons (binding sites) of the P19 Rep protein removed.	This Study
pMTL-AMH101	*pyrE* ACE knockout vector designed for use in *C. autoethanogenum*	(Liew et al., 2017)
pMTL-JM101	*pyrE* ACE knockout vector designed for use in *E. limosum*.	This Study
pSBRCas9	CRISPR-Cas9 vector designed for use in clostridia.	P. Rowe (SBRC, Nottingham, UK)
pMTL431511::ORFD	CRISPR-Cas9 knockout vector targeted against *E. limosum* Type I restriction	This Study

	endonuclease *R.Eli8486ORFDP*.	
pMTL-Tet3nO	CatP reporter plasmid with TetR tetracycline-inducible system.	SBRC, Nottingham, UK
pMTL-JM102	As pMTL-Tet3nO, with the Gram-positive replicon replaced by pCB102 for use in *E. limosum*.	This Study
pMTL-CW22	Inducible transposase vector designed for use in clostridia. Possesses the pCD6 Gram-positive replicon.	C. Woods (SBRC, Nottingham, UK)
pMTL-CW21	As pMTL-CW22, with the Gram-positive replicon replaced with pCB102 for use in *E. limosum*.	This Study
pMTL-CW23	As pMTL-CW22, with the Gram-positive replicon replaced with P19 for use in *E. limosum*.	This Study
pMTL-CW30	Non-replicative “suicide” transposon vector based on pMTL-CW22, with the Gram-positive replicon removed.	This Study

2.2 Oligonucleotide Primers

Table 2.3: A complete list of oligonucleotide primers used in this study. Underlined sections denote restriction sites, with the specific enzyme being noted in the function column.

Oligonucleotide	Sequence	Function
P19_FWD	ATATATGGCGCGCCTTGATATGAAAAATTTAAGGAATAATTTTGGAAAAGTAAG	P19 replicon forward modularisation primer with AscI site.
P19_REV	ATATATGGCCGGCCTTAACCAAACCTAATTATAGCCATTTGGG	P19 replicon reverse modularisation primer with FseI site.
P19_noProm_FWD	ATATATGGCGCGCCATGGACAACTTTGCATATTGTGATCATTTATTTGATAG	P19A forward truncation primer with AscI site. Removes putative promoter sequence and RBS from P19 *rep* gene.
P19_noRepeats_REV	ATATATGGCCGGCCCTATTTTAATAATCTTATAACCTTTGTTCTATTTATATCTAAAGCTCTTGCTATTTC	P19B reverse truncation primer with FseI site. Removes both iterons from the modular replicon.
P19_oneRepeat_REV	ATATATGGCCGGCCTCTATATCTTTTTTAATAAAAAACAAAAATCAATTACGAGTACTTTGC	P19C reverse truncation primer with FseI site. Removes a single iteron from the modular replicon.
Trans_P19_fwd	TCTTTTAGGTGAAATACCTAGAAGACTTTTTTATTG	Forward sequencing primer binding to pMTL80000 backbone and reading through Gram-positive replicon.
Trans_P19_rev	CATCTAAGTTCCCTCTCAAATTCAAGTTTATC	Reverse sequencing primer binding to pMTL80000 backbone and reading through Gram-positive replicon.

P19 replicon 179-201 FWD	CGCTTAGCCAAAATG CGGGAACA	Plasmid qPCR forward primer for replicon P19.
P19 replicon 230-257 REV	GTTCTCTGGAAGTCG TTTAGGTACCTGT	Plasmid qPCR reverse primer for plasmid P19.
pMTL Rep3 1065-1087 FWD	TGAGAGTGCCGACA CAGTATGCA	Plasmid qPCR forward primer for replicon pCB102.
pMTL Rep3 1173-1194 REV	TCGTCGTGAAGTCCC ACCGACA	Plasmid qPCR reverse primer for replicon pCB102.
ispE_FWD	TCGGGTCGCTGAAAA GCTGCAG	Chromosomal reference forward primer for *E. limosum* qPCR.
ispE_REV	CAGCCAGTCCCGCTT CACTTGG	Chromosomal reference reverse primer for *E. limosum* qPCR.
Cbei59B_aldH_ LF65_04317 654-675 FWD	TCTTTGCGGAACTGG AGGTCCA	Chromosomal reference forward primer for *C. beijerinckii* qPCR.
Cbei59B_aldH_ LF65_04317 711-732 REV	TCCAGCGCCAGCACC TATAGCT	Chromosomal reference reverse primer for *C. beijerinckii* qPCR.
Cbxi_G_FWD_2	AGGCTGGTGGACAA GTTGGTGA	Chromosomal reference forward primer for *C. carboxidivorans* qPCR.
Cbxi_G_REV	TGTCCATGTCGTTGC ATCTGGC	Chromosomal reference reverse primer for *C. carboxidivorans* qPCR.
catP_INV_F1	TATTGTATAGCTTGG TATCATCTCATCATA TATCCCCAATTCACC	Mini-transposon forward primer for inverse PCR.
ELIM-LHAF-v2	ATATATGCTAGCTGC GGGTGTTGCTGCTAT TATTTTATC	Long homology arm (LHA) forward primer for *E. limosum pyrE* knockout vector. Contains NheI site.

ELIM-LHAR	ATATATGGCGCGCCGTGCTACAATTGGTCAGGAAAAG	LHA reverse primer for *E. limosum pyrE* knockout vector. Contains AscI site.
ELIM-SHA-F	ATATATCCGCGGGACCTTTGGCGATTTCGTG	Short homology arm (SHA) forward primer for *E. limosum pyrE* knockout vector. Contains SacII site.
ELIM-SHA-R	ATATATGCGGCCGCGATGGCCTTGTCGCCGTCCTGGGGCTTGTAGCC	SHA reverse primer for *E. limosum pyrE* knockout vector. Contains NotI site.
catP_INV_R1	TATTTGTGTGATATCCACTTTAACGGTCATGCTGTAGGTACAAGG	Mini-transposon reverse primer for inverse PCR.
catP-INV-R2	GGCAAGTGTTCAAGAAGTTATTAAGTCGGGAGTGCAGTCGAAGTGG	Mini-transposon sequencing primer.

2.3 Other Oligonucleotides

Table 2.4: A complete list of non-primer oligonucleotides used in this study.

Oligonucleotide	**Sequence**	**Function**
sgRNA_D1	ATCTTAAGGAGGAGTTTTCGTCGACGGGATGGACAGAAAAAGTGGGTTTTAGAGCTAGAAATAGCAAGTT	sgRNA oligonucleotide for Eli8486ORFDP CRISPR knockout vector

2.4 Chemicals and Suppliers

Table 2.5: Chemical suppliers and associated chemicals used in this study.

Organisation/Supplier	**Products Used**
New England Biolabs	Monarch® Plasmid Miniprep and Gel Extraction Kits, Restriction Endonucleases, Q5® High-Fidelity DNA Polymerase, HiFi® DNA Assembly Master Mix, 5-alpha® *E. coli* competent cells, Antarctic Phosphatase, GeneRuler® 1Kb Plus DNA Ladder, GpC (M.CviPI) and CpG (M.SssI) *in vitro* methyltransferases.
Promega	T4 DNA ligase
Oxoid	Compounds used for preparation of growth media.
Sigma-Aldrich	Compounds used for preparation of growth media, GenElute™ Bacterial Genomic DNA Kit.
QIAGEN Ltd.	Proteinase K.
Sartorius	Minisart® syringe filters
Thermo Fisher	Compounds used for preparation of growth media, SYBR Safe DNA Gel Stain

2.5 Bio-informatics Tools

2.5.1 DNA Sequence Analysis

Visualisation of DNA sequences was routinely conducted within ApE (A Plasmid Editor, developed and freely distributed by M. Wayne Davis) and Benchling (Benchling Ltd). The BLASTn algorithm (NCBI) was used to

conduct searches with DNA sequences against a DNA nucleotide database. For construction of 16S ribosomal RNA trees, 16S rRNA sequences were downloaded from RDP (https://rdp.cme.msu.edu/index.jsp) and assembled into trees using the online tool suite at http://www.phylogeny.fr/.

2.5.2 Protein Sequence Analysis

The BLASTp algorithm (NCBI) was used to conduct searches with amino acid sequences against a protein sequence database, and to identify functional domains within amino acid sequences. The PFAM online tool (http://pfam.xfam.org/) by EMBL-EBI was used to identify functional domains within an amino acid sequence.

2.5.3 Plasmid Design and Visualisation

Routine plasmid design and visualisation in this study was conducted within the Benchling graphical interface (Benchling Ltd). Plasmid maps were prepared using SnapGene and edited with graphical software including Microsoft PowerPoint and Inkscape[1].

2.5.4 Oligonucleotide Design

Annealing regions for oligonucleotide primers used in this study were selected manually from DNA sequences. They were assessed for suitable melting

1 https://inkscape.org/en/

temperature using Benchling (Benchling Ltd). Where possible, primers were sized to possess a melting temperature of 60^{o}C.

2.6 Data Analysis and Presentation

Statistical analysis of results was performed using GraphPad Prism. Graphs for the presentation of data were also constructed in GraphPad Prism.

2.7 General Microbiological Techniques

2.7.1 Growth Media

In all cases, agar medium was prepared by adding No.1 Bacteriological Agar (Oxoid) to 1.5% w/v to liquid media prior to autoclaving. All media was sterilised by autoclaving at 121^{o}C, unless specifically stated.

2.7.2 Media Supplements

All media supplements were prepared in stock solutions as shown in Table 2.6, below, and stored at the appropriate temperature. All stock solutions were sterilised via filtration, using a 0.2-µm filter.

Table 2.6: Media supplements used in this study.

Supplement	**Stock (mg ml^{-1})**	**Solvent**	**Working conc. (*E. coli*) (µg ml^{-1})**	**Working conc. (clostridia) (µg ml^{-1})**
Erythromycin	50	100% EtOH	500	-
Clarithromycin	6	DMSO	-	6
Chloramphenicol	25	100% EtOH	25	-
Thiamphenicol	15	50% EtOH	-	7.5
Anhydrotetracycline	0.1	EtOH	-	0.1
Ampicillin	100	H_2O	100	-

2.7.3 Anaerobic culture of microorganisms

Unless specified, all anaerobic culture of microorganisms took place in an anaerobic cabinet (Don Whitley Scientific Limited, Shipley, UK). Growth took place at 37^{o}C, in an atmosphere consisting of N_2, CO_2 and H_2 in the ratio 80:10:10.

2.7.4 Routine Culture of *Escherichia coli*

For the purposes of routine culture, *E. coli* was grown in LB broth (Sigma-Aldrich, St. Louis, Missouri, USA) at 37^{o}C. Selective antibiotics were used when appropriate at the following concentrations: erythromycin 500 µg ml^{-1}; chloramphenicol 25 µg ml^{-1}; ampicillin 100 µg ml^{-1}.

2.7.5 **Routine Culture of *Eubacterium limosum***

E. limosum was grown in anaerobic conditions at 37°C, within an anaerobic cabinet (Don Whitley Scientific Limited, Shipley, UK). *E. limosum* was cultured in a modified *A. woodii* medium (AWM) (Dr. Philippa Strong, personal communication) consisting of KH_2PO_4 (1.76 g L-1), K_2HPO_4 (8.44 g L-1), NH_4Cl (1.0 g L^{-1}), l-Cysteine hydrochloride (Sigma; 0.5 g L^{-1}), NaCl (2.9 g L^{-1}), yeast extract (oxoid; 2.0 g L^{-1}), $KHCO_3$ (6.0 g L^{-1}), resazurin (0.1 ml L^{-1} of 1% v/v solution), trace Element Solution SL9 (1.0 ml L^{-1}), selenite-tungstate solution (1.0 ml L^{-1}), and vitamin solution DSMZ 141 (2.0 ml L^{-1}). Media components were dissolved in dH_2O, made up to a final volume of 1 L, and autoclaved at 120°C. Following autoclaving, media were placed within an anaerobic cabinet for a minimum of 12 hours. Once anaerobic, fructose was added from a sterile, anaerobic 2 M stock to a final concentration of 20 mM, and $MgSO_4$ was added from a sterile, anaerobic 0.75 M stock to a final concentration of 1.5 mM. All chemicals were sourced from Sigma-Aldrich unless otherwise noted.

Trace Element Solution SL9 consists of nitriloacetic acid (12.8 g L^{-1}), $FeCl_2.4H_2O$ (2.0 g L^{-1}), $ZnCl_2$ (0.070 g L^{-1}), $MnCl_2.4H_2O$ (0.1 g L^{-1}), H_3BO_3 (0.006 g L^{-1}), $CoCl_2.6H_2O$ (0.19 g L^{-1}), $CuCl_2.2H_2O$ (0.002 g L^{-1}), $NiCl_2.6H_2O$ (0.024 g L^{-1}), and $Na_2MoO_4.2H_2O$ (0.036 g L^{-1}). Nitriloacetic acid was dissolved prior to all other components, and the solution adjusted to pH 6.0 with NaOH. Selenite-tungstate solution consists of NaOH (0.5 g L^{-1}), $Na_2SeO_3.5H_2O$ (0.003 g L^{-1}), and $Na_2WO_4.2H_2O$ (0.004 g L^{-1}).

Vitamin solution DSMZ 141 consists of biotin (2 mg L^{-1}), folic acid (2 mg L^{-1}), pyridoxine hydrochloride (10 mg L^{-1}), thiamine hydrochloride dihydrate (5 mg L^{-1}), riboflavin (5 mg L^{-1}), nicotinic acid (5 mg L^{-1}), calcium pantothenate (5 mg L^{-1}), vitamin B12 (0.1 mg L^{-1}), p-aminobenzoic acid (5 mg L^{-1}), and lipoic acid (5 mg L^{-1}).

For plates, agar was added to 1.5% w/v, as noted above. Fructose and $MgSO_4$ were added to plates immediately prior to pouring, under aerobic conditions, from sterile, aerobic stocks. Once poured, plates were placed in an anaerobic cabinet for a minimum of 6 hours prior to inoculation. For selective media, antibiotics were added to the following concentrations when appropriate: thiamphenicol 7.5 or 15 µg ml^{-1}; clarithromycin 6 µg ml^{-1} (as described in Table 2.6).

During generation of *pyrE* KO mutants, *E. limosum* was grown on phosphate-buffered basal media (PBBM). This consists of the following (per L): NaCl (0.9 g), $MgSO_4.2H_2O$ (0.2 g), $CaCl_2.H_2O$ (0.2 g), NH_4Cl (1 g), Trace Mineral Solution (10 ml), Vitamin Solution (10 ml), Phosphate Buffer (1M, pH 7.2; 10 ml), L-Cysteine HCL (0.04% W/V), Yeast Extract (2 g), and distilled water (to 1 l). The pH of PBBM was adjusted to 7.2.

PBBM Trace Mineral Solution consists of the following (per L): NTA (1.5 g), $FeSO_4.7H_2O$ (0.1 g), $MnSO_4.4H_2O$ (0.1 g), $CoCl_2.6H_2O$ (1.7 g), $ZnCl_2$ (0.1 g), $CaCl_2.2H_2O$ (0.1 g), $CuCl_2.2H_2O$ (0.02 g), H_3BO_3 (0.01 g), Na_2SeO_3 (0.017 g), $NiSO_4.6H_2O$ (0.026 g), and NaCl (1 g).

PBBM Phosphate Buffer consists of the following: KH_2PO_4 (1M), Na_2HPO_4 (1M), mixed to yield pH 7.2.

PBBM Vitamin Solution consists of the following (per L): Biotin (2 mg), Folic acid (2 mg), Pyridoxine HCL (10 mg), Thiamine HCL (5 mg), Riboflavin (5 mg), Nicotinic acid (5 mg), Pantothenic acid (5 mg), Cyanocobalamin (0.1 mg), *p*-aminobenzoic acid (5 mg), and Lipoic acid (5 mg).

2.7.6 **Routine Culture of *Clostridium carboxidivorans***

C. carboxidivorans was grown in YTF (Yeast, Tryptone, Fructose) medium at 37^{o}C under anaerobic conditions within an anaerobic workstation, as described above. YTF consisted of yeast extract (10 g l^{-1}), tryptone (16 g l^{-1}), fructose (10 g l^{-1}), NaCl (0.2 g l^{-1}), acidic trace element solution (1 ml), basic trace element solution (1 ml), vitamin solution DSMZ 141 (1 ml). Acidic trace element solution consists of (per l) HCl (50 mM), H_3BO_3 (100 mg), $MnCl_2.4H_2O$ (230 mg), $FeCl_2.4H_2O$ (780 mg), $CoCl_2.6H_2O$ (103 mg), $NiCl_2.6H_2O$ (602 mg), $ZnCl_2$ (78 mg), $CuSO_4.5H_2O$ (50 mg), $AlK(SO_4)_2.12H_2O$ (50 mg). Basic trace element solution consists of (per l) 10 mM NaOH, Na_2SeO_3 (58 mg), Na_2WO_4 (53 mg), $Na_2MbO_4.2H_2O$ (52 mg). DSMZ vitamin solution 141 was as described in section 2.7.5, above. YTF was not pH-adjusted prior to culture. All medium was placed overnight in an anaerobic cabinet to de-gas prior to inoculation.

2.7.7 Routine Culture of *Clostridium beijerinckii*

C. beijerinckii was grown in 2xYTG (Yeast, Tryptone, Glucose) medium at 37ºC under anaerobic conditions within an anaerobic workstation, as described above. All medium was placed overnight in an anaerobic workstation to de-gas prior to inoculation. 2x YTG consisted of 2x YT powder (31 g l^{-1}) (Sigma-Aldrich) with D-glucose added after autoclaving to a final concentration of 16 mM.

2.7.8 Routine Culture of *Bacillus subtilis*

B. subtilis was cultured in LB broth (Sigma-Aldrich) within a shaking incubator at 37ºC under aerobic conditions and vigorous agitation (270 RPM).

2.7.9 Growth Curves

To generate turbidity measurements of clostridial or *E. coli* growth, culture samples were taken at appropriate time points and diluted 1:1 or 1:3 (v/v) in appropriate media. Sample turbidity was measured at 600 nm using a Jenway spectrophotometer.

2.7.10 Strain Storage of *E. coli*

E. coli strains harboring plasmid DNA were stored at -80ºC using the Microbank Long Term Bacterial and Fungal Storage System (Pro-Lab Diagnostics).

2.7.11 Strain Storage of Gram-positive organisms

Cells of Gram-positive species (including *C. carboxidivorans, E. limosum, C. beijerinckii*, *C. ljungdahlii* and *C. autoethanogenum*) were prepared for -80ºC storage by the addition of dimethyl sulfoxide (DMSO; Sigma-Aldrich) to a concentration of 10% v/v. Strains and competent cell aliquots of *E. limosum* were stored for several months with no obvious loss of viability.

2.7.12 Plasmid Segregational Stability Assay

To assess the segregational stability of plasmids, competent cells of *E. limosum* were transformed with 5 µl of pMTL83151, pMTL87151 or pMTL87151C. Following electroporation and recovery (performed as described in section 2.9.2), cells were plated onto AWM supplemented with thiamphenicol. Transformant colonies were picked in triplicate and grown overnight under selective conditions. Following overnight growth, dilution series were prepared and 50 µl 10^{-2}-10^{-6} dilution were spread on selective and non-selective plates for each replicate. Following 48 hours of growth, colonies were enumerated on these T0 plates. Plasmid retention was assessed as the percentage of thiamphenicol-resistant colonies relative to the number of colonies on non-selective plates. After plating of the initial overnight cultures, subcultures of non-selective media were inoculated with 1% V/V inoculum from the initial overnight cultures. Following overnight growth of the subcultures, dilution series from these cultures were plated in the way described above, with colonies being enumerated for both media conditions after allowing 48 hours for growth. Additional passages in non-selective

media were conducted and enumerated in the same way for a one-week period.

2.8 General Molecular Techniques

2.8.1 Quantification of DNA

Quantification of DNA was carried out using a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Scientific).

2.8.2 Preparation of Genomic DNA from Clostridia

For routine purposes such as inverse PCR, genomic DNA was prepared using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich) according to the manufacturer's instructions.

For genome re-sequencing of clostridial strains (including *E. limosum*) via Illumina MiSeq or PacBio instruments, genomic DNA was prepared using phenol-chloroform extraction as follows: 5 ml of *E. limosum* culture at an OD_{600} of >0.5 was collected by centrifugation at 6000 × *G* and the supernatant discarded. The pellet was resuspended in a lysis buffer consisting of 45 mg/ml lysozyme (Sigma-Aldrich) dissolved in PBS. The mixture was incubated at 37°C for 30 minutes, with occasional agitation. 85 µl of dH_2O, 25 µl Proteinase K (Qiagen), and 110 µl of 10% w/v sodium dodecyl sulfate (SDS) were then added to the sample and the sample mixed by inversion. The sample was then incubated at 65°C for 30 minutes, with occasional agitation. 400 µl

phenol chloroform-isoamyl alcohol (Sigma-Aldrich) were then added, and the entire sample added to a Phase Lock Gel Heavy tube (Eppendorf). The sample was then centrifuged at 17000 × *G* for 3 minutes and the top phase in the tube transferred to a fresh 1.5 ml centrifuge tube. 5 µl of RNAse A (Sigma-Aldrich) was then added, and the sample was incubated at room temperature for 30 minutes. The sample was then transferred to a fresh phase lock tube containing 400 µl of phenol chloroform-isoamyl alcohol and mixed by inversion. The phase lock tube containing the sample was then centrifuged at 17000 × *G* for 3 minutes, and the top phase transferred to a fresh phase lock tube containing 400 µl phenol chloroform-isoamyl alcohol. Sample was mixed and centrifuged as before, and the top layer transferred to a microcentrifuge tube containing 40 µl of 3 M sodium acetate and 800 µl of ethanol, which had previously been chilled at -80^{o}C. If necessary, samples were stored for extended periods at this time. To continue sample preparation, samples were returned to room temperature and centrifuged at 17000 × *G* for 15 minutes. Supernatant was carefully discarded, and 1 ml of 70% ethanol added, the sample mixed by inversion, and then the tube centrifuged at 17000 × *G* for 3 minutes. Supernatant was discarded, the sample centrifuged again as before, and any remaining ethanol removed by pipetting. The pellet was then air-dried for 45 minutes under sterile conditions, and then re-dissolved in 50 µl sterile H_2O. Purity and concentration of the resulting genomic DNA preparation was assessed spectrophotometrically using a NanoDrop UV-Vis spectrophotometer.

2.8.3 Preparation of Plasmid DNA

Plasmid DNA was prepared from *E. coli* using the Monarch Plasmid Miniprep Kit (NEB), according to the manufacturer's instructions.

2.8.4 Restriction Enzyme Digests

Restriction digests were performed using enzymes from New England Biolabs, with reaction conditions as per the manufacturer's instructions. Reactions typically consisted of 1-10 µl of plasmid or genomic DNA, 0.5-1.0 µl of each restriction enzyme being employed, and a 1X concentration of the appropriate restriction endonuclease buffer (typically NEB CutSmart©). Reactions were made up to 20-50 µl with sterile H_2O. Restriction digests were timed to allow digestion of 1 µG of DNA per unit of enzyme per hour, as recommended by the manufacturer.

2.8.5 De-phosphorylation of Linear DNA

De-phosphorylation of digested plasmid DNA or of PCR-generated DNA fragments was performed using Antarctic Phosphatase (NEB) according to the manufacturer's instructions. De-phosphorylation reactions were incubated at 37^{o}C for 20 minutes to 2 hours, depending on the concentration of input DNA.

2.8.6 Ligation Reactions

Ligation reactions were performed using T4 DNA Ligase (Promega or NEB), according to the manufacturer's instructions. Reactions were carried out at 22^{o}C (with samples placed in a heat block), or overnight at 4^{o}C, or on melting

ice in order to provide a temperature gradient. DNA fragments were typically included at an insert to vector ratio of 3:1. Ligations were typically conducted in volumes of 20 µl. Ligation reactions were not typically dialysed prior to electroporation.

2.8.7 Agarose Gel Electrophoresis

Gels for DNA electrophoresis were made using agarose powder suspended in TAE buffer (40 mM Tris, 1 mM EDTA, 0.1% Glacial Acetic Acid) at a concentration of 0.8% or 1% w/v. SYBRSafe DNA Gel Stain (Thermo Fisher) was added at a concentration of 0.01% v/v. Samples for gel electrophoresis were mixed with NEB loading buffer prior to loading into the gel. Samples were resolved on the gel at 111-120 V for 15-40 minutes, with the gel submerged in TAE buffer. Gels were run with 5 ul of GeneRuler 1 kb Plus (Thermo Fisher) as a marker.

2.8.8 DNA Extraction from Agarose Gel

DNA fragments were visualised under blue light and extracted using a scalpel. DNA was purified from gel fragments using a Monarch® Gel Purification Kit (NEB) as per the manufacturer's instructions. Purified DNA was eluted in 20 µl of sterile dH_2O.

2.8.9 Polymerase Chain Reaction

PCRs were conducted using Q5® DNA polymerase (NEB), using a "touchdown" PCR programme. PCR primers were designed as previously described. For these touchdown PCRs, an initial annealing temperature of

70°C was employed. Subsequent cycles had decreasing annealing temperatures (decrementing at 1.5°C per cycle) until the annealing temperature was 60°C. Cycling then continued at this annealing temperature for 30-55 cycles. Extension times of 30 seconds per Kb were employed. This programme is displayed diagrammatically in Table 2.7.

Table 2.7: A typical PCR programme as used in this study. * Note that during touchdown PCR, annealing temperatures were decremented by 1.5°C after every cycle, until they reached 60°C.

Cycle Step	**Temperature (°C)**	**Duration (s)**	**Program Cycle**
1. Initial Denaturation	95	300	
2. Denaturation	95	30	
3. Annealing	70-60*	60	
4. Elongation	72	30 / Kb	Return to step 2 (30-55X)
5. Final Elongation	72	300	
6. Hold	15	-	

2.8.10 Splicing by Overlap Extension (SOEing) Polymerase Chain Reaction

SOEing PCRs were performed as above, utilising a touchdown protocol with elongation times adjusted to correspond to the expected spliced product length. In addition to the fragments to be spliced together, the primers used to generate the distal ends of the two fragments were also added to each reaction.

2.8.11 DNA Fragment Sequencing

Sanger sequencing of DNA was performed by Eurofins Genomics (Ebersberg, Germany).

2.9 Transformation Procedures

2.9.1 Preparation of *Eubacterium limosum* Electro-competent Cells

To generate competent cell aliquots for electro-transformation, overnight precultures of the liquid medium described above were inoculated from -80ºC stocks of *E. limosum*. A dilution series was created from the initial culture prior to overnight incubation. The following morning, turbidity of the precultures was determined spectrophotometrically at 600 nm. The most turbid preculture with an OD_{600} of <0.5 was used to inoculate 400 ml of liquid medium. A 1% v/v inoculum was used. The larger culture was grown to an optical density of 0.28-0.3, and then dispensed into ice-cold 50 ml centrifuge tubes. These were capped tightly, transferred out of the anaerobic cabinet, and centrifuged at 7000 × *G* for 10 minutes at 4ºC. Tubes were returned to the anaerobic cabinet, supernatant discarded under anaerobic conditions, and pellets resuspended in ice-cold electroporation buffer (EPB) consisting of sucrose (270 mM), Na_2HPO_4 (5 mM) and NaH_2PO_4 (5 mM), with a pH of 7.4. Cells were pelleted and resuspended twice more in identical conditions, and finally resuspended in 2 ml of EPB with 10% v/v DMSO. The resulting cell suspension was pipetted into pre-chilled microcentrifuge tubes in 100 ul aliquots. Competent cell aliquots were frozen at -80ºC or used immediately. Flash-freezing of cell aliquots in liquid nitrogen was not found to improve transformability. Aliquots of 500 µl were employed periodically for multiple

simultaneous electroporations, and the extra pipetting to divide the thawed cells was not found to effect transformability.

2.9.2 Electroporation of *Eubacterium limosum*

Aliquots of *E. limosum* competent cells were thawed on ice for 5 minutes under anaerobic conditions, and kept ice-cold until immediately prior to electroporation. 5 µl of plasmid (adjusted to final volume with H_2O as necessary) was then added directly to the thawed cells, and the mixture agitated gently by flicking. Cells were incubated with plasmid for five minutes, and then transferred to an ice-cold electroporation cuvette. Cells were electroporated using a Bio-Rad Gene Pulser Xcell system (Bio-Rad Laboratories, California, USA) using Time Constant Mode, with the following settings: 1.5 kV, 6 ms (unless otherwise noted above). Cells were immediately recovered in 1 ml liquid medium, and incubated for 4 hours at 37^{o}C in a microcentrifuge tube. Following recovery, cells were plated on selective media as appropriate. Depending on the construct, transformant cells were observed 24-72 hours following plating.

2.9.3 Transformation of *Bacillus subtilis*

A transformation protocol based on the method of Spizizen *et al* was employed (Anagnostopoulos *et al*., 1961). *B. subtilis* CU2189 (derived from *B. subtilis* 168) was streaked out onto LB agar medium and grown overnight at 37^{o}C. Following overnight development of colonies, total colony material was collected from the plate and inoculated into 4.5 ml of Medium A. Medium A consists of (per 100 ml): 81 ml sterile H_2O, 10 ml 10x Medium A base, and 9

ml 10x Bacillus Salts. 10x Medium A base consists of: 10 g L^{-1} yeast extract (Sigma), 2 g L^{-1} casamino acids (Sigma), and sterile H_2O to 900 ml. These constituents were mixed and autoclaved, and then filter-sterilised 50% w/v glucose was added, to 1 l. 10x Bacillus salts consists of $(NH_4)_2SO_4$ 20 g L^{-1}, $K_2HPO_4.3H_2O$ 183 g L^{-1}, KH_2PO_4 60 g L^{-1}, sodium citrate 2 g L^{-1}, and sterile H_2O to 1 L. The optical density of the culture was measured at 600 nm, and the culture diluted as necessary to adjust the OD to 0.1-0.2. The culture was removed afterwards to maintain the volume of the culture at 4.5 ml. The culture was then incubated with vigorous aeration at 37°C. OD_{600} measurements were taken every 30 minutes, and incubation was continued for 90 minutes after the end of the exponential phase. 0.05 ml of this grown culture was transferred to 0.45 ml of pre-warmed Medium B, and incubated at 37°C with vigorous aeration for 90 minutes. At this point, DNA was added to the cells and they were incubated for a further 30 minutes. Transformant colonies were spread onto LB agar supplemented with the appropriate antibiotic, such as chloramphenicol at 25 µg/ml.

Medium B consists of 10 ml Medium A, 0.1 ml of 50 mM $CaCl_2.2H_2O$, and 0.1 ml of 250 mM $MgCl_2.6H_2O$.

2.9.4 **Preparation of *Clostridium beijerinckii* Electro-competent Cells**

10 ml of 2xYTG medium were inoculated with a single colony of *C. beijerinckii*. This inoculated culture was serially diluted 10-fold to give 5 cultures, and grown overnight. The lowest dilution to grow was used to inoculate 300 ml of 2xYTG broth, and grown to an OD_{600} of 0.2. The culture

was then divided into centrifuge tubes and harvested by centrifugation at 4000 × *G* for 10 minutes, in a centrifuge pre-cooled to 4ºC. The supernatant was discarded under anaerobic conditions and each pellet gently washed in 40 ml ice-cold Sucrose Electroporation Buffer (SEB). SEB (Sucrose Electroporation Buffer) (pH 7.4) consists of sucrose 270 mM, Na_2HPO_4 0.5 g 500 ml^{-1}, NaH_2PO_4 0.42 g 500 ml^{-1}, $MgCl_2$ 1 mM, and sterile H_2O to 500 ml. SEB was filter-sterilised. The centrifugation was repeated and the supernatant discarded as before. Pellets were suspended in 20 ml ice-cold SEB each, and pairs of tubes combined. The centrifugation was repeated again, and each pellet resuspended in 2 ml SEB buffer supplemented with 10% v/v DMSO. Cells were gently mixed and dispensed in 280 µl aliquots into pre-chilled microcentrifuge tubes. Cells were either used immediately or stored at -80ºC.

2.9.5 Transformation of *Clostridium beijerinckii*

One aliquot of competent cells was used per electroporation. Cells were thawed on ice under anaerobic conditions for 5 minutes. Once cells were thawed, 20 µl DNA was added, and the mixture incubated on ice for a further 5 minutes. The cells and plasmid were then transferred to an ice-cold 0.2 cm cuvette, and electroporated under anaerobic conditions using a Bio-Rad Gene Pulser Xcell in the Exponential Decay mode, with the following parameters: 1.25 kV, 200 Ω, 25 µF. Cells were immediately recovered in 3 ml of 2xYTG for 3 hours and then harvested by centrifugation at 4000xG for 10 minutes. The resulting pellet was then resuspended in 1 ml 2xYTG and spread in 100 µl

aliquots on 2xYTG agar plates supplemented with appropriate antibiotic (thiamphenicol, 15-30 µg/ml).

2xYTG consists of the following: Tryptone (Sigma) 16 g L^{-1}, Yeast Extract 10 g L^{-1}, NaCl 5 g L^{-1}, Glucose 5 g L^{-1}, and sterile H_2O to 1 L. 2xYTG was then adjusted to pH 7.4 and filter-sterilised.

SEB (Sucrose Electroporation Buffer) (pH 7.4) consists of sucrose 270 mM, Na_2HPO_4 0.5 g 500 ml^{-1}, NaH_2PO_4 0.42 g 500 ml^{-1}, $MgCl_2$ 1 mM, and sterile H_2O to 500 ml. SEB was filter-sterilised.

2.9.6 **Transformation of *Clostridium pasteurianum***

Plasmid DNA for electroporation into *C. pasteurianum* was prepared from an *E. coli* HB101 derivative expressing the M.BepI methyltransferase (an isoschizomer of FnuDII) on a plasmid conferring kanamycin resistance. Plasmid DNA was obtained from A. Pech-Canul (SBRC Nottingham). Prepared plasmid was incubated with 580 µL of *C. pasteurianum* competent cells (obtained from A. Pech-Canul), mixed with 30 µL of ice-cold 96% ethanol and incubated on ice for 5 minutes. Cells were electroporated with exponential decay settings at 1.8 kV, 25 µF, and ∞Ω, generating a time constant of 10-18 ms. Cells were immediately rescued with 2xYTG (consisting of tryptone 16 g L^{-1}, yeast extract 10 g L^{-1}, and NaCl 5 g L^{-1}; adjusted to pH 6.2 using HCl) supplemented with 0.2 M sucrose (1 ml 2M sucrose per 10 ml 2xYTG). Cells were recovered for 4-6 hours at 37°C. Following recovery, cells were plated in aliquots of 100 µL on RCM (Reinforced Clostridial Medium; Oxoid; prepared according to the

manufacturer's instructions) agar plates supplemented with 15 μg/ml thiamphenicol. Colonies were observed after 1-2 days.

2.10 Plasmid Vector Construction

2.10.1 Generation of the Clostridial Shuttle Vector pMTL87151

The plasmid backbone of pMTL87151 was constructed by removing the Gram-positive replicon from pMTL83151 (obtained from the culture collection of the Synthetic Biology Research Centre, Nottingham) via AscI/FseI restriction digest. In tandem with this step, the minimal replicon of the *C. carboxidivorans* native plasmid P19 was amplified from *C. carboxidivorans* genomic DNA using primers P19_FWD and P19_REV (see above) to generate a fragment of 1,963 bp. This was digested with AscI/FseI and ligated to the pMTL8-151 fragment. The resulting ligated construct was transformed into 5-Alpha *E. coli* (NEB), transformant colonies selected on chloramphenicol (25 μg ml^{-1}). Putative transformant colonies were grown overnight under selective conditions and plasmid DNA purified via miniprep. Plasmid DNA was verified via sequencing.

2.10.2 Generation of pMTL87151 Variants pMTL87151A, B, and C

The construction of truncated P19 replicon variants proceeded as with the construction of pMTL87151, above, but with alternative primers employed as necessary to amplify the truncated replicons from *C. carboxidivorans* genomic

DNA. For pMTL87151A, the alternative forward primer P19_noProm_FWD was employed. For pMTL87151B and pMTL87151C, the alternative reverse primers P19_noRepeats_REV and P19_oneRepeat_REV were employed. Subsequent cloning steps were carried out as described for pMTL87151, above.

2.10.3 Generation of the Clostridial Shuttle Vector pMTL87251C

To generate the clostridial shuttle vector pMTL87251C, pMTL87151C was digested with FseI/PmeI to remove the *catP* selective marker. The *ermB* gene was prepared for ligation via digestion of pMTL84251 with FseI/PmeI. The pMTL backbone and *ermB* gene were ligated together and transformed into 5-alpha *E. coli*. Transformants were selected on erythromycin (500 µg ml^{-1}) and putative transformant colonies grown overnight under selective conditions. Plasmid DNA was prepared and cloning verified via sequencing.

2.10.4 Generation of *pyrE* Knockout Vector pMTL-JM101

To generate a vector for gene knockout of the orotate phosphoribosyltransferase gene (*pyrE*) in *E. limosum*, the vector pMTL-AMH101 (designed for *pyrE* knockout in *C. autoethanogenum*) was modified. Firstly, the Gram-positive replicon was removed by AscI/FseI digestion, and the resulting backbone fragment ligated to a fragment containing the Gram-positive replicon pCB102 (obtained via AscI/FseI digestion of pMTL83151). This intermediate plasmid was transformed into *E. coli* 5-alpha ® (NEB) (identical to *E. coli* DH5α) and putative transformants selected on LB supplemented with chloramphenicol.

Following *E. coli* colony outgrowth, putative mutants were grown overnight in liquid medium with selection as above. Plasmid DNA was prepared from overnight cultures via miniprep, and cloning verified via restriction digest.

The short homology arm (SHA) of the vector was designed to comprise an internal fragment of the *E. limosum pyrE*, lacking the first 50 bp and last 300 bp. The appropriate PCR fragment was generated with primers ELIM-SHAF and ELIM-SHAR, possessing SacII and NotI sites, respectively. Because the long homology arm (LHA) region contains two SacII sites, it was necessary to conduct this step prior to cloning of the LHA. These enzymes were used to linearise the intermediate vector and digest the SHA PCR product, and the two fragments were ligated together and transformed as above.

Following verification, a successful clone was digested further with NheI/AscI to remove the LHA of the existing vector. The resulting vector was ligated to a NheI/AscI-digested PCR product amplified from *E. limosum* genomic DNA using primers ELIM-LHAF-v2 and ELIM-LHAR (see Table 2.3, above). This comprises 1001 bp downstream of the *E. limosum pyrE*, except that a 215 bp stretch is omitted because of presumed toxicity issues during cloning in *E. coli*.

2.10.5 **Construction of CRISPR-Cas9 Knockout Vector pMTL431511::ORFD**

Homology arms were designed to generate amplicons consisting of 700-750 base pairs prior to the start codon of the target ORF (the “left homology arm”)

and following the stop codon of the ORF (the "right homology arm"), respectively. Homology arms were amplified using PCR in a "touchdown" protocol, as described above in section 2.8.9. Homology arms were joined together using splicing by overlap extension (SOEing) PCR, with a PCR program identical to that used to generate the arms, but with a longer extension time. Homology arms were digested with appropriate restriction enzymes, and ligated using T4 DNA Ligase (NEB) into a correspondingly-digested backbone (pSBRCas9, obtained from P. Rowe, SBRC Nottingham) which was treated with Antarctic Phosphatase (NEB) prior to ligation. Ligation was conducted in a thermocycler block using a program which cycled 60 times between 22ºC for 1 minute and 10ºC for 1 minute, resulting in a program of roughly 2 hours duration. In other cases, ligations were carried out at 4ºC overnight. Ligated products were transformed via heat-shock into *E. coli* 5-alpha competent cells (NEB), as per the manufacturer's instructions. Following recovery, transformant cells were plated onto LB agar supplemented with appropriate antibiotic (chloramphenicol 25 µg ml^{-1}).

In order to insert the sgRNA responsible for targeting the CRISPR vector into the vector, HiFi DNA assembly was used. DNA oligonucleotides encoding the desired sgRNAs were synthesised and resuspended in H_2O at 20 mM. These oligonucleotides consist of a 20 bp targeting region, flanked by 25 bp regions homologous to the CRISPR-Cas9 vector. Resuspended oligonucleotides were diluted in NEBuffer 2 (NEB) to a final concentration of 0.2 µM. A reaction was set up with HiFi DNA Assembly Master Mix (NEB), consisting of the following components: 5 µl of 0.2 µM oligonucleotides, 30 ng of linearised

and Antarctic Phosphatase-treated CRISPR-Cas9 vector, and sterile H_2O to a final volume of 10 µl. This reaction was incubated at 50^{o}C for 1 hour. 2 µl of this reaction mixture was then used to transform 5-alpha competent cells, as per the manufacturer's instructions. Transformant colonies were plated on LB agar supplemented with appropriate antibiotic selection (chloramphenicol 25 µg/ml). Following colony formation, colonies were grown in LB medium with antibiotic selection as above. Plasmid DNA was purified as above, and sequencing conducted to identify correctly-assembled vectors, using primers Cas_seq_Fwd and Cas_seq_Rev (see Table 2.3)

2.11 Reporter Assays

2.11.1 CatP Activity Assay in *E. limosum*

Cell pellets were collected by centrifugation and resuspended in BugBuster Protein Extraction Reagent with 1:99 Calbiochem Protease Inhibitor Cocktail VII (Merck). Resuspended cell pellets were sonically disrupted using a BioRupter Plus sonication device (Diagenode Inc, Liege, Belgium), during ten 30-second cycles of on/off sonication. Pellets were kept on ice prior to and after sonication, which occurred at 4^{o}C. Disrupted cells were then pelleted by centrifugation (17,000 x *G* for 5 minutes) and the supernatant reserved for use in the CatP assay. To conduct the assay, 10 µl of supernatant was added to 90 µl of Tris pH 7.8. To this diluted sample, 100 µl of CatP master mix was added, and the resulting colour change monitored spectrophotometrically

using an Infinite M1000 Pro microplate reader (Tecan, Mannedorf, Switzerland), which was also used to dispense the CatP master mix. This master mix consisted of 5, 5'-Dithio-bis (2-Nitrobenzoic Acid) (DTNB) (0.083 mM), chloramphenicol (0.005% w/v), and acetyl CoA sodium salt (0.19 mM). A_{415nm} measurements were taken at 10-second intervals for 190 seconds.

The maximum rate of ΔA_{415} over the period of measurement was determined and used to calculate CatP activity in each protein extract assayed. The following equation was employed, based on the CatP assay protocol available on the Sigma-Aldrich website[2]:

$$Units\ per\ mL\ enzyme = \frac{((\Delta 415\ \min Test) - (\Delta 415\ \min Blank))(0.2)(1)}{(0.0136)(0.01)}$$

0.2 refers to the volume of each reaction in ml. 0.01 is the volume in ml of protein sample used. The blank sample refers to a well containing only the CatP master mix, and no protein extract supernatant. Activities per ml extract were compared to protein content (in µg ml^{-1}) of each extract as determined by bicinchoninic acid (BCA) assay (see below), and an activity per mg protein calculated for each sample in duplicate or triplicate.

