Forward Genetics in Clostridial Acetogens as a Route to Process Improvements

Craig Alexander Woods

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.

Craig Alexander Woods,

September 2017

Abstract

Clostridium autoethanogenum is able to use CO or CO_2 + H_2 as a sole carbon and energy source and natively produces the industrially relevant compounds ethanol, 2,3-butanediol and acetate. Increasing concerns about energy security, environmental impacts of fossil fuel usage and long term fuel and chemical production have led to the commercialisation of biological production of fuels and chemicals. *C. autoethanogenum* has been adopted as the organism of choice by a leading biotechnology company which has now reached full commercial-scale ethanol production using industrial waste gases as feedstock in a process termed gas fermentation. Strain development is a crucial process to the further adoption of this gas fermentation technology.

An important step in strain development is establishing reliable and efficient DNA transfer protocols. In this study a novel *E. coli* conjugal donor strain was constructed which offered up to 744-fold increases in conjugation efficiency over an established donor strain. Improvements in DNA transfer aided forward genetics using the *mariner* element *himar1*. Transposon mutants were identified including strains deficient in amino acid biosynthesis and motility as well as several mutants with significantly improved ethanol production.

Transposon Directed Insertion-site Sequencing (TraDIS) involves the creation of a large transposon mutant pool followed by high-throughput sequencing to simultaneously map every insertion site. TraDIS is capable of producing a comprehensive list of candidate essential genes as well as gene targets implicated in product tolerance, as such it is a valuable tool for directing future strain engineering. Transposon-delivery vector and DNA sequencing library innovations have been demonstrated which will facilitate TraDIS experiments in this industrially important organism.

Acknowledgements

Acknowledgements must principally go to my supervisors Professor Nigel Minton, Dr Christopher Humphreys, Dr Anne M. Henstra and Dr Klaus Winzer for research opportunities and support. The wide diversity of support is not something I had anticipated when embarking on this study, innumerable conversations with many members of the Clostridia Research Group and later the Synthetic Biology Research Centre Nottingham have aided my progress. Particular thanks goes to Dr Christopher Humphreys whose seemingly endless willingness to engage in scientific discussions was of immense benefit. Thanks must also go to friends and family who have offered support throughout. I am lucky enough to have been brought up in an environment where learning and the ability to think correctly were strongly valued.

Abbreviations

%	Percentage
µl	Microlitres
˚C	Degrees Celsius
5-FOA	5-Fluorooratic Acid
ACE	Allele-Coupled Exchange
aTc	Anhydrotetracycline
ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CFU	Colony Forming Unit
cm	Centimetre
Clr	Clarithromycin
Cm	Chloramphenicol
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase pair
LB	Luria-Bertani medium
LHA	Left Homology Arm
OD	Optical Density
ORF	Open Reading Frame
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
qPCR	Quantitative Polymerase Chain Reaction
RBS	Ribosome Binding Site
RHA	Right Homology Arm
RNA	Ribonucleic Acid
TAE	Tris Base-Acetic Acid-EDTA Buffer
Tm	Thiamphenicol
w/v	Weight per volume
YTF	Yeast Tryptone Fructose

List of figures

Figure 1.1: Allele-Coupled Exchange..... 32
Figure 1.2: Transposase Families..... 42
Figure 1.3: Saturation and essentiality..... 48
Figure 3.1 Common methylation patterns in *E. coli*. 76
Figure 3.2: Conjugation efficiency comparison between Sexpress and CA434. 82
Figure 3.3: RM systems of *C. autoethanogenum* REBASE archive..... 86
Figure 3.4: Recombineering *C. autoethanogenum* MTases into R702. 91
Figure 3.5: pCW10..... 93
Figure 3.6: pCW24..... 95
Figure 3.7: Sexpress, InterstellaR and InterstellaR+MTases conjugative efficiency comparison. 98
Figure 3.8: Sexpress and Sexpress+MTases conjugation efficiency comparison..... 99
Figure 4.1: Orthogonal expression from P_{tcdB} in *C. autoethanogenum* $TcdR^+$ and wild type..... 121
Figure 4.2: Orthogonal expression in two $TcdR^+$ *C. autoethanogenum* strains..... 122
Figure 4.3: Orthogonal expression from P_{tcdB} in an inducible $TcdR^+$ *C. autoethanogenum* strain..... 123
Figure 4.4: Conditional replicon demonstration in *C. autoethanogenum*..... 125
Figure 4.5: *C. ljungdahlii* $TcdR^+$ and Δ*pyrE* with *pyrE* flanking primers..... 128
Figure 4.6: Conditional replicon demonstration in *C. ljungdahlii*. 129
Figure 4.7: pMTL-YZ14..... 133
Figure 4.8: Histidine biosynthesis. 136
Figure 4.9: Insertion-site in *potE*. 136
Figure 4.10: Insertion-site in *hisD*. 136
Figure 4.11: Motility assay on wild type *C. autoethanogenum* and putative flagella mutants..... 138
Figure 4.12: HPLC analysis of wild type compared to two transposon mutants after 72 hours growth. 140
Figure 4.13: Ethanol percentage as a total of detected metabolites. 141
Figure 4.14: HPLC analysis of *C. autoethanogenum* wild type and two transposon mutants 142
Figure 4.15: Insertion site of a 5-FOA^R mutant. 144
Figure 4.16: Insertion site of a 5-FOA^R mutant. 145
Figure 4.17: Final OD_{600} of wild type *C. autoethanogenum* in varying concentrations of ethanol, acetone or isopropanol. 146
Figure 5.1: Effect of plasmid copy number..... 161
Figure 5.2: Read distribution. 164
Figure 5.3: P_{thl}-Tn-*codA* stability strategy. 166
Figure 5.4: Lactose induction of P_{bgaL} 174
Figure 5.5: Effect of induction on pMTL-CW17 transposition frequency. 177
Figure 5.6: pMTL-CW20 179
Figure 5.7: Diagnostic digest of pMTL-YZ14..... 180
Figure 5.8: Diagnostic digest of pMTL-CW20. 181
Figure 5.9: Effect of 5FC on pMTL-CW20 Transposition Frequency. 183
Figure 5.10: Tetracycline inducible transposon-delivery plasmids. 185
Figure 5.11: pMTL-CW22 186
Figure 5.12: pMTL-CW21. 189
Figure 5.13: Transposition frequency comparison across transposon-delivery vectors. 191
Figure 6.1: I-SceI digestion in sequencing library preparation. 204
Figure 6.2: Transposon insertions across *C. autoethanogenum* genome and pMTL-CW20..... 210

Figure 6.3: Read distribution 211
Figure 6.4: Sequencing library preparation summary. 214
Figure 6.5: Size distribution of DNA fragments after adapter ligation. 220
Figure 6.6: Size distribution of DNA fragments after PCR enrichment. 222
Figure 6.7: Expected adapter-containing fragments for sequencing. 223
Figure 6.8: Alignment of TA cloning sequencing fragments. 224
Figure 6.9: Growth curve of transposon mutant libraries. 228
Figure 6.10: Library primer re-design. 229
Figure 6.11: Optimisation of transposon-chromosomal enrichment PCR. 232
Figure 6.12: Size distribution of DNA fragments prior to sequencing, untreated libraries. 234
Figure 6.13: Size distribution of DNA fragments prior to sequencing, treated libraries. 235
Figure 6.14: Splinkerette adapters. 239
Figure 8.1 Read Distribution 256
Figure 8.2 Histogram of Genes and Insertion Index 257

List of tables

Table 1.1: ACE selection 34
Table 2.1: Bacterial strains used in this study. 52
Table 2.2 Yeast Tryptone Fructose (YTF) medium. 56
Table 2.3: 1000x Acid trace element solution. 56
Table 2.4: Basic trace element solution. 57
Table 2.5: B-vitamin solution. 57
Table 2.6: Semi-defined Minimal Media (SDMM). 57
Table 2.7: PETC 58
Table 2.8: Trace metal solution 100X stock. 58
Table 2.9: Wolfe's vitamin solution 100X stock. 59
Table 2.10: LB 59
Table 2.11: Supplements 65
Table 3.1 Conjugal transfer efficiency comparison, CA434 and Sexpress. 81
Table 3.2: Type IV systems in selected clostridia. 83
Table 3.3: Conjugation efficiency comparison between CA434 and Sexpress in *C. difficile* DH1916. 84
Table 3.4: BLAST results of two *C. autoethanogenum* MTases against several clostridia. 87
Table 3.5: Comparison of *C. ljungdahlii* and *C. autoethanogenum* RM systems. 88
Table 3.6: Primers and templates used to construct pCW10. 92
Table 3.7: Primers and templates used to construct pCW24. 94
Table 4.1: Insertion sites for randomly chosen transposon mutants. 134
Table 4.2: Ethanol to acetate ratios for *C. autoethanogenum* C24 and two transposon mutants. 141
Table 5.1: Primers and templates used for construction of pMTL-CW14. 169
Table 5.2: Assessment of pMTL-CW14. 170
Table 5.3: Primers and templates used for the construction of pMTL-CW17. 175
Table 5.4: Assessment of pMTL-CW17. 176
Table 5.5: Primers and templates used for construction of pMTL-CW20. 178
Table 5.6: Assessment of pMTL-CW20 182
Table 5.7: Primers and templates used for construction of pMTL-CW22. 187
Table 5.8: Assessment of pMTL-CW22. 187
Table 5.9: Assessment of pMTL-CW21 190
Table 5.10: Assessment of pMTL-CW21 and pMTL-CW23 in *E. limosum*. 191
Table 5.11: Plasmid copy number of p19 and pCB102. 195
Table 6.1: Sequencing primers. 209
Table 6.2: TraDIS reads summary. 209
Table 6.3: Size selection required for libraries. 222
Table 6.4: Sequencing primer comparison between the two sequencing library preparations. 229
Table 6.5: PCR recipe for transposon-chromosomal junction enrichment. 232
Table 6.6: PCR program for transposon-chromosomal junction enrichment. 232
Table 6.7: Summary of changes to sequencing library preparation. 233
Table 8.1 A comparison of the two library preparations performed in this work. 254
Table 8.2 Effect of methods to remove plasmid based reads 255
Table 8.3 List of Candidate Essential Genes of *C. autoethangenum* for Growth on YTF 258

Table of Contents

1 Introduction 13

1.1 GHG Emissions and Climate Change 13

1.2 A Role for Biofuels 14

1.3 Syngas and Gas Fermentation 15

1.4 Fischer-Tropsch Process 16

1.5 Acetogens 17

1.6 Wood-Ljungdahl Pathway 18

1.7 Gas Fermentation Advantages 19

1.8 Commercialisation of Gas Fermentation Technology 20

1.9 Gas Fermentation Commercial Process Overview 21

1.10 Challenges for Gas Fermentation 21

1.11 Media Optimisation 22

1.12 Biological Catalyst 23

1.13 *C. autoethanogenum* 24

1.14 Notable Alternative Technologies 25

1.14.1 Consolidated Bioprocessing Process 25

1.14.2 Algae 26

1.15 Genetic Tools in *Clostridium* 26

1.15.1 Summary of Genetic Tools Available in *Clostridium* 26

1.15.2 Chromosomal DNA Integration and Allele-Coupled Exchange (ACE) 28

1.15.3 Overview of the Process 30

1.15.4 Pseudo-suicide 32

1.15.5 Homology Arms 33

1.15.6 Acetogen Genetic Manipulations 34

1.16 DNA Transfer 35

1.16.1 General Methods for DNA Transfer 35

1.16.2 Restriction Systems 36

1.17 Transposon Mutagenesis 38

1.17.1 Transposons 38

1.17.2 The Mariner Transposase Himar1C9 43

1.17.3 Transposon Insertion Sequencing (TIS) 44

1.18 Summary 50
1.19 Aims 51
2 Materials and Methods 52
2.1 Bacterial Strains and Growth Conditions 52
2.1.1 Bacterial Strains 52
2.1.2 Plasmids 53
2.1.3 Growth Conditions 54
2.2 Chemicals and Suppliers 54
2.3 Bioinformatic Tools 55
2.4 General Microbiology Techniques 55
2.4.1 Clostridial Growth Media 56
2.4.2 PETC 58
2.4.3 *E. coli* Growth Media 59
2.4.4 Strain Storage and Revival 60
2.4.5 Preparation of Electrocompetent *E. coli* 60
2.4.6 Transformation of *E. coli* by Electroporation 61
2.4.7 Plasmid Transfer to *C. autoethanogenum* via Conjugative Transfer 61
2.4.8 Preparation of Electrocompetent *C. ljungdahlii* 62
2.4.9 Preparation of Electrocompetent *C. autoethanogenum* 63
2.4.10 Transformation of *C. ljungdahlii* 63
2.4.11 Lambda Red Recombineering 64
2.4.12 Supplements 64
2.5 General Molecular Techniques 65
2.5.1 Phenol-chloroform Extraction of Genomic DNA from *C. autoethanogenum* or *C. ljungdahlii* 65
2.5.2 Preparation of Plasmid DNA from *E. coli* 66
2.5.3 Restriction Enzyme Digests 66
2.5.4 Dephosphorylation 66
2.5.5 Ligations 67
2.5.6 Agarose Gel Electrophoresis and Visualisation 67
2.5.7 DNA extraction from Agarose Gels 67
2.5.8 DNA Fragment Sequencing 68
2.5.9 HiFi Assembly 68
2.6 Generation of Clostridial Mutants via Allele Coupled Exchange 68
2.6.1 Generation of the *pyrE* Repair Vector Inserting *tcdR*. 68

2.6.2 Restoration of *pyrE* Function and Insertion of *tcdR* at the *pyrE* Locus in *C. autoethanogenum*. ... 69
2.6.3 Restoration of *pyrE* Function and Insertion of *tcdR* at the *pyrE* Locus in *C. ljungdahlii*. 69
2.7 Transfer of R702 Between *E. coli* Strains ... 70
2.7.1 Conjugation of Modified R702 from InterstellaR to NEB Express ... 70
2.7.2 Conjugation of R702 from CA434 to NEB Express ... 71
2.8 Chloramphenicol Acetyl Transferase (CAT) Assay ... 71
3 Improving DNA Transfer ... 73
3.1 Introduction ... 73
3.1.1 DNA Transfer to *C. autoethanogenum* ... 73
3.1.2 Comparison to *C. ljungdahlii* and Importance of Methylation ... 74
3.1.3 Restriction-Modification Systems of *C. autoethanogenum* ... 74
3.1.4 Type IV Systems and Native *E. coli* Methylation ... 75
3.1.5 Rationale for Attempting to Improve DNA Transfer ... 77
3.2 Results ... 79
3.2.1 Conjugation Technique ... 79
3.2.2 Creation of a Novel Conjugal Donor Strain - Sexpress. ... 79
3.2.3 Mimicking *C. autosethanogenum* Methylation Patterns ... 84
3.3 Discussion ... 100
3.3.1 Importance of Donor *dcm* Methylation Status ... 101
3.3.2 Expression of MTases in Sexpress ... 107
3.3.3 Expression of MTases from InterstellaR ... 108
3.3.4 Further Work ... 110
3.4 Conclusion ... 112
3.5 Results Summary ... 112
4 Implementation of Transposon Mutagenesis in *C. autoethanogenum* and *C. ljungdahlii* ... 114
4.1 Introduction ... 114
4.1.1 Transposon Mutagenesis in *Clostridium* ... 114
4.1.2 Transposon Delivery ... 114
4.1.3 Conditional Replicon ... 116
4.1.4 Transcriptional Control of Transposase ... 116
4.1.5 Orthogonal System TcdR – P_{tcdB} ... 117
4.1.6 Aims ... 118
4.2 Results ... 119
4.2.1 Insertion of *tcdR* via Allele-coupled Exchange (ACE) to *C. autoethanogenum* ... 119

4.2.2 Plasmid Loss Demonstration in *C. autoethanogenum* 123
4.2.3 Conjugation of pMTL-YZ14 into *C. autoethanogenum* 126
4.2.4 Creation of a TcdR$^+$ Strain of *C. ljungdahlii* 127
4.2.5 Plasmid Loss in *C. ljungdahlii* 128
4.2.6 pMTL-YZ14 in *C. ljungdahlii* 129
4.2.7 pMTL-YZ14 in *C. autoethanogenum* 131
4.2.8 Phenotypic Screens and Mutant Characterisation 135
4.2.9 Gamma Glutamyl Hydrazide (GGH) Screening 143
4.2.10 5-FOA Screening 143
4.2.11 Solvent Tolerance. 145
4.3 Discussion 147
4.3.1 Conditional Replication and Orthogonal Expression 147
4.3.2 Conditional Replicon P_{fac}-pCB102 148
4.3.3 Construction of *C. ljungdahlii* TcdR$^+$ and Transposon Mutagenesis 149
4.3.4 Transposon Mutagenesis of *C. autoethanogenum* with pMTL-YZ14 150
4.3.5 Characterisation of Sequenced Mutants 151
4.3.6 Screening Mutant Pools 153
4.4 Conclusion 157
4.4.1 Results Summary 157
5 Developing Vectors for Transposon Insertion Sequencing 158
5.1 Introduction 158
5.1.1 The Importance of Unique, Chromosomal Reads. 159
5.1.2 Plasmid Read-contamination 160
5.1.3 I-SceI Strategy 162
5.1.4 Early Transposition Events 162
5.1.5 Unwanted Transposition Events in Donor and Storage Strains 164
5.1.6 P_{thl}-Tn-*codA* Strategy 165
5.2 Results 167
5.2.1 Suicide Vector pMTL-GL15 167
5.2.2 pMTL-CW14, Allowing Removal of Plasmid-based Reads 168
5.2.3 pMTL-CW17 A Lactose-inducible Transposon-delivery Plasmid 172
5.2.4 pMTL-CW20, Increased Plasmid Stability. 178
5.2.5 pMTL-CW22, A Tetracycline-inducible Transposon-delivery Plasmid 184
5.2.6 pMTL-CW21 188

5.2.7 Transposon-delivery Vectors in *C. autoethanogenum* 191
5.3 Discussion 192
5.4 Conclusion 200
5.4.1 Results Summary 200
6 TraDIS 202
6.1 Introduction 202
6.1.1 TraDIS 202
6.1.2 Aims 203
6.1.3 Sequencing Library Preparation and Inclusion of I-SceI Step 203
6.1.4 Essential Genes 205
6.1.5 Identification of Isopropanol Resistance Genes 206
6.1.6 Biological Library Creation Protocol 207
6.2 Results 208
6.2.1 TraDIS First Attempt 208
6.2.2 Troubleshooting 212
6.2.3 TraDIS Second Attempt 225
6.2.4 Sequencing Library Preparation 228
6.3 Discussion 235
6.3.1 Overview 235
6.3.2 The Importance of Library Complexity 236
6.3.3 Removal of Plasmid-based Reads with a Restriction Digest Step 238
6.3.4 Splinkerette Adapter 239
6.3.5 Future Work 240
7 General Discussion 242
7.1 Context 242
7.2 DNA Transfer 244
7.3 Transposon Mutagenesis 245
8 Appendix 251
8.1 Oligonucleotide Primers Used 251
8.2 TraDIS Results Summary 254
9 References 271

1 Introduction

1.1 GHG Emissions and Climate Change

The links between the use of fossil fuels, increasing greenhouse gas (GHG) emissions and global climate change have been established (Ipcc, 2013; Tol, 2014). As a consequence 175 states, including the UK, signed the Paris Climate Accord in 2016 which aims to limit global mean temperature to "*well below 2°C*" above pre-industrial levels as well as to "*foster... low greenhouse gas emissions development in a manner which does not threaten food production*". The objectives of the accord were set in light of the possible environmental effects associated with the continuation of an upward trend of GHG emissions which would result in a doubling of the emission rate by 2050 and a CO_2 concentration of 500ppm[3] leading to a global mean temperature increase of 2°C (Neill, Oppenheimer, Neill, & Oppenheimer, 2002; Pacala, 2004). A 2°C increase in global mean temperature has been predicted to cause a wide variety of environmental changes, perhaps most notably the degradation of the West Antarctic Ice Sheet which could cause sea level rises of 4-6 meters (Neill et al., 2002). There is a degree of scientific uncertainty on the timing, probability and magnitude of the environmental impacts of climate change, but the unavoidable reasonable concern for the possibility of adverse effects make the adoption of the precautionary principle a wise measure (van der Sluijs & Turkenburg, 2006).

1.2 A Role for Biofuels

As a means to achieving the aims of the Paris Climate Change Accord Treaty a range of renewable energy generation technologies are in development including wind, solar, tidal, geothermal and biofuels. Biofuels have a niche in this market due to their production of compounds with a high energy density and their compatibility with existing infrastructure. High energy density fuels are still required for infrastructure and as such biofuels are set to be relevant for the foreseeable future (Liao, Mi, Pontrelli, & Luo, 2016). In 2009 the EU committed to 10% of transportation fuels being biofuels, this compares to the 3.5% of global oil demand for road transport met by biofuels in 2013 (European Academies Science Advisory Council, 2012; OECD & IEA, 2014).

Traditionally biofuels have been produced using crop feedstocks which cause food-fuel competition. First-generation biofuels such as these are usually derived from corn starch or cane sugar and have the advantage of needing no biomass degradation step as the inputs are already digestible by the microorganisms used to produce the biofuel. First-generation biofuels are predominately produced in the USA from corn or in Brazil from sugar cane. In 2007-08 the prices of many food crops doubled and the National Academy of Sciences estimated that biofuel expansion accounted for 20-40% of these price increases (Wise, 2012), although analyses taking into account the consequent fuel, and therefore food production costs, estimates this at the lower value of 10-15% (Hochman, Rajagopal, & Zilberman, 2010). Since primary biofuels require the same resources for production as food – energy, land and water, using dedicated crops as feedstock for biofuels puts a strain on global food production and causes environmental issues through the increase in cropland

at the expense of set-aside grasslands and forests. For instance, 850,000 hectares of grassland was converted to cropland due to the corn-ethanol boom between 2005-08 with resultant damage to ecosystem services and wildlife which are typically excluded from economic analyses of the process (Fargione et al., 2009).

Global population growth exceeds increases in grain production, consequently food availability per capital declined continuously between 1961-2006 (Pimentel et al., 2009), though a 2017 WHO report examining child malnutrition showed that in most developing regions the proportion of overweight children is rising while the proportions of stunted or wasting children continues to fall (UNICEF/WHO/World Bank, 2016). Despite this, undernutrition contributes to "*nearly half*" of deaths of children under the age of 5 (Pimentel et al., 2009). Some of the problems of first-generation biofuels have been addressed by using non-food parts of crops as a feedstock in so-called second generation biofuels, though environmental concerns remain, for instance the removal of crop residues can intensify soil erosion, acidification and eutrophication due to increased runoff (Pimentel et al., 2009).

1.3 Syngas and Gas Fermentation

The organisms examined in this text can utilise CO or $CO_2 + H_2$ as a sole carbon and energy source. This enables the use of industrial waste gases as a feedstock, for example, from steel manufacture, oil refining, coal or natural gas extraction. Alternatively, synthesis gas (syngas) can be used as a feedstock allowing a myriad of waste streams to ultimately provide the carbon and energy for growth, since essentially any hydrocarbon compound can be converted to

syngas via partial combustion including lignocellulosic biomass, municipal solid waste or coal. The conversion of syngas to liquid hydrocarbons has historically been achieved through the series of chemical reactions which comprise the Fischer-Tropsch process, but recently commercial interests have centred around biological production of liquid fuels and chemicals using syngas as a feedstock, in a process known as Gas Fermentation.

1.4 Fischer-Tropsch Process

The Fischer-Tropsch process was developed during the 1920s in Germany (Demirbas, 2007) and was understood to be of immense importance in the event of oil, gas and solid organic matter depletion (Schulz, 1999a). The conversion of CO and H_2 into ordered composition hydrocarbons is a significant step which may have allowed the prospect of continual depletion of fossil fuel reserves to be less worrying than otherwise it might have been. Since the inception of the technology the FT process has been used sporadically geographically and temporarily depending on environmental, strategic or parochial economic factors. In WWII Germany the FT process provided strategically important liquid fuel from coal, but in peacetime economic imperatives reduced the capacity from a zenith of 600,000 t per annum (Schulz, 1999b). After WWII cheap and plentiful oil supply limited the use of the FT process in most parts of the world, one notable exception being in South Africa where local coal production and gasification was encouraged by state policy. Worrying forecasts of a dwindling oil supply in the 1970-80s caused a resurgence of interest in the FT process again using gasified coal in the US, Japan and Europe (Schulz, 1999a). FT synthesis has been implemented in the main using ‘stranded gas’ which is essentially unused or wasted gas. FT

synthesis today is largely an environmentally-friendly alternative to flaring (Schulz, 1999a).

Following the history of the FT process exemplifies the importance of the wider societal, political and economic context to the use of syngas. In global politics the overall trend in the 21st century has so far been toward an increasing commitment to mitigating human harm to the natural environment and so projecting this trend leads to the conclusion that syngas fermentation will grow in industrial importance. Furthermore, although global oil reserves are continually being discovered, long-term there will be necessity to find alternative routes to the estimated 30,000 products currently derived from fossil fuels as well as to produce high-energy density fuel molecules. At the turn of the millennium interest in the FT process was high and an increasing adoption of the process was thought likely (Schulz, 1999a), however, biological use of gasified waste has emerged as an possible alternative – gas fermentation technology.

1.5 Acetogens

Gas fermentation technology is reliant on carbon fixation via the Wood-Ljungdahl pathway which fixes carbon from CO or CO_2 to acetate. Organisms which possess this capability are termed acetogens, although sometimes the term is applied to any bacteria with acetate as a fermentation product (Ljungdahl, 1986). In this text “acetogens” will refer to bacteria which can fix carbon from a gaseous mix of CO or $CO_2 + H_2$ and produce acetate using the Wood-Ljungdahl pathway. A variety of industrial waste gases derived from sources such as steel manufacture, oil refinery or coal as well as syngas

discussed earlier can serve as growth substrates for acetogenic bacteria. For around 90% of acetogens, acetate is the sole fermentation product (Köpke et al., 2014) but within the remaining 10% are organisms which additionally naturally produce such industrially relevant compounds as ethanol, butyrate, butanol and 2,3 butanediol. Notable organisms which have been reported to produce ethanol using CO as their sole carbon and energy source include *Clostridium autoethanogenum* (Abrini, Naveau, & Nyns, 1994), *Clostridium ljungdahlii* (Tanner, Miller, & Yang, 1993; Vega, Antorrena, Clausen, & Gaddy, 1989), *Clostridium carboxidivorans* (Liou, Balkwill, Drake, & Tanner, 2005), *Eubacterium limosum* (Worden, Grethlein, Jain, Datta, 1993; Chang, Kim, Lovitt, & Bang, 2001), *Clostridium ragsdalei* (strain P11) (Kundiyana, Huhnke, Maddipati, Atiyeh, & Wilkins, 2010; Kunbandra, Atiyeh, Wilkins, Huhnke), *Alkalibaculum bacchi* (Allen, Caldwell, Lawson, Huhnke, & Tanner, 2010), mesophilic bacterium p7 (Rajagopalan, P. Datar, & Lewis, 2002) and *Butyribacterium methylotrophicum* (Shen, Shieh, Grethlein, Jain, & Zeikus, 1999). Currently all known acetogens capable of producing ethanol are mesophilic. Genetic manipulations in these organisms are relatively undeveloped, but commercial interest has fuelled research efforts in recent years opening up a range of possibilities for industrial use and elucidating the underlying biology involved in these important pathways.

1.6 Wood-Ljungdahl Pathway

The Wood-Ljungdahl pathway is the defining feature of all acetogens and it is the most energy efficient carbon fixation pathway as it is the only pathway which produces rather than consumes ATP (Cotton, Edlich-Muth, & Bar-Even, 2017). Moreover, it is also considered to have played an important role in the early

evolution of life where chemoautotrophs dominated (Braakman & Smith, 2012). It is a relatively simple, linear pathway in which electrons are provided by CO or H_2 and carbon is provided by CO or CO_2. A CO dehydrogenase (CODH) allows electron production from CO and a hydrogenase does the same for H_2. Carbon fixation is achieved by the enzyme acetyl-CoA synthase which combines a carboxyl group from CO (or from CO_2 via reduction), a methyl group and a coenzyme A to form acetyl-coA (Ljungdahl, 1986).

1.7 Gas Fermentation Advantages

Syngas or industrial waste gases can be used as the feedstock for a variety of microorganisms which utilise the carbon for CO or CO_2 and the energy from CO or H_2 to produce fuel or commodity chemicals. This process is termed gas fermentation and is the process which this work aims to improve. Gas fermentation using a biological catalyst instead of the chemical FT process offers several advantages. The Fischer-Tropsch process relies on large amounts of catalyst which can be either cobalt-based or iron-based. Cobalt-based catalysts are preferred if the feedstock has a high H_2/CO ratio as is provided by natural gas and is the more expensive of the two catalysts (Munasinghe & Khanal, 2011). Gas fermentation instead uses a biological catalyst which inherently self-replicates. The FT process requires elevated temperatures and pressures while gas fermentation proceeds at ambient temperatures and atmospheric pressure allowing significant energy and cost savings (Liew et al., 2016). Biological fermentation is irreversible allowing greater conversion efficiencies than the FT process offers (Liew et al., 2016). Enzyme specificity allows product selectively and therefore a product of higher purify and fewer side-products (Munasinghe & Khanal, 2011). Gas feedstock

for the FT process must be pre-treated to remove contaminants which negatively affect the process, while biological conversion is more tolerant to these gaseous impurities (Köpke, Noack, & Dürre, 2011; Mohammadi et al., 2011). Biological production of chemicals can also be extremely efficient at producing compounds stereo-specifically. For instance, an NADPH-dependent primary- secondary alcohol dehydrogenase (CaADH) from *C. autoethanogenum* was shown to install an R stereo centre in the reduction of acetoin to 2,3-butanediol (Köpke et al., 2014).

1.8 Commercialisation of Gas Fermentation Technology

Three companies have operated pilot plants using acetogens for gas fermentation. Of these, INEOS Bio and Lanzatech have proceeded to commercial scale production, in the USA (INEOS Bio & Lanzatech), Taiwan and Belgium (Lanzatech). It is estimated that if Lanzatech scaled up to the maximum potential across Europe 1.6 million barrels of fossil fuel-derived gasoline could be displaced (Liew et al., 2016). Life Cycle Assessments are used to ensure accurate evaluation of environmental concerns and estimate the total reduction (or otherwise) of GHG emissions of a given technology, taking into account the entire process inputs and outputs. Several LCAs have been published for the Lanzatech process with reductions in GHG emissions estimated at between 50% to 98% depending on the country the LCA was performed and the feedstock (Handler, Shonnard, Griffing, Lai, & Palou-Rivera, 2016; Ou, Zhang, Zhang, & Zhang, 2013).

1.9 Gas Fermentation Commercial Process Overview

Gasification of starting products proceeds at elevated temperature (600-1000°C) (Griffin, Schultz, Irving, & Road, 2012) and pressure (>30 bar) (Breault, 2010). The composition of the resulting syngas is not as important to the biological catalyst as it would be for entry into the FT process where it is sometimes necessary to perform additional reactions (Munasinghe & Khanal, 2011). Additionally the presence of contaminants is not as deleterious to the process, although the presence of NO, acetylene and cyanide inhibit key enzymes in the Wood-Ljungdahl pathway and pre-treatment of gas is undertaken prior to gas fermentation (Liew et al., 2016). The Lanzatech process is a continuous fermentation with media pumped into the bioreactor and fermentation broth continually output at an equal rate. The bioreactor is designed to optimise the transfer of gases to the aqueous phase where it is accessible to the microorganism. Fermentation products are extracted using a variety of techniques as outlined in (Liew et al., 2016).

1.10 Challenges for Gas Fermentation

There are several problems with gas fermentation technology for which improvements would likely cause a more rapid and widespread commercial adoption. These include a low gas-to-liquid mass transfer, slow cell rates, low productivity, product toxicity and product recovery (Abubackar, Veiga, & Kennes, 2012; Bengelsdorf, Straub, & Dürre, 2013; Latif, Zeidan, Nielsen, & Zengler, 2014; Wilkins & Atiyeh, 2011).

Reactor design is important to maximise the transfer of gaseous CO to liquid phase where it is accessible to the bacteria. The majority of research has been

performed using stirred tank reactors which have been the subject of mechanical engineering refinements to optimise the mass transfer coefficient verses power demand (Ungerman & Heindel, 2007). Scale-up using a Continuous Stirred-Tank Reactor (CSTR) is undesirable because of the high operating electricity requirements needed. Alternative designs include bubble column reactors, monolithic biofilm reactors, trickle-bead reactors, microbubble dispersion stirred-tank reactors and membrane-based systems (Munasinghe & Khanal, 2011).

1.11 Media Optimisation

The medium for industrial use should be as cheap as possible while supporting stable cultures with maximal production rates. The question of whether to use a minimal, defined medium or a cheap, undefined but rich medium is of particular interest. One study compared growth of *Clostridium* strain P11 using a minimal, defined medium, PETC, verses a medium supplemented with cotton-seed extract which is obtainable far cheaper than yeast extract $0.91/kg compared to $183/kg according to (Kundiyana et al., 2010). Yeast extract is a required supplement for the growth of many acetogens, but an elucidation of the exact vitamin and mineral requirements of the organism is preferable for maintaining culture stability. If the cost of providing an undefined supplement such as cotton-seed extract or corn-steep liquor is lower than that of a defined, minimal media then the process economics will force adoption of the former option. Media composition can also have a strong impact on ethanol production (Abubackar, Veiga, & Kennes, 2015; Guo et al., 2010; Saxena & Tanner, 2011, 2012). It may be beneficial for research to identify a strong candidate for a fermentation medium early on so that subsequent genetic research can be of

increased relevance. At present the use of a range of media types as well as a range of reactor types makes cross comparison of key performance indicators of different biological catalysts very difficult (Kundiyana et al., 2010).

1.12 Biological Catalyst

Selection and optimisation of a biological chassis is an important aspect of gas fermentation technology. There are relatively few species known with both the capacity to fix carbon via the Wood-Ljungdahl pathway and produce ethanol, all of which are strict anaerobes (Munasinghe & Khanal, 2011). Environmental isolations of novel strains may provide greater productivities and robustness than the suite of currently known microorganisms provide, or genetic modifications of known strains may extend their capabilities to the extent that novel strains are unnecessary. No single organism has dominated the literature as of yet with syngas fermentations frequently being performed using *C. ljungdahlii, C. autoethanogenum, Acetobacterium woodii, C. carboxidivorans* and *Clostrdium aceticum* (Munasinghe & Khanal, 2011).

As previously mentioned, strain comparisons between laboratories are difficult leading to one study which aimed to revive six strains reported to convert syngas to ethanol in the literature and compare their performance in a two-stage CSTR fermentation (Martin, Richter, Saha, & Angenent, 2016a). Of these six strains only three were successfully grown to steady state in defined medium without the addition of yeast extract – *C. ljungdahlii PETC, C. ljungdahlii ERI-2, and C. autoethanogenum JA1-1*. Of the three successful strains the two *C. ljungdahlii* strains demonstrated superior culture stability and ethanol production rates over *C. autoethanogenum* (Martin et al., 2016a). However, this

study may have limited relevance to the evolved, lab-adapted strains which are generally used by commercial entities. The power of directed genetic modifications and synthetic biology approaches should make the innate properties of various strains relatively unimportant. The natural variance between species will hopefully be dwarfed by the variation between a genetically modified strain and any natural strain. Biological catalyst optimisation is a complex and multi-factorial problem with traits such as titre, selectivity, biomass density, yield, percentage conversion of carbon to ethanol and culture stability all being potentially relevant to the selection of a strain for industrial scale production (Latif et al., 2014; Martin, Richter, Saha, & Angenent, 2016b). The present work aims toward improvement of an acetogenic strain to facilitate efforts to commercialise gas fermentation technology.

1.13 *C. autoethanogenum*

C. autoethanogenum was first isolated in 1994 from enriched rabbit feces. It was isolated from other enriched species by its ability to grow using CO as a carbon and energy source. Genetic studies have been aided by its ability to also use various sugars as an alternative carbon and energy source, though being strictly anaerobic has made laboratory manipulation more cumbersome. Growth is optimal at 37°C and between pH 5.8-6. *C. autoethanogenum* is of interest principally because it has been adopted as the organism of choice for the biotechnology company Lanzatech which is currently the only gas fermentation company running commercial scale facilities. After the initial isolation publication in 1994, the next published article concerning *C. autoethanogenum* was not released until 2009. Since then there have been around 31 published articles concerned principally with manipulations of *C. autoethanogenum*. Of

these, only four have involved genetic manipulation rather than modification of some aspect of the fermentation or metabolic modelling. The four genetics papers have all been published in 2016 or later (Bengelsdorf et al., 2016; Liew et al., 2016; 2017; Nagaraju et al., 2016). A full genome of *C. autoethanogenum* was published in 2014 (Brown et al., 2014) with improvements to the sequence being made using manual annotation and sanger sequencing to validate disputed SNPs (Humphreys et al., 2015). The complete genome sequence is a requirement for TraDIS which is attempted in this text since transposon reads must be mapped onto a reference genome.

1.14 Notable Alternative Technologies

1.14.1 Consolidated Bioprocessing Process

Non-food components of crops and other waste materials are one potential feedstock for biofuel production. The use of lignocellulosic feed stocks such as switchgrass, poplar or wheat straw avoids some of the problems of first generation biofuels (Latif et al., 2014). A significant portion of this waste biomass is resistant to biological degradation by established ethanol fermenting organisms and requires chemical processing to make nutrients available for fermentation. There is a complex array of methods to convert raw lignocellulosic biomass into fermentable monosaccharides as detailed in (Geddes, Nieves, & Ingram, 2011). The costly pre-treatment steps before fermentation can occur makes this approach less economically viable. To overcome this problem a Consolidated Bioprocessing Process (CBP) approach is being developed which takes advantage of cellulolytic microorganisms capable of performing the hydrolysis and subsequent fermentation of lignocellulosic biomass (Raftery &

Karim, 2017). CBP is an alternative to gas fermentation which would convert lignocellulosic biomass into syngas before fermentation.

1.14.2 Algae

Third generation biofuels are those generated by macro or micro algae (Behera et al., 2015). Use of microalgae avoids many drawbacks of first and second generation biofuels. Similar to the FT process, the scientific basis of using microalgae as a potential replacement of fossil fuels was outlined in the 1940s but serious attention was only given in the oil-crisis in the early 1970s (Barry, Wolfe, English, Ruddick, & Lambert, 2016). Some microalgae species accumulate up to 50% of their dry cell weight in the form of lipids (Hu et al., 2008). Genetic modification of algal strains is necessary to attempt to free natural strains from the compromise between cell growth and high lipid content (John, Anisha, Nampoothiri, & Pandey, 2011). The costs and difficulties of cultivation are considered to be a major constraint to the prospect of commercialisation and it is as yet unclear whether low-cost, low-productivity open-pool systems will dominant over the expensive, but higher productivity closed bioreactor systems (Ho, Ngo, & Guo, 2014). No economically viable system has been deployed at commercial scale and there is still a great deal of research required to fully commercialize the algae-to-biofuels process (Barry et al., 2016).

1.15 Genetic Tools in *Clostridium*

1.15.1 Summary of Genetic Tools Available in *Clostridium*

Genetic tools are relatively underdeveloped in *Clostridium* largely as a result of an inability to reliably and efficiently transfer DNA (Pyne, Bruder, Moo-Young,

Chung, & Chou, 2014). Inefficient DNA transfer necessitates standard molecular cloning to be performed in *E. coli* before introduction to the clostridial species. A range of shuttle vectors containing both Gram-negative and Gram-positive replicons has been developed for this purpose (Heap, Pennington, Cartman, & Minton, 2009). Early work utilised anti-sense RNA to downregulate targeted genes, this approach requires only a suitable plasmid-vector and promoter to control RNA production as well as sequence data for the targeted gene(s) (Desai & Papoutsakis, 1999; Tummala, Welker, & Papoutsakis, 2003). Insertional knockouts can be created using ClosTron technology (Heap et al., 2011), directed in-frame deletions and chromosomal integration of exogenous DNA can be achieved using replicative plasmids harbouring homology arms and a counter-selection marker on the backbone. A range of counter-selection markers have been demonstrated in clostridia including *pyrE*, *pyrF*, *codA*, *mazF* and *upp* (Al-Hinai, Fast, & Papoutsakis, 2012; Cartman, Kelly, Heeg, Heap, & Minton, 2012b; Croux et al., 2016; Ng et al., 2013). Counter-selection markers can be omitted if the Allele-Coupled Exchange technique (Cartman, Kelly, Heeg, Heap, & Minton, 2012; Heap et al., 2012; Ng et al., 2013) is used as outlined below. Recently, CRISPR-cas9 mediated genome editing has been demonstrated (Nagaraju et al., 2016; Wang et al., 2015; Xu et al., 2015) which allows single-step, marker-less genome editing, increasing the ease and rapidity of the mutant-generation process. An alternative one-step technique to CRISPR involves selection of double crossover events utilising three homology arms, one positive selection marker and two negative selection markers and has been used to generate an in-frame deletion in *C. autoethanogenum* (Walker & Köpke, 2015).

1.15.2 Chromosomal DNA Integration and Allele-Coupled Exchange (ACE)

The ability to insert exogenous DNA into the chromosome of an organism is important for genetic studies. Many origins of replication have been demonstrated to function in *Clostridium* allowing expression of foreign DNA on self-replicating plasmids, however the physiological relevance of expression of genes from a plasmid is lower than the equivalent expression via the chromosome due to varying copy-number between plasmid and chromosome and even more so if a coding sequence replaces another gene using the latter gene's transcriptional control mechanisms. Expression from plasmid DNA is often experimentally inferior to chromosomal expression due to the possibility of unwanted homologous recombination, the need for continual antibiotic supplementation to select for plasmid maintenance and non-physiological gene dosage caused by plasmid copy-number variation.