To determine total protein content of *E. limosum* protein extract supernatants, the Pierce BCA Protein Assay Kit (Thermo Fisher) was used. 25 µl of protein

[2] http://www.sigmaaldrich.com/technical-documents/protocols/biology/enzymatic-assay-of-chloramphenicol-acetyltransferase.html. Accessed 5/6/2017.

extract or BSA standard was incubated with 200 µl of Pierce BCA Working Reagent for 30 minutes. Measurements were taken at 562 nm using an Infinite M1000 Pro (Tecan). A standard curve for the BCA assay was constructed as per the manufacturer's instructions, and used to determine the protein concentration of protein extract samples. All standard and sample measurements were taken within 2 seconds of each other.

2.12 Plasmid Copy Number Assay

2.12.1 Plasmid Copy Number Determination via Quantitative PCR (qPCR)

Genomic DNA was purified from pelleted cells of *E. limosum* (or other bacterial species as noted), using a GenElute Bacterial Genomic DNA kit (Sigma-Aldrich, St. Louis, Missouri, USA). Cells were collected from cultures at the growth phase noted in the relevant figures. Quantitative PCRs were conducted in 20 µl volume, using a LightCycler 480 thermal cycler (Roche Life Sciences, Indianapolis, Indiana, USA) and the LuminoCt SYBR Green qPCR ReadyMix (Sigma-Aldrich), as per the manufacturer's instructions. To prepare genomic DNA for qPCR analysis, 10 µl per sample was digested in AscI (New England Biolabs, Ipswich, Massachusetts, USA), in a 20 µl reaction for two hours at 37^{o}C. Afterwards, the reactions were inactivated at 80^{o}C for 25 minutes. The digested genomic DNA was diluted 1:99 in de-ionised H_2O. 5 µl of diluted, digested DNA sample was used per qPCR reaction. Each reaction consisted of 5 µl DNA sample, 1 µl of forward primer,

1 µl of reverse primer (both from 0.5 µM stocks; final concentration 0.025 µM), 3 µl of H_2O, and 10 µl of LuminoCT ReadyMix. For each digestion reaction (biological replicate), three technical replicates were analysed via qPCR, unless otherwise noted. Primers were designed using the Benchling online bioinformatics suite (www.benchling.com), with target amplicon lengths of 100 bp.. Reaction programmes began with a 10 minute incubation step at 95ºC. This was followed by 50 cycles at 95ºC for 20 seconds (denaturation), 55ºC for 20 seconds (annealing), and 72ºC for 20 seconds (extension). Fluorescence data was acquired during the extension phase of each cycle. Once cycling was complete, melt curves were produced by incubation at 95ºC, with fluorescence data collected continuously during melting. This PCR programme is displayed below (Table 2.8).

Table 2.8: Typical qPCR programme.

Cycle Step	**Temp (ºC)**	**Duration**	**Program Cycle**	**Data Capture**
1. Initial Denaturation	95	10 min		
2. Denaturation	95	20 sec		
3. Annealing	55	20 sec		
4. Extension	72	20 sec	Return to step 2 (50x)	Single
5. Melting	95			Continuous
6. Hold	30			

The software provided with the LightCycler 480 was used to determine C_T (threshold cycle) values for reference and target genes in each genomic DNA sample. Copy number was calculated with the LightCycler480 software, using

the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). Mean copy number values were calculated using Microsoft Excel.

Gel electrophoresis of no-template controls (NTCs) revealed product formation corresponding to the expected product sizes, indicating reagent contamination (data not shown). However, examination of fluorescence results throughout the PCR program indicated that product formation in the NTCs was significantly delayed (>10 cycles) relative to the samples being assessed, indicating that they would not impact on copy number calculations. Melting temperature analysis of the NTCs indicated that formation of products similar in size to the intended target sequences was responsible for the delayed fluorescence observed during qPCR, rather than primer dimerisation (which would be expected to form a product with a distinctly lower melting temperature than that observed for 75-130 bp products in this assay).

2.13 Generation of *E. limosum* Insertional Mutants via Transposon Mutagenesis

Competent *E. limosum* cells were electroporated with 5 µl of either pMTL-CW21, pMTL-CW23, or pMTL-CW30 (see sections 5.2.1 and 5.2.2). Transformant colonies were selected on AWM supplemented with clarithromycin (6 µg/ml). Following 48-72 hours for colony outgrowth, all colonies on selective medium were collected and frozen at -80^{o}C in nonselective liquid medium supplemented with DMSO (10% v/v). To promote activity of the transposase and induce plasmid loss simultaneously, dilution

series of the total transformant collection mixture were plated on media supplemented with anhydrotetracycline (100 ng/ml).

In some cases, initial transformant colonies were selected on clarithromycin supplemented with 1 mM 5-fluorocytosine (5-FC). 5-FC was prepared as a 100 mM aqueous solution, and protected from light during solution preparation to avoid spontaneous deamination to toxic 5-FU. Clarithromycin + 5-FC plates were protected from light with aluminium foil.

2.13.1 Inverse PCR of *E. limosum* Transposon Mutants

Putative *E. limosum* transposon mutants were isolated by plating, and grown overnight in 1 ml AWM cultures supplemented with the appropriate antibiotic (thiamphenicol, 7.5 µg/ml). Cultures were divided 50:50 for storage at -80^{o}C (with 10% v/v DMSO) and for inverse PCR analysis. Genomic DNA was isolated from samples as described above, and samples digested with HindIII (NEB) for 2 hours. Digestion reactions were inactivated as per the manufacturer's instructions, and 2 µl of each digestion was used in a ligation reaction using T4 DNA ligase (Promega) at room temperature for one hour. 1 µl ligation reaction was used as a template in a subsequent PCR reaction using Q5 2X Master Mix (NEB). Primers binding to the *catP* gene of the transposon were employed; catP-INV-R1 (5' tattgtatagcttggtatcatctcatcatatatccccaattcacc 3') amd catP-INV-F1 (5' taaatcatttttagcagattatgaaagtgatacgcaacggtatgg 3'). PCR reactions were conducted with a "touchdown" programme, as described above (section 2.8.9, above). An extension time of 6 minutes was employed. PCR products were either purified on an agarose gel to assess product size, or

purified directly from the PCR. In both cases, the Wizard SV Gel and PCR Clean-Up System (Promega) was used to purify DNA for subsequent sequencing. Purified PCR products were sequenced via Sanger sequencing (Eurofins MWG Operon, Ebersberg, Germany) using primer catP-INV-R2 (5' ggcaagtgttcaagaagttattaagtcgggagtgcagtcgaagtgg 3').

2.14 Generation of *E. limosum pyrE* Deletion Mutant via ACE

E. limosum competent cells were transformed with the *pyrE* KO vector pMTL-JM101, as described above. Mutants were selected on PBBM minimal medium supplemented with thiamphenicol. Transformants were restreaked onto PBBM supplemented with 5-fluoroorotic acid (5-FOA) at a concentration of 2 mg/ml. FOA^R colonies were restreaked on the above selection medium again, and analysed via colony PCR across the *pyrE* locus. Clones displaying bands characteristic of *pyrE* truncation were verified by sequencing.

2.15 Generation of *E. limosum* Deletion Mutants via CRISPR-Cas9

To generate in-frame deletion mutants in selected target genes, competent cells of *E. limosum* were electroporated with a CRISPR-Cas9 knockout vector from the pSBRCas9 series (construction of which is described in section 2.10.5). Colonies were selected on thiamphenicol. The total recovery volume was plated in 100 µl aliquots. <50 transformants per electroporation were typically obtained.

Following colony outgrowth (48-72 hours), cells were picked into liquid selective medium (thiamphenicol), and colony PCR was conducted following

overnight growth. In the case of putative *R.Eli8486ORFD* knockout mutants, the primers ORFD-EXT-F/-R (binding outside of the recombination locus) and ORFD-RHAR/-LHAF (binding at the ends of the recombination locus) were employed in various combinations. The presence of knockout clones and the purity of putative knockout mutants was assessed by observing the presence of wild-type and truncated bands (with the latter indicating knockout). Where necessary, mixed colonies containing wild-type and knockout clones were purified by repeated restreaking on selective medium, in order to maintain the Cas9 vector and encourage counterselection against wild-type clones.

CHAPTER THREE

Electroporation Optimisation and Initial Tools Development for *E. limosum* ATCC 8486

3 Culture of *E. limosum* ATCC 8486 and notes on Electroporation Optimisation

3.1 Introduction

During this study, several optimisation steps were undertaken in order to improve the efficiency of DNA transfer to *E. limosum*. The resistance of *E. limosum* to antibiotic selection was assessed. The electroporation parameters used to transform *E. limosum* were altered and the results assessed in terms of transformant colonies obtained per µg plasmid. The growth behaviour of *E. limosum* while maintaining standard shuttle vectors was noted. Most transformations in this chapter were undertaken with the shuttle vector pMTL87151, which possesses a Gram-positive replicon that was discovered and exploited during the course of the work described here (see Chapter 4).

Following the establishment of gene transfer, and the characterisation of Gram-positive plasmid replicons, the creation of a *pyrE* mutant is a foundational step in the genetic modification of clostridia (Minton et al., 2016).

The *pyrE* gene encodes orotate phosphoribosyltransferase, which is involved in pyrimidine biosynthesis, converting orotic acid into orotidine 5'-monophosphate (OMP). PyrE also converts 5-FOA (an analogue of orotic acid) into 5-fluoroorotidine monophosphate (5-FOMP), a precursor to the toxic 5-fluorouridine monophosphate (5-FUMP). Accordingly, resistance to 5-

FOA can be used to select in favour of a *pyrE* mutant. The use of 5-FOA as a selective agent originated in the study of yeast (Boeke et al., 1987).

Because of the role of *pyrE* in pyrimidine biosynthesis, disruption of *pyrE* confers uracil auxotrophy. The *pyrE* mutant can therefore be selected against by growth on minimal media lacking uracil. Wild-type organisms have the reverse phenotype, being uracil prototrophs that are sensitive to 5-FOA.

The dual and opposite phenotypes conferred by functional and non-functional *pyrE* have been exploited in clostridial genetics through the development of the ACE method (Heap et al., 2012) (Figure 3.1). This method involves the use of integrative knockout plasmids with replication-deficient replicons. Replication deficiency of the knockout vector means that single-integrant clones will grow more quickly than those with an un-integrated plasmid. Asymmetrical homology arms direct the order in which integration events are likely to occur, linking the second recombination event to a selectable phenotype (FOA^R, Ura-).

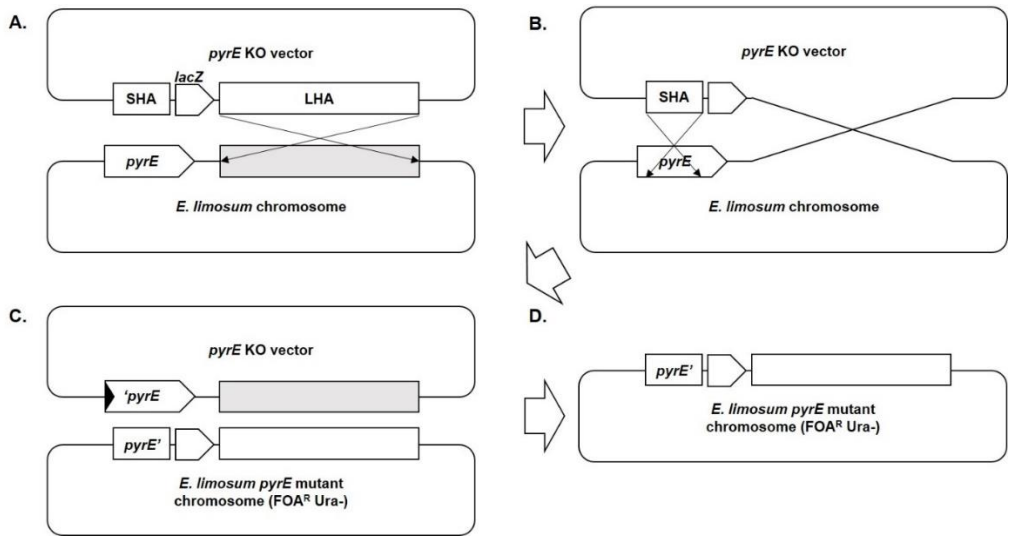


Figure 3.1: Illustration of *pyrE* knockout via Allele-Coupled Exchange (ACE). Arrows indicate crossover during homologous recombination. **A**: The order of recombination events is dictated by the asymmetry of the homology arms; the LHA (Long Homology Arm) will usually integrate first. **B**: Integration of the knockout vector results in faster colony outgrowth, as the single-integrant is no longer limited by the speed of plasmid replication. **C**: A second recombination event results in truncation of the chromosomal *pyrE*. Double-crossover mutants are therefore coupled to a selectable phenotype. **D**: Mutants are restreaked under 5-fluoroorotic acid (FOA) selection to discourage additional recombination events while promoting loss of the plasmid.

In this study, as in other examples of *pyrE* knockout via ACE, a knockout vector was constructed with asymmetrical homology arms. The long homology arm (LHA) is likely to recombine first due to its greater length. Single-crossover mutants can then be selected by their larger colony size relative to transformants with un-incorporated plasmid. For this reason, the Gram-positive replicon selected for the knockout vector should be as replication-deficient as possible (causing the greatest delay to cell growth). Two plasmid replicons have been demonstrated to be functional in *E. limosum*. These are P19 (developed in this study) and pCB102. The latter is

the more deficient in replication (see Chapter 4), and was therefore selected for the knockout vector.

3.2 Results

3.2.1 Antibiotic Tolerance of *Eubacterium limosum*

To determine inhibitory concentrations of antibiotic selective agents commonly used in clostridia, a growth experiment was conducted in which *E. limosum* was subcultured into various concentrations of thiamphenicol or clarithromycin. Both of these antibiotics are used as selective markers for plasmids of the pMTL80000 range (Heap et al., 2009), with resistance encoded by the *catP* and *ermB* genes, respectively. Both antibiotics were found to be inhibitory at all concentrations tested after 24 hours (see Figure 3.2). Concentrations of 7.5 and 6 µg/ml for thiamphenicol and clarithromycin respectively were employed throughout the rest of this study.

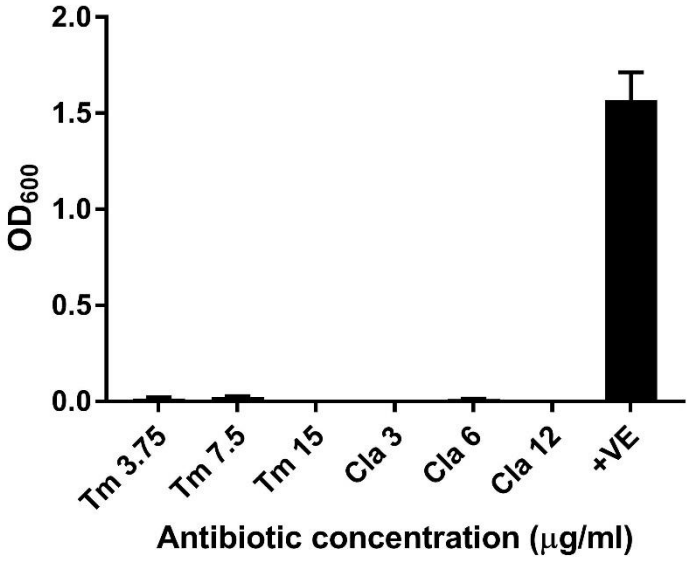


Figure 3.2: Antibiotic resistance of *E. limosum*. Growth of *E. limosum* wild type after 24 hours in liquid medium supplemented with varying concentrations of thiamphenicol (Tm), clarithromycin (Cla), or without antibiotic selection (+VE). Measurements are mean average OD_{600} readings from technical triplicates. Error bars represent the standard deviation.

3.2.2 Electroporation Optimisation

Many techniques of interest in synthetic biology are dependent upon (or benefit from) a transformation efficiency that is as high as possible. While factors such as endogenous CRISPR systems and restriction-modification (R-M) systems have a significant impact on transformability of many microbial species, gains in electroporation efficiency can also be derived from optimisation of electroporation conditions. Therefore, prior to attempts to knock out endogenous restriction-modification systems in *E. limosum*, attempts were made to optimise the existing electroporation protocol. Following the establishment of an electroporation protocol for *E. limosum*

(Rob Mansfield, personal communication), efforts were undertaken in this study to improve transformation efficiency for the type strain. Parameters such as recovery time, electroporation voltage, and optimum amount of input DNA were explored. Throughout this study, all electroporations were undertaken under anaerobic conditions using a Bio-Rad Gene Pulser Xcell system, operating in Time Constant Mode to reduce arcing at high voltages.

The effect of recovery times on transformation efficiency was assessed. It was found that transformants can be recovered after 4 hours of recovery, in quantities within an order of magnitude of those for 6 and 8 hour recoveries (Figure 3.3). Increasing recovery of CFU following prolonged incubation may be attributable to replication of the initial transformants, which may then obscure the actual initial transformation rate. Accordingly, 4 hour recoveries were settled on as a convenient recovery period for transformations undertaken during the remainder of this study.

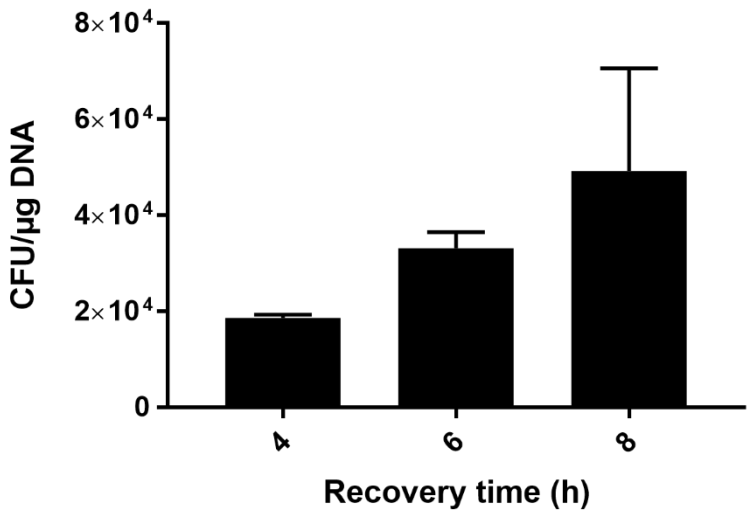


Figure 3.3: Effect of recovery time on transformant recovery in *E. limosum*. Bars show mean averages of three transformations. Error bars represent standard deviation. All transformations were conducted with identical aliquots of pMTL87151.

The voltage of the initial transformation protocol was set at 1.5 kilovolts (kV), and voltages below this were explored (Rob Mansfield, personal communication). In this study, electroporations were conducted at 1.0, 1.5, 1.8 and 2.0 kV (see Figure 3.4). Electroporation was attempted at 2.5 kV, but arcing was encountered in both Time Constant and Exponential Decay modes, and greatly reduced numbers of viable transformant colonies were obtained.

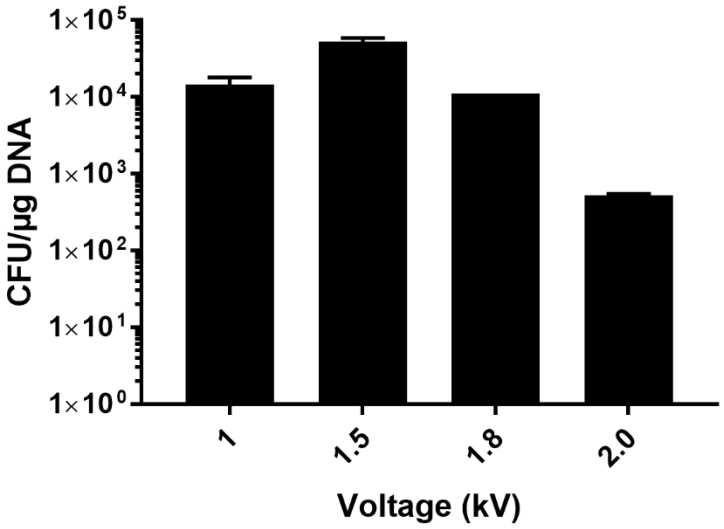


Figure 3.4: Transformation efficiency of *E. limosum* at varying voltages. Bars show mean transformation efficiency of three transformations. Error bars represent standard deviation. All transformations were conducted with identical aliquots of pMTL87151, with roughly 800 ng per aliquot, as determined spectrophotometrically.

A peak in transformation efficiency was observed at 1.5 kV, with decreasing efficiency observed at lower and higher voltages, and a significant decrease in transformation efficiency when electroporating at 2.0 kV. Electroporations were conducted at 1.5 kV for the remainder of this study.

To determine the extent to which DNA transfer efficiency (CFU/µg DNA) varied with the amount of DNA per transformation, transformation efficiency was assessed for varying amounts of input DNA (see Figure 3.5).

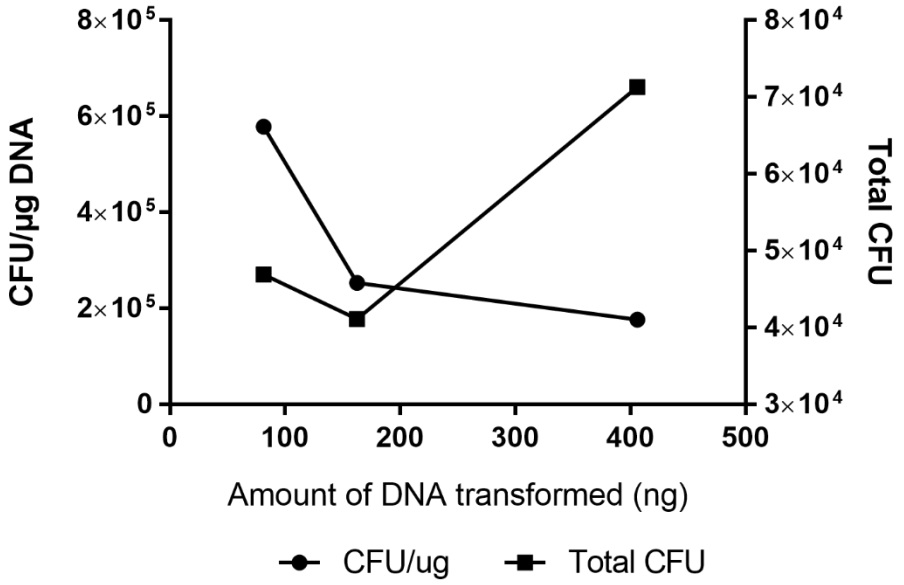


Figure 3.5: Effect of input DNA amounts on transformation of *E. limosum*. Each data point represents a single transformation with pMTL87151. All transformations were conducted at 1.5 kV, with a 4 hour recovery time.

This experiment reveals a correlation between input DNA amount and apparent transformation efficiency. Across the range of plasmid concentrations investigated, a roughly 3-fold decrease in efficiency was observed with increasing plasmid concentration. Note that electroporations prior to this display an efficiency of 10^4 transformants per ug DNA, due to generally being conducted with 0.5-1.0 ug of plasmid, whereas efficiences of 10^5 cells per ug DNA were achieved when less plasmid was used. However, apparent transformation efficiencies appear to be strongly dependent on the size of the plasmid being transformed, with vectors of over 8 Kb transferring with much lower efficiencies (see Chapter 5).

3.2.3 Replicon Performance in *E. limosum*

Over the course of this study, it was apparent that transformant colonies appeared with different timings, which were dependent on the Gram-positive replicon of the transformed plasmid. Plasmids based on P19 were generally found to appear overnight, whereas those based on pCB102 required a minimum of 48 hours to appear (see Table 3.1). Electroporation of plasmids carrying other Gram-positive replicons (pBP1, pCD6, and pIM13) did not result in transformant colony formation (data not shown).

Table 3.1: *E. limosum* colony outgrowth under selection.

Replicon	Time to colony formation (h)
P19	18-48
pCB102	48-72

Note that colony outgrowth times were noticeably different depending on the plasmid replicon employed.

3.2.4 Extracellular Polysaccharides

During this study, *E. limosum* ATCC 8486 was observed to produce mucoidal colonies, complicating restreaking and resuspension of colony material. At higher optical densities, the efficiency of centrifugation was inhibited somewhat. For the sake of expedience, during the preparation of electrocompetent cells harvesting was therefore performed at a low optical density (0.28-0.30) and from a large culture volume (see Materials and Methods, section 2.9.1).

3.2.5 **5-FOA Toxicity in *E. limosum***

To determine the concentration of 5-FOA suitable for selection of *pyrE* mutants, a range of 5-FOA concentrations were assessed in a growth assay. All concentrations tested were found to significantly inhibit or prevent growth (Figure 3.6).

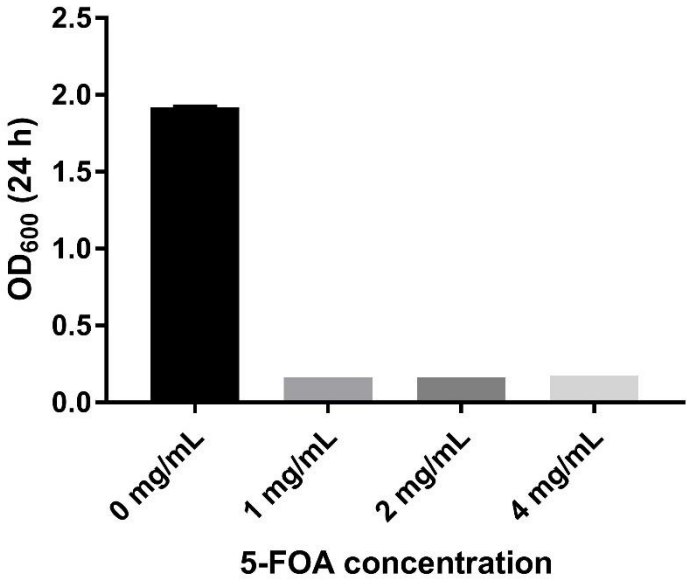


Figure 3.6: 5-FOA toxicity assay. OD_{600} measurements were taken 24 hours after inoculation, and represent the mean average of biological triplicates for 0 mg/ml, and single cultures for higher concentrations. Error bars represent the standard deviation.

3.2.1 **Construction of *pyrE* knockout vector**

The construction of the *pyrE* knockout vector employed in this study is described fully in Materials and Methods. The *pyrE* knockout vector was initially designed with a full-length (1200 bp) long homology arm (Figure 3.7). It was not possible to construct this vector in *E. coli*, possibly due to

expression of a toxic gene within the long homology arm. However, a truncated long homology arm was successfully cloned into the vector backbone, completing the knockout vector.

As a consequence of the second homologous recombination event, the *lacZ* fragment and multiple cloning site of the knockout vector is inserted into the chromosome, downstream of the truncated *pyrE*.

In common with other *pyrE* ACE knockout vectors, pMTL-JM101 possesses a heterologous *pyrE* (in this case, from *C. acetobutylicum*) for counterselection purposes.

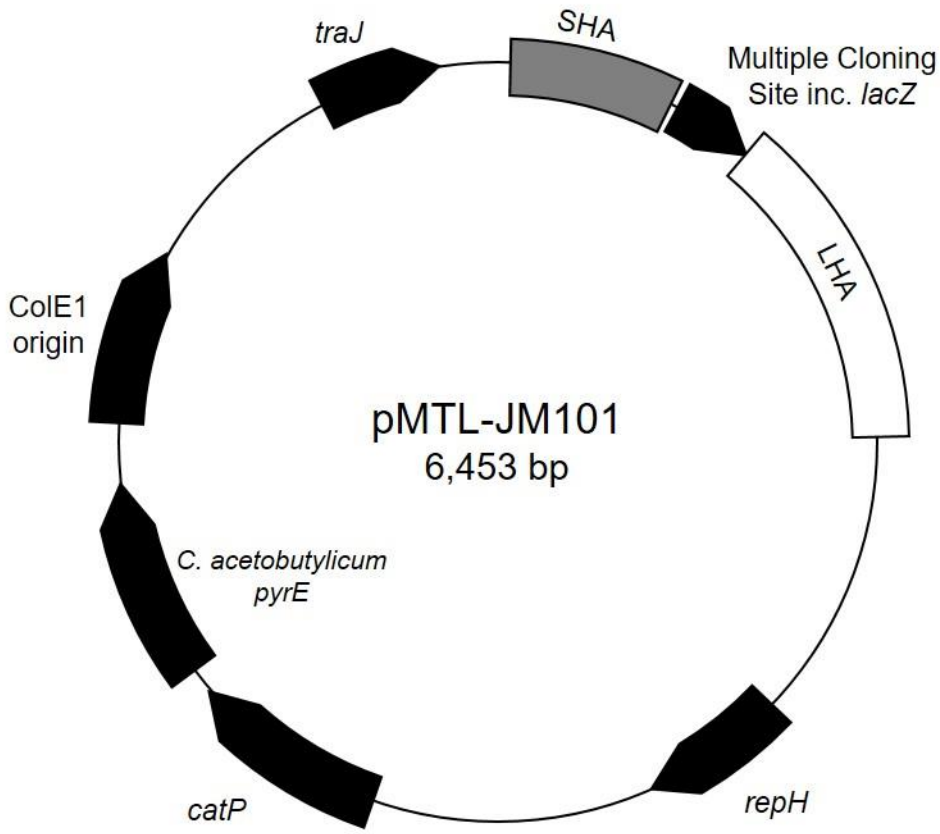


Figure 3.7: The ACE *pyrE* knockout vector employed in this study. SHA: Short Homology Arm, consisting of an *E. limosum pyrE* internal fragment lacking the first 50 bp and last 300 bp of the gene. LHA: Long Homology Arm, consisting of 1001 bp downstream of *pyrE*, excluding a 215 bp sequence that could not be cloned in *E. coli*.

3.2.2 **Knockout of *pyrE* in *E. limosum***

Following verification of successful cloning, the knockout vector was electroporated into *E. limosum*, and mutants selected for on agar supplemented with thiamphenicol (Tm). Following colony outgrowth, putative mutants were restreaked to 5-FOA, to select in favour of double-crossover mutants in which no functional *pyrE* was present.

Following outgrowth of 5-FOA^R colonies, these were restreaked several times on non-selective media to promote loss of the knockout vector and minimise the risk of phenotype reversion. After restreaking, 5-FOA^R putative mutants were restreaked to Tm selective media to confirm plasmid loss. Purified 5-FOA^R, Tm^S clones were analysed via colony PCR with primers ELIM-SHA-F and ELIM-LHA-R, binding at each end of the recombination locus (Figure 3.8).

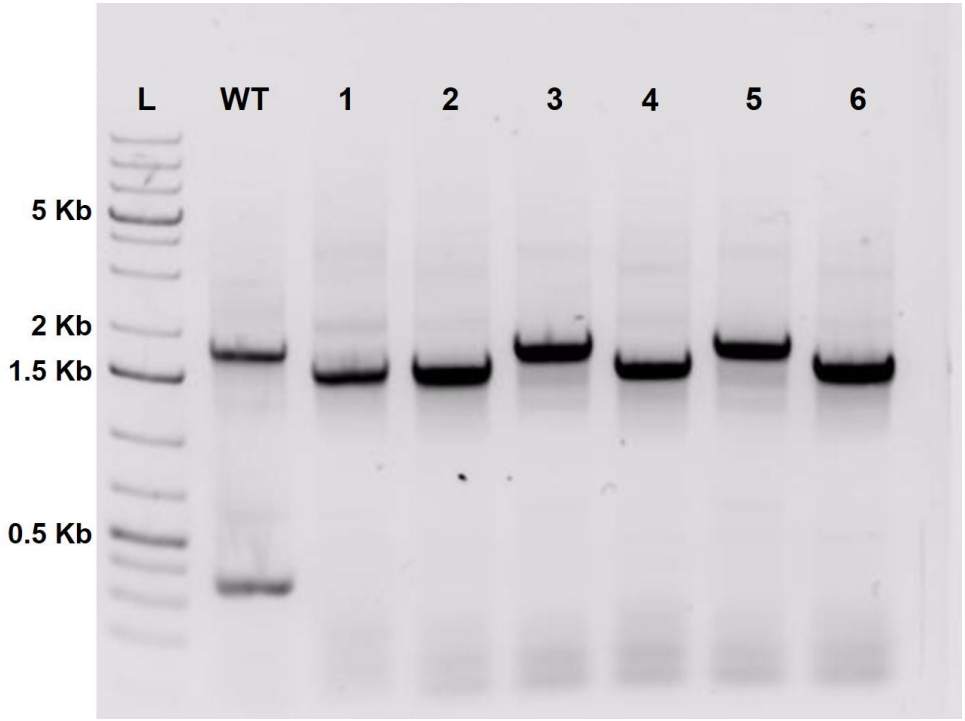


Figure 3.8: Colony PCR verification of *pyrE* deletion in putative mutants. Lane "WT" uses wild-type genomic DNA as a template, and shows a wild-type band of 1816 bp. Lanes 3 and 5 likewise show wild-type band sizes. Lanes 1, 2, 4 and 6 show truncated bands of 1599 bp, indicative of *pyrE* knockout. Bands indicating knockout were excised and verified by sequencing.

Colony PCR indicated that 4/6 clones analysed were *pyrE* knockouts, with the remaining two clones appearing to be wild-type. Bands were excised for these clones and the truncation of *pyrE* confirmed via sequencing.

3.3 Discussion

In this chapter, the optimal voltage for electroporation of *E. limosum* ATCC 8486 (using electro-competent cells as described in Materials and Methods) was determined. Additionally, a variety of recovery times were tested to establish whether lengthy recoveries were necessary. It was found that four-hour recovery times were sufficient. Greater recovery times result in higher apparent transformation efficiencies, but this may be an artefact of transformant replication within the recovery media.

To establish reverse genetics in *E. limosum*, an Allele-Coupled Exchange (ACE) strategy has been employed to generate a *pyrE* deletion mutant in *E. limosum*. This mutant strain may form the basis of subsequent genetic modification, with the *pyrE* locus corrected to wild-type in conjunction with the insertion of a gene of interest, with selection for both steps being linked to reversion to the wild-type (FOA^S, Ura^+) phenotype.

This study made use of a 5-FOA concentration of 2 mg/ml, which was found to prevent the growth of wild-type *E. limosum*. Differing concentrations of 5-FOA have been employed in previous *pyrE* knockouts in clostridia, with a range of 0.8-3.5 mg/ml being described (Minton et al., 2016). The ‘gold standard’ isolation of *pyrE* mutants via ACE would involve the initial

knockout of *pyrE* via ClosTron and subsequent experimental determination of the exact 5-FOA concentration required to separate wild-type from *pyrE* mutants. This approach was not taken in this study, and 2 mg/ml 5-FOA was only partially successful in isolating *pyrE* mutants; 2/6 FOA^R colonies screened were found to be wild-type at the *pyrE* locus (Figure 3.8).

The *pyrE* knockout mutant generated during this study has subsequently been used to insert the orthogonal transcription factor *tcdR* at the *pyrE* locus concomitant with repair of *pyrE* (Rob Mansfield, personal communication).

3.4 Chapter Summary

- 1.5 kV was found to be the optimum voltage for electroporation of *E. limosum* ATCC 8486, although good transformation frequencies can be obtained at 1.0-1.8 kV.
- It is likely that sibling effects start to dominate transformant population when recovery times are extended to up to 8 hours. For applications where the generation of unique transformants is a priority, recovery times should be kept to a minimum.
- When preparing competent cells, a low final OD_{600} was selected (0.28-0.30) to minimise the impact of extracellular polysaccharides produced by *E. limosum* during the late-exponential and stationary growth phases.
- Plasmids based on the P19 Gram-positive replicon allow much faster growth of *E. limosum* under selective conditions than those based on pCB102.

- Gene editing via Allele-Coupled Exchange (ACE) was exemplified in *E. limosum*, generating a mutant with a truncated *pyrE* gene, resistant to 5-fluoroorotic acid (5-FOA).
- This provides the basis for the addition of genes using homologous recombination vectors designed to complement the deficient uracil metabolism of the *pyrE* mutant.

CHAPTER FOUR

Generation of a Novel Gram-positive Plasmid Replicon from *C. carboxidivorans*

4 Generation of a Novel Gram-positive Plasmid Replicon from *C. carboxidivorans*

4.1 Introduction

This study began as an attempt to establish synthetic biology techniques in *Clostridium carboxidivorans* P7, an acetogen of considerable industrial interest due to its range of reduced fermentation products, including acetate, butyrate, and butanol, and formerly the process organism of the US-based company Coskata. However, *C. carboxidivorans* proved recalcitrant in the face of various attempts at transformation. This was attributed to its possession of a large number and variety of restriction-modification systems (of Types I, II, III and IV), capable of restricting incoming DNA if un-methylated, over-methylated or incorrectly methylated. In the course of these attempts, in the search for molecular tools of use in *C. carboxidivorans* and other clostridia, the replicon of *C. carboxidivorans*' native plasmid was isolated and modularised. The properties of this plasmid replicon are described in the following study. General features of Gram-positive plasmid replicons will also be discussed, to contextualise the features of the P19 replicon.

4.1.1 General Features of Plasmid Replicons

Plasmids are commonly defined as extra-chromosomal genetic elements that are capable of self-controlled replication within their host (Del Solar et al.,

1998). Plasmid origins of replication ("replicons") can be defined either as the minimal area required to allow replication of the plasmid, or more narrowly as the specific binding sites for host- or plasmid-encoded proteins or RNAs involved in replication.

Plasmid replication mechanisms are generally categorised in three groups, based on key structural features: Rolling-circle (or sigma), strand displacement, and theta-type (reviewed in Del Solar et al., 1998). The latter two classes are grouped together in some reviews (Dmowski and Jagura-Burdzy, 2013). Plasmids can also be grouped into families via sequence homology (Carattoli et al., 2014).

General features of theta origins of replication include plasmid-encoded replication (Rep) proteins and repeat sequences known as iterons. Iterons are essential sites at which the Rep protein binds and which are responsible for replication and maintenance of the plasmid. Iterons are not specific to theta plasmids, being also found in strand-displacement and rolling-circle plasmids (Del Solar et al., 1998). Iterons are generally arranged as tandem sets of repeat sequences. The spaces between these sets have frequently been noted to be a multiple of 11 base pairs, corresponding to the helical periodicity of DNA. Not all iterons in a given origin of replication are essential. For example, studies on the R6K Gram-negative replicon in *E. coli* observed that a single replicon (of seven) could be removed without effect, whereas replication was abolished if three or more were removed (Kolter and Helinski, 1978). While iterons are considered to be characteristic of theta-replicating plasmids, some theta

origins of replication do not possess them, such as the well-studied ColE1 replicon in *E. coli* (reviewed in Del Solar et al., 1998).

Interactions between iterons and Rep initiator proteins are believed to be one of the mechanisms responsible for the control of copy number in plasmids that possess both of these features (Nordström, 1990). Rep proteins have been observed to auto-regulate their expression, and are known to interact not only with their cognate iterons, but also with other Rep-iteron complexes they encounter. These interactions, in turn, are thought to lead to steric interference that inhibits plasmid replication until cellular division dilutes the total number of plasmids per cell. Mutations in the Rep protein have been observed to affect plasmid copy number. In one example, mutations in the Rep protein of the *E. coli* R6K replicon resulted in an increase in copy number (Filutowicz et al., 1994). Apart from their role in plasmid replication, iterons are also implicated in the segregation of plasmids between bacterial daughter cells (see below).

The stability of native plasmids can be ensured by a variety of means. Low-copy plasmids generally possess some form of active partitioning system, ensuring that each daughter cell receives at least one plasmid. Alternatively, post-segregational killing systems can destroy cells that do not inherit a plasmid. Additionally, conjugative machinery found on some plasmids can allow them to re-establish themselves within cells which had previously lost them (reviewed in Million-Weaver and Camps, 2014).

Bacterial plasmids can vary widely in size. Some large plasmids (‘megaplasmids’, second chromosomes or ‘chromids’) (Harrison et al., 2010)

harbour essential genes and are maintained accordingly. Others do not appear to possess essential genes, and are only maintained as a result of plasmid-borne stability factors (reviewed in Million-Weaver and Camps, 2014).

4.1.2 *C. carboxidivorans* and Plasmid P19

C. carboxidivorans is a Gram-positive anaerobe. The type strain, *C. carboxidivorans* P7, was isolated from an agricultural settling lagoon (Liou et al., 2005). It is notable for its acetogenic metabolism and ability to produce butanol during gas fermentation, which is a rare capability among the acetogens.

During analysis of the first published genome of *C. carboxidivorans* $P7^T$, it was noted to contain a single 19 Kb cryptic plasmid (Bruant et al., 2010). The plasmid was noted to contain no genes essential for butanol production, unlike the megaplasmid observed in *C. acetobutylicum*.

4.2 Results

4.2.1 Sequence Analysis of the P19 Cryptic Plasmid

The P19 plasmid of *C. carboxividorans* P7 is a 19,781 base pair covalently-closed circular plasmid (see Figure 4.1). Sequence analysis of the plasmid reveals that it possesses numerous phage-related genes. These include a phage head fiber protein, a phage tail length tape measure protein, two phage

integrases, and a hybrid phage integrase/recombinase. The phage tail tape measure protein has similarities with hypothetical proteins in *Clostridium magnum* (60% identity over 99% coverage, WP_066630751.1), and homology suggests a membership of the *yqbO* phage minor tail proteins.

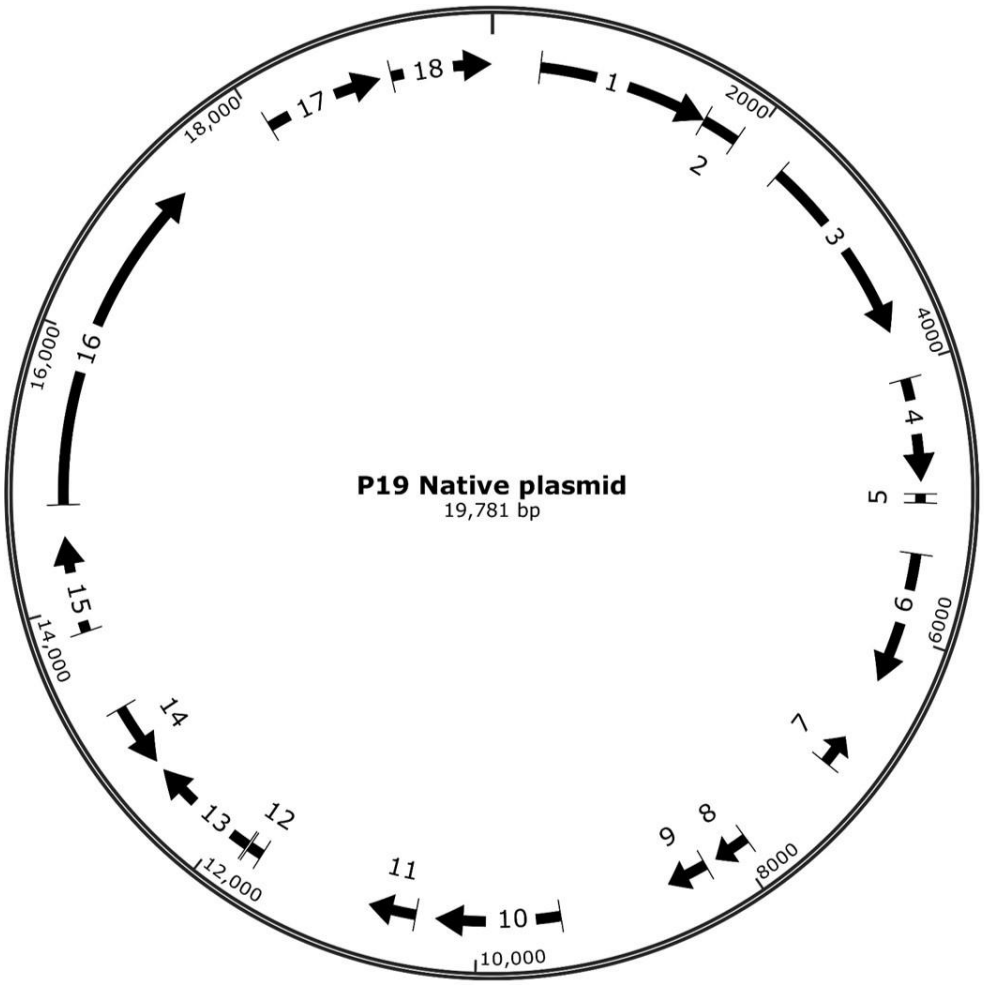


Figure 4.1: Map of *C. carboxidivorans* cryptic plasmid P19. Plasmid features are: 1. *rep*; 2. Iterons; 3. DUF2971; 4. *parA* chromosomal segregation protein; 5. Iterons II; 6. SMC chromosomal segregation protein; 7. Phage CI repressor; 8. Glycyl-tRNA synthetase beta subunit; 9. DUF3784; 10. Phage integrase-recombinase; 11. Phage plasmid stability protein; 12. Iterons III; 13. Phage integrase; 14. Phage head fibre protein; 15. SMC chromosomal segregation protein; 16. Phage tail protein; 17. Phage integrase; 18. Phage anti-repressor protein.

P19 does not possess any genes related to solvent production. Further, it does not possess any genes that appear to confer a selective advantage on *C. carboxidivorans*, such as antibiotic resistance or heavy metal tolerance.

The P19 backbone possesses a *parA* chromosomal segregation protein and a pair of genes encoding SMC chromosomal segregation proteins. Genes of the *par* family are involved in the active partitioning of low-copy plasmids (Gerdes et al., 1985) (Million-Weaver and Camps, 2014). They are subdivided based on their partitioning mechanism; in *par* I systems, polymerised motor proteins pull plasmids towards the poles of a dividing cell, while *par* II systems push plasmids apart instead. In both systems, a plasmid-encoded adaptor protein binds to a specific plasmid sequence (the centromere) and then binds to the polymerised motor protein. While *parA* was detected on the P19 backbone via sequence analysis, other *par* genes such as *parB* (responsible for killing of daughter cells which do not receive the plasmid) were not observed. Of note, the centromere in *par* systems consists of tandem direct or inverted repeats, to which the *par* adaptor protein binds. Two such sets of tandem repeats have been detected on P19, and are described below.

The Gram-positive replicon of the cryptic P19 plasmid contains a 1272 bp ORF encoding a Rep gene. This putative gene has 80% homology to an annotated DNA-binding response regulator in *C. botulinum* (NCBI reference sequence WP_003363180.1), and a 73% identity to an annotated replication protein in *Clostridium coskatii* (WP_063602457.1). It also has 58% and 57% identity with a hypothetical protein in *Paraclostridium bifermentans*

(WP_021430645.1) and an annotated replication protein in *Paeniclostridium sordeliii* (WP_057554672.1).

Downstream of the *rep* gene are two sets of repeat sequences in tandem. These are likely to be the iterons to which the Rep protein binds to initiate DNA replication. Each consists of a set of one 35- 32- and 18-mer sequence (i.e. the entire set is arranged in the pattern ABC, ABC). The two regions are separated by 89 base pairs. This is significant because, as noted above, such repeat regions are often spaced at intervals related to the 11-base pair helical periodicity of DNA. These putative iterons were included in the modularisation of the P19 replicon for use in the pMTL80000 shuttle vector system (see below).

Interestingly, there is another set of repeat sequences on the P19 plasmid, consisting of three repeats of the sequence TGATACTAATA. These repeat sequences are positioned immediately downstream of the *parA* chromosomal segregation protein. As noted above, *par* segregation systems require a centromere (consisting of tandem repeat sequences) for the *parA* protein to bind (reviewed in Million-Weaver and Camps, 2014). This set of repeats likely constitutes the centromere of the wild-type P19 plasmid. They may also act as an alternative site for DNA melting and initiation of replication. Many plasmids have been noted to possess multiple sets of iterons with various functions (reviewed in Del Solar et al., 1998). The P19 replicon region (consisting of the *rep* gene and its associated iterons, and including the

putative second iteron site just described) is enclosed by phage-related repressor and anti-repressor proteins.

Judging by the preponderance of phage-related proteins on the P19 backbone, it may be a prophage existing in circular form within *C. carboxidivorans*. No attempts were made in this study to induce *C. carboxidivorans* with mitomycin C to check whether cell lysis was induced.

With the identification of the putative replicon of the P19 plasmid, efforts were made to determine the minimal functional replicon, in order to utilise it in a modular shuttle vector for use in clostridia. A representation of the P19 replicon can be seen in Figure 4.2. The plasmid replicon defines the host range of the plasmid (Bardaji et al., 2017), and the replicon of a cryptic plasmid that was evidently functional in *C. carboxidivorans* seemed likely to be functional in other clostridia of industrial interest. As noted above, the replicon of plasmid P19 appears to display several discrete functional units: A gene encoding a replication initiation (Rep) protein, a promoter region upstream of that gene, and a set of repeat sequences (iterons) to which the Rep protein binds. A putative second set of iterons downstream of the first set may also be involved in plasmid replication and maintenance.

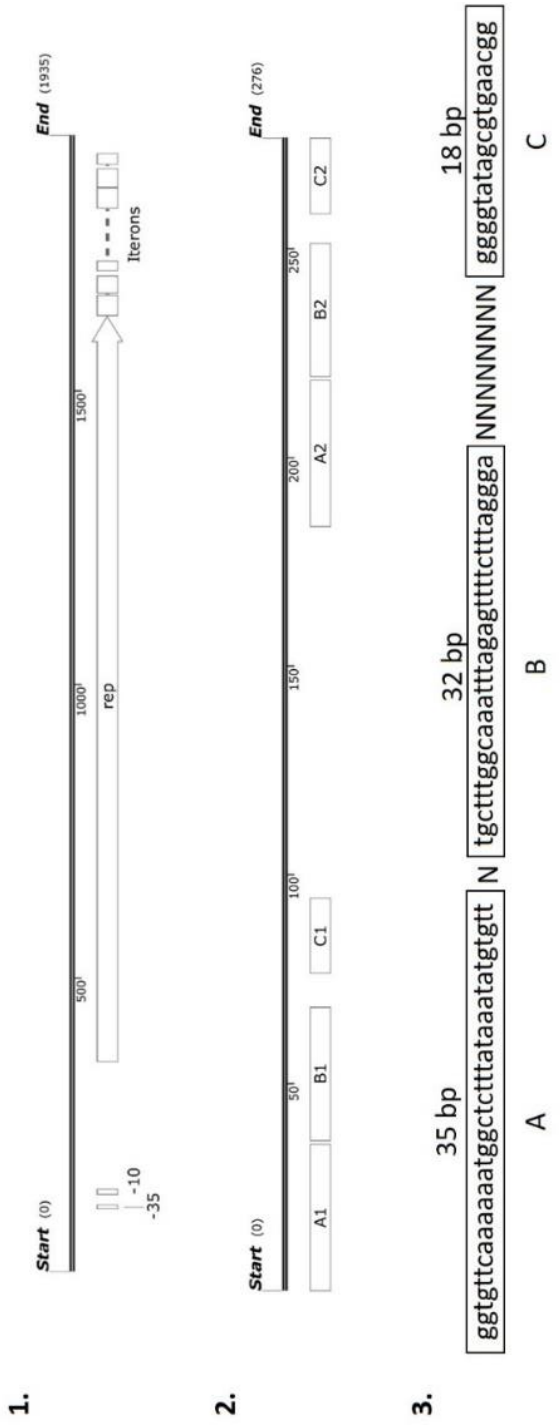


Figure 4.2: Structure of the P19 minimal replicon. **1.** Modular replicon. **-35, -10**: Putative promoter sequence for P19 rep gene. Rep: Replication initiator protein gene, rep. Iterons: Repeat sequences to which Rep protein binds to initiate plasmid replication. **2**. Detailed map of iteron structure. Direct repeats are numbered, e.g. A1, A2. **3**. Nucleotide sequence of iterons. Only one repeat is shown.

4.2.2 **Creation of the Modular P19 Replicon**

To test the functionality of the P19 replicon, it was cloned into the pMTL80000 clostridial shuttle vector system. The pMTL80000 series are modular plasmids with components separated by a standardised set of restriction sites (Heap et al., 2009). An overview of the pMTL80000 configuration can be seen in Figure 4.3.

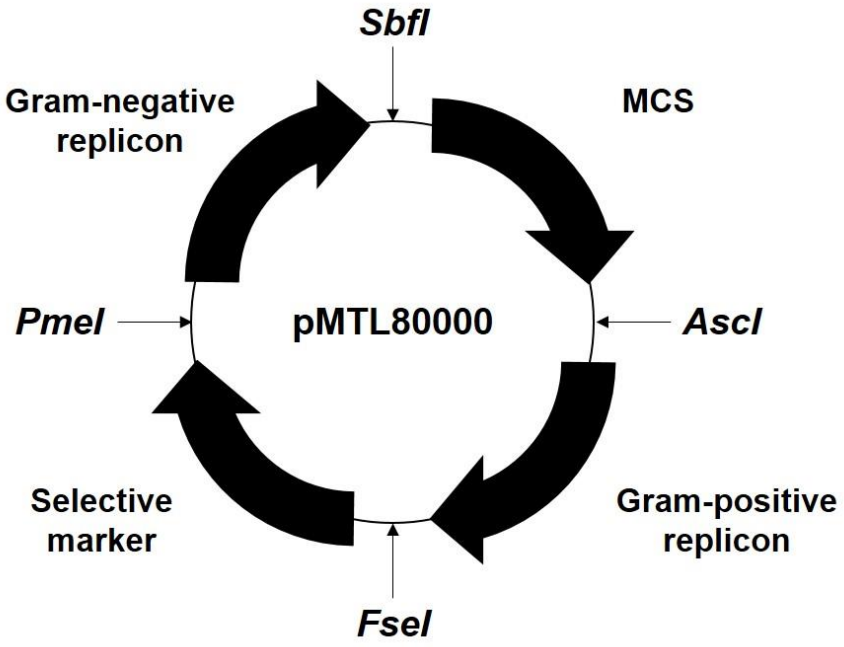


Figure 4.3: The pMTL80000 modular shuttle vector system.

The pMTL80000 modular shuttle vector system consists of a set of four plasmid components joined by unique restriction sites in a specific order (Gram-positive replicon, selective marker, Gram-negative replicon with optional combined conjugal function, and application-specific module, with

the latter typically containing a Multiple Cloning Site (MCS)) (Heap et al., 2009). It was created to facilitate and standardise the genetic modification of clostridia, and forms a framework in which to test the qualities of the modularised P19 replicon. The system can be extended simply by modularising the desired biological component using the appropriate restriction sites, as was done in this study. In the naming scheme of the pMTL80000 vector system, the second numeral denotes the Gram-positive replicon. For example, pMTL83151 contains the replicon pCB102 (the third Gram-positive replicon in the system).

When modularising the P19 plasmid replicon for use in a shuttle vector, it was useful to determine the minimal functional replicon size, in order to minimise the size of the resulting replicon module. Therefore, only the "core" replicon (i.e. the *rep* gene, its putative promoter region, and the iterons) was initially amplified via PCR. Appropriate AscI and FseI restriction sites were added to the ends of this amplicon, and it was cloned into the pMTL80000 backbone (Heap et al., 2009), yielding pMTL87151 (Figure 4.4). The 7 representing the novel replicon, and the remaining numbers indicating that the backbone possesses the *catP* selective marker conferring resistance to chloramphenicol/thiamphenicol, the ColE1 Gram-negative replicon, and that the vector can be transferred via conjugation due to the *traJ* mobilisation system. The inclusion of a Gram-negative replicon including *traJ* was a deliberate choice, to facilitate conjugal transfer of the plasmid between *E. coli* conjugative donors and clostridial species which are only receptive to DNA

transfer via conjugation. Construction of the new vector was verified via restriction digest (Figure 4.5).

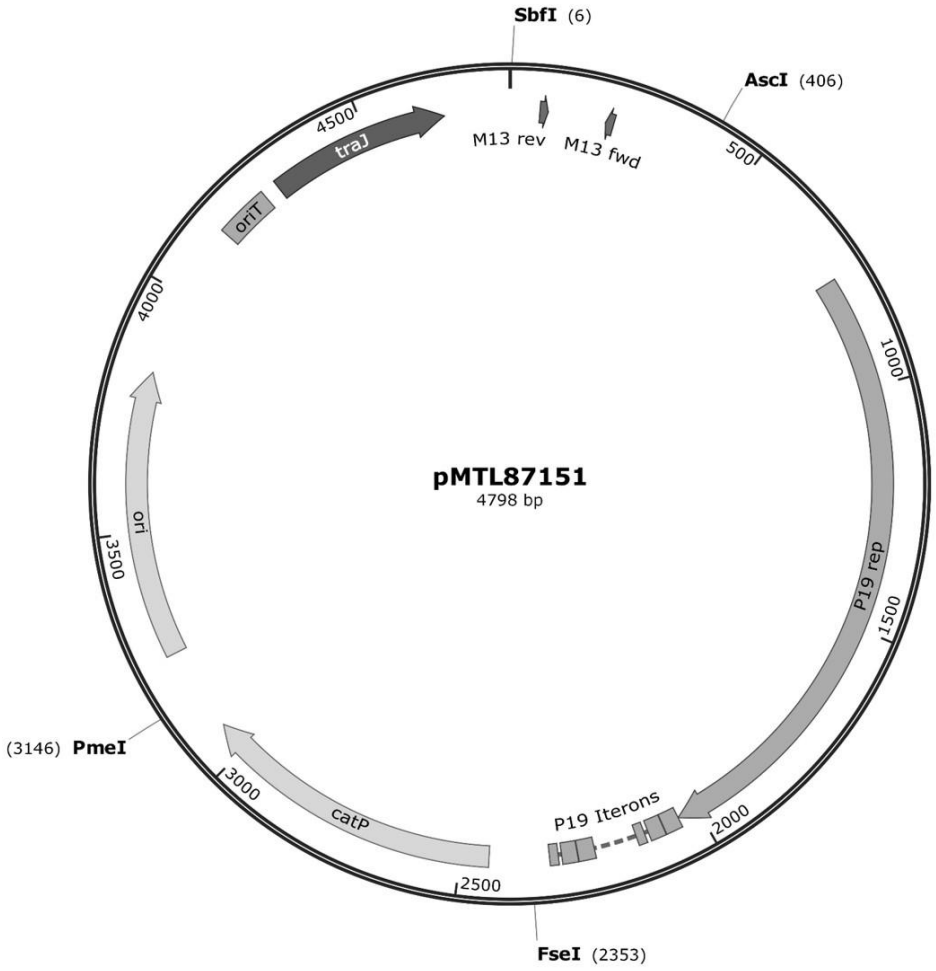


Figure 4.4: The modular shuttle vector pMTL87151. This plasmid carries the P19 Gram-positive origin of replication between the AscI and FseI restriction sites. Other features include the Gram-negative ColE1 replicon, *catP* selective marker conferring chloramphenicol/thiamphenicol resistance, and *traJ* conjugative transfer gene. The multiple cloning site (MCS) is located between the SbfI and AscI sites (and see Figure 4.3).

The amplicon generated by PCR, containing the *rep* gene and extending up- and downstream to the nearest annotated ORFs, is 1935 bp in length. The *rep* gene is therefore controlled by its native promoter, and both sets of putative

iteron sequences are included in the modular replicon. The complete sequence of the modularised P19 replicon can be found in the Appendix (see section 8.1).

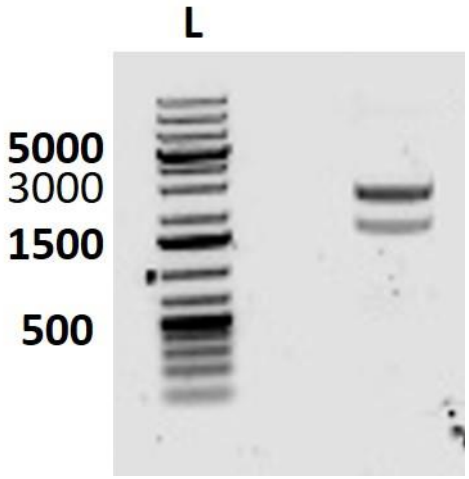


Figure 4.5: Diagnostic digest of pMTL87151. Digestion with AscI and FseI resulted in the expected 1957 and 2851 bp fragments.

4.2.2.1 Gram-positive replicon pCB102

Many of the experiments below refer to the shuttle vector pMTL83151, which pMTL87151 was derived from via replicon-swap. Gram-positive replicon pCB102 was derived from a cryptic plasmid harboured by *Clostridium butyricum* NCIB 7423 (Minton and Morris, 1981). The mechanism of function of this origin of replication is unknown, as the putative ORF it contains does not resemble any characterised *rep* protein gene (Heap et al., 2009). Plasmids replicating via pCB102 were found to be the least-stable option available in studies on *C. difficile* 630Δ*erm* (Heap et al., 2009), making it of great interest for biotechnological applications. The Gram-positive replicon derived from

pCB102 has been the subject of two published modifications aimed at disrupting its stability, and hence increasing the speed at which the plasmid can be cured. In the first case, 6.5 kb of phage lambda DNA was inserted (Heap et al., 2012). In the second example, in a more directed approach, a frame-shift mutation was introduced into the 3' end of the putative *repH* gene of the pCB102 replicon (Ng et al., 2013). The resulting plasmid, designated pMTL83*251, was found to be significantly less stable than its precursor, pMTL83251, with a 100-fold decrease in resistant CFU observed over a 24-hour period, during which no observable decrease occurred for cells harbouring pMTL83251. Decreasing plasmid stability was seen as desirable in the context of the study by Ng *et al*, as such "pseudo-suicide" vectors provided another means by which to assess the growth of single-crossover versus double-crossover integrants.

4.2.3 Comparison of Transfer Frequencies Between pMTL83151 and pMTL87151

During the creation of the pMTL80000 vector system, various Gram-positive replicons were assessed in terms of their transfer frequency. Generally, this did not vary between the replicons being assessed, though a clear increase in transfer frequency was observed in *C. botulinum* ATCC 3502 for plasmids carrying the pBP1 or pCB102 replicons in comparison to those with the pIM13 and pCD6 replicons (Heap et al., 2009).

In my study, comparable transfer frequencies were observed in *E. limosum* for plasmids pMTL83151 and pMTL87151. However, plasmid pMTL87151C (bearing a truncated version of the P19 replicon and derived in this study; see below) was observed to transfer at significantly higher frequencies. It has previously been noted that plasmid transfer frequencies can be affected by plasmid size (Heap et al., 2009), and these results must be interpreted accordingly, as pMTL83151 and pMTL87151 are relatively small (4476 and 4798 base pairs, respectively). However, the significantly increased transfer frequency of pMTL87151C seems unlikely to be purely due to a shortening of the entire plasmid length by 397 base pairs, given that pMTL83151 is of comparable size. A possible explanation for this increase in transfer efficiency is the removal of cleavage sites for *E. limosum* native restriction enzymes. This possibility is explored further in this and subsequent chapters.

4.2.4 **Segregational Stability Analysis**

Assessing the segregational stability of available plasmid replicons is an essential step when designing vectors for chromosomal modification of an organism. For many applications, replication-deficient plasmids are desirable.

During the creation of the pMTL80000 modular shuttle vector system, the stability of the system's Gram-positive replicons was assessed in a number of host species (Heap et al., 2009). In that study, the plasmids used were identical except in their Gram-positive replicons, with plasmids pMTL82151, pMTL83151, pMTL84151, and pMTL85151 assessed. The stability of those

plasmids was assessed in terms of their percentage stability per generation, indicative of the speed at which they were lost. Considerable differences were observed between the same replicon in different species, and between different replicons within the same species. It was concluded that particular Gram-positive replicons would be suited for different roles in a species-specific fashion. At the time of writing, two plasmid replicons have been demonstrated in *E. limosum*. These are P19 (this study) and pCB102. These are found on the shuttle vectors pMTL87151 and pMTL83151, respectively.

Before the construction of plasmids for mutagenesis of *E. limosum*, it is necessary to identify the most deficient plasmid replicon available. Accordingly, the stability of plasmids pMTL83151 and pMTL87151 in *E. limosum* was assessed. As in the study by Heap *et al*, these shuttle vectors provide a useful vehicle for studying plasmid maintenance in a strain of interest. Plasmid segregational stability is likely to vary with plasmid size, and the data generated here must be interpreted accordingly.

As in the study by Heap and coworkers, a modified form of a previously-established method was used to assess segregational stability of the plasmids being studied (Bron and Luxen, 1985). In that study, it was observed that increasing the size of plasmids caused a significant decrease in their segregational stability in *Bacillus subtilis*. In the study by Bron and Luxen, increasing plasmid size was also correlated with a decrease in copy number.

4.2.5 **Segregational Stability of pMTL87151 and pMTL83151 in *E. limosum***

The stability of pMTL87151 and pMTL83151 in the absence of selective pressure was assessed in *E. limosum* (

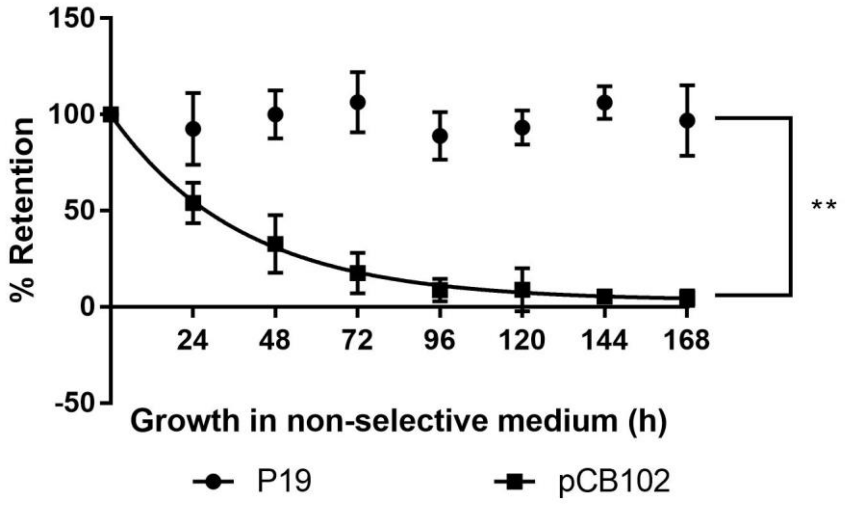


Figure **4.6**).

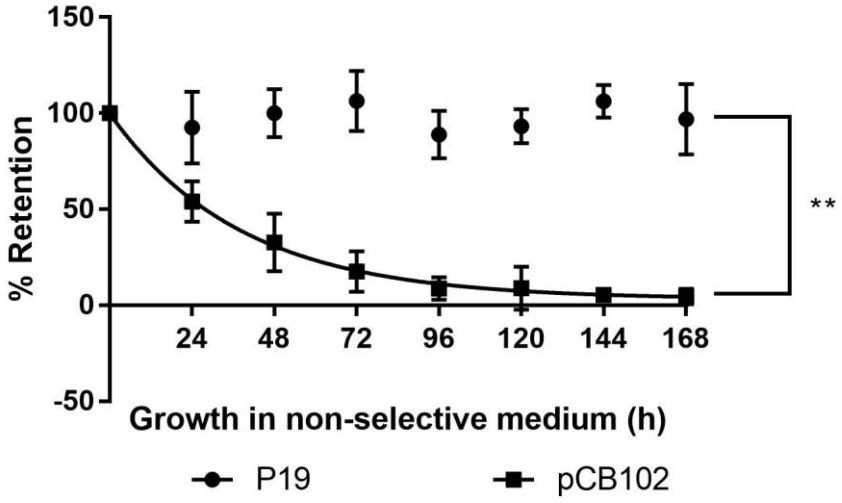


Figure 4.6: Segregational stability assay for pMTL83151 and pMTL87151 in *E. limosum*. Retention is expressed as the proportion of thiamphenicol-resistant (Tm^R) CFU versus total colonies over successive passages through non-selective liquid media. Passages occurred every 24 hours. Data points are the average of four biological replicates. Error bars represent the standard deviation. **P19:** pMTL87151. **pCB102:** pMTL83151. Differences in final stability measurements were assessed using Welch's T test with a two-tailed P value (P value = 0.0015).

Comparing the percentage of Tm^R CFU across several generations of passage reveals that while pMTL83151 is lost from the *E. limosum* population at a steady rate, no loss of pMTL87151 can be observed. The pCB102 replicon of pMTL83151 is therefore the more deficient replicon. The varying stability of the two plasmids is reflected in their impact on the growth of *E. limosum* and *C. beijerinckii* (see Figure 4.9 and Figure 4.10, below).

4.2.6 Host Range of pMTL87151

Determination of the host range of the P19 replicon is important when assessing its potential impact as a molecular biology tool. The viability of P19 was assessed in several Gram-positive species of interest, including the acetogens *C. autoethanogenum*, *C. ljungdahlii*, and *A. woodii* (

Table **4.1**). A 16S phylogenetic tree was constructed which includes the species assessed in this study, as well as *C. carboxidivorans*, the origin of the P19 replicon (Figure 4.7).

Table 4.1: Host range of the Gram-positive P19 replicon. Where given, transformation efficiency measurements are the mean average of experiments repeated in triplicate. See Materials and Methods for DNA transfer protocols for the species below. Transformation efficiency values could not be obtained for *C. pasteurianum* because competent cells were electroporated with a mixture of pMTL87151 and a plasmid expressing the methyltransferase M.BepI; instead, total transformant CFU is reported. ND = Not Determined.

Organism	**Source**	**Viable?**	**DNA transfer method**	**Transformation efficiency**
E. limosum	DSMZ	Yes	Electroporation	$1x10^5$
C. beijerinckii	G. Little	Yes	Electroporation	ND
C. autoethanogenum	DSMZ	Yes	Conjugation	ND
C. pasteurianum	A. Pech-Canul, A. Grosse-Honebrink	Yes	Electroporation	$1x10^4$ CFU in one transformation (see caption)
A. woodii	DSMZ	No	Electroporation	N/A
B. subtilis	DSMZ	Yes	Natural competence	ND
G. thermoglucosidasius	DSMZ	No	Electroporation	N/A

C. beijerinckii is a solventogenic clostridium of industrial interest. It is a close relative of *C. acetobutylicum*, and was formerly classified within that species. Transformation revealed that pMTL87151 was functional within this species.

C. pasteurianum was selected as a target for transformation of pMTL87151 because it is a solventogenic organism of considerable industrial interest, in which several tools for reverse and forward genetics have been employed (Pyne et al., 2013, Pyne et al., 2014, Zhang et al., 2015, Schwarz et al., 2017). The electroporation described in

Table **4.1** was conducted in a hyper-transformable mutant strain (Grosse-Honebrink et al., 2017).

While transformants of the ethanologenic acetogen *C. ljungdahlii* could not be obtained, pMTL87151 was found to be functiional in the very closely-related *C. autoethanogenum* when introduced via conjugation. Lack of function in *C. ljungdahlii* may be due to species-specific factors (notably, the two species have different restriction-modification system profiles), or may simply represent a failure of the transformation attempt.

While *E. limosum* and *A. woodii* are closely related, no transformants were observed during electorporation of *A. woodii* with pMTL87151, using a highly-robust transformation protocol (Jon Baker, personal communication). Given the high transformability of *A. woodii*, lack of transformation with pMTL87151 may represent the effect of host factors on plasmid replication and maintenance (which may also explain the species-to-species variability of plasmid copy number observed in this study). While *E. limosum* and *A. woodii* are closely related (see Figure 4.7), they evidently do not interact with the P19 replicon in the same way.

Additionally, the functionality of P19 was assessed in the aerobic Gram-positive species *B. subtilis* and *G. thermoglucosidasius*. The former is a model organism for studies in Gram-positive organisms, and has the advantage of being naturally competent. *G. thermoglucosidasius* is a thermophilic organism of industrial interest. No transformants were observed during attempted transformation of *G. thermoglucosidasius*, while transformants were observed

during transformation of *B. subtilis*. The fact that pMTL87151 is compatible with the aerobe *B. subtilis* may facilitate future screening of temperature-sensitive mutants (or other high-throughput experiments), because large-scale automated experiments are much more practical when an anaerobic environment is not necessary.

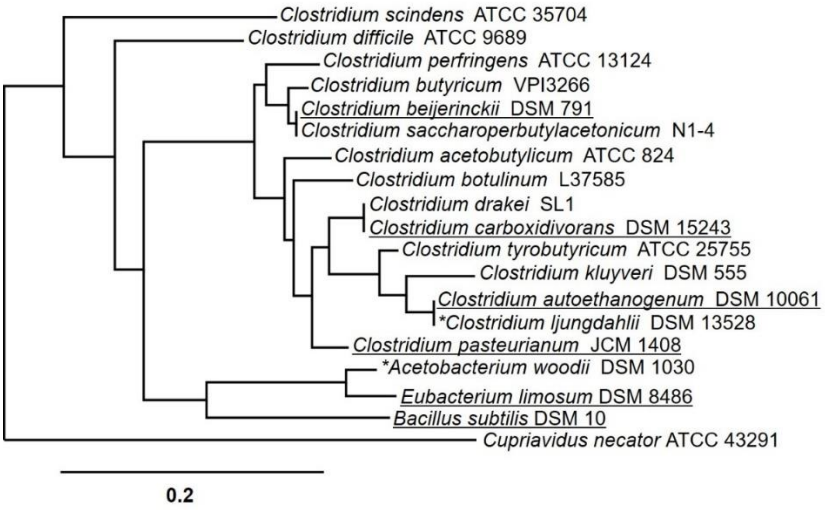


Figure 4.7: 16S rRNA phylogenetic tree of Gram-positive species of interest, including acetogens and solventogenic clostridia. Underlined: Species in which pMTL87151 has been tested and is functional (except *C. carboxidivorans*, which harbors the native P19 plasmid). Species in which pMTL87151 has been found to be non-functional are marked with asterisks.

4.2.7 Plasmid Copy Number in *E. limosum*

As noted previously, comparative data on the properties of Gram-positive replicons is limited (Heap et al., 2009). To-date, the copy numbers of Gram-positive replicons have remained relatively unstudied. Plasmid copy numbers have previously been assessed in several Gram-negative bacterial species through the use of quantitative PCR (qPCR) (Providenti et al., 2006). This

method is advantageous because it is self-calibrating; with an appropriately-selected amplicon in the chromosome as a reference, the copy number of a plasmid-borne amplicon can be determined. The absolute quantity of DNA in the sample does not need to be determined. A limitation of this approach is that both reference and target amplification signals must be derived from the same genomic DNA sample in order for comparisons to be meaningful.

To better understand the behaviour and dynamics of the modularised P19 replicon, the copy number of pMTL87151 was assessed via qPCR in *C. beijerinckii* and *E. limosum*. To provide a point of comparison, the copy number of pMTL83151, bearing the pCB102 Gram-positive replicon, was also assessed for these species. The copy number of the native P19 plasmid was also assessed in *C. carboxidivorans*, its native context. Finally, the copy number of the P19-derived truncated replicon P19C (refer to section 4.2.11) in the plasmid pMTL87151C was also assessed in *E. limosum* to determine whether the increased transformation frequency observed for that replicon was correlated with a difference in copy number.

Prior to running qPCR experiments, the primer pairs utilised in the study were validated by agarose gel electrophoresis to determine that they were producing unique amplicons (Figure 4.8). Conditions for the test PCR were adjusted to approximate to those found during the qPCR itself (i.e. an annealing temperature of 55^{o}C was used, as it was during the qPCR). Product formation was observed at the expected sizes for the chromosomal primer sets for *E.*

limosum, *C. beijerinckii*, and *C. carboxidivorans*, as well as for the plasmid-specific primers (130 bp for pMTL83151, 75 bp for pMTL87151).

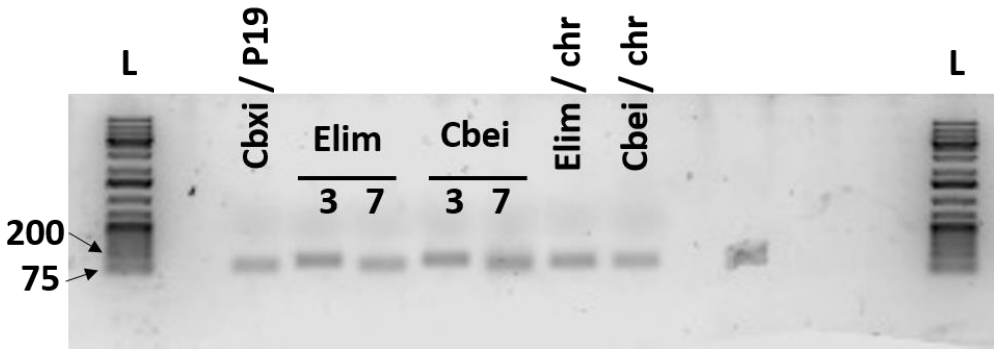


Figure 4.8: Validation of primer sets used in qPCR experiments in this study. DNA template and primers concentrations are identical to those employed in qPCR experiments in this study. **L:** Ladder. **Elim:** *E. limosum* 87151/83151 genomic DNA template. **Cbei:** *C. beijerinckii* 87151/83151 genomic DNA template. **Cbxi:** *C. carboxidivorans* genomic DNA template. **P19:** Primers amplifying region of native P19 plasmid. **3/7:** Primers amplifying replicon sequence of pMTL87151 and pMTL83151 respectively **Chr:** Primers amplifying chromosomal region of *E. limosum* or *C. beijerinckii*.

To assess the plasmid copy numbers of pMTL87151 and pMTL83151 throughout the growth cycle of *E. limosum*, samples were collected at intervals through a growth curve (Figure 4.9). The impact of pMTL83151 on host growth is immediately apparent, with growth of *E. limosum* pMTL83151 being significantly delayed. The copy number of pMTL83151 appears to increase throughout the period measured, increasing from 8 ± 3 in mid-exponential phase to 25 ± 5 in stationary phase.

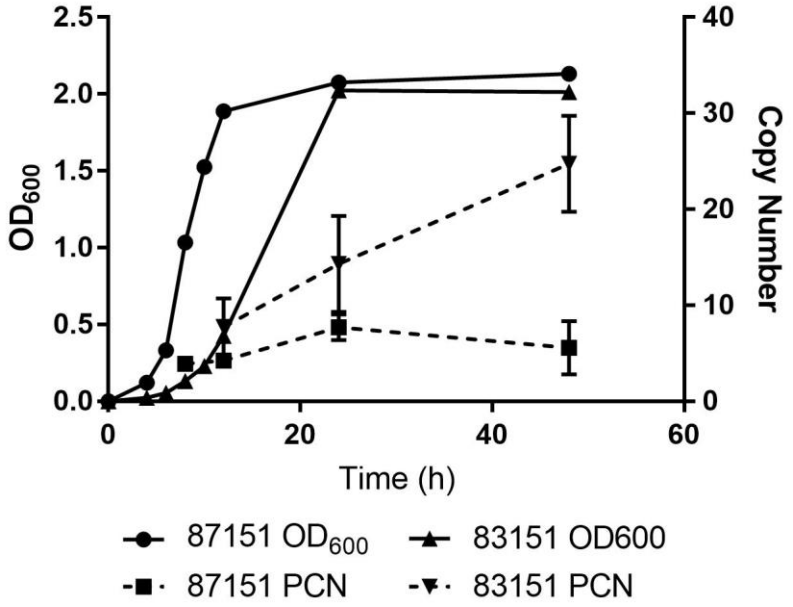


Figure 4.9: Plasmid copy number (PCN) of pMTL87151 and pMTL83151 in *E. limosum* ATCC 8486. Solid lines represent bacterial growth, dashed lines represent PCN. Error bars on growth curves represent the mean average of three biological replicates, and are not visible at this scale. Error bars on PCN measurements represent the mean average of three biological replicates (of three technical replicates each). All growth occurred under antibiotic selection.

In contrast, pMTL87151 does not appear to hinder *E. limosum* growth relative to cultures not harbouring a plasmid. It is also apparent that pMTL87151 is maintained at a lower copy number than pMTL83151, beginning at 4 ± 0.5 in mid-exponential phase, and finishing at 6 ± 3 in stationary phase. While the copy number of pMTL87151 appears to remain relatively stable, the copy number of pMTL83151 was observed to increase throughout the growth experiment, beginning at 8 ± 3 and ending at 26 ± 5.

4.2.8 **Plasmid Copy Number in *C. beijerinckii***

In order to assess the copy number of pMTL87151 throughout the growth cycle of *C. beijerinckii*, a growth experiment was conducted, and cell samples withdrawn during exponential and stationary phases (Figure 4.10).

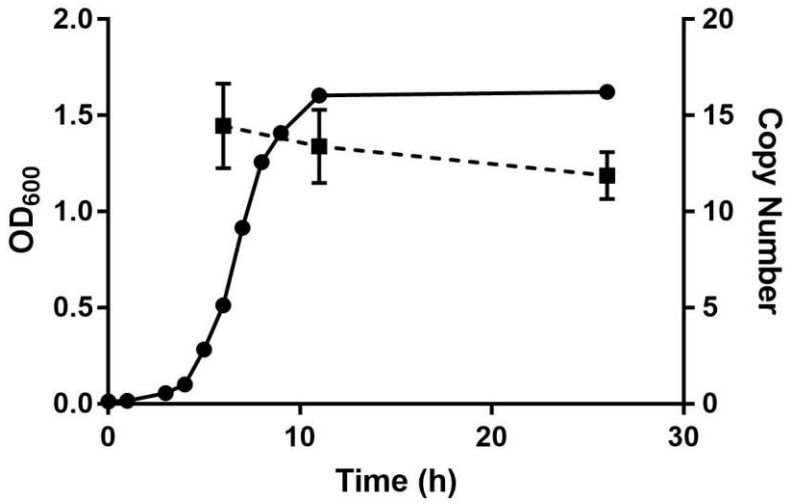


Figure 4.10: Plasmid copy number (PCN) of pMTL87151 in *C. beijerinckii*. The solid line represents optical density (OD). The dotted line represents PCN. Error bars of growth curve represent the mean average of three biological replicates. Error bars of the PCN measurement represent mean average of three biological replicates (of three technical replicates each).

Genomic DNA samples were collected from three biological replicates of *C. beijerinckii* during this growth experiment, and used in a subsequent qPCR experiment. The copy number of pMTL87151 in *C. beijerinckii* was seen to vary between 14 ± 2.2 in mid-exponential phase to 12 ± 1.2 in stationary phase. Accordingly, pMTL87151 can be considered a low-copy plasmid in *C. beijerinckii*. To provide a point of comparison, qPCR was used to determine the copy number of pMTL83151 in *C. beijerinckii* (Figure 4.11).

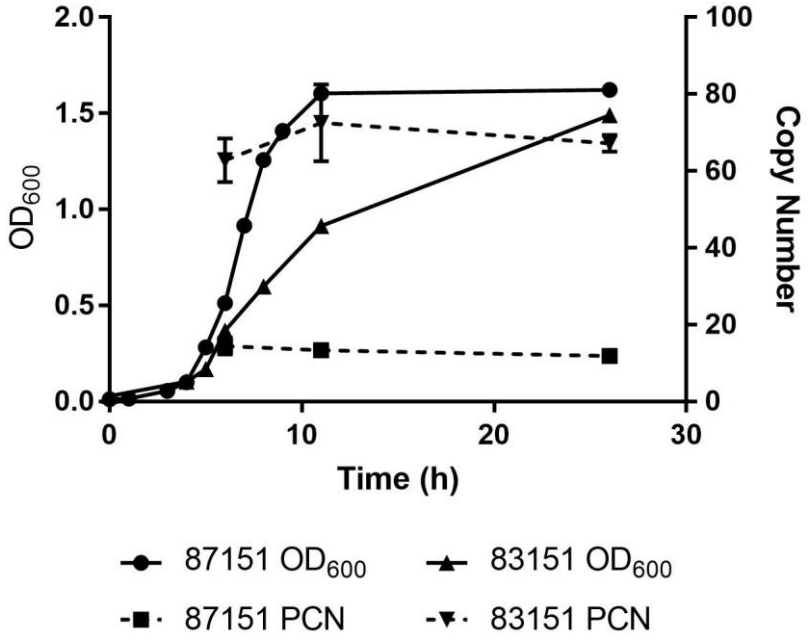


Figure 4.11: Comparison of growth and plasmid copy number of *C. beijerinckii* harboring pMTL83151 and pMTL87151. PCN are represented by dashed lines, bacterial growth by solid lines. OD measurements for *C. beijerinckii* pMTL87151 are the mean average of three biological replicates. Error bars represent the standard deviation, and are not all visible at this scale. The growth curve for *C. beijerinckii* pMTL83151 represents a single biological replicate. Error bars for PCN measurements are the mean of three biological replicates (of three technical replicates each) for pMTL87151 and one biological replicate (of three technical replicates) for pMTL83151.

The impact of pMTL83151 on growth of *C. beijerinckii* can be clearly discerned in Figure 4.11, with growth being delayed as was observed in *E. limosum*. Samples were collected at the same time-points as during the growth of *C. beijerinckii* pMTL87151, and plasmid copy numbers vary from 63 ± 6 in mid-exponential phase to 67 ± 2 in stationary phase. As with pMTL87151, the copy number of pMTL83151 is seen to be relatively stable throughout the growth stages of *C. beijerinckii*. The relative stability of copy number for both plasmids in *C. beijerinckii* is in contrast to that observed in *E. limosum*, in

which pMTL83151 appeared to increase in copy number over the course of the growth experiment. Interestingly, the copy numbers achieved by both replicons were seen to vary depending on the organism being studied. Both plasmids are maintained at higher copy numbers in *C. beijerinckii* than in *E. limosum*.

4.2.9 P19 Copy Number in *C. carboxidivorans*

To better characterise the cryptic plasmid from which the P19 replicon was derived, qPCR was conducted on genomic DNA isolated from *C. carboxidivorans* (Figure 4.12). This experiment also provides a point of comparison between the wild-type P19 replicon and the minimal replicon present in pMTL87151.