For organisms transformable with linear DNA the incorporation of exogenous DNA into the chromosome can be achieved using allele exchange by positioning a selectable antibiotic marker alongside the sequence of interest. Integrants can then be selected for directly using the antibiotic. Alternation between two selectable markers can then allow insertion of multiple large sequences (Itaya, Fujita, Kuroki, & Tsuge, 2008).

The clostridia are not transformable with linear DNA meaning that circular plasmid vectors must be used instead. Knockouts can be obtained by the inclusion of homology arms with which a single homologous recombination event can integrate the entire plasmid, however this is an unstable conformation since a second recombination event will revert the chromosomal locus to WT.

The benefit of linear DNA transformation using an antibiotic resistance gene flanked by two homology arms over using replicative plasmids and single crossover events was well exemplified in *Acinetobacter baumanii*. Knockout strains obtained using transformation with a PCR product retained absolute stability while those using a plasmid single crossover integration had a 40% wild type reversion rate (Aranda et al., 2010).

Directed homologous recombination can be achieved by the inclusion of a pair of homology arms on a plasmid. In order to create 'clean' knockout strains, a homology arm pair must be designed including flanking regions of the gene, but omitting the gene itself, with the desired mutant occurring as a result of a double-recombination event. Since recombination events are rare, a selection mechanism must be employed so as to avoid an unfeasible amount of screening, although a deletion obtained by homologous recombination has been found in this way in *C. acetobutylicum* (Harris, Welker, & Papoutsakis, 2002a). In *Clostridium*, reports of deletion via homologous recombination were initially limited to clostridia with high transformation frequencies in which suicide plasmids could be employed (Awad, Bryant, Stevens, & Rood, 1995). The incorporation of counter-selection markers on the plasmid-backbone allowed selection for double-crossover integration events and hence the use of replicative plasmids in less readily transformable species. Several counter-selection markers have been described for this purpose, but use of a counter-selection marker can be avoided using Allele-coupled Exchange (ACE) (Heap et al., 2012).

Selection for rare double crossover events without a plasmid-based counter selection marker in *C. acetobutylicum* has been described (Heap et al., 2012),

and has since proven applicable to many other clostridia. ACE is capable of producing clean, directed knockouts or insertion of selected DNA sequences into the host chromosome. The method depends on the function of the orotate phosphoribosyl transferase enzyme encoded by the *pyrE* gene. It can be used as both a positive or negative selection marker allowing stepwise selection of crossover events. A *pyrE* knockout strain is created using a replicative plasmid harbouring homology arms designed to delete a portion of *pyrE* which will render it inactive.

1.15.3 Overview of the Process

To create a *pyrE* knockout strain a replication plasmid is constructed with one homology arm internal to *pyrE* excluding the 3' end and a second, larger homology arm is designed downstream of *pyrE.* Transfer of the homology arm plasmid occurs with selection for the *catP* marker located on the plasmid backbone. PCR screening of large colonies from thiamphenicol plates which are likely to represent single crossover mutants as explained in section 1.14.4 proceeds and successful single crossover mutants are carried forward to the next stage of selection. Screened single crossover mutants are subjected to negative selection of a functional *pyrE* via the addition of 5-fluoroorotic acid (5-FOA) which is converted by the product of *pyrE* - orotate phosphoribosyl transferase to the highly toxic 5-fluorouridine monophosphate (5-FUMP). Selection plates for *pyrE* knockouts are supplemented with uracil since the deletion of *pyrE* renders the strain auxotrophic for uracil. The creation of a *pyrE* knockout strain is a necessary first step using ACE technology, which can then be used as a base strain to either create 'clean' in-frame deletions or

incorporate exogenous DNA to the chromosome. An overview of ACE strain generation is provided in Figure 1.1.

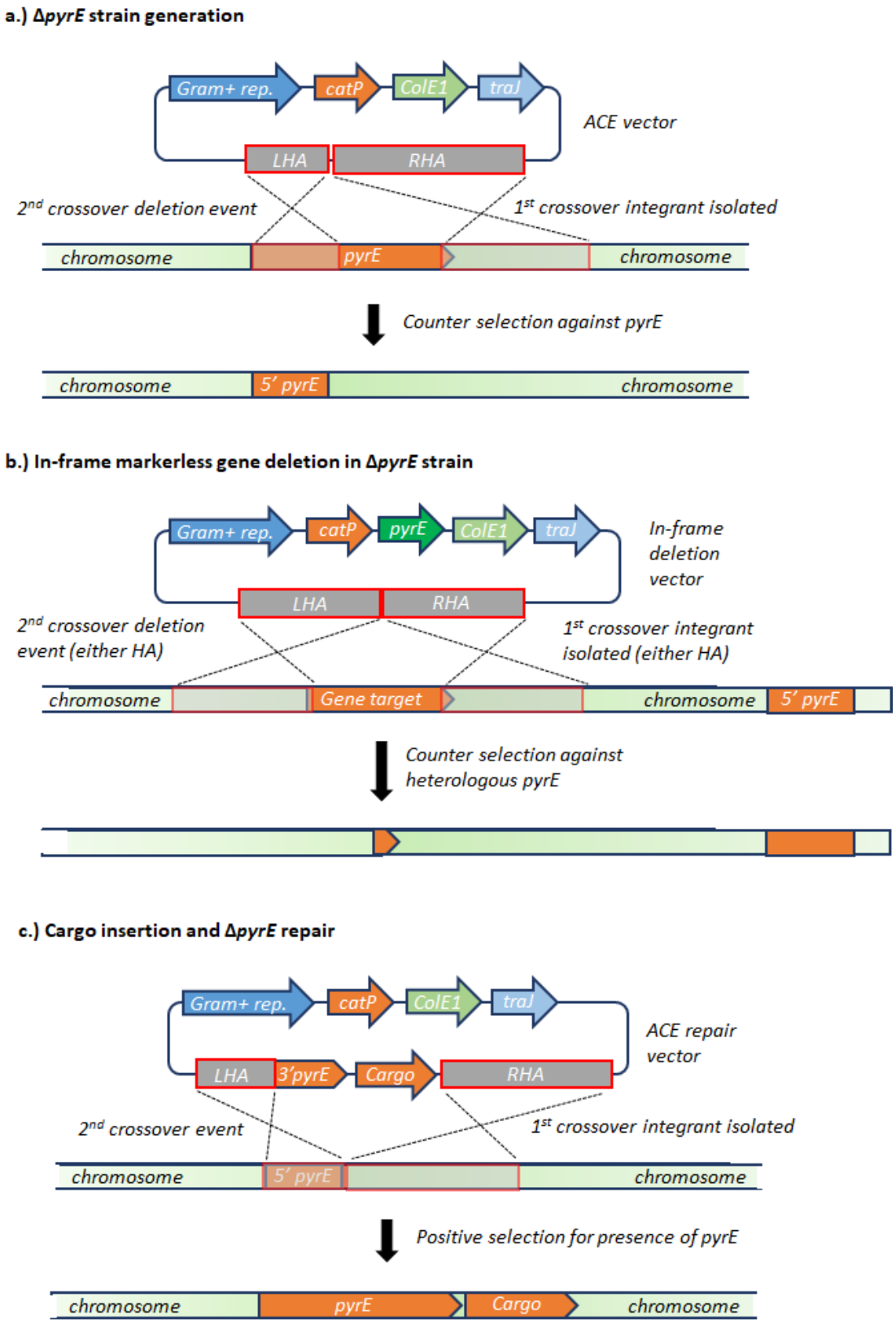


Figure 1.1: Allele-Coupled Exchange.

A summary of ACE. a.) A Δ*pyrE* strain is created by transformation with a plasmid harbouring a left homology arm (LHA) and a right homology arm (RHA) which recombine with the chromosome to produce a non-functional *pyrE* lacking its 3' end. A first crossover event is selected for by media supplementation with thiamphenicol which plasmid-harbouring and single crossover mutants are resistant to due to the presence of the *catP* resistance marker. Single crossovers have a fitness advantage over plasmid-harbouring cells since the plasmid origin of replication is less efficient than that chromosomal origin and as such single crossover mutants are likely to appear as larger colonies on agar plates. b.) The Δ*pyrE* strain can be used as a base strain to produce alternative in-frame deletion mutants. In this procedure homology arms are designed flanking the target chromosomal locus and single crossover mutants are selected for by media supplementation with thiamphenicol. Double crossover mutants are then selected for in a second selective stage by the inclusion of 5-FOA which counter-selects the functional orthologous *pyrE* located on the deletion plasmid backbone. c.) Insertion of genetic cargo at the *pyrE* locus can also be achieved in the Δ*pyrE* base strain by first selecting for the plasmid-located *catP* marker with thiamphenicol supplementation followed by a second selection stage on medium lacking uracil. A functional *pyrE* is required for uracil biosynthesis and so double crossover mutants are selected for since a second recombination event restores the truncated chromosomal *pyrE*.

1.15.4 Pseudo-suicide

The selection of colonies in which a single crossover event has occurred is aided by use of a pseudo-suicide replicon. The homology arm-containing vector possesses the least segregationally stable origin of replication specific for the clostridial species to which genetic alterations are being made. The pseudo-suicide replicon should replicate less effectively than the chromosome meaning that cell division can occur with a daughter cell unable to inherit an antibiotic-resistance providing plasmid causing cell death. At a population level this will slow growth. The functionality of pseudo-suicide replicons lies in the preferential selection of single crossover event colonies by choosing larger colonies which are not slowed by the rate of plasmid replication.

1.15.5 Homology Arms

Transformants are initially selected for by antibiotic selection for the plasmid-located marker, the intention is to produce as many single crossover event mutants as possible which do not disrupt *pyrE.* The homology arms are of different lengths in an attempt to control the order of recombination events. The first recombination event is intended to occur downstream of *pyrE* without affecting PyrE function. The majority of single crossovers maintain the integrity of *pyrE* since the homology arm complementing *pyrE* is a quarter of the size of the homology arm designed outside of *pyrE* making homologous recombination more likely to occur using the larger homology arm external to *pyrE*. Large colonies on the first selection stage for antibiotic resistance to thiamphenicol are therefore likely to be single crossover event mutants which are Tm^R $5\text{-}FOA^S$. This is important because the second stage of selection involves plating to 5-FOA to counter-select *pyrE*, assuming the first crossover event used the larger homology arm downstream of *pyrE*, the second cross-over event will confer resistance to 5-FOA by truncating the chromosomal *pyrE*. If the smaller homology arm directed the first recombination event then selection of 5-FOA resistant mutants would not select against these single crossover event mutants. PCR screening for correct single crossover events avoids this occurrence, these unwanted $5\text{-}FOA^R$ colonies can also easily be identified as they will also be thiamphenicol resistant due to the incorporation of the entire plasmid backbone.

The knockout *pyrE* strain can be used to insert DNA cargo at the *pyrE* locus, selection for this event is possible due to the reformation of *pyrE* and subsequent restoration to uracil prototrophy. In creating the *pyrE* knockout only

the 3’ end of the gene is removed, this allows simple restoration to wild type via homologous recombination with a homology arm containing the missing portion of the truncated *pyrE*.

In-frame deletions can be created in the *pyrE* knockout background strain by incorporation of homology arms flanking the chromosomal region of interest and incorporation of functional orthologous *pyrE* from a closely related organism on the plasmid backbone where it functions as a negative selection marker.

Table 1.1: ACE selection

Summary of the selection stages for in-frame deletions or chromosomal integration.

Aim	Background strain	First selection	Second Selection
Integration of DNA	*pyrE* knockout	Thiamphenicol, single crossovers.	Uracil deficient medium
In-frame deletion	*pyrE* knockout	Thiamphenicol, single crossovers.	5-FOA

1.15.6 Acetogen Genetic Manipulations

Genetic studies in acetogens are an even more recent reality except in the case of *C. difficile* which has been studied for its role as a human pathogen. *C. difficile* contains an intact Wood-Ljungdahl pathway and was found capable of growing autotrophically using H_2 and CO_2 (Köpke, Straub, & Dürre, 2013) consequently research initially performed due to the organism’s role as a human pathogen is also relevant to studies on acetogens. It has been suggested that the autotrophic growth capabilities of *C. difficile* contribute to the persistence of the strain (Köpke et al., 2013). Interestingly, a proteome analysis demonstrated upregulation of Wood-Ljungdahl pathway genes during maximal toxin production (Karlsson, Burman, & Åkerlund, 2008). DNA transfer protocols for other acetogens were developed only since 2010 (Köpke et al., 2010) and

genetic studies in acetogens are still considered to be difficult (Burk & Van Dien, 2016).

1.16 DNA Transfer

1.16.1 General Methods for DNA Transfer

Naturally competent cells can uptake extracellular DNA from their environment, the transfer of DNA between cells with direct cell-to-cell contact can be achieved with conjugation or bacteriophage infections can transduce DNA into bacterial cells. Adding to these three natural DNA transfer mechanisms are artificial transformation techniques including chemotransformation, electroporation, sonoporation, liposome-mediated DNA transfer, chitosan-mediated transformation, biolistic transformation, tribos transformation, protoplast fusion, spheroplast transformation and the freeze-thaw technique (Erik, Aune, & Aachmann, 2010). Of the artificial methods, chemotransformation and electroporation have the most widespread applicability. In general the bacterial cell membrane must be induced to be permeable to DNA, which then enters the cell and is then processed and condensed. Gram-positive bacteria tend to be more challenging to transform due to the presence of a thick cell wall limiting opportunities for DNA to interact with the cell membrane.

Conjugation or electroporation are typically used to transform clostridia. Methods to establish DNA transfer protocols for each species and strain can often require dedicated studies. Factors such as cell wall composition, cell fragility, presence of extracellular nucleases, susceptibility to electroporation enhancing molecules, presence of native CRISPR systems and presence of

native restriction systems can all impact on the potential to transform a strain (Marraffini & Sontheimer, 2008).

1.16.2 Restriction Systems

Restriction-Modification (RM) systems have two activities, restriction and modification of DNA, enzymes responsible for both recognise the same sequence. For Type I-III RM systems, if the sequence is not modified then DNA cleavage will occur via the endonuclease either at the recognition sequence or elsewhere. Restriction and modification functions can be performed by separate enzymes or in some cases a single fused enzyme can perform both function on different domains or subunits. There are four types of restriction systems. Types I, II and III all exhibit nucleolytic activity on unmodified DNA, typically unmethylated DNA. Type IV restriction systems exhibit nucleolytic activity against incorrectly modified DNA.

REBASE is a publicly accessible online database for DNA restriction and modification enzymes (Roberts, Vincze, Posfai, & Macelis, 2010, 2015). Published genomes are analysed for putative methyltransferases and their cognate endonucleases, and if possible, a sequence specificity is assigned. The advent of Single Molecule Real Time (SMRT) sequencing has aided this effort because DNA modifications can be detected during sequencing (Clark et al., 2012). Genome sequences and associated DNA modification information is uploaded to the REBASE database where the sequence specific DNA modifications can often be attributed to a methyltransferase identified within the genome. REBASE identifies all likely candidates for DNA methyltransferases and so often includes RNA and protein methyltransferases initially until these can be confidently reassigned. Restriction enzymes are always found nearby to

their cognate methyltransferase, but are more difficult to recognise due to rapid evolution. The rapid evolution of methyltransferases and restriction enzymes is consistent with their evolutionary role in host-parasite relationships and speciation. Acquisition of RM systems provides a barrier to DNA exchange and hence preserves the genetic identity of a strain allowing further adaptive evolution. Some methyltransferases have no cognate restriction enzyme and are termed orphan methyltransferases. There are also a minority of restriction enzymes which have no associated methyltransferase, in these cases protection of genomic DNA is reliant on the underrepresentation of the recognition sequence in the genome (Shen et al., 2010). In addition to their role as in cellular defence against invading DNA, RM systems can aid in the maintenance of chromosomal stability, allow homologous or non-homologous recombination, provide a genetic barrier allowing speciation, regulation of conjugation, DNA replication, cell cycle control, DNA repair and gene expression regulation both transcriptional and posttranscriptional (Bestor, 1990; Jeltsch, 2003; Low & Casadesús, 2008; Vasu & Nagaraja, 2013; Wion & Casadesús, 2006).

Type IV restriction enzymes exhibit the opposite behaviour to the other restriction enzyme types in that they recognize and cleave methylated DNA, hence they are also known as Methylation-Dependent Restriction Enzymes. Methylation is the most common modification known to activate Type IV restriction enzymes, but other DNA modification types can also be targeted (Loenen & Raleigh, 2014). Restriction of modified DNA is achieved generally without the sequence specificity associated with Types I-III. The best studied Type IV restriction enzyme is McrBC from *E. coli* which has a sequence

requirement of two methylcytosine or hydroxymethylcytosine residues which are separated by 40-3000bp (Loenen & Raleigh, 2014).

Type IV systems likely evolved to counter-act a loading of restriction modification pairs in the genome and as part of an evolutionary arms-race with bacteriophages, low sequence specificity may aid in protection against a broad range of invasive DNA (Loenen & Raleigh, 2014). Various modification types can be recognised by Type IV systems including modification of adenine (m6A) and cytosine (m5C, m4C, hm5C, ghm5C) residues as well as phosphothioate (PT) modification of the sugar-phosphate backbone (Liu et al., 2010; Loenen & Raleigh, 2014). One subgroup of Type II RM systems termed Type IIM has activity against methylated DNA as Type IV systems do and has recently been suggested to be reclassified as Type IV systems since activity against modified DNA is the uniting property of Type IV systems and there is no meaningful distinction between Type IIM and Type IV systems (Loenen & Raleigh, 2014). However, it seems that the distinction between systems which cleave unmodified DNA and systems which cleave incorrectly modified DNA may become increasingly blurred as more enzymes are found for which both conditions play a role. For instance BamHI preferentially attacks DNA with a modified adenine (m6a) in its recognition sequence GGATCC (Whitaker et al., 1999).

1.17 Transposon Mutagenesis

1.17.1 Transposons

Transposons are mobile genetic elements found in all domains of life in almost all organisms and have had important impacts on the evolutionary process

(Ayarpadikannan & Kim, 2014). Transposon insertions may activate or inactivate genes depending on their site of insertion relative to the gene in question. Inversions or deletions can result directly from transposition or due to the homology provided by multiple-copy transposons. A rearrangement of non-transposon located host genes can be induced by the transportation of flanking DNA alongside transposon with certain elements (Exon Shuffling by L1 Retrotransposition). In prokaryotes transposons are important to studies on dissemination of antibiotic resistance and virulence factors through bacterial populations. Eukaryotes have gene expression functions dependent on transposon-associated mechanisms including V(D)J rearrangements (Hiom, Melek, & Gellert, 1998) telomere regeneration (Nakamura & Cech, 1998) and nuclear intron splicing (Curcio & Derbyshire, 2003). Transposon mutagenesis is also a key tool in molecular biology allowing the creation of a mutant library which can be screened for a desired phenotype. Mutants can be isolated on the basis of phenotype and the underlying genetics of the trait identified. This approach, termed forward genetics has been used to identify gene function for a wide range of biological phenomena (Curcio & Derbyshire, 2003; Hickman & Dyda, 2016).

Transposons are the only type of mobile genetic element which require no homology between the section of DNA to be inserted and the target site of insertion (Hickman & Dyda, 2016). The enzyme responsible for movement of a transposon is termed the transposase. Transposases are typically multi-domain proteins with a dedicated domain for nucleolytic activity to excise the transposon from the initial site and create an insertion site elsewhere, where further domains join the broken sections of DNA. In some cases multiple enzymes are

required for the whole transposition process (Hickman & Dyda, 2016). Transposases can be encoded internally within the transposon as in autonomous transposons or elsewhere in the genome. The ability of transposases to act *in trans* makes them of particular use in functional genomics.

Eukaryotic genomes are often replete with transposons with inactivated or absent transposases, with around 40% of the genome in higher eukaryotes being transposon-derived. A lower proportion exists in lower eukaryotes and bacteria of between approximately 1-5%, where selection pressures and generation times work against their presence in the genomic equilibrium (Curcio & Derbyshire, 2003). Transposons often contain genes which are not involved in transposition, termed passenger genes, which can cause an acceleration of antibiotic resistance among bacterial populations. For mobility a transposon requires access to an active transposase and generally the presence of two specific recognition sequences peripheral to the DNA section which is to be moved. The flanking recognition sequences are termed either Inverted Repeats (IRs), Inverted Terminal Repeats (ITRs) or Terminal Inverted Repeats (TIRs), though in this text ITRs is used throughout. Evolutionary pressures on transposons are to possess a high enough transposition frequency so as to proliferate, but not so high so as to be overly deleterious to the host genome. Similarly the site specificity for the ITRs will tend to be far more stringent than the insertion-site specificity so that the transposase does not cause large-scale DNA cleavage and genomic rearrangements, but will be capable of inserting the target DNA into a variety of locations (Hickman & Dyda, 2016), though certain classes of retrotransposons lack terminal repeat regions (Curcio &

Derbyshire, 2003). The acceptable target site for insertion varies between transposases and is relevant to the application of transposons for mutagenesis studies where a relaxed or common target site is preferred, so as to allow the generation of more varied mutant library with minimal effort. The target site preference can be highly specific such as for IS*492* which requires a 7bp sequence (Higgins, Popkowski, Caruana, & Karls, 2009), or have little sequence preference such as with the Tn*5* transposase common in transposon mutagenesis studies (Sasakawa, Carle, & Berg, 1983). Tn*7* contains two separate transposition pathways, one highly-specific for the *att*Tn*7* site which operates at high efficiency and one non-specific transposition which inserts randomly within the genome (McKown, Orle, Chen, & Craig, 1988). It has been hypothesised that this combination of specificities is evolutionarily beneficial as the specific insertion to a non-deleterious site allows stable vertical transmission while the random transposition facilitates horizontal transmission (Craig, 1991; Peters & Craig, 2001).

Transposons are classified by their nucleotide intermediate type (RNA or DNA), their nuclease domain (with four main classes) or their general mechanism of transposition (cut and paste, copy and paste or copy out paste in). Class I elements use an RNA intermediate while Class II elements have a DNA intermediate. Cut and paste transposons have an intermediate which is a physically separated from the donor strand and re-inserted at a different location. These elements proliferate through a genome when transposition occurs during DNA replication, when the excision event is from replicated DNA, and the insertion is in DNA yet to be replicated. In copy and paste transposition the transposon is not excised from the donor site but is copied into a nicked

target site. Copy out paste in transposition creates an intermediate copy which is physically separated from the donor strand which then integrates into cleaved target DNA. There is little variety in the types of nucleolytic domains found in transposases allowing them to form the basis of another classification scheme (Hickman & Dyda, 2016). The four main classes are named based on the identity of conserved residues in the activate site, these are DD (E/D) transposases, serine transposases, tyrosine transposases and HUH transposases (Hickman & Dyda, 2016). Other reviews list five protein families and include a family with reverse transcriptase and endonuclease activities (RT/En) (Curcio & Derbyshire, 2003).

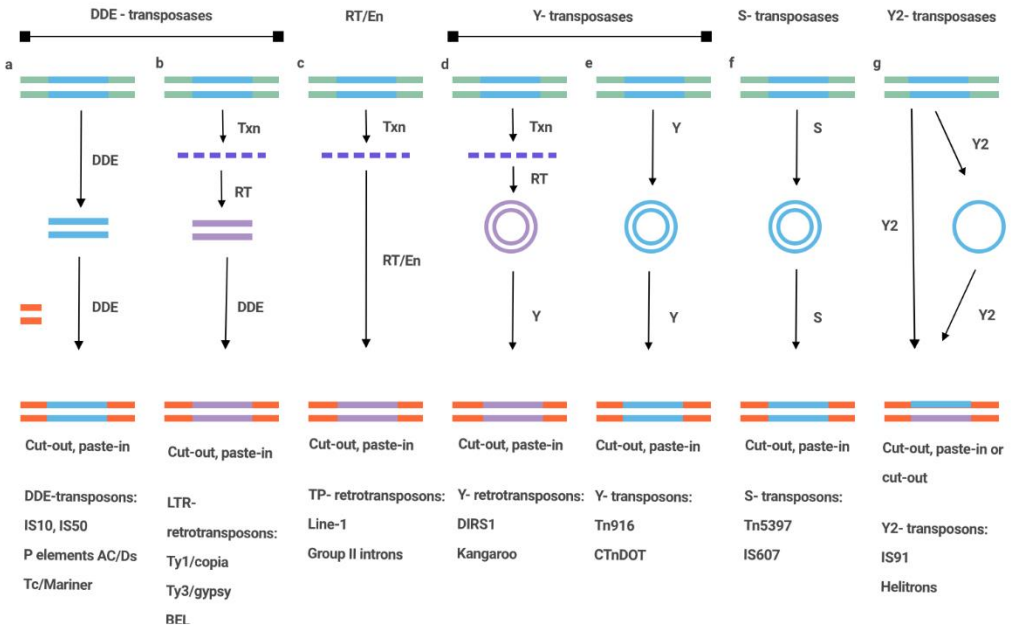


Figure 1.2: Transposase Families

A summary of the diversity of transposase families adapted from (Curcio & Derbyshire, 2003).

1.17.2 The Mariner Transposase Himar1C9

This study involves the use of the *mariner* element *Himar1*. *Mariner* elements are composed of a transposase gene flanked by inverted terminal repeats, significantly transposition can be achieved solely by the transposase with no host factor requirement. Host-factor independence is a key factor rendering it useful as a genetic tool and is responsible for the broad range of species in which *mariner* elements can be found natively including eukaryotes (Plasterek et al., 1999) and bacteria (Picardeau, 2009). Another beneficial feature of *mariner* elements is their lack of sequence specificity beyond the dinucleotide sequence “TA” (Lampe et al., 1996). A bias in insertion sites necessitates the creation of a large mutant pool to achieve the same genetic diversity. The Tn*917* transposase for instance has been shown to have a strong preference for insertion into the replication terminus region of certain *Firmicutes* due to interaction with the termination protein RTP (Shi et al., 2009). The conjugative transposon Tn*916* has been demonstrated in *C. acetobutylicum* to use an insertional hotspot (Woolley, Pennock, Ashton, Davies, & Young, 1989). In some cases, if the sequence requirement is sufficiently stringent the generation of a particular mutant may be impossible if it does not contain the required sequence.

The transposase used in this study was developed when the importance of increasing the transposition activity of the native system from *Haemotobia irritans (Himar1)* was recognised (Lampe, Akerley, Rubin, Mekalanos, & Robertson, 1999). Two mutations were found to act synergistically to increase *in vivo* transposition rates 50-fold over the wild type, this mutant transposase was termed *Himar1C9.* A *catP*-based mini transposon with Inverted Terminal

Repeats (ITRs) from *Himar1* flanking the *catP* gene which confers thiamphenicol resistance was used throughout this study (Cartman & Minton, 2010).

1.17.3 Transposon Insertion Sequencing (TIS)

Microbial genetics aims to uncover the genetic basis for a given phenotypic trait. Classical forward genetics requires the laborious creation and individual screening of a population of mutants. A mutant population is created and individually subjected to a test condition and a control condition, for example rich verses minimal media to detect auxotrophs. Once a desired phenotype is detected in this way the genotype is examined and a function of that gene inferred. The mutant population in this case can be created in a variety of means of which transposon mutagenesis is merely an example. Forward genetics is often the first step when analysing the genetic basis of a phenotype of interest by allowing initial identification of relevant genes.

Deficiencies in the methods of novel gene function discovery were illuminated by the breakthroughs in sequencing technologies. Even with only a selection of microorganisms' whole genome sequence available it was clear that an embarrassingly large proportion of genes had no known function. The need to upgrade methods for gene function determination were called for as early as 1995 when genetic footprinting was developed (V. Smith, Botstein, & Browntt, 1995). Efforts to create a minimal genome further exemplify the inadequacies of functional genomics. A minimal genome for *Mycoplasma genitalium* was published with only 382 predicted essential genes remaining after global transposon mutagenesis and of these still 28% were proteins of unknown function, when published this was the smallest known genome of an organism

able to grow in pure culture (Glass et al., 2006). Recently a genome which is smaller than that of any autonomously replicating cell found in nature thus far has been chemically synthesised and installed into a receptive cellular environment to create JCVI-syn3.0 (C. A. Hutchison et al., 2016). JCVI-syn3.0 is 531kbp and has 473 genes, 149 (32%) of which have unknown biological function.

Defined mutant libraries have been created for several well studied organisms which can then be screened for survival in a condition of interest. Creation of random mutant libraries requires less effort and assuming it is sufficiently saturated can achieve the same result. The surviving mutants have been identified in several different ways defining different experimental techniques. Genetic footprinting involved the creation of a transposon library followed by growth in a condition of interest and a control condition. Genomic DNA was extracted for each condition and used as PCR template using one primer from the transposon and one primer from the gene of interest. A lack of PCR product from this PCR was taken to indicate a lack of transposon insertions into that gene of interest while an abundance of different sized bands was interpreted as the presence of a diversity of different insertions into that gene (Smith, Botstein, & Brown, 1995). In this way gene essentiality could be assessed one gene at a time on a pool of mutants transferring the effort of screening a mutant pool from agar plates to the thermal cycler (Hare et al., 2001). Alternatively uniquely tagged transposon mutants have been identified using DNA hybridization (Hensel et al., 1995). This method was followed by microarray-based techniques which used identical transposons and detected the genomic DNA flanking the insertion rather than the transposon itself. Microarray-based

insertion site determination techniques include Transposon Site Hybridization TraSH (Sassetti, Boyd, & Rubin, 2001), MicroArray Tracking of Transposon mutants (MATT) (Salama, Shepherd, & Falkow, 2004), Transposon Mediated Differential Hybridization (TMDH) (Charles and Maskell, 2001; Chaudhuri et al., 2009) and Designer Arrays for Defined Mutants Analysis (DeADMAn) (Lamichhane, Tyagi, & Bishai, 2005) which together succeeded in setting an incredibly low bar for acronym denomination. In all four techniques DNA is extracted from transposon mutant pools before and after selection so that each mutation can be amplified and labelled before subjection to a genomic microarray. All of these microarray-based methods suffer from the same problems of relatively low precision, accuracy and resolution compared to transposon sequencing (Langridge et al., 2009) as well as the need to develop a microarray for each strain they are applied to (van Opijnen, Bodi, & Camilli, 2009).

Sequencing was first used to identify transposon insertion-sites in 1999 by extraction of DNA from a mutant pool followed by digestion and circularisation with a DNA ligase as in inverse PCR. Ligation products were amplified by PCR, the products of which were digested and cloned into pUC18 vectors which were then used to sequence the insert (Clyde A Hutchison et al., 1999). The sequencing data was used to map the insertion-site to the genome which had recently been sequenced (Fraser et al., 1995). A lack of high-throughput sequencing technology meant that sequencing to identify transposon insertion-sites was less scalable than microarray-based techniques. However a return of sequencing was encouraged by the advent of next-generation sequencing

technologies and the problems associated with microarray-based methods (Barquist, Boinett, & Cain, 2013).

Transposon Insertion Sequencing (TIS) is a term for a group of techniques initially developed by four different groups in 2009 each utilising high-throughput sequencing from an amplicon sequencing library enriching for transposon-chromosomal DNA junctions. TraDIS (Langridge et al., 2009), HITS (Goodman et al., 2009), Tn-Seq (van Opijnen et al., 2009) and INSeq (Gawronski, Wong, Giannoukos, Ward, & Akerley, 2009) all involve the creation of a large transposon mutant pool followed by second-generation sequencing of the pooled DNA sample. The techniques have minor differences at the sequencing library stage. Adaptations and improvements to these techniques have been made since these initial publications, all of which sequence the transposon-genomic DNA junction in order to identify the quantity of transposon mutants for each gene in a mutant pool. Transposon sequencing is major step forward in identifying the link between genotype and phenotype by effectively allowing quantitative assessment of the composition of a transposon mutant pool. In this way a comparison between before and after selection can be performed and mutants which are significantly over or under-represented after selection can be identified.

TIS studies aim to identify the essential gene set for a given environment and to quantify fitness contributions of every gene in the genome simultaneously. To achieve these aims a large transposon mutant pool is created in which every non-essential gene is interrupted by an insertion. The DNA from this mutant pool is then extracted and PCR is used to enrich for transposon-chromosomal junctions which are then sequenced typically using a primer which hybridises to

the transposon sequence. The resulting sequencing data is then analysed and reads are mapped onto a reference genome. Any genes into which a transposon insertion event was lethal will be unrepresented or severely under-represented in the read mapping stage as lethal insertion mutants would by definition not proliferate in the library creation process. In this way a list of candidate essential genes can be produced from the sequencing data.

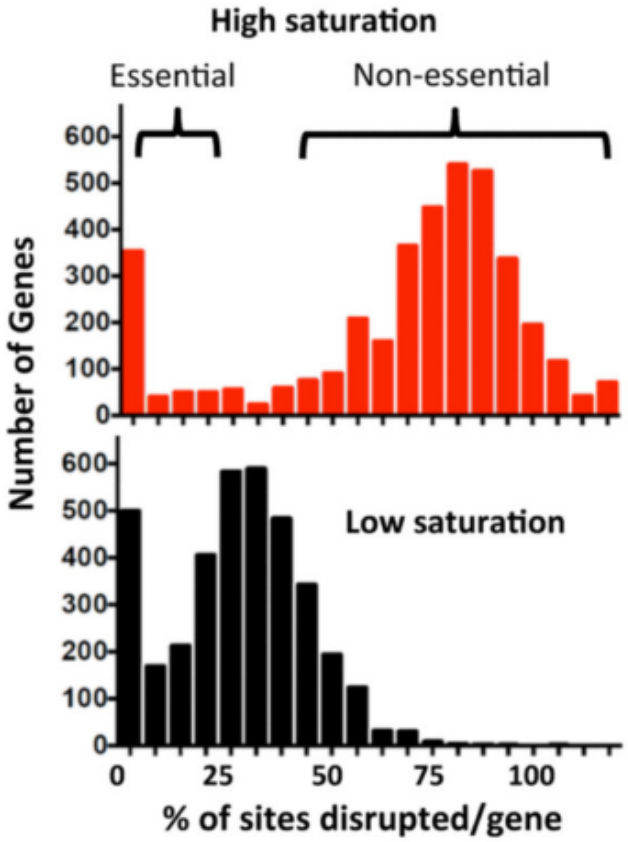


Figure 1.3: Saturation and essentiality.

A representation of the importance of achieving library saturation in order to identify essential genes with confidence. Essential genes cluster on the left of the graph with no or few insertions and form the first peak. The second peak represents non-essential genes. The space between the two characteristic peaks represents the ambiguous genes which may or may not be essential or be quasi-essential in that mutants are viable but have severely reduced fitness. Greater library saturation leads to a clearer delineation between essential and non-essential genes. (Chao, Abel, Davis, & Waldor, 2016).

To measure fitness contribution of each gene simultaneously it is necessary to subculture the mutant pool and take DNA samples after each growth period. Since the number of sequencing reads from a given gene corresponds with the relative number of cells of that mutant in the mutant pool, the growth rate of every gene can be assessed simultaneously. The correlation between sequencing reads and mutant prevalence is only belied by several biases present during sequencing including the variable amenity of DNA regions to be amplified in PCR and in relation to ambiguity of mapping certain insertion sites.

The ability to identify essential gene sets and quantitatively track transposon mutants through periods of growth in different environmental conditions allows the investigation of a range of biological phenomena. The high-throughput nature of these investigations is necessary in order for the discovery of gene function to keep pace with the volume of novel sequencing data which high-throughput sequencing allows. Diverse biological functions can be investigated by careful experimental design but in each case essentially an initial transposon mutant pool is inoculated into two or more conditions (including a control condition) and genomic DNA is extracted and sequenced after growth, allowing the differential growth success of a range of mutants to be inferred. Such biological traits as antibiotic resistance (Gallagher, Shendure, & Manoil, 2011), bile tolerance (Langridge et al., 2009), cholesterol utilization (Griffin et al., 2011), infection of murine models such as by *Yersinia pseudotuberculosis* (Crimmins et al., 2012), sporulation (Dembek et al., 2015) and heavy metal stress (Opijnen, Camilli, Opijnen, & Camilli, 2012) have been explored using TIS.

The most convincing studies demonstrating the power of TIS are those which examine already well-understood biological phenomena. DNA repair mechanisms in *E. coli* were examined using TIS subjecting a transposon pool to ionizing radiation and quantitatively determining insertion sites before and after selection. This approach identified forty-six genes significantly under-represented in the IR exposure compared to control, twenty of which had never been identified using traditional forward genetics approaches (Byrne, Chen, Wood, Cabot, & Cox, 2014). Another informative example is the discovery of twenty-four genes never previously implicated in sporulation being identified in a TIS experiment in *Bacillus subtilis* (Meeske et al., 2016). Both DNA repair and sporulation are traits which have been studied seemingly exhaustively over several decades and yet TIS studies revealed further genes implicated in these traits (Burby, Nye, Schroeder, & Simmons, 2017; Errington, 2003).

1.18 Summary

Gas fermentation enables the use of waste products to produce commodity chemicals and fuels in any geographical context without competing for food or land. Commercialisation of gas fermentation technology has begun and improvements of the biological catalyst will hasten widespread adoption and secure it as a means to a sustainable fuel economy, reducing the dependence on fossil fuels. *C. autoethanogenum* is the species employed by the leading company employing gas fermentation technology. It natively produces ethanol, acetate and 2,3-butanediol. Synthetic biology approaches aim to extend the product range to include more lucrative compounds and create increased productivity strains with respect to the native products. DNA transfer into *C. autoethanogenum* has relatively low efficiency and reliability and improvements

would facilitate strain development. Transposon mutagenesis can identify gene function and screen for potentially useful mutants. In this case useful mutants are likely to be those which have increase production levels of industrially relevant fermentation products or increased resistance to toxic compounds found in industrial settings. Transposon Insertion-site Sequencing (TIS) utilises high-throughput sequencing to determine essential gene sets and can provide quantitative assessment of fitness contribution for every gene simultaneously. TIS can therefore be used to identify candidate essential genes which cannot be inactivated via directed methods such as ACE or ClosTron. TIS can also be used to identify genes implicated in product tolerance using a comparison of mutant fitness contribution in a challenge condition compared to a control. Both of these approaches demonstrate transposon mutagenesis and sequencing as valuable tools for future strain engineering.

1.19 Aims

The research aims were to increase the efficiency of reliability of transformation of transformation of *C. autoethanogenum*. This increase in transformation rates will facilitate the implementation of transposon mutagenesis. A conditional plasmid loss system and orthogonal transposase expression system previously demonstrated in related clostridia will be tested in *C. autoethanogenum*. Ultimately a transposon mutant library suitable for Transposon Insertion Sequencing (TIS) will be created and transposon sequencing employed to determine an essential gene set for *C. autoethanogenum*.

2 Materials and Methods

2.1 Bacterial Strains and Growth Conditions

2.1.1 Bacterial Strains

Table 2.1: Bacterial strains used in this study.

Strain	Source/Reference	Description/Genotype	Genotype
Escherichia coli TOP 10	Invitrogen	Plasmid cloning and storage	F– mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galK rpsL (StrR) endA1 nupG
E. coli XL-1 Blue	Promega UK Ltd	Plasmid cloning and storage	endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_K^- m_K^+)
E. coli DH5 alpha	New England Biolabs (NEB)	Plasmid cloning and storage	F$^–$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG pur B20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^–m_K^+$), λ$^–$
E. coli HB101	Promega UK Ltd	Plasmid cloning and storage	F$^-$ mcrB mrr hsdS20(r_B^- m_B^-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(Sm^R) glnV44 λ$^-$
E. coli NEB Express	New England Biolabs (NEB)	Conversion to conjugal donor strain	fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10
E. coli CA434	M. Young, UCW, Aberystwyth, UK (Purdy et al., 2002)	Conjugative transfer strain. HB101 (unknown genotype) with R702.	F$^-$ mcrB mrr hsdS20(r_B^- m_B^-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(Sm^R) glnV44 λ$^-$ R702
E. coli InterstellaR	J. Millard	Conjugative transfer strain	F–, ara,Δ(lac-proAB) [Φ80d lacZΔM15], rpsL(str), thi, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, dam, dcm. R702
E. coli Sexpress	This study	Conjugative transfer strain	fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10. R702
InterstellaR_CLAU_0514 +CLAU_1608	This study	Conjugative transfer strain	F–, ara,Δ(lac-proAB) [Φ80d lacZΔM15], rpsL(str), thi, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, dam, dcm. R702_CLAU_0514_CLAU_1608

E. coli Sexpress_CLAU_0514+ CLAU_1608	This study	Conjugative transfer strain	fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10. R702. R702_CLAU_0514_CLAU_1608
C. autoethanogenum	DSM 10061 DSMZ Braunschweig, Germany.	Wild type	WT
C. ljungdahlii	DSM 13528T DSMZ Braunschweig, Germany.	Wild type	WT
C. autoethanogenum Δ*pyrE*	R. Breitkopf	Δ*pyrE*	Δ*pyrE*
C. ljungdahlii Δ*pyrE*	A. M. Henstra	Δ*pyrE*	Δ*pyrE*
C. autoethanogenum TcdR$^+$	This study	TcdR$^+$	TcdR$^+$
C. autoethanogenum C24	C. Humphreys	P_{bgaL}-*tcdR*	P_{bgaL}-*tcdR*
C. ljungdahlii TcdR$^+$	This study	TcdR$^+$	TcdR$^+$

2.1.2 Plasmids

Plasmid	Description	Source/Reference
pMTL-YZ14	Transposon-delivery	Zhang *et al.,* 2015
pMTL-83151	Conjugation test	Heap *et al*., 2009
pMTL-84151	Conjugation test	Heap *et al*., 2009
pMTL-83*151	Conjugation test	Heap *et al*., 2009
pMTL-86151	Conjugation test	F. Annan
pMTL-87151	Conjugation test	J. Millard
pMTL-pCB101	Conjugation test	P. Rowe
pBMTBX-4	Source of araC-Pbad	N.P. Minton.
pACYC184	Backbone for pCW10	N.P. Minton.
pCW10	MTase expression	This study
pCW24	MTase expression with apra and homology arms	This study
pKD46	Recombineering	R. Mansfield
pREC-P_{ptb}-catP	Source of apra and R702 homology arms	R. Mansfield
pMTL-CH20	*pyrE* correction	C. Humphreys
pMTL-CW20_*tcdR*	*pyrE* correction inserting *tcdR*	This study
pMTL82254-PtcdB	PtcdB reporter	Zhang et al., 2015
pMTL83251-lacI	Conditional replicon test without terminator	Zhang et al., 2015
pMTL83251-lacI-T	Conditional replicon test with terminator	Zhang et al., 2015
pMTL-CW14	Transposon-delivery	This study
pMTL82254::HZ1	PbgaL reporter	H. Wang
pMTL-CW17	Transposon-delivery	This study
pMTL-CW20	Transposon-delivery	This study
pMTL-CW21	Transposon-delivery	J. Millard
pMTL-CW22	Transposon-delivery	This study
pMTL-CW23	Transposon-delivery	J. Millard

pMTL-GL15	Suicide transposon-delivery	G. Little
pMTL-84–JH29-rbs-TcdR	Source of *tcdR* for pMTL-CH20_*tcdR* construction	N.P. Minton.
pMTL_Tet3nO_RFtetR (catP assay)	Source of Tet3nO promoter	N.P. Minton.