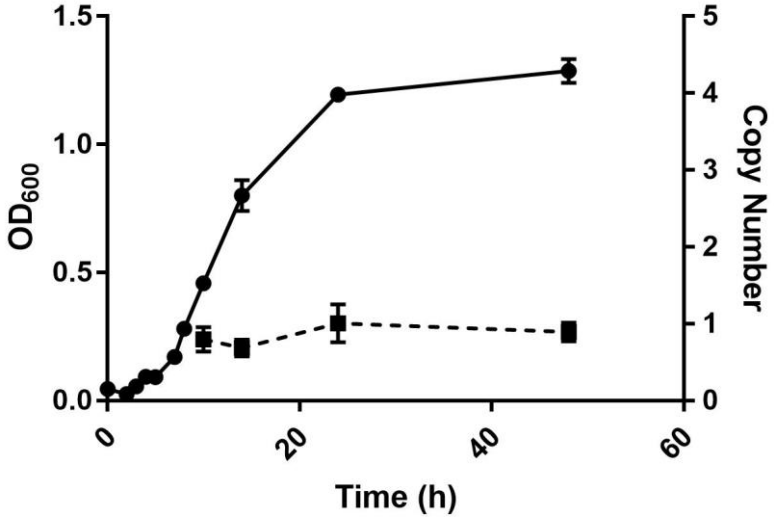


Figure 4.12: Plasmid copy number of cryptic plasmid P19 in *C. carboxidivorans*. Solid line represents bacterial growth. Dashed line represents PCN. Error bars of growth curve represent the mean average of three biological replicates. Error bars of PCN measurements represent the mean average of two biological replicates (of three technical replicates each).

Analysis of the copy number of the cryptic P19 plasmid in *C. carboxidivorans* indicates that the plasmid is maintained at a single copy throughout growth. This is supported by the presence of *par* genes on the plasmid backbone, as noted above. These genes encode active partitioning systems which are characteristic of low-copy plasmids (Million-Weaver and Camps, 2014). The *parA*-possessing plasmid R1 has previously been noted to be single-copy (Nordstrom, 2006).

Alternatively, this may indicate that in its native context, P19 spends time integrated into the genome of *C. carboxidivorans*. The presence of several phage-derived integrases on the backbone of P19 may support this idea. To

determine whether P19 existed as a free plasmid, a PCR was conducted with primers reading across the ends of the assembled P19 sequence. This produced a product, indicating that P19 does exist at least partially in a covalently-closed form within *C. carboxidivorans* (data not shown). This conclusion is further supported by the fact that all genome shotgun sequences of *C. carboxidivorans* at the time of writing feature P19 as a 19 Kb scaffold, whereas if it were integral to the chromosome, shotgun assembly algorithms would presumably not have sorted into identically-sized fragments in all three cases.

P19 does not appear to contain genes essential to survival such as antibiotic resistance or heavy metal tolerance genes. Its stability as a low- or single-copy plasmid is therefore likely to depend on the *par* genes observed during sequence analysis. A role may also be attributed to the two distinct regions of direct repeats (iterons) which have been identified via sequence analysis in this study. The iterons directly downstream of the *rep* protein were implicated in function of the plasmid replication origin in the course of this study (see below), but the other set of iterons may function as the centromere of the *par* system, to which an adaptor protein and polymerised motor protein would bind. This is supported by their close proximity to the *par* genes on the P19 backbone (see Figure 4.1).

4.2.10 **Functional Dissection of the P19 Replicon**

Truncation of plasmid replicons can be used to define the minimum functional replicon. Some, such as the RepJ replicon of plasmid pPSV48c found in *Pseudomonas*, can be viable after several functional components have been removed (Bardaji et al., 2017). To determine whether the module previously generated by PCR was the minimal functional replicon, the sequence was truncated by PCR, creating three variants. These variants removed 1) the putative promoter region, 2) both sets of iterons, or 3) a single set of iterons (Figure 4.13). Generation of truncation products was assessed via gel prior to ligation (Figure 4.14). Truncation of the putative Rep gene promoter reduced transformation efficiency by several orders of magnitude, as did truncation of both sets of iterons (Figure 4.15).



Figure 4.13: Functional dissection of the P19 replicon via truncation. **P19:** Original modularised P19 replicon, including all components. **-35, -10**: Putative promoter components driving expression of P19 *rep* gene. **Rep**: gene encoding replication initiation protein. **Iterons:** Repeat sequences to which Rep protein binds during initiation of plasmid replication. **A:** P19A variant replicon in which putative promoter region is removed. **B:** P19B variant replicon in both sets of iterons have been removed. **C:** P19C variant replicon in which one set of iterons has been removed.

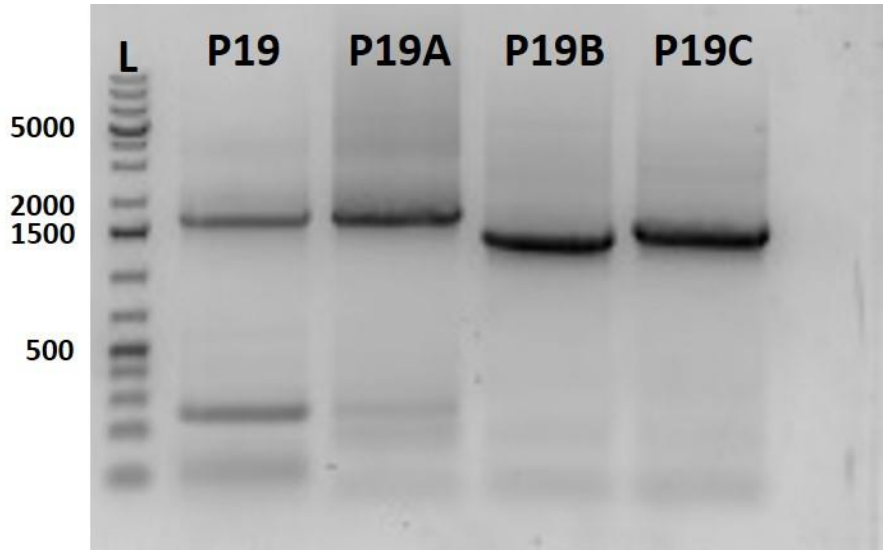


Figure 4.14: PCR across the Gram-positive replicon region of pMTL87151 and variants. **L:** GeneRuler 1 Kb Plus DNA Ladder (NEB). Note varying lengths of variant replicons due to truncation. Amplicons were sequenced to verify truncated replicon cloning.

However, removal of only one set of iterons appeared to enhance transformation efficiency by two orders of magnitude (Figure 4.15). The reduced length of the replicon may contribute to enhanced transformability. Plasmid length has been noted to influence both the transfer frequency and copy number of plasmids in Gram-positive bacteria (Bron and Luxen, 1985), and in this study large differences in transformation efficiency have been observed as plasmid sizes increase. However, this is unlikely to explain the observed increase in this case, as the difference in length between P19 and P19C is only 276 bp.

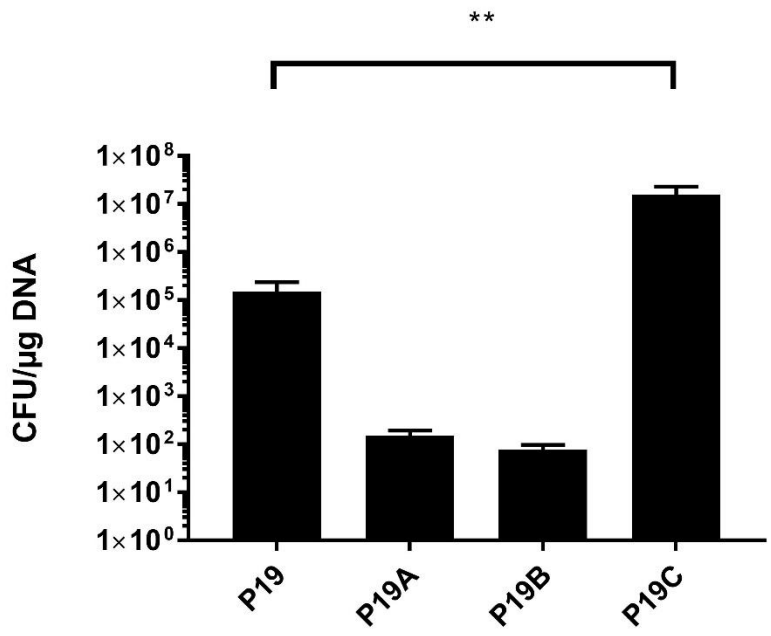


Figure 4.15: Transformation efficiency of full-length and truncated P19 replicon variants into *E. limosum* ATCC 8486. **P19A**: Putative rep promoter region removed. **P19B**: Second set of iterons removed. **P19C**: Both sets of iterons removed. Bars represent mean averages of duplicate transformations (triplicate transformations for P19 and P19C). Error bars represent the standard deviation. All transformations were conducted with the same batch of *E. limosum* competent cells. Transformation efficiencies for P19A and P19B were extrapolated from single-digit numbers of putative transformant colonies. One-way ANOVA was used to assess the significance of the gene transfer differences observed (P-value 0.0017).

The increase in transformation efficiency observed when transforming pMTL87151C may be attributable to the absence of cleavage sites for restriction endonucleases native to *E. limosum* (see Discussion). Methylation motifs associated with the *E. limosum* ATCC 8486 genome have been identified (see Chapter 7). Two putative cleavage sites for *E. limosum* ATCC 8486 native restriction systems can be found within the iteron region that is truncated in P19C.

4.2.11 **Characterisation of the truncated replicon P19C**

To investigate the possibility that truncation of the P19 replicon had resulted in an alteration in copy number, qPCR was conducted, using the same methods employed previously for the study of P19 and pCB102 copy number. Copy number analysis via qPCR revealed no significant difference in the copy number of P19C as compared to P19 (see Figure 4.16, below). Additionally, no difference in growth was observed between *E. limosum* harboring pMTL87151 and pMTL87151C (Figure 4.16).

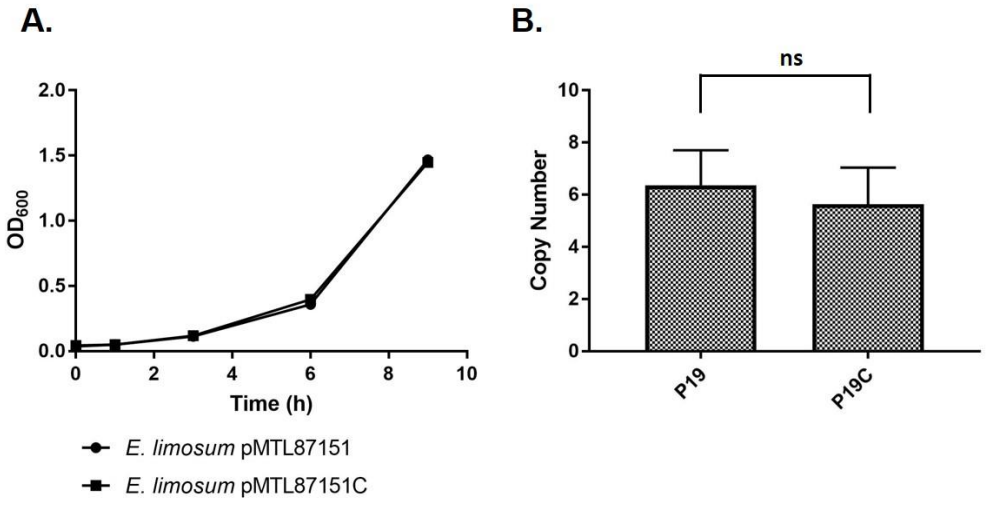


Figure 4.16: Comparison of bacterial growth and plasmid copy number of pMTL87151 (P19) and pMTL87151C (P19C) in *E. limosum*. **A.** Growth curve comparing both strains (only one biological replicate of two was measured) extending into late exponential phase. **B.** Plasmid copy number assay. Samples for PCN measurement were collected at 9 h following inoculation. Column heights represent the mean average of two biological replicates of three technical replicates each. Error bars represent the standard deviation. No significance was observed between the mean PCN values for P19 and P19C (Unpaired Welch's T test with a two-tailed P value; P value 0.6463).

While truncation of the P19 replicon to produce P19C serves to increase transfer efficiency of the new plasmid, this does not appear to be a result of altered copy number. Additionally, the segregational stability of pMTL87151C, harbouring the truncated replicon, was not significantly less than that of the replicon from which it was derived (

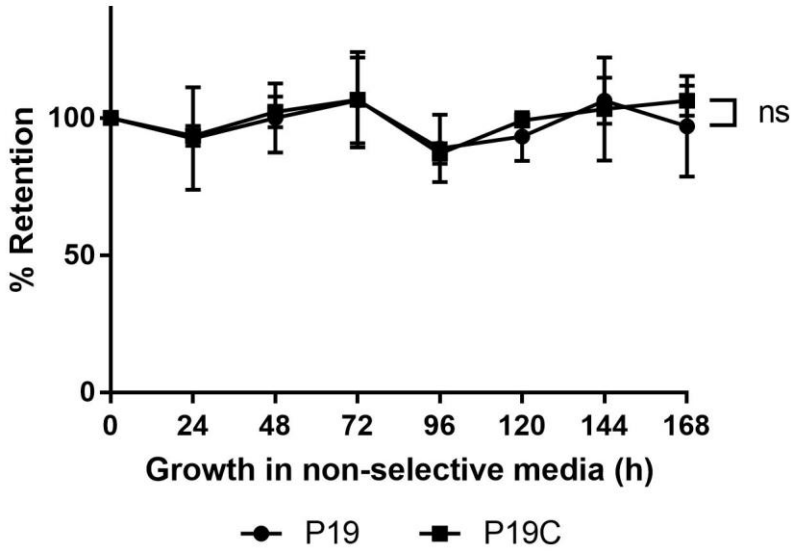


Figure **4.17**). This indicates that the sequence removed by truncation is dispensable in terms of both copy number control and segregational stability.

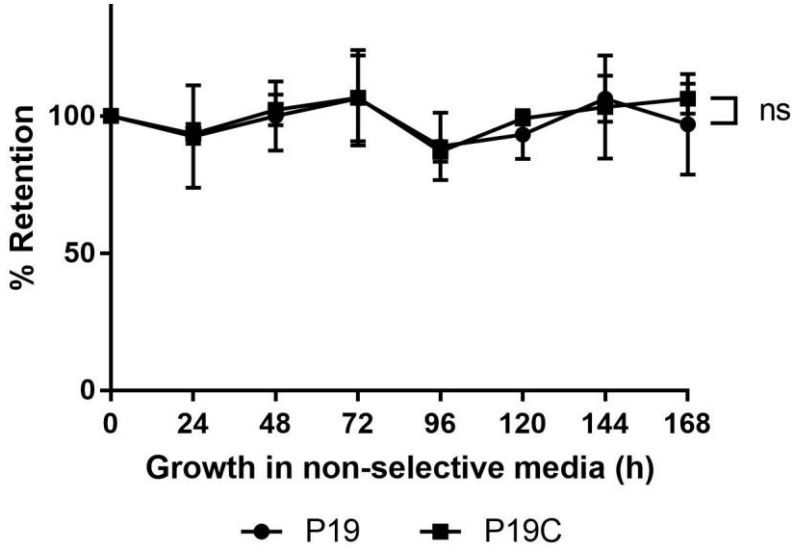


Figure 4.17: Comparison of segregational stability over time for plasmids pMTL87151 and pMTL87151C. Data for pMTL87151 is identical to that in

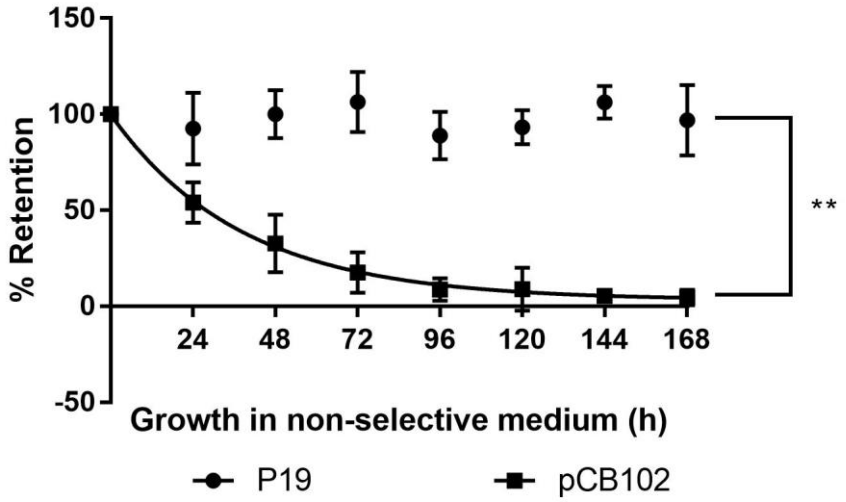


Figure 4.6, above, and is included only for comparative purposes. % Retention indicates the number of Tm^R colonies vs number of colonies on non-selective medium. Data points are the mean average of three biological replicates for P19C and four biological replicates for P19. Error bars represent the standard deviation. Differences in final stability values were assessed with Welch's T test with a two-tailed P value (P value = 0.3931).

4.2.12 Employment of the P19 replicon in other studies

Including the transposon mutagenesis work described in this study (see Chapter 5), the modularised P19 replicon has been employed in other studies involving *E. limosum* and other acetogenic bacteria. It has been used as the Gram-positive replicon for large synthetic constructs expressing genes for acetone and isopropanol production in *E. limosum* (R. Mansfield, personal communication). It has also been used in studies involving fluorescently-tagged proteins in *C. autoethanogenum* in which P19-based expression vectors were successfully maintained in the absence of antibiotic selection (P. Piatek, personal communication). In that same study, P19-based vectors expressing *C. autoethanogenum* genes were used to complement knockout mutants. These studies further demonstrate the utility of this Gram-positive replicon in the study of *E. limosum* and *C. autoethanogenum*, both acetogenic species of considerable industrial interest (the latter has been utilised in industrial-scale production of ethanol from steel mill waste gases).

4.3 Discussion

This study establishes the P19 Gram-positive replicon as a useful tool for plasmid construction and use in Gram-positive organisms. The structure of the replicon has been characterised via functional dissection, and the copy number determined with reference to the widely-used pCB102 Gram-positive replicon. Additionally, a variant replicon, designated pMTL87151C, has been identified

which is of particular interest in studies involving *E. limosum* due to its apparent increased efficiency of transfer.

Synthetic plasmids with low and consistent copy numbers have previously been a subject of interest in industrial contexts. It has been observed that unstable, high-copy plasmids have a deleterious effect on the yield of recombinant proteins during industrial fermentation (Grabherr et al., 2002). In the context of gas fermentation by modified acetogens, however, the emphasis is on production of chemicals via novel genes, rather than the yield of the proteins catalysing production of those chemicals *per se*. In addition, current studies mainly focus on direct chromosomal modification of these organisms of interest, removing the need for plasmid maintenance via antibiotics and providing a stable copy number. However, plasmid-based strategies may offer advantages in industrial contexts. The expression of heterologous proteins at greater than chromosomal copy numbers may be useful in some contexts.

Current strategies for chromosomal modification of clostridia involve the use of replication-deficient plasmid vectors, which can be readily lost in the absence of selection (Minton et al., 2016). Identifying the most deficient replicon available for a particular organism is therefore a priority. In this study, it can be seen that pCB102 is the more deficient of the two available replicons. In contrast, P19 displays a high degree of stability.

Growth experiments with *E. limosum* harboring pMTL87151 and pMTL83151 show an obvious difference in growth rates between the two cultures, with pMTL83151 having a greater impact on growth. Metabolic stress caused by

plasmid replication has been implicated in plasmid instability (Silva et al., 2012), though systematic study of the physiological effects of plasmid maintenance has apparently only been conducted in *E. coli*. The copy number differences observed for P19- and pCB102-based plasmids offer a ready explanation for the observed differences in growth rate and stability. Differences in segregational mechanism (active partitioning versus random distribution between daughter cells) may also explain these differences. This would be unusual in the case of P19, because the modularised replicon dispenses with many of the components that would presumably be involved in partitioning of the wild-type P19 plasmid (e.g. the *par* genes).

The variance of plasmid segregational stability is similar to that observed between replicons of the pMTL80000 system, with the characteristics of each replicon being inconsistent between clostridia (Minton et al., 2016). The presence of a highly-stable replicon may be advantageous in future studies in *E. limosum*. A comparable case of 100% segregational stability has previously been observed in plasmid pLB2 in *Bacillus subtilis* (Bron and Luxen, 1985). Notably, in that study the high segregational stability of pLB2 was seen to decrease in proportion to the length of DNA inserts added to it from a variety of sources. The exponential decay in plasmid retention observed for replicon pCB102 during this study is characteristic of slow-growing plasmid-harboring cells being overtaken by the non-plasmid-bearing population (Lee et al., 1993).

The best analogue to the P19 replicon appears to be the pIP404 replicon of *C. perfringens*, which also consists (at a minimum) of a *rep* gene and downstream iteron region (Garnier and Cole, 1988), with a dispensable *cop* gene which is implicated in the control of copy number. As in the present study, the minimal pIP404 replicon was isolated from a native plasmid and analysed via deletion analysis. The *rep* genes of P19 and pIP404 are of comparable size, and their amino acid sequences possess significant similarity, although the P19 *rep* is 50 bp longer and their associated iteron regions are entirely dissimilar. The *rep* genes of pIP404-related plasmids in *C. perfringens* isolates are generally ~1200 bp in length (Miyamoto et al., 2015), as is the *rep* of P19. Interestingly, the deletion analysis of pIP404 conducted by Garnier and Cole (1988) took place in *B. subtilis*. Given the structural and sequence similarities between P19 and pIP404, it is unsurprising that P19 was found to be functional in *B. subtilis* in this study.

Plasmid copy numbers have been obtained for clostridial shuttle vectors in the past, in studies of *C. perfringens* (Allen and Blaschek, 1990) and *C. acetobutylicum* (Lee et al., 1993). These studies employed gel electrophoresis-based quantification methods. Both studies additionally studied the segregational stability of the plasmids concerned, using methods comparable to those employed in this study. This study is the first to employ quantitative PCR to determine plasmid copy number in clostridia, and is the first study of plasmid copy number in *E. limosum*. At the time of writing, more sophisticated methods such as droplet digital PCR (ddPCR) are being employed to measure plasmid copy number in *E. coli* (Plotka et al., 2017).

These methods may be useful in future studies of clostridial plasmid copy number.

The varying copy number observed between the Gram-positive species tested (*E. limosum*, *C. beijerinckii* and *C. carboxidivorans*) indicates that copy number control of pMTL87151 depends at least to some extent on host factors. While no obvious RNA regulatory elements were identified within the P19 replicon during this study, they may be present within the modularised replicon, and have been noted within comparable Gram-positive replicons (Grohmann et al., 2016). The difference between the copy numbers of the wild-type and modularised P19 replicons is interesting. This difference may be an artefact of differing plasmid size, or may indicate additional replicon components on the wild-type plasmid. The copy number of the minimal P19 replicon is probably determined by interactions between the Rep protein and the iteron sequences. Once multiple plasmids are available in the cell, Rep-iteron complexes may form between multiple plasmids, resulting in "steric handcuffing" and inhibiting further plasmid replication (Nordström, 1990).

It is interesting to note that the P19C replicon is functional even after one set of iterons has been removed. Unsurprisingly, complete removal of the iterons (in P19B) or the promoter for the *rep* gene (in P19A) reduces transformation. Iterons are generally understood to be responsible for Rep protein binding, and play a role in initiation of DNA melting and plasmid replication. The fact one set is dispensable in this case is not unprecedented, with previous studies showing continued (though altered) function of plasmid replicons after one or

more sets of iterons has been removed (Kolter and Helinski, 1978). It appears that a single P19 iteron binds Rep proteins as well as a duplicate pair of iterons.

Iterons are also known to a play a role in the segregation of low-copy plasmids, acting as a centromere to which motor proteins will bind (reviewed in Million-Weaver and Camps, 2014). The segregational stability of the P19 modular replicon may be due to its iterons acting as a centromere, although the *par* adaptor which would normally bind to the centromere is encoded by the native P19 plasmid and is not present on the modularised replicon. Additionally, the segregational machinery encoded by the P19 plasmid probably binds to another set of iterons, located immediately downstream of the *parA* gene on the native plasmid (see above). The *rep* gene present within the P19 minimal replicon may play a role in plasmid segregation. It has previously been noted that in Gram-positive plasmid replicons, the genetic elements responsible for plasmid segregation and replication are closely associated physically, and may overlap (Dmowski and Jagura-Burdzy, 2013).

The apparent increase in transformability observed during transformation of pMTL87151C has implications for future genetic engineering of *E. limosum* particularly. The truncated P19C replicon does not appear to differ from P19 in copy number or segregational stability, and therefore these factors cannot explain its efficiency of transfer to *E. limosum*. Methylome analysis of *E. limosum* ATCC 8486 indicates that the motif CAAAAAR is methylated. Three such motifs are found in the P19 replicon, and two of these are within

the iteron region. One of these is removed in P19C. It is tempting to assign the removal of a restriction site as the reason for the observed increase in DNA transfer efficiency. However, the 100-fold increase observed appears very high for the removal of only one third of such sites on a plasmid. Additionally, the pCB102 replicon found on pMTL83151 possesses no CAAAAAR motifs, and is not transferred into *E. limosum* with the same high efficiency as pMTL87151C. The implications of these findings are discussed further in Chapter 7, which deals with the deletion of an endogenous restriction-modification system in *E. limosum*. The only difference between P19 and P19C is the iteron region, which is a binding site for DNA-interacting proteins. The altered iteron structure of P19C may result in more efficient interaction with surface proteins which bind DNA (resulting in more efficient import of the plasmid during electroporation), although this hypothesis has not been investigated further in this study. The interaction of iteron-based plasmid replicons with membrane-bound proteins has been noted before in the pIP404 plasmid of *C. perfringens* and in the staphylococcal plasmid pUB110 (Garnier and Cole, 1988), but there has been no suggestion that iterons may play any role in horizontal gene transmission. Interestingly, an increase in the number of iterons associated with the replicon of the pAD1 plasmid of *Enterococcus faecalis* has been shown to affect conjugative efficiency (Heath et al., 1995).

Future directions for this work may include determining the segregational stability and plasmid copy number of the P19 replicon in *C. autoethanogenum*. This would be of interest given the industrial interest demonstrated for that species. During this project, the stability of pMTL87151 was tested in *C.*

beijerinckii via serial restreaking on non-selective medium (data not shown). The plasmid was found to be 100% stable, as observed during serial passage of *E. limosum*. Testing this replicon (and determining its segregational stability) in a greater variety of acetogenic and solventogenic bacteria may be a useful direction for research.

Another useful experiment would be to characterise the stability of the P19 replicon when maintaining large constructs. It would be interesting to see whether the segregational stability noted for pMTL87151 applies to the maintenance of larger plasmids. Maintaining artificial operons on plasmid backbones in the absence of antibiotic selection would be a useful capability for strain engineering in *E. limosum*. Conversely, segregational instability would be a useful quality for vectors used for in-frame gene deletions and transposon mutagenesis.

4.4 Chapter Summary

- A novel Gram-positive plasmid replicon has been isolated from the native plasmid P19 of *C. carboxidivorans* P7 and introduced into the pMTL80000 modular vector system, yielding pMTL87151.
- The host range of this plasmid among Gram-positive organisms has been explored. It is functional in the acetogens *E. limosum* and *C. autoethanogenum*, the solventogenic clostridial species *C. beijerinckii*, as well as the Gram-positive aerobe *B. subtilis*.

- The copy number of pMTL87151 has been determined in *E. limosum*, *C. beijerinckii*, and *B. subtilis*. It was found that pMTL87151 has a low copy number in comparison to pMTL83151, whose copy number was also characterised for the first time in this study.

- The modularised P19 replicon was truncated in order to identify functionally indispensible components. Truncated variant pMTL87151C, lacking one set of iterons, was found to not only be functional, but to transfer to *E. limosum* with a higher frequency that pMTL87151. At the time of writing, this increase in transfer efficiency has not been assessed for other bacterial species. This may be attributable to removal of a cleavage motif for one of *E. limosum*'s restriction-modification systems, although pMTL83151 (lacking all such sites) does not transfer with comparably high frequency.

CHAPTER FIVE

Exemplification of an aTc-inducible *mariner* Transposon System in *Eubacterium limosum*

5 Exemplification of an aTc-inducible *mariner* Transposon System in *Eubacterium limosum*

5.1 Introduction

In this chapter, insertional mutagenesis of *E. limosum* ATCC 8486 via a *mariner* transposon is demonstrated. The total number of unique mutants obtained was limited by the relative inefficiency of DNA transfer. This chapter serves not only to demonstrate the function of an inducible transposon vector in *E. limosum*, but to illustrate the use of an anhydrotetracycline (aTc) – induced gene expression system for other purposes. The TetR inducible system was found to be time- and dose-dependent in *E. limosum*, and relatively inactive in the absence of induction according to the results of a CatP assay. However, the appearance of spontaneous transposon mutants in the absence of induction indicates that the basal activity of the TetR system in *E. limosum* is sufficient to reduce the number of unique transposon mutants that can currently be obtained.

5.1.1 Transposons

Transposon mutagenesis is an important tool for forward genetic studies, in which mutants are generated, screened for desirable phenotypes, and only subsequently characterised in terms of their genotype.

Creation of a knockout transposon mutant library requires both random insertion into the target genome, and the insertion of only one transposon per mutant strain (Klein et al., 2015). Insertional mutagenesis studies can be hampered by the preferential insertion of transposons into particular chromosomal regions. This effect has been observed in the transposon mutagenesis of *Staphylococcus aureus* using the Tn*917* and Tn*551* systems (Yajjala et al., 2016). Transposition using *mariner*-based systems was found to avoid this insertional bias. This is because the *mariner* transposon's only sequence requirement for insertion is the dinucleotide TA, resulting in a high degree of randomness of insertion, even in GC-rich regions (Bryan et al., 1990).

5.1.2 *Mariner* Transposons

The *mariner* transposable element was originally identified in *Drosophila mauritiana* (Jacobson et al., 1986) (Bryan et al., 1990). The originally-identified *mariner* is a 1.3 kilobase sequence consisting of a transposase gene flanked by inverted terminal repeat (ITR) sequences. *Mariner*-family transposons which differ by as few as 6 out of 348 amino acids have been observed in fly species which diverged over 200 million years ago (Lampe et al., 1996). Noting the broad species distribution of *mariner* transposons, Lampe *et al* suggested that they functioned independently of host factors, suggesting that each *mariner* element was a self-contained transposing system. The *mariner* element *Himar1*, isolated from the horn fly *Haematobia irritans*,

was demonstrated to be active *in vitro* when purified and capable of transposing itself from two donor plasmids into radio-labelled target DNA (Lampe et al., 1996). The donor plasmids possessed the same *mariner* transposon, but differed in the length of their plasmid backbones. This allowed Lampe *et al* to distinguish between the formation of integrated plasmid-target DNA structures (which could be distinguished based on size) and simple cut-and-paste insertion from the donor plasmid into the target DNA. Simple insertions into the target DNA were observed, indicating both that the *mariner* transposon inserted via a cut-and-paste mechanism, and that it was able to do so *in vitro*, independent of host factors.

5.1.3 Transposon Mutagenesis in Clostridia

Transposon mutagenesis of clostridia using *in vitro* transposases has been demonstrated in *C. perfringens*. Transposons used include the bacteriophage Mu transposon (Lanckriet et al., 2009) and the EZ-Tn*5* transposon (Vidal et al., 2009).

More recently, *mariner* transposon mutagenesis systems have been developed in which the transposase is active *in vivo*. In these systems, the transposase gene is placed onto a shuttle vector, and is expressed within the bacterium. The transposase is separated from the inverted terminal repeats (ITRs) that would normally flank it, and these are instead used to flank a selective marker. Such systems have been employed relatively recently in members of the genus *Clostridium* (Zhang et al., 2015). These are of particular interest to the current

study, due to the similarity between *E. limosum* and acetogenic clostridia such as *C. autoethanogenum*, which has itself been the subject of transposon mutagenesis concomitant with this study (C. Woods, personal communication).

An important consideration in plasmid-based transposon mutagenesis strategies that depend on *in vivo* activity of the transposase is that it should not be active during routine cloning. It is essential that the vector should not be mutagenising *E. coli* while it is being constructed. In a study by Zhang *et al*, in which transposon mutagenesis was exemplified in the solventogenic clostridia *C. acetobutylicum* and *Clostridium sporogenes*, the activity of the transposase was driven by the P_{tcdB} promoter of the *C. difficile* toxin B gene (*tcdB*). This promoter in turn is non-functional except in the presence of the *C. difficile* sigma factor TcdR. For mutagenesis of species other than *C. difficile*, this strategy relies on the insertion of the *tcdR* gene into the species targeted for mutagenesis. Thus, transposase activity is only possible in the target strain. In this case, the *tcdR* gene was inserted at the *pyrE* locus using ACE (Zhang et al., 2015).

To aid high-throughput mutagenesis strategies, it is useful to be able to cure the transposase-bearing plasmid following transposition. In the study of Zhang and coworkers (Zhang et al., 2015), the transposase-bearing vector possessed a conditional replicon which was unable to replicate in the presence of an inducer, isopropyl β-D-1-thiogalactopyranoside (IPTG). An alternative strategy was employed during mutagenesis of *C. perfringens*, in which the

mutagenesis vector possessed a galactokinase gene (*galK*) which had previously been deleted in the host strain. Expression of *galK* prevents the cell from growing in the presence of galactose, meaning that only transposon mutants that had lost the mutagenesis plasmid could grow on selective media supplemented with 3% w/v galactose (Liu et al., 2013). Both of these systems share the requirement for a host strain in which specific chromosomal modifications (deletion of *galK* for Liu *et al*, insertion of *tcdR* for Zhang *et al*) have previously been made. A transposon mutagenesis vector in which induction and plasmid loss mechanisms were both inherent to the vector would be more beneficial, allowing forward genetics in a wild-type background.

This requirement is satisfied by a tetracycline-inducible system that has previously been employed in transposon mutagenesis of *C. difficile* (Dembek et al., 2015). This system proved to be highly stringent, allowing the generation of large mutant pools. In the transposon vector used by Dembek *et al*, both transposase activity and plasmid loss are driven by anhydrotetracycline induction. This is due to a divergent promoter system that reads into the *Himar1* transposase, the *tetR* repressor protein, and the pCD6 Gram-positive replicon. Transcriptional read-through into the latter results in plasmid instability. This plasmid loss system is based on the observation that anhydrotetracycline induced plasmid instability in an inducible gene expression system for *C. difficile* (Fagan and Fairweather, 2011).

5.1.4 **The pMTL-CW22 Transposon Vector Series**

The present study makes use of two tetracycline-inducible transposon vectors derived from pMTL-CW22 (Craig Woods, SBRC Nottingham) (Figure 5.1). This vector carries the *Himar1 C9* transposase and *catP*-based mini-transposon previously employed in *C. pasteurianum* (Zhang et al., 2015), and possesses a divergent tetracycline-induced promoter system derived from pMTL-tet3nO (Walker and Koepke, 2016), and similar to that employed by Dembek *et al*, described above.

As the pCD6 replicon is non-functional in *E. limosum*, it was necessary to swap the Gram-positive replicon used (see Figure 5.3). Because the induced plasmid loss system demonstrated in *C. difficile* was based on transcriptional readthrough into the pCD6 replicon (Fagan and Fairweather, 2011), it was necessary to assess the effectiveness of this mechanism against the pCB102 and P19 replicons which were known to be functional in *E. limosum*.

To prevent spontaneous transposition during routine cloning, the transposable element in pMTL-CW22 is sited between the *codA* gene and an upstream promoter, preventing transcription (a terminator is present immediately downstream of the first ITR). Transposition of the mini-transposon results in expression of the *codA*. An active CodA acts as a counter-selective marker, converting 5-fluorocytosine (5-FC) into 5-fluorouracil (5-FU), which is toxic. It has been noted that even in the absence of 5-FC, assembly of an intact *codA* is deleterious to *E. coli*, and premature transposition within *E. coli* is therefore much reduced (C. Woods, personal communication). However, the primary

purpose of the *codA* is to counter-select premature transposition in the target organism (e.g. *C. autoethanogenum*) prior to plating, in order to maximise variability in a mutant population following induction of transposition.

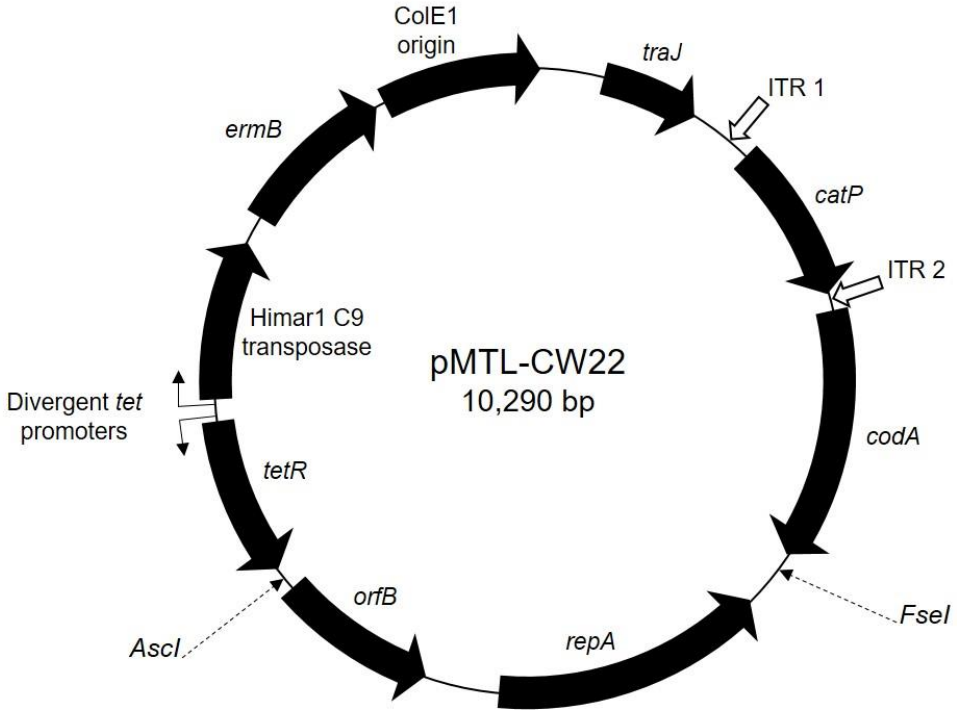


Figure 5.1: Plasmid map of pMTL-CW22. This is the vector from which the transposon vectors used in this study were derived (see Figure 5.9). Note divergent promoter system driving both the *Himar1 C9* transposase, *tetR*, and induced plasmid loss via transcriptional read-through into the Gram-positive replicon (*orfB*, *repA*). Antibiotic resistance is provided against clarithromycin/erythromycin via *ermB* and against thiamphenicol/chloramphenicol via *catP*. In both cases, the former antibiotics were used during selection of *E. limosum*, and the latter during selection of *E. coli* during cloning steps. Note that *ermB* is used to select for the pMTL-CW22 backbone specifically, whereas *catP* is located on a transposable element (selection on thiamphenicol will not distinguish between plasmid-harboring cells and spontaneous transposon mutants). ITR: Inverted Terminal Repeats, defining the ends of the *catP*-containing transposon. ColE1 and traJ: Gram-negative replicon. The cytosine deaminase gene (*codA*) provides counterselection against spontaneous transposition events, provided growth medium is supplemented with 5-fluorocytosine (5-FC).

5.1.5 **Anhydrotetracycline-inducible Gene Expression Systems**

Tetracycline-inducible systems have been employed for controlled gene expression in a wide variety of organisms, ranging from prokaryotes to mammalian cell systems (reviewed in Bertram and Hillen, 2008).

Such systems often employ TetR(B) from the Tn*10* transposon of *E. coli* (Hillen and Berens, 1994) (Chopra and Roberts, 2001), which is the best-characterised member of the TetR family (Grkovic et al., 2002). In this system, the *tetR* gene encoding the repressor protein is oriented divergently to the tetracycline (Tc) antiporter gene *tetA*. TetR forms a homodimer, and in the absence of (anhydro-) tetracycline, it binds to *tet* operator (*tetO*) sites in the three Tet-responsive promoters found within the *tetR-tetA* intergenic region (see

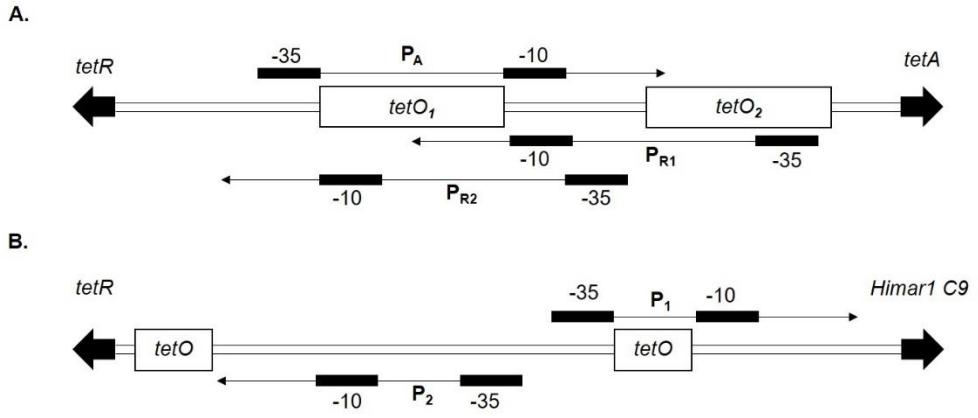


Figure **5.2**, below). In the presence of (a)Tc, TetR is unable to bind *tetO* sites, allowing expression of TetA and conferring Tc resistance (Bertram and Hillen, 2008). This system is required to be highly sensitive; TetA is harmful to the cell in the absence of Tc, but must be expressed before intracellular levels of

Tc are sufficient to halt protein synthesis (reviewed in Berens and Hillen, 2003).

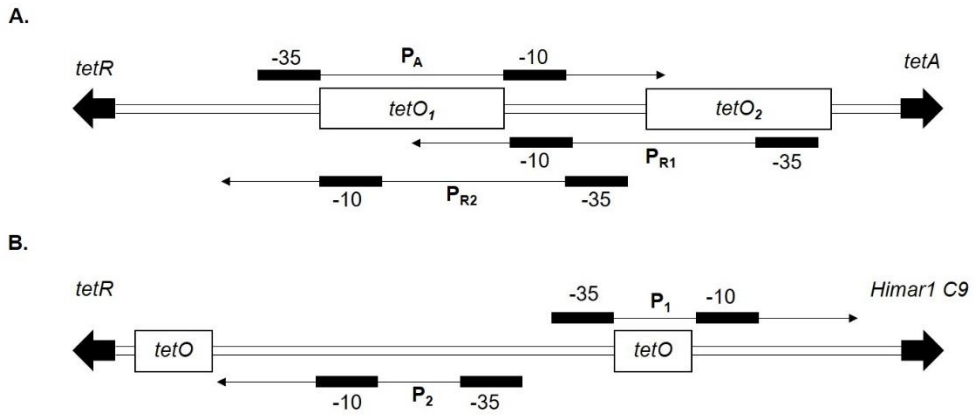


Figure 5.2: Comparison of tetracycline-inducible systems. Note that these systems are not shown to scale with one another. **A:** *tetR-tetA* system of Tn10 (image adapted from Bertram & Hillen, 2008). **B:** Divergent promoter system from pMTL-CW22 transposon mutagenesis vectors employed in this study. **TetO:** *tet* operator sequences, to which TetR binds.

When induced gene expression systems based on the Tet system are assembled, they are often induced with anhydrotetracycline. This is advantageous because aTc is less toxic than Tc, and is bound more readily by TetR (Gossen and Bujard, 1993). Both Tc and aTc are able to cross membranes passively by diffusion, not being limited to transport through dedicated channels (Ehrt et al., 2005). Tetracyclines are able to penetrate both prokaryotic and eukaryotic cell membranes by diffusion, which has been exploited by artificial gene induction systems which are effective on bacteria growing *in vivo* in mammalian cells (Obrist and Miller, 2012). The stringency

of the Tet system is due to the fact that TetR has a high affinity both for (a)Tc and for *tetO*. Affinity for *tetO*, rather than non-specific binding to other genomic targets, maximises TetR's occupancy of *tetO* during Tc- conditions (Bertram and Hillen, 2008).

In the *tetR-tetA* system found in the *E. coli* Tn*10* transposon, expression of TetR is self-regulating. Because expression of *tetA* is deleterious to the cell, it is not surprising that TetR binds to $tetO_2$ (blocking transcription of *tetA*, but not *tetR*), with roughly twice the efficiency of its binding to $tetO_1$ (reviewed in Grkovic et al., 2002) (refer to

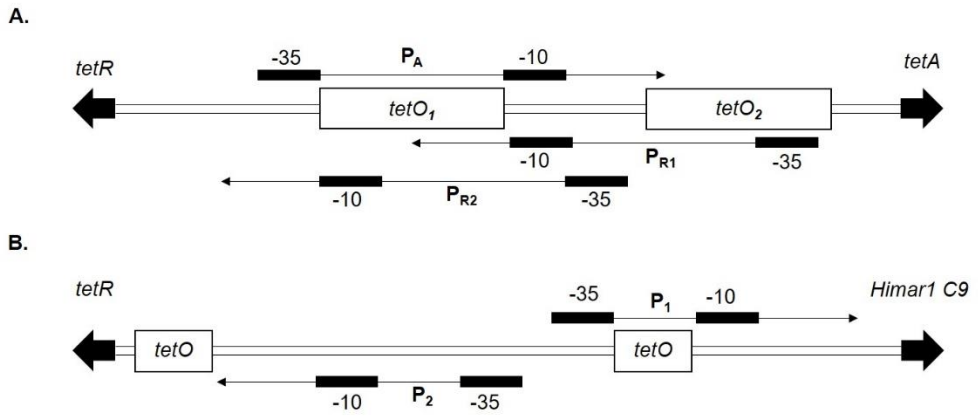


Figure **5.2**, above). Greater affinity of TetR for $tetO_2$ also means that $tetO_2$ is more likely than $tetO_1$ to be occupied even if intracellular TetR is diluted, for example by cell division. This contributes to tight regulation of TetA. Insertion of *tetO* sites is a method by which Tet-inducible promoters can be created from promoters native to a species of interest (Chubiz et al., 2013) (McClain et al., 2013) (Li and Liao, 2015).

The stringency of a Tet-sensitive gene expression system has been demonstrated in *C. acetobutylicum*. An aTc-driven CatP expression system

analogous to that tested in this study was found to require continuous induction of CatP to confer chloramphenicol resistance (Dong et al., 2012). Additionally, the system was found to require induction prior to chloramphenicol exposure; cells bearing the reporter plasmid were unable to grow on plates containing both aTc and chloramphenicol if they had not been exposed to aTc beforehand. The Tet system utilised in this work, Tet3nO, has been employed in the acetogen *C. autoethanogenum* (Walker and Koepke, 2016).

5.2 Results

5.2.1 Generation of Transposon Mutagenesis Vector pMTL-CW21

An overview of the replicative transposon mutagenesis vectors employed in this study can be seen below (Figure 5.9). The transposon mutagenesis vector pMTL-CW22 was obtained from C. Woods (Synthetic Biology Research Centre, Nottingham, UK). The Gram-positive replicon of pMTL-CW22 was removed via digestion with restriction endonucleases AscI and FseI (Figure 5.3, below). The Gram-positive replicon pCB102 of pMTL83151 was removed via digestion with the same restriction endonucleases. The pMTL-CW22 linear backbone was purified from the digest via gel, and de-phosphorylated. The pMTL-CW22 backbone was then ligated to the pCB102 Gram-positive replicon fragment, and the ligation reaction transformed into chemically-competent *E. coli* 5-alpha® (NEB). Transformant cells were plated onto LB agar supplemented with both erythromycin and chloramphenicol. Putative transformant colonies were cultured overnight in LB under the same selective conditions. Plasmid DNA was extracted via miniprep, and a diagnostic digest was performed with EcoRV to validate the prepared plasmid DNA (Figure 5.4).

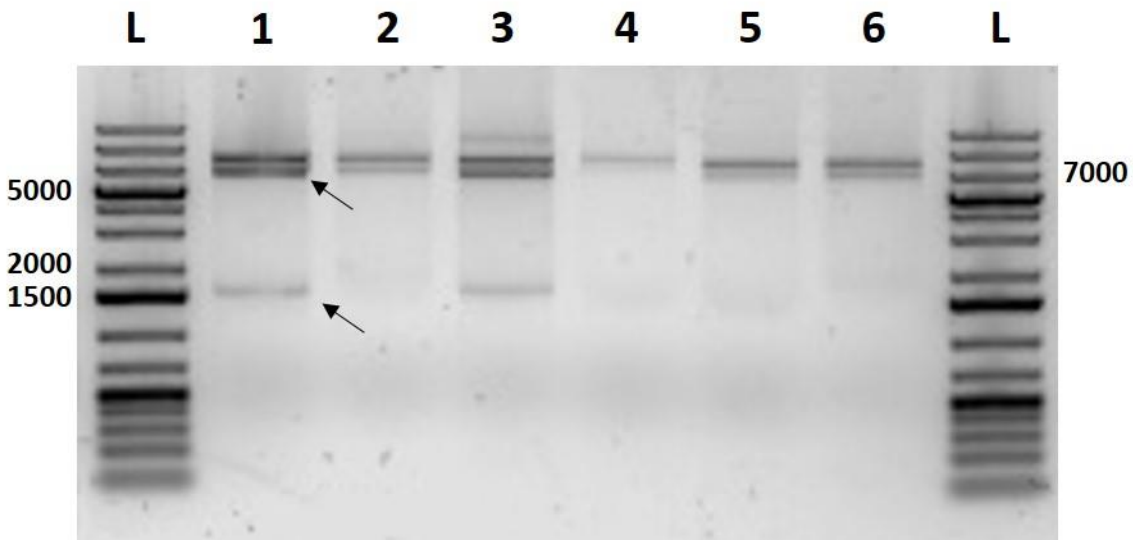


Figure 5.3: pMTL-CW21 diagnostic digest. Six putative clones were digested with AscI/FseI to separate the plasmid backbone and the pCB102 Gram-positive replicon. Arrows in lane 1 indicate the 1625 bp replicon fragment and the 6875 bp plasmid backbone. Undigested plasmid backbone can be seen above the backbone fragment. L = ladder.

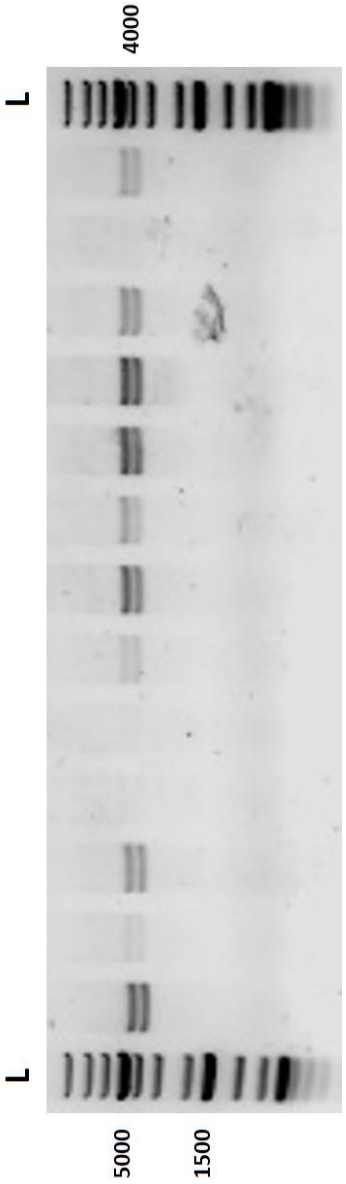


Figure 5.4: pMTL-CW21 diagnostic digest with EcoRV. Because EcoRV sites are found within both the transposon and the plasmid backbone, they can be used to assess whether the transposon has spontaneously transposed during cloning. In this example, all visible lanes show correct fragments (4677 and 3815 bp, respectively). L = ladder.

5.2.2 Generation of Transposon Mutagenesis Vector pMTL-CW23

Vector pMTL-CW23 was generated in the same way described for pMTL-CW21 above, except that the Gram-positive P19 replicon from pMTL87151 was used instead of pCB102 from pMTL83151. As with pMTL-CW21, pMTL-CW23 was validated via *EcoRV* digestion.

5.2.3 aTc Growth Impact in *E. limosum*

Prior to conducting a chloramphenicol acetyltransferase (CatP) assay using an aTc-inducible reporter plasmid, a toxicity test was undertaken (see Figure 5.5, below) to determine inhibitory and/or lethal concentrations of anhydrotetracycline (aTc). Concentrations of less than 1000 ng/ml were found to have little impact on the growth of *E. limosum*. The highest concentration tested (1000 ng/ml) was not lethal, only delaying growth. A concentration of 100 ng/ml was used for further growth experiments and for induction of transposition and plasmid loss following transformation of competent *E. limosum* with transposon vectors pMTL-CW21 and pMTL-CW23.

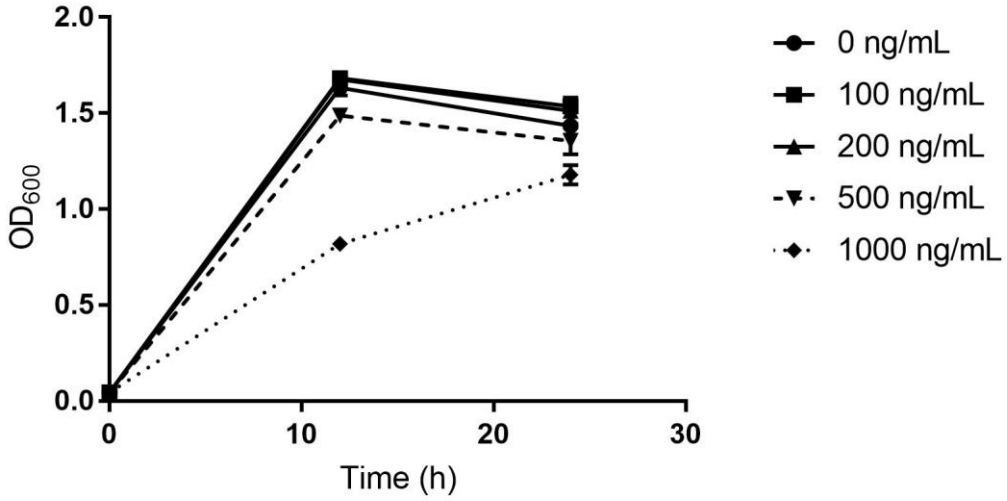


Figure 5.5: Growth of *E. limosum* in medium supplemented with varying concentrations of anhydrotetracycline (aTc). Points represent the mean averages of three biological replicates. Error bars represent the standard deviation.

5.2.4 Construction of Reporter Plasmid pMTL-JM102

The TetR reporter plasmid pMTL-Tet3nO was obtained from the culture collection of the Synthetic Biology Research Centre (Nottingham, UK). The vector was digested with FseI and AscI (NEB) to remove the Gram-positive replicon. The linear plasmid backbone was then de-phosphorylated with Antarctic Phosphatase (NEB) and gel purified. The plasmid backbone was then ligated to the Gram-positive replicon pCB102, prepared from pMTL83151 via digestion with AscI and FseI and gel purified. The ligation mixture was transformed into *E. coli* 5-alpha (NEB) and transformant colonies selected on media supplemented with erythromycin. Putative transformant colonies were grown overnight in LB supplemented with erythromycin, and

plasmid DNA obtained by miniprep. Prepared plasmid DNA was validated via restriction digest with AscI and FseI.

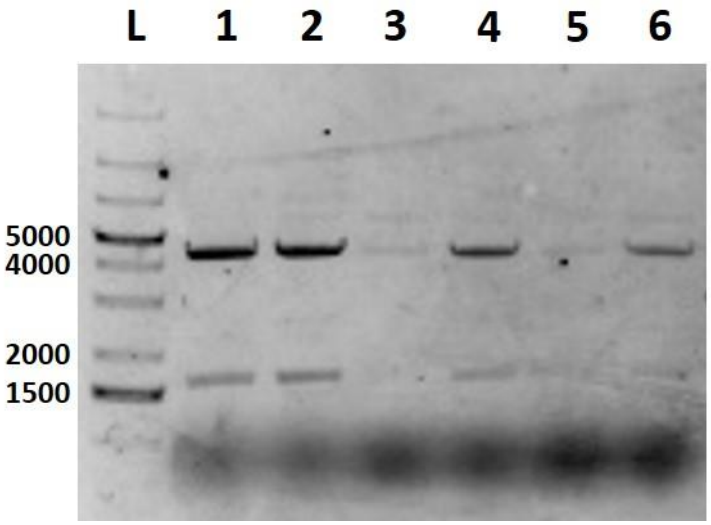


Figure 5.6: pMTL-JM102 diagnostic digest. Six clones (lanes 1-6) were digested with AscI/FseI, with successful clones separating into fragments of 1625 and 4402 bp in length. These represent the Gram-positive replicon and plasmid backbone, respectively. L = ladder.

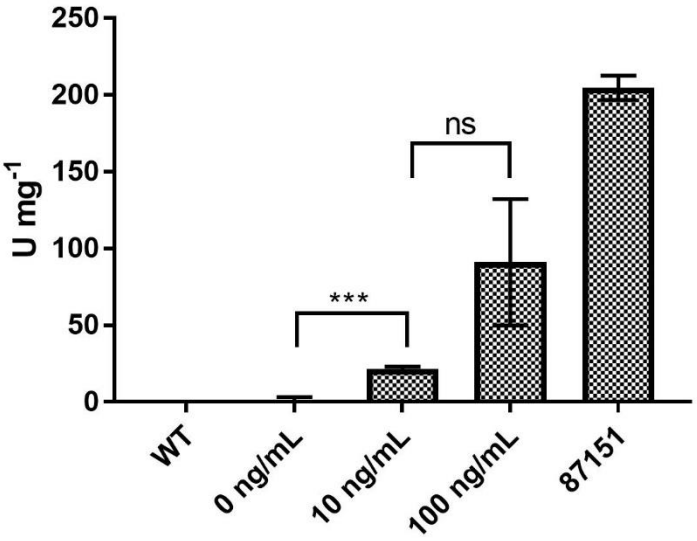
5.2.5 CatP Reporter Assay in *E. limosum*

Numerous studies involving aTc-induced reporter systems have been undertaken. Reporter systems employed include GFP in *Mycobacterium smegmatis* and *M. tuberculosis* (Ehrt et al., 2005) and GusA in *C. acetobutylicum* (Dong et al., 2012). The study involving *M. smegmatis* and *M. tuberculosis* noted maximum reporter gene activity occurring at 5-10% minimal inhibitory concentration of aTc (Ehrt et al., 2005), with significant GFP activity occurring within an hour of aTc induction. The *C. acetobutylicum* study only assessed induction after three hours of incubation.

A CatP reporter assay was conducted to determine the sensitivity and stringency of Tet-induced systems in *E. limosum.* This experiment utilised the *catP* reporter plasmid pMTL-JM102 (see section 5.2.4, above). This is based on the CatP reporter construct pMTL-Tet3nO (Walker and Koepke, 2016), with the Gram-positive origin of replication exchanged for pCB102 to allow replication within *E. limosum*. The Tet-inducible cassette in pMTL-JM102 is identical to that found in pMTL-CW22 and all vectors derived from it.

To provide a positive control, protein was extracted from *E. limosum* harbouring plasmid pMTL87151, containing a constitutively-active *catP*, used as a selection marker to confer thiamphenicol resistance. Negative controls were provided by *E. limosum* wild-type cells and by *E. limosum* harbouring pMTL-JM102 in the absence of anhydrotetracycline. This served to indicate whether *E. limosum* protein extracts have any basal chloramphenicol acetyltransferase activity, and to determine whether any spontaneous activity of the reporter construct was possible in the absence of induction.

An initial experiment was performed in which *E. limosum* starter cultures carrying pMTL-JM102 were inoculated into medium supplemented with 0, 10



or 100 ng/ml aTc (

Figure **5.7**). This concentration was selected because it was the maximum concentration previously observed to not cause growth inhibition (Figure 5.5). Samples were grown under induction conditions until they reached an OD_{600} of 0.2, and 5 ml samples were subsequently collected. This low harvesting OD was selected to preclude problems with mucoidal material inhibiting centrifugation and protein extraction. CatP activity was determined spectrophotometrically using a Tecan M1000 Pro plate reader, in accordance with the protocol described in Materials and Methods (section 2.11.1). Readings were obtained in triplicate for each biological replicate. Little to no CatP activity was observed in the negative controls, which consisted of wild-type *E. limosum* and un-induced *E. limosum* pMTL-JM102. Significant CatP activity (approximately 200 U/mg) was observed in the positive control, a sample derived from *E. limosum* transformed with pMTL87151.

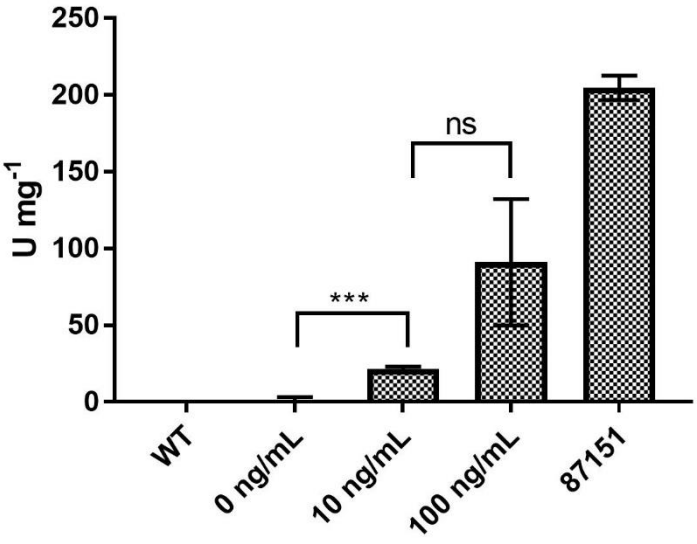


Figure 5.7: Preliminary CatP activity assay in *E. limosum.* Wild-type (WT; negative control), *E. limosum* pMTL87151 (positive control), and *E. limosum* harbouring TetR reporter plasmid pMTL-JM102 in the presence of 0, 10 or 100 ng/ml anhydrotetracycline (aTc). Samples for 10 and 100 ng/ml were prepared in biological duplicate. Column heights represent average enzyme activities of three technical replicates. Error bars represent the standard deviation. Differences between sample activities were assessed using Welch's T test with two-tailed P values (0.0004 and 0.0992, respectively).

Following this initial experiment, which confirmed the functionality of the TetR system in *E. limosum*, a further experiment was conducted to characterise the time- and dose-dependency of the system. This would indicate whether there was a minimum or maximum effective dose for aTc induction, and whether full induction of the system was apparent within an hour of induction, or whether it was slower to develop.

The subsequent CatP assay was conducted with a lengthier pre-culture step prior to induction (Figure 5.8). In this case, an *E. limosum* preculture

inoculated from a pool of several hundred clones bearing the reporter plasmid was grown to an OD_{600} of approximately 0.5. The preculture was then aliquoted into four daughter cultures, each of which was exposed to different concentrations of aTc (0, 10, 50 and 100 ng/ml). Cells were harvested in 1 ml volumes at 1, 3, and 5 hours after induction. To aid collection of cell material, harvested cultures were mixed 1:1 with 1.5 M NaCl prior to centrifugation. CatP activity of the resulting protein extracts was examined and found to be time- and dose-dependent (Figure 5.8). In contrast to the preliminary experiment, little activity was observed during induction with 10 ng/ml aTc, with no activity observed at 1 hour, 2.8 ± 1.6 U/mg at 3 hours, and 8.1 ± 0.5 U/mg observed at 5 hours. No measurable activity was observed in the un-induced cultures, except at the 5 hour time point, at which 6.3 ± 0.3 U/mg was detected. This may represent noise in the assay, or a basal level of *catP* expression occurring in the absence of aTc. This is consistent with the overnight appearance of single colonies on the aTc-negative selection plates inoculated from cultures carrying the catP reporter plasmid pMTL-JM102 in Figure 5.13, below.

Significant induction of *catP* was observed at all time points during induction at 50 and 100 ng/ml aTc. A significant increase in CatP activity was observed between the 1- and 3-hour time points for both of these higher concentrations, indicating a development of *catP* induction during this period. At 50 ng/ml, a 2.1–fold increase was observed (from 61.7 ± 11 U/mg to 131 ± 9.2 U/mg). At 100 ng/ml, a 3-fold increase was observed (from 75.7 ± 2.2 U/mg to 226 ± 17.2 U/mg). CatP activity was observed to decline at both of these induction

concentrations between the 3- and 5-hour time points. There is a significant increase in induction at 100 ng/ml aTc relative to 50 ng/ml.

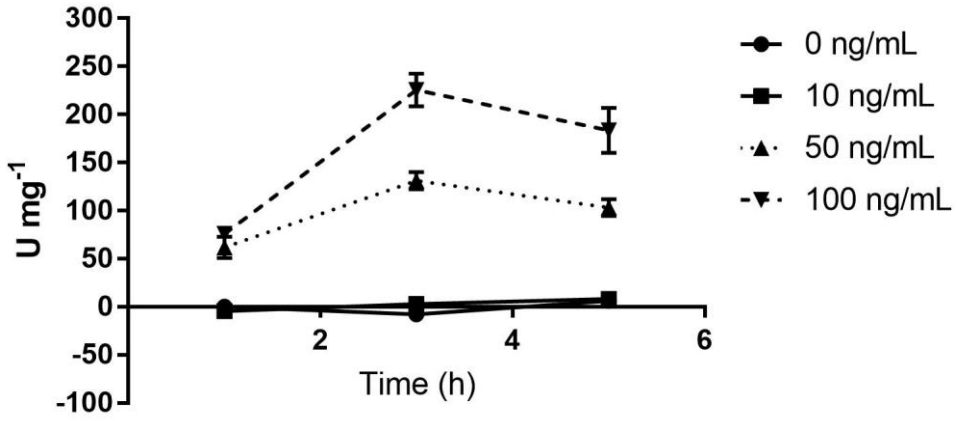


Figure 5.8: CatP time-course assay. CatP activity expressed in units per mg protein (U/mg) extracted from *E. limosum* cultures harbouring the TetR reporter plasmid pMTL-JM102. Data points are the mean averages of three biological replicates. Error bars represent the standard deviation. Data series indicate different concentrations of anhydrotetracycline (aTc).

This assay illustrates the functionality of the aTc-induced gene expression system in *E. limosum*. Induction of *catP* is seen to develop over the course of the assay, and a decline in CatP activity is seen to occur at some time between 3 and 5 hours post-induction. Unlike the preliminary experiment, this CatP assay shows little to no activity during induction with 10 ng/ml aTc.

5.2.6 **Transposon Mutagenesis Vectors pMTL-CW21 and –CW23**

As noted above, the constructs used for transposon mutagenesis in this study are based on vector pMTL-CW22, which was developed for use in *Clostridium autoethanogenum*. Vector pMTL-CW22 and the derivatives employed in this study differ only in their Gram-positive replicons. While pMTL-CW22 replicates via a *C. difficile*-derived replicon, pMTL-CW21 and –CW23 rely on the pCB102 and P19 vectors respectively. The characteristics of these replicons in *E. limosum* have been detailed previously in this study (see Chapter 4, above). These Gram-positive replicons were selected for this work in order to investigate any differences in induced plasmid loss between the two. Because the P19 replicon, as found on pMTL87151, has been observed to be extremely stable in the absence of selective pressure, its interaction with the induced plasmid loss mechanism on pMTL-CW22 was a subject of interest. Additionally, significant differences in the copy numbers of the pCB102 and P19 replicons had previously been observed in this study, and it was thought possible that this may result in different behaviours in the context of a transposon vector.

The pMTL-CW22 vector series are similar in concept to a universal *mariner* transposon system exemplified in solventogenic clostridia previously (Zhang et al., 2015), except for their tet-inducible system, which is based on one previously employed in *C. autoethanogenum* (Walker and Koepke, 2016). The main advantage of this system is that it does not require prior modification of

the target organism, and can therefore be used for transposon mutagenesis in a wild-type background.

The pMTL-CW22 vector was designed and constructed in accordance with the pMTL80000 modular shuttle vector system for use in clostridia. Accordingly, the Gram-positive replicon is flanked by AscI and FseI sites. To adapt pMTL-CW22 for use in *E. limosum*, it was necessary to swap out the pre-existing Gram-positive replicon (the *C. difficile*-derived pCD6) with either pCB102 or P19, both having been demonstrated to be functional within *E. limosum*. This yielded the constructs pMTL-CW21 and pMTL-CW23, respectively (see Figure 5.9 below, and sections 5.2.1 and 5.2.2 above).

To verify that the transposon remained in place within the new transposon vectors, EcoRV digests were performed. On the pMTL-CW22 plasmid backbone, EcoRV sites are present within both the *Himar1C9* transposase gene and the *catP* within the mini-transposon.

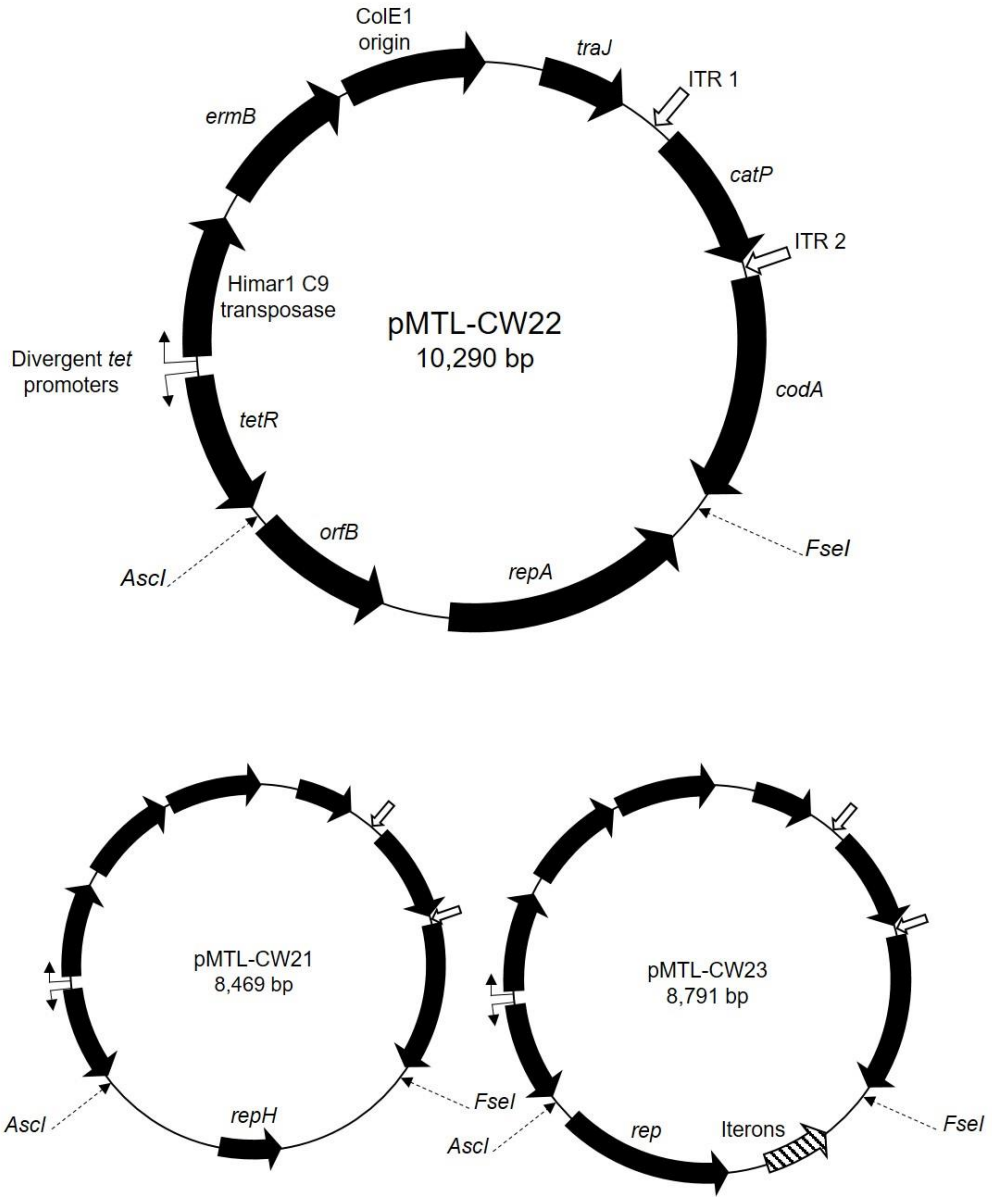


Figure 5.9: Transposon mutagenesis vectors employed in this study. All vectors are identical except for their Gram-positive replicons. Gram-positive replicons are flanked in all vectors by AscI/FseI restriction sites. The original vector, pMTL-CW22, bears the pCD6 Gram-positive replicon. Vectors pMTL-CW21 and pMTL-CW23 bear the replicons pCB102 and P19, respectively. **ITR**: Inverted Terminal Repeats which define the ends of the *catP*-bearing mini-transposon. ***ErmB*:** Selective marker conferring resistance to erythromycin and clarithromycin. ***CatP:*** Selective marker conferring resistance to thiamphenicol.

5.2.7 **Creation of Transposon Mutant Libraries in *E. limosum***

The creation of transposon mutagenesis vectors pMTL-CW21 and pMTL-CW23 is described above (sections 5.2.1 and 5.2.2). Following cloning, these plasmids were transferred into *E. limosum* via electroporation. Transformant colonies were selected on clarithromycin supplemented with 5-fluorocytosine (see below) in the absence of anhydrotetracycline, to avoid selection in favour of spontaneous mutants.

During transformation of pMTL-CW21 and pMTL-CW23 into *E. limosum*, greatly reduced DNA transfer efficiencies were observed for both plasmids, with 10^3-10^4 transformants per ug DNA, resulting in 10^2-10^3 colonies observed per transformation event.

This is in contrast to the DNA transfer efficiencies of 10^5-10^6 transformants per µg observed during transformation of smaller vectors such as pMTL83151 and pMTL87151. This may be attributable to the greater size of the transposon vectors versus the smaller shuttle vectors; pMTL-CW21 and –CW23 are 8.5 and 8.8 Kb in length, in comparison to pMTL83151 and pMTL87151, which are 4.5 and 4.8 Kb in length, respectively. Alternatively, reduced transfer efficiencies may be due to the presence of restriction sites on the larger plasmids which are cleaved by *E. limosum*'s native restriction systems. If transposase activity is stringently regulated and no early transposition events occur, then reduced DNA transfer is irrelevant to the final size of a transposon mutant library, because un-induced cells can be grown to the desired OD, and then induced to generate insertional mutant libraries of arbitrary size.

Conversely, the more active the transposase is in the absence of induction, the more the initial population of transformant cells will be dominated by early mutants.

5.2.8 Efficiency of Transposase Induction and Plasmid Loss

To provide efficiency figures for aTc-induced transposition and plasmid loss, collected transformant libraries of pMTL-CW21 and pMTL-CW23 were plated on various selective media. These were (1) agar supplemented with clarithromycin (to assess plasmid-bearing CFU, which represents the entire collected population), (2) thiamphenicol plus aTc (which should represent a pure insertional mutant population, bearing the *catP* selective marker in their chromosomes and unable to replicate the transposon vector), and (3) clarithromycin plus aTc (to indicate possible ineffectiveness of the plasmid loss mechanism; in case of 100% efficient plasmid loss, this would be an entirely lethal condition barring integration of the transposon vector or mutation in the *tetR* cassette)(Figure 5.10).

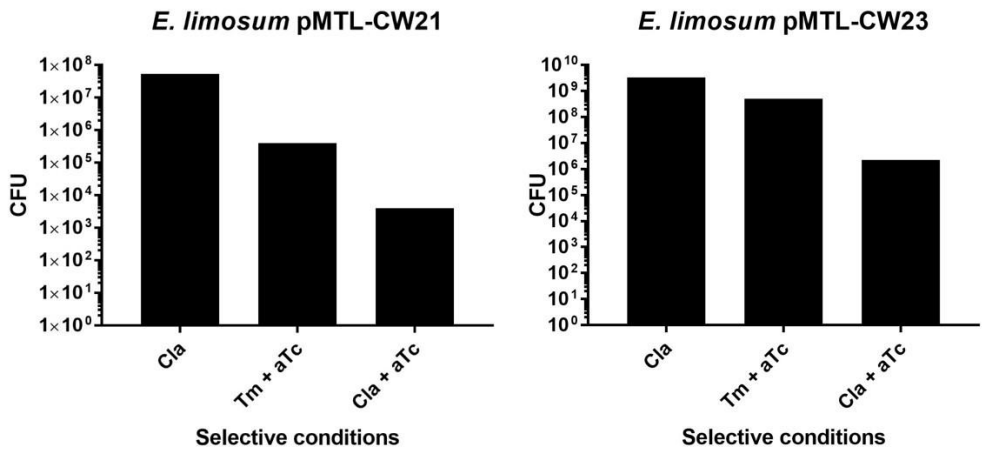


Figure 5.10: Transposase activity and plasmid loss efficiency in *E. limosum* transformant pools harboring pMTL-CW21 and pMTL-CW23. **Cla:** Clarithromycin selection, representing total transformant pool. **Tm + aTc:** Cell population following induction of transposase activity and induction of plasmid loss. **Cla + aTc:** Cell population following induction of transposase and plasmid loss, with selection for plasmid backbone. **Cla:** Clarithromycin (selection for plasmid backbone); **Tm:** Thiamphenicol (selection for mini-transposon).

The percentages of CFU observed in these media conditions should represent the following: The percentage of Tm+aTc colonies relative to the total Cla^R population should indicate transposition frequency, with the missing CFU in the Tm+aTc condition representing those cells which failed to transpose successfully prior to plasmid loss. The percentage of Cla+aTc colonies should be a proxy for the effectiveness of the plasmid loss mechanism, with CFU on the Cla+aTc plates representing cells which have resisted the induction of plasmid loss and have therefore retained their Cla^R phenotype. However, other factors can contribute to these CFU counts. For example, spontaneous integration of the transposon vector into the chromosome may account for some of the $Cla+aTc^R$ population. Additionally, the growth behaviour of cells

harboring these plasmids under the influence of aTc suggests that these figures should be interpreted cautiously (see section 5.2.9, below).

Figure 5.10 reveals differences between the transposon vectors. The apparent plasmid loss rate for pMTL-CW21 is higher by an order of magnitude than that for pMTL-CW23 (99.995% vs 99.94%). The apparent effectiveness of transposition is higher in pMTL-CW23 than in pCB102, by roughly a factor of 20 (14.45% versus 0.75%) (Refer to Figure 5.10, above).

The differences observed between the total transformant CFU (Cla^R) and the Tm+aTc and Cla+aTc conditions indicate that aTc induction has a profound effect on these transformant populations. This suggests that the aTc-induced transcriptional readthrough into the Gram-positive replicon P19 is indeed capable of causing plasmid instability, as has previously been demonstrated for pCD6 (Fagan and Fairweather, 2011) and pCB102 (Zhang et al., 2015) previously. While induction of plasmid loss by aTc appears to abolish clarithromycin resistance in the majority of cases (Figure 5.10), it is not completely effective. Inverse PCR performed on colonies grown on clarithromycin and aTc revealed at least 6/10 still possessed the mutagenesis vector (Figure 5.11). As has previously been observed in *mariner* transposon mutagenesis, the inverse PCR products depicted in Figure 5.11 all appear to be single insertions.

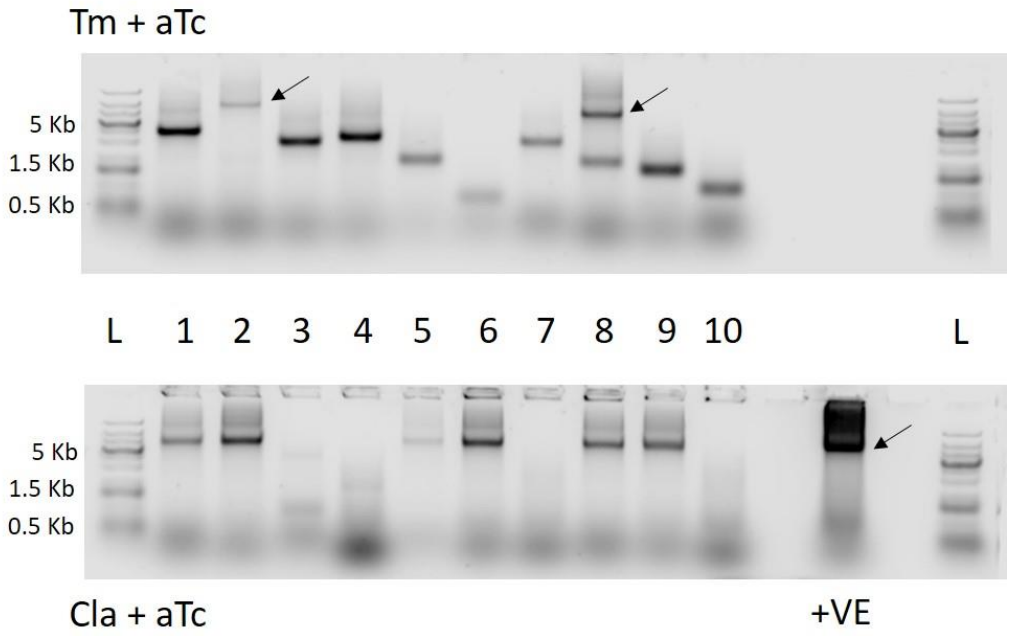


Figure 5.11: Inverse PCR products from mutants grown on Tm + aTc (top gel) or Cla + aTc (bottom gel). Numbers indicate lanes; identically numbered clones are not related across gels. **+VE:** Inverse PCR product amplified from transposon mutagenesis vector pMTL-CW23. The Cla + aTc clones are dominated by products similar in size to the positive control. In contrast, a variety of product sizes are observed in the Tm + aTc clones, and plasmid loss can be observed, although lanes 2 and 8 contain vector-sized bands. Thin black arrows indicate PCR products corresponding in size to that generated by the positive control. L = ladder.

While greatly decreased CFU in Tm+aTc conditions would appear to indicate that the plasmid loss system in pMTL-CW21/-CW23 is highly effective, this is predicated on plasmid loss being a benign process, which host cells survive without injury. To illustrate the effect of aTc on the growth of *E. limosum* cultures bearing pMTL-CW21 and –CW23, growth experiments were conducted (see below).

5.2.9 **Differentiating between Induced Plasmid Loss and Cell Death**

An experiment was conducted to determine whether the CFU observed in Tm+aTc conditions were indicative of true plasmid loss, or whether they were influenced by cell death occurring during plasmid loss. The resulting growth experiment conducted is illustrated in Figure 5.12, below. *E. limosum* overnight precultures harbouring pMTL-CW21 and –CW23 (both of OD_{600} 1.6) were subcultured into non-selective liquid medium (with a 1% v/v inoculum) and grown to OD_{600}s of 0.1-0.2 before being divided into aTc +/- conditions. An aTc concentration of 100 ng/ml was used, as this had previously been observed to cause no inhibition of growth (see Figure 5.5, above).

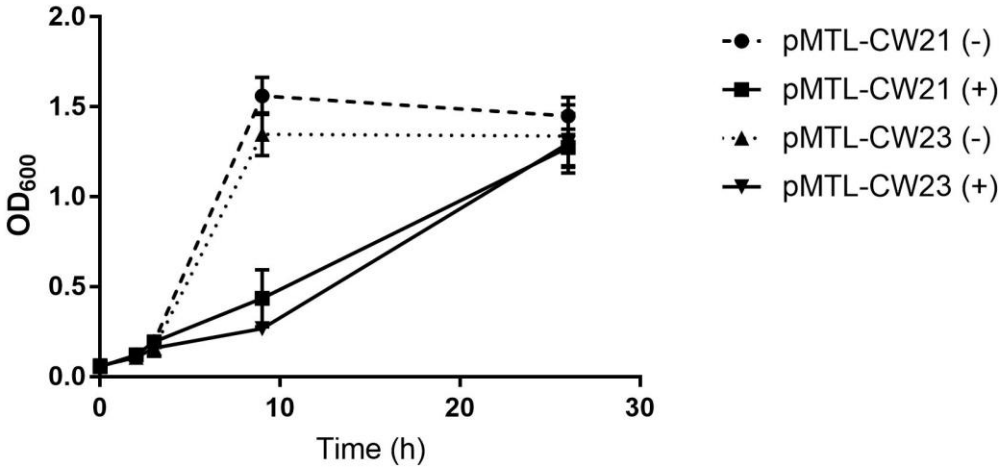


Figure 5.12: Growth of *E. limosum* harbouring transposon vectors during aTc-induced plasmid loss. **pMTL-CW21 (+):** Growth of *E. limosum* pMTL-CW21 in the presence of 100 ng/ml aTc; **pMTL-CW21 (-):** Growth of sibling culture in aTc-free conditions; **pMTL-CW23 (+):** Growth of *E. limosum* pMTL-CW23 in the presence of aTc, as above; **pMTL-CW23 (-):** Growth of sibling culture in aTc-free conditions. Data points indicate the mean average of three replicates. Error bars represent the standard deviation. Not all error bars are visible at the scale used.

During the growth experiment, a significant impact on growth of *E. limosum* harbouring the transposon vectors was observed in the aTc-positive cultures. Their biological replicates in the aTc-free cultures grew without inhibition. This delay in growth is greater than that observed during growth of the wild-type in the presence of 1000 ng/ml aTc (Figure 5.5). This suggests that aTc-induction, instead of resulting in a benign shutdown of plasmid replication, instead has a traumatic effect on the host cells. This in turn is of interest where the generation of insertional mutant libraries is concerned. If the growth of plasmid-bearing clones is retarded until the plasmid is lost, a clonal population under aTc induction may become dominated by clones which have serendipitously lost the plasmid early. The relative proportion of cells which have spontaneously lost the plasmid will rise as those still possessing the vector are killed. This early domination of mutants may explain the apparent domination of early transposition events noted during inverse PCR studies (below).

5.2.10 Plate Testing of the CatP Reporter

E. limosum cultures bearing the CatP reporter plasmid pMTL-JM102 were grown in selective media (supplemented with clarithromycin at 6 µg/ml) with and without 100 ng/ml aTc and plated to selective media (supplemented with thiamphenicol at 7.5 µg/ml) with and without aTc. The purpose of this was to provide a preliminary indication that aTc-induced expression of the *catP* gene was functioning as desired, and to indicate whether a sustained induction of

catP was necessary to provide resistance, or whether a temporary induction would provide sufficient CatP to allow growth in selective media. As a negative control, un-induced culture was also plated to thiamphenicol selective media. Patch growth was observed, and indicated that induction of CatP via aTc was effective on-plate, not requiring pre-induction in liquid media (Figure 5.13). Growth of pre-induced culture on a plate lacking aTc demonstrates that sufficient induction of CatP occurs such that growth can occur for a time even in the absence of ongoing induction. Single colonies were observed on the negative control, indicating a degree of CatP activity even in the absence of induction. Note that thiamphenicol-resistant colonies are not typically observed when conducting transformations of *E. limosum* in which much greater concentrations of cells are plated, so these are unlikely to be spontaneous mutants. This indicates a degree of "leakiness" in the Tet system, demonstrating that in a minority of cells, *catP* is transcribed despite repression by TetR.