2.1.3 Growth Conditions

E. coli strains were grown at 30°C or 37°C in a static incubator or shaking incubator at 220 RPM. Clostridial strains were incubated at 37°C using a MG1000 Mark II Anaerobic Work Station (Don Whitley Scientific Ltd). Gas composition within the anaerobic cabinet was maintained as 80% nitrogen, 10% carbon dioxide and 10% hydrogen. Agar plates for the growth of clostridial strains was left in the anaerobic workstation for at least four hours prior to addition of cultures to ensure anaerobicity. Liquid media intended for growth of clostridial cultures was left to reduce overnight before inoculation.

C. autoethanogenum was cultured on either YTF (table 2.2) or PETC (table 2.7) media. *C. ljungdahlii* was cultured exclusively on YTF.

2.2 Chemicals and Suppliers

Plasmid mini-prep kits and genomic DNA extraction kits were sourced from Sigma-Aldrich or Promega. Gel extraction kits were supplied by Zymogen or New England Biolabs. Several polymerases were used throughout including DreamTaq (Sigma-Aldrich), Q5 (New England Biolabs), KOD (Sigma-Aldrich) and Kapa Taq (Sigma-Aldrich). Oligonucleotides were sourced from Sigma-Aldrich. All other enzymes for molecular biology were sourced from New England Biolabs as well as 2-log and 1kb+ DNA ladders. All chemicals were ordered from Sigma-Aldrich Company Ltd.

2.3 Bioinformatic Tools

DNA sequences were routinely visualised using the browser-based platform Benchling. Basic Local Alignment Searches for nucleotides and proteins were performed using the algorithms provided by the NCBI. DNA sequence alignments were performed using EMBOSS Needle provided by the European Bioinformatics Institute or using Benchling. Oligonucleotides were designed with the aid of Oligoevaluator (Sigma Aldrich) and Oligoanalyzer (Integrated DNA Technologies) for prediction of secondary structure, formation of primer dimers and melting temperature calculation. Artemis genome browser was used for visualisation of mapped transposon insertion reads. Statistical analyses were performed using GraphPad Prism.

2.4 General Microbiology Techniques

For solid media 1.5g/100ml No. 1 Bacteriological Agar (oxoid) was added except in the case of media for motility assays which used 0.25% w/v/ agar. Media sterilization proceeded at 121°C for 20 minutes.

2.4.1 Clostridial Growth Media

2.4.1.1 YTF

Table 2.2 Yeast Tryptone Fructose (YTF) medium.

Components for 1L, pH adjust to 5.8 using HCL.

Component	Quantity per litre
Yeast Extract	10 g
Tryptone	16 g
Fructose	10 g
NaCl	0.2 g
Acid 1000x trace element solution	1 ml
Basic 1000x trace element solution	1 ml
Vitamin 1000x stock solution	1 ml

2.4.1.2 1000x Acidic trace element solution

Table 2.3: 1000x Acid trace element solution.

Component	Concentration
HCl	50 mM
H_3BO_3	100 mg/L
$MnCl_2.4H_2O$	230 mg/L
$FeCl_2.4H_2O$	78 mg/L
$CoCl_2.6H_2O$	103 mg/L
$NiCl_2.6H_2O$	602 mg/L
$ZnCl_2$	78 mg/L
$CuSO_4.5H_2O$	50 mg/L
$AlK(SO_4)_2.12H_2O$	50 mg/L

2.4.1.3 1000x Basic trace element solution

Table 2.4: Basic trace element solution.

Component	Quantity
NaOH	10 mM
Na_2SeO_3	58 mg/L
Na_2WO_4	53 mg/L
$Na_2MbO_4.2H_2O$	52 mg/L

2.4.1.4 1000x B-vitamin solution

Table 2.5: B-vitamin solution.

Component	Concentration mg/L
p-aminobenzoate	114
Riboflavin	104
Thiamine	200
Nicotinate	206
Pyridoxin	510
Pantothenate (calcium)	104
Cyanocobalamin	78
d-biotin	22
Folate	48
Lipoate/thioctic acid	50

2.4.1.5 Semi-defined minimal media (SDMM)

Table 2.6: Semi-defined Minimal Media (SDMM).

Component	Concentration
MES	5 g/L
D-fructose	7.2 g/L
Casein hydrolysate	10 g/L
5M NaOH	0.2% v/v
Resazurin (5g/L)	0.05% v/v
Acid 1000x trace element solution	0.1% v/v
Basic 1000x trace element solution	0.1% v/v
Vitamin 1000x stock solution	0.1% v/v

pH adjust to 5.8 using HCl

2.4.2 PETC

Table 2.7: PETC

Component	Concentration (g/L)
NH_4Cl	1.00
KCl	0.10
$MgSO_4.7H_2O$	0.20
KH_2PO_4	0.20
$CaCl_2$	0.02
Nitrilotriacetic Acid	0.05
$Fe(SO_4)_2(NH_4)_2.6H2O$	0.05
CH_3COONa	0.25
MES buffer	20.00
Stock additions	**Concentration (% v/v)**
Trace metals (100x)	1%
Wolfes vitamins (100x)	1%
Resazurin (2 gL^{-1})(2000x)	0.05%

pH to 5.8 with NaOH

Table 2.8: Trace metal solution 100X stock.

Trace metal solution 100X	Grams per litre
Nitrilotriacetic Acid	2.00
MnSO4.H2O	1.00
Fe (SO4)2(NH4)2.6H2O	0.80
CoCl2.6H2O	0.20
ZnSO4.7H2O	0.0002
CuCl2.2H2O	0.02
NaMoO4.2H2O	0.02
Na2SeO3	0.02
NiCl2.6H2O	0.02
Na2WO4.2H2O	0.02

pH to 6 with KOH

Table 2.9: Wolfe's vitamin solution 100X stock.

Wolfe's vitamin solution 100X	Grams per litre
Biotin	0.004
Folic acid	0.004
Pyridoxine hydrochloride	0.002
Thiamine.HCl	0.010
Riboflavin	0.010
Nicotinic acid	0.010
Calcium D-(+)-pantothenate	0.010
Vitamin B12	0.0002
p-Aminobenzoic acid	0.010
Thioctic acid	0.010

Wolfe's vitamin solution was filtered sterilised and then stored at 4°C.

The base PETC medium was supplemented with 0.8M pyruvate or 1.5g per L casein hydrolysate if specified.

2.4.3 *E. coli* Growth Media

2.4.3.1 LB

Table 2.10: LB

Component	Amount
Tryptone	10 g/L
Yeast Extract	5 g/L
NaCl	5 g/L

2.4.3.2 Conjugation Selection Medium

M9 Salts: $Na_2HPO_4 \cdot 7H_2O$, KH_2PO_4, NaCl, NH_4Cl

Vitamins: Biotin, Thiamine.

Other salts: $CaCl_2$, $MgSO_4$

Carbon Source: 20ml per L media of 20% w/v glucose.

Amino Acids: Glycine, Alanine, Serine, Threonine, Cysteine, Valine, Methionine, Phenylalanine, Aspartic Acid, Glutamic Acid, Asparagine, Histidine, Lysine, Arginine.

(No Isoleucine, Glutamine, Leucine or Proline.)

2.4.4 Strain Storage and Revival

Bacterial stocks were stored in -80°C chest freezers. *E. coli* strains utilised the Microbank™ Long Term Bacterial and Fungal Storage System provided by Pro-lab Diagnostics. Clostridial stocks were generated by the addition of 20% glycerol or 20% DMSO to 1ml of clostridial culture and added to the -80°C freezer without snap freezing.

2.4.5 Preparation of Electrocompetent *E. coli*

5ml overnight cultures were prepared with the relevant antibiotics after inoculation from a single colony from an LB plate. 200ml of LB contained in a conical flask with relevant antibiotics was inoculated with a 2% inoculum from the overnight. The optical density (OD_{600}) of the culture was monitored until between OD_{600} 0.2 and OD_{600} 0.5 was obtained. The culture was split into 50ml falcon tubes and handled on ice from this point onwards. Cells were centrifuged at 4000x g for 10 minutes in a pre-cooled centrifuge at 4°C. Cell pellets from

each 50ml falcon tube were then washed with 30ml 10% glycerol and centrifuged again. Finally the washed pellets were resuspended in between 50µl-500µl sterile, chilled 10% glycerol depending on whether the cells would be needed for challenging transformations or whether more vials of cells was a priority.

2.4.6 Transformation of *E. coli* by Electroporation

Aliquots of electrocompetent cells were thawed on ice before addition of 0.5-5µl of plasmid or ligation mix. Ligation mixtures were purified using drop dialysis whereby a Type-VS Millipore membrane with 0.025µm mean pore size is floated in distilled water and the ligation mix pipetted onto the membrane and left for one hour before retrieval. Cell-DNA mixtures were then transferred to a pre-chilled Biorad electroporation cuvette with a 0.2cm gap. A biorad gene pulsar was used to deliver a pulse of 2.5kV, 200 Ω and 25 µF. Recovery media (either LB or SOC) was added post pulsing and recovery was allowed for one hour at 37°C shaking at 220 RPM. The recovery mixture was plated to selective LB agar plates and incubated overnight at 37°C or 30°C.

2.4.7 Plasmid Transfer to *C. autoethanogenum* via Conjugative Transfer

C. autoethanogenum glycerol stocks were restored onto YTF agar plates and colonies used to inoculate liquid cultures. After 48 hours culturing the liquid cultures were subcultured 24 hours prior to the mating plate stage. *E. coli* donor strains containing the target plasmid were cultured overnight before being subcultured on the morning of the mating plate stage. Donor cultures were grown for approximately 3-4 hours and the optical density monitored until it reached OD_{600} 0.2-0.4. 1ml of *E. coli* donor strain was added to a 1.5ml Eppendorf tube using a wide bore pipette tip and centrifuged at 3000 x g for 2

minutes. The pellet was washed with 500µl PBS before addition to the anaerobic cabinet. YTF agar mating plates were allowed to degas for at least four hours prior to plating. The *E. coli* donor pellet was resuspended in 200µl of *C. autoethanogenum* and spread or spotted onto the YTF agar mating plate. Mating plates were left for 20-24 hours before all growth was harvested by flooding the plate with PBS and the mating plate mixture was either applied directly to plasmid selection plates or serially diluted in PBS before plating. Plasmid selection plates were composed of YTF agar supplemented with D-cycloserine (DCS) to counter-select the *E. coli* donor as well as a relevant antibiotic to select for the target plasmid. Putative transconjugant colonies were observed after 48-72 hours.

2.4.8 Preparation of Electrocompetent *C. ljungdahlii*

200ml of unsupplemented YTF was inoculated to an OD of 0.05 from an overnight culture and the growth of this culture OD monitored until it reached between OD_{600} 0.2-0.4. At this point the 200ml culture was split into four 50ml falcon tubes and cells were pelleted in a pre-cooled centrifuge at 4°C at 4,000xg for 5 minutes. Supernatant was decanted off and the cell pellets resuspended in 30ml pre-cooled Sodium Mono Phosphate (SMP) pH 5.8 buffer. Cells were centrifuged again with previous settings and the supernatant was removed before the cell pellets were resuspended in 10ml SMP per tube. The 10ml cell mix from each tube was then pooled into one falcon tube and centrifuged as previously. The supernatant was first decanted off and then the remainder removed by pipetting carefully around the cell pellet. Finally the cell pellet was resuspended in 2ml SMP with 10% DMSO and competent cell aliquots of 100µl

were made. All centrifugation steps took place inside an anaerobic coy tent and cells were handled on ice.

2.4.8.1 SMP Buffer

A 1000x stock solution of sodium buffer was made up by adding 7.9 ml of Na_2HPO_4 to 92.1ml of NaH_2PO_4. 0.5ml of 1000x buffer solution was then added to 500ml ddH_2O alongside 1mM 1mM $MgCl_2$ and 270mM sucrose.

2.4.9 Preparation of Electrocompetent *C. autoethanogenum*

Frozen stocks of *C. autoethanogenum* were revived into YTF, after two-three days this culture was subcultured into fresh YTF medium and allowed to grow for one day before being used to inoculate 100ml YTF medium supplemented with 5ml 1.6M D, L-threonine solution. A 1% inoculum was used which usually resulted in the 200ml culture reaching OD of between 0.2-0.5 by the following morning. The 200ml overnight culture was then prepared the same as with *C. ljungdahlii* except on some occasions a lower volume of SMP/DMSO was used in an effort to increase the likelihood of obtaining transformants.

2.4.10 Transformation of *C. ljungdahlii*

A vial of prepared electrocompetent cells were removed from the -80°C freezer and gently thawed on ice. 5-10µl plasmid DNA was added to the thawed cells and left for 30 minutes. 2mm electroporation cuvettes were used and the cells were pulsed with 1000V, 50µF, 200Ω using exponential pulses for 5ms. After the pulse was delivered cells were recovered with unsupplemented YTF added immediately to the cuvette. The cuvette contents were then added to 14ml of recovery media (YTF) in a falcon tube. Recovery was allowed for 20 hours before cells were centrifuged at 4,000x g and the supernatant discarded, the

pellet was then resuspended in 200µl YTF and then plated to selective agar plates. Agar plates were incubated in the anaerobic cabinet for 48 hours.

2.4.11 Lambda Red Recombineering

Cells to be modified by recombineering were first transformed with pKD46 which contains lambda Red genes *exo*, *bet* and *gam* under the control of the arabinose-inducible promoter P_{araB} as well as a temperature sensitive replicon (repA101ts) for which growth at 37°C is the non-permissive condition. Electrocompetent cells harbouring pKD46 were made as detailed in this chapter except for growth occurring at 30°C and the inclusion of the recombinase induction step when cells reached an OD_{600} of approximately 0.2 which involved the addition of 10mM L-arabinose. Electrocompetent cells were transformed using PCR products with the intended insertion and an apramycin resistance marker between two homology arms and transformants were selected for on LB agar plates supplemented with apramycin.

2.4.12 Supplements

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

Table 2.11: Supplements

Supplement	Solvent	Working concentration (µg/ml)	
		E.coli	**Clostridia**
Ampicillin	H_2O	100	-
Chloramphenicol	100% Ethanol	25	-
Erythromycin	100% Ethanol	500	-
Clarithromycin	100% DMSO	-	6
Kanamycin	H_2O	50	-
Spectinomycin	H_2O	250	-
Tetracycline	70% Ethanol	10	-
Thiamphenicol	100% DMSO	-	15
Trimethoprim	100% DMSO	10	-
IPTG	H_2O	-	238
D-cycloserine	H_2O	250	-
5-Fluoroorotic acid (light sensitive)	100% DMSO	-	1000
50x uracil	H_2O	-	20

2.5 General Molecular Techniques

2.5.1 Phenol-chloroform Extraction of Genomic DNA from *C. autoethanogenum* or *C. ljungdahlii*

Between 1ml and 5ml of culture was centrifuged at 4000 x g for 3 minutes at room temperature. Supernatant was removed by decanting followed by pipetting. Bacterial pellets were re-suspended in 200µl PBS with 10mg/ml lysozyme. Cells were lysed at 37°C for 30 minutes with occasional mixing. After the lysis step, 25µl proteinase K, 85µl dH2O and 100µl 10% SDS solution were added and vortexed. Incubation of the mixture proceeded at 55°C for 30 minutes with occasional vortexing. 400µl Chloroform: Isoamyl alcohol (25:24:1) with 10mM Tris, pH 8.0, 1mM EDTA was added to phase-lock gel tubes and after proteinase treatment the samples were added to the prepared phase-lock tubes. Samples were then centrifuged at 14,000x g for 2 minutes at room temperature allowing the extraction of the top layer to which 20µl 10mg/ml

RNAse (Sigma Aldrich) was added and incubated for 2 minutes at room temperature. After RNAse treatment the samples were purified with another three phase-lock tubes with phenol-chloroform purifications transferring the top layer each time. Finally the top layer was transferred to a pre-chilled 1.5ml eppendorf containing 40 µl of 3M Sodium acetate and 800 µl of 100% ethanol. Samples were mixed and then placed at -80°C for at least 30 minutes, often overnight. A centrifuge was pre-chilled to 4°C before the chilled sample was placed in it and centrifuged at 14,000 x g for 15 minutes. The supernatant was carefully removed by aspiration. 1ml of 70% ethanol was added to the pellet and a further centrifugation step was performed at 14,000 x g for 5 minutes. The supernatant was poured off and the remainder removed carefully using a P200 pipette. The pellet was air-dried for around 45 minutes before being resuspended in 50-100µl sterile dH_2O.

2.5.2 Preparation of Plasmid DNA from *E. coli*

Plasmid DNA was recovered from overnight *E. coli* cultures using mini-prep kits from either sigma-aldrich or NEB following the manufacturer's instructions.

2.5.3 Restriction Enzyme Digests

Restriction enzymes and buffers were obtained from New England Biolabs (NEB) and performed as per the associated protocol.

2.5.4 Dephosphorylation

Preventing of backbone vector recircularisation was achieved using Antarctic phosphatase from NEB following the manufacturer's instructions.

2.5.5 Ligations

For routine cloning T4 DNA ligase was obtained from either NEB or Promega and the manufacturer's instructions were followed. Ligation mixtures were incubated overnight in thawing ice to provide a temperature gradient and purified using 0.025 µm Millipore membrane discs over dH_2O for 15-60 minutes prior to electroporation.

2.5.6 Agarose Gel Electrophoresis and Visualisation

DNA fragments were visualised after separation on agarose gels ranging in concentration from 0.7%-2.5% (w/v) using agarose powder supplied by MP Biomedicals dissolved in TAE buffer (40mM Tris, 1mM EDTA, 0.1% glacial acetic acid). Sybr safe DNA gel stain (thermofisher) was used for binding and visualisation of DNA either by immediate addition to poured agarose gels or by post-staining agarose gels. Low melt agarose was used to purify chromosomal DNA and separate it from plasmid DNA prior to TraDIS sequencing library preparations. During TraDIS sequencing library preparation the final transposon-chromosomal junction enrichment PCR products were extracted from a post-stained ultra low range agarose gel. Imaging of agarose gels was achieved using a transilluminator or an Odyssey® Fc Imaging System.

2.5.7 DNA extraction from Agarose Gels

Agarose gels stained with sybr safe were visualised using a blue light box and target fragment bands were excised using a scalpel blade. Gel fragments were then dissolved and DNA purified using gel extraction kits from either sigma-aldrich, zymo research or NEB following the manufacturer's instructions.

2.5.8 DNA Fragment Sequencing

Sequencing of DNA either plasmids or purified PCR products was performed by Source Bioscience or Eurofins. Sequencing primers were supplied to the companies or in the case of TA-cloning in troubleshooting of TraDIS the T7 primer sequencing was requested.

2.5.9 HiFi Assembly

The HiFi assembly kit from NEB was used to construct the majority of constructs in this text. DNA fragments were generated either via PCR or via restriction digest of a plasmid which would serve as the backbone for the constructed vector. PCR primers were designed to include overlap regions of 20-25bp. DNA fragments were quantified using a nanodrop and combined in the recommended ratios according the manufacturers instructions. The exonuclease, DNA polymerase and DNA ligase activities occurred in a single tube containing all the composite fragments and was left at 50°C for between 1-4 hours before being used to transform *E. coli* cloning strains.

2.6 Generation of Clostridial Mutants via Allele Coupled Exchange

2.6.1 Generation of the *pyrE* Repair Vector Inserting *tcdR.*

In order to create a strain of *C. autoethanogenum* expressing the *tcdR* gene from *C. difficile* a *pyrE* knockout strain was obtained from Ronja Breikopf into which a *pyrE* correction vector was transferred via conjugation. The *pyrE* correction vector contains homology arms designed to restore the function of *pyrE.* The *pyrE* correction vector pMTL-CH20 can be adapted to insert genetic

cargo into the *pyrE* locus while restoring *pyrE* function. The restoration of *pyrE* can be selected for via growth on minimal media lacking uracil.

pMTL-CH20_*tcdR* was created using the pMTL-CH20 PyrE repair vector as a template and inserting the *tcdR* gene from a second plasmid pMTL-84 –JH29-rbs-TcdR. The *tcdR* gene with its native ribosome binding site excised from pMTL-84-JH29-rbs-TcdR as a NotI – NheI fragment and ligated into the equivalent site of the pMTL-CH20 pyrE repair vector which was also digested with NotI and NheI. The ligation was used to transform Invitrogen one shot top ten chemically competent cells which were grown to amplify the desired plasmid which was then used to transform the conjugation donor strain E. coli CA434.

2.6.2 Restoration of *pyrE* Function and Insertion of *tcdR* at the *pyrE* Locus in *C. autoethanogenum*.

Conjugation proceeded as outlined with transconjugants being selected on YTF plates containing thiamphenicol (15µg/ml) and uracil (20µg/ml). Thiamphenicol resistant colonies at this stage could be as a result of a single cross-over event with the entire plasmid backbone incorporate including the *catP* resistance marker, alternatively they could simply harbour the plasmid with no cross-over events having occurred. Colonies appearing on thiamphenciol, uracil plates were then streaked onto semi-defined minimal media plates deficient in uracil to select for double crossover mutants with restored *pyrE* function.

2.6.3 Restoration of *pyrE* Function and Insertion of *tcdR* at the *pyrE* Locus in *C. ljungdahlii*.

The homology arms within the *pyrE* restoration vector pMTL-CH20 share 100% sequence identity with the corresponding region in the *C. ljungdahlii* genome

allowing the same vector to be used in *C. ljungdahlii* as was used in *C. autoethanogenum* to restore the function of *pyrE* and insertion *tcdR* at the *pyrE* locus. The mutant generation procedure was the same between the two organisms, except that in the case of *C. ljungdahlii* the pMTL-CH20_*tcdR* plasmid was transferred via electroporation rather than conjugation. The *C. ljungdahlii pyrE* knockout strain was obtained from Anne Henstra.

2.7 Transfer of R702 Between *E. coli* Strains

2.7.1 Conjugation of Modified R702 from InterstellaR to NEB Express

R702 is the R factor containing transfer function responsible for conjugal transfer from *E. coli* to *C. autoethanogenum.* In this study, it was modified to incorporate two *C. autoethanogenum* methyltransferases encoded by CLAU_0514 and CLAU-1608. Initially this modification took place on the R702 harboured by the *E. coli* strain InterstellaR giving the strain InterstellaR_ CLAU_0514_CLAU_1608. Later, the modified R702 was transferred via conjugal transfer from InterstellaR to Sexpress. This section describes that process. Overnight cultures of InterstellaR_CLAU_0514_CLAU_1608 in LB supplemented with kanamycin and apramycin and overnight cultures of NEB Express in unsupplemented LB were used to inoculate fresh media with 2% inoculum. After 3 hours growth the OD_{600} was measured and cultures mixed in liquid culture at an OD adjusted 1:1. The conjugation mix was incubated statically at 37°C for one hour. Transconjugants were selected for on the conjugation selection medium detailed in 0 which selects against InterstellaR (which is a proline auxotroph) and was supplemented with apramycin and kanamycin to select for the presence of the modified R702. Colonies appearing

on this selective medium should therefore be NEB Express cells which can grow without proline supplementation and have attained apramycin and kanamycin resistance from the modified R702 which was transferred by InterstellaR_CLAU_0514_CLAU_1608.

2.7.2 Conjugation of R702 from CA434 to NEB Express

Overnights of CA434 in LB supplemented with spectinomycin and NEB Express in unsupplemented LB were used to inoculate fresh media with 2% inoculum. After 3 hours growth the OD_{600} was measured and cultures mixed in liquid culture at an OD adjusted 1:1. The conjugation mix was incubated statically at 37°C for one hour. Transconjugants were selected for on the on the conjugation selection medium detailed in section 2.4.3.2 which selects against the proline/leucine auxotroph CA434 and was supplemented with tetracycline to select for the presence of R702.

2.8 Chloramphenicol Acetyl Transferase (CAT) Assay

The *catP* gene encodes a chloramphenicol acetyltransferase (CAT) which confers resistance to chloramphenicol/thiamphenicol as well as serving as a reporter gene. As a reporter gene its level can be measured with the method of (Shaw, 1967). The amount of CAT is calculated by measuring the initial rate of reaction as determined by the appearance of a free CoA sulfhydryl group. 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) reacts with the free CoA sulfhydryl group to produce 5-thio-2-nitrobenzoate which has a molar extinction coefficient of 13,600 at 412nm making it easy to detect using a spectrophotometer. A solution was prepared containing chloramphenicol (0.005%), acetyl CoA (0.167mM), DTNB (0.0833mM) and 100mM Tris buffer pH 7.8. 1ml cell samples were

centrifuged and then resuspended in 1ml 10mM Tris pH7.8 with 20µl calbiochem protease inhibitor cocktail VII. Samples were sonicated with 5 cycles of 30 seconds on 30 seconds off before centrifugation at full speed, 4°C for 30 minutes. Samples were placed in a 96-well plate and the Tecan Infinite® M1000 PRO microplate reader used to inject the master mix and then measure absorbtion at 412nm. Levels of CAT were normalised to the level of protein as determined with a Bradford assay. The Bradford assay reagent was sourced from Sigma-Aldrich and used as per the manufacturer's instructions.

3 Improving DNA Transfer

3.1 Introduction

3.1.1 DNA Transfer to *C. autoethanogenum*

DNA transfer protocols have been established in *C. autoethanogenum* including the use of both electroporation and conjugation to deliver exogenous DNA (Köpke & Liew, 2011). Of these, conjugal transfer is considered to be more reliable. The *E. coli* strain CA434 which harbours the R factor R702 responsible for plasmid transfer has been used successfully (Mock et al., 2015; Williams, Young, & Young, 1990). However, the introduction of a transposon-delivery vector to *C. autoethanogenum* via electroporation is preferable to conjugation for a number of reasons. In the first instance, the utilisation of frozen competent cell stocks and frozen plasmid preparations means there is no need to carefully monitor donor and recipient cultures. Another advantage specific to transposon work is that electroporation negates the risk of unwanted transposition events during the growth of donor cells. Unfortunately *C. autoethanogenum* is far more amenable to DNA transfer via conjugation rather than electroporation. This is somewhat surprising given that *C. ljungdahlii,* which shares 100% sequence identity in 16S rRNA, is readily transformable via electroporation. The pCB102 origin of replication used for the majority of the work in this text has an electroporation efficiency of only 0.2 cfu/µg DNA in *C. autoethanogenum* compared to 2,667 cfu/µg in *C. ljungdahlii* (Liew, 2015).

3.1.2 Comparison to *C. ljungdahlii* and Importance of Methylation

There are two likely explanations for the relative intractability of *C. autoethanogenum* compared to *C. ljungdahlii*. The first is the presence of 12 extra putative Restriction-Modification (R-M) related genes and the second is the presence of a CRISPR system in the *C. autoethanogenum* genome. Evidence for the relevance of the R-M systems of *C. autoethanogenum* comes from the observation that plasmid DNA recovered from *C. autoethanogenum* show 20 to 80-fold increases in transformation efficiencies when re-transformed (Liew, 2015). The acquisition of the methylation pattern of *C. autoethanogenum* in the recovered plasmids is the likely cause of the increased transformation rates.

3.1.3 Restriction-Modification Systems of *C. autoethanogenum*

The methylation profile of *C. autoethanogenum* has been determined through PacBio sequencing and deposited in the REBASE database (Utturkar et al., 2015). Three sequences have been identified at which 6-mA methylation occurs and the REBASE database lists four putative methyltransferases (MTases) which could be responsible. In addition to the two predicted Type II restriction enzymes there are five ORFs predicted to be Type IV restriction enzymes. Type IV restriction enzymes are relatively poorly characterised, but typified by their ability to cut DNA with foreign modification patterns (Loenen & Raleigh, 2014). From this cursory examination of the RM systems of *C. autoethanogenum* the strategy to circumvent these systems presents itself – avoid foreign modification patterns which the Type IV systems recognise, then identify which methylation patterns are recognised by the Type II restriction enzymes and either mimic the methylation pattern by modification of the *E. coli* donor strain or create a

knockout *C. autoethanogenum* strain for the relevant restriction enzymes. There are merits to either strategy and both are regularly employed in studies of this kind; for knockouts see (Cui et al., 2012) and for replication of methylation pattern see (Lesiak, Liebl, & Ehrenreich, 2014). In general the creation of a knockout strain lacking the implicated restriction enzyme is simpler assuming that reasonable DNA transfer protocols are already in place but this results in a base strain which is genetically further from wild type, which is undesirable for further studies. The restriction enzyme may have secondary functions and the knockout strain may be more susceptible to phage attack, which is of particular relevance when considering long duration bioreactor studies as is common for *C. autoethanogenum*. In some cases, including the case of *C. autoethanogenum*, the organism possesses a fused Type II system in which the same enzyme encompasses both restriction and methylation functions. A knockout strain of a fused Type II system will therefore have a different methylation pattern to wild type which can lead to global transcription and/or DNA replication changes due to the regulatory role of the methylation (Vasu & Nagaraja, 2013). In this study an attempt was made to clone the methyltransferases of *C. autoethanogenum* and express them from the *E. coli* resistance transfer factor R702 allowing delivery of target plasmids with similar methylation patterns to *C. autoethanogenum* and circumvention of native R-M systems.

3.1.4 Type IV Systems and Native *E. coli* Methylation

The majority of *E. coli* laboratory strains contain at least the two methyltransferases, *dam* and *dcm*. The methylase *dam* is responsible for adding a methyl group onto the N6 position of adenine residues in the sequence

G**A**TC, while *dcm* causes the carbon at position 5 to be methylated in cytosine residues contained in the sequence C**C**WGG. *dam* and *dcm* are the only methyltransferases found in the conjugal transfer strain CA434 or the storage strains from which plasmids were prepared in electroporation attempts, accordingly no other *E. coli* methylation types will be discussed here.

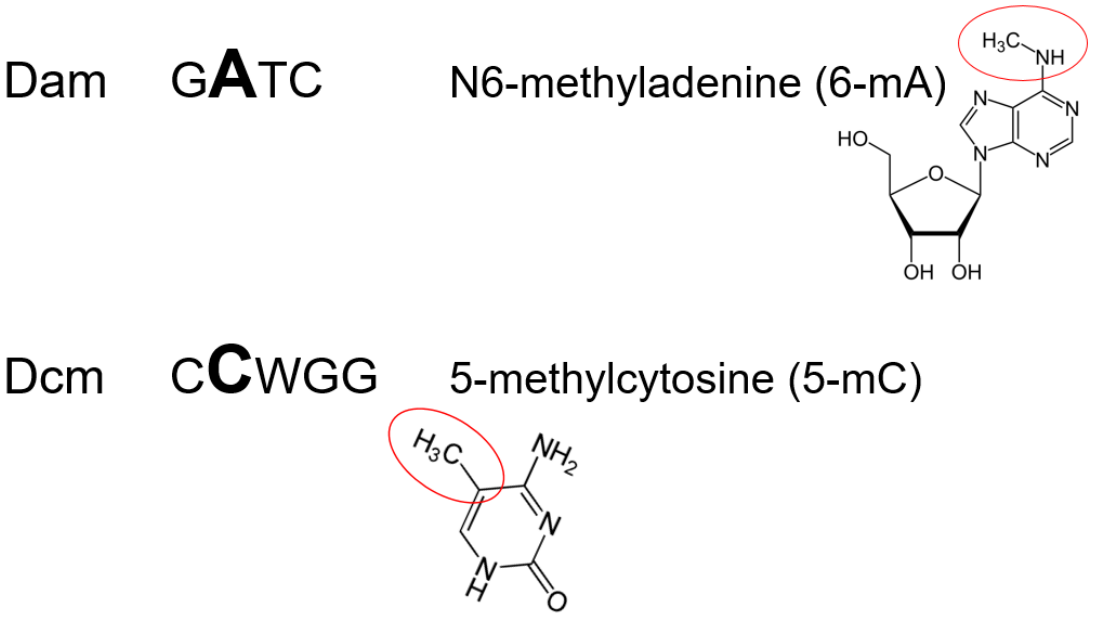


Figure 3.1 Common methylation patterns in *E. coli*.

The methylated bases and two common methylation patterns in *E. coli* produced by the MTases Dam and Dcm.

These methylation patterns are very common in plasmids relevant to this study, for instance there are 16 GATC sequences and 13 CCWGG sequences in the first transposon plasmid used (pMTL-YZ14). However, the Type IV systems of *C. autoethanogenum* are annotated as being cytosine specific so this being true, only the CCWGG sequences should concern us. This is fortuitous since *dam*- strains of *E. coli* have a mutation rate approximately 200-fold higher than

wild type (Marinus, 2010). A *dam*+ *dcm*- conjugal donor strain was created containing the R702 R factor responsible for plasmid transfer.

3.1.5 Rationale for Attempting to Improve DNA Transfer

The transposon delivery plasmid intended for use in this study (pMTL-YZ14) is larger than most plasmids routinely used previously (8.5kb) and has a tendency for the transposon to insert into the Gram-positive origin of replication, these two factors likely led to the persistent inability to transform *C. autoethanogenum* with pMTL-YZ14 even after more than 30 attempts. This inability to begin transposon mutagenesis led to efforts to increase the transformability of this organism as well as the decision to simultaneously perform transposon mutagenesis in the closely related organism *C. ljungdahlii.*

Increasing the frequency of DNA transfer to *C. autoethanogenum* is valuable for a variety of applications. Specific to transposon mutagenesis, a high DNA transfer efficiency is a requirement of the use of suicide vectors or electroporation with transposomes (transposase-transposon complexes), both of which are highly beneficial in Transposon Insertion Sequencing (TIS) experiments. If an organism is not readily transformable then replicative delivery vectors are required for transposon library generation, which can create significant problems at the DNA sequencing library preparation stage since they result in a large proportion of sequencing reads into the delivery plasmid rather than the genome. In addition the use of replicative transposon delivery plasmids tends to produce transposon libraries with low complexity since it is possible for a single mutant to form a large portion of the final mutant pool. These issues are discussed in more detail in chapters 5 and 6. An effective DNA transfer protocol circumvents these difficulties if replication-deficient DNA is used to

create the library. An increased transformation rate is still of use with replicative plasmids as it mitigates the problem of early transposition events dominating the mutant pool by combining many distinct colonies from initial plasmid selection plates.

Another application for which increasing the DNA transfer rate could be beneficial is CRISPR genome editing which has recently been exemplified in *C. autoethanogenum* (Nagaraju et al., 2016). In this process a replicative plasmid expressing *cas9* and a guide RNA specific to the target is introduced to the organism allowing for the direct selection of rare double crossover events dictated by the presence of homology arms on the vector. Expression of *cas9* and the guide RNA will in theory result in double-stranded breaks in the host chromosome. As *C. autoethanogenum* does not possess a non-homologous end joining (NHEJ) activity it is unable to repair these breaks, meaning that in theory the only viable cells are those in which a double crossover has occurred and consequently the target sequence has been removed. If an appropriate inducible promoter system is unavailable then constitutive promoters will be used to express *cas9,* in which case a high DNA transfer rate is highly valuable since the chance of obtaining the desired double cross-over event is a function of the number of plasmid-harbouring cells multiplied by the probability of a double cross-over event occurring.

3.2 Results

3.2.1 Conjugation Technique

The first approach to increasing the conjugation efficiency into *C. autoethanogenum* was to experiment with various aspects of the conjugation protocol including growth stage of the donor and recipient, ratio of donor to recipient, growth media and time on mating plate. Technical alterations were made including centrifuging donor strains at a lower speeds, allowing the donor pellet to become anaerobic before resuspension in the recipient culture and avoiding pipetting the donor strain at any point prior to addition to the mating plate to avoid excessive shearing forces on the donor cells. All of these efforts used the *E. coli* CA434 donor strain.

None of these experiments produced statistically significant improvements in conjugation efficiency. One persistent problem was that of reproducibility between biological replicates. The tendency for conjugation attempts to fail to produce any colonies frustrated many research efforts in this laboratory.

3.2.2 Creation of a Novel Conjugal Donor Strain - Sexpress.

The standard conjugal donor used in this laboratory at the start of this study was CA434 - an HB101 strain of *E. coli* carrying the R factor R702 which contains the transfer function. A salient genotype of this strain is that it is *dam*+, *dcm*+ meaning any plasmids passaged through it will acquire the methylation pattern indicated in Figure 3.1. This methylation pattern may leave DNA susceptible to attack by the Type IV restriction systems of *C. autoethanogenum*. A suitable *E. coli* strain to serve as a basis for a novel conjugal donor with a preferable methylation status was found in NEB Express. NEB Express has no

dcm methylation, though it still has *dam* methylation. This strain was chosen as all of the Type IV system genes of *C. autoethanogenum* are predicted to exhibit cytosine-specific activity, rendering them ineffective on *dam* methylation sites. In addition, *dam*- strains are known to have high mutation rates which is an undesirable phenotype for a conjugal donor (Marinus, 2010).

A novel conjugal donor strain was created via conjugal transfer of the R factor R702 from CA434 into a strain of NEB Express. Donor and recipient cultures were incubated statically together for one hour at 37°C before being plated to media lacking leucine and proline to select against the donor CA434 and supplemented with tetracycline to select for R702. Transconjugants were re-streaked on the same selective media and screened for the presence of R702 and identity as NEB Express, through acquisition of antibiotic resistance and absence of auxotrophies, respectively. All transconjugants screened were shown to be NEB Express cells harbouring R702 as expected and this strain was subsequently named Sexpress. Sexpress was predicted to serve as an effective conjugal donor to *C. autoethanogenum* because of the presence of R702 from the proven CA434, but the lack of *dcm* methylation was hoped to allow target plasmid DNA to avoid restriction by the Type IV restriction systems of *C. autoethanogenum*.

3.2.2.1 Conjugation Efficiency Measurement of Sexpress vs. CA434 in *C. autoethanogenum*

To determine if the novel conjugal donor strain Sexpress produced a significant improvement in performance over the established strain CA434 a conjugation efficiency experiment was performed, using a range of Gram-positive origins of replication including pCD6, pCB102, pCB102 frame shift version, pCB101,

pCACET, pIP404 and p19. The protocol for this experiment is detailed in the materials and methods chapter 2. Briefly, donor strains containing pMTL8x151 (*catP* antibiotic marker) were subcultured from overnight cultures before being centrifuged at 3,000x g, washed in anaerobic PBS and mixed with recipient 24 hour *C. autoethanogenum* wild type cultures at an OD_{600} of approximately 1. Donor and recipient were incubated together anaerobically on YTF plates without selection for 20 hours at which point the plates were flooded with PBS and all growth dislodged from the surface of the agar with a wedge shaped spreader, aspirated by pipetting and transferred to solid selective media for selection of the target plasmid. Sexpress proved to be a superior conjugal donor to CA434 in every replicon tested for which transconjugants were obtained (Table 3.1)

Table 3.1 Conjugal transfer efficiency comparison, CA434 and Sexpress.

Data represents averages of conjugations performed in triplicate from the point of *E. coli* donor strain colonies. Conjugations were performed as detailed in materials and methods 2.4.7 by C. Woods, C. Humphreys, P. Rowe and F. Annan.

	Transconjugant Colonies			
Replicon	**CA434**	**Sexpress**	**No. of CCWGG Sites**	**Fold-increase**
pCACET	0	0	9	-
pCB101	0	2.1	9	-
pCD6	0.3	4.6	9	15
pCB102	20.5	180.4	9	9
pCB102FS	18.1	201.8	9	11
p19	53.5	1974.3	10	37
pIP404	3.25	2418.4	12	744

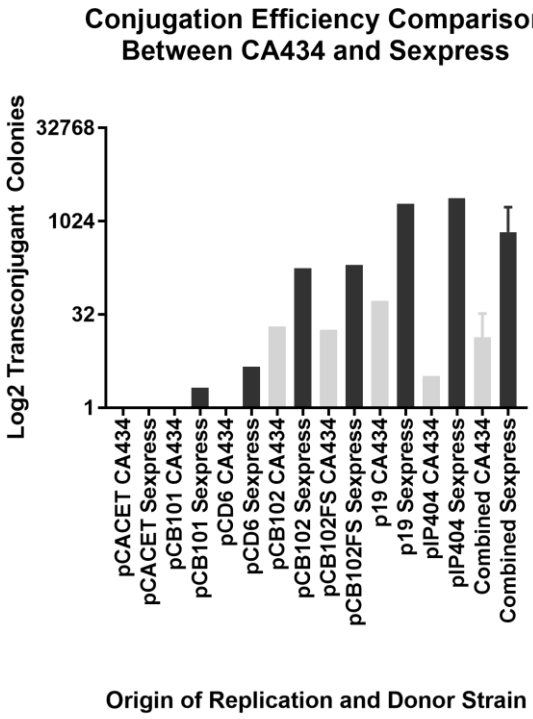


Figure 3.2: Conjugation efficiency comparison between Sexpress and CA434.