A. **B**.

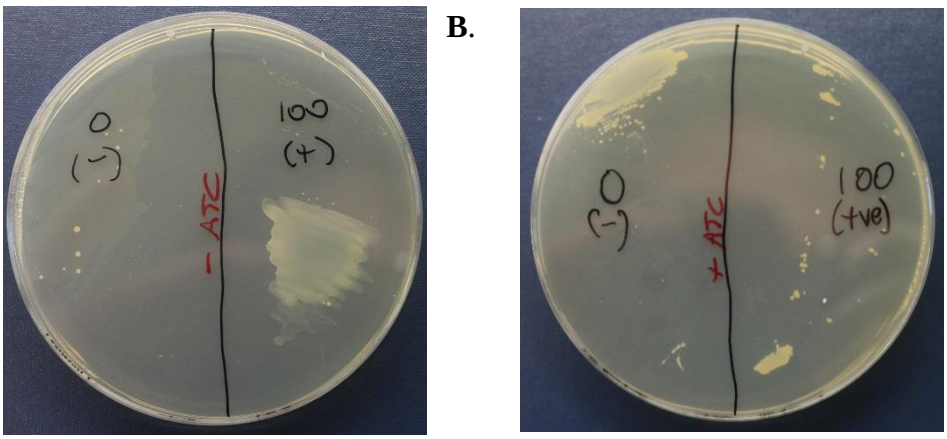


Figure 5.13: Plate test of the CatP reporter. *E. limosum* with TetR reporter plasmid pMTL-JM102 was plated onto agar supplemented with thiamphenicol (7.5 ug/ml) ± aTc (100 ng/ml) with or without prior induction of CatP (chloramphenicol acetyltransferase). Note that thiamphenicol resistance will not be conferred without CatP activity. **A**: *E. limosum* with reporter plasmid plated onto thiamphenicol plate which has not been supplemented with aTc. 0 (-): No aTc in preculture; 100 (+): 100 ng/ml aTc in preculture. **B**: *E. limosum* with reporter plasmid plated onto thiamphenicol plate supplemented with aTc (100 ng/ml). 0 (-): No aTc in preculture; 100 (+): 100 ng/ml aTc in preculture.

5.2.11 **Determining the Spontaneity of Transposition**

The pMTL-CW22 transposon vector series contains a safeguard against spontaneous transposase activity in the form of a cytosine deaminase (*codA*) 'dead man's switch', in which transcription of *codA* is blocked by the mini-transposon on the vector, and allowed following transposition, leading to expression of CodA. Transposition results in reconstruction of the *codA* promoter. This, in turn, is lethal in the presence of 5-fluorocytosine (5-FC), which it converts to toxic 5-fluorouracil (5-FU). In previous plasmid-based strategies for transposon mutagenesis of clostridia, spontaneous transposition

has been prevented by coupling expression of the transposase to an orthogonal transcription factor expressed in the clostridial host (Zhang et al., 2015). During this study, inverse PCR was conducted to assess the rate of spontaneous transposition with and without 5-FC.

Following initial transformation of pMTL-CW21, a single transformant colony (Cla^R, no 5-FC) was resuspended in liquid medium without selection or aTc, and a dilution series plated to Tm+aTc to induce transposition and plasmid loss. Inverse PCR was conducted on 20 individual clones as described in Materials and Methods. Sequencing was successful for 13 of these clones, and after mapping of their insertion sites, they were found to represent three mutant lineages (see section 8.2,

Table **5.1**). This provided an early indication of spontaneous transposase activity in the absence of aTc. The basal activity of the divergent *Tet* promoters is apparently sufficient to allow the early activity of the transposase, while not being sufficient to cause noticable plasmid instability in the absence of aTc. A list of all transposon mutants sequenced during this study can be found in the Appendix (see section 8.2).

Table 5.1: Insertion sites for initial inverse PCR sequencing of transposon mutants. All mutants were derived from a single Cla^R colony, which was resuspended and plated to Tm+aTc to isolate individual transposon mutants.

Clone	Locus	BLAST
1	Ga0111251_111929	DNA Polymerase III Subunit Epsilon
2	Ga0111251_111929	DNA Polymerase III Subunit Epsilon
3	Ga0111251_111929	DNA Polymerase III Subunit Epsilon
4	*rnr* gene	Ribonuclease R
5	Ga0111251_112082	Hypothetical Protein [*Eubacterium limosum*]
6	Ga0111251_111929	DNA Polymerase III Subunit Epsilon
7	Ga0111251_111929	DNA Polymerase III Subunit Epsilon
8	Ga0111251_112082	Hypothetical Protein [*Eubacterium limosum*]
9	Ga0111251_112082	Hypothetical Protein [*Eubacterium limosum*]
10	Ga0111251_112082	Hypothetical Protein [*Eubacterium limosum*]
11	*rnr* gene	Ribonuclease R
12	Ga0111251_111929	DNA Polymerase III Subunit Epsilon
13	Ga0111251_112082	Hypothetical Protein [*Eubacterium limosum*]

To confirm this observation, three additional clones were picked from the initial transformation plate, resuspended, and plated onto Tm+aTc to isolate individual colonies. The diversity of mutants represented by each colony was again assessed via sequencing of inverse PCR products for five descendent colonies of each parent

Table **5.2**). Sequencing revealed that one colony was entirely dominated by one mutant, indicating a very early transposition event in the parent cell (prior to division, or prior to division of a daughter cell containing the transposon vector). A second colony was found to contain at least two mutants, though one was overwhelming represented in sequencing (4/5). The third, in contrast, appeared to be much more diverse, with every sequenced clone being a unique mutant. These results, taken together with the lack of diversity shown in

Table **5.1**, suggest that spontaneous activity of the transposase is significant. In turn, this implies that the collections of transformants generated purely by selection on clarithromycin can only be expected to generate a limited number of mutants following plating onto induction media (Tm+aTc). That is, the total number of unique mutants represented may not be much greater than the initial number of transformant colonies (100-1000).

Table 5.2: Insertion site sequencing of 15 transposon mutants derived from three parent cells selected on clarithromycin. No valid sequencing data was obtained for clone C4.

Clone	Locus	BLAST	Unique
A1	Ga0111251_113033	tryptophan-rich sensory protein [Eubacterium limosum]	0/5
A2	Ga0111251_113033	tryptophan-rich sensory protein [Eubacterium limosum]	
A3	Ga0111251_113033	tryptophan-rich sensory protein [Eubacterium limosum]	
A4	Ga0111251_113033	tryptophan-rich sensory protein [Eubacterium limosum]	
A5	Ga0111251_113033	tryptophan-rich sensory protein [Eubacterium limosum]	
B1	Ga0111251_11268	IclR family transcriptional regulator [Eubacterium limosum]	1/5
B2	Ga0111251_11268	IclR family transcriptional regulator [Eubacterium limosum]	
B3	Ga0111251_11268	IclR family transcriptional regulator [Eubacterium limosum]	
B4	Ga0111251_11268	IclR family transcriptional regulator [Eubacterium limosum]	
B5	Ga0111251_113816	hypothetical protein [Eubacterium limosum]	
C1	Ga0111251_11995	hypothetical protein [Eubacterium limosum]	≥ 4/5
C2	msrA CDS	peptide-methionine (S)-S-oxide reductase [Eubacterium limosum]	
C3	Ga0111251_113774	SH3 domain-containing protein [Eubacterium callanderi]	
C4	-	-	
C5	Ga0111251_11917	AI-2E family transporter [Eubacterium limosum]	

5.2.12 Selecting against Spontaneous Transposition via *codA* Reassembly

The sequencing results above illustrate the problem posed by spontaneous transposition prior to induction. If transposase activity cannot be stringently controlled, then the total number of mutants generated per transformation may

not exceed by very much the total number of transformants. Given the low transfer efficiencies observed for the transposase vector in this study, this poses a challenge towards generation of large mutant libraries in *E. limosum*.

As noted above, the pMTL-CW22 vector series has been designed with a *codA* counterselection marker immediately downstream of the transposable element (see Figure 5.1). This gene is not transcribed initially; transcription is only permitted once the transposon is cut-and-pasted out of the plasmid. This system is therefore intended to provide a safeguard against premature transposition in clostridia, and has been noted to greatly reduce premature transposition in *E. coli*, even in the absence of 5-FC (C. Woods, personal communication).

5-FC may prove toxic in itself if the exposed organism has a functional endogenous *codA*. However, *E. limosum* ATCC 8486 possesses no homologue of *codA*, and growth of *E. limosum* on agar supplemented with 1 mM 5-FC showed no reduction or inhibition of growth (data not shown).

To assess the spontaneity of transposition in the presence of 5-FC, a further transformation was conducted, and transformants selected on clarithromycin supplemented by 1 mM 5-FC. Selection plates were protected from light to prevent spontaneous deamination of 5-FC to 5-FU. Three colonies were resuspended as described above, and dilution series plated to Tm+aTc to isolate individual colonies. These three lineages were then assessed for diversity of mutation via inverse PCR and sequencing ().

Table **5.3**).

Table 5.3: Insertion site sequencing of 30 transposon mutants, derived from three parent cells selected on clarithromycin and supplemented with 1 mM 5-FC. No valid sequencing data was obtained for clones B1, B2, B5 and B7.

Clone	Locus	BLAST	Proportion unique
A1	Ga0111251_113869	diguanylate cyclase [Eubacterium limosum]	7/10
A2	Ga0111251_113869	diguanylate cyclase [Eubacterium limosum]	
A3	Ga0111251_11171	GGDEF domain-containing protein [Eubacterium limosum]	
A4	Ga0111251_114076	hypothetical protein ELI_4407 [Eubacterium limosum KIST612]	
A5	Intergenic		
A6	Ga0111251_111575	PTS galactitol transporter subunit IIC [Eubacterium limosum]	
A7	Ga0111251_113869	diguanylate cyclase [Eubacterium limosum]	
A8	nfo gene	endonuclease IV [Eubacterium limosum]	
A9	Ga0111251_11908	OmpR DNA-binding response regulator [Eubacterium limosum]	
A10	Repeat region	(CRISPR repeat region)	
B1	-	-	≤4/10
B2	-	-	
B3	Intergenic		
B4	Intergenic (as B3)		
B5	-	-	
B6	Ga0111251_111503	tellurium resistance protein [*Eubacterium limosum*]	
B7	-	-	
B8	Ga0111251_111503	tellurium resistance protein [*Eubacterium limosum*]	
B9	Ga0111251_111503	tellurium resistance protein [*Eubacterium limosum*]	

B10	Ga0111251_111503	tellurium resistance protein [*Eubacterium limosum*]	
C1	Ga0111251_111971	EamA domain-containing membrane protein RarD [Eubacterium callanderi]	10/10
C2	Ga0111251_113850	NADPH-dependent 2,4-dienoyl-CoA reductase, sulfur reductase [Eubacterium callanderi]	
C3			
C4	Ga0111251_111413	HAD family hydrolase [Eubacterium limosum]	
C5	Ga0111251_113962	MULTISPECIES: uroporphyrinogen decarboxylase [Robinsoniella]	
C6	Ga0111251_114030	hypothetical protein [Eubacterium limosum]	
C7	Ga0111251_111657	hypothetical protein [Eubacterium limosum]	
C8	pre-Ga0111251_11973		
C9	Ga0111251_11317	MULTISPECIES: Fis family transcriptional regulator [Clostridiales]	
C10	Ga0111251_11980	NAD/NADP octopine/nopaline dehydrogenase [Eubacterium limosum]	

Of the three clonal lineages assessed (A-C), one appeared to have entirely avoided spontaneous transposition during clarithromycin selection (C, with 10 unique mutants). In a second, 3/10 colonies assessed were identical, indicating a single early transposition event. In the third and least diverse lineage, 6/10 colonies assessed were found to represent two mutants. While it is difficult to assign a definitive improvement when compared to colonies selected without 5-FC supplementation, 5-FC appears to have had at least some effect on the frequency of spontaneous transposition.

5.2.13 **Maximising Diversity in Transposon Mutant Libraries**

In accordance with the results above, subsequent transformations were conducted with initial selection on clarithromycin and 5-FC. Following colony outgrowth, all colonies obtained from a given electroporation were pooled and stored at -80^{o}C as described in Materials and Methods. These frozen stocks were retained for future generation of larger transposon mutant libraries.

5.2.14 **Plotting of Transposon Insertion Sites.**

To visualise the insertion mutants sequenced in the course of this study, the insertion sites of 61 unique mutants were plotted on a diagram of the *E. limosum* genome (Figure 5.14). As expected of the *Himar1C9* transposase, all sequenced mutants were found to possess inserted transposons at TA dinucleotide sites (data not shown).

The insertion site sequencing workflow employed in this study has previously been employed in larger, multi-person efforts to map the insertion sites of thousands of mutants in *Staphylococcus aureus* (Bae et al., 2004).

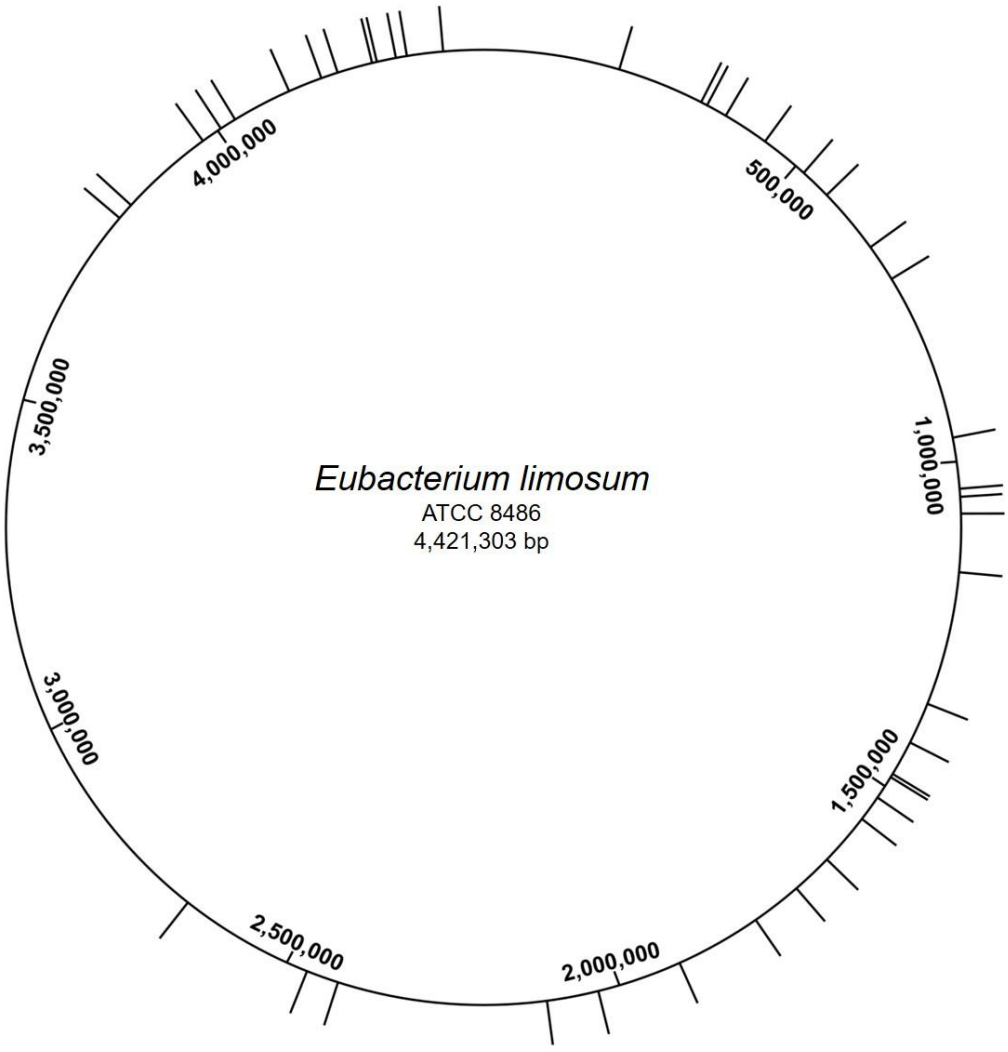


Figure 5.14: Transposon insertion sites on the *E. limosum* ATCC 8486 genome. Markers on the circular genome indicate insertion sites. Bases are numbered with reference to the first nucleotide of the start codon of *dnaA*, to which the rest of the genome is indexed. Plotted insertion sites represent mutants generated via transformation of both pMTL-CW21 and pMTL-CW23. Plotting of some insertion sites has been omitted due to their proximity to other insertion sites. For a complete list of insertional mutants generated during this study and sequenced after inverse PCR, see Appendix (section 8.2, Table 8.1).

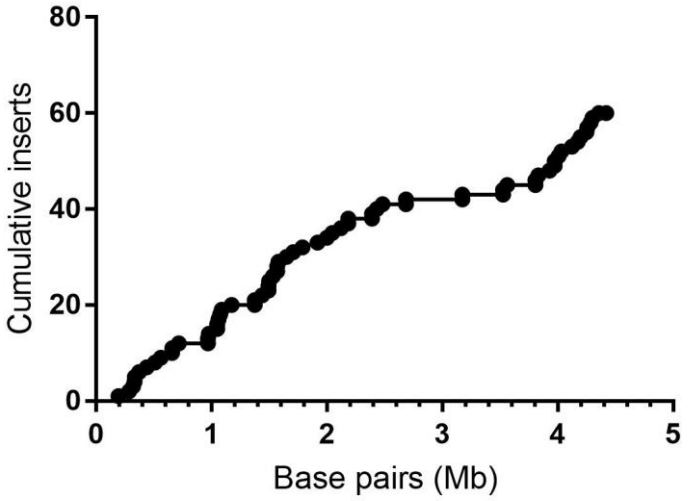


Figure 5.15: Cumulative total of transposon insertion sites plotted against total genome length of *E. limosum* ATCC 8486. Insertion sites are identical to those plotted in Figure 5.14.

An alternative visualisation of transposon insertion sites in the *E. limosum* genome can be seen in Figure 5.15, which is a plot of cumulative insertion sites per megabase. Such plots have been used to assess the randomness of transposon insertion sites in *S. aureus* (Bae et al., 2004), although in that study over 500 unique mutants were sequenced (note that the axes are reversed in Figure 5.15 versus Bae et al, to better illustrate the gaps in insertion coverage; additionally, the plot in Bae et al includes multiple insertion mutants per gene due to their greater coverage of the genome).

5.2.15 **Creation of Suicide Vector**

To explore the possibility of transposon mutant generation via a non-replicating ("suicide") vector, the existing pMTL-CW22 vector was modified. The existing Gram-positive replicon was removed and the remaining vector fragment was blunted and self-ligated to form pMTL-CW30. This was transformed into *E. limosum* and found to result in far fewer colonies than the previous vectors (10^0-10^1 colonies per transformation). Exposure to aTc during the recovery period (4 hours) was not found to rescue transformability, implying that this was within the window of effective induction of transposition. This may point to loss of the vector shortly after initial entry into the cell and prior to selection on thiamphenicol. Alternatively, transposase expression from single vectors may be insufficient to allow transposition prior to aTc induction and cell death.

5.3 Discussion

As reviewed earlier in this study, Tetracycline-induced gene expression systems have been employed in industrially-relevant species before, such as *C. acetobutylicum* (Dong et al., 2012). This is the first demonstration of a Tet-inducible gene expression system in *E. limosum*. Inducible gene expression systems have previously been employed in acetogenic bacteria such as *C. ljungdahlii* (Banerjee et al., 2014). They provide a useful means of complementing knockout mutants and assessing the strength of heterologous promoters. The work in this study constitutes one of the first tests of an

inducible gene expression system in *E. limosum*, with the other being an IPTG/lactose-inducible system assessed concomitantly to this study (R. Mansfield, personal communication). The CatP assay results described above indicate that the TetR system remains quite stringently controlled in *E. limosum*, in common with its performance in numerous other microorganisms. A degree of time- and dose-dependency of the TetR system's response to anhydrotetracycline have been noted. The inhibitory effects of aTc on the growth of *E. limosum* appear to be minimal below concentrations of 1000 ng/ml.

The diversity of transposon mutant libraries generated in this study have been hampered by the low numbers of unique mutants obtained per transformation attempt. Experimental approaches such as Transposon-Directed Insertion Site Sequencing (TraDIS) typically require large mutant pools containing hundreds of thousands or millions of unique insertion mutants (Langridge et al., 2009), although TraDIS has been conducted with mutant libraries containing tens of thousands of mutants (Dembek et al., 2015).

Two obstacles exist to generating large mutant pools in *E. limosum*: (1) Spontaneous transposition activity in *E. limosum* cells harboring the transposon vector and (2) low DNA transfer efficiency when transforming the vector. These effects act in synergy; neither would be an issue in the absence of the other. These issues have previously been identified as inhibiting the generation of large mutant libraries in *C. difficile* (Dembek et al., 2015).

The spontaneous transposase activity noted in this study (in the absence of aTc induction) means that transposon mutants may arise early in colony formation. Early mutants may therefore be overrepresented (potentially by over a million-fold). Given perfect stringency of transposase induction, DNA transfer efficiency would be a non-issue, because a single transformant could be grown to whatever volume was required, and the resulting population then induced into transposition. Alternative induction systems have been demonstrated in clostridia, including lactose- and IPTG-inducible systems (Zhang et al., 2015). It may be worthwhile to investigate the stringency of these systems in *E. limosum*, although the tetR system was expected to be of high stringency relative to other systems. Supplementation of selection plates with 5-FC appeared to have an effect on the spontaneity of tranposition. While 1 mM 5-FC was not particularly effective in preventing the appearance of early mutants, it may be worth investigating the effect of higher concentrations, up to the physiological limit of *E. limosum*.

The relative inefficiency of DNA transfer when electroporating large transposon vectors into *E. limosum* is the second factor inhibiting the generation of large mutant libraries. This low efficiency of transfer is likely due to the physical size of the transposon vectors and the presence of endogenous restriction barriers in *E. limosum*. Methylome data recently generated for *E. limosum* ATCC 8486 (the subject of this study) indicates that the motif CACNNNNRTANNG is recognised by one of *E. limosum*'s restriction-modification systems (see Chapter 7). That motif appears twice within *catP*, as found within the mini-transposon employed in this study.

Knockout of the corresponding R-M system (achieved during this study) may provide a more suitable host in which to conduct transposon mutagenesis, provided that higher DNA transfer efficiencies can be achieved for the transposon vector.

Alternatively, a non-plasmid-based system (consisting of transposable elements attached *in vitro* to purified transposase) may be effective in achieving a higher transformation efficiency than is achievable by a plasmid-based system. Such systems have been used to create large mutant libraries in *Salmonella typhimurium* (Barquist et al., 2013). The relative efficiency of electroporating enzymes versus plasmids into *E. limosum* has yet to be assessed.

A third issue is that the distribution of insertion sites observed during sequencing of transposon mutants during this study suggests that *mariner* may not insert randomly in *E. limosum*. If experiments such as TraDIS are to be attempted in the future, it may be necessary to employ an alternative transposon system, such as Tn*5*. However, this apparent inequality of distribution may be a statistical artefact of the relatively low number of unique mutants assessed during this study; *Mariner* has been noted to insert randomly due to its TA dinucleotide insertion site specificity (Chao et al., 2016).

5.4 Chapter Summary

- This study demonstrates the activity of an aTc-induced gene expression system in *E. limosum* ATCC 8486. This study represents the first transposon mutagenesis of *E. limosum*, using a *mariner*-based *in vivo* system. In all sequenced examples, the transposon was found to generate single insertions. Attempts to characterise the mechanism of induced plasmid loss in this system suggest that induced cell death may play a role.

- The efficacy of the anhydrotetracycline-induced promoter system has been demonstrated in *E. limosum*. Induction of CatP expression through the TetR system was shown to be time- and dose-dependent, with little to no measurable CatP activity in the absence of induction. 10 ng/ml was found to have an inconsistent effect between assays, and 50-100 ng/ml was found to have a significant effect.

- aTc was found to be inhibitory to growth only at concentrations in excess of 500 ng/ml, with delayed growth permitted at concentrations up to and including 1000 ng/ml.

- The TetR system was found to be insufficiently stringent to prevent spontaneous activity of the *mariner* transposase in vector pMTL-CW21/-CW23, even with supplementation by 5-FC. The formation of spontaneous mutants was confirmed via inverse PCR of un-induced transformant colonies.

- Clustering was observed when insertion sites of the *mariner* mini-transposon were plotted. A significant contiguous region of the *E. limosum* ATCC 8486

genome was not affected by insertional mutagenesis in any of the mutants sequenced during this study. This may be an artefact of the low number of transposon mutants generated and mapped during this study. Alternatively, it may reflect the conformation or replicative state of the bacterial genome during transposition.

- A previously-reported plasmid loss mechanism based on transcriptional read-through into the Gram-positive replicon appears to be applicable to the P19 replicon, in addition to the pCD6 and pCB102 replicons for which this behaviour has been previously described.

- Transformation of the transposon vectors used in this study was limited to 1 x 10^2 to 1 x 10^3 transformants per µg DNA. When considering the high observed incidence of spontaneous transposition from the vector, this means that obtaining the number of unique insertion mutants for high-throughput techniques such as TraDIS may be impractical. High-throughput insertional mutagenesis of *E. limosum* must wait for the establishment of higher DNA transfer efficiencies or the demonstration of more complete control over transposase activity.

CHAPTER SIX

Knockout of a Restriction Endonuclease in *E. limosum* via CRISPR-Cas9

6 Knockout of a Restriction Endonuclease in *E. limosum* via CRISPR-Cas9

Eubacterium limosum ATCC 8486 possesses a number of restriction-modification (R-M) systems, which are believed to inhibit DNA transfer. In this study, CRISPR-Cas9 gene editing was employed to generate an in-frame deletion of a type I restriction endonuclease in *E. limosum*, with a corresponding increase in DNA transfer efficiency being observed. A brief overview of R-M systems is presented here, as well as an overview of CRISPR-Cas9 gene editing technology. The R-M systems identified within *E. limosum* ATCC 8486 are characterised *in silico*, and their associated methylation motifs mapped against the plasmids previously employed in this study. Construction of CRISPR-Cas9 knockout vectors against these target genes is also described.

6.1 Introduction

Restriction-modification (R-M) systems have long been recognised as barriers to gene transfer. There are four types of R-M system. Type II restriction endonucleases were the earliest of such systems to be discovered and characterised, a process starting in the 1970s. They have since been extensively employed in biotechnological applications (reviewed in Pingoud et al., 2014). Type II restriction endonucleases are grouped together due to their ability to cut DNA at specific positions within or near specific recognition sites. The best-characterised sub-class of the Type II systems is Type IIP,

which recognises palindromic DNA sequences. EcoRI and EcoRV are examples of this class.

Restriction enzymes bind non-specifically to DNA, and they diffuse linearly down the DNA molecule until they encounter their specific recognition sequence, at which point cleavage occurs at or near the recognition site (Pingoud and Jeltsch, 1997). Some restriction endonucleases only create single-strand nicks in DNA, with full cleavage requiring two separate restriction events.

Cleavage of DNA by any restriction enzyme results in the formation of 5'-phosphoryl and 3'-hydroxyl groups (reviewed in Pingoud et al., 2014). The realisation that compatible ends could be generated by digesting disparate DNA molecules with the same restriction endonuclease, which could then be ligated together, marked a foundational moment of genetic engineering (reviewed in Berg and Mertz, 2010).

Restriction endonucleases are prevented from cleaving host DNA by their cognate methyltransferases, which may methylate cytosine or adenine. The restriction endonuclease and methyltransferase activities may occur in the same polypeptide chain, or they may be physically separate. Type II systems which possess both activities are classed as Type IIG or Type II Gamma restriction endonucleases. As with Type I R-M systems (see below), Type IIG R-M systems possess specificity subunits, which are often incorporated into the C-terminus of the combined R-M protein (reviewed in Pingoud et al.,

2014). These specificity (S) modules determine the target sequence of the R-M system.

Restriction-modification systems, with both activities taken together, have been described as selfish genetic elements, in which restriction endonucleases kill their host cell if their cognate methyltransferases are lost (Nakayama and Kobayashi, 1998). This hypothesis was supported by experiments in which plasmids bearing R-M systems could not be displaced by incompatible plasmids, even if those incoming plasmids were not vulnerable to restriction. This is an alternative to the hypothesis of restriction systems as purely anti-bacteriophage defences, and has been proposed as an explanation of the tight association of many restriction endonucleases with their cognate methyltransferase genes.

Unlike Type II systems, Type I R-M systems have not been widely employed as molecular biology tools (Loenen et al., 2014). They consist of distinct R, M and S subunits. Because both the R and M activities of the Type I system are guided by the same S subunit, a Type I R-M system can change sequence specificity in one coordinated step, which can counteract bacteriophage adaptation to host defences (Loenen et al., 2014). The genes encoding the R, M and S activities are often designated *hsd* ('host specificity determinant') (*hsdR*, *hsdM*, *hsdS* in *E. coli*). The best-studied Type I system is EcoKI of *E. coli* K12. It is co-located with that organism's Type IV R-M systems in a distinct chromosomal region called the 'immigration control region' (Loenen et al., 2014).

Type III R-M systems are unusual in that they recognise two inversely-repeated sites, and will only initiate cleavage (at a region downstream of one of the sites) if both are unmethylated (reviewed in Rao et al., 2014). Due to this specificity, they are unlikely to inhibit the transfer of small recombinant plasmids during genetic engineering.

Type IV restriction systems are distinct from Types I, II and III in that they restrict modified DNA (e.g. methylated or glucosylated), rather than DNA which is unmethylated at specific sites (reviewed in Loenen and Raleigh, 2014). Because of this activity, Type IV systems are referred to as modification-dependent systems (MDS) rather than restriction-modification systems (R-M) (Bair and Black, 2007). Because they do not affect un-modified DNA, DNA transfer efficiency can be improved in some organisms by using entirely un-methylated DNA for transformations, or conducting conjugation from a methylation-negative donor strain. A relevant example of this occurs in the acetogenic clostridium *C. autoethanogenum*, in which the absence of *dam* methylation improves DNA transfer efficiency significantly (C. Woods, SBRC Nottingham, personal communication).

6.2 Restriction-Modification Systems as a Barrier to Gene Transfer

Because they cleave incoming un-methylated DNA, restriction-modification systems have long been implicated in the reduction of exogenous DNA transfer efficiency (Aune and Aachmann, 2010, Pyne et al., 2014).

Several strategies have been developed to increase DNA transfer efficiency in the face of host restriction systems. Plasmid DNA can be passaged through a methylation strain of *E. coli* prior to transformation into a Gram-positive target strain. An example of this approach was utilised in *C. pasteurianum*, in which *E. coli* expressing the heterologous methyltransferase M.FnuDII was used to protect DNA prior to electrotransformation of *C. pasteurianum* (Pyne et al., 2013).

Alternatively, DNA may be modified *in vitro* by commercially-available methyltransferases (Groot et al., 2008). Approaches such as these, which do not require chromosomal modification of the target strain, are particularly of use when the host organism is otherwise entirely recalcitrant to transformation. A third approach involves the removal of putative restriction sites from plasmids, although this approach depends on knowledge of the restriction sites recognised by endogenous R-M systems, which may be determined based on homology to characterised systems, or following methylome sequencing.

Finally, restriction endonucleases may be directly disrupted via genetic engineering, whether by insertional mutagenesis vectors (e.g. ClosTron) or through homologous recombination. This approach has been applied in several clostridia. For example, two Type I restriction endonucleases in *C. saccharobutylicum* were disrupted via ClosTron, with an increase in conjugation efficiency being noted as a result (Lesiak et al., 2014).

All of the above approaches may be combined, with cumulative effects; in *C. pasteurianum*, improvements in transformation efficiency were observed following removal from the plasmid undergoing transformation of cleavage sites for the Type I restriction-modification system CpaAII, and additionally following disruption of the Type II restriction endonuclease CpaAI (Pyne et al., 2014).

During this study, low DNA transfer efficiency to *E. limosum* has been noted as a potential factor hindering high-throughput transposon mutagenesis (see Chapter 5). The transfer efficiency of plasmids is seen to fall as they increase in size, resulting in a 10,000- to 100,000-fold decrease in the number of observed colonies after transformation with constructs of 8 Kb rather than 5 Kb. This reduction in transfer efficiency is likely due to a combination of physical factors and the increased number of restriction sites on the larger vector. It follows that inactivation of *E. limosum*'s restriction endonucleases should improve DNA transfer efficiency.

The deletion of restriction-modification systems provides a useful demonstration of reverse genetic methods. The removal of a restriction endonuclease subunit is unlikely to be lethal. Even when a combined restriction endonuclease-methyltransferase ORF is removed, cells are likely to survive because no additional restriction endonuclease is produced to cleave the newly-synthesised un-methylated DNA. Secondly, removal of R-M systems is known to enhance DNA transfer efficiency, and R-M knockouts are therefore more amenable to further genetic manipulation. Thirdly, knockout of

R-M systems provides an easy-to-assess phenotype; differences between the mutant and wild-type can be determined by comparing DNA transfer efficiency.

6.3 Results

6.3.1 Sequence Analysis of R-M Systems in *E. limosum* ATCC 8486

E. limosum ATCC 8486 possesses three putative functional restriction systems. These are the Type I R-M system Eli8486ORFDP ("ORFD"), the Type II R-M system Eli8486ORFAP ("ORFA"), and the Type II methyltransferase M.Eli8486ORFBP ("ORFB") (see Figure 6.1). It also contains two putative methyltransferases which are unlikely to be functional due to their short length. In this study, DNA sequences for the first three R-M systems were obtained from a closed genome developed from work by R. Mansfield (R. Mansfield, SBRC Nottingham, personal communication). The sequences obtained from this closed genome are in perfect agreement with another *E. limosum* ATCC 8486 closed genome that was uploaded to the RefSeq database in February 2017 (Song., et al; unpublished).

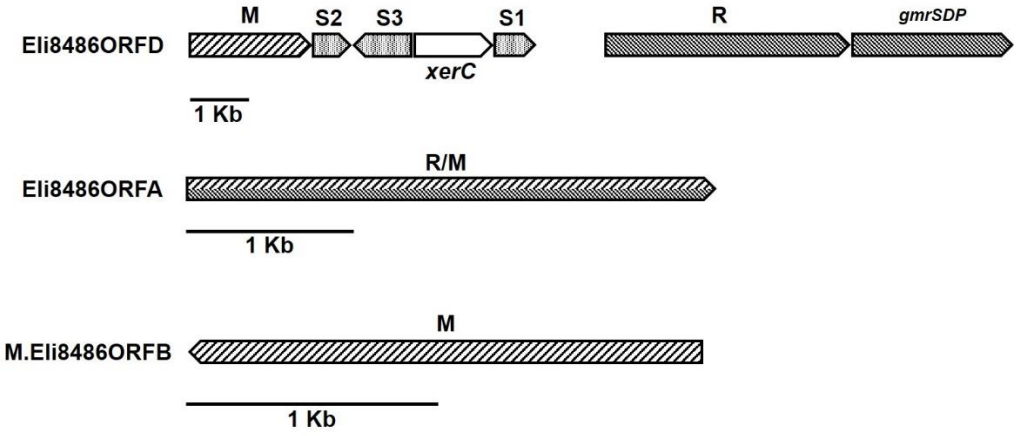


Figure 6.1: Overview of restriction-modification (R-M) systems in *E. limosum* ATCC 8486. **M:** Methyltransferase domain; **S1/S2/S3:** Specificity domains; **R:** Restriction endonuclease domain. ***xerC:*** Integrase. ***gmrSDP*:** Putative Type IV R-M system associated with *R.Eli8486ORFD*. Note scale bars accompanying each system. Additional prophage-borne R-M systems have been excluded from this figure because they are unlikely to be functional (see text for further details).

The Type I R-M system of *E. limosum* ATCC 8486 consists of three subunits (Restriction, Methylation, and Specificity), as is typical for Type I systems. The M and R subunits are 1506 bp and 3188 bp in length, respectively. Three specificity subunits (S1, S2, and S3) are found within this system, located adjacent to a *xerC* integrase. The S1 subunit was assessed to be the active component in this system (Prof. Richard Roberts, NEB, personal communication). A BLASTp search with the amino acid sequence of the methyltransferase (M.Eli8486ORFDP) reveals 89% and 88% identity with methyltransferases encoded by the human commensal and ruminant commensal species *Bariatricus massiliensis* (Bessis et al., 2016) and *Ruminococcus flavefaciens*, respectively. This is unsurprising as *E. limosum* is

associated with the mammalian gastrointestinal tract and restriction-modification systems are good candidates for horizontal gene transmission.

The S1 specificity subunit of Eli8486ORFDP (S1.Eli8486ORFDP) is more unique, having only a 67% identity to a specificity subunit from *Lachnoanaerobaculum saburreum*. Because this subunit directs both the methyltransferase and restriction endonuclease activities of this R-M system, mutations which change the target can occur safely, as the new sites will be methylated in advance of cleavage. This can serve to confound incoming bacteriophages which may have had their DNA methylated in a host with the obsolete R-M system.

The restriction endonuclease associated with Eli8486ORFDP (R.Eli8486ORFDP) bears an 88% amino acid sequence identity to an uncharacterised protein from *Clostridium* sp. CAG: 169. It bears 87% and 86% identity to Type I restriction endonucleases from *Clostridium lactatifermentans* and Lachnospiraceae bacterium NK4A179, the former being a lactate-fermenting commensal organism isolated from the gastrointestinal tract of a chicken (van der Wielen, 2002). The R subunit of ORFD is located in close proximity to a putative Type IV MDS (modification-directed system) with homology to the *E. coli gmrSD* system. Close association of Type I and Type IV systems has been noted previously in *E. coli* K12 (Loenen et al., 2014).

A BLASTp search of the amino acid sequence of the Type II R-M system of *E. limosum* ATCC 8486 (Eli8486ORFAP), reveals a 74% identity with a

methyltransferase from *Neglecta timonensis*. Lower levels of sequence identity (64% and 63%) are observed for methyltransferases from *Prevotella ruminicola* and *Ruminococcus flavefaciens*.

Analysis of the amino acid sequence of M.Eli8486ORFBP, the final intact methyltransferase gene of *E. limosum* ATCC 8486, indicates 95% identity with a multi-species clostridial DNA methyltransferase, and with an adenine-specific DNA methyltransferase from *Eubacterium callanderi*. This protein possesses a 50% sequence identity with a hypothetical protein from *Acetobacterium bakii*. It does not appear in any of the ruminant bacterial species mentioned above.

Interestingly, none of the R-M systems above appear in the other sequenced strains of *E. limosum* (KIST612 and SA11). This may be a result of relatively recent acquisition of these R-M systems from environmental sources via horizontal gene transmission. Analysis of the closed genome of *E. limosum* ATCC 8486 produced concomitantly to this study (R. Mansfield, SBRC, personal communication) indicates that the two additional methyltransferases (M.Eli8486ORFCP and M.Eli8486ORFGP) present in the genome are present on prophages, indicating previous horizontal transmission. Both are roughly 200 bp long and are not expected to be expressed (Prof Richard Roberts, NEB, personal communication).

Next-Generation sequencing of the *E. limosum* ATCC 8486 genome has revealed the following methylation motifs, constituting the organism's methylome: CAAAAAR, CACNNNNRTANNG, and GCGRAG (see Table

6.1). In all cases, methylation is found on A or T residues. This obviates protection of plasmid DNA by commercially-available CpG and GpC methyltransferases (e.g. M.CviPI and M.SssI), which only methlyate C residues. Recognition of the motif GCGRAG has been attributed to the Type IIG restriction-modification system Eli8486ORFAP, the motif CAAAAAR has been attributed to the Type IIG orphan methyltransferase M.Eli8486ORFBP, and the motif CACNNNNNRTANNG has been tentatively attributed to the Type I restriction-modification system Eli8486ORFDP (Prof. Richard Roberts, personal communication). The classification IIG (or II Gamma) denotes a combination of restriction endonuclease and methylase activities. M.Eli8486ORFBP was not noted to possess any restriction endonuclease activity during sequence analysis (hence the "M" prefix), and was not targeted for knockout via CRISPR-Cas9 in this study.

Table 6.1: Methylation motifs associated with R-M systems of *E. limosum* ATCC 8486.

R-M System	**Classification**	**Putative Methylation Motif**
Eli8486ORFD	I	CACNNNNNRTANNG
Eli8486ORFA	IIG	GCGRAG
M.Eli8486ORFB	IIG	CAAAAAR

As previously noted, Eli8486ORFDP constitutes a typical Type I restriction-modification system in that it includes R, M and S subunits. Interestingly, it also includes a GmrSD-type gene, designated Eli8486GmrSDP, immediately downstream of the R subunit of the Type I system. GmrSD is an *E. coli* Type

IV restriction system, which digests glucosylated bacteriophage DNA and acts as a defensive system (Bair and Black, 2007). The *E. limosum gmr* gene may provide an analogous defence. Such modification-dependent restriction systems are ubiquitous in bacteria, and are often associated with Type I restriction endonucleases (reviewed in Loenen and Raleigh, 2014).

It is worth noting that the transposon vectors (pMTL-CW21 and –CW23) employed in this study possess 6/9/3 and 6/10/3 of each of the motifs described above, respectively. It is therefore plausible that cutting of incoming plasmid DNA by endogenous R-M systems has a deleterious effect on transfer efficiency of those plasmids. For reference, the plasmid pMTL87151, which transfers at a much higher rate into *E. limosum* than the aforementioned transposon vectors, possesses 5/6/2 of the above-mentioned sites. However, it is also much smaller than the transposon vectors, and the size of plasmids has been observed to be inversely proportional to their efficiency of transformation (reviewed in Aune and Aachmann, 2010). A combination of cleavage by restriction endonucleases and physical size inhibiting transport across transient membrane pores is probably responsible for the lower transfer efficiencies observed for large plasmids in this study. A summary of putative *E. limosum* ATCC 8486 methylation motifs within the pMTL80000 components used in this study can be seen below (Table 6.2).

Table 6.2: Putative restriction-modification motifs within the pMTL80000 modular shuttle vector system. Only pMTL components used in this study are shown.

Component	**Methylation motifs present**		
	ORF D	ORF A	ORF B
pCB102 Gram-positive replicon			2
P19 Gram-positive replicon			3
ColE1 Gram-negative replicon (inc. traJ)		4	2
catP selective marker	2		1
ermB selective marker			
Multiple cloning site (inc. lacZ)		1	

Interestingly, two of the CAAAAAR motifs on pMTL87151 lie within the direct repeat (iteron) region of the P19 Gram-positive replicon. On the variant plasmid pMTL87151C, in whose Gram-positive replicon this region has been truncated, a roughly 100-fold increase in transformation frequency has been observed (see Chapter 4), which does not appear attributable to increases in host replication speed or any perturbation of plasmid copy number. The increased efficiency of pMTL87151C transfer could be attributed to alleviation of restriction. However, the orphan methyltransferase to which this motif is attributed, M.Eli8486ORFBP, has not been annotated as possessing a restriction activity.

During this study, two restriction methylation systems were initially identified as possessing unambiguous restriction activities, and were assigned provisional target motifs. These were the Type I R-M system Eli8486ORFD (Ga0111251_112370 to Ga0111251_112376) and the Type II R-M system Eli8486ORFA (Ga0111251_11256). Accordingly, these two systems were the initial targets for knockout via CRISPR-Cas9. The knockout approach taken in

this study can be contrasted with the approach of heterologously expressing *E. limosum* methyltransferases within an *E. coli* donor strain, an approach which was pursued concomitantly to this study (R. Mansfield, SBRC, personal communication).

6.3.2 The Endogenous CRISPR System of *E. limosum* ATCC 8486

The native CRISPR system of *E. limosum* ATCC 8486 may form an additional barrier to DNA transfer. To determine whether CRISPR-mediated cleavage of plasmid DNA is likely, the sequence of the native CRISPR locus was examined. As is typical, the *E. limosum* CRISPR locus consists of a number of *cas* genes and an associated repeat region containing the protospacers (Figure 6.2). While *cas1-cas8* are present, there is no native *cas9*. The presence of *cas5-8* in the *E. limosum* CRISPR locus, and the absence of *cas9*, allow the classification of this system as a Type I system, as described above (Cass et al., 2015). The remaining "core" *cas* genes are arranged *cas3-cas4-cas1-cas2*, which is the most common arrangement of these genes (Haft et al., 2005).

Located downstream of the *cas* genes, the repeat region consists of 30 repeat sequences, each with the nucleotide sequence "GTTGAAGATTAACATGAGATGTATTTAAAT". Between these repeats are the captured spacer regions which form the "memory" of the CRISPR system. These spacer regions are 34-37 bp in length, with the majority being 35 or 36 bp in length.

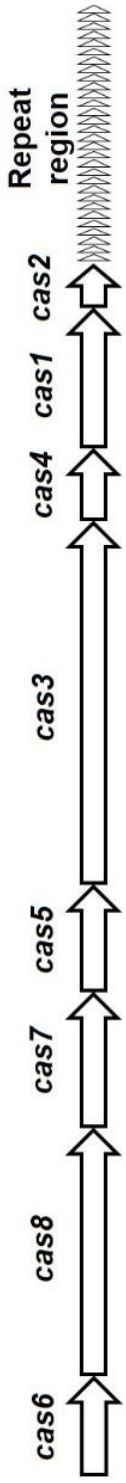


Figure 6.2: Diagram of *E. limosum* ATCC 8486 native CRISPR locus. The genes *cas1-4* constitute the "core" system. The remaining *cas* genes constitute the effector complex. The repeat region contains the CRISPR repeat motifs and spacer sequences. ***Cas6:*** endoribonuclease**;** ***Cas3:*** helicase**;** ***Cas4:*** exonuclease**;** ***Cas1:*** dsDNA endonuclease**;** ***Cas2:*** endoribonuclease.

To determine whether the native CRISPR system may function as a barrier to transformation, sequence alignment was conducted between the spacers in the CRISPR repeat region and the backbones of pMTL87151 and the transposon vector pMTL-CW21, as representative of plasmids used in this study. No correspondence was found between the CRISPR spacers and either of the two plasmids examined. Therefore, it is unlikely that the *E. limosum* native CRISPR system has a deleterious effect on the transfer efficiency of the plasmids utilised in this study. Simultaneously with this study, a *cas3* knockout mutant was obtained in *E. limosum* via CRISPR-Cas9 gene deletion. No effect on transformability was reported (Pete Rowe, SBRC, personal communication).

6.3.3 Evaluating R-M System Efficiency

The methylome data recently made available for *E. limosum* ATCC 8486 allows the tentative assignment of specificity sites to R-M systems (see Table 6.1). To assess the potential impact of R-M system knockouts prior to the creation of mutants, the transfer efficiencies of plasmids lacking putative restriction sites were compared to those possessing them. This approach takes advantage of the modular pMTL80000 system, which allows the easy generation of plasmids variant in the region containing putative restriction sites, but otherwise identical.

As noted in Chapter 4, truncation of the P19 replicon lead to a 100-fold increase in DNA transfer efficiency. This serendipitous result suggested that the Type II system M.Eli8486ORFB might possess a restriction endonuclease activity. However, the difference in CAAAAAR sites between pMTL87151 and pMTL87151C is only slight (6 versus 5), and pMTL83151 contains the same number (5) as pMTL87151C. There is no comparable difference in transfer efficiency of pMTL87151/83151 to that between pMTL87151/87151C (see Figure 6.2). This means that the increased transfer efficiency of pMTL87151C cannot be attributed to alleviation of restriction.

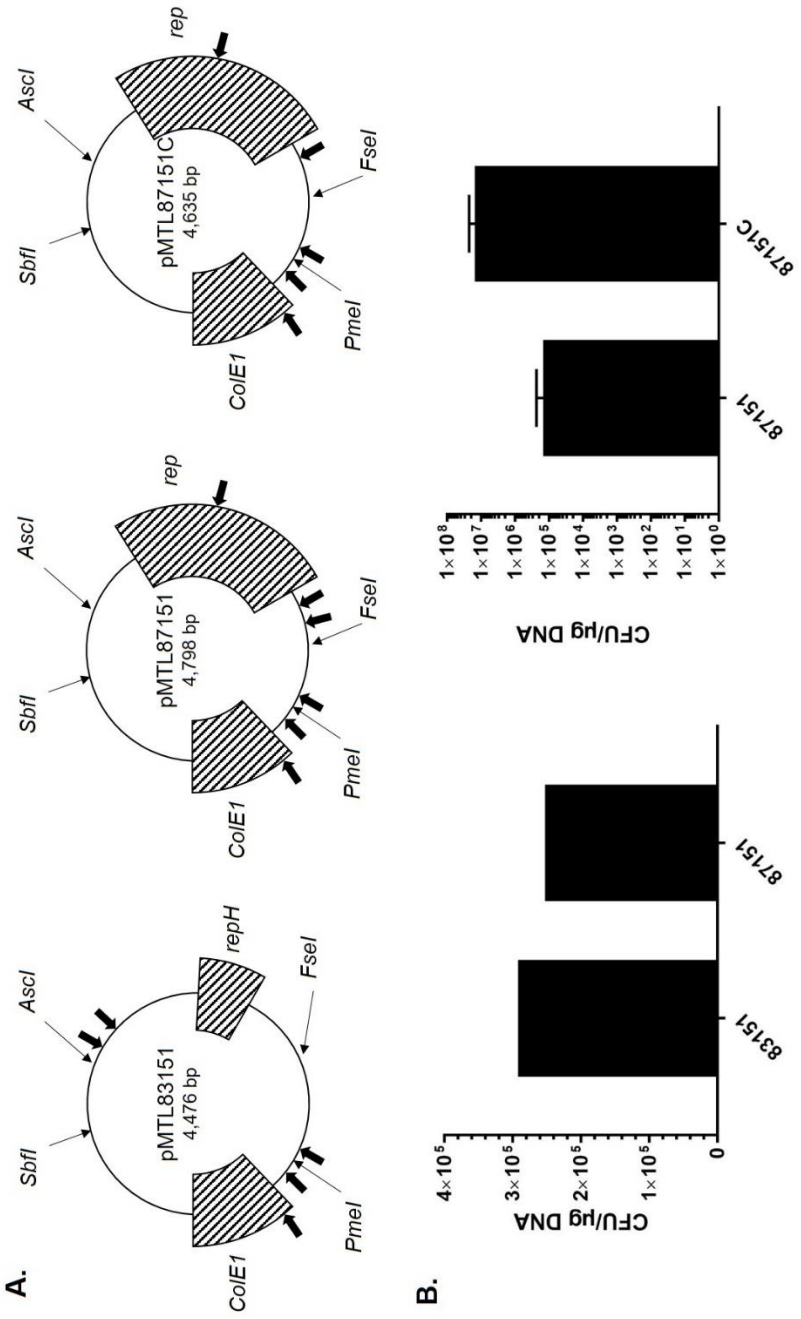


Figure 6.3: Comparative DNA transfer efficiencies of pMTL83151, 87151 and 87151C. **A.** Plasmid maps of the constructs used for this comparison. Thick black arrows indicate putative methylation sites (CAAAAAR) for M.Eli8486ORFB. Components of the pMTL series are separated by characteristic restriction sites; AscI/FseI: Gram-positive replicon. FseI/PmeI: Selective marker. PmeI/SbfI: Gram-negative replicon. **B.** Representative transformation efficiencies of the above plasmids. Note that while

pMTL83151 and pMTL87151 transfer at roughly $1x10^5$ CFU/µg DNA, pMTL87151C transfers at $1x10^7$ CFU/µg DNA.

This is consistent with the annotation of M.Eli8486ORFB as an orphan methyltransferase, lacking a restriction endonuclease activity. However, this leaves the increase in transfer efficiency observed for pMTL87151C without a ready explanation.

A second R-M system was assessed in this way during this study. Based on available methylome data for *E. limosum* ATCC 8486, R.Eli8486ORFD was assigned the methylation motif CACNNNNRTANNG and should therefore cut twice within the *catP* selective marker on pMTL80000 series plasmids (see Table 6.2). Accordingly, the effect of the restriction activity of ORFD was assessed by comparing the transfer efficiencies of pMTL87151C and pMTL87251C, which differ only in their selective marker. The latter plasmid should provide no target for restriction by ORFD.

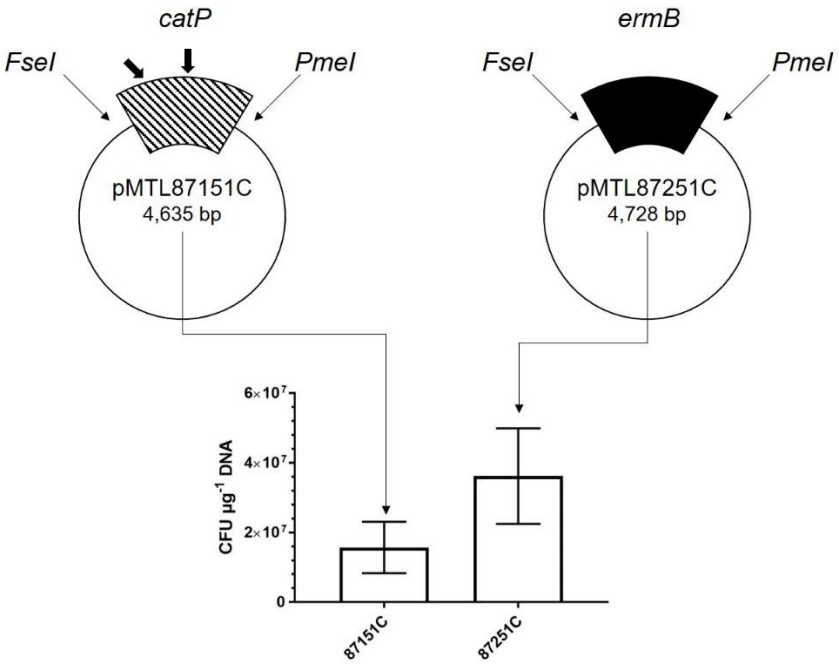


Figure 6.4: The effect of removing putative restriction sites for R.Eli8486ORFD. Transformations were conducted in *E. limosum*. Solid black arrows indicate putative methylation motifs (CACNNNNRTANNG) for M.Eli8486ORFD. The two plasmids differ only in their selective marker (*catP* conferring resistance to thiamphenicol/chloramphenicol, and *ermB* conferring resistance to clarithromycin). Column heights represent the average transformation efficiencies of 3 separate transformations. Error bars represent the standard deviation. No significant difference was detected between the CFU obtained for each condition (using an unpaired Welch's T test with a two-tailed P value; P value = 0.1054).

Comparison of the transfer efficiencies of pMTL87251C versus pMTL87151C revealed a roughly two-fold increase in transfer efficiency in favour of pMTL87251C. However, statistical analysis indicated that the observed difference in CFU was not significant (Figure 6.4). Because of the high transformant CFU obtained under these conditions, dilution error may obscure differences in DNA transfer efficiency.

6.3.4 **Construction of CRISPR KO Vectors**

All plasmids and primers employed in this study are detailed in Materials and Methods. The pSBRCas9 vector was employed for CRISPR-Cas9 gene deletion in this study (obtained from and developed by Pete Rowe, SBRC, Nottingham). It possesses a *S. pyogenes cas9* driven by a strong P_{thl} thiolase promoter from *C. acetobutylicum*. The *cas9* on the pSBRC vector possesses a frameshift mutation in the 5' poly-T region, but has been shown to be functional. Constructs based on pSBRCas9 are named pMTL4NNNNN (see Figure 6.6). The determining factor in the counterselective strength of the CRISPR vector is the activity of the promoter driving transcription of the sgRNA. On the pSBRCas9 vector, this is driven by the *C. acetobutylicum araE* promoter, which has been assessed as being of moderate-to-high strength (Pete Rowe, personal communication). The vector was initially conceived for use in *C. autoethanogenum*, but is functional in a number of clostridia. A plasmid map of the pMTL431511::ORFD knockout vector employed in this study is shown below (Figure 6.6). This vector has been designed to accept modularised components, with the recombination cassette flanked by AsiSI and AscI sites, and the Gram-positive replicon and selective marker flanked by AscI/FseI and FseI/PmeI, respectively.

To exemplify CRISPR-Cas9 gene deletion in *E. limosum*, and to render the strain more amenable to genetic manipulation, *E. limosum*'s endogenous restriction-modification systems were chosen as targets for gene knockout. In parallel to this study, efforts were made to express *E. limosum*

methyltransferases in an *E. coli* background (R. Mansfield, personal communication), approaching the same problem from a different direction.

The construction of CRISPR-Cas9 knockout vectors is described more fully in Materials and Methods. Briefly, each knockout vector varies only in the target-specific recombination cassette and sgRNA targeting sequence it possesses. The homology arms of the recombination cassette are designed to remove the target gene from the host chromosome following a double-crossover event. Homology arms were designed to be a minimum of 750 bp in length. While longer homology arms have a beneficial effect on recombination frequencies (Fujitani et al., 1995), total size of the knockout vector is inversely related to transformation efficiency.

In the event of a double crossover, the target gene is removed from the host chromosome and the plasmid backbone de-integrates. When this has occurred, CRISPR-Cas9 (guided by an RNA sequence transcribed from the sgRNA sequence on the knockout vector) is unable to cut within the host chromosome (barring off-target events). If a double crossover event does not occur, the CRISPR-Cas9 is able to catalyse a lethal double-stranded break in the host chromosome (both of these scenarios are illustrated in Figure 6.5). Even following a successful double-crossover, it is possible for off-target cutting to occur. The risk of off-target cutting was minimised by selecting candidate sgRNAs with low off-target scores, as defined by the Benchling software suite (see Materials and Methods).

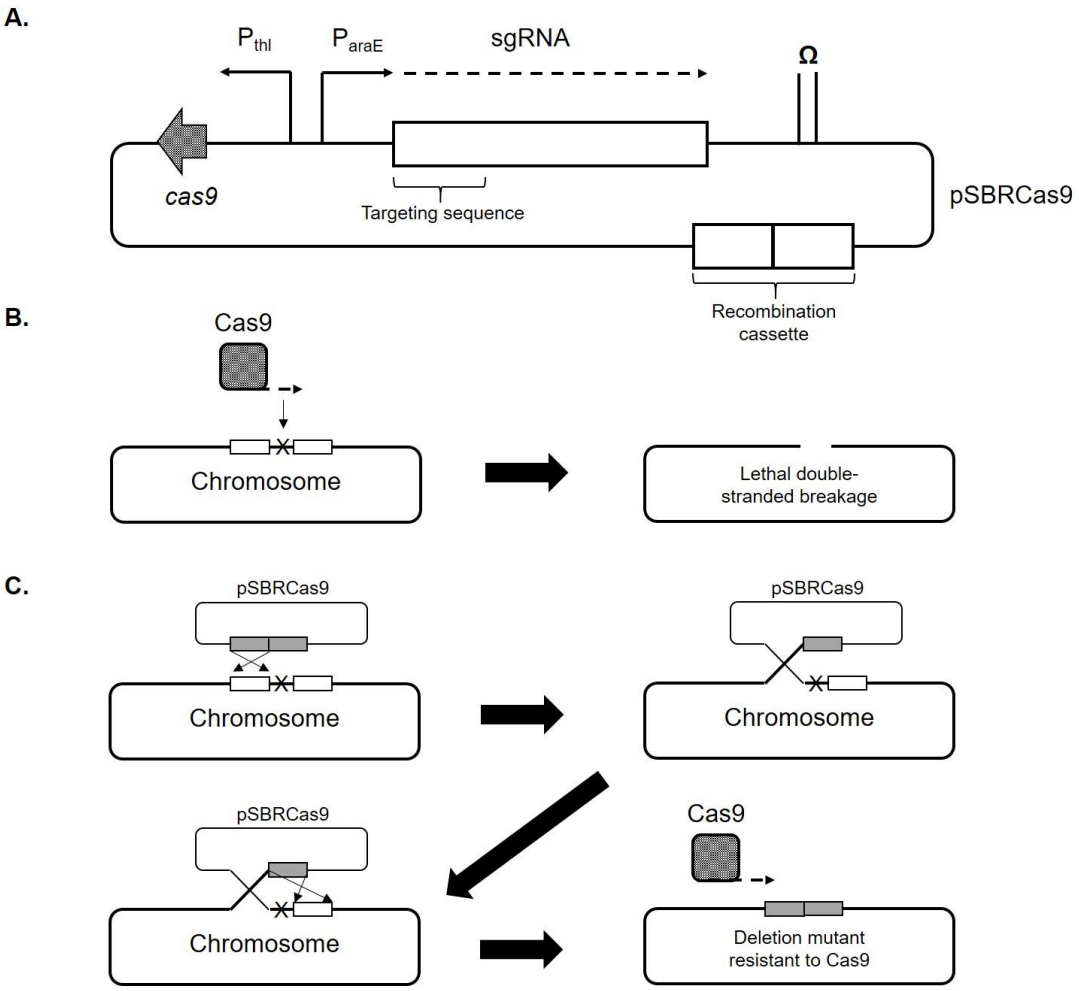


Figure 6.5: Illustration of gene editing via CRISPR-Cas9. This figure shows the method of action of the pSBRCas9 vector (see plasmid map, Figure 6.6). **A:** Schematic of the pSBRCas9 vector. The short guide RNA (sgRNA) is shown as a dashed line. Vector elements are not shown to scale. **B:** Counterselection by Cas9. Guided by the sgRNA (dashed line) encoded by the mutagenesis vector, Cas9 catalyses a double-stranded breakage at the target DNA sequence (X), killing wild-type cells. **C:** A double-crossover event rescues the mutant cell from Cas9 counterselection by removing the target site (X). The sequence of events depicts an initial crossover event which integrates the pSBRCas9 vector into the chromosome, followed by a second crossover event which excises the vector, removing the target gene (and the sequence to which Cas9 is targeted). For the purposes of clarity, features of the pSBRCas9 backbone (including *catP* selective marker and pCB102 Gram-positive replicon) have been omitted from this illustration (see Figure 6.6 for detailed map).

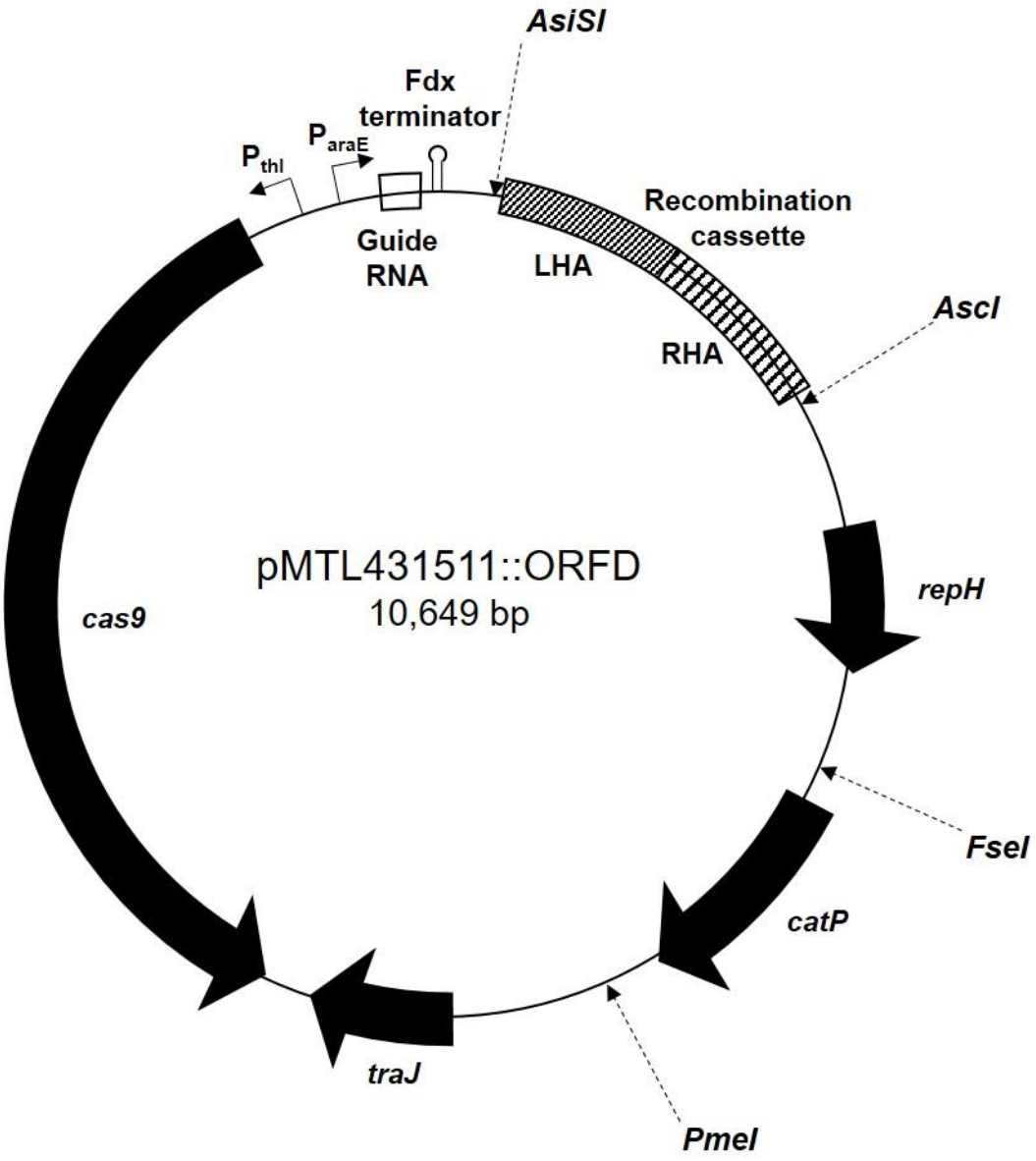


Figure 6.6: Schematic overview of the pMTL431511::ORFD knockout vector employed in this study. **P_{thl}:** *Clostridium acetobutylicum* thiolase promoter driving expression of Cas9. **P_{araE}**: *C. acetobutylicum araE* promoter driving expression of sgRNA. **Cas9:** *cas9* gene from *Streptococcus pyogenes*. The nomenclature for pMTL CRISPR vectors is similar to that employed for pMTL80000 shuttle vectors. The first digit (4) indicates the use of the SBRC *S. pyogenes cas9*. The second digit (3) indicates use of the pCB102 Gram-positive replicon. The third digit (1) denotes the *catP* selective marker. The fourth digit (5) denotes the ColE1 Gram-negative replicon with *traJ*. The fifth digit (1) denotes P_{araE} driving expression of the sgRNA. The sixth digit (1) denotes P_{thl} driving expression of the *cas9*.

6.3.4.1 Construction of pMTL431511::ORFD

In this study, the restriction subunit of the Type I system, *R.Eli8486ORFD*, was targeted for in-frame deletion. To assemble a CRISPR Cas9 knockout vector targeted against *R.Eli8486ORFD*, a suitable sgRNA was selected as described in Materials and Methods (section 2.10.5). Homology arms were created via PCR, spliced together via SOEing PCR, and ligated into the CRISPR-Cas9 vector (pSBRCas9) following digestion of both insert and vector with AsiSI and AscI. Of note is that the recombination cassette was designed to remove both *R.Eli8486ORFD* and the Type IV system *gmrSD*, immediately downstream. This gene was identified as being associated with the restriction subunit of Eli8486ORFD (Prof Richard Roberts, NEB, personal communication), and appears to be co-transcribed.

6.3.5 Knockout of *R.Eli8486ORFD*

Following cloning of the CRISPR knockout vector against *R.Eli8486ORFD*, it was transformed into *E. limosum* via electroporation. In contrast to the efficiencies observed when transforming 5-6 Kb shuttle vectors of the pMTL series (10^4-10^5 transformants per µg DNA), less than 100 colonies were typically observed following transformation of the CRISPR vectors. This may be attributable to the large size (over 10 Kb) of the knockout vectors, and possibly to early activity of the Cas9 system as well. A similar reduction in DNA transfer efficiency was observed when transforming in large transposon vectors (8 Kb), resulting in transformation efficiencies of 10^3-10^4

transformants per µg DNA. Low transformation frequencies have previously been observed for CRISPR knockout vectors of a similar design in *C. acetobutylicum* (Li et al., 2016), and during transformation of pSBRCas9 in *C. autoethanogenum* (Pete Rowe, SBRC Nottingham personal communication).

6 of the initial transformant colonies were restreaked on selective medium and analysed via colony PCR. The primers employed were ORFD-EXT-F (binding on the *E. limosum* chromosome immediately outside of the knockout locus) and ORFD-RHAR (the primer used to generate the Right Homology Arm of the knockout vector; binding roughly 700 bp downstream of the knockout locus). Following the initial restreak, 2/6 colonies screened showed only mutant bands, with 3/6 showing a mixture of WT and mutant bands, and the final colony showing only a WT band (Figure 6.7). The two entirely mutant clones were restreaked on selective medium and a further diagnostic PCR performed with primers ORFD-EXT-F and ORFD-EXT-R on three replicate colonies from each plate (Figure 6.8). Bands of roughly 2 Kb were produced, consistent with removal of *R.Eli8486ORFD* and *gmrSD*. These bands were extracted and sequenced using the aforementioned primers to confirm gene deletion (Figure 6.9). Note that because both of these screening primers lie outside of the recombination locus, any observed truncations must be due to a double-crossover event, rather than a single-crossover leading to an integrated plasmid.

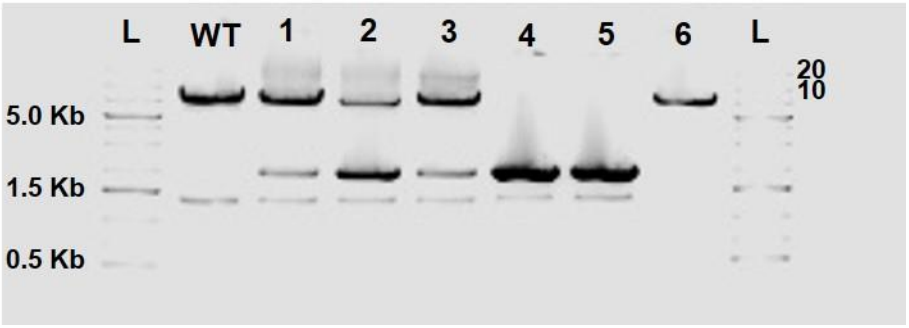


Figure 6.7: Diagnostic colony PCR of *E. limosum* putative transformant colonies following transformation with ORFD knockout construct pMTL431511::ORFD. Products amplified using primers ORFD-EXT-F (external to recombination locus) and ORFD-RHAR (binding to 3' end of recombination locus). Lanes 1-6: putative knockout mutants. WT**:** Wild-type *E. limosum* ATCC 8486. Lanes 1-3 represent mixed colonies in which part of the colony has undergone a double-crossover event. Lane 6 represents a colony in which no recombination events have yet occurred. Lanes 4 and 5 represent colonies in which recombination events have occurred and in which no wild-type copy of the recombination locus remains. L = ladder.

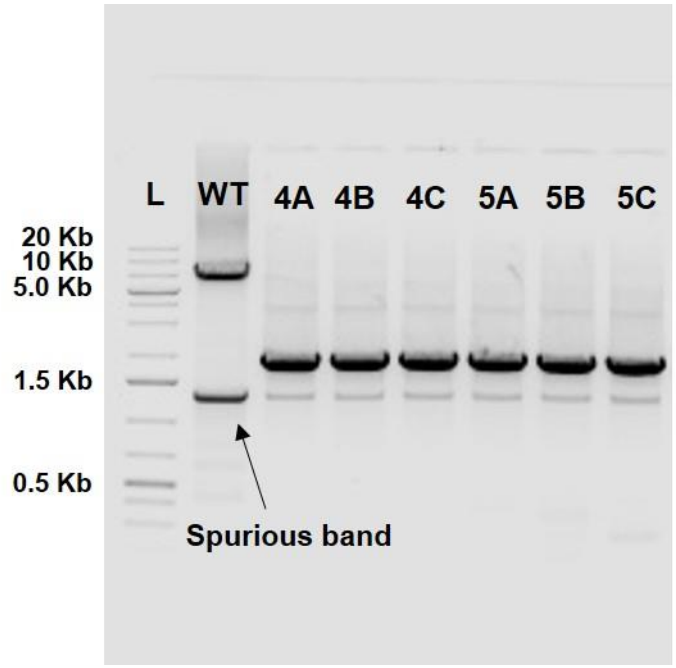


Figure 6.8: PCR across both ends of the *R.Eli8486ORFD* locus confirming double-crossover and knockout. PCR products amplified with ORFD-EXT-F/-R, binding externally to the recombination locus. WT: Wild-type *E. limosum* (negative control). 4A-C, 5A-C: Replicates of colonies 4/5 in Figure 6.7, above. L = ladder.

AATTGATAGTATAAATTCTTGTTTAAAATGAAACGATAAATAAAAA
TCAAGCCTCAGATGGAGGCGTTGCGATAGAAAAAACGACAAAGGA
GTGATAACTATGTAAAATCAGAAAAAACAATTACAGTCAGATGAA
ATATTTAGGATATTATTTTGAGGTTCAGATAAATCGATATTATGTAG
ATAGATGCTTTAATTATAGATAC

Figure 6.9: Sequence of the *R.Eli8486ORFD-gmrSD* locus following deletion of *R.Eli8486ORFD* and *gmrSD*. 100 bp of the left and right homology arms adjacent to the deletion site have been highlighted. The remaining start codon of *R.Eli8486ORFD* and stop codon of *gmrSD* have been underlined.

Following knockout of the target gene, several additional rounds of restreaking on non-selective medium were employed in an attempt to remove the knockout vector. However, no obvious reduction in thiamphenicol resistance was observed after several rounds of restreaking. Liquid passage in non-selective medium was conducted, and dilution series plated after each passage in order to enumerate Tm^R CFU and determine plasmid loss. This thiamphenicol resistance may represent an integration event by the CRISPR knockout vector at a location away from the target site.

Competent cells were then prepared from the ORFD knockout clones and their transformability assessed. Because of lingering thiamphenicol resistance, transformant selection on clarithromycin was necessary. However, because both putative motifs of ORFD lie within *catP*, a plasmid was selected that possessed a *catP* but which did not rely on it as a selective marker. For reasons of expedience, the TetR reporter plasmid employed in the previous chapter (pMTL-JM101; 6,027 bp) was used. This plasmid features an *ermB* selective

marker conferring clarithromycin resistance, as well as an inducible *catP* to provide target sites for *R.Eli8486ORFD*.

Following transformation of pMTL-JM102 into *E. limosum* wild-type and *ORFD*, a significant increase in transformability was observed (Figure 6.10). The increase observed was from an average of $7.05x10^3$ CFU/µg DNA (wild-type) to $7.22x10^4$ CFU/µg DNA. Note that the transformation efficiencies observed in this experiment (10^3-10^4 CFU/µg DNA) are lower than those observed when transforming pMTL87151 and pMTL83151 (10^6 CFU/µg DNA). This is probably due to the greater size of the reporter plasmid as compared to those shuttle vectors, resulting in a greater number of potential restriction sites and greater physical difficulty in transporting the plasmid into the cell. The increase in transformability observed in this experiment is in contrast to that observed when assessing the transformation efficiency of pMTL87151C versus that of pMTL87251C (section 4.2.10). In that case, a difference in transformant CFU may have been obscured by cumulative dilution errors arising from the need to plate high dilutions to obtain distinct CFU.

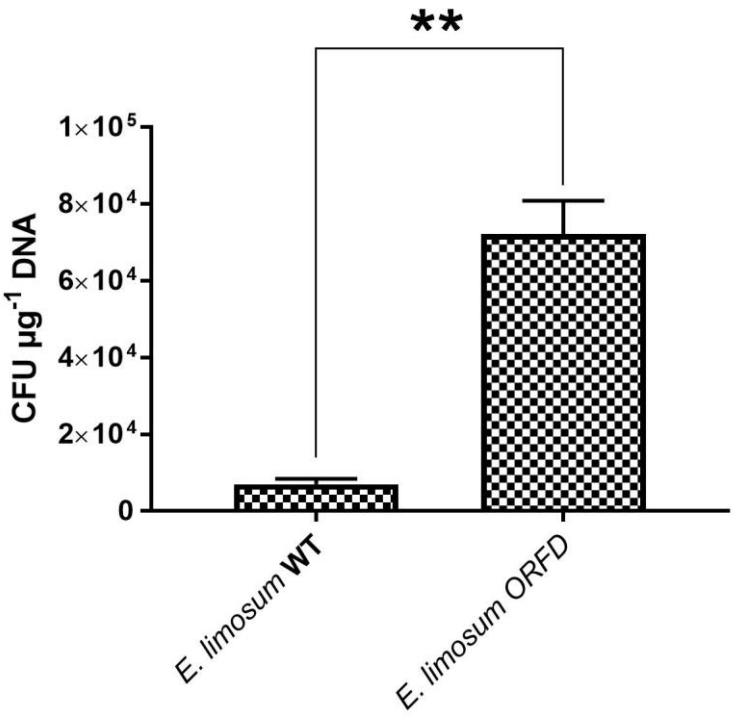


Figure 6.10: Comparison of transformation efficiency between wild-type *E. limosum* ATCC 8486 **(WT)** and *E. limosum* ATCC 8486 *ORFD* following transformation with pMTL-JM102. Column heights represent the mean average of three transformations. Error bars represent the standard deviation. All transformations were conducted using aliquots from the same miniprep of pMTL-JM102. Statistical analysis indicated that the observed difference between the two average transformation efficiencies was significant (using Welch's unpaired T test with two-tailed P value; P value = 0.0049).

Transformation of *E. limosum ORFD* with the reporter construct pMTL-JM102, based on the same Gram-positive replicon as the knockout vector, might be expected to provoke loss of that vector (unless the lingering thiamphenicol resistance observed is due to chromosomal integration of the knockout vector). To test this, *E. limosum ORFD* transformants harboring the reporter vector were restreaked onto Tm selective medium. 4/5 clones were observed to retain thiamphenicol resistance. The last clone was passaged in

non-selective medium and plated to confirm thiamphenicol sensitivity. A diagnostic PCR was performed (as Figure 6.8, above) to confirm that the cells represented a knockout mutant. This Tm^S clone was stored at -80°C.

6.3.6 SNP within the guide RNA of the knockout vector

Sequencing of pMTL431511::ORFD during cloning revealed an A→T SNP in the guide RNA region. This does not appear to have affected the initial effectiveness of counter-selection; a previous employment of the pSBRC knockout vector (without sgRNA SNP) in *C. autoethanogenum* resulted in 90% knockouts following the initial restreak (comparable to the >80% observed for knockout in this study, if all colonies containing mixed wild-type and mutant populations are included; see Figure 6.7) (P. Rowe, SBRC, personal communication).

6.4 Discussion

This chapter provides a demonstration of CRISPR-Cas9 gene editing in *E. limosum*. As employed in this study, CRISPR-Cas9 does not require any media supplementation or specific nutritional background to enable counterselection. Pure double-crossover mutants can be enriched by restreaking of colonies harbouring the knockout vector on selective media.

Knockout of the restriction endonuclease subunit of the Type I system Eli8486ORFD was seen to result in a roughly 10-fold increase in transformability. Sequencing across the ORFD locus confirmed that the restriction endonuclease *R.Eli8486ORFD* and adjacent Type IV restriction endonuclease *gmrSD* had been deleted. Deletion of this restriction endonuclease is of particular interest because two of its putative cleavage sites are found within the *catP* gene, which is widely used as a selective marker and has been used on a variety of plasmids within this study. Activity of *R.Eli8486ORFD* may therefore impact the function of the mini-transposon employed in this study, which consists of a *catP* flanked by transposase recognition sites. This knockout strain may therefore enable greater efficiency of transformation of the transposon vector (and eliminate the possibility of self-killing following transposition of an unmethylated mini-transposon into the chromosome). The 10-fold increase in DNA transfer efficiency observed in this study is comparable to other studies in which Type I restriction endonucleases were deleted. In *C. saccharobutylicum*, deletion of a pair of

Type I restriction endonucleases resulted in 8-fold and 10-fold increases in DNA transfer efficiency, respectively (Lesiak et al., 2014).

In two independent *R.ORFD* knockout mutants assessed in this study, it proved difficult to lose the thiamphenicol resistance conferred by the CRISPR knockout vector, although a thiamphenicol-sensitive clone was obtained following transformation of a plasmid maintained by the same replicon and selected for by clarithromycin. The fact that the plasmid was very difficult to lose in two independent mutants may indicate an integrative event at a locus outside of the locus defined by the homology arms of the knockout vector. This event may be characteristic of employment of this vector in *E. limosum*. Recombination between the plasmid-borne *cas9* and the *E. limosum* native CRISPR locus is unlikely, because the native system does not possess a *cas9*. It may be worthwhile to conduct inverse PCR on the existing thiamphenicol-resistant clones to determine the vector integration site, which may lead to re-designing of the present knockout vector to reduce the incidence of integration.

Because of the CRISPR-Cas9 counterselection method, cumulative knockout of all R-M systems in *E. limosum* should be feasible. A double knockout lacking ORFD and ORFA should possess considerably higher transformability than the wild-type and *ORFD* strains. Single knockouts in these genes could be compared with a double knockout to assess the relative importance of these R-M systems in inhibiting DNA transfer.

6.5 Chapter Summary

- Sequence analysis reveals that *E. limosum* ATCC 8486 possesses three restriction-modification (R-M) systems that are not shared by the other sequenced strains of this species, KIST612 and SA11.
- All three R-M systems analysed possess multiple putative methylation sites within the standard pMTL backbone.
- The truncated Gram-positive replicon P19C was previously found to result in a 100-fold increase in transformation efficiency. While P19C is missing a single putative methylation site for M.Eli8486ORFB, the observed increase in DNA transfer efficiency is unlikely to be due to relief from restriction. This is because no such increase is observed during transfer of pMTL83151, which possesses the same number of sites as pMTL87151C.
- Knockout of the Type I restriction endonuclease R.Eli8486ORFD resulted in a ten-fold increase in transformability. This restriction endonuclease possessed two putative cleavage sites within the widely-used *catP* selective marker.
- R.Eli8486ORFD's ability to cut within *catP* may have implications for the use of *catP*-based mini-transposon systems in *E. limosum* ATCC 8486, as seen in Chapter 5. The *ORFD* knockout mutant might be expected to be a superior host for the transposon vector, assuming plasmid loss can be confirmed.
- The effectiveness of a plasmid-based CRISPR-Cas9 knockout vector was demonstrated in *E. limosum* ATCC 8486, exemplifying this method of genome editing in this strain.

- *E. limosum* ATCC 8486 possesses a native Type I CRISPR system, but due to lack of homology between the spacer sequences it possesses and the plasmids employed in this study, it is unlikely to affect DNA transfer efficiency.

CHAPTER SEVEN

GENERAL DISCUSSION

7 General Discussion

Prior to this study, no molecular biological work had been published in any strain of *E. limosum*, and no electroporation protocol had been established in the literature. A complete genome had previously been published for *E. limosum* KIST612 (Roh et al., 2011). During the course of this study, two additional genomes have been published, representing *E. limosum* SA11 (Kelly et al., 2016), and *E. limosum* ATCC 8486 (the type strain which is examined in this study) (Song and Cho, 2015). Simultaneously or slightly before the draft genome of Song *et al*, a closed genome of ATCC 8486 was produced at the Synthetic Biology Research Centre, Nottingham (Rob Mansfield). This genome (and associated methylome data) provided a guide for the forward and reverse genetic methods conducted in this study.

Following the establishment of an electroporation protocol, optimisation steps were taken to determine the maximum DNA transfer efficiency achievable with the clostridial shuttle vectors available.

This study began as an attempt to establish advanced genetic methods in *Clostridium carboxidivorans*, a species which proved refractory to DNA transfer. This is likely due to the number and variety of restriction-modification systems present in that organism. However, following production of a draft genome via Illumina sequencing, a sequence analysis of *C. carboxidivorans'* native plasmid P19 was conducted. From this analysis, the putative origin of replication was identified and successfully modularised into a small plasmid replicon. This replicon was found to be suitable for use in a

variety of clostridia of industrial interest, and proved to be remarkably stable in the absence of antibiotic selection. Quantitative PCR revealed the novel replicon to possess a low copy number across multiple species. This work also established the copy number of the widely-used pCB102 Gram-positive replicon, which is the other plasmid replicon capable of efficient replication within *E. limosum*. Segregational stability assays of P19 and pCB102 in *E. limosum* revealed the latter to be more deficient in replication, and it was accordingly employed in the integrative vectors used for reverse genetics in this study.

During the course of this study, the function of the anhydrotetracycline-inducible TetR system was characterised in *E. limosum*. A CatP assay was conducted and revealed the system to be dose-responsive, with an apparent minimum inducer concentration between 0-10 ng/ml. The greatest system induction was observed at 100 ng/ml, which was the highest level of induction tested. Minimal CatP activity was observed in the absence of anhydrotetracycline induction. The Tet system demonstrated in this study may form the basis for future inducible gene expression systems in *E. limosum*.

Transposon mutagenesis of *E. limosum* was exemplified for the first time in this study. This work employed the *Mariner* mini-transposon system previously employed in a number of clostridia (Cartman and Minton, 2010, Zhang et al., 2015), placed within a novel vector developed concomitantly with this study for use in *Clostridium autoethanogenum* (Craig Woods). After replacing the Gram-positive replicon of the *C. autoethanogenum* transposon

vector, electroporation was conducted at low efficiency. Inverse PCR was conducted to map the insertion sites of the *Mariner* transposon across the *E. limosum* genome. The *mariner* transposon was found to insert at TA loci, as previously reported, and to generate single-insertion mutants. However, lack of double insertions was only assessed via inverse PCR, and preferential amplification of one insertion site over another may result in false negatives. Southern Blots could be performed to determine whether the transposon mutants observed thus far are in fact single insertion mutants.