A graphical representation of the data presented in Table 3.1

3.2.2.2 Selected members of *Clostridium* and Type IV Systems

Table 3.2: Type IV systems in selected clostridia.

Selected clostridial species detailing the presence of Type IV restriction systems.

Species	REBASE Organism Number	Type IV System Present?	Number of Type IV Genes Predicted by REBASE
C. aceticum	15750	No	0
C. acetobutylicum 824	1061	No	0
C. autoethanogenum	11223	Yes	5
C. beijerinckii 59B	14798	Yes	1
C. botulinum A str. Hall	4821	No	0
C. butyricum	11074	Yes	1
C. carboxidivorans P7	6499	Yes	3
C. difficile 630	3249	Yes	1
C. difficile R20291	5600	Yes	1
C. ljungdahlii	6104	Yes	1
C. pasteurianum	1072	No	0
C. perfringens	1074	No	0
C. saccharobutylicum	11415	Yes	1
C. sporogenes	15809	Yes	1
C. tetani	3692	Yes	1
Eubacterium limosum	14951	No	0
Eubacterium limosum ATCC8486	23903	No	0
Eubacterium limosum KIST612	6384	Yes	1

As shown in Table 3.2 there are a range of clostridial species for which a *dcm*- donor strain might prove beneficial to increase conjugation efficiencies. Alternatively if electroporation is preferred the plasmid DNA should be prepared from *dcm*- strains.

3.2.2.3 Sexpress in *C. difficile*

One such strain for which it was hypothesised that Sexpress may prove valuable was the clinical isolate of *C. difficile* (DH1916). Unpublished work

performed by P. Ingle demonstrated the value of Sexpress when working with this strain. Five Gram-positive origins of replication were trialled using test plasmids containing the chloramphenicol resistance marker *catP*. Conjugations were performed in triplicate from either CA434 or Sexpress. As in *C. autoethanogenum*, Sexpress proved to be a more efficient conjugal donor than CA434 in each replicon tested (Table 3.3).

Table 3.3: Conjugation efficiency comparison between CA434 and Sexpress in *C. difficile* DH1916. Data comparing conjugation efficiency between CA434 and Sexpress in *C. difficile* DH1916 produced by P. Ingle (unpublished).

Test plasmid	Gram-positive origin of replication	Conjugation Efficiency (Transconjugants per Recipient Cell)	
		CA434	Sexpress
pMTL81151	None	0	0
pMTL82151	pBP1	$4.94x10^{-8}$	$7.41x0^{-8}$
pMTL83151	pCB102	$1.09x10^{-9}$	$1.08x10^{-8}$
pMTL84151	pCD6	$3.02x10^{-8}$	$1.27x10^{-6}$
pMTL85151	pIM13	0	0
pMTL86151	pIP404	0	0

3.2.3 Mimicking *C. autosethanogenum* Methylation Patterns

3.2.3.1 Analysis of the *C. autoethanogenum* Methylation Profile and Native MTases

Having increased the conjugation efficiency through alteration of the methylation status of the donor strain to avoid the Type IV systems of *C. autoethanogenum,* it was thought beneficial to further mimic the recipient methylation pattern via the cloning and expression of *C. autoethanogenum*

methyltransferases in a donor strain. REBASE predicts a maximum of four possible methyltransferases in the *C. autoethanogenum* genome though PacBio data has revealed only three methylation patterns. There is only one clearly defined Type II restriction enzyme in the genome, a fused Restriction-Modification enzyme designated CLAU_0514. The three sequences which were found to have methylated residues by PacBio sequencing were CAAAA**A**R, GTTA**A**T, SNNGCA**A**T with the (modified base highlighted in bold).

CAAAA<u>A</u>R

The CAAAAAR methylation motif has also been identified in *Acetobacterium woodii* DSM 1030, *C. botulinum* CDC_67190, *C. pasteurianum* ATCC 6013 (DSM525) and *C. estertheticum subsp. estertheticum*. A protein Basic Local Alignment Search Tool (BLAST) search was performed against each of these species using the translated product of CLAU_1258 (ALU35687.1) as well as for CLAU_2858 (ALU37285.1) in order to show the query coverage and identity that might be expected from an enzyme with a similar function, but a different motif recognition. Each protein BLAST result showed a clear homolog of CLAU_1258 in each species which possess the CAAAA**A**R motif, but not in *Clostridium thermocellum DSM 1313* which lacks it making it the candidate methylase responsible for CAAAA**A**R. REBASE lists CLAU_1258 as M.Cau10061I and states it has been cloned and the CAAAA**A**R pattern confirmed.

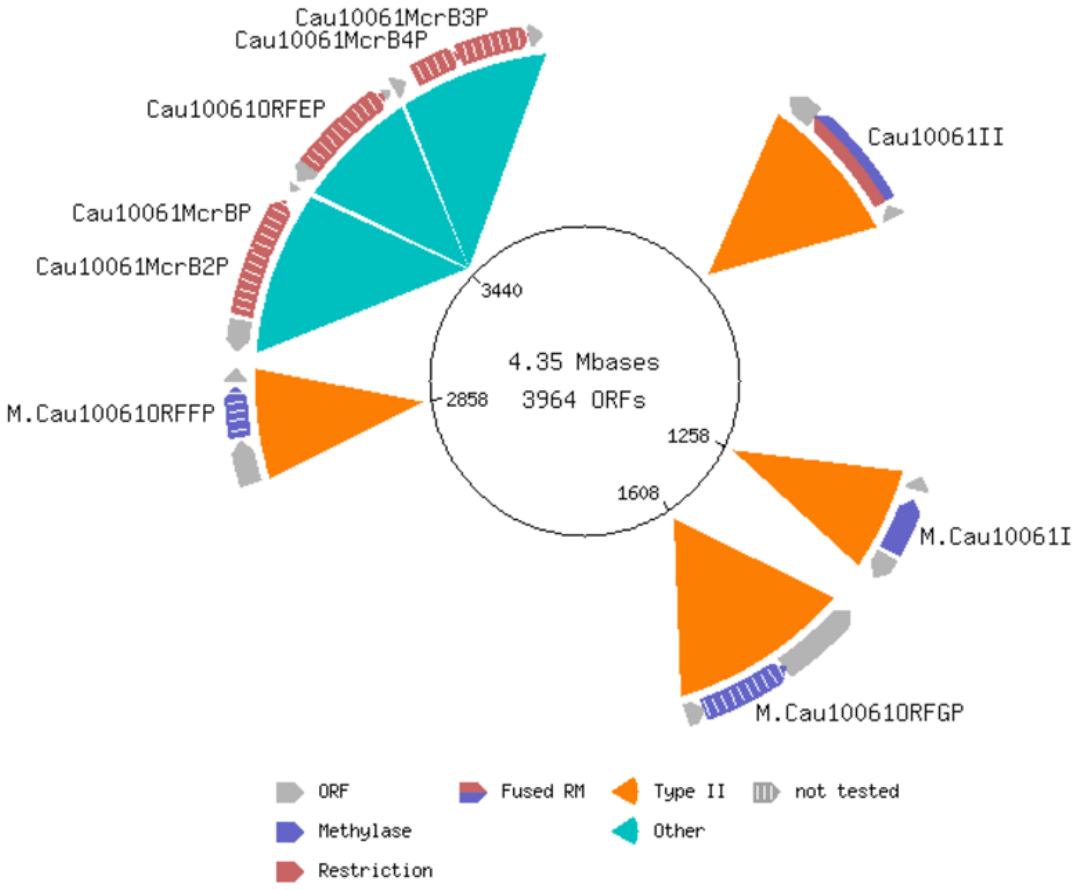


Figure 3.3: RM systems of *C. autoethanogenum* REBASE archive.

C. autoethanogenum has a range of restriction systems as predicted by REBASE. These include type IV systems (blue highlight) and four predicted methyltransferases including one fused restriction/modification system (Cau10061II).

Table 3.4: BLAST results of two *C. autoethanogenum* MTases against several clostridia.

	CLAU_1258, ALU35687.1		CLAU_2858, ALU37285.1	
Organism	**Max Query Coverage**	**Identity (%)**	**Max Query Coverage**	**Identity (%)**
Acetobacterium woodii DSM 1030	85	38	55	23
C. pasteurianum ATCC 6013*	99	54	65	29
Clostridium estertheticum	97	50	54	28
Clostridium thermocellum	6	33	10	29

GWTAAT

CLAU_0514 is named RM.Cau10061II in REBASE where it is reported to methylate the motif GTTAAT which makes it the candidate for the GWTA**A**T motif found in *C. autoethanogenum.* Interestingly *C. sporogenes* DSM795 has a PacBio entry in REBASE (ref# = 20306) with the sequence GATA**A**T. A protein BLAST of CLAU_0514 was performed against *C. sporogenes* taxid:1509 and a homologue was identified with 96% query coverage and 42% sequence identity. *C. ljungdahlii* also possesses an identical homolog to RM.Cau10061II and so this enzyme is unlikely to account for the transformation differences between the two species.

SNNGCAAT

There are two potential candidates responsible for this methylation pattern, CLAU_2858 (ORF FP) and CLAU_1608 (ORF GP). None of the genes flanking either methyltransferase are annotated as being restriction enzymes, and neither are any nearby genes. Bioinformatic inquiries were made into the four flanking genes. Of these, CLAU_1607 appeared to be a reasonable candidate restriction enzyme. An analysis of the protein domains of CLAU_1607 was

performed. The presence of an N-terminal putative catalytic domain of an HKD family nuclease (cd09179) suggests that this is the most likely of the four flanking genes to have restriction activity.

CLAU_2858 (ORF FP) has an identical gene in *C. ljungdahlii* and homologs in many clostridia. CLAU_1608 has no homolog in *C. ljungdahlii* but does have homologs throughout the Genus. The two genes thought most likely to be beneficial to express in a donor strain were therefore CLAU_0514 encoding a fused Type IIG Restriction-Modification enzyme and CLAU_1608 a methyltransferase not found in *C. ljungdahlii* located upstream of a potential cognate restriction enzyme.

3.2.3.2 Comparison to *C. ljungdahlii* RM Systems

Table 3.5: Comparison of *C. ljungdahlii* and *C. autoethanogenum* RM systems.

Type	Activity*	Name	Predicted Recognition	*C. autoethanogenum* homolog
I	M	M.CljORF3310P	-	None. Similar to CLAU_1258
I	S	S.CljORF3310P	-	None
I	R	CljORF3310P	-	None
II	M	M.CljORF3480P	-	None
II	RM	CljORF8460P	-	CLAU_2858
II	RM	CljORF24700P	GTTAAT	CLAU_0514
II	M	M.CljORF30980P	-	None
II	RM	CljORF32690P	-	None. Similar to CLAU_0514
II	M	M.CljORF33980P	CAAAAAR	CLAU_1258
IV	R	CljMcrBP	-	CLAU_3527

* Methylation (M), Specificity (S) or Restriction (R)

Perhaps the relative lack of transformability of *C. autoethanogenum* compared to *C. ljungdahlii* is due to the presence of an extra R-M system, in which case it makes sense to avoid cloning genes which have *C. ljungdahlii* homologs and instead concentrate on those unique to *C. autoethanogenum*.

CljORF8460P is the homolog of CLAU_2858 (ORF FP) which is one of two candidates which methylate the sequence SNNGCA**A**T. The other candidate CLAU_1608 (ORF GP) appears more likely to be relevant to efforts to improve DNA transfer since a nearby ORF seems to be a reasonable candidate for a cognate restriction enzyme, whereas CLAU_2858 has no nearby ORFs which look likely to have restriction activity. CljORF3310P is a Type I restriction system not found in *C. autoethanogenum.* The enzymes responsible for GTA**A**T and CAAAA**A**R methylation motifs in *C. autoethanogenum* are also found in *C. ljungdahlii*.

3.2.3.3 Expression of *C. autoethanogenum* methyltransferases in *E. coli* Donor Strains

3.2.3.4 Overall Strategy

It was decided to express the methyltransferases of *C. autoethanogenum* via the R factor responsible for plasmid transfer – R702. This strategy allows the transfer of the altered R factor between *E. coli* strains and avoids the potential problems associated with a donor strain harbouring three plasmid types – R702, the plasmid of interest and the methylation plasmid. Lambda red recombineering was utilised to insert the methyltransferases into the R factor, causing an insertional disruption of a spectinomycin resistance gene it contains, a useful downstream phenotypic screen in the assembly process. The overall

strategy is outlined in Figure 3.4. Firstly the strategy involved creation of a plasmid (pCW10) containing P_{BAD}-araC upstream of CLAU_0514 and CLAU_1608 with their native ribosome binding sites (RBS). The promoter-MTase cassette was then amplified from pCW10 and cloned into a second plasmid pCW24, where it was flanked by homology arms designed to insert the MTase cassette into R702 disrupting the spectinomycin resistance gene. Both pCW10 and pCW24 were created using HiFi Assembly and pCW24 was used as a PCR template for the production of the linear DNA used in lambda red recombineering to create the modified R702.

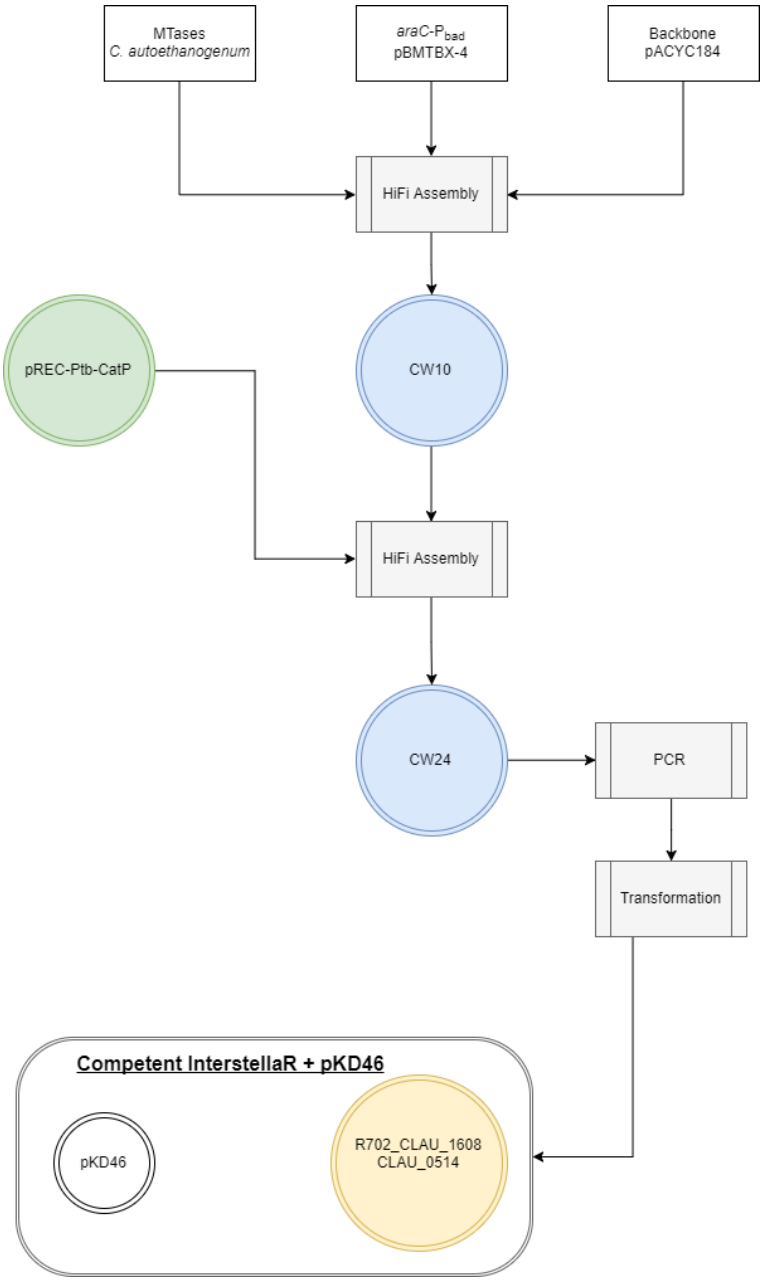


Figure 3.4: Recombineering *C. autoethanogenum* MTases into R702.

The overall strategy for modifying R702 to express two *C. autoethanogenum* MTases encoded by CLAU_1608 and CLAU_0514. MTase sequenced were amplified from *C. autoethanogenum* wild type genomic DNA and combined with the *araC*-P_{bad} promoter system found on pBMTBX-4 and the backbone pACYC184 to create pCW10. pCW10 was then used as template in the creation of a second plasmid, pCW24, which added the apramycin resistance marker and two homology arms for insertion into R702 flanking the promoter MTases cassette. pCW24 could then be used as template to create the final PCR product used to transform recombineering-ready electrocompetent InterstellaR cells which harboured the recombineering plasmid pKD46. Transformants were selected for on LB agar supplemented with apramycin and kanamycin.

3.2.3.5 Creation of pCW10

The pCW10 plasmid was created based on pACYC184 and so harbouring the p15a origin of replication, making it compatible for co-harbouring target plasmids with the colE1 origin of replication frequently used in the pMTL80000 modular shuttle plasmids. pCW10 was generated using HiFi Assembly from NEB. pACYC184 was linearised with a HindIII restriction digest and served as the vector backbone. CLAU_1608 and CLAU_0514 were amplified from *C. autoethanogenum* wild type genomic DNA (gDNA) and cloned alongside pACYC184 and P_{BAD}-*araC* fragments using hi-fi assembly to create the final vector pCW10. pCW10 places both methyltransferases along with their respective native RBS downstream of the arabinose-inducible P_{BAD} promoter. DNA fragments were generated using primers and templates as indicated in Table 3.6 and combined as per the manufacturer's instructions. Transformation of Top10 *E. coli* followed HiFi Assembly with transformants selected for with chloramphenicol.

Table 3.6: Primers and templates used to construct pCW10.

pCW10 was constructed using HiFi assembly. This table presents the primers and relevant templates used to generate the DNA fragments with overlapping regions allowing the HiFi reaction to form a completed molecule.

Primer pair	Template	Product
82_pBAD-araC_fwd/83_pBAD-araC_rev	pBMTBX-4	araC-P_{BAD} promoter
84_CLAU_0514_fwd/85_CLAU_0514_rev	*C. autoethanogenum* genomic DNA	CLAU_0514
86_CLAU_1608_fwd/87_CLAU_1608_rev	*C. autoethanogenum* genomic DNA	CLAU_1608

E. coli strains harbouring the pCW10 methylation plasmid and a plasmid of interest can be used to prepare plasmid DNA for transformation using arabinose

to induce expression of the two methyltransferases. This approach was not attempted in this study, instead pCW10 was used as a template to generate fragments for a second plasmid pCW24 which inserts the promoter-methyltransferase cassette between homology arms for recombineering into R702.

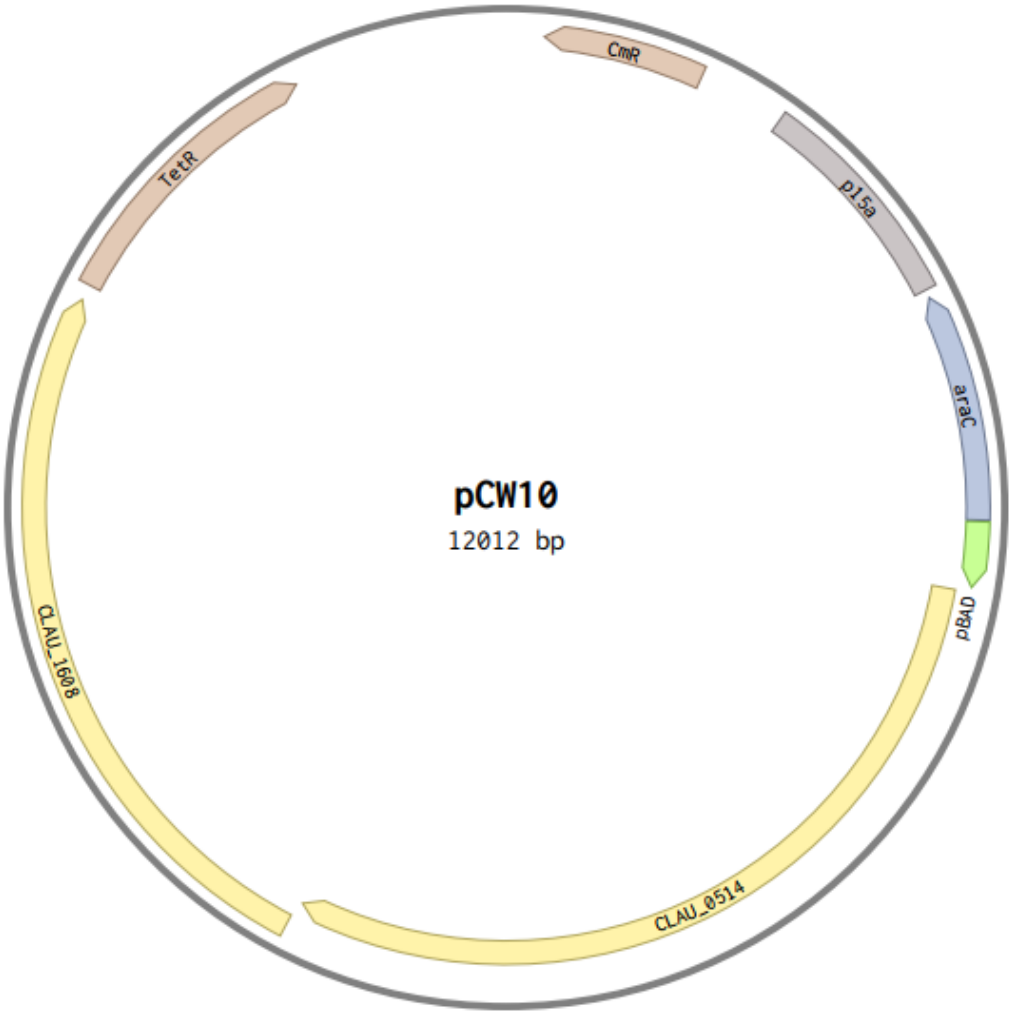


Figure 3.5: pCW10

Plasmid map of pCW10 which is the first plasmid in the strategy to express MTases from R702, but can also be used as a co-expression plasmid with a target plasmid reliant on a *ColEI* origin of replication since pCW10 is based on the compatible p15a origin of replication.

3.2.3.6 Creation of pCW24

In order to clone the promoter-methyltransferases cassette onto R702 a plasmid was created which inserts the cassette between two homology arms directed against R702 as well as an apramycin resistance marker. This plasmid termed pCW24 was then used as template in a PCR to generate the linear fragment for recombineering. pCW24 was created again using HiFi Assembly with DNA fragments generated as indicated in Table 3.7.

Table 3.7: Primers and templates used to construct pCW24.

pCW24 was constructed using HiFi assembly. This table presents the primers and relevant templates used to generate the DNA fragments with overlapping regions allowing the HiFi reaction to form a completed molecule.

Primer pair	Template	Product
194-cw24-backbone-reverse/201-cw24-backbone-forward	pREC-ptb-catP	Vector backbone.
195-cw24-LHA-fwd/196-cw24-LHA-rev	pREC-ptb-catP	Left homology arm.
197-cw24-meth-fwd/198-cw24-meth-rev	pCW10	Promoter and MTases of pCW10.
199-cw24-RHA-fwd/200-cw24-RHA-rev	pREC-ptb-catP	Right homology arm.

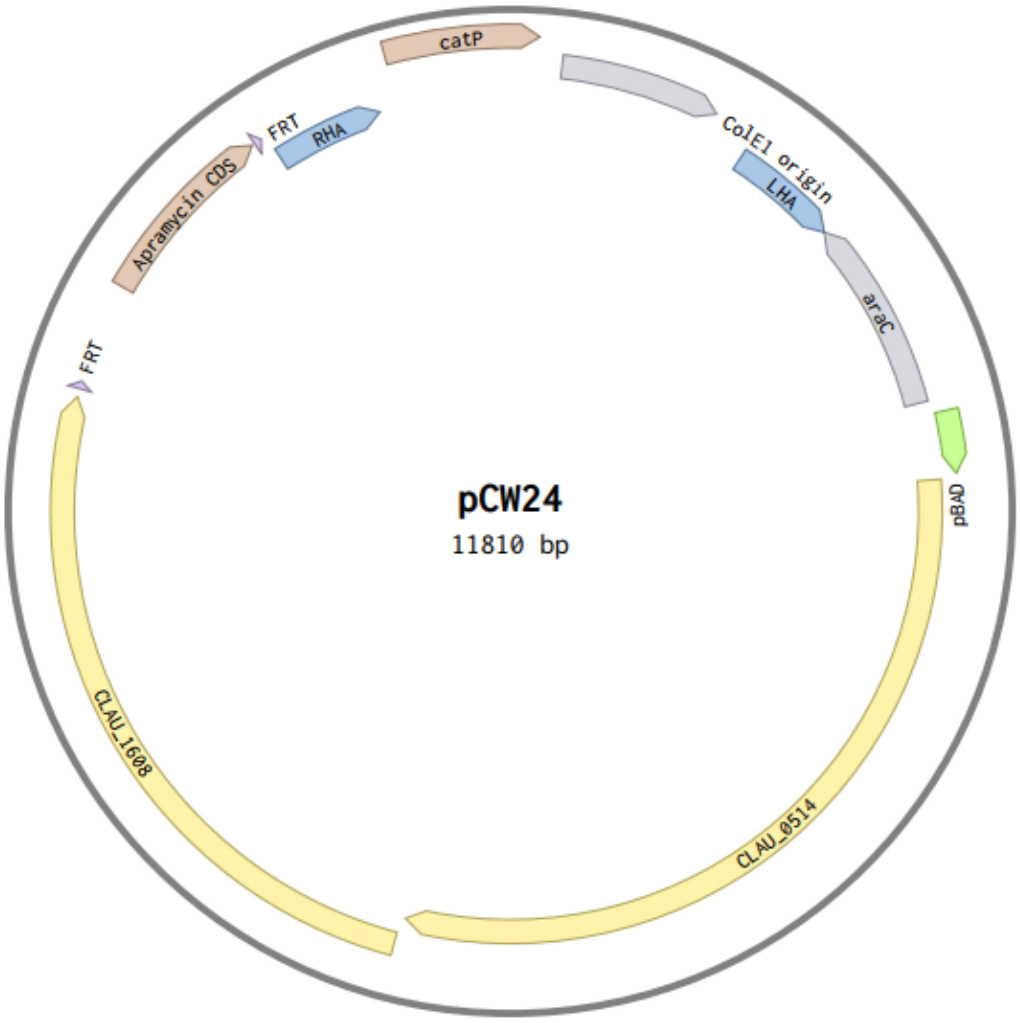


Figure 3.6: pCW24

Plasmid map of pCW24 which was used as a template for the production of the eventual PCR product allowing modification of R702. Two *C. autoethanogenum* MTases are controlled by the arabinose-inducible promoter P_{bad}.

The pCW24 plasmid then served as the template for a PCR to generate the single stranded product for recombineering the methyltransferases into R702 as described below.

3.2.3.7 Lambda Red Recombineering to Create InterstellaR_CLAU_1608+ CLAU_0514.

InterstellaR is an *E. coli* conjugal donor strain analogous to the previously described Sexpress. InterstellaR was created by James Millard in this laboratory (unpublished) following the protocol established for the creation of Sexpress. InterstellaR was based on a *dam- dcm-* commercial strain sold by Promega named Stellar (note that multiple versions of Stellar exist with different genotypes, but only one with *dam- dcm-*). InterstellaR was transformed with the rKD46 lambda red recombineering plasmid and competent cells prepared as described in chapter 2, inducing the recombinase with arabinose. Competent InterstellaR + rKD46 was then transformed with a PCR product generated from pCW24 using primers 266_RecoV2-LHM-Fw and 267_RecoV2-RHM-Rv. This PCR product contained the promoter and methyltransferases region with an apramycin resistance marker all flanked by homology arms directed against the spectinomycin gene of R702. Transformants were recovered at 37°C to force loss of the temperature-sensitive pKD46 plasmid and all subsequent growth stages performed at 37°C. Transformants were selected on LB agar plates supplemented with apramycin. Transformants were inoculated into apramycin and kanamycin to select for the desired genotype; Spectinomycin to verify that the spectinomycin gene function had been lost by insertion of the methylation cassette; Chloramphenicol to verify that the pKD46 plasmid had been lost.

For all six transformants selected growth was observed in the first culture supplemented with apramycin and kanamycin only, suggesting that the apramycin resistance gene had inserted into the spectinomycin gene of R702 and that the kanamycin resistance originating elsewhere on R702 remained

intact. The failure to grow in chloramphenicol suggested that the chloramphenicol resistant temperature-sensitive pKD46 recombineering plasmid had been successfully lost. A PCR using an internal – external primer pair was performed to verify the insertion of the desired methylation cassette into R702. The resulting bands were gel-purified and sequenced revealing successful insertion as predicted.

3.2.3.8 Assessment of the Novel Strain InterstellaR R702+MTases for Conjugal Transfer

The newly created version of InterstellaR containing two methyltransferases of *C. autoethanogenum* under control of P_{BAD} was evaluated as a conjugal donor strain in comparison to InterstellaR and Sexpress. pMTL83151 was used as a test plasmid harbouring the pCB102 origin of replication and the chloramphenicol resistance marker *catP*. Conjugations were performed in triplicate for each strain following the protocol outlined in the materials and methods chapter, section 2.4.7. The donor cultures were treated identically except for the addition of 10 mM L-arabinose to the “induced” cultures. The results presented in Figure 3.7 show that InterstellaR proved to be a far inferior donor strain to Sexpress, but the induction of the two *C. autoethanogenum* MTases appeared to revive some ability to transfer plasmid DNA.

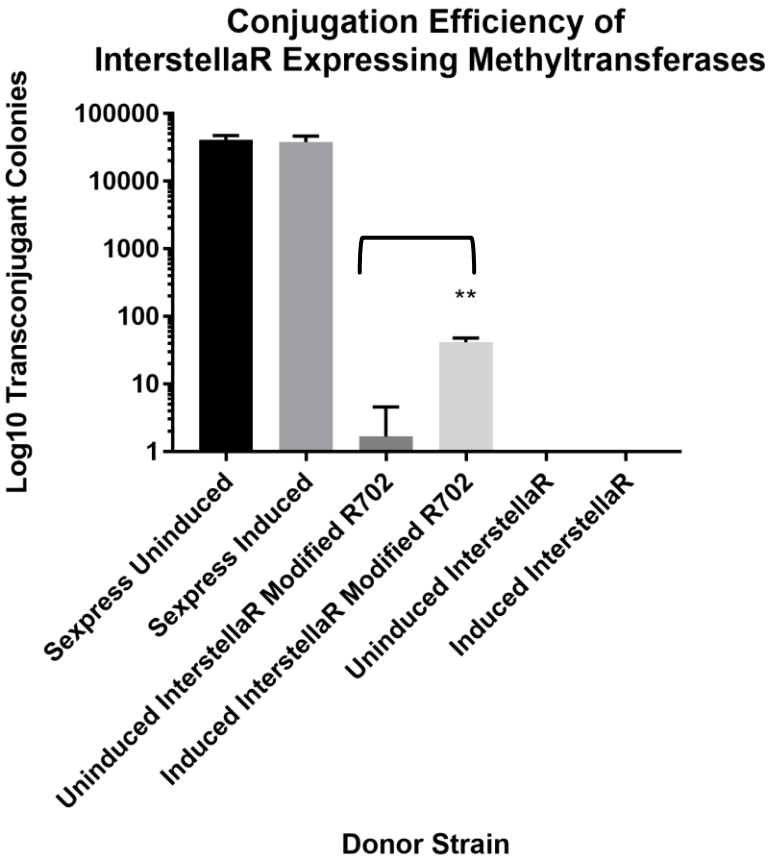


Figure 3.7: Sexpress, InterstellaR and InterstellaR+MTases conjugative efficiency comparison. Data represents conjugations performed in triplicate following the protocol outlined in section 2.4.7. Induced and uninduced cultures were treated identically except for the addition of 10mM arabinose to the subculture of 'induced' cultures. A significant difference ** (p value=0.0039) was found between the number of transconjugants between induced and uninduced modified InterstellaR cultures.

3.2.3.9 Transfer of the Modified R702 into Sexpress

The modified R702 expressing the *C. autoethanogenum* MTases CLAU_0514 and CLAU_1608 was transferred from InterstellaR to NEB Express via conjugal transfer. The donor strain InterstellaR was selected against by plating on a medium lacking proline whilst selecting for the presence of the modified R702 with apramycin. Transconjugant colonies represented NEB Express with the modified R702.

3.2.3.10 Assessment of the Novel Strain Sexpress R702_CLAU_0514_CLAU_1608

pMTL82151 was used to test the conjugal abilities of the novel donor strain. It contains one SNNGCAAT site, six CAAAAAR sites and one GTTAAT site. Expression of CLAU_1608 was thought to offer protection against restriction at SNNGCAAT sites while CLAU_0514 should protect GTTAAT sites. The CAAAAAR sites were not considered at this point since there was no obvious cognate restriction enzyme associated with the methyltransferase CLAU_1258. Sexpress R702_CLAU_0514_CLAU_1608 produced a higher, but not statistically significantly higher, number of transconjugants (Figure 3.8).

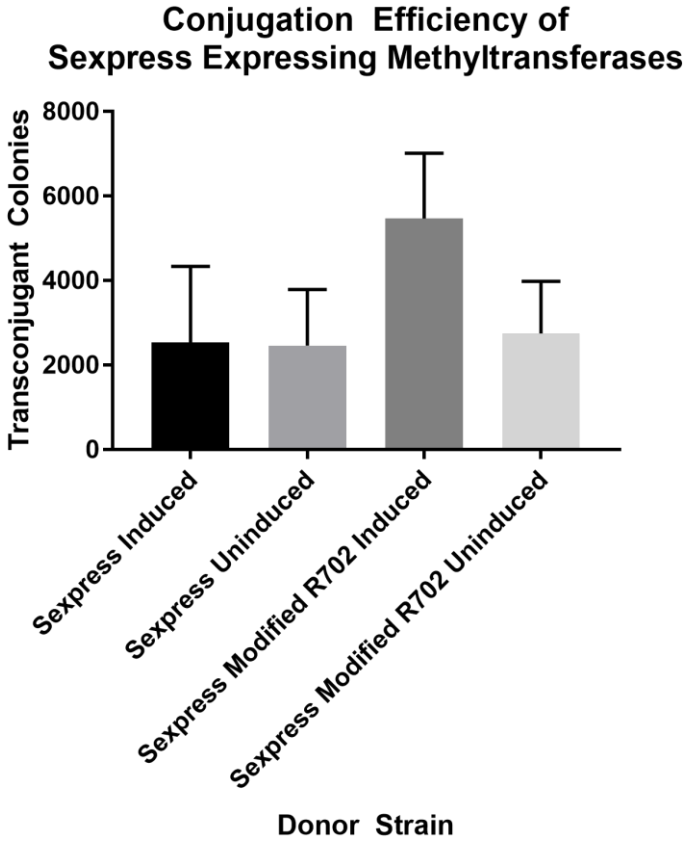


Figure 3.8: Sexpress and Sexpress+MTases conjugation efficiency comparison.

Data represents conjugations performed in triplicate following the protocol outlined in section 2.4.7. Induced and uninduced cultures were treated identically except for the addition of 10 mM arabinose to the

morning subculture of ‘induced’ cultures. No significant difference was measured between induced and uninduced cultures.

3.3 Discussion

Genetic modification of industrially relevant organisms is highly advantageous in order to allow improvement of the native properties of the organism, introducing synthetic pathways to extend the natural product range or imbuing the organism with heightened tolerance to toxic compounds likely to be found in an industrial context. Moreover, genetic manipulations allow the study of the underlying biology of the traits which make the strain useful for example the ability to use CO/CO_2+H_2 as a sole carbon and energy source. Exogenous DNA is generally introduced to bacterial cells via transformation, conjugation or transduction. Gram-positive bacteria tend to be more resistant to transformation efforts due to the barrier presented by the thick peptidoglycan layer and protein S-layer as well as the presence of only a single cytoplasmic membrane which leaves them more susceptible to homeostatic shock. Members of the clostridia are often particularly challenging to establish a robust DNA transfer protocol for (Pyne et al., 2014). It is common for new strains to require lengthy studies to optimise and establish DNA transfer and the range of techniques used seems more diverse than the genetic differences would warrant. The most common methods of transforming clostridia are electroporation and conjugation, though recently sonoporation was demonstrate to be superior to either conjugation or electroporation in *C. pasteurianum* (Kolek, Sedlar, Provaznik, & Patakova, 2016).

3.3.1 Importance of Donor *dcm* Methylation Status

The importance of the methylation status of the DNA being introduced has been demonstrated in this study by the superior performance of Sexpress (*dam+, dcm-*) over CA434 (*dam+, dcm+*) in *C. autoethanogenum* and further supported by work in *C. difficile* DH1916 by P. Ingle.

3.3.1.1 Analysis of Sexpress for *C. autoethanogenum*

The results presented in Table 3.1 show that Sexpress outperformed CA434 with every replicon tested, with the exception of cases whereby both strains failed to produce any transconjugants, as with pCACET. The most remarkable increase in efficiency was observed with the pIP404 replicon increasing the average colonies per conjugation from 3 with CA434 to 2,419 with Sexpress. For pCB101 the only transconjugants observed were when using Sexpress while CA434 failed entirely. Most relevant to transposon mutagenesis in this study is the 9-fold increase observed using pCB102 since this replicon is used in the established transposon mutagenesis system intended to be implemented. The increased efficiency observed for pCD6 is crucial also, since this replicon is used in ACE vectors necessary for strain engineering, including the production of a $TcdR^+$ strain which is required in order to use the established pMTL-YZ14 transposon delivery vector. The poor performance of pCD6 had previously been a persistent problem in mutant generation, and when compared to transformation rates with other replicons the reasons for this became evident, with an average colonies per conjugation of only 0.3 when using CA434. Although the average number of colonies per conjugation with Sexpress was still a modest 4.6, this represents a significant 15-fold increase over CA434.

3.3.1.2 Analysis of Sexpress for *C. difficile*

Sexpress achieved higher conjugation efficiencies using all origins of replication tested as shown in Table 3.3. Sexpress was used in all future conjugations into DH1916. Ideally the possibility of other genotypic differences between strains would be ruled out by creating a *dcm* knockout in a CA434 background. NEB Express is a B-strain derivative and as such exhibits non-negligible differences with the K-12 derived CA434 (HB101 strain harbouring R702).

3.3.1.3 Importance of *dcm* Status in Other Species

Reductions in DNA transfer rates with methylated DNA have been documented in various bacterial species such as *Bacillus anthracis* (Sitaraman & Leppla, 2012), *Streptomyces* (González-Cerón, Miranda-Olivares, & Servín-González, 2009; Kieser & Hopwood, 1991) and *Lactobacillus* (Spath, Heinl, & Grabherr, 2012). More recently, several studies on clostridia have come to similar conclusions concerning methylation status including *C. thermocellum* DSM1313 (Guss, Olson, Caiazza, & Lynd, 2012b) *C. ljungdahlii* DSM13528 (Leang, Ueki, Nevin, & Lovley, 2013a) and *C. pasteurianum NRRL B-598* (Kolek et al., 2016).

C. pasteurianum

C. pasteurianum NRRL B-598 is a solventogenic non-type strain for which a DNA transfer protocol was recently established (Kolek et al., 2016). After the failure of initial attempts varying conjugation/electroporation parameters the R-M systems were examined via Single Molecule Real Time (SMRT) sequencing which reveals the methylation patterns present in the genome. Two methylated motifs were detected, both of which exhibited 6-mA methylation. Two Type I

Candidate R-M systems were identified for these patterns via homology searching. In addition, two candidate Type II systems were identified which were not predicted to produce any methylation pattern. However, SMRT sequencing is more proficient at detecting 6-mA and 4-mC type methylation which produce stronger kinetic signals than 5-mC methylation. To detect 5-mC a strand coverage of 250x is recommended in a technical note of Pacific Biosciences while this study only provided 79x strand coverage meaning 5-mC methylation motifs could have been missed. The SMRT sequencing of *C. autoethanogenum* achieved a strand coverage of 179x (Utturkar et al., 2015) so this could also be true for *C. autoethanogenum*. In the *C. pasteurianum* paper the native restriction systems were not manipulated, instead changing the *dam/dcm* methylation patterns of input DNA was found to allow reliable transformation (Kolek et al., 2016). Plasmid DNA was prepared from *E. coli* strains with varying *dam*/*dcm* status. Interestingly the best results were obtained when DNA with neither *dam* nor *dcm* methylation was used, yielding 118 colonies per µg DNA compared to 14 colonies per µg DNA for *dam*+ *dcm*- and 13.5 colonies per µg DNA for *dam*- *dcm*+. Lack of either *dam* or *dcm* methylation therefore resulted in roughly equal efficiency increases over *dam*+ *dcm*+. It was suggested that since the test plasmid used (pMTL83253) contained more *dcm* sites (18) than *dam* sites (10) and the efficiency was the same when removing either methylation pattern that the *dam* methylation is more problematic to DNA transfer. This is at odds with the findings for *C. autoethanogenum* where the *dam*- *dcm*- strain (InterstellaR) performed far more poorly than the *dam*+ *dcm*- strain (Sexpress), though other genetic differences between these two strains could be relevant. *C. pasteurianum*

NRRL B-598 has one annotated Type IV restriction system. One explanation for the results obtained is, therefore, that this restriction system is responsible for attacking incorrect methylation patterns both at *dam* and *dcm* sites. However, when exposing plasmid DNA with various *dam/dcm* methylation status to cell lysates no difference in band pattern was observed. An active Type IV restriction enzyme might be expected to attack *dam*$^{+}$ *dcm*+ DNA with greater efficacy than unmethylated DNA, producing a smear when visualised on an agarose gel, but no such observation was found. It is possible that the Type IV enzyme simply does not function effectively in these *in vitro* conditions. Another possibility is that the Type IV system is not in fact important to DNA transfer and instead the methylation status of the origin of replication is crucial to successful replication, this will be discussed further below in section 3.3.1.7.

C. thermocellum

C. thermocellum is another organism for which the *dam*/*dcm* methylation status of transmitted DNA has been reported as significant. In 2004 research (M V Tyurin, Desai, & Lynd, 2004) stated that a *dam*+ *dcm*+ strain (DH5 alpha) outperforms a *dam*- *dcm*- (GM272) strain and suggested the likely reason was a partial protection from the "*MboI*-like restriction system" identified in (Klapatch, Demain, & Lynd, 1996) which recognises the *dam* sequence GATC. The efficiency increase observed when using *dam*+ *dcm*+ DNA over *dam*- *dcm*- was thought to be surprisingly low (only 7-fold for strain ATCC 27405 and 40-fold in DSM 1313).

“The relatively modest effect of dam methylation observed here for C. thermocellum is not the result of a low number of restriction sites, since there are 23 MboI sites in pIKm1, and awaits explanation” (Tyurin et al., 2004)

An explanation was provided in 2012 when it was demonstrated that *dcm* -methylation is detrimental to transformation (Guss et al., 2012b). This explains the “*relatively modest*” effect of *dam* methylation observed in 2004 since the *dam*- strain used (GM272) was also *dcm*- while the *dam*+ strain was also *dcm*+ so in effect the benefit of the strain being *dam*+ was masked by the negative impact of it also being *dcm*+. The benefit of removing *dcm* methylation in *C. thermocellum* was confirmed in this later work 1000-fold increases in transformation efficiency observed using *dam*+ *dcm*- over *dam*+ *dcm*+ DNA (Guss, Olson, Caiazza, & Lynd, 2012a).

C. ljungdahlii

The methylation status of transmitted DNA was reported as significant in *C. ljungdahlii* (Leang et al., 2013a). In 2010 a DNA transfer protocol was published for *C. ljungdahlii* involving *in vivo* methylation using the strain ER2275. ER2275 is an *E. coli* strain formerly sold by NEB which harbours a p15a-based methylation plasmid expressing the *Bacillus subtilis* phage phi3T I methyltransferase, this strain has been shown to circumvent the Cac824 restriction endonuclease of *C. acetobutylicum* (Mermelstein & Papoutsakis, 1993) accordingly it was hoped that the methyltransferase would offer protection against the Type I restriction endonuclease of *C. ljungdahlii* (Köpke et al., 2010). However, a later publication in 2013 reported that *in vivo* methylation reduced transformation efficiency (Leang, Ueki, Nevin, & Lovley,

2013b). The later publication also stated that DNA isolated from *dam*+ *dcm*- strains transformed at higher efficiency than when from *dam*+ *dcm*+, but no data was presented to this effect. This observation is in line with the findings made in the current work, and suggests that methylation status is important whether DNA is transferred via electroporation or conjugation. It is noteworthy that *C. ljungdahlii* also is more transformable with *dcm*- DNA since it has only one predicted Type IV ORF compared to the five of *C. autoethanogenum* according to the REBASE information for *C. ljungdahlii,* however, it also has a homolog of a predicted Type IV enzyme of *C. autoethanogenum* (Cau10061ORFEP) so it may possess two Type IV systems. The Type IV system of *C. ljungdahlii* has a homolog in *C. autoethanogenum* which is not a predicted Type IV enzyme in the REBASE genome entry (reference # 21199) but is in the two older entries (reference # 16692 and # 16407).

3.3.1.4 Hemimethylation

The importance of *dam*/*dcm* methylation may not be due to the effect of Type IV restriction systems. In *Bacillus anthracis* it was found that knocking out all the putative Type IV systems improved transformation frequency with *dam*+ *dcm*+ DNA and yet *dcm*- *dam*- DNA still transformed with greater efficiency (Sitaraman & Leppla, 2012). With no more obvious Type IV targets it was thought more likely that there must be an alternative mechanism by which methylated DNA is less able to transform *B. anthracis*. This phenomenon is also documented in both *dam*- strains *E. coli* (Russell & Zinder, 1987) and *Salmonella enterica* (Aloui, Tagourti, El May, Joseleau Petit, & Landoulsi, 2011) which are easily transformed by unmethylated DNA and suffer efficiency losses with *dam* methylated DNA. Plasmids with *dam*+ *dcm*+ methylation entering *C.*

autoethanogenum will form hemimethylated products after their first replication. Hemimethylated DNA has been shown to be a worse substrate for DNA replication in *dam- E. coli* strains where hemimethylated plasmids accumulate (Russell & Zinder, 1987). Perhaps the efficiency gains by avoiding *dcm* methylation are not entirely due to avoidance of Type IV systems but due to a more suitable substrate for further replication cycles after the first replication event. In addition perhaps a *dam*- donor strain could confer benefit to strains even which lack any adenine-directed Type IV systems.

3.3.2 Expression of MTases in Sexpress

Sexpress modified to express CLAU_0514 and CLAU_1608 was hypothesised to modify target plasmid DNA in the conjugal donor strain and so confer protection against *C. autoethanogenum* restriction systems. The results of the conjugation efficiency experiment show that there is no significant difference in the number of transconjugant colonies obtained when using the modified strain and no significant difference between induced and uninduced conditions of the modified strain Figure 3.8. There are several likely explanations for this. The methylation motif conferred by the cloned methyltransferases may give no additional protection against restriction systems of *C. autoethanogenum*. Other methyltransferases discussed previously such as CLAU_2858 and CLAU_1258 may be better suited to this effort. Conjugation may be an inappropriate means of DNA delivery to test this phenomenon if restriction enzymes fail to act on the single stranded intermediate involved in conjugal transfer. Transformation of plasmids prepared through the modified methyltransferase expressing strains should be attempted. Previous work in *C. autoethanogenum* demonstrated an increase in transformation rates with DNA prepared from *C. autoethanogenum*

(Liew, 2015). It may be that the methyltransferases failed to act in *E. coli* and strategies to optimise gene expression such as changing the RBS are necessary. SMRT sequencing of the novel methyltransferase expressing strains would reveal if modification of target plasmid DNA is occurring successfully.

3.3.3 Expression of MTases from InterstellaR

InterstellaR contains the R factor (R702) containing the transfer function responsible for the capabilities of the well documented strain CA434. The transfer of R702 to NEB Express conferred this strain with an ability to transfer DNA via conjugation superior to that of CA434. CA434 is *dam*+ *dcm*+, Sexpress is *dam*+ *dcm*-, Stellar is *dam*- *dcm*- and was used to create the novel donor strain InterstellaR in a process analogous to that used to create Sexpress. InterstellaR was assessed as a conjugal donor and found to produce only one transconjugant over six separate conjugation attempts whilst Sexpress in parallel conjugations produced an average of 4.0×10^4 (Figure 3.7). Future work on this topic should aim to determine if this difference in abilities is due to the *dam*- genotype of the strain or an alternative unmentioned genotype. If *dam*- DNA is a poor substrate for conjugal transfer then there are several possible explanations. In other clostridia *dam*+ DNA has been shown to be essential for DNA transfer (Chen, Boucle, & Blaschek, 1996), in the case of *C. thermocellum* this is due to the protection it confers from a native Type II restriction system which recognises the GATC motif (Michael V Tyurin, Desai, & Lynd, 2004). If this were the case in *C. autoethanogenum* we would expect the methylation data obtained from SMRT sequencing to highlight a motif that overlaps the GATC motif. This is not the case with any of the three motifs identified. If *C.*

autoethanogenum possesses a restriction enzyme which recognises the motif GATC it must also methylate that same sequence in its own DNA. To argue for the protective benefits of *dam* methylation therefore requires appealing to the uncertainties of a) the SMRT sequencing results and the possibility of an undiscovered motif or b) more generic protective effects of *dam* methylation such as restricting access to the DNA. As previously discussed, the strand coverage of the SMRT sequencing performed on *C. autoethanogenum* is below that which is advised for the detection of 5-mc residues, but 6ma methylation of adenines only requires a strand coverage of 25x meaning motifs are unlikely to have been missed by the sequencing. A more unlikely possibility is that a methylation motif was missed because the MTase responsible was inactivate at the time of DNA extraction for sequencing. An alternative explanation for the positive effect of *dam* methylation on conjugal transfer into *C. autoethanogenum* is that methylated DNA serves as a better substrate for DNA replication. Replication proteins of *C. autoethanogenum* may act poorly upon unmethylated DNA. This explanation is more tempting in light of the recovery of conjugal function of InterstellaR with a modified R702 which expresses MTases. InterstellaR has been demonstrated to be a suitable donor strain in *C. botulinum* strains 3502 and 19397 (D. Groothuis, unpublished) so it seems that the donor strain does not require *dam* methylation in order to recognise and transfer plasmid.

InterstellaR with modified R702 containing P_{BAD}-CLAU_0514+CLAU_1608 was able to conjugate to *C. autoethanogenum* at significantly (P=0.0039 **) different frequencies depending on whether the P_{BAD} promoter was induced or uninduced. Transcriptional control of the MTases is therefore possible. The

advantage of using an inducible promoter is to avoid deleterious effects to the donor strain before induction, however, no difference in growth rates was observed between induced and uninduced cultures (data not shown). The pertinent question now is by what mechanism does expression of the *C. autoethanogenum* MTases act to improve conjugation frequencies? Initially the intention was to offer protection against *C. autoethanogenum* restriction systems my mimicking the native pattern, but in Sexpress the modified R702 offered no statistically significant increase in conjugation efficiency implying that in this context they do not provide protection. It may be that more replicates are required to detect a significant effect and the reason we see a significant effect with InterstellaR and not Sexpress is merely because InterstellaR had incredibly low numbers of transconjugants, so any increase is likely to be detected, whilst Sexpress already produces $>1x10^4$ transconjugant colonies per conjugation making moderate effects more difficult to observe.

3.3.4 Further Work

To determine the relevance of *E. coli* native MTases a *dam*- *dcm*+ strain harbouring R702 should be constructed and tested for ability to conjugate into *C. autoethanogenum*. This strain would be expected to be a poor conjugal donor from the work in this study which demonstrated that *dcm* methylation negatively affects DNA transfer, but lacking both *dam* and *dcm* methylation almost completely removes the donor strain ability to conjugate. To disregard the possibility of other genetic differences between the donor strains employed in this study, single knockout versions of CA434 for *dam* and *dcm* could be created as well as the double knockout.

The induction of *C. autoethanogenum* methyltransferases in InterstellaR proves beneficial implying their activity could increase conjugation efficiencies in other genetic backgrounds including that of Sexpress, but this result was not replicated in this strain. More replicates giving more degrees of freedom may be necessary to detect an increase in efficiency for Sexpress. Only two MTases from *C. autoethanogenum* were cloned in this study, therefore further work will involve expression of the remaining MTases. Of particular interest would be expression of CLAU_1258 which is predicted to be responsible for the methylation motif CAAAAAR. In addition each of the four MTases could be expressed separately to determine their individual effects and discover any synergistic interactions.

Conjugation was used to assess the importance of methylation status in this chapter. Several efforts were made to transform *C. autoethanogenum* via electroporation, but the resultant colonies were too few in number for this technique to be considered appropriate for studying DNA transfer. These attempts were prior to the demonstration of the relevance of *dcm* methylation and so should be repeated using DNA prepared from a *dcm*- strain.

Beyond studying the effect of methylation status there are a huge number of variables in the conjugation or electroporation protocols which could prove to be relevant as well as considering the sonoporation technique which proved so beneficial to *C. pasteurianum* (Kolek et al., 2016). Seemingly minor alterations to protocols can have effects spanning orders of magnitude, for example in *Geobacillus thermodenitrificans* T12 electroporation efficiency was increased from 3-4 colonies per µg DNA to 1.7 × 104 by changing the concentration of

K_2HPO_4 in the growth medium (Daas, van de Weijer, de Vos, van der Oost, & van Kranenburg, 2016).

3.4 Conclusion

The gains in efficiency offered by use of the novel Sexpress strain ameliorated efforts to perform transposon mutagenesis in *C. autoethanogenum* using a replicative transposon-delivery vector. Further gains in the efficiency of DNA transfer may allow the use of a non-replicative transposon-delivery mechanism which would greatly enhance the potential for Transposon insertion sequencing experiments as well as other genetic studies in this organism. The importance of *dam* and *dcm* methylation in the *E. coli* donor has been demonstrated. The expression of native MTases significantly improved transfer from InterstellaR, but not from Sexpress. This partial success warrants further investigation using a similar strategy.

3.5 Results Summary

The main outcomes from the results section of the chapter can be summarised as follows:

- A novel conjugal donor strain Sexpress was created by conjugation of R702 from CA434 to NEB Express.
- Sexpress (*dam*+ *dcm*- R702) is a superior conjugal donor strain to CA434 (*dam*+ *dcm*+ R702) for *C. autoethanogenum* and *C. difficile* DH1916.
- InterstellaR (*dam*- *dcm*- R702) has a severely inferior ability to produce transconjugants of *C. autoethanogenum* compared to Sexpress.

- A modified R702 was created which expresses two MTases from *C. autoethanogenum* CLAU_0514+CLAU_1608 from the inducible P_{BAD} promoter.
- InterstellaR with modified R702 (CLAU_0514+CLAU_1608) exhibited a significant increase in conjugation efficiency when the MTases were induced over the uninduced state.
- The modified R702 (CLAU_0514+CLAU_1608) was transferred from InterstellaR to Sexpress via conjugal transfer.
- Sexpress harbouring the modified R702 failed to produce a statistically significant difference in conjugation efficiency either between induced/uninduced states or when compared to the unmodified R702.

4 Implementation of Transposon Mutagenesis in *C. autoethanogenum* and *C. ljungdahlii*

4.1 Introduction

4.1.1 Transposon Mutagenesis in *Clostridium*

Forward genetics using random transposon mutagenesis is a valuable tool in molecular microbiology. A basic application of transposon mutagenesis involves generating a pool of random mutants, selection for a desired phenotype followed by determining the genetic basis of the trait. A forward genetics approach such as this therefore has the ability to discover novel genes implicated in a relevant phenotype. Recent success in related species using a *mariner* system has been reported in *C. difficile* (Cartman & Minton, 2010), *C. perfringens* (Liu et al., 2013), *C. acetobutylicum* and *C. sporogenes* (Zhang et al., 2015). Earlier efforts were hampered by use of transposable elements which gave rise to multiple insertions per genome and non-random insertions (Bertram & Dürre, 1989) (Transfer of Tn*1545* and Tn*916* to *C. acetobutylicum*). Recently, the first Transposon Insertion Sequencing (TIS) experiment in clostridia was performed in the two *C. difficile* strains 630Δ*erm* and R20291 allowing identification of a core set of 404 essential genes and 798 genes implicated in spore production (Dembek et al., 2015).

4.1.2 Transposon Delivery

Transposon mutagenesis studies ideally deliver the transposon via a transposon-transposase complex (transposome) or via a non-replicating suicide plasmid. These approaches have the advantage that any colony

appearing on selection plates (except in the case of plasmid integration) must be the result of a transposition event, therefore all colonies can be carried forward to a phenotypic selection with reliable estimates of the number of unique mutants and without any need for further screening such as colony PCR to confirm their identity as transposon mutants. In addition, the presence of transposon-harbouring plasmids during high-throughput insertion site sequencing can severely reduce the numbers of genuine transposon insertion sites identified, conferring further advantage to the use of non-replicative delivery systems. Unfortunately, suicide-delivery is only a feasible option in organisms which are highly amenable to transformation since the number of mutants obtained is a function of the transformability and the transposition rate. For instance, in *E. coli* K12 MC1061 the plasmid electroporation efficiency was reported as 1.8×10^9 CFU/µg DNA compared to only 1.1×10^8 CFU/µg DNA for a Tn*5* transposome complex (6.1% of the electroporation efficiency) (Goryshin, Jendrisak, Hoffman, Meis, & Reznikoff, 2000). An attempt to optimise transposome delivery into Gram-positive organisms was undertaken using mini-Mu transposomes reporting a transposition efficiency of 0.05-0.5% (Pajunen, Pulliainen, Finne, & Savilahti, 2005). Typically the efficiency of transposome integration is around 3-orders of magnitude lower than the electroporation efficiency (Lamberg, Nieminen, Qiao, & Savilahti, 2002). *C. autoethanogenum* has been transformed using a plasmid harbouring the pCB102 replicon at a frequency of 0.2 cfu/µg DNA meaning around 200µg of transposome DNA would be required to produce one transposon mutant using these estimations.

4.1.3 Conditional Replicon

Low transformability of *C. autoethanogenum* requires the use of a replicative plasmid harbouring both transposase and transposon. This allows a population of cells all of which harbour the transposon plasmid to grow, each of which can undergo transposition events. The difficulty of using a replicating plasmid is that there is often no way to discern between plasmid-harbouring cells and transposon mutants on selection plates since the *catP* marker is active in either case. Use of a pseudo-suicide vector in which plasmid replication is slower than chromosomal replication can increase the likelihood of colonies representing transposon mutants. Preferable to this is a conditional replicon which allows selection of the transposon plasmid, growth of a large population of cells undergoing transposition reactions followed by plating to non-permissive conditions selecting for transposon mutants.

A replicative plasmid based on the pCB102 replicon native to *C. butyricum* (Minton & Morris, 1981) was chosen as the means of transposon-delivery. This replicon is rendered unstable by transcriptional read-through from the IPTG-inducible P_{fac} promoter which is located 5' to pCB102. The P_{fac}-pCB102 system forms a conditional replicon with the non-permissive condition being the presence of IPTG which induces P_{fac} (Zhang et al.,2015). Thiamphenicol resistance arises due to the activity of the transposon-located *catP* gene whether on the plasmid or integrated in the genome so a conditional replicon is necessary to discern true transposition events.

4.1.4 Transcriptional Control of Transposase

The promoter controlling the transposase is important and should fit two criteria. Firstly, it must be sufficiently strong as to produce a high transposition frequency

to generate a large transposon mutant pool. Secondly the control of transposition should be controlled at the level of transcription by means of an inducible or orthogonal promoter active only in the target strain. Control of the transposase is desirable to avoid unwanted transposition events in the *E. coli* storage and donor strains. Inducible transposase control may also provide benefits in terms of mutant library complexity in organisms where replicative plasmid transposon delivery is required. If transposition events can be prevented until the inducible promoter is activated a large population of plasmid-harbouring cells can grow each of which has the potential to undergo unique transposition events.

4.1.5 Orthogonal System TcdR – P_{tcdB}

Orthogonal transcriptional control fulfils both criteria of high expression in the target strain, but low expression elsewhere. High levels of transcription can be achieved due to the isolation of the promoter from native regulatory processes by instead having a dedicated *trans* regulatory element. An effective orthogonal promoter system has been demonstrated in clostridia based on an alternative sigma factor from *C. difficile* which natively activates transcription from toxin gene promoters P_{tcdA} and P_{tcdB} both of which are poorly recognised by the *E. coli* RNA polymerase (Minton et al., 2016), but which in *C. difficile* are the only transcriptional signals recognised by TcdR. TcdR has no homolog in most clostridia, including *C. autoethanogenum* making this orthogonal system an appropriate choice for this application.

Transposase control is allowed by the use of the P_{tcdB} promoter from *C. difficile* which is active in the presence of the alternative sigma factor TcdR (Mani and Dupuy, 2001). The TcdR gene was inserted into *C. autoethanogenum* at the

pyrE locus via Allele-coupled exchange (ACE) to render the transposase active in this strain. P_{tcdB} driving the transposase should therefore be the only promoter in the *C. autoethanogenum* genome activated by TcdR hence increasing expression. In *E. coli* the transposase is poorly expressed due to the weak recognition of P_{tcdB} by the *E. coli* RNA polymerase in the absence of TcdR.

Both conditional plasmid loss utilizing P_{fac}-pCB102 and orthogonal transposase expression with chromosomally-located *tcdR* allowing P_{tcdB} expression were demonstrated in *C. acetobutylicum* and *C. sporogenes* (Zhang, Grosse-Honebrink, & Minton, 2015). In this study an attempt was made to replicate this approach in *C. autoethanogenum* and *C. ljungdahlii*.

4.1.6 Aims

- Create a strain of *C. autoethanogenum* expressing *tcdR* from the chromosome.
- Verify activity of the TcdR-activated P_{tcdB} promoter.
- Assess the conditional replicon P_{fac}-pCB102 in *C. autoethanogenum*
- Use the transposon-delivery plasmid pMTL-YZ14 to create transposon mutants in *C. autoethanogenum*.
- Determine the effectiveness of pMTL-YZ14 in terms of transposition frequency and plasmid-loss.
- Perform phenotypic screens on *C. autoethanogenum* transposon mutants.

4.2 Results

4.2.1 Insertion of *tcdR* via Allele-coupled Exchange (ACE) to *C. autoethanogenum*

A protein BLAST was performed for TcdR against the *C. autoethanogenum* genome and no homolog was found, the *tcdR* gene from *C. difficile* was therefore inserted onto the *C. autoethanogenum* genome at the *pyrE* locus using Allele-coupled Exchange (ACE). ACE was described in principle in the introductory chapter and the experimental details are provided in materials and methods section 2.6.2. In short, a Δ*pyrE* mutant strain of *C. autoethanogenum* is restored to prototrophy after a double-crossover event with two plasmid-homology arms which flank the target exogenous gene to be inserted.

Plasmid pMTL-CH20_*tcdR* was cloned inserting promoterless *tcdR* between two homology arms designed to restore a Δ*pyrE* strain of *C. autoethanogenum* to uracil prototrophy through simultaneous restoration of a full *pyrE* gene and insertion of *tcdR* immediately downstream of this. pMTL-CH20_*tcdR* plasmid was conjugated to *C. autoethanogenum* Δ*pyrE* and selected for with thiamphenicol. Colonies were screened for single cross-over events via colony PCR. Successful single cross-over mutants were then streaked onto minimal media lacking uracil to select for double cross-over events restoring *pyrE.* Double crossover mutants were verified with colony PCR followed by sequencing. One sequence verified mutant strain was used to prepare glycerol stocks and was henceforth termed *C. autoethanogenum* TcdR^{+}.

4.2.1.1 Orthogonal Expression Demonstration in *C. autoethanogenum*

The newly created strain *C. autoethanogenum* TcdR$^+$ in theory will be a context in which the TcdR-mediated P_{tcdB} promoter will be active. Plasmid pMTL82254-P_{tcdB} contains the orthogonal promoter P_{tcdB} upstream of *catP* - encoding a chloramphenicol acetyl transferase which can be quantitatively detected in a CAT (Chloramphenicol Acetyl Transferase) assay as well as provide resistance to chloramphenicol *in vivo*. Plasmid pMTL82254-P_{tcdB} was transferred to *C. autoethanogenum* wild type and *C. autoethanogenum* TcdR$^+$ via conjugal transfer. Three separate colonies were grown in liquid culture and 1ml samples taken after 48 hours growth. Promoter activity was assessed via a CAT assay in the TcdR$^+$ strain and compared to expression in the wild type strain (Figure 4.1). The CAT assay methodology is detailed in the materials and methods chapter, section 2.8. As shown in Figure 4.1, the TcdR+ strain exhibits a statistically significantly increase in expression from PtcdB compared to wild type (p value=0.0137).

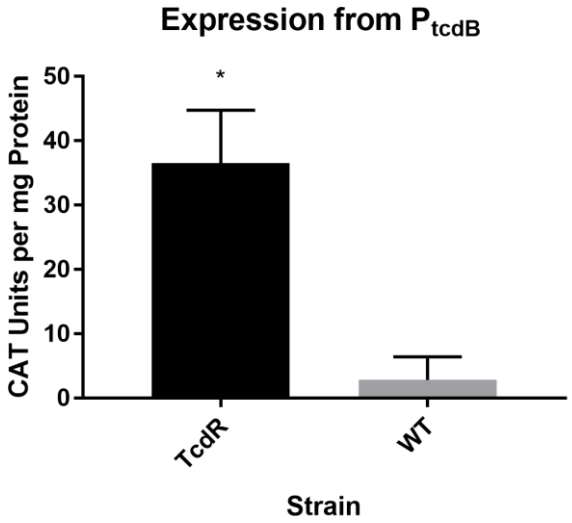


Figure 4.1: Orthogonal expression from P_{tcdB} in *C. autoethanogenum* TcdR$^+$ and wild type.

Expression from P_{tcdB} was quantified using a CAT assay as detailed in section 2.8. Plasmid pMTL-82254-P_{tcdB} was transferred to *C. autoethanogenum* TcdR$^+$ or wild type via conjugal transfer. Three independent colonies were inoculated into liquid YTF which was then subcultured to the same medium and 1ml samples taken after 48 hours. Data represents biological triplicates.

4.2.1.2 Inducible Orthogonal System bgaR-PbgaL-TcdR

The TcdR$^+$ strain of *C. autoethanogenum* was effective at allowing expression from the P_{tcdB} promoter with CAT activity 40-fold higher than in the WT strain. However, in this system TcdR is continually produced causing P_{tcdB} to be constitutively activated. Advantages in terms of overall promoter activity and greater control of the timing of expression would be allowed by placing *tcdR* under the control of an inducible promoter. In order to achieve this another strain of *C. autoethanogenum* termed C24 was created by C. Humphreys in this institute (SBRC Nottingham) via allele-coupled exchange using the same methodology as for *C. autoethanogenum* TcdR$^+$. C24 is identical to *C. autoethanogenum* TcdR$^+$ except for the addition of the *bgaR*-P_{bgaL} lactose inducible promoter system driving *tcdR* instead of being reliant of read-through

transcription from the *pyrE* operon of *C. autoethanogenum* (Hartman et al., 2011). The *pyrE* locus of these two strains is shown in Figure 4.2. A CAT assay was performed in *C. autoethanogenum* C24 as in *C. autoethanogenum* TcdR$^+$ except for the inclusion of ß-lactose at 10 mM or 100 mM concentrations for induced cultures. Either ß-lactose concentration was sufficient to give significant increases in expression over the uninduced culture and the expression was similar to that of the positive control, strong promoter P_{thl} (Figure 4.3).

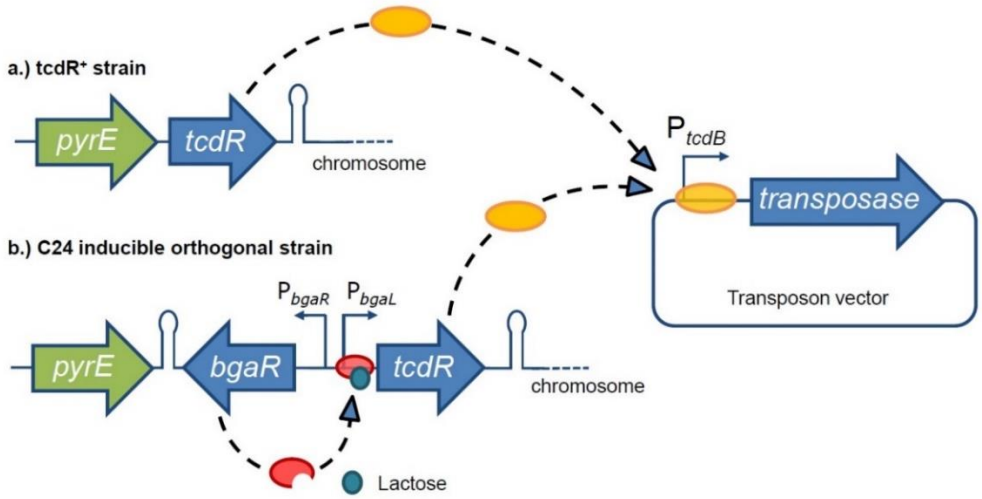


Figure 4.2: Orthogonal expression in two TcdR$^+$ *C. autoethanogenum* strains.

Two strains of *C. autoethanogenum* were used to express proteins via the plasmid-located P_{tcdB} promoter in this study, (a) *C. autoethanogenum* TcdR$^+$ and (b) *C. autoethanogenum* C24. (a) *C .autoethanogenum* TcdR$^+$ expresses the alternative sigma factor TcdR native to *C. difficile* using transcriptional read-through of the *pyrE* operon. TcdR recruits RNA polymerases to the P_{tcdB} promoter. (b) In *C. autoethanogenum* C24 *tcdR* is under the control of the lactose-inducible promoter system *bgaR*-P_{bgaL} from *C. perfringens*. In this way the P_{tcdB} promoter can be induced indirectly via the inducible expression of *tcdR* from the chromosome.

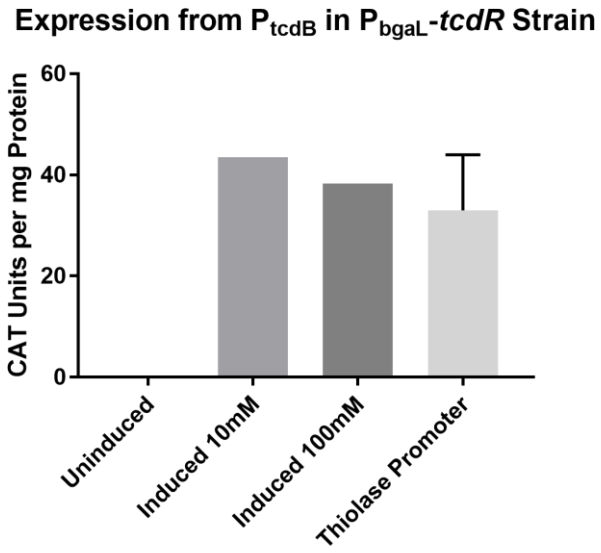


Figure 4.3: Orthogonal expression from P_{tcdB} in an inducible TcdR$^+$ *C. autoethanogenum* strain.

Expression from P_{tcdB} was quantified using a CAT assay as detailed in section 2.8. Plasmid pMTL-82254-P_{tcdB} was transferred to *C. autoethanogenum* C24 via conjugal transfer. Three independent colonies were inoculated into liquid YTF which was then subcultured in the same medium and 1ml samples taken after 48 hours. Simultaneously a control plasmid with *catP* under the control of the known strong promoter P_{thl} was examined and served as a positive control. Data represents biological triplicates for the thiolase promoter or single samples for pMTL-82254-P_{tcdB}.

4.2.2 Plasmid Loss Demonstration in *C. autoethanogenum*

With orthogonal expression demonstrated the next feature of pMTL-YZ14 necessary for efficient mutant creation was that of conditional plasmid loss. Conditional plasmid replication in pMTL-YZ14 is achieved by the placement of an inducible promoter (P_{fac}) upstream of the Gram-positive replicon pCB102. To determine if this system would be effective in *C. autoethanogenum* a direct comparison between two test plasmids was performed. pMTL83251-lacI contains the P_{fac}-pCB102 replicon and so should exhibit instability in the presence of the P_{fac} inducer IPTG and pMTL83251-lacI-T which is identical except for the addition of a transcriptional terminator positioned between P_{fac} and pCB102 preventing transcription-mediated plasmid instability. The P_{fac}

promoter is active in the presence of IPTG and transcription into pCB102 has been shown to cause plasmid loss, the non-permissive condition is therefore the presence of IPTG. The inclusion of pMTL83251-lacI-T allowed the hypothesis that the mechanism of plasmid loss is transcription into the pCB102 replicon to be tested, if the addition of IPTG fails to cause plasmid loss then transcription must be implicated. Figure 4.4 shows the significantly lower CFU on plasmid-selective plates containing IPTG compared to those without IPTG when using pMTL83251-lacI, but no reduction in CFU with the addition of IPTG if the terminator-containing plasmid pMTL83251-lacI-T is used.

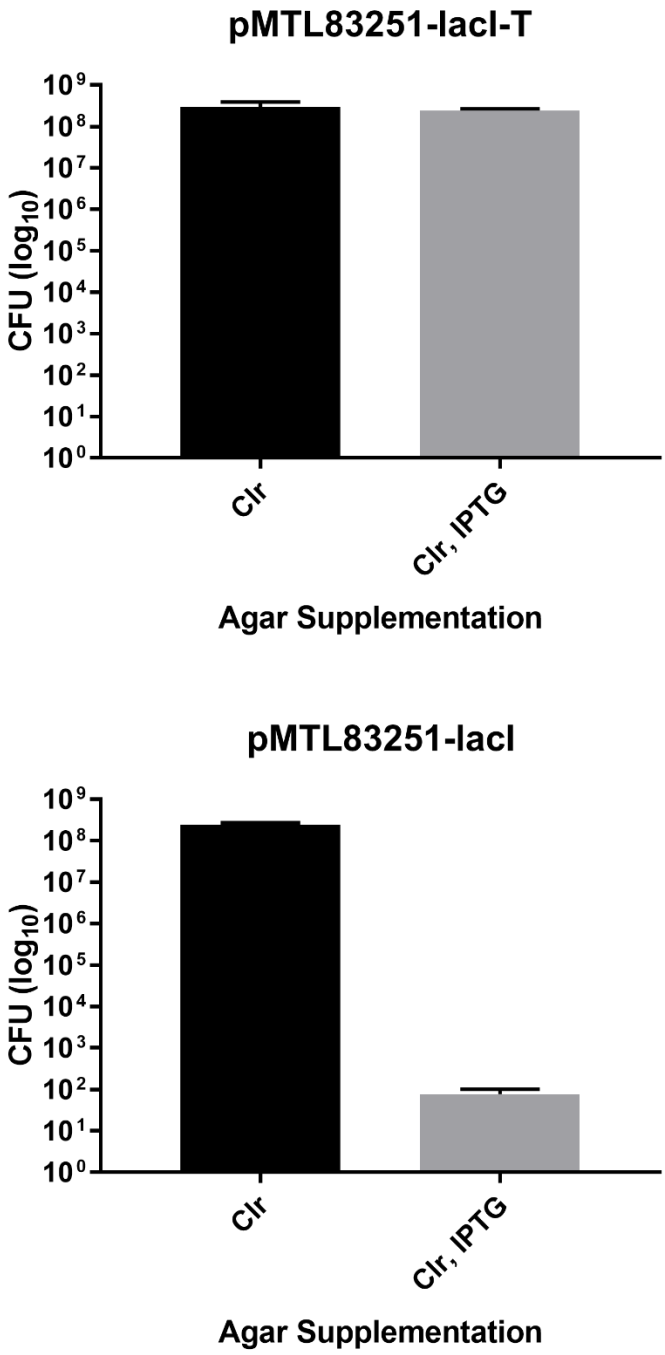


Figure 4.4: Conditional replicon demonstration in *C. autoethanogenum*.

The two test plasmids pMTL83251-lacI and pMTL83251-lacI-T were transferred to wild type *C. autoethanogenum* via conjugal transfer and three independent transconjugants inoculated to liquid growth in YTF for each plasmid. Liquid cultures were grown for 48 hours before being serially diluted and plated to permissive (IPTG-ve) and non-permissive (IPTG+ve) YTF agar plates both supplemented with clarithromycin (Clr) to select for the presence of the plasmid. Data represents biological triplicates.

4.2.3 Conjugation of pMTL-YZ14 into *C. autoethanogenum*

Having established an effective conditional replicon and an inducible, orthogonal expression system the transposon-delivery plasmid pMTL-YZ14 containing both P_{fac} -pCB102 and P_{tcdB} -*Himar1C9* was transformed into the conjugal donor *E. coli* strain CA434. More than 30 separate attempts were made to conjugate pMTL-YZ14 to *C. autoethanogenum* $TcdR^+$ or *C. autoethanogenum* C24 none of which produced transconjugants. Diagnostic restriction digests were frequently performed on the plasmid when harboured by *E. coli* storage or donor strains which in the majority of cases revealed unexpected band patterns indicative of plasmid instability as detailed in chapter 5. These patterns suggested that transposition had occurred internally within the transposon plasmid. Sequencing outwards from the transposon with primer '46_ Himar1C9_Seq_F' demonstrated insertions in a variety of different locations around the plasmid. These results suggest that the orthogonal system does not control the transposase expression sufficiently well to prevent unwanted transposition events in *E. coli.* Many of the unwanted insertions were found within the Gram-positive origin of replication pCB102, likely rendering these plasmids unable to replicate efficiently in *C. autoethanogenum*. Unwanted insertions into the Gram-positive replicon alongside generally poor conjugation frequencies were thought to be responsible for the inability to produce any transconjugants. Both of these issues were addressed in this study: chapter 5 includes a genetic system to confer greater stability to transposon plasmids while chapter 2 detailed efforts to improve DNA transfer frequency. Simultaneous to these efforts it was decided to drive forward system

development working within the genetically similar (100% 16S rRNA sequence similarity), but more genetically accessible *C. ljungdahlii*.

4.2.4 Creation of a TcdR$^+$ Strain of *C. ljungdahlii*

C. ljungdahlii shares 100% sequence identity for the homology arm regions of the plasmid used to create the TcdR$^+$ strain of *C. autoethanogenum* allowing the re-use of this plasmid for the same purpose in *C. ljungdahlii*. CH20_TcdR was introduced to *C. ljungdahlii* Δ*pyrE* (created by Anne Henstra in this laboratory) via electroporation and transformants were selected for on YTF supplemented with thiamphenicol. After 72 hours of growth large colonies were chosen and streaked onto Semi-Defined Minimal Medium (SDMM) agar plates which lack uracil. Growth on SDMM was re-streaked onto the same medium before being inoculated into liquid broth SDMM from which glycerol stocks were made and genomic DNA extracted for verification. Genomic DNA was used as template in a PCR using primers flanking the *pyrE* locus (Figure 4.5). The resultant band was visualised on an agarose gel and found to be of the expected size for a restoration of *pyrE* and insertion of *tcdR*, the band was gel extracted and sent for Sanger sequencing to confirm the insert.

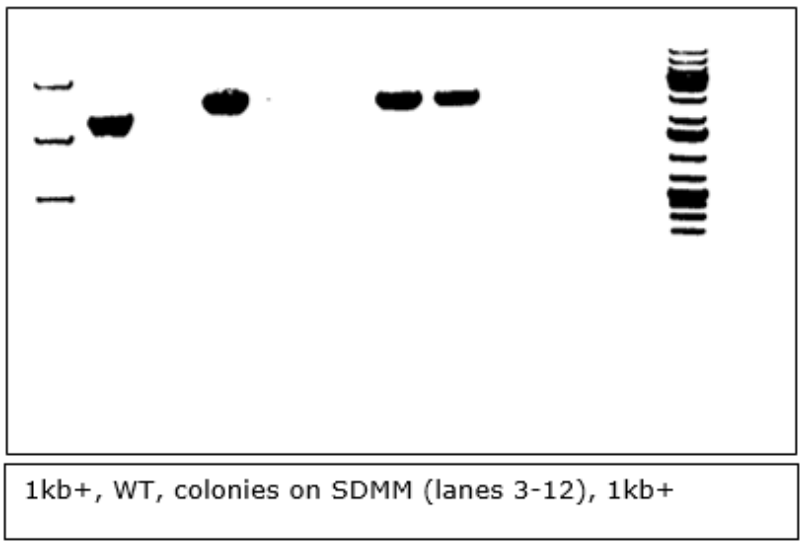


Figure 4.5: *C. ljungdahlii* TcdR$^+$ and Δ*pyrE* with *pyrE* flanking primers.

Genomic DNA was extracted from eight separate colonies appearing on SDMM medium. Genomic DNA from putative TcdR$^+$ colonies as well as the base strain Δ*pyrE* was used as template using *pyrE* flanking primers '25_ACE_pyrE_FSP' and '26_ACE_pyrE_RSP'. *pyrE* flanking primers acting on wild type genomic DNA (lane 2) or putative *C. ljungdahlii* TcdR$^+$ colonies (lanes 3-9). 1kb$^+$ ladder (lanes 1 and 13). Colonies two, five and six show the expected band size for a TcdR$^+$ strain with *tcdR* inserted at the *pyrE* locus.

4.2.5 Plasmid Loss in *C. ljungdahlii*

The conditional replicon P_{fac} -pCB102 has been demonstrated to be effective in *C. autoethanogenum* and the genetic similarities of these species suggests the results will be similar. The same process as was undertaken for *C. autoethanogenum* was repeated for *C. ljungdahlii* but with plasmids pMTL83251-lacI and pMTL83251-lacI-T delivered via electroporation instead of conjugation. Transformants were inoculated to liquid culture before serial dilution and plating after 48 hours. Similar results were obtained for *C. ljungdahlii* as for *C. autoethanogenum* with a substantial reduction in CFU in the presence of IPTG with pMTL83251-lacI, but not for pMTL83251-lacI-T (Figure 4.6).

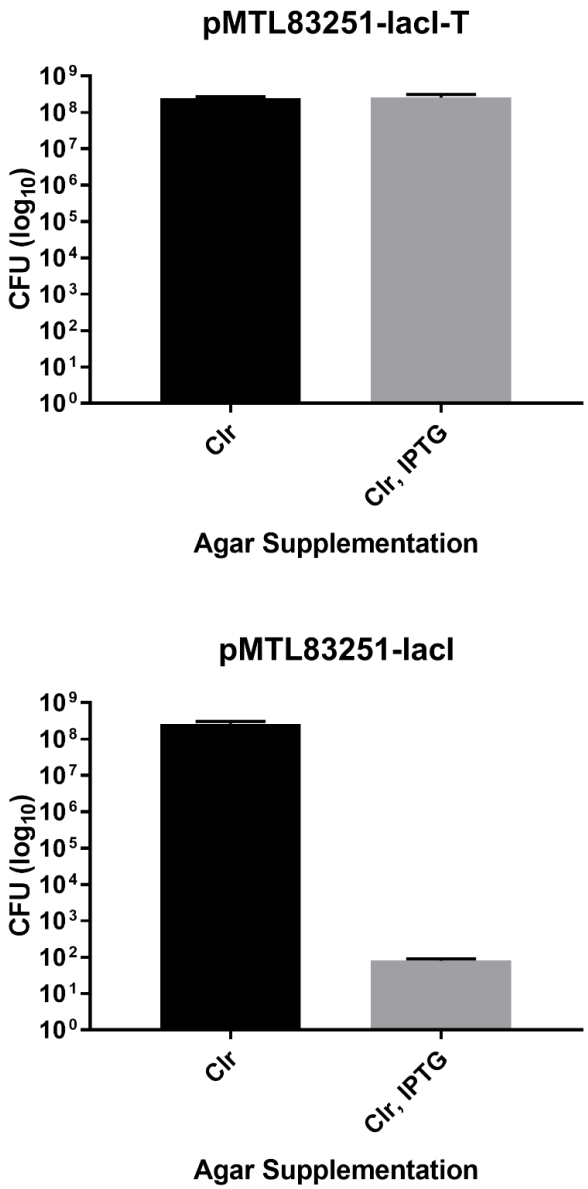


Figure 4.6: Conditional replicon demonstration in *C. ljungdahlii*.