Despite the apparently stringent regulation conferred by the Tet system, spontaneous transposase activity was observed following inverse PCR of clonal lineages carrying the transposon vector prior to induction with anhydrotetracycline. This is significant because spontaneous transposition on the initial selection plates reduced the total diversity of a transposon mutant library. The occurrence of spontaneous transposition means that the relative inefficiency of DNA transfer limits the richness of mutant populations generated during this study. Essentially, the combination of a large mutagenesis vector and insufficiently stringent control of the transposase means that only small mutant libraries can currently be generated. However, if the transposase gene (and associated regulatory system) were placed on a separate plasmid to the transposon, and the two transformed in tandem, both of these issues may be overcome. Pre-transformation with a vector expressing one component (transposon or transposase) would be followed by transformation with the second component (hosted on a non-replicative "suicide" vector). Following mutant selection, loss of the first vector could be

prompted via the aTc-mediated plasmid loss mechanism demonstrated in this study (refer to section 5.2.8, above). It remains to be seen whether the *Himar1C9* transposase operates efficiently in *trans* in the Gram-positive bacterial context. However, this strategy has been employed with *Himar1C9* in human cells (Keravala et al., 2006).

The methylome data generated for *E. limosum* ATCC 8486 in the course of this study indicated the presence of two restriction sites within the mini-transposon employed in this study, and this made the deletion of the restriction endonuclease responsible a high priority.

This study saw the exemplification of advanced genome editing methods in *E. limosum*, with the deleting of a restriction endonuclease via CRISPR-Cas9. The phenotype of the resulting deletion mutant was assessed via electroporation, and a ten-fold increase in DNA transfer efficiency was observed. This result is significant for future transposon mutagenesis of *E. limosum* ATCC 8486, as the restriction endonuclease deleted (*R.Eli8486ORFD*) possesses two cleavage sites within the *catP* gene, a widely-used selective marker which also forms the core of the mini-transposon system employed in this study, as noted above. Activity of this restriction endonuclease is likely to reduce initial establishment of the transposon vector, and may also cause post-transposition killing of the host cell.

Functional dissection of the P19 Gram-positive replicon generated an unintuitive result. A truncated variant of P19, P19C, was found to transfer into *E. limosum* roughly a hundred times as efficiently as P19. As the copy

number, stability and impact upon growth of this variant plasmid are no different to those of its progenitor, this increase in transfer efficiency was initially attributed to the absence of a cleavage site for an *E. limosum* endogenous methyltransferase, M.Eli8486ORFA. However, this gene is an 'orphan' methyltransferase with no associated restriction endonuclease activity. Transformation of pMTL83151, which possesses the same number of ORFA sites as pMTL87151C, does not display the same elevated transformation efficiency. This effect therefore remains unexplained, but may be related to the DNA-binding properties of the truncated iteron region.

During this study, both forward and reverse genetic methods have been demonstrated in *E. limosum* ATCC 8486. This strain, which is of considerable industrial interest due to its acetogenic metabolism and ability to metabolise methanol, is therefore readily available for future genome editing and transposon mutagenesis, both techniques having been demonstrated in this study. With refinements to the generation of transposon mutant libraries outlined in this study, the generation of large mutant libraries may be feasible. The functionality of the TetR inducible system has been demonstrated in this study, and may be incorporated into inducible gene systems for future synthetic biology work in *E. limosum*. The restriction endonuclease deletion mutant generated during this study, aside from demonstrating the utility of CRISPR-Cas9 genome editing in *E. limosum*, will also prove useful for future genetics work requiring efficient transformation of *E. limosum*.

Over the course of my project, *E. limosum* ATCC 8486 has been developed into a strain with relatively well-developed tools. Genetic modification of acetogens has so far been focused upon *C.ljundahlii* and *C. autoethanogenum*, with studies of *E. limosum* restricted to its fermentative performance and its genome. With the tools outlined above, extensive gene editing of this species should be possible, providing another option for industrial processes involving acetogenic bacteria. It is likely that large-scale transposon mutant libraries will be developed in the near future. Sequencing of these libraries will allow for the determination of essential genes under various culture conditions, and for comparison of their contributions to organism fitness under different conditions. This data may in turn guide further directed modification.

While a lesser contribution that the transposon mutagenesis and CRISPR tools outlined in this study, the P19 replicon has proved useful in a variety of clostridial species, and in *Bacillus subtilis*. Preliminary evidence suggests that even large plasmids replicating via P19 remain stable in the absence of selective pressure, which may facilitate the testing of synthetic operons in species in which chromosomal integration methods are not well-developed.

My project was conducted in parallel with that of Robert Mansfield, who has frequently been cited in this text.

Chapter Eight

Appendix

8 Appendix

8.1 P19 Modular Replicon Sequence

GGCGCGCCTTGATATGAAAAATTTAAGGAATAATTTTGGAAAAGTAAGGT
GTGAAAGATATAGTATAATTTATTTTAACAAATATTTTCAATTTTAGTTAA
GTGATATGACAATTTTTATTCCATAATGTTCATTTATGGTTTATAATAAGG
AAAAAAGGAAATAAAAAAAGTCCTCTATCGCTTAGCCAAAATGCGGGAA
CATTTTAGTTAAGTATATCGATCAGGTATAACAGGTACCTAAACGACTTC
CAGAGAACTTATTTTTACTATTTTTATATAGTAAATATACCAAATAAGTTA
CTGAAATTCAAGTCTTTTAGGTGAAATACCTAGAAGACTTTTTTATTGGAA
TGGAGTGGGATAGATGGACAACTTTGCATATTGTGATCATTTATTTGATA
GCAGCACAGATGGATACATACAATTGATGCAGATAAATGAAAAGAAAAT
AAAAGTTTATAATACCAAATATAAAGCACTAAGAACTGTTATAGAAGAG
GTAGAAGGGGAAAAAGATATTTACATATCAAGCAATACATATTACAAGC
CTAAACGTTCTACGAGCTATATAAGACAGTTTAGAGCATTATACATAGAT
TTGGATTTAAAGAAGTATGGAAAAAGTGAAGCTGTATATATAATATATGA
TTTAGTAAATAAAAATAAAATACCTACTCCCAGTATGATAGTTGATTCAG
GAAGAGGAATTCATTTGTATTGGAGAATTAAAAATGCACCTTATGGAGCT
ATACATACATGGCAAGAATTAGAAGACTACCTCTATAATCAATTAAAAGT
TTTAGGGGCAGATATTCAAGCTACAGATAGTGTAAGAGTTTTAAGATTAC
CAAATACTGTAAATAGTAGAATAAATGAAGAATGTAAAGTATTAATTATA
AATGATGATATAGAATACTCTATGTATGATTTAAGAGAAGAATTCCTGAA
TTACAGTTCAAAAAAAGCGTATCAATTAGAGTTTGAGCAAACTAAGATAA
ATAAACAGAAGAAACAAAGTGTAATAAGTAGATTATTTAATAGTTATACA
CTACATTTAGCAAGGGTTGGAGATATAGAAACACTTTGTAAGTTGAGAAG

CTATAATGTTAAAGGATATAGAAATTTTATTTTACATTGTTATGCATATTG

GAAGGGAATATATACAAGAGATGTTGAAGAGCTTGAAAAAGAAGTAATA

GAGTTCAATAATTCATTTAGAAATCCGTTAAGATCTTCAGAAGTAAATGC

AATATTAAGATGTATTCCAAAGGCTATAAGTAAATTTGTAGAGTATGAAC

AGGGATTAAGAAATGGTGAAACTAAAAGAGTAACTAAGGGCATGAGCGA

CAAAGGTGGCTACTGGTATAAAAATCAAACATTAATTGAAAGATTAGATA

TAACATTAGATGAACAAAAGCATATGAAAACAATCATAAGCCAGGAAGA

GAAATATAGAAGAAATAATGATAAAAGAAAGGCCAGTAGAAGGAATGA

GGAAGGATTGACACAAAAGCAATTAGAAATTAAGCAATTAATGGAGCAG

GTAAAGAAATTAAAGGAACAGGGGTTAAATAATAGTGAAATAGCAAGAG

CTTTAGATATAAATAGAACAAAGGTTATAAGATTATTAAAATAGGGTGTT

CAAAAAATGGCTCTTTATAAATATGTGTTGTGCTTTGGCAAATTTAGAGTT

TTCTTTAGGGAGATCTACTGGGGTATAGCGTGAACGGGCAAAGTACTCGT

AATTGATTTTTGTTTTTTATTAAAAAAGATATAGAATAAATAGCTATATCT

TTTTATTTATTCTTACTTTTTTATGGGTGTTCAAAAAATGGCTCTTTATAAA

TATGTGTTATGCTTTGGCAAATTTAGAGTTTTCTTTAGGGAAATCTAGGGG

GTATAGCGTGAACGGCCCAAATGGCTATAATTAGGTTTGGTTAAGGCCGG

CC

Figure 8.1: Sequence of the P19 modularised replicon. AscI and FseI sites are highlighted in red. Iteron sequences A, B and C are highlighted in cyan, yellow and green, respectively. The sequence of the *rep* gene, from start codon to end codon, is underlined. The P19A truncated variant replicon is identical to P19 as depicted in this figure, except that the promoter region (5' of the *rep* ORF) has been removed. The P19C truncated variant replicon is identical to the sequence depicted in this figure, except that all iterons are removed and the stop codon of the *rep* is adjacent to the FseI site.

GGTGTTCAAAAAATGGCTCTTTATAAATATGTGTTGTGCTTTGGCA

AATTTAGAGTTTTCTTTAGGGAGATCTACTGGGGTATAGCGTGAAC

GGGCAAAGTACTCGTAATTGATTTTTGTTTTTTATTAAAAAAGATAT

AGAGGCCGGCC

Figure 8.2: Sequence of the iteron region of the P19B modularised replicon. Sequences are highlighted as above. The 5' region of this replicon is identical to that of P19, depicted above.

8.2 Transposon Insertion Sites Sequenced during this Study

Table 8.1: Transposon insertion sites identified during sequencing of individual transposon mutants via inverse PCR. Insertion sites are positioned relative to an annotated locus (5', 3' or within the ORF of a gene).

CLONE	5'/3'/ORF	LOCUS	BLASTP
1	5'	Ga0111251_111929	DNA Polymerase III Subunit Epsilon
2	5'	rnr gene	Ribonuclease R
3	ORF	Ga0111251_112082	Hypothetical protein [E. limosum] (87% ident)
4	ORF	Ga0111251_113869	diguanylate cyclase [Eubacterium limosum]
5	ORF	Ga0111251_11171	GGDEF domain-containing protein [Eubacterium limosum]
6	ORF	Ga0111251_114076	hypothetical protein ELI_4407 [Eubacterium limosum KIST612]
7	Inter		
8	ORF	Ga0111251_111575	PTS galactitol transporter subunit IIC [Eubacterium limosum]
9	ORF	nfo gene	endonuclease IV [Eubacterium limosum]
10	ORF	Ga0111251_11908	OmpR DNA-binding response regulator [Eubacterium limosum]
11	N/A	"Repeat region"	
12	ORF	Ga0111251_111503	tellurium resistance protein [Eubacterium limosum]
13	ORF	Ga0111251_111971	EamA domain-containing membrane protein RarD [Eubacterium callanderi]

14	ORF	Ga0111251_113850	NADPH-dependent 2,4-dienoyl-CoA reductase, sulfur reductase [Eubacterium callanderi]
15			
16	ORF	Ga0111251_111413	HAD family hydrolase [Eubacterium limosum]
17	ORF	Ga0111251_113962	MULTISPECIES: uroporphyrinogen decarboxylase [Robinsoniella]
18	ORF	Ga0111251_114030	hypothetical protein [Eubacterium limosum]
19	ORF	Ga0111251_111657	hypothetical protein [Eubacterium limosum]
20	5'	Ga0111251_11973	
21	ORF	Ga0111251_11317	MULTISPECIES: Fis family transcriptional regulator [Clostridiales]
22	ORF	Ga0111251_11980	NAD/NADP octopine/nopaline dehydrogenase [Eubacterium limosum]
23	ORF	Ga0111251_111354	Sporulation protein YtaF
24	ORF	Ga0111251_112554	[Transketolase]
25	ORF	Ga0111251_11352	Membrane protein
26	ORF	Ga0111251_111494	Hypothetical protein
27	ORF	Ga0111251_114106	Aspartate ammonia-lyase
28	ORF	Ga0111251_114069	Putative ABC exporter
29	ORF	Ga0111251_11609	propionaldehyde dehydrogenase PduP
30	ORF	Ga0111251_11686	capsular biosynthesis protein
31	ORF	Ga0111251_11313	ABC transporter ATP-binding protein
32	ORF	Ga0111251_111294	Hypothetical protein
33	Inter		
34	ORF	dsdA gene	D-serine ammonia lyase
35	5'	Ga0111251_111418	MULTISPECIES: membrane protein [Clostridiales]
36	ORF	Ga0111251_111445	Predicted DNA-binding transcriptional regulator YafY, contains an HTH and WYL domains [Eubacterium callanderi]
37	ORF	Ga0111251_113815	hypothetical protein [Eubacterium limosum]
38	ORF	Ga0111251_113673	hypothetical protein [Eubacterium limosum]
39	ORF	Ga0111251_114120	ABC transporter ATP-binding protein [Eubacterium limosum]
40	??	??	??

41	5'	Ga0111251_111413	HAD family hydrolase [Eubacterium limosum]
42	ORF	Ga0111251_112361	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase [Eubacterium limosum]
43	ORF	Ga0111251_111850	GntR family transcriptional regulator [Eubacterium limosum]
44	ORF	Ga0111251_111494	Hypothetical protein
45	ORF	queA gebe	S-adenosylmethionine:tRNA ribosyltransferase-isomerase [Eubacterium limosum]
46	ORF	Ga0111251_112319	FMN reductase [Eubacterium limosum]
47	ORF	Ga0111251_111097	hypothetical protein [Eubacterium limosum]
48	ORF	Ga0111251_111723	hypothetical protein [Eubacterium limosum]
49	ORF	Ga0111251_11528	cobalamin B12-binding domain protein [Eubacterium limosum KIST612]
50	ORF	Ga0111251_114177	ABC transporter [Eubacterium limosum]
51	ORF	Ga0111251_11296	hypothetical protein [Eubacterium limosum]
52	5'	Ga0111251_11611	hypothetical protein [Eubacterium limosum]
53	ORF	Ga0111251_113418	UDP-glucose 6-dehydrogenase [Eubacterium limosum]
54	ORF	Ga0111251_112082	hypothetical protein [Eubacterium limosum]
55	5'	Ga0111251_113033	tryptophan-rich sensory protein [Eubacterium limosum]
56	3'	Ga0111251_11268	IclR family transcriptional regulator [Eubacterium limosum]
57	ORF	Ga0111251_113816	hypothetical protein [Eubacterium limosum]
58	5'	Ga0111251_113774	SH3 domain-containing protein [Eubacterium callanderi]
59	ORF	Ga0111251_11917	AI-2E family transporter [Eubacterium limosum]
60	ORF	Ga0111251_11995	hypothetical protein [Eubacterium limosum]
61	ORF	msrA CDS	peptide-methionine (S)-S-oxide reductase [Eubacterium limosum]

References

9 References

AL-HINAI, M. A., FAST, A. G. & PAPOUTSAKIS, E. T. 2012. Novel system for efficient isolation of *Clostridium* double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration. *Appl Environ Microbiol,* 78**,** 8112-21.

ALLEN, S. P. & BLASCHEK, H. P. 1990. Plasmid copy number and stability determination in *Clostridium perfringens* transformants. *FEMS Microbiol Lett,* 72**,** 323-327.

ANAGNOSTOPOULOS, C. & SPIZIZEN, J. 1961. Requirements for transformation in *Bacillus subtilis*. *J Bacteriol*, 81(5), 741-746.

AUNE, T. E. & AACHMANN, F. L. 2010. Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Appl Microbiol Biotechnol,* 85**,** 1301-13.

BAE, T., BANGER, A. K., WALLACE, A., GLASS, E. M., ASLUND, F., SCHNEEWIND, O. & MISSIAKAS, D. M. 2004. *Staphylococcus aureus* virulence genes identified by bursa aurealis mutagenesis and nematode killing. *Proc Natl Acad Sci U S A,* 101**,** 12312-7.

BAIR, C. L. & BLACK, L. W. 2007. A Type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J Mol Biol,* 366**,** 768-778.

BANERJEE, A., LEANG, C., UEKI, T., NEVIN, K. P. & LOVLEY, D. R. 2014. Lactose-inducible system for metabolic engineering of *Clostridium ljungdahlii*. *Appl Environ Microbiol,* 80**,** 2410-6.

BARDAJI, L., ANORGA, M., RUIZ-MASO, J. A., DEL SOLAR, G. & MURILLO, J. 2017. Plasmid Replicons from *Pseudomonas* Are Natural Chimeras of Functional, Exchangeable Modules. *Front Microbiol,* 8**,** 190.

BARQUIST, L., LANGRIDGE, G. C., TURNER, D. J., PHAN, M. D., TURNER, A. K., BATEMAN, A., PARKHILL, J., WAIN, J. & GARDNER, P. P. 2013. A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res,* 41**,** 4549-64.

BERENS, C. & HILLEN, W. 2003. Gene regulation by tetracyclines. Constraints of resistance regulation in bacteria shape TetR for application in eukaryotes. *Eur J Biochem,* 270**,** 3109-3121.

BERG, P. & MERTZ, J. E. 2010. Personal reflections on the origins and emergence of recombinant DNA technology. *Genetics,* 184**,** 9-17.

BERTRAM, R. & HILLEN, W. 2008. The application of Tet repressor in prokaryotic gene regulation and expression. *Microb Biotechnol,* 1**,** 2-16.

BERTSCH, J. & MULLER, V. 2015. Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. *Biotechnol Biofuels,* 8**,** 210.

BESSIS, S., AMADOU, T., DUBOURG, G., RAOULT, D. & FOURNIER, P. E. 2016. "*Bariatricus massiliensis*" as a new bacterial species from human gut microbiota. *New Microbes New Infect,* 12**,** 54-5.

BOEKE, J. D., TRUEHEART, J., NATSOULIS, G. & FINK, G. R. 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol,* 154**,** 164-175.

BOLOTIN, A., QUINQUIS, B., SOROKIN, A. & EHRLICH, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology,* 151**,** 2551-61.

BRON, S. & LUXEN, E. 1985. Segregational Instability of pUB110-Derived Recombinant Plasmids in *Bacillus subtilis*. *Plasmid,* 14**,** 235-244.

BRUANT, G., LEVESQUE, M. J., PETER, C., GUIOT, S. R. & MASSON, L. 2010. Genomic analysis of carbon monoxide utilization and butanol production by *Clostridium carboxidivorans* strain P7. *PLoS One,* 5**,** e13033.

BRYAN, G., GARZA, D. & HARTL, D. 1990. Insertion and Excision of the Transposable Element *mariner* in *Drosophila*. *Genetics,* 125**,** 103-114.

CARATTOLI, A., ZANKARI, E., GARCIA-FERNANDEZ, A., LARSEN, M. V., LUND, O., VILLA, L., AARESTRUP, F. M. & HASMAN, H. 2014. *In Silico* Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrob. Agents Chemother,* 58**,** 3895-3903.

CARTMAN, S. T., KELLY, M. L., HEEG, D., HEAP, J. T. & MINTON, N. P. 2012. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. *Appl Environ Microbiol,* 78**,** 4683-90.

CARTMAN, S. T. & MINTON, N. P. 2010. A mariner-based transposon system for in vivo random mutagenesis of *Clostridium difficile*. *Appl Environ Microbiol,* 76**,** 1103-9.

CASS, S. D., HAAS, K. A., STOLL, B., ALKHNBASHI, O. S., SHARMA, K., URLAUB, H., BACKOFEN, R., MARCHFELDER, A. & BOLT, E. L. 2015. The role of Cas8 in type I CRISPR interference. *Biosci Rep,* 35.

CHANG, I. S., KIM, B. H., LOVITT, R. W. & BANG, J. S. 2001. Effect of CO partial pressure on cell-recycled continuous CO fermentation by *Eubacterium limosum* KIST612. *Process Biochem,* 37**,** 411-421.

CHAO, M. C., ABEL, S., DAVIS, B. M. & WALDOR, M. K. 2016. The design and analysis of transposon insertion sequencing experiments. *Nat Rev Microbiol,* 14**,** 119-28.

CHARPENTIER, E. & MARRAFFINI, L. A. 2014. Harnessing CRISPR-Cas9 immunity for genetic engineering. *Curr Opin Microbiol,* 19**,** 114-9.

CHOPRA, I. & ROBERTS, M. 2001. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol Rev,* 65**,** 232-60 ; second page, table of contents.

CHUBIZ, L. M., PURSWANI, J., CARROLL, S. M. & MARX, C. J. 2013. Novel inducible expression vectors for use in *Methylobacterium extorquens. BMC Res Notes,* 6.

COBB, R. E., WANG, Y. & ZHAO, H. 2015. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth Biol,* 4**,** 723-8.

DEL SOLAR, G., GIRALDO, R., RUIZ-ECHEVARRIA, J., ESPINOSA, M. & DIAZ-OREJAS, R. 1998. Replication and Control of Circular Bacterial Plasmids. *Microbiol Mol Biol Rev,* 62**,** 434-464.

DEMBEK, M., BARQUIST, L., BOINETT, C. J., CAIN, A. K., MAYHO, M., LAWLEY, T. D., FAIRWEATHER, N. F. & FAGAN, R. P. 2015. High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *mBio,* 6**,** e02383.

DMOWSKI, M. & JAGURA-BURDZY, G. 2013. Active Stable Maintenance Functions in Low Copy-Number Plasmids of Gram-Positive Bacteria. *Pol J Microbiol,* 62**,** 3-16.

DONG, H., TAO, W., ZHANG, Y. & LI, Y. 2012. Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering. *Metab Eng,* 14**,** 59-67.

DRAKE, H. L., GOSSNER, A. S. & DANIEL, S. L. 2008. Old acetogens, new light. *Ann N Y Acad Sci,* 1125**,** 100-28.

EGGERTH, A. H. 1935. The Gram-positive Non-spore-bearing Anaerobic Bacilli of Human Feces. *J Bacteriol,* 30**,** 277-299.

EHRT, S., GUO, X. V., HICKEY, C. M., RYOU, M., MONTELEONE, M., RILEY, L. W. & SCHNAPPINGER, D. 2005. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res,* 33**,** e21.

FAGAN, R. P. & FAIRWEATHER, N. F. 2011. *Clostridium difficile* has two parallel and essential Sec secretion systems. *J Biol Chem,* 286**,** 27483-93.

FILUTOWICZ, M., DELLIS, S., LEVCHENKO, I., URH, M., WU, F. & YORK, D. 1994. Regulation of Replication of an Iteron-containing DNA Molecule. *Prog Nucleic Acid Res Mol Biol*, 48**,** 239-273.

FUJITANI, Y., YAMAMOTO, K. & KOBAYASHI, I. 1995. Dependence of Frequency of Homologous Recombination on the Homology Length. *Genetics,* 140**,** 797-809.

GARNIER, T. & COLE, S. T. 1988. Identification and molecular genetic analysis of replication functions of the bacteriocinogenic plasmid pIP404 from *Clostridium perfringens*. *Plasmid,* 19**,** 151-160.

GAWRONSKI, J. D., WONG, S. M., GIANNOUKOS, G., WARD, D. V. & AKERLEY, B. J. 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proc Natl Acad Sci U S A,* 106**,** 16422-7.

GENTHNER, B. R. S. & BRYANT, M. P. 1982. Growth of *Eubacterium limosum* with Carbon Monoxide as the Energy Source. *Appl Environ Microbiol,* 43**,** 70-74.

GENTHNER, B. R. S., DAVIS, C. L. & BRYANT, M. P. 1981. Features of Rumen and Sewage Sludge Strains of *Eubacterium limosum*, a Methanol- and H2-C02-Utilizing Species. *Appl Environ Microbiol,* 42**,** 12-19.

GERDES, K., LARSEN, J. E. L. & MOLIN, S. 1985. Stable Inheritance of Plasmid R1 Requires Two Different Loci. *J Bacteriol,* 161**,** 292-298.

GOODMAN, A. L., MCNULTY, N. P., ZHAO, Y., LEIP, D., MITRA, R. D., LOZUPONE, C. A., KNIGHT, R. & GORDON, J. I. 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe,* 6**,** 279-89.

GOSSEN, M. & BUJARD, H. 1993. Anhydrotetracycline, a novel effector for tetracycline controlled gene expression systems in eukaryotic cells. *Nucleic Acids Res,* 21.

GRABHERR, R., NILSSON, E., STRIEDNER, G. & BAYER, K. 2002. Stabilizing Plasmid Copy Number to Improve Recombinant Protein Production. *Biotechnol. Bioeng,* 77**,** 142-147.

GRKOVIC, S., BROWN, M. H. & SKURRAY, R. A. 2002. Regulation of Bacterial Drug Export Systems. *Microbiol Mol Bio Rev,* 66**,** 671-701.

GROHMANN, E., GOESSWEINER-MOHR, N. & BRANTL, S. 2016. DNA-Binding Proteins Regulating pIP501 Transfer and Replication. *Front Mol Biosci,* 3**,** 42.

GROOT, M. N., NIEBOER, F. & ABEE, T. 2008. Enhanced transformation efficiency of recalcitrant *Bacillus cereus* and *Bacillus weihenstephanensis* isolates upon in vitro methylation of plasmid DNA. *Appl Environ Microbiol,* 74**,** 7817-20.

GROSSE-HONEBRINK, A., SCHWARZ, K. M., WANG, H., MINTON, N. P. & ZHANG, Y. 2017. Improving gene transfer in *Clostridium pasteurianum* through the isolation of rare hypertransformable variants. *Anaerobe*.

HAFT, D. H., SELENGUT, J., MONGODIN, E. F. & NELSON, K. E. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol,* 1**,** e60.

HARRISON, P. W., LOWER, R. P., KIM, N. K. & YOUNG, J. P. 2010. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol,* 18**,** 141-8.

HEAP, J. T., EHSAAN, M., COOKSLEY, C. M., NG, Y. K., CARTMAN, S. T., WINZER, K. & MINTON, N. P. 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res,* 40**,** e59.

HEAP, J. T., PENNINGTON, O. J., CARTMAN, S. T., CARTER, G. P. & MINTON, N. P. 2007. The ClosTron: a universal gene knock-out system for the genus Clostridium. *J Microbiol Methods,* 70**,** 452-64.

HEAP, J. T., PENNINGTON, O. J., CARTMAN, S. T. & MINTON, N. P. 2009. A modular system for *Clostridium* shuttle plasmids. *J Microbiol Methods,* 78**,** 79-85.

HEATH, D. G., AN, F. Y., WEAVER, K. E. & CLEWELL, D. B. 1995. Phase Variation of *Enterococcus faecalis* pAD1 Conjugation Functions Relates to Changes in Iteron Sequence Region. *J Bacteriol,* 177**,** 5453-5459.

HILLEN, W. & BERENS, C. 1994. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu Rev Microbiol,* 48**,** 345-369.

HORVATH, P. & BARRANGOU, R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science,* 327**,** 167-70.

HUSSAIN, H. A., ROBERTS, A. P. & 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**,** 137-41.

JACOBSON, J. W., MEDHORA, M. M. & HARTL, D. 1986. Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc Natl Acad Sci U S A,* 83**,** 8684-8688.

JANG, N., YASIN, M., PARK, S., LOVITT, R. W. & CHANG, I. S. 2017. Determination of volumetric gas-liquid mass transfer coefficient of carbon monoxide in a batch cultivation system using kinetic simulations. *Bioresour Technol,* 239**,** 387-393.

JEONG, J., BERTSCH, J., HESS, V., CHOI, S., CHOI, I. G., CHANG, I. S. & MULLER, V. 2015. Energy Conservation Model Based on Genomic and Experimental Analyses of a Carbon Monoxide-Utilizing, Butyrate-Forming Acetogen, *Eubacterium limosum* KIST612. *Appl Environ Microbiol,* 81**,** 4782-90.

JIANG, W., BIKARD, D., COX, D., ZHANG, F. & MARRAFFINI, L. A. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol,* 31**,** 233-9.

JIANG, Y., CHEN, B., DUAN, C., SUN, B., YANG, J. & YANG, S. 2015. Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl Environ Microbiol,* 81**,** 2506-14.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science,* 337**,** 816-821.

KARBERG, M., GUO, H., ZHONG, J., COON, R., PERUTKA, J. & LAMBOWITZ, A. M. 2001. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat Biotechnol,* 19**,** 1162-1167.

KELLY, W. J., HENDERSON, G., PACHECO, D. M., LI, D., REILLY, K., NAYLOR, G. E., JANSSEN, P. H., ATTWOOD, G. T., ALTERMANN, E. & LEAHY, S. C. 2016. The complete genome sequence of *Eubacterium limosum* SA11, a metabolically versatile rumen acetogen. *Stand Genomic Sci,* 11**,** 26.

KERAVALA, A., LIU, D., LECHMAN, E. R., WOLFE, D., NASH, J. A., LAMPE, D. J. & ROBBINS, P. D. 2006. Hyperactive *Himar1* Transposase Mediates Transposition in Cell Culture and Enhances Gene Expression *In Vivo*. *Hum Gene Ther,* 17**,** 1006-1018.

KLEIN, B. A., DUNCAN, M. J. & HU, L. T. 2015. Defining essential genes and identifying virulence factors of *Porphyromonas gingivalis* by massively parallel sequencing of transposon libraries (Tn-seq). *Methods Mol Biol,* 1279**,** 25-43.

KOLTER, R. & HELINSKI, D. R. 1978. Activity of the Replication Terminus of Plasmid R6K in Hybrid Replicons in *Escherichia coli*. *J. Mol. Biol.,* 124**,** 425-441.

KOPKE, M., HELD, C., HUJER, S., LIESEGANG, H., WIEZER, A., WOLLHERR, A., EHRENREICH, A., LIEBL, W., GOTTSCHALK, G. & DURRE, P. 2010. *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc Natl Acad Sci U S A,* 107**,** 13087-92.

LAMM, L., HORIG, J. A., RENZ, P. & HECKMANN, G. 1980. Biosynthesis of Vitamin B12. *Eur J Biochem,* 109**,** 115-118.

LAMPE, D. J., CHURCHILL, M. E. A. & ROBERTSON, H. M. 1996. A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. *EMBO J,* 15**,** 5470-5479.

LANCKRIET, A., TIMBERMONT, L., HAPPONEN, L. J., PAJUNEN, M. I., PASMANS, F., HAESEBROUCK, F., DUCATELLE, R., SAVILAHTI, H. & VAN IMMERSEEL, F. 2009. Generation of single-copy transposon insertions in *Clostridium perfringens* by electroporation of phage mu DNA transposition complexes. *Appl Environ Microbiol,* 75**,** 2638-42.

LANGRIDGE, G. C., PHAN, M. D., TURNER, D. J., PERKINS, T. T., PARTS, L., HAASE, J., CHARLES, I., MASKELL, D. J., PETERS, S. E., DOUGAN, G., WAIN, J., PARKHILL, J. & TURNER, A. K. 2009. Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res,* 19**,** 2308-16.

LEANG, C., UEKI, T., NEVIN, K. P. & LOVLEY, D. R. 2013. A Genetic System for *Clostridium ljungdahlii*: a Chassis for Autotrophic Production of Biocommodities and a Model Homoacetogen. *Appl Environ Microbiol,* 79**,** 1102-1109.

LEE, S. Y., MERMELSTEIN, L. D. & PAPOUTSAKIS, E. T. 1993. Determination of plasmid copy number and stability in *Clostridium acetobutylicum* ATCC 824. *FEMS Microbiol Lett,* 108**,** 319-323.

LESIAK, J. M., LIEBL, W. & EHRENREICH, A. 2014. Development of an in vivo methylation system for the solventogen *Clostridium saccharobutylicum* NCP 262 and analysis of two endonuclease mutants. *J Biotechnol,* 188**,** 97-9.

LI, H. & LIAO, J. C. 2015. A synthetic anhydrotetracycline-controllable gene expression system in *Ralstonia eutropha* H16. *ACS Synth Biol,* 4**,** 101-6.

LI, Q., CHEN, J., MINTON, N. P., ZHANG, Y., WEN, Z., LIU, J., YANG, H., ZENG, Z., REN, X., YANG, J., GU, Y., JIANG, W., JIANG, Y. & YANG, S. 2016. CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnol J,* 11**,** 961-72.

LIEW, F., HENSTRA, A. M., KPKE, M., WINZER, K., SIMPSON, S. D. & MINTON, N. P. 2017. Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Metab Eng,* 40**,** 104-114.

LIEW, F., MARTIN, M. E., TAPPEL, R. C., HEIJSTRA, B. D., MIHALCEA, C. & KOPKE, M. 2016. Gas Fermentation-A Flexible Platform for Commercial Scale Production of Low-Carbon-Fuels and Chemicals from Waste and Renewable Feedstocks. *Front Microbiol,* 7**,** 694.

LIOU, J. S., BALKWILL, D. L., DRAKE, G. R. & TANNER, R. S. 2005. *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *Int J Syst Evol Microbiol,* 55**,** 2085-91.

LIU, H., BOUILLAUT, L., SONENSHELM, A. L. & MELVILLE, S. B. 2013. Use of a Mariner-Based Transposon Mutagenesis System To Isolate *Clostridium perfringens* Mutants Deficient in Gliding Motility. *J Bacteriol,* 195**,** 629-636.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.

LJUNGDAHL, L. G. 1986. The Autotrophic Pathway of Acetate Synthesis in Acetogenic Bacteria. *Annu Rev Microbiol,* 40**,** 415-450.

LOENEN, W. A., DRYDEN, D. T., RALEIGH, E. A. & WILSON, G. G. 2014. Type I restriction enzymes and their relatives. *Nucleic Acids Res,* 42**,** 20-44.

LOENEN, W. A. & RALEIGH, E. A. 2014. The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res,* 42**,** 56-69.

MAKAROVA, K. S., GRISHIN, N. V., SHABALINA, S. A., WOLF, Y. I. & KOONIN, E. V. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct,* 1**,** 7.

MCCLAIN, M. S., DUNCAN, S. S., GADDY, J. A. & COVER, T. L. 2013. Control of gene expression in *Helicobacter pylori* using the Tet repressor. *J Microbiol Methods,* 95**,** 336-41.

MILLION-WEAVER, S. & CAMPS, M. 2014. Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? *Plasmid,* 75**,** 27-36.

MINTON, N. P., EHSAAN, M., HUMPHREYS, C. M., LITTLE, G. T., BAKER, J., HENSTRA, A. M., LIEW, F., KELLY, M. L., SHENG, L., SCHWARZ, K. & ZHANG, Y. 2016. A roadmap for gene system development in *Clostridium. Anaerobe,* 41**,** 104-112.

MINTON, N. P. & MORRIS, J. G. 1981. Isolation and Partial Characterization of Three Cryptic Plasmids from Strains of *Clostridium butyricum*. *J Gen Microbiol,* 127**,** 325-331.

MIYAMOTO, K., SEIKE, S., TAKAGISHI, T., OKUI, K., ODA, M., TAKEHARA, M. & NAGAHAMA, M. 2015. Identification of the replication region in pBCNF5603, a bacteriocin-encoding plasmid, in the enterotoxigenic *Clostridium perfringens* strain F5603. *BMC Microbiol,* 15**,** 118.

MOJICA, F. J., DIEZ-VILLASENOR, C., GARCIA-MARTINEZ, J. & ALMENDROS, C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology,* 155**,** 733-40.

MUNASINGHE, P. C. & KHANAL, S. K. 2010. Biomass-derived syngas fermentation into biofuels: Opportunities and challenges. *Bioresour Technol,* 101**,** 5013-22.

NAKAYAMA, Y. & KOBAYASHI, I. 1998. Restriction-modification gene complexes as selfish gene entities: Roles of a regulatory system in their establishment, maintenance, and apoptotic mututal exclusion. *Proc Natl Acad Sci U S A,* 95**,** 6442-6447.

NG, Y. K., EHSAAN, M., PHILIP, S., COLLERY, M. M., JANOIR, C., COLLIGNON, A., CARTMAN, S. T. & MINTON, N. P. 2013. Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using pyrE alleles. *PLoS One,* 8**,** e56051.

NORDSTROM, K. 2006. Plasmid R1--replication and its control. *Plasmid,* 55**,** 1-26.

NORDSTRÖM, K. 1990. Control of Plasmid Replication - How Do DNA Iterons Set the Replication Frequency? *Cell* 63**,** 1121-1124.

OBRIST, M. W. & MILLER, V. L. 2012. Low copy expression vectors for use in *Yersinia* sp. and related organisms. *Plasmid,* 68**,** 33-42.

PAPOUTSAKIS, E. T. 2008. Engineering solventogenic clostridia. *Curr Opin Biotechnol,* 19**,** 420-9.

PINGOUD, A. & JELTSCH, A. 1997. Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur J Biochem,* 246**,** 1-22.

PINGOUD, A., WILSON, G. G. & WENDE, W. 2014. Type II restriction endonucleases--a historical perspective and more. *Nucleic Acids Res,* 42**,** 7489-527.

PLOTKA, M., WOZNIAK, M. & KACZOROWSKI, T. 2017. Quantification of Plasmid Copy Number with Single Colour Droplet Digital PCR. *PLoS One,* 12**,** e0169846.

PROVIDENTI, M. A., O'BRIEN, J. M., EWING, R. J., PATERSON, E. S. & SMITH, M. L. 2006. The copy-number of plasmids and other genetic elements can be determined by SYBR-Green-based quantitative real-time PCR. *J Microbiol Methods,* 65**,** 476-87.

PYNE, M. E., MOO-YOUNG, M., CHUNG, D. A. & CHOU, C. P. 2013. Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. *Biotechnol Biofuels,* 6.

PYNE, M. E., MOO-YOUNG, M., CHUNG, D. A. & CHOU, C. P. 2014. Expansion of the genetic toolkit for metabolic engineering of *Clostridium pasteurianum*: chromosomal gene disruption of the endogenous CpaAI restriction enzyme. *Biotechnol Biofuels,* 7.

PYNE, M. E., MOO-YOUNG, M., CHUNG, D. A. & CHOU, C. P. 2015. Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in Escherichia coli. *Appl Environ Microbiol,* 81**,** 5103-14.

RAGSDALE, S. W. 2008. Enzymology of the wood-Ljungdahl pathway of acetogenesis. *Ann N Y Acad Sci,* 1125**,** 129-36.

RAO, D. N., DRYDEN, D. T. & BHEEMANAIK, S. 2014. Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res,* 42**,** 45-55.

RODE, L. M., GENTHNER, B. R. S. & BRYANT, M. P. 1981. Syntrophic Association by Cocultures of the Methanol- and C02-H2-Utilizing Species *Eubacterium limosum* and Pectin Fermenting

Lachnospira multiparus During Growth in a Pectin Medium. *Appl Environ Microbiol,* 42**,** 20-22.
ROH, H., KO, H. J., KIM, D., CHOI, D. G., PARK, S., KIM, S., CHANG, I. S. & CHOI, I. G. 2011. Complete genome sequence of a carbon monoxide-utilizing acetogen, *Eubacterium limosum* KIST612. *J Bacteriol,* 193**,** 307-8.
SAKAI, Y., KARAZIMOTO, T. & TANI, Y. 1991. Transformation System for an Asporogenous Methylotrophic Yeast, *Candida boidinii*: Cloning of the Orotidine-5'-Phosphate Decarboxylase Gene (*URA3*), Isolation of Uracil Auxotrophic Mutants, and Use of the Mutants for Integrative Transformation. *J Bacteriol,* 173**,** 7458-7463.
SCHIEL-BENGELSDORF, B. & DURRE, P. 2012. Pathway engineering and synthetic biology using acetogens. *FEBS Lett,* 586**,** 2191-8.
SCHUCHMANN, K. & MULLER, V. 2014. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol,* 12**,** 809-21.
SCHWARZ, K. M., GROSSE-HONEBRINK, A., DERECKA, K., ROTTA, C., ZHANG, Y. & MINTON, N. P. 2017. Towards improved butanol production through targeted genetic modification of Clostridium pasteurianum. *Metab Eng,* 40**,** 124-137.
SILVA, F., QUEIROZ, J. A. & DOMINGUES, F. C. 2012. Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. *Biotechnol Adv,* 30**,** 691-708.
SONG, Y. & CHO, B. K. 2015. Draft Genome Sequence of Chemolithoautotrophic Acetogenic Butanol-Producing Eubacterium limosum ATCC 8486. *Genome Announc,* 3.
TRACY, B. P., JONES, S. W., FAST, A. G., INDURTHI, D. C. & PAPOUTSAKIS, E. T. 2012. Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Curr Opin Biotechnol,* 23**,** 364-81.
VAN DER WIELEN, P. 2002. *Clostridium lactatifermentans* sp. nov., a lactate-fermenting anaerobe isolated from the caeca of a chicken. *International Journal of Systematic and Evolutionary Microbiology,* 52**,** 921-925.
VAN OPIJNEN, T., BODI, K. L. & CAMILLI, A. 2009. Tn-seq; high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods,* 6**,** 767-772.
VAN OPIJNEN, T. & CAMILLI, A. 2013. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol,* 11.
VIDAL, J. E., CHEN, J., LI, J. & MCCLANE, B. A. 2009. Use of an EZ-Tn5-based random mutagenesis system to identify a novel toxin regulatory locus in *Clostridium perfringens* strain 13. *PLoS One,* 4**,** e6232.
WALKER, D. J. F. & KOEPKE, M. 2016. Method of producing a recombinant microorganism. Google Patents.
WANG, H., SMITH, M. C. & MULLANY, P. 2006. The conjugative transposon Tn5397 has a strong preference for integration into its *Clostridium difficile* target site. *J Bacteriol,* 188**,** 4871-8.
WANG, S., HUANG, H., KAHNT, J., MUELLER, A. P., KOPKE, M. & THAUER, R. K. 2013. NADP-Specific Electron-Bifurcating [FeFe]-Hydrogenase in a Functional Complex with Formate Dehydrogenase in *Clostridium autoethanogenum* Grown on CO. *J Bacteriol,* 195**,** 4373-4386.
WARD, D. M., MAH, R. A. & KAPLAN, I. R. 1978. Methanogenesis from Acetate: a Nonmethanogenic Bacterium from an Anaerobic Acetate Enrichment. *Applied and Environmental Microbiology,* 35**,** 1185-1192.
WIEDENHEFT, B., ZHOU, K., JINEK, M., COYLE, S. M., MA, W. & DOUDNA, J. A. 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure,* 17**,** 904-12.
XU, D., TREE, D. R. & LEWIS, R. S. 2011. The effects of syngas impurities on syngas fermentation to liquid fuels. *Biomass Bioenergy,* 35**,** 2690-2696.

YAJJALA, V. K., WIDHELM, T. J., ENDRES, J. L., FEY, P. D. & BAYLES, K. W. 2016. Generation of a Transposon Mutant Library in *Staphylococcus aureus* and *Staphylococcus epidermidis* Using bursa aurealis. *Methods Mol Biol,* 1373**,** 103-10.

ZHANG, Y., GROSSE-HONEBRINK, A. & MINTON, N. P. 2015. A universal mariner transposon system for forward genetic studies in the genus *Clostridium*. *PLoS One,* 10**,** e0122411.