The two test plasmids pMTL83251-lacI and pMTL83251-lacI-T were transferred to wild type *C. ljungdahlii* via electroporation and three independent transformants were inoculated to liquid growth in YTF for each plasmid. Liquid cultures were grown for 48 hours before being serially diluted and plated to permissive (IPTG-ve) and non-permissive (IPTG+ve) YTF agar plates both supplemented with clarithromycin (Clr) to select for the presence of the plasmid. Data represents biological triplicates.

4.2.6 pMTL-YZ14 in *C. ljungdahlii*

C. ljungdahlii proved to be more amenable to transformation with pMTL-YZ14 (Figure 4.7), electroporation was used as a means of DNA delivery in this

organism which has the benefit of avoiding a growth stage in *E. coli* and so skips a stage of verification for plasmid stability which would be necessary if taken through a conjugal donor strain. *E. coli* Top10 were transformed with pMTL-YZ14 and a number of independent overnights were used to prepare plasmid DNA, to mitigate against the risk that some cultures may become dominated by plasmid in which transposon rearrangement has occurred. Plasmid DNA was subjected to a diagnostic digest and sequencing for verification against unwanted transposition events before being used to transform *C. ljungdahlii* $TcdR^{+}$. Transformants were selected on clarithromycin and 8 colonies were re-streaked onto the same media. Growth from re-streak plates was harvested and serially diluted down to 10^{-3} using PBS and plated to IPTG, thiamphenicol to select for transposon mutants. Replica plating to thiamphenicol and clarithromycin was used to estimate the proportion of colonies appearing on thiamphenicol, IPTG plates representing transposon mutants. Colony PCR using primers hybridizing to the transposon plasmid backbone as well as internal transposon primers were used to further verify intended transposition events. For each initial colony appearing on selection plates the percentage of transposon mutants on IPTG, thiamphenicol plates according to replica plate results was between 76% – 100%. A transposition frequency of 1.21×10^{-5} was calculated as the ratio between the number of estimated transposon mutants and the number of plasmid-harbouring cells. Sixty Cla^{S} Tm^{R} colonies were inoculated into liquid culture and cultured for 48 hours before glycerol stocks were prepared and the remainder of the culture used for a genomic DNA extraction in order to determine the insertion site for each mutant. Inverse PCRs on gDNA from twenty mutants was used to verify

that transposition events had occurred, but only eight of these successfully generated products. The eight PCR products were gel extracted and sent for Sanger sequencing using primer '30_catP-INV-R2'. Sequencing results allowed the insertion sites of four mutants to be determined successfully demonstrating that transposition had occurred. Two mutants had identical insertion sites in CLJU_c40490: a thioredoxin reductase. The remaining two mutants contained insertions in CLJU_30660 and CLJU_30440 which encode a predicted signal transduction protein and a putative membrane protein respectively. Ineffectual inverse PCRs notwithstanding, conditional plasmid replication and transposition have been demonstrated in *C. ljungdahlii,* implying that pMTL-YZ14 can be used successfully in the closely related *C. autoethanogenum*. Successes in increasing DNA transfer in *C. autoethanogenum* detailed in chapter 3 led to a renewal of efforts to conjugate the transposon-delivery plasmid pMTL-YZ14 into that species which is of higher academic and industrial interest due to its adoption by Lanzatech, a company currently operating gas fermentation operations at a commercial scale.

4.2.7 pMTL-YZ14 in *C. autoethanogenum*

To increase the conjugation frequency into *C. autoethanogenum* attempts were made to optimise the conjugation protocol, including creation of a novel conjugal donor strain as detailed in chapter 3. Having improved DNA transfer to *C. autoethanogenum* another attempt was made to conjugate pMTL-YZ14 into *C. autoethanogenum C24* and *C. autoethanogenum* TcdR^{+} from Sexpress. Transconjugants were obtained and two individual colonies re-streaked onto clarithromycin selecting for the plasmid backbone. Growth was harvested from re-streak plates, serially diluted and plated to IPTG, thiamphenicol to select for

transposon mutants. Replica plating from mutant selection plates (IPTG Tm) onto clarithromycin (Clr) and thiamphenicol (Tm) agar plates was performed to determine the proportion of colonies on mutant selection plates which represent true genomic insertion transposon mutants. Two colonies were treated separately from the initial conjugation selection plates, for colony 1 100% of colonies on mutant selection plates had the expected phenotype of transposon mutants ($Tm^{R}Clr^{s}$), for colony 2 this value was 89%. Thirty-five separate Tm^{R} Cla^{S} colonies were chosen which should represent transposon mutants with no retained plasmid and subjected to genomic DNA extraction and inverse PCR to identify the transposon insertion site. All 35 colonies contained genomic insertion sites, thirty-one of these colonies represented unique insertion events with three different pairs of the same mutant, 91% therefore represented unique genomic insertion transposon mutants (Table 4.1).

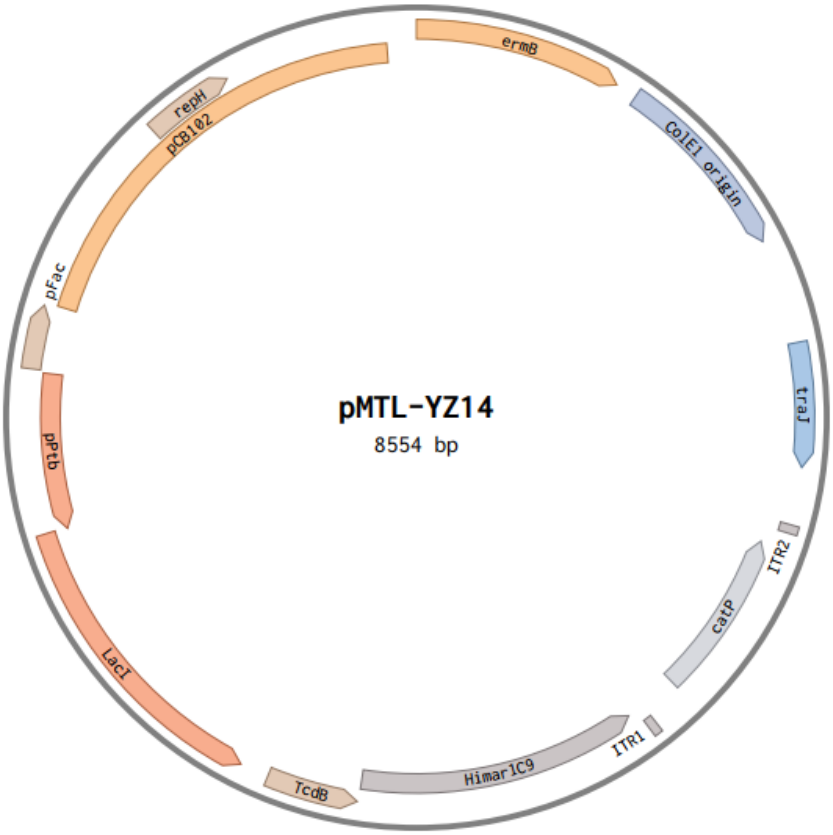


Figure 4.7: pMTL-YZ14

A plasmid map of pMTL-YZ14, a transposon-delivery vector used in *C. autoethanogenum* and *C. ljungdahlii* in this study. pMTL-YZ14 has orthogonal transposase expression from P_{tcdB} and a conditional replicon P_{fac}-pCB102 for which the presence of IPTG is the non-permissive condition.

Table 4.1: Insertion sites for randomly chosen transposon mutants.

Transposon mutagenesis of *C. autoethanogenum* C24 using pMTL-YZ14 produced putative transposon mutants with thiamphenicol resistance from the *catP* marker located within the *mariner* ITRs, but lacking clarithromycin resistance conferred via *ermB* located on the pMTL-YZ14 backbone. Putative transposon mutants were inoculated to liquid culture from which genomic DNA was isolated after 48 hours growth. Inverse PCR was used to identify the transposon insertion-site. Sibling mutants are represented in bold.

Locus tag	Description
CLAU_0595	Putative secreted protein SpoIID domain, at end of gene interrupting last amino acids.
CLAU_3176	HisD, Histidinol dehydrogenase
CLAU_3569	Arsenite methyltransferase
CLAU_3645	Membrane transporter.
CLAU_3004	AAA-ATPase-like protein
CLAU_1148	Sigma-54 specific transcriptional regulator Fis family
CLAU_1961	Purine operon repressor.
CLAU_2233-CLAU_2243	Intergenic
CLAU_2024	AguA, agmatine deiminase
CLAU_3033	Flagellar protein export ATPase FliI
CLAU_0135	**Uroporphyrinogen decarboxylase (URO-D)**
CLAU_0135	**Uroporphyrinogen decarboxylase (URO-D)**
CLAU_1670	ATP-binding region ATPase domain protein putative histidine kinase
CLAU_0697	4Fe-4S ferredoxin iron-sulphur binding domain-containing protein
CLAU_3045	FlhA Flagellar biosynthesis protein
CLAU_2853	GDSL-like Lipase/Acylhydrolase family
CLAU_3049	**RNA polymerase sigma factor. Homologous to FliA in E. coli (flagellar sigma factor)**
CLAU_2965	Flagella hook-associated protein
CLAU_3030	Flagellar M-ring protein
CLAU_3049	**RNA polymerase sigma factor. Homologous to FliA in E. coli (flagellar sigma factor)**
CLAU_3043	FliQ flagella biosynthesis gene
CLAU_2953	CheD chemotaxis protein
CLAU_0255	Membrane protein
CLAU_0524	Transcriptional regulator XRE family
CLAU_3041	FliO flagella biosynthesis gene
CLAU_3965-CLAU_3966	Intergenic
CLAU_3654	Hypothetical protein
CLAU_4059-CLAU_4060	Intergenic, between two proteins of unknown function.
CLAU_3965-CLAU_3966	Intergenic
CLAU_3653-CLAU_3654(adhe1)	Intergenic
CLAU_0715	Hypothetical protein
CLAU_2792	**phosphonate-transporting ATPase**
CLAU_1245	Beta-lactamase domain protein
CLAU_2792	**phosphonate-transporting ATPase**
CLAU_2801	Hypothetical membrane protein

4.2.8 Phenotypic Screens and Mutant Characterisation

In order to demonstrate the value of transposon mutagenesis as a genetic tool several phenotypic screens were performed including amino acid auxotrophy, motility, resistance to Gamma Glutamyl Hydrazide (GGH), resistance to 5-FOA and solvent tolerance.

4.2.8.1 Amino Acid Auxotrophy

Although *C. autoethanogenum* is capable of growing without amino acid supplementation in liquid PETC, all attempts to grow WT *C. autoethanogenum* on PETC agar plates with fructose, but no amino acid supplementation failed. Often a long lag phase is noted when *C. autoethanogenum* is transferred from rich media to a medium lacking amino acids, it is possible that this lag phase causes agar plates to become overly desiccated before colonies can emerge. Liquid PETC with or without casamino acid supplementation was, therefore, employed to screen for amino acid auxotrophs. Two of the sequenced mutants generated previously appeared candidates for amino acid auxotrophs.

CLAU_3645 appears to have a transposon insertion in the first half of a gene annotated *potE* due to similarity to other putrescine transporters (Figure 4.9). Putrescine is produced via the breakdown of amino acids and is toxic in high concentrations, the transporter encoded by *potE* is responsible for both uptake and excretion of putrescine. The second mutant to be investigated contains an insert in CLAU_3176 *hisD* (Figure 4.10) which encodes a histidinol dehydrogenase and is the final enzyme in the biosynthetic pathway toward histidine production according to KEGG pathway prediction (Figure 4.8).



Figure 4.8: Histidine biosynthesis.

The predicted C. autoethanogenum biosynthetic pathway for the production of histidine from KEGG with the position of the interrupted *hisD* gene highlighted.



Figure 4.9: Insertion-site in *potE*.

The transposon insertion-site in *potE* of an amino acid auxotroph. BLAST result of the inverse PCR product. The insertion site is indicated by the black line.

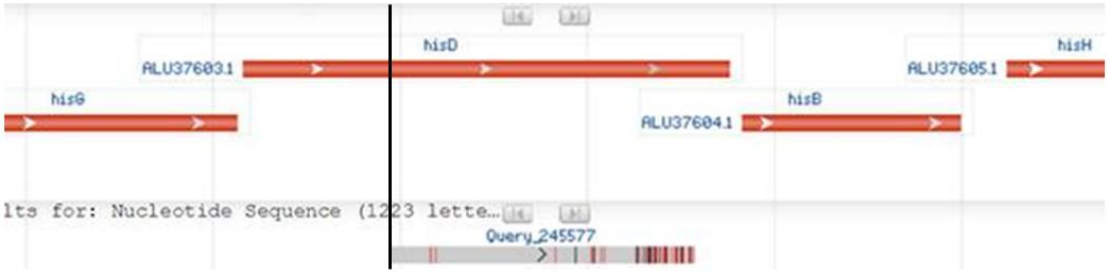


Figure 4.10: Insertion-site in *hisD*.

The transposon insertion-site in *hisD* of an amino acid auxotroph. BLAST result of the inverse PCR product. The insertion site is indicated by the black line.

The two mutants of interest from random sequencing as well as thirty colonies from IPTG thiamphenicol mutant selection plates were inoculated into PETC with or without amino acids and growth observed after 5 days. All colonies inoculated from mutant selection plates grew to OD > 0.8 within 5 days without casamino acid supplementation while the two mutants of interest failed to grow beyond OD 0.8, but did grow with casamino acid supplementation (data not shown).

4.2.8.2 Motility Assay

Several randomly sequenced mutants contained insertions within genes predicted to be implicated in flagella assembly or control. Mutants with insertions in CLAU_3049 *fliA* and CLAU_3045 *flhA* were restored from frozen stock into liquid YTF cultures alongside *C. autoethanogenum* C24 which served as the reference strain. After 72 hours of growth a toothpick was dipped into the culture before being stabbed into the centre of a 0.25% YTF agar plate. After 48 hours of growth the wild type and both mutant strains remained concentrated in the centre of the agar plate, but after 72-96 hours the wild type samples spread diffusely through the agar while both mutant strains remained concentrated centrally (Figure 4.11). At least three repeats were employed for each mutant strain and wild type with consistent results obtained throughout the experimentation.

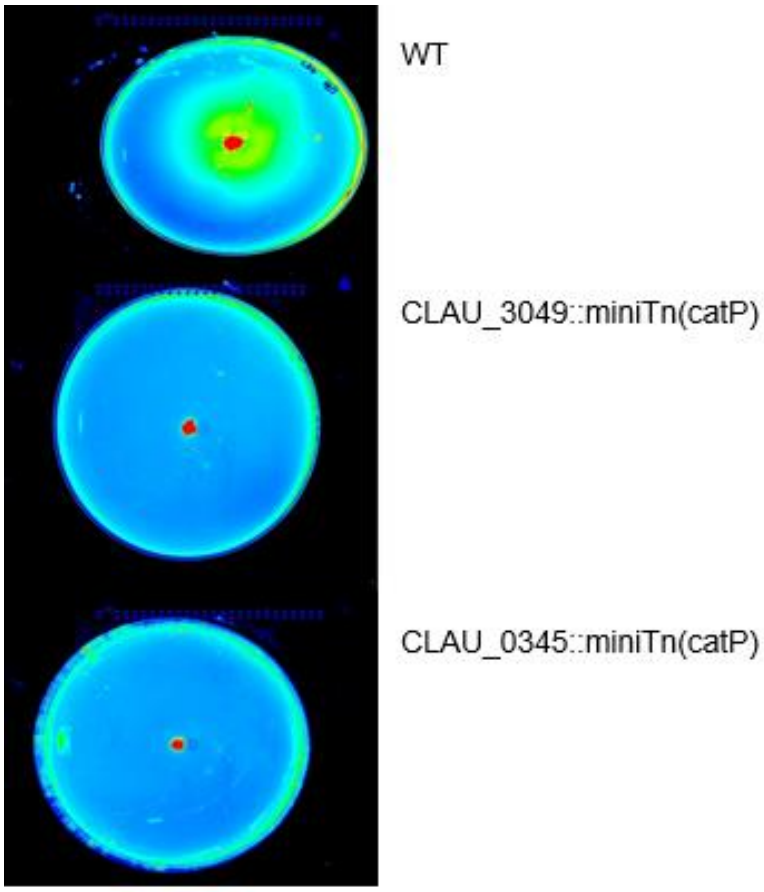


Figure 4.11: Motility assay on wild type *C. autoethanogenum* and putative flagella mutants.

Motility assay comparing wild type to two transposon mutants with insertions in *fliA* and *flhA*. *C. autoethanogenum* C24 and two suspected motility mutants were stab inoculated into the centre of a 0.25% YTF agar plate from liquid culture. After 72 hours growth the agar plates were visualised using an Odyssey® Fc Imaging System. Dense bacterial growth is represented in red whilst more diffuse growth appears in greens and yellows as can be seen for 'WT' surrounding the central inoculation site.

4.2.8.3 Product Formation

Sequencing of a selection of mutants to determine random transposition also revealed several genotypes which could conceivably produce interesting phenotypes in terms of product formation. Two mutants were investigated, the first contained an insertion in CLAU_1148, a predicted Fis family sigma-54 specific transcriptional regulator. Disruption of a major transcription factor was thought a reasonable candidate to produce a difference in product formation.

The second mutant of interest was CLAU_0697::miniTn(catP) with an insertion into a gene coding for a protein with a ferredoxin iron-sulphur binding domain. Interfering with electron exchange was thought to be reasonably likely to change the flux of carbon through central metabolism, potentially diverting toward industrially relevant compounds.

To assess the impact on product formation both mutants were grown in liquid YTF; Samples were taken after 24, 48 and 72 hours and the supernatant analysed by High-Performance Liquid Chromatography (HPLC) which allowed the quantitative detection of ethanol, acetate and lactate. At the 72 hour timepoint there was a significant difference in acetate:ethanol ratio for both mutant strains compared to wild type (Figure 4.12). In addition, ethanol as a proportion of total detected metabolites was increased in the two mutant strains compared to wild type (Figure 4.13).

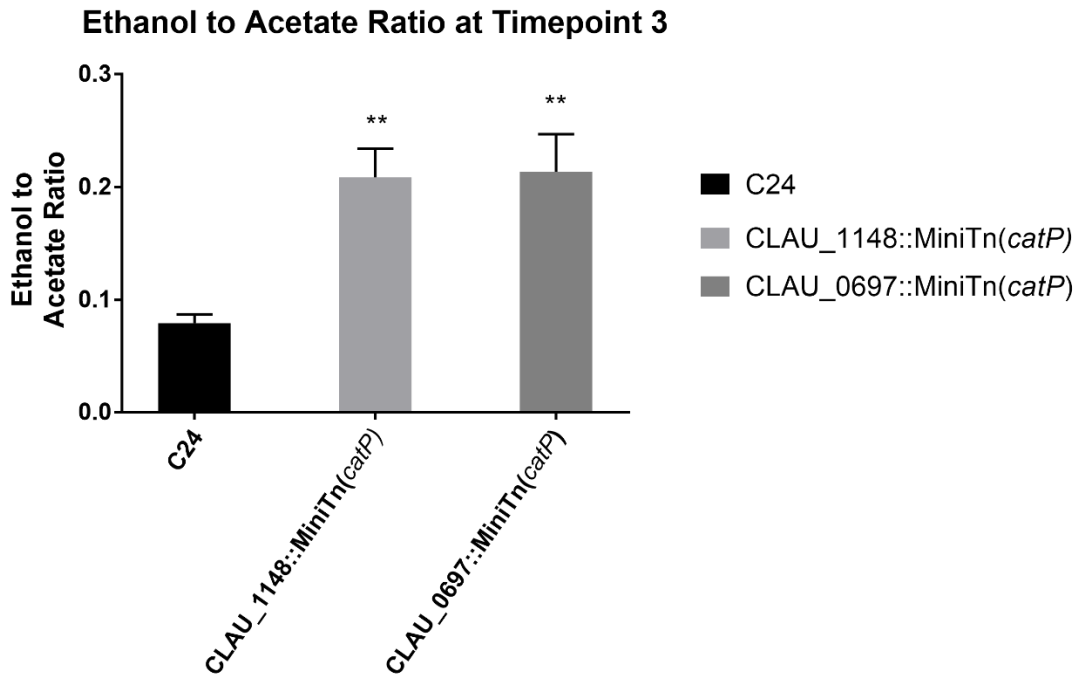


Figure 4.12: HPLC analysis of wild type compared to two transposon mutants after 72 hours growth.

HPLC analysis of the product profile of *C .autoethanogenum* C24 and two transposon mutant strains. 1ml samples were taken at three time-points of liquid culture. This graph represents biological triplicate data only from the 72 hour time point. The ethanol to acetate ratio was calculated and compared to *C. autoethanogenum* C24 for both mutants and in both cases found to be statistically significant: CLAU_1148::mini*Tn*(*catP*) (p value=0.0011 and CLAU_0697::mini*Tn*(*catP*) p value=0.0031.

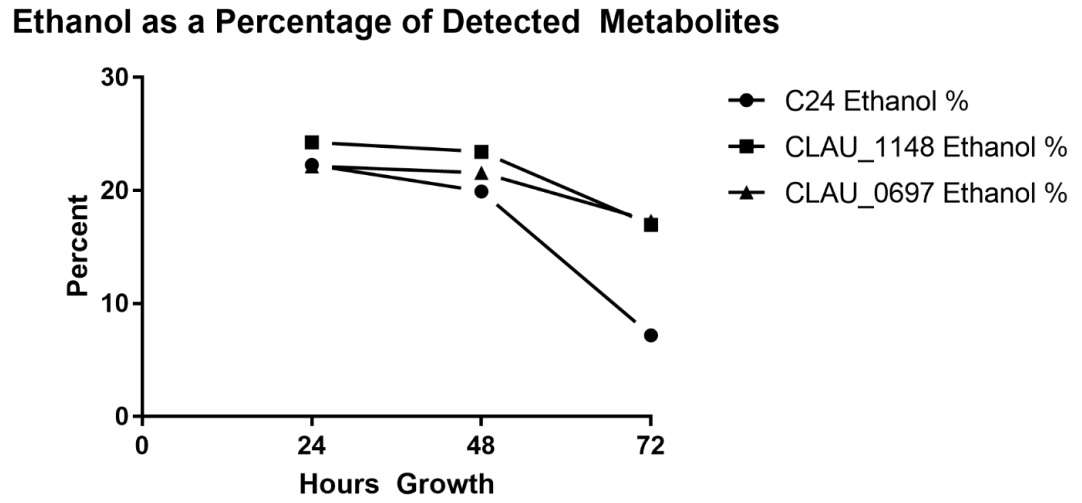


Figure 4.13: Ethanol percentage as a total of detected metabolites.

HPLC analysis of wild type compared to two transposon mutants over three timepoints showing only the percentage of ethanol compared to other detected metabolites lactate and acetate. 1ml samples were taken in triplicate at three time-points of liquid culture.

Transposon Insertion Locus tag	Description of Coding Sequence Disrupted	Improvement in Ethanol to Acetate Ratio over wild type.
CLAU_1148	Sigma-54 specific transcriptional regulator Fis family	102%
CLAU_0697	4Fe-4S ferredoxin iron-sulphur binding domain-containing protein	110%

Table 4.2: Ethanol to acetate ratios for *C. autoethanogenum* C24 and two transposon mutants.

Improvement of the ethanol to acetate ratio over wild type *C. autoethanogenum* for two transposon mutants derived from HPLC analysis of the product profile with 1ml samples taken at three time-points of liquid culture. Data presented in this table is exclusively derived from the final 72 hour timepoint and represents biological triplicates.

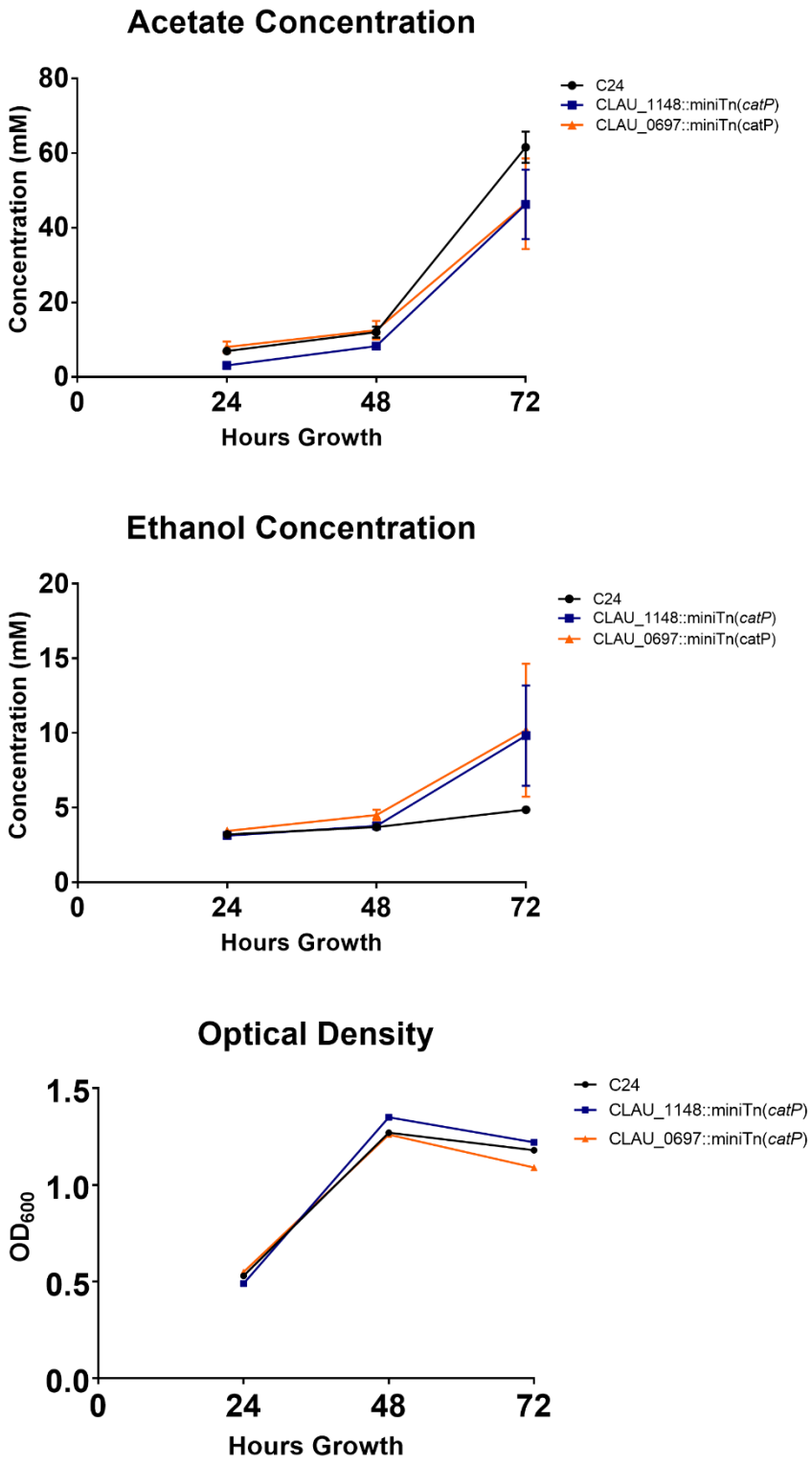


Figure 4.14: HPLC analysis of *C. autoethanogenum* wild type and two transposon mutants

HPLC analysis of wild type compared to two transposon mutants over three timepoints

1ml samples were taken in triplicate at three time-points of liquid culture and analysed using HPLC.

4.2.9 Gamma Glutamyl Hydrazide (GGH) Screening

Gamma glutamyl hydrazide is a toxic analogue of glutamine which has been used to select against a glutamine transport gene *glnQ* in Group B Streptococci where *glnQ* acts as a counter-selectable marker for allele exchange (Tamura, Bratt, Yim, & Nittayajarn, 2005). Screening of a mutant pool of *C. autoethanogenum* could reveal a GGH-resistant mutant with an insertion into an equivalent glutamine transport gene to *glnQ* potentially allowing development of a novel counter-selection in this organism as well as details of amino acid uptake mechanisms.

Growth of *C. autoethanogenum* WT was found to be completely inhibited by GGH at a concentration of 0.5% w/v in YTF medium. A pool of transposon mutants created previously was restored from stock into YTF liquid culture and allowed to grow for 48 hours. The restored mutant pool was then spread onto twenty YTF agar plates supplemented 0.5% GGH agar plates. After 96 hours no colonies were observed suggesting none of the mutants present in the pool were able to prevent uptake of the toxic analogue.

4.2.10 5-FOA Screening

5-fluoroorotic acid (5-FOA) is converted to the highly toxic compound 5-fluoro-uridine monophosphase (5-FUMP) by orotate phosphoribosyl transferase which is encoded by *pyrE*. 5-FOA is used to counter-select the presence of *pyrE* in allele-coupled exchange as described previously (Heap et al., 2012). On rare occasions resistance to 5-FOA is found in colonies which lack the expected genotype, i.e. they retain a functional *pyrE*. In these cases resistance to 5-FOA must have been acquired through a mechanism other than loss of a functional *pyrE*. Screening a transposon pool for resistance to 5-FOA could yield

suggestions of alternative 5-FOA resistance mechanisms which could hamper efforts to perform allele-coupled exchange (ACE).

Growth of wild type *C. autoethanogenum* was found to be completely inhibited by 1mg/ml 5-FOA. A pool of transposon mutants was restored from frozen stock alongside wild type *C. autoethanogenum* and spread onto 20 agar plates of YTF supplemented with 1mg/ml 5-FOA and thiamphenicol, the wild type strain was spread onto one plate. After 72 hours approximately fifty colonies were observed over the twenty mutant pool plates of which eight were inoculated into liquid YTF from which genomic DNA was extracted for inverse PCR and insertion site identification. No colonies were observed on the control wild-type 5-FOA plate.

Five out of eight samples originating from separate 5-FOA^R colonies successfully generated a product in an inverse PCR. Sequencing of the gel extracted product revealed four out of five of these mutants to have the same insertion site intergenically between a predicted multi-drug resistance protein and a hypothetical protein (Figure 4.16), the other affected locus was CLAU_2385 (Figure 4.15).

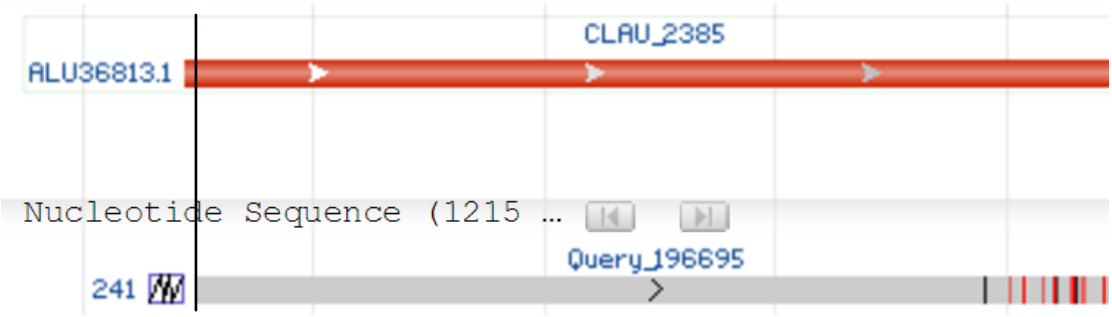


Figure 4.15: Insertion site of a 5-FOA^R mutant.

The transposon insertion site in CLAU_2385 of a 5-FOA resistant mutant.

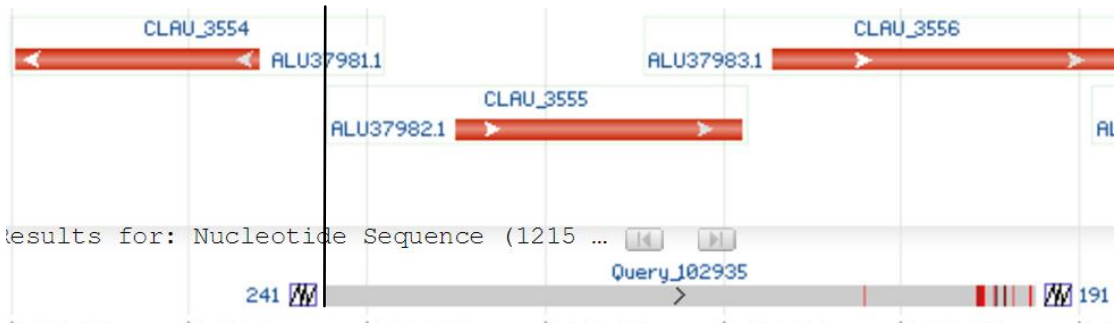


Figure 4.16: Insertion site of a 5-FOA^R mutant.

The transposon insertion site in the intergenic region between CLAU_3554 and CLAU_3555 of a 5-FOA resistant mutant.

4.2.11 Solvent Tolerance.

4.2.11.1 Establishment of Wild Type Resistance Levels

Three potential targets for production by a commercial strain of *C. autoethanogenum* are ethanol, isopropanol and acetone. In order to screen for resistance it was necessary to first establish the tolerance of wild type *C. autoethanogenum*. Wild type *C. autoethanogenum* was restored from frozen stock onto an YTF agar plate and three independent colonies inoculated into liquid YTF. After 48 hours of growth the culture was used as a 1% inoculum into 2ml screw-lock tubes containing liquid YTF with various amounts of ethanol, isopropanol or acetone. Screw-lock tubes were tightly sealed to prevent escape of volatile solvents. After one week, growth was assessed by vortexing the screw-lock tubes and determining the OD_{600} for each culture.

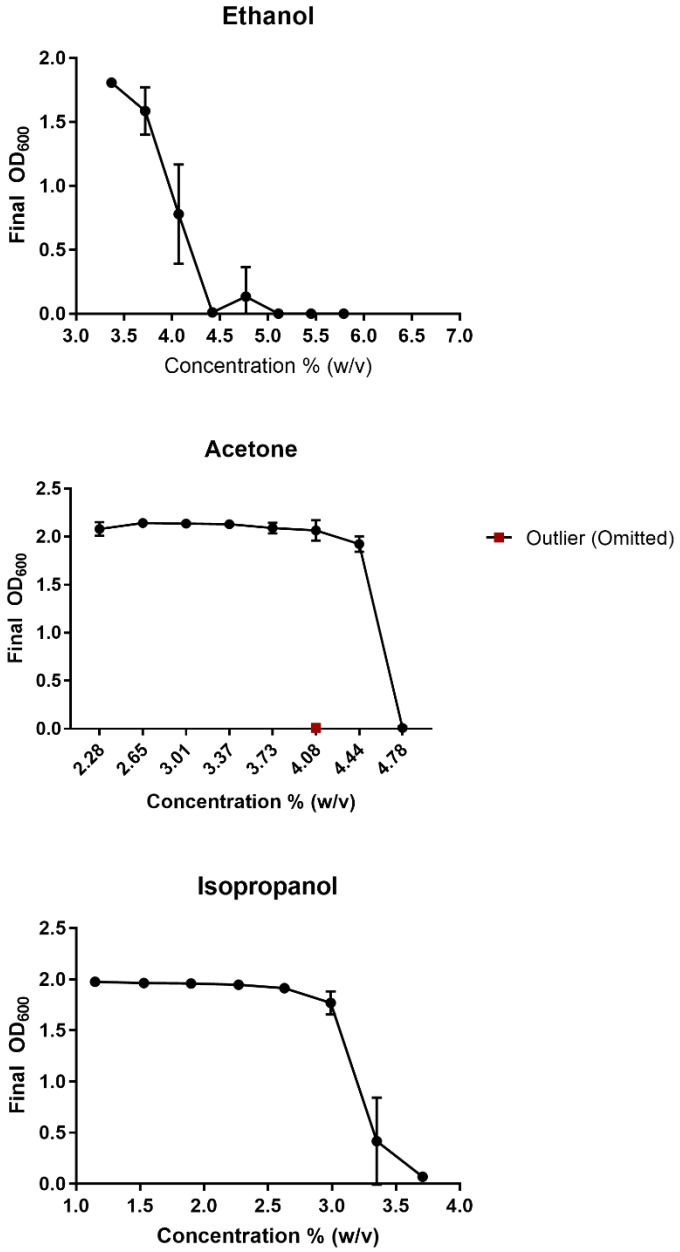


Figure 4.17: Final OD_{600} of wild type *C. autoethanogenum* in varying concentrations of ethanol, acetone or isopropanol.

Wild type *C. autoethanogenum* was restored from frozen stock onto an YTF agar plate and three independent colonies inoculated into liquid YTF. After 48 hours of growth the culture was used as a 1% inoculum into 2ml screw-lock tubes containing liquid YTF with various amounts of ethanol, isopropanol or acetone. Screw-lock tubes were tightly sealed to prevent escape of volatile solvents. After one week, growth was assessed by vortexing the screw-lock tubes and determining the OD_{600} for each culture.

Previously generated transposon mutant pools were restored from frozen stock into liquid YTF and allowed to grow for 72 hours. 1% inoculums were taken from restored pools into screw-lock tubes containing 1ml YTF and a concentration of ethanol, acetone or isopropanol previously shown to completely inhibit growth of the wild type strain which was 4.42% for ethanol, 4.78% for acetone and 3.71% for isopropanol (Figure 4.17). None of the mutant pools in challenge conditions showed any signs of growth implying that no mutants therein had an ability to grow at higher alcohol concentrations. An alternative approach was then taken whereby restored mutant pools were inoculated to YTF with 2.99% isopropanol which was shown to reduce the final OD to an average of 1.77 compared to 1.91 for 2.63% isopropanol cultures. It was thought that this approach may be more forgiving more mutants which have more modest increases in isopropanol tolerance, however at the end of growth no significant difference was noted in final OD_{600} values.

4.3 Discussion

4.3.1 Conditional Replication and Orthogonal Expression

Two genetic systems which facilitate transposon mutagenesis, previously demonstrated in related clostridial species (Zhang et al., 2015), have been shown to be functional in *C. autoethanogenum*: orthogonal expression and a conditional replicon.

The *tcdR* gene encoding an alternative sigma factor native to *C. difficile* was inserted at the *pyrE* locus of *C. autoethanogenum* and *C. ljungdahlii* via Allele-coupled Exchange. Orthogonal expression from P_{tcdB} was demonstrated in *C. autoethanogenum* whereby the $TcdR^{+}$ strain exhibited 40-fold increased activity

over wild type (Figure 4.1). *C. autoethanogenum* C24 with *tcdR* under the control of a lactose-inducible promoter system from *C. perfringens* was shown to support orthogonal expression in a similar manner to the TcdR$^+$ strain (C. Humphreys, unpublished) (Figure 4.3). When induced with 10 mM lactose expression is higher than the strong thiolase promoter positive control and has an induction/repression ratio of 40 (Harris, Welker, & Papoutsakis, 2002b; Tummala, Welker, & Papoutsakis, 1999). The inducible orthogonal system therefore should offer a suitable expression system for the transposase. In *E. coli* there should be little expression due to the lack of the alternative sigma factor TcdR whilst in the target strain with *bgaR-*P_{bgaL}*-tcdR* the addition of lactose leads to strong expression from the P_{tcdB} promoter.

4.3.2 Conditional Replicon P_{fac}-pCB102

The P_{fac}-pCB102 system creates a highly effective way to force plasmid loss via the addition of IPTG. On average around 1 in 16.5 million cells harbouring plasmid will form colonies on IPTG-containing plates in both *C. autoethanogenum* and *C. ljungdahlii*. The positioning of a transcriptional terminator between the inducible promoter and pCB102 appears to entirely inhibit the plasmid-loss effects of the addition of IPTG. Taken together these results suggest that P_{fac} is sufficiently tightly controlled in the absence of IPTG to allow pCB102 to function uninhibited to the same degree as the placement of a terminator after the induced promoter. The effectiveness of the transcriptional terminator of the ferredoxin gene from *C. pasteurianum* is also suggested by the complete negation of the effect of IPTG addition.

If non-replicative transposon delivery is not feasible then the ability to force plasmid loss in highly advantageous to transposon mutagenesis studies and

functionally essential to TIS experiments. There are two mechanisms by which plasmid-loss is attained, either the plasmid copy-number control mechanisms are interfered with leading to segregational instability and loss through a dilution in the population or, cells harbouring plasmids are counter-selected directly. Which of these mechanisms is acting in this case is relevant to TIS studies, with each having an advantage. If segregational instability is the dominant mechanism then a sufficient number of generations of growth must be allowed for plasmid-loss to occur, this reduces the library complexity since mutants arising prior to forcing plasmid-loss will amplify exponentially during these generations. If on the other hand plasmid-harbouring cells are counter-selected directly then cells in which a transposition reaction has occurred will be lost since they will still contain the transposon-delivery plasmid vector, this will reduce library size since the only colonies appearing on mutant selection plates will be those resistant to the toxic effects of forced plasmid-loss or those which have by chance inherited no plasmid. This subject will be discussed further in the following chapter.

4.3.3 Construction of *C. ljungdahlii* TcdR$^+$ and Transposon Mutagenesis

C. ljungdahlii Δ*pyrE* was restored to uracil prototrophy with the simultaneous insertion of *tcdR* encoding the alternative sigma factor TcdR from *C. difficile* via ACE. The resultant strain *C. ljungdahlii* TcdR$^+$ was transformed with pMTL-YZ14 allowing the transposition frequency, proportion of transposon mutants on selection plates and plasmid loss values to be calculated. The transposition frequency of 1.21×10^{-5} is comparable to similar studies (Cartman & Minton, 2010; Zhang et al., 2015). The percentage of colonies on mutant selection plates representing transposon mutants without plasmid averaged 88.9% using

replica plating results or calculated as 98.6% using CFU counts. This places *C. ljungdahlii* between *C. acetobutylicum* (80%) and *C. sporogenes* (100%) for this value according to the literature (Zhang et al., 2015). Transposition was verified by inverse PCR and sequence verification of the insertion site of several mutant strains. The results obtained from mutagenising *C. ljungdahlii* with pMTL-YZ14 suggested that this transposon-delivery vector would also be appropriate in *C. autoethanogenum* due to their genetic similarity.

4.3.4 Transposon Mutagenesis of *C. autoethanogenum* with pMTL-YZ14

After initial attempts to conjugate pMTL-YZ14 into *C. autoethanogenum* TcdR^{+} failed, efforts were made to improve DNA transfer. The use of a novel conjugal donor strain Sexpress (chapter 3) allowed conjugation of pMTL-YZ14 into strains of *C. autoethanogenum*. A strain of *C. autoethanogenum* with *tcdR* under the control of the lactose-inducible promoter system bgaR-P_{bgaL} (C24) was used to generate transposon mutants. The transposition frequency ($1.21x10^{-5}$) appears to be similar to studies in *C. acetobutylicum, C. sporogenes, C. difficile* (Cartman & Minton, 2010; Zhang et al., 2015) and *C. ljungdahlii*. All mutants analysed with inverse PCR produced only one distinct band on the resultant gel, though Southern blotting would more accurately demonstrate single-insertion mutants. Of the thirty-five mutants with insertion-sites identified there were three pairs of identical mutants implying that transposition occurs prior plating on mutant selection plates. This is non-ideal, but not unexpected when working with a replicative vector for transposon-delivery. Attempts to ameliorate this effect are outlined in chapter 5.

4.3.5 Characterisation of Sequenced Mutants

4.3.5.1 Amino Acid Auxotrophs

Two mutant strains with previously identified insertion sites were inoculated into PETC with or without amino acid supplementation alongside thirty other random mutant colonies. The previously identified strains both proved unable to grow in PETC lacking amino acid supplementation, but grew in the control condition implying the acquisition of an amino acid auxotrophy. Of the two strains the mutant with an insertion in *hisD* (CLAU_3176) is less surprising since *hisD* encodes a histindol dehydrogenase which is predicted to be the final enzyme required for histidine biosynthesis in a KEGG pathway. The second strain is more interesting since it interrupts *potE* which may encode a membrane transporter for putrescine. The toxic effects of putrescine have not been documented in *C. autoethanogenum*, but this is perhaps the first suggestion that its accumulation could be lethal in this organism. Why a knockout strain is viable with amino acid supplementation, but not without requires further speculation since putrescine is typically a product of the breakdown of amino acids. Perhaps the network of amino acid biosynthesis which is required without supplementation produces a higher concentration of toxic by-products which *potE* helps to remove. Alternatively it may be the uptake abilities of PotE which are required in minimal media causing the lack of viability for the knockout strain (Kashiwagi, Shibuya, Kuraishi, & Igarashi, 1997).

4.3.5.2 Motility-deficient Mutants

Two transposon mutant strains were assessed for their ability to swim through 0.25% agar, both proved unable to do so confirming genes *fliA* (CLAU_3049) and *flhA* (CLAU_3045) as being essential for flagella functionality. Several more

flagella-related mutant strains were identified giving a high overall representation within the sequenced mutants. This is interesting in light of the discovery in several TIS experiments of an over-representation of flagella mutants in subcultured libraries suggesting the flagella may be a particularly energetically expensive addition to the bacterial cell (Fels, Zane, Blake, & Wall, 2013; Langridge et al., 2009). This is not to imply that transposition into flagella related areas of the genome is inherently more likely to occur. If flagella mutants were found to be over-represented the proposed mechanism would have to rely on an increase fitness value in a proliferation stage prior to plating on mutant selection plates.

4.3.5.3 Product Formation

Two sequenced mutants were analysed in terms of their product formation profile in rich media using HPLC. The first two timepoints at 24 and 48 hours showed no significant difference to the wild type profile, but at 72 hours both strains exhibited significantly improved ethanol/acetate ratios. At the third timepoint the ethanol concentration was higher and the acetate concentration was lower in both mutant strains compared to wild type.

When growing using fructose as a carbon source both CLAU_1148::mini*Tn*(*catP*) and CLAU_0697::mini*Tn*(*catP*) had statistically significant (p value=0.0011 and 0.0031, respectively) improvements in terms of ethanol/acetate ratios over the background strain *C. autoethanogenum* C24 at the 72 hour timepoint (Figure 4.12). This was not the case for samples taken at 24 and 48 hours where ethanol constitutes a similar percentage of the total metabolites at these two timepoints, only at 72 hours does C24 have a reduction

in ethanol percentage while the two mutant strains maintain a similar value as earlier timepoints (Figure 4.13).

Transposon Insertion **Locus tag**	**Description of Coding Sequence Disrupted**	**Improvement in Ethanol to Acetate Ratio over wild type.**
CLAU_1148	Sigma-54 specific transcriptional regulator Fis family	102%
CLAU_0697	4Fe-4S ferredoxin iron-sulphur binding domain-containing protein	110%

Table 4.2 shows that both mutant strains produce more than double the concentration of ethanol after 72 hours growth and both produce significantly less acetate than C24. Optical density measurements imply that all three strains were at a similar phase of growth for the three timepoints sampled (Figure 4.14).

Together these results warrant future work. Acetate is of lower commercial value than ethanol and as such this result is promising in terms of future work characterising these mutants in more industrially relevant conditions including growing in a minimal, defined medium and utilising CO or $CO_2 + H_2$ as a sole carbon and energy source. If the product profile in these conditions is of interest then the genes implicated should be studied further with directed, clean mutation techniques and overexpression from a plasmid-based vector.

4.3.6 Screening Mutant Pools

pMTL-YZ14 was then used to generate several pools of transposon mutants which were subjected to screens for amino acid auxotrophy, GGH resistance, 5-FOA resistance and solvent tolerance. Screening of mutant pools failed to identify mutants resistant to GGH nor with significantly greater tolerance to

ethanol, acetone or isopropanol, but two unique 5-FOA^R transposon mutants have been isolated and the transposon insertion-site determined.

4.3.6.1 GGH

The mutant pool may have been insufficiently complex to contain a mutant with GGH resistance. Alternatively there may be redundancy in GGH uptake mechanisms and several genes would need to be knocked out to find the phenotype of interest. The hypothetical multiple mutant knockout may be unlikely to exist in a transposon pool primarily consisting of single insertions, *B. subtilis* is known to have multiple systems for glutamine uptake (Satomura, Shimura, Asai, Sadaie, & Hirooka, 2005). Another possibility is that disruption of the uptake mechanism for GGH is lethal under these experimental conditions.

4.3.6.2 5-FOA

Five separate colonies appearing on 5-FOA selection plates were subjected to inverse PCR and Sanger sequencing into the insertion-site. Four of the five colonies represented identical insertions intergenically between genes encoding a small multidrug resistance protein and a hypothetical protein. The hypothetical protein returned no results with a pfam search. The small multidrug resistance protein contains (PF00893) a domain thought to confer drug resistance via efflux. It is possible that the transposon insertion 5' to this efflux protein increases expression hence exporting enough 5-FOA from the cell to produce non-lethal intracellular concentrations. This could be due to interference with the native transcriptional control mechanisms upstream of CLAU_3555 or the divergently oriented hypothetical protein CLAU_3554 which may itself regulate CLAU_3555. Transcriptional read-through from the transposon itself could also be directly responsible for an upregulation of the

small multidrug resistance protein, this is consistent with the orientation of the insertion.

A spontaneous mutation elsewhere in the genome could also have conferred 5-FOA^R, though this must have occurred prior to exposure to 5-FOA since multiple independent colonies exhibit this resistance. To explore this further directed knockout strains for the two genes flanking the insertion could be created and assessed for 5-FOA^R. Additionally the efflux protein CLAU_3555 and the hypothetical protein CLAU_3554 could be expressed from a strong promoter on a heterologous plasmid in a WT background and the sensitivity to 5-FOA assessed. The mutant strain CLAU_3554/CLAU_3555::miniTn(catP) could also be analysed via whole genome sequencing to search for mutations elsewhere on the genome which could explain the 5-FOA^R phenotype. The other insertion-site linked to a 5-FOA^R phenotype was found to be within CLAU_2385 a predicted hydroxyacid dehydrogenase. This has no obvious link to 5-FOA resistance and a link to that phenotype must be established in the same way as described for CLAU_3554/CLAU_3555::miniTn(catP). (zfc::miniTn(catP)). Insertions into *pyrE* or *pyrF* were expected to be found since the deletion of these genes is well documented to confer resistance to 5FC and is a crucial part of the ACE method (Heap et al., 2012; Ng et al., 2013). The absence of these mutants suggests that the library used for screening was insufficiently complex or that too few colonies were examined.

4.3.6.3 Solvent Tolerance

Ethanol is a native product of *C. autoethanogenum*, while proof of concept production of both acetone and isopropanol has been demonstrated via incorporation of exogenous genes (Bengelsdorf et al., 2016; Koepke, Simpson,

Liew, & Chen, 2012). Wild type tolerance to ethanol, acetone and isopropanol was studied and data is presented in Figure 4.17. Growth of wild type *C. autoethanogenum* was completly inhibited by 4.42% (w/v) ethanol, 4.78% acetone and 3.71% isopropanol. Figure 4.17 also shows that ethanol seems to have a greater range of concentrations at which it is inhibitory to growth without preventing growth entirely compared to acetone and isopropanol which exhibit a noticeably precipitous decline in final OD_{600} values. The difference between completely inhibited growth and unimpeded growth is >1.75% for ethanol and only 1.05% or 1.43% for acetone and isopropanol respectively. The range of concentrations tested for ethanol should have been extended at the least concentrated end since the lowest concentration tested produced a final optical density (1.81) lower than that of unimpeded isopropanol (1.98) or acetone (2.08) cultures. In order to analyse the range of semi-inhibitory concentrations the ranges tested should all be extended until they include final optical densities which match cultures with no added solvent.

Several mutant pools were subjected to the inhibitory concentration of ethanol, acetone and isopropanol as determined in this study. No mutant pool was able to grow reliably in these challenge conditions depriving us from a simple identification of a potentially useful and revealing mutant strain. This was likely either due to a lack of mutants which could grow at that concentration of solvent or because of the inhibitory effect of the non-viable mutants in the pool. A lower concentration of isopropanol was then used which caused wild type to grow to a lower final OD but mutant pools failed to produce a significant difference in final OD values.

4.4 Conclusion

4.4.1 Results Summary

- Transcription into pCB102 via the inducible promoter P_{fac} causes plasmid instability in *C. autoethanogenum* and *C. ljungdahlii*.
- ACE was used to insert *tcdR* to the chromosomes of *C. autoethanogenum* and *C. ljungdahlii*
- P_{tcdB} was shown to be active in the presence of TcdR.
- pMTL-YZ14 was demonstrated to be an effective means of transposon delivery in *C. autoethanogenum* and *C. ljungdahlii*.
- Two amino acid transposon mutant auxotrophs were isolated.
- Two transposon mutants were demonstrated to have a significant increase in ethanol/acetate concentrations.
- Insertions into two chromosomal loci were linked with tolerance to 5-FOA.
- Two transposon mutants with inserts in *fliA* and *flhA* were shown to be non-motile.

5 Developing Vectors for Transposon Insertion Sequencing

5.1 Introduction

Transposon Insertion Sequencing (TIS) is a powerful technique capable of determining the essential gene set for a given growth condition. A major obstacle preventing the technique from being applied to a greater number of different organisms is the technical challenge of creating a sufficiently large mutant pool. This is especially problematic in organisms with a low efficiency of DNA transfer. High complexity libraries have multiple, unique insertions in each locus, with a greater number of unique mutants comes increased statistical power to determine regional essentially and higher precision mapping of the genome. With a sufficiently high insertion density the subunits within genes can be determined to be essential or non-essential giving clues as to the reason why the gene is essential.

The majority of TIS studies have created libraries with either pre-assembled transposon-transposase complexes (transposomes) or suicide constructs. In these cases, the efficiency of mutant generation is a combination of the efficiency of DNA transfer and the transposition efficiency within the cell. With replication-deficient delivery, transposition frequency is usually approximately three orders of magnitude lower than the plasmid electroporation frequency (Lamberg et al., 2002) meaning that library creation is only feasible in readily transformable organisms. More recently, there have been several TIS studies

using a replicative plasmid as a means of transposon-delivery (Dembek et al., 2015; Charbonneau et al., 2017).

A replicative vector harbouring transposon and transposase allows the outgrowth from a low number of colonies on transformation plates to create a large population of cells which harbour transposon-delivery plasmids, each of which are capable of hosting independent transposition events. This effectively creates a comparable number of cells with the transposon and transposase as could be achieved using a highly transformable organism. However, this approach presents two major new challenges to TIS experiments: plasmid-based transposon read-contamination and domination of early transposition event mutants within the mutant pool. At the biological library stage this manifests as a decoupling of colony number and number of unique mutants. At the sequencing library stage both problems result in the same symptom - a depletion of unique, chromosomal reads.

5.1.1 The Importance of Unique, Chromosomal Reads.

Genomic DNA prepared from transposon pools is used to prepare an amplicon sequencing library which is then sequenced on an Illumina platform (usually a MiSeq or HiSeq). A custom sequencing primer is used which hybridises within the transposon. Sequencing reads are scanned for the presence of a 'transposon tag' which represents the final 8-12bp of the transposon. If a read contains a transposon tag it is treated as a genuine transposon read. The neighbouring sequence downstream of the transposon tag is then mapped against the reference genome to identify the transposon insertion site. During sequencing library preparation the primary objective is to maximise the number of unique, chromosomal reads and avoid bias toward any particular insertion-

site. Bias can occur during the PCR amplification stage resulting in ‘jackpot’ amplification products which decouple the crucial link between read number and mutant prevalence. High quality TIS experiments will exhibit multiple separate insertion sites in every non-essential gene and usually several unimpactful insertions into the 3’ end of essential genes which do not disturb protein function. Detection of reads originating from multiple sites within a gene allows greater confidence in ascribing non-essential status to that gene, with each separate insertion-site essentially serving as a biological replicate.

5.1.2 Plasmid Read-contamination

The transposon-delivery vectors detailed in this chapter use conditionally-replicative Gram-positive origins of replication for which the presence of IPTG is the non-permissive condition. Transposon mutants are selected for using thiamphenicol (to select for the presence of the transposon-located *catP* marker) and IPTG (to select against plasmid-retention). The conditional system was described and demonstrated in chapter 4 where it was demonstrated that only around 1 in 7 million plasmid-harbouring cells are able to form colonies on IPTG-containing plates. For TIS experiments the proportion of chromosomal transposon mutants to plasmid-harbouring cells is the relevant value. Colonies appearing on IPTG, thiamphenicol (mutant selection) plates are either the result of a transposition event or due to rare plasmid-retention effects in the presence of IPTG. The proportion of mutants to plasmid-harbouring cells is therefore a function of the transposition frequency and the efficiency of plasmid-loss. Vector development in this chapter will therefore aim to improve both of these values.

Transposon mutagenesis of *C. autoethanogenum* using pMTL-YZ14 was described in the previous chapter and succeeded in producing 89-100%

chromosomal insertion mutants on mutant selection plates. This compares to values of 80% and 100% for *C. acetobutylicum* and *C. sporogenes* respectively (Zhang et al., 2015). If the consequence of this plasmid-contamination was that only 0-20% of reads were plasmid-based rather than chromosomal this would not be of large consequence, however the problem of plasmid-contamination is compounded by the impact of copy number. Figure 5.1 exemplifies this problem with various copy numbers. To counteract the effects of plasmid-contamination a range of different transposon-delivery systems were constructed and characterised with the aims of improving plasmid loss efficiency, increasing transposition frequency and reducing plasmid copy number in order to ultimately decrease the number of sequencing reads originating from the transposon plasmid.

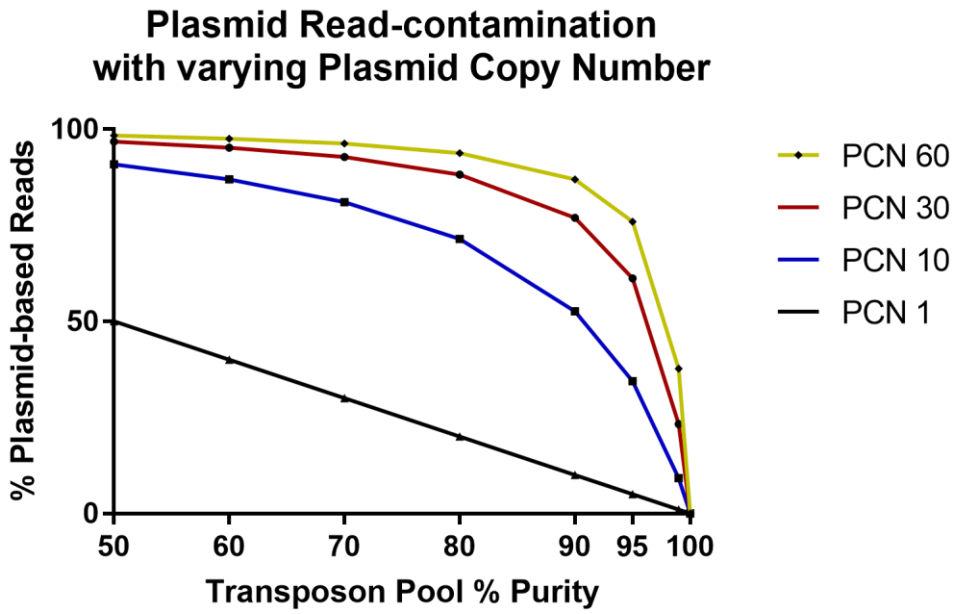


Figure 5.1: Effect of plasmid copy number.

The effect of plasmid copy number (PCN) on the percentage of the sequencing library expected to represent target genomic insertions (% purity).

5.1.3 I-SceI Strategy

It has been suggested that the problem of plasmid read-contamination could be addressed at the sequencing library preparation stage by the incorporation of an additional restriction digest step using the I-SceI endonuclease (R. Chaudhuri, personal communication). A restriction digest on the adapter-ligated products would cleave any strands representing transposon-plasmid junctions and yet leave transposon-chromosomal junctions intact with the effect that chromosomal junctions will be amplified in the PCR-enrichment stage and plasmid junctions will not since they will lack binding sites for the adapter hybridising primer. A modified transposon-delivery plasmid was therefore constructed with I-SceI recognition sequences flanking the transposon.

5.1.4 Early Transposition Events

Creation of transposon mutant pools using non-replicating DNA renders the problem of early transposition events dominating the transposon pool insignificant. With this approach it can be assumed that every colony appearing on selection plates is likely to be the result of a unique transposition event. Exceptions to this can occur if a long recovery time with several generations of growth are allowed before plating, or non-integrative carriage of the transposon occurs, but in general non-replicative transposon delivery ensures a direct correspondence between colonies obtained on initial selection plates and the number of unique insertion mutants in the library. For this reason the number of colonies on selection plates is a reliable estimate of not only total number of mutants, but also total number of unique mutants i.e. library complexity. Unfortunately this is not the case when using a replicative plasmid for transposon-delivery. In this instance a transposition event which occurs early

on in the mutant pool creation process can represent a significant proportion of the eventual colonies on mutant selection plates and therefore dominate the number of reads obtained from sequencing. The inability to confidently link the number of colonies on selection plates to the complexity of the library is a major disadvantage of replicative plasmids. Estimates of library complexity can be derived by randomly selecting a number of mutants and identifying the insertion-site by inverse PCR, or through Sanger sequencing from the sequencing library preparation prior to Illumina sequencing, but these methods are time-consuming and relatively inaccurate.

Several approaches to mitigate the impact of early transposition events were investigated during this study. The most obvious approach is to limit the number of generations in which cells harbour the transposon plasmid before they are plated to mutant selection plates, and with this in mind various mutant generation procedures were trialled. Another more ambitious approach involves transcriptional control of the transposase with inducible promoter systems. Finally the possibility of conditional lethality of transposition events was examined.

Recently, TraDIS (a version of TIS) was attempted using pMTL-YZ14 in *C. beijerinckii* (G. little, unpublished). In this experiment, both domination of early transposition event mutants and plasmid read-contamination combined resulting in a low complexity sequencing library. Plasmid read-contamination was evident as 54% of reads mapped to the plasmid and only 4.9% of the total reads contained the correct transposon tag and mapped to the reference genome. Domination of early transposition event mutants was also observed as of these 4.9% reads which successfully mapped to the chromosome, a large

proportion mapped to only a few sites (Figure 5.2). The net result of these two issues is that costly further sequencing runs would be required to acquire a sufficiently saturated sequencing library to make confident predictions of gene essentiality.

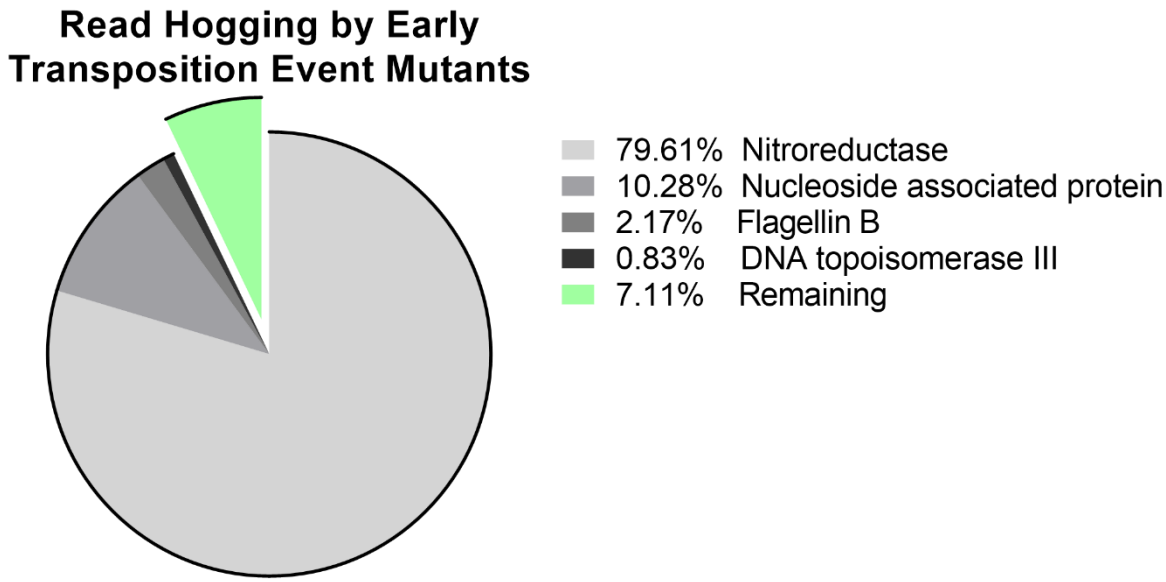


Figure 5.2: Read distribution.

Percentage read occupation showing the top 3 represented genes and the percentage which represents reads in all other genes.

5.1.5 Unwanted Transposition Events in Donor and Storage Strains

Unwanted transposition events in *E. coli* storage and donor strains can cause a variety of problems when attempting to perform transposon mutagenesis, especially when using a replicative plasmid. *Mariner* elements replicate via a cut and paste mechanism, which consequently could result in a transposon-less vector which clearly will not be useful when transferred to the target strain. Alternatively, the transposon could insert into any of the other components of the plasmid rendering it ineffective in several different ways. The Gram-positive

replicon could be disrupted leading to a lack of successful DNA transfer. The inducible P_{fac} promoter responsible for conditional plasmid stability could be affected, leading either to an inability to force plasmid loss, or a constantly active P_{fac} promoter if the repressor gene *lacI* is affected, again resulting in a lack of successful DNA transfer. Alternatively transposase activity could be abolished altogether via an insertion into the transposase. Unwanted insertions can therefore affect DNA transfer as well as the performance of the vector when harboured in the target strain.

5.1.6 P_{thl}-Tn-*codA* Strategy

The *codA* gene from *E. coli* encodes a cytosine deaminase responsible for deaminating cytosine to uracil as an integral part of the pyrimidine salvage pathway. It has also been used as a counter selection within eukaryotic and prokaryotic organisms as this enzyme converts 5-fluorocytosine (5FC) into the highly toxic 5-fluorouracil (5FU). Attempts to clone *codA* downstream of a strong promoter in this research group have been problematic (N. Minton, personal communication). This suggests that strong expression of *codA* in *E. coli* is deleterious to the cell even in the absence of 5FC. It was thought that the placement of a strong promoter upstream of *codA* with the *mariner* transposon in-between the two would produce a vector with enhanced stability (Figure 5.3). The *catP* mini-transposon possesses a terminator sequence which should prevent transcriptional read-through from the strong promoter (P_{thl}) to *codA*, protecting the cell from the deleterious effects of *codA* expression. The cut and paste mechanism of *mariner* elements is exploited here, as removal of the transposon will reform the deleterious conformation of P_{thl} – *codA*. Cells in which transposition events have occurred should therefore have significant fitness

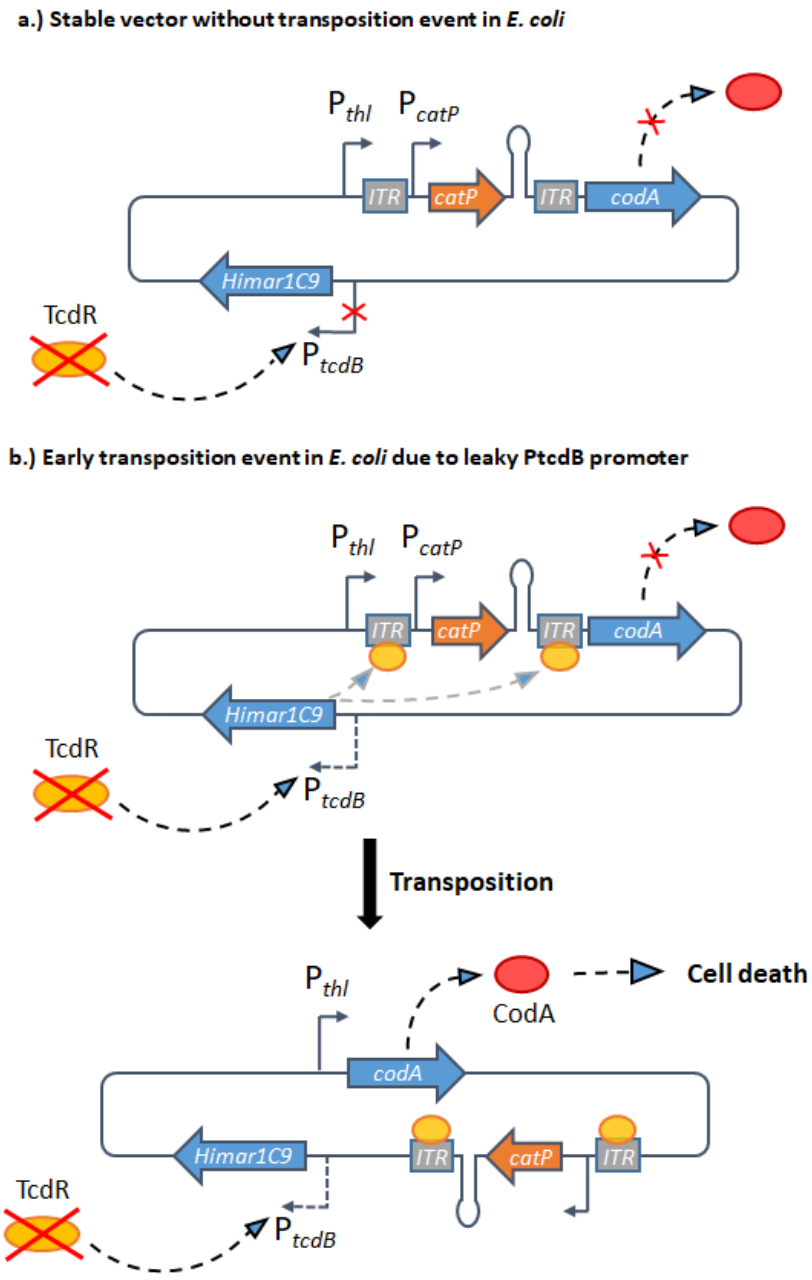


Figure 5.3: P_{thl}-Tn-*codA* stability strategy.

An overview of the P_{thl}-Tn-*codA* strategy. Expression of *codA* from the strong P_{thl} promoter is thought to be deleterious to *E. coli*. a.) The presence of the transposon leaves *codA* promoterless protecting the cell from the deleterious effects of the encoded cytosine deaminase. In addition, the *E. coli* strain will have no TcdR sigma factor to activate expression from P_{tcdB} meaning transposase concentrations are low. b.) A transposition event causes the *codA* to be downstream of the strong thiolase promoter P_{thl} since the *mariner* element replicates via a cut-and-paste mechanism. This conformation is thought to be deleterious to the *E. coli* and as such reduces the fitness of any strains in which the original, protecting position of the transposon has been lost. Vectors harbouring the transposon in the intended position are expected to out-compete any cells in which transposition has occurred, hence increasing the proportion of plasmid prepared from *E. coli* which has the intended transposon location.

reductions relative to those with the desired conformation with the transposon in the initial location, and therefore come to represent a lower proportion of the plasmids transferred to *C. autoethanogenum*. This will likely result in an increased conjugation efficiency and perhaps more robust performance when in the recipient strain.

5.2 Results

5.2.1 Suicide Vector pMTL-GL15

Suicide-delivery of the transposon is preferable over the use of a replicative vector, and therefore pMTL-GL15 was used to determine if use of a suicide vector was feasible for TIS library construction. pMTL-GL15 is identical to pMTL-YZ14 described in chapter 4 except for lacking a Gram-positive replicon. The DNA transfer efficiency gains observed by the use of the novel conjugal strain Sexpress detailed in chapter 3 made this approach worth attempting.

Transfer of pMTL-GL15 and a control replicative plasmid to *C. autoethanogenum* C24 proceeded in triplicate via conjugal transfer using the donor strain Sexpress. The control plasmid (pMTL83151) produced an average of 2,016 colonies on selection plates while conjugations using the suicide transposon delivery vector pMTL-GL15 produced an average of 15.5 colonies. Transconjugants of pMTL-GL15 were selected for on YTF agar plates supplemented with thiamphenicol to select for the *catP* marker internal to the transposon and ß-lactose to induce transposition via the chromosomally located bgaR-P_{bgaL}-*tcdR* expression system of the C24 strain. Transconjugants of pMTL-GL15 were streaked onto both clarithromycin and thiamphenicol and

found to have the expected phenotype of Clr^S Tm^R associated with transposon mutants. The ratio of transposon mutants to the control conjugation with a replicative plasmid was 0.007 which is in line with other studies (Lamberg et al., 2002).

Despite the increases in conjugation efficiency allowed by the use of Sexpress the use delivery of the transposon via this suicide plasmid produces too few colonies to be a feasible means of creating a mutant library for TIS studies. Further work in this text is reliant upon replicative plasmids with conditional replicons to deliver the transposon.

5.2.2 pMTL-CW14, Allowing Removal of Plasmid-based Reads

A novel transposon-delivery vector was created termed pMTL-CW14, which is identical to pMTL-YZ14 except for the addition of I-SceI recognition sequences flanking the transposon.

5.2.2.1 Construction of pMTL-CW14

pMTL-CW14 is identical to pMTL-YZ14 except for the addition of I-SceI sites 5'...tagggataacagggtaat... 3' flanking the transposon. HiFi assembly was used to construct this vector. Four DNA fragments were generated by PCR amplification with primers as shown in Table 5.1 using pMTL-YZ14 as a template. I-SceI sites were incorporated into the HiFi Assembly primers used to amplify the transposon-containing fragments. The final vector was created via HiFi assembly as detailed in materials and methods (2.5.9). The plasmid was verified by restriction digest with EcoRV and Sanger sequencing for the transposon and transposase regions.

Table 5.1: Primers and templates used for construction of pMTL-CW14.

pMTL-CW14 was constructed using HiFi assembly. This table presents the primers and relevant templates used to generate the DNA fragments with overlapping regions allowing the HiFi reaction to form a completed molecule.

PCR Template	Primers	Notes
pMTL-YZ14	109, 110	*ermB*, cole1+traJ
pMTL-YZ14	111,112	Adds I-SceI sites to transposon
pMTL-YZ14	113,114	Transposase, part of conditional replicon
pMTL-YZ14	115,116	Remainder of conditional replicon

5.2.2.2 Assessment of pMTL-CW14

pMTL-CW14 was transferred to *C. autoethanogenum* C24 via conjugal transfer. Initial selection plates were flooded with PBS to harvest all growth after 72 hours before being serially diluted and plated to agar plates with various supplementation so as to calculate the transposition efficiency (Tm, IPTG), number of plasmid-harbouring cells (Tm) and resistance to IPTG (Clr, IPTG). As expected, the vector performed similarly to pMTL-YZ14 by these metrics (Table 5.2).

Table 5.2: Assessment of pMTL-CW14.

Summary of the relevant values for assessment of pMTL-CW14 as a transposon-delivery vector. pMTL-CW14 was transferred to *C. autoethanogenum* C24 via conjugal transfer. The transconjugant selection plate was flooded with PBS and the entire cell matter harvested. The transconjugant cell mixture was serially diluted down to 10-7 and plated to a) Clr, b) Clr, IPTG, c) Tm IPTG, d) Tm and the respective CFU counts per ml were calculated. CFU counts were used to calculated transposition frequency ((Tm, IPTG – IPTG, Clr) ÷ Tm; Plasmid loss efficiency (Tm – Tm, IPTG) ÷ Tm × 100; Plasmid retention in presence of IPTG (Tm ÷ Tm, IPTG) and Mutant selection plate purity which is the calculated percentage of colonies of mutant selection plates (IPTG, Tm) expected to represent transposon mutants (Tm, IPTG – Clr, IPTG) ÷ Tm, IPTG. This value was also estimated by re-streaking putative transposon mutant colonies onto YTF agar plates supplemented with Clr and then Tm and the percentage of Clr^S Tm^R colonies calculated.

Transposition Frequency	Plasmid Loss Efficiency (%)	Plasmid Retention in Presence of IPTG	Mutant Selection Plate Purity	
			Replica plating	Calculated
6.52×10^{-5}	99.99999	16,799,755.80	96.5	99.9

5.2.2.3 Creation of a Transposon Library for TraDIS using pMTL-CW14 and Liquid Selection

Liquid selection involves direct inoculation of transconjugants from initial selective plates into liquid cultures in which transposon mutants are selected for. It is a relatively easy and rapid alternative to selecting for transposon mutants on agar plates. Avoiding an agar plate mutant selection stage may also allow less generations for early transposition event mutants to dominate, offering advantages for the number of unique reads in TIS results.

The newly created pMTL-CW14 plasmid was used to create two mutant pools. Transcriptional control is to some extent possible using the *C. autoethanogenum* C24 strain with inducible expression of *tcdR* on the chromosome. In addition the plasmid is conditionally stable with the addition of IPTG forcing plasmid loss. Finally, pMTL-CW14 has I-SceI recognition

sequences flanking the transposon allowing the elimination of plasmid-based reads at the DNA sequencing library stage.

Approximately 1,100 colonies from conjugation selection plates were harvested and used to inoculate 400ml YTF with 1mM IPTG, 10mM ß-lactose, thiamphenicol and D-cycloserine to an OD_{600} of 0.1. The optical density (OD_{600}) was monitored until the culture reached exponential phase after 72 hours at which point several glycerol stocks were created and the remaining culture was harvested for subsequent genomic DNA extractions.

The length of time taken to reach exponential phase in the creation of the first mutant pool created concern that relatively few mutants had dominated the pool. Immediate forcing of plasmid loss possibly selected against the vast majority of plasmid-containing cells before transposition occurred, this left a minority of cells which were already transposon mutants or that were resistant to the plasmid-destabilising effect of IPTG to dominate the pool. This created a long lag phase by effectively reducing the viable inoculum. It was thought worthwhile subsequently to allow a period of transposase induction before the forcing of plasmid loss. To achieve this, a second library was generated in a similar way with 3,500 transconjugant colonies used as inoculum to an OD_{600} of 0.1. In this case 1mM IPTG was added six hours after inoculation rather than immediately. In this case the culture reached exponential phase within 24 hours. CFU counts from the end point of this culture revealed that it consisted of ~90% plasmid-harbouring cells and 10% transposon mutants which represented approximately 110,000 mutants in total.

5.2.2.4 pMTL-CW14 in Liquid Selection Analysis

Generation of transposon libraries in liquid culture was remarkably easy to achieve experimentally, but produced far inferior ratios of transposon mutant to plasmid-harbouring cells according the CFU counts taken. The 18 hours between IPTG-induction forcing plasmid loss and the serial dilution and plating for the CFU counts was not sufficiently long for effective forcing plasmid loss at the population level. *C. autoethanogenum* has a generation time of around three hours in these conditions so the estimated six generations allowed for plasmid-loss was insufficient to attain the high (>99%) plasmid loss values calculated when performing an analogous experiment using IPTG on agar plates. Liquid selection therefore appears to be suitable as a means of rapidly creating large mutant pools for phenotypic screening, but less suited for the creation of libraries for high-throughput insertion-site sequencing.

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

5.2.3.1 *bgaR*-P_{bgaL} CAT Promoter Assay

A lactose-inducible promoter system (*bgaR*-P_{bgaL}) was assessed for suitability for controlling transposase expression. *C. perfringens* harbours a transcriptional regulator gene *bgaR* which is located divergently to *bgaL* which encodes a beta-galactosidase. *bgaR* has been shown to be a positive regulator of P_{bgaL} in *C. perfringens* (Hartman, Liu, & Melville, 2011), *C. acetobutylicum* (Al-Hinai et al., 2012) and *C. ljungdahlii* (Banerjee, Leang, Ueki, Nevin, & Lovley, 2014). Having previously been used to express *tcdR* within the *C. autoethanogenum* genome, it was thought useful to measure the strength and inducibility of this promoter alone and if appropriate, construct a new transposon-delivery vector based on it. A reporter plasmid pMTL82254::HZ1 which harbours *catP* under the control

of P_{bgaL} was used to measure the effectiveness of lactose induction. Promoter activity increases with the addition of lactose at least until 100mM and is significantly higher than the uninduced state, it is therefore an effective inducible promoter system (Figure 5.4).

5.2.3.2 Construction of pMTL-CW17

Having established that the bgaR-PbgaL promoter system is effective in *C. autoethanogenum*, a novel transposon-delivery vector was constructed based on the pMTL-YZ14 plasmid but with the *Himar1C9* transposase under the control of the lactose-inducible *bgaR*-P_{bgaL} promoter, instead of the orthogonally-controlled P_{tcdB}. The benefit of this vector is that it can be used in any unmodified strains in which the *bgaR*-P_{bgaL} promoter is active negating the need to insert *tcdR* onto the chromosome. The change in transposase promoter was hoped to allow pMTL-CW17 to produce an improved proportion of chromosomal transposon mutants on mutant selection plates over the orthogonal system. pMTL-CW17 is still reliant on transcriptional read-through into pCB102 for forced plasmid loss but since the transposon ratio is dependent

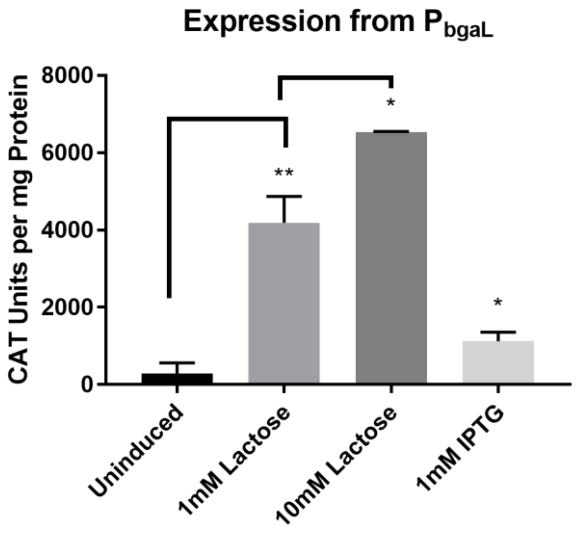


Figure 5.4: Lactose induction of P_{bgaL}

A reporter plasmid (pMTL82254::HZ1) containing the *catP* gene downstream of BgaR-PbgaL was conjugated into a wild type strain of *C. autoethanogenum* and three independent colonies were inoculated into YTF liquid culture. Inoculations from liquid culture were taken after 48 hours into YTF liquid with or without varying levels of ß-lactose, selecting for the presence of plasmid with thiamphenicol throughout. ß-lactose was added at 10mM or 100mM concentrations. The 1mM ß-lactose and 10mM ß-lactose induced cultures were both significantly different to the uninduced culture ** 0.0026 and **** <0.0001 respectively. There was a significant difference between 1mM and 10mM induction * (0.0401). Induction with IPTG also offered a significant increase over the uninduced culture * (0.0154), but expression was significantly lower than when ß-lactose was used in equimolar amounts ** (0.0046).

on transposition frequency as well as plasmid loss effectiveness it was hoped that the lactose-inducible promoter could increase the transposition frequency sufficiently to increase the ratio of transposon mutants to plasmid-harbouring cells.

To construct pMTL-CW17 primers '129_Tn_I-SceI_F' and 130_Tn_I-SceI_R' were used to amplify the transposon from pMTL-YZ14 whilst adding flanking I-SceI sites. This fragment was used as one of the templates in the subsequent HiFi assembly process which combined fragments generated using primers and templates as shown in Table 5.3.

Table 5.3: Primers and templates used for the construction of pMTL-CW17.

pMTL-CW17 was constructed using HiFi assembly. This table presents the primers and relevant templates used to generate the DNA fragments with overlapping regions allowing the HiFi reaction to form a completed molecule.

PCR Template	Primers	Notes
PCR product of primers 129 and 130 (step 1)	139, 140	Transposon
pMTL82254::HZ1	133, 134	Lactose inducible promoter
pMTL-YZ14	131, 132	Backbone
pMTL-YZ14	135, 136	Himar1C9
pMTL-YZ14	137, 138	Pfac-pCB102

5.2.3.3 Assessment of pMTL-CW17

Assessment of pMTL-CW17 proceeded using the same protocol as for pMTL-CW14 using wild type *C. autoethanogenum* instead of the *C. autoethanogenum* C24 strain. Briefly, transconjugants were selected for with clarithromycin using D-cycloserine to counter-select the *E. coli* donor strain. Selection plates were flooded with PBS and all cell matter was harvested, serially diluted and plated to agar plates with various supplements in order to calculate the relevant values for assessing the effectiveness of the vector. The plasmid loss efficiency was similar to that of pMTL-CW14 and pMTL-YZ14, but the transposition frequency was lower: 1.99×10^{6} as opposed to 1.21×10^{-5} (Table 5.4).

Table 5.4: Assessment of pMTL-CW17.

Summary of the relevant values for assessment of pMTL-CW17 as a transposon-delivery vector. pMTL-CW17 was transferred to *C. autoethanogenum* wild type via conjugal transfer. The transconjugant selection plate was flooded with PBS and the entire cell matter harvested. The transconjugant cell mixture was serially diluted down to 10-7 and plated to a) Clr, b) Clr, IPTG, c) Tm IPTG, d) Tm and the respective CFU counts per ml were calculated. CFU counts were used to calculated transposition frequency (Tm, IPTG – IPTG, Clr) ÷ Tm; Plasmid loss efficiency (Tm – Tm, IPTG) ÷ Tm × 100; Plasmid retention in presence of IPTG (Tm ÷ Tm, IPTG) and Mutant selection plate purity which is the calculated percentage of colonies of mutant selection plates (IPTG, Tm) expected to represent transposon mutants (Tm, IPTG – Clr, IPTG) ÷ Tm, IPTG.

Transposition Frequency	Plasmid Loss Efficiency (%)	Plasmid Retention in Presence of IPTG	Mutant Selection Plate Purity	
			Replica plating	Calculated
1.99×10^{-6}	99.99999	9,064,153	87.5	89.8

5.2.3.4 Induced vs. Uninduced Transposition Frequency

Having established pMTL-CW17 as being capable of producing transposon mutants with a similar efficiency to when using the orthogonal P_{tcdB} promoter, the difference between induced and uninduced states was investigated. With perfect transcriptional control there should be no transposon mutants produced if the cells are not exposed to lactose at any point during the procedure. Initially this was investigated by plating an uninduced liquid culture to mutant selection plates with or without lactose. There was no significant difference between the two conditions, each producing around 3,350 CFU per ml on mutant selection plates. It was however observed that colonies on mutant selection plates which lacked lactose were smaller and visible after 72 hours growth rather than 48 hours growth. This may be due to transposition occuring earlier on the lactose-containing plates than on the plates lacking inducer.

An alternative approach was then taken to assess the effect of lactose-induction. Transconjugants were selected for with clarithromycin and D-

cycloserine before single colonies were inoculated into either clarithromycin and D-cycloserine (uninduced) or clarithromycin, D-cycloserine and 10mM ß-lactose (induced). Three cultures for each state, either induced or uninduced, were inoculated from separate colonies. Seven hours were allowed for transposition to occur before cultures were serially diluted and plated to mutant selection plates (IPTG, thiamphenicol) to assess the number of transposon mutants. The induced cultures contained an average of 2,277 transposon mutants compared to only 1,390 for the uninduced cultures (Figure 5.5). The proportion of Tm^R Cla^S colonies on mutant selection plates was 87% as assessed via replica plating.

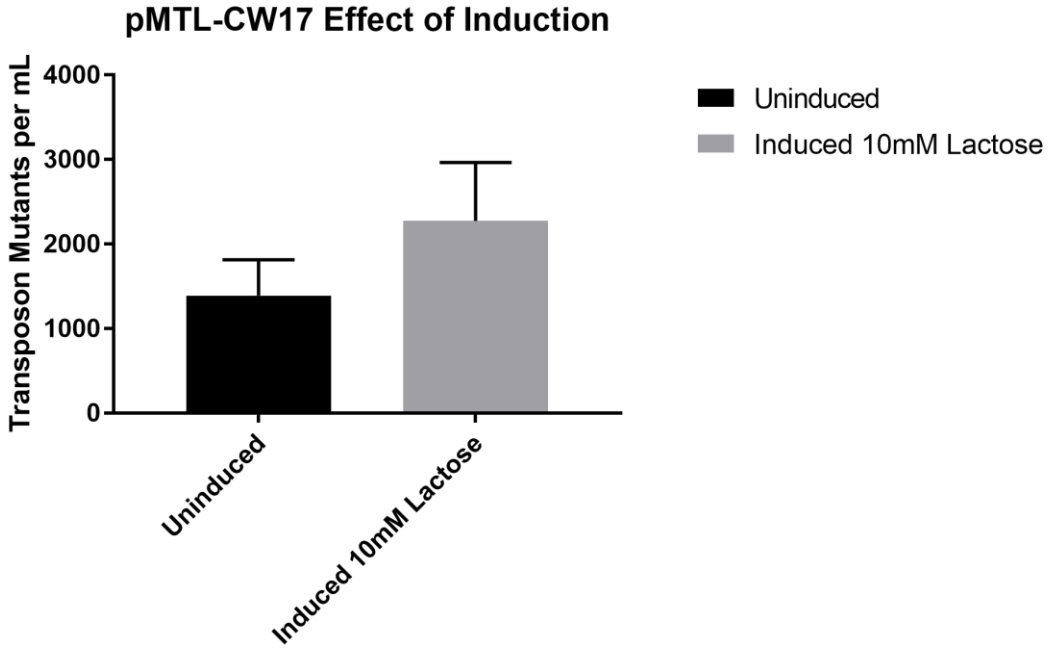


Figure 5.5: Effect of induction on pMTL-CW17 transposition frequency.

The transposition frequency of pMTL-CW17 was assessed with and without induction with 10mM ß-lactose. pMTL-CW17 was transferred to *C. autoethanogenum* wild type via conjugal transfer. Six transconjugant colonies were inoculated into YTF liquid culture supplemented with clarithromycin and D-cycloserine and for three of the six (induced cultures) with 10mM ß-lactose in addition. After seven hours cultures were serially diluted down to 10^{-7} in PBS and plated to mutant selection plates supplemented with

thiamphenicol and IPTG as well as thiamphenicol alone to estimate the number of plasmid-harbouring cells and Clr, IPTG to determine the number of plasmid-harbouring cells resistant to the effects of IPTG. Transposition frequency was calculated from the CFU counts: (Tm, IPTG – IPTG, Clr) ÷ Tm.

5.2.4 pMTL-CW20, Increased Plasmid Stability.

5.2.4.1 Construction of pMTL-CW20

pMTL-CW20 was constructed to investigate the effectiveness of the P_{thl}-Tn-*codA* conformation to select against unwanted transposition events in transposon delivery plasmids. HiFi assembly was used to construct the vector with templates and primers detailed in Table 5.5. Fragments were combined and treated with HiFi master mix as per the manufacturer's instructions. HiFi products were used to transform *E. coli* Top10 and plasmid DNA was verified from transformants using EcoRV diagnostic digests and sequencing of the transposon site and transposase.

Table 5.5: Primers and templates used for construction of pMTL-CW20.

pMTL-CW20 was constructed using HiFi assembly. This table presents the primers and relevant templates used to generate the DNA fragments with overlapping regions allowing the HiFi reaction to form a completed molecule.

Template	Primers	Notes
pMTL-CW14	162,163	Transposase PtcdB
pMTL82252	164,165	P_{thl}
pMTL-CW14	166,167	Tn part 1
pMTL-CW14	168,169	Tn part 2
pMTL-YZ17	170,171	*CodA*

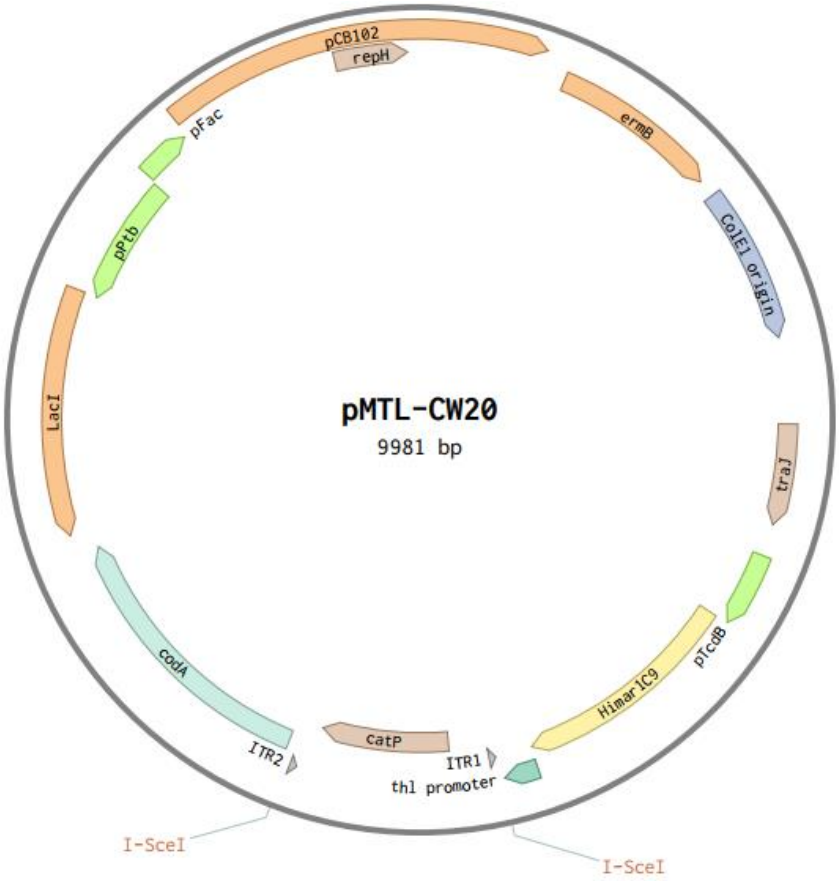


Figure 5.6: pMTL-CW20

A plasmid map of pMTL-CW20 which exhibits orthogonal expression of the himar1C9 transposase via P_{tcdB} and a conditional Gram-positive replicon, Pfac-pCB102. The vector is similar to pMTL-YZ14 as documented in chapter 4, but contains I-SceI recognition sites flanking the transposon and has the conformation Pthl-Tn-codA to encourage plasmid stability.

5.2.4.2 Evidence of Instability

Transposon-delivery plasmid instability was suggested to be one contributory reason for an inability to conjugate pMTL-YZ14 into *C. autoethanogenum* since similar plasmids using the pCB102 replicon including plasmids with the P_{fac}-pCB102 region were able to be transferred relatively easily. To investigate this possibility a verified pMTL-YZ14 plasmid was used to transform *E. coli* Top10 and transformant colonies used to inoculate overnight cultures. Plasmids prepared from overnight cultures were extracted and treated with SbfI.

Movement of the transposon into various other parts of the vector was frequently found to have occurred as illustrated in Figure 5.7 with only one of the six plasmids exhibiting the expected band pattern. An analogous procedure was later followed using pMTL-CW20 instead of pMTL-YZ14. In this case all six plasmids exhibited the expected band pattern (Figure 5.8).

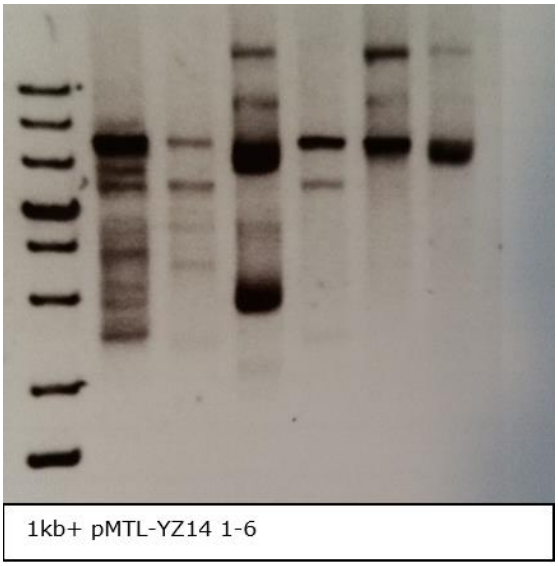


Figure 5.7: Diagnostic digest of pMTL-YZ14.

A diagnostic digest of six pMTL-YZ14 plasmids derived from independent colonies via an overnight culture. Plasmid DNA was treated with SbfI. The expected band pattern is observed in only one of six (lane 5). Lane 1 contains 1kb+ ladder, lanes 2-7 contain digested plasmid samples.

5.2.4.3 Evidence of Increased Stability

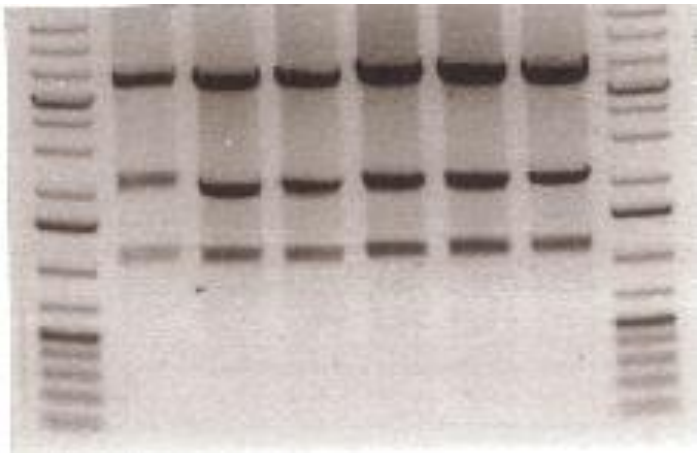


Figure 5.8: Diagnostic digest of pMTL-CW20.

A diagnostic digest of six pMTL-CW20 plasmids derived from independent colonies via an overnight culture. Plasmid DNA was treated with EcoRV which has three recognition sequences within pMTL-CW20 including one internal to the transposon. The expected band pattern is observed in all six samples with no extra bands observed (lanes 2-7). Lanes 1 and 8 contain 1kb+ ladder.

5.2.4.4 Assessment of pMTL-CW20

pMTL-CW20 appears to have increased stability in *E. coli*, but it was thought possible that the increased stability is merely due to the lower transposition rate which would render the plasmid undesirable. This apparent stability is likely to be the result of a selection against transposition events rather than an increase in stability per se. The transposition frequency of pMTL-CW20 was assessed to ensure the novel conformation which appears to reduce plasmid instability in *E. coli* does not significantly impair the ability of the vector to produce transposon mutants in clostridia. pMTL-CW20 was transferred to *C. autoethanogenum* via conjugal transfer from Sexpress and selected for on YTF agar plates supplemented with clarithromycin and D-cycloserine. Colonies were harvested from selection plates by flooding with PBS and the entire cell matter was serially

diluted and spread onto YTF agar plates supplemented as previously to determine the pertinent values to assess the vector. Colonies appearing on mutant selection plates were streaked onto clarithromycin and thiamphenicol plates separately to estimate the percentage of transposon mutants to plasmid-harbouring cells. The transposition frequency was calculated as 1.90×10^{-4} which is comparable to other transposition frequencies in this study .

Table 5.6: Assessment of pMTL-CW20

Summary of the relevant values for assessment of pMTL-CW20 as a transposon-delivery vector. pMTL-CW20 was transferred to *C. autoethanogenum* C24 via conjugal transfer. The transconjugant selection plate was flooded with PBS and the entire cell matter harvested. The transconjugant cell mixture was serially diluted down to 10-7 and plated to a) Clr, b) Clr, IPTG, c) Tm IPTG, d) Tm and the respective CFU counts per ml were calculated. CFU counts were used to calculated transposition frequency ((Tm, IPTG – IPTG, Clr) ÷ Tm; Plasmid loss efficiency (Tm – Tm, IPTG) ÷ Tm × 100; Plasmid retention in presence of IPTG (Tm ÷ Tm, IPTG) and Mutant selection plate purity which is the calculated percentage of colonies of mutant selection plates (IPTG, Tm) expected to represent transposon mutants (Tm, IPTG – Clr, IPTG) ÷ Tm, IPTG.

Transposition Frequency	Plasmid Loss Efficiency (%)	Plasmid Retention in Presence of IPTG	Mutant Selection Plate Purity	
			Replica plating	Calculated
1.90×10^{-4}	99.99999	31,202,501	94.2	99.9

5.2.4.5 Effect of 5FC on Transposition Rate

Having established that the ability of the plasmid to create transposon mutants was not impaired, the effect of the presence of 5FC on mating and selection plates was established. The expression of *codA* from a strong promoter on a high copy-number plasmid may be expected to increase the susceptibility of *C. autoethanogenum* to 5FC. This increased susceptibility could be advantageous

and negate the need to create a *codA* knockout strain. Growth of wild type *C. autoethanogenum* is inhibited at 50µg/ml 5FC (Liew, 2015), and it was thought that a sub-lethal concentration could select against P_{thl} – *codA* as this conformation could lower the lethal concentration of 5FC. The transposition frequency of pMTL-CW20 is lower than that of pMTL-CW14 in the presence of 5FC, but the difference was not statistically significant (Figure 5.9).

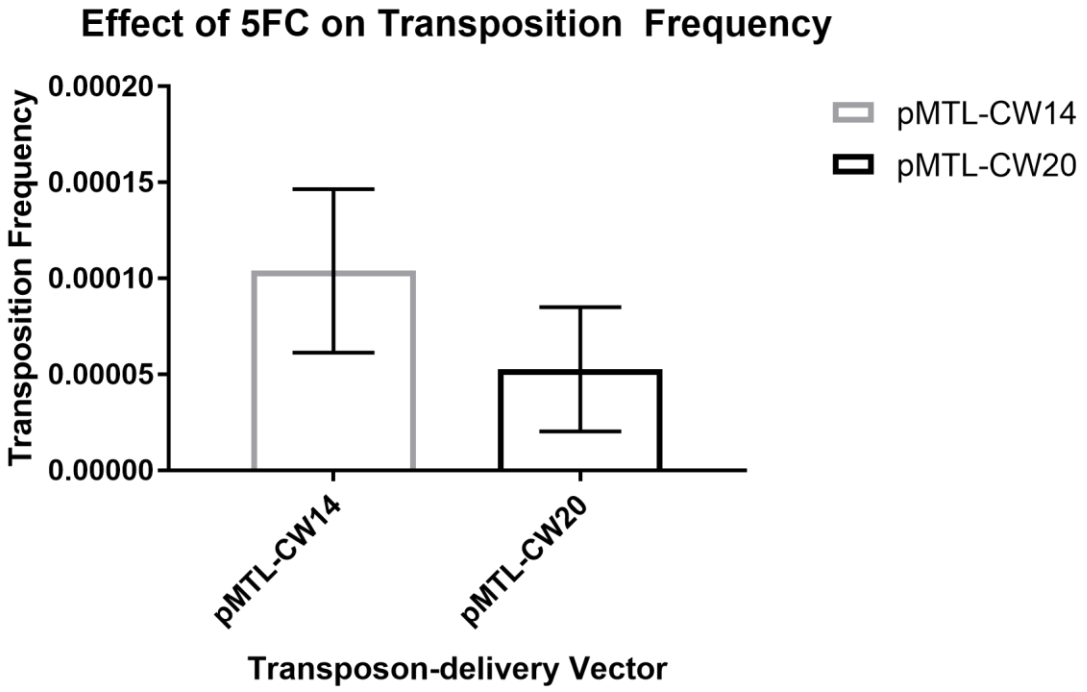


Figure 5.9: Effect of 5FC on pMTL-CW20 Transposition Frequency.

pMTL-CW20 and pMTL-CW14 were transferred to *C. autoethanogenum* via conjugal transfer. The transconjugant selection plate was flooded with PBS and the entire cell matter harvested. The transconjugant cell mixture was serially diluted down to 10-7 and plated to a) Clr, b) Clr, IPTG, c) Tm IPTG, d) Tm and the respective CFU counts per ml were calculated. CFU counts were used to calculated transposition frequency ((Tm, IPTG – IPTG, Clr) ÷ Tm; Plasmid loss efficiency (Tm – Tm, IPTG) ÷ Tm × 100; Plasmid retention in presence of IPTG (Tm ÷ Tm, IPTG) and Mutant selection plate purity which is the calculated percentage of colonies of mutant selection plates (IPTG, Tm) expected to represent transposon mutants (Tm, IPTG – Clr, IPTG) ÷ Tm, IPTG. 5FC (32.5µg/ml) was added to conjugation mating and selection plates as well as plates for CFU counts.

5.2.5 pMTL-CW22, A Tetracycline-inducible Transposon-delivery Plasmid

Recently, simultaneous transposase expression and forced plasmid-loss were demonstrated in *C. difficile* using divergently orientated tetracycline-inducible promoters (Dembek et al., 2015). This transposon delivery system was used to produce a library on which TraDIS was performed, identifying 404 core essential genes in *C. difficile* R20291 (Dembek et al., 2015). It was thought that this promoter system could offer improvements in terms of a higher proportion of transposon mutants on mutant selection plates via more efficient plasmid-loss or a higher transposase activity. To assess the system a novel vector was created based the divergent promoter system, Gram-positive replicon and transposase described as well as the P_{thl}-Tn–*codA* system previously demonstrated to increase plasmid stability in *E. coli* and with I-SceI recognition sequences flanking the mini(catP) transposon to eliminate plasmid reads at the DNA sequencing library stage.

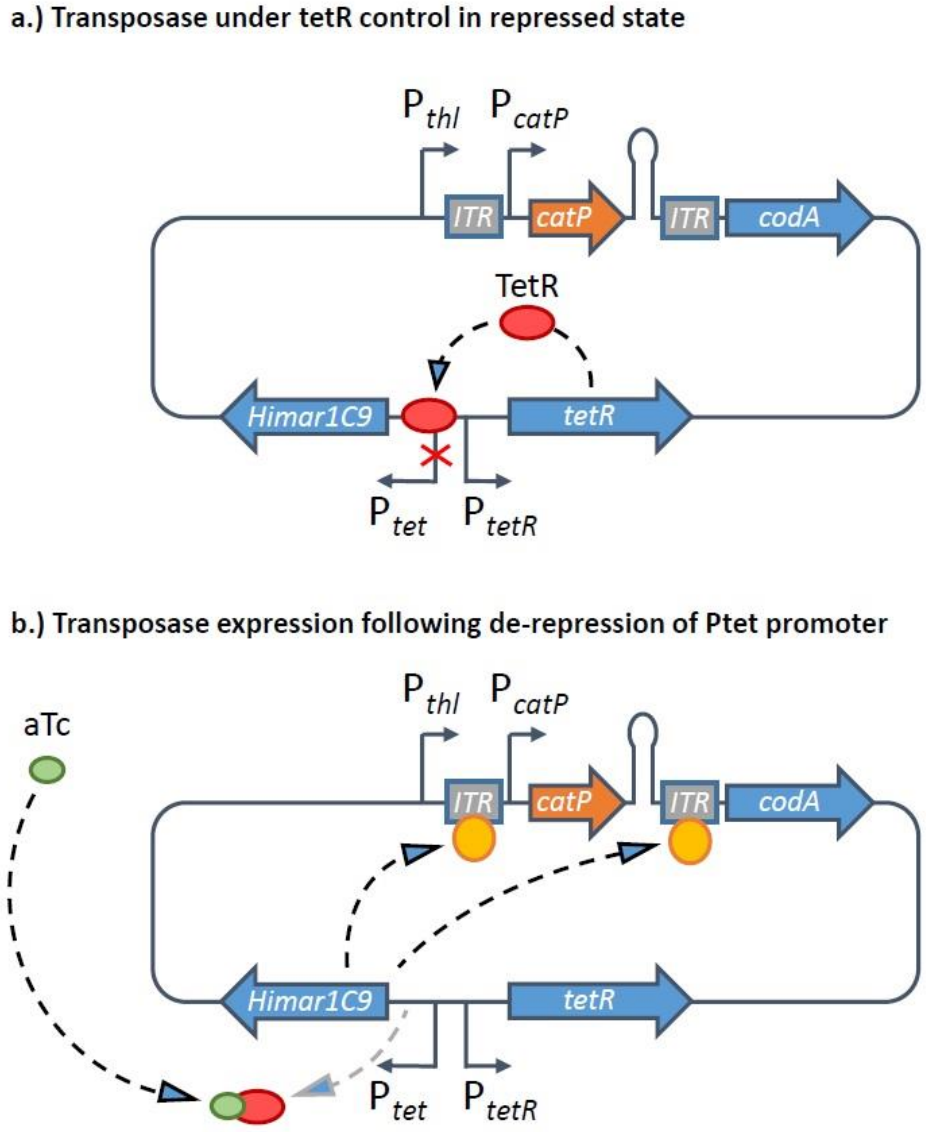


Figure 5.10: Tetracycline inducible transposon-delivery plasmids.

A genetic schematic of the tetracycline-inducible transposon-delivery plasmids created in this text. The *himar1C9* transposase is controlled by the tetracycline-inducible promoter P_{tet} which is inhibited by the binding of TetR. TetR is expressed from the divergent promoter P_{tetR} which is also inhibited by TetR. In the presence of tetracycline or the analog anhydrotetracycline (aTc) TetR is released from promoter binding and both P_{tet} and P_{tetR} are de-repressed. The addition of aTc therefore causes transposase expression and an increased expression of the TetR protein. Importantly, transcriptional read-through from P_{tetR} also causes plasmid instability via transcription into the Gram-positive origin of replication. The Gram-positive origin of replication is positioned downstream of *tetR* and is the only source of variation between pMTL-CW21 (pCB102), pMTL-CW22 (pCD6) and pMTL-CW23 (p19). The P_{thl}-Tn-*codA* module which confers plasmid stability in *E. coli* donor and storage strains is also present on the tetracycline-inducible transposon-delivery plasmids as well as I-SceI recognition sites flanking the transposon.

5.2.5.1 Construction of pMTL-CW22

pMTL-CW22 was cloned using NEB HiFi assembly, the necessary fragments were generated via PCR using the templates and primers as shown in Table 5.7. Transformants were selected for on LB agar supplemented with chloramphenicol and erythromycin and inoculated to overnight cultures with the same supplementation. Plasmid DNA was prepared from overnight cultures and verified with diagnostic digests and Sanger sequencing.

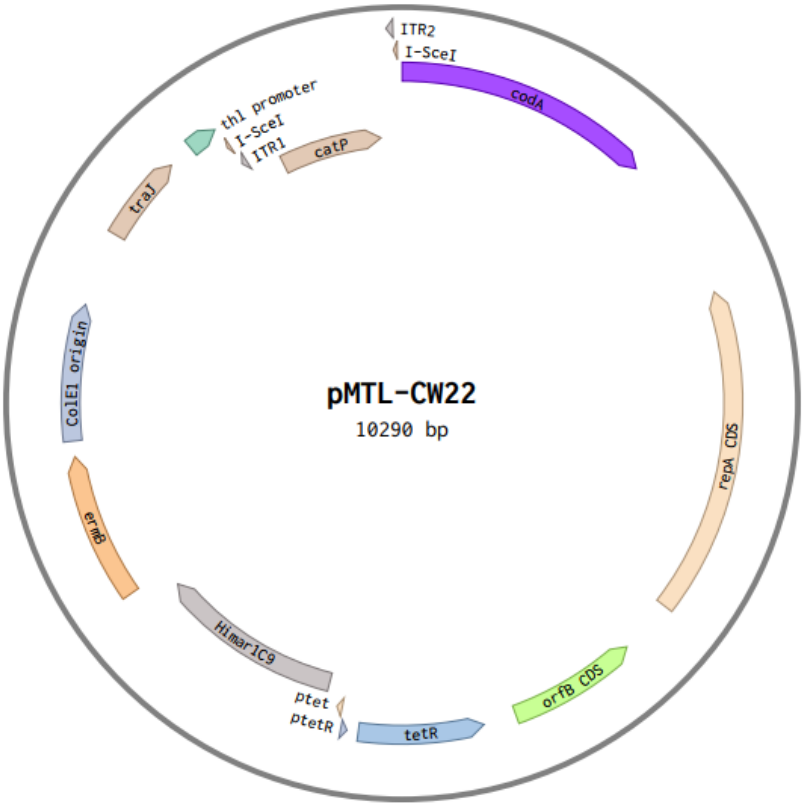


Figure 5.11: pMTL-CW22

A plasmid map of pMTL-CW22. The divergent tetracycline promoters control the transposase and the TetR protein with transcriptional read-through also causing plasmid instability. I-SceI recognition sequences flank the transposon as in pMTL-CW20 and the conformation P_{thl}-Tn-*codA* is present for increased plasmid stability in *E. coli* storage and donor strains.

Table 5.7: Primers and templates used for construction of pMTL-CW22.

pMTL-CW22 was constructed using HiFi assembly. This table presents the primers and relevant templates used to generate the DNA fragments with overlapping regions allowing the HiFi reaction to form a completed molecule.

Template	Primers	Description
pMTL_Tet3nO_RFtetR (catP assay)	172,173	Promoters
pMTL-CW20	174,175	Transposase
pMTL83151	176,177	Backbone
CW20	178,179	Transposon and *codA*
pMTL84151	180,181	pCD6

5.2.5.2 Assessment of pMTL-CW22 in *C. autoethanogenum*

Transposon mutant production was performed as for assessment of the other vectors in this chapter (Table 5.8) but mutant selection plates were supplemented with anhydrotetracycline (aTc) (100ng/ml) rather than IPTG as the non-permissive condition for plasmid replication.

Table 5.8: Assessment of pMTL-CW22.

Summary of the relevant values for assessment of pMTL-CW22 as a transposon-delivery vector. Anhydrotetracycline (aTc) (100ng/ml) was used to induce transposition and plasmid-loss simultaneously. pMTL-CW22 was transferred to C. autoethanogenum wild type via conjugal transfer. The transconjugant selection plate was flooded with PBS and the entire cell matter harvested. The transconjugant cell mixture was serially diluted down to 10-7 and plated to a) Clr, b) Clr, aTc, c) Tm aTc, d) Tm and the respective CFU counts per ml were calculated. CFU counts were used to calculated transposition frequency ((Tm, aTc – Clr, aTc) ÷ Tm; Plasmid loss efficiency (Tm - Tm, aTc) ÷ Tm × 100; Plasmid retention in presence of aTc (Tm ÷ Tm, aTc) and Mutant selection plate purity which is the calculated percentage of colonies of mutant selection plates (Tm, aTc) expected to represent transposon mutants (Tm, aTc - Clr, aTc) ÷ Tm, aTc.

Transposition Frequency	Plasmid Loss Efficiency (%)	Plasmid Retention in Presence of aTc	Mutant Selection Plate Purity	
			Replica plating	Calculated
0	99.99477	25,277.78	0	0

No transposon mutants were observed using pMTL-CW22 (Table 5.8), this effect was thought to be due either to a failure of the P_{tet} promoter in *C. autoethanogenum* or the impact of the pCD6 replicon since previous transposon-delivery plasmids were all based on the pCB102 origin of replication. Transcriptional read-through from P_{tet} through *tetR* into pCD6 appears to cause plasmid instability in *C. autoethanogenum* as predicted, though with a lower efficiency than the P_{fac}-pCB102 conditional replicon.

5.2.6 pMTL-CW21

An alternative version of pMTL-CW22 was constructed by J. Millard in this laboratory called pMTL-CW21 which is identical to pMTL-CW22 except for the identity of the Gram-positive replicon, pMTL-CW21 contains pCB102 instead of the pCD6. Since pMTL-CW22 proved unable to produce transposon mutants in *C. autoethanogenum* it was thought worthwhile to investigate if changing the Gram-positive replicon would produce a significant effect.

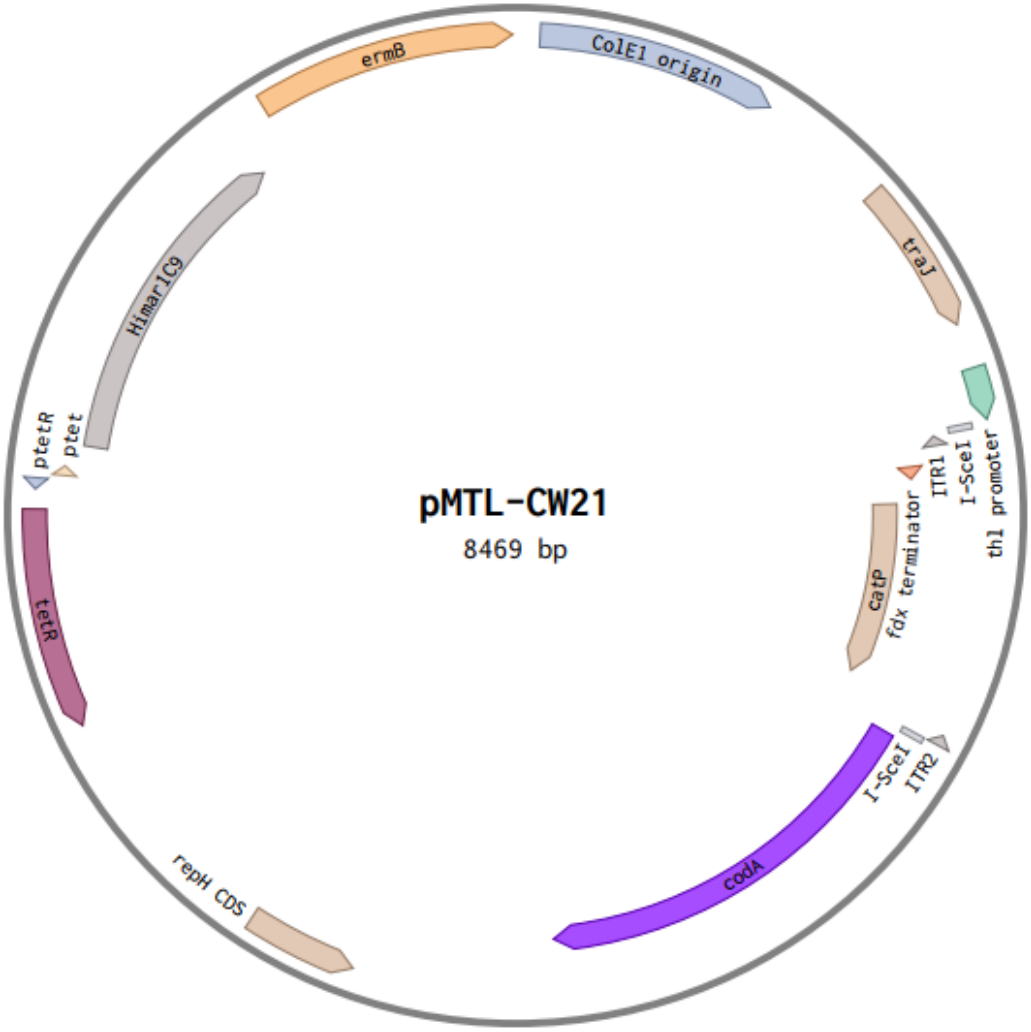


Figure 5.12: pMTL-CW21.

A plasmid map of pMTL-CW22. The divergent tetracycline promoters control the transposase and the TetR protein with transcriptional read-through also causing plasmid instability. I-SceI recognition sequences flank the transposon as in pMTL-CW20 and the conformation P_{thl}-Tn-*codA* is present for increased plasmid stability in *E. coli* storage and donor strains.

pMTL-CW21 was transferred to *C. autoethanogenum* and the mutant generation procedure was carried out as for pMTL-CW22. pMTL-CW21 produced a similar transposition frequency to that observed for pMTL-YZ14, pMTL-CW14 and pMTL-CW20, but the plasmid loss efficiency was lower (Table 5.9).

Table 5.9: Assessment of pMTL-CW21

Summary of the relevant values for assessment of pMTL-CW21 as a transposon-delivery vector. Anhydrotetracycline (aTc) (100ng/ml) was used to induce transposition and plasmid-loss simultaneously. pMTL-CW21 was transferred to C. autoethanogenum wild type via conjugal transfer. The transconjugant selection plate was flooded with PBS and the entire cell matter harvested. The transconjugant cell mixture was serially diluted down to 10-7 and plated to a) Clr, b) Clr, aTc, c) Tm aTc, d) Tm and the respective CFU counts per ml were calculated. CFU counts were used to calculated transposition frequency ((Tm, aTc – Clr, aTc) ÷ Tm; Plasmid loss efficiency (Tm - Tm, aTc) ÷ Tm × 100; Plasmid retention in presence of aTc (Tm ÷ Tm, aTc) and Mutant selection plate purity which is the calculated percentage of colonies of mutant selection plates (Tm, aTc) expected to represent transposon mutants (Tm, aTc - Clr, aTc) ÷ Tm, aTc.

Transposition Frequency	Plasmid Loss Efficiency (%)	Plasmid Retention in Presence of aTc	Mutant Selection Plate Purity	
			Replica plating	Calculated
1.75×10^{-4}	99.99997	9,507,389	97.5	99.84

5.2.6.1 Tetracycline-inducible Promoters in *E. limosum*

pMTL-CW23 is another variant of pMTL-CW22 with a different Gram-positive replicon. In pMTL-CW23 pCD6 is replaced by the origin of replication native to *Clostridium carboxidivorans* termed p19. pMTL-CW21 and pMTL-CW23 were used to transform *E. limosum* and assessed in a similar way to that of *C. autoethanogenum* described previously. All experiments were conducted by J. Millard.

Table 5.10: Assessment of pMTL-CW21 and pMTL-CW23 in *E. limosum*.

Calculated transposition frequency, plasmid loss efficiency and plasmid retention in the presence of the non-permissive aTc in *Eubacterium limosum* according to J. Millard (unpublished). Two plasmid-delivery vectors were assessed based on pMTL-CW22, but with the pCD6 replicon replaced by either pCB102 (pMTL-CW21) or p19 (pMTL-CW23).

Vector	Transposition Frequency	Plasmid Loss Efficiency (%)	Plasmid Retention in Presence of aTc
pMTL-CW21	7.47×10^{-3}	0.99996	23,950
pMTL-CW23	1.35×10^{-1}	0.99935	1,539

5.2.7 Transposon-delivery Vectors in *C. autoethanogenum*

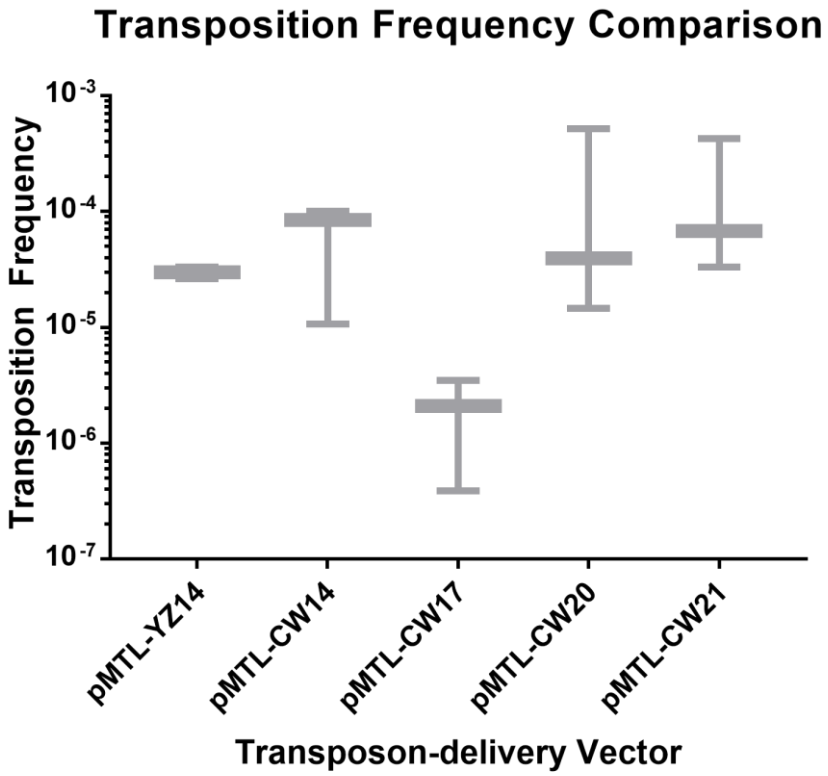


Figure 5.13: Transposition frequency comparison across transposon-delivery vectors.

A summary of the transposition frequencies calculated for the vectors assessed in this chapter. Data for individual vectors can be found in the corresponding vector assessment table.

5.3 Discussion

5.3.1.1 Overview

Two innovations have been made to existing transposon vectors which make TIS studies more practical. The addition of I-SceI recognition sequences flanking the transposon will allow an attempt to remove plasmid read contamination and the conformation of the strong thiolase promoter 5' of the mini transposon with *codA* downstream creates a more stable delivery vector. Two alternative promoters P_{tet} and P_{bgaL} driving transposase expression have been found to be effective in *C. autoethanogenum* with transposase activity comparable to P_{tcdB}. A conditional vector system (Y. Zhang et al., 2015) has been demonstrated in *C. autoethanogenum* and transcriptional read-through has also been shown to be effective at forcing plasmid loss using three different Gram-positive replicons – pCB102 and p19 in several different organisms – *C. autoethanogenum, E. limosum* and *C. sporogenes*.

5.3.1.2 Transposition Frequency and Inducible Promoters

The average transposition frequency does not vary significantly between any of the transposon-delivery vectors assessed in this study (Figure 5.13). The average for pMTL-CW17 appears to be lower, perhaps due to a relatively inactive P_{bgaL} promoter. The activity of P_{bgaL} was measured with a CAT assay. Figure 5.4 shows the effectiveness of the system with peak activity at 10mM ß-lactose and a peak induction-repression ratio of 18 was calculated. Higher activity levels may be achieved with a higher concentration of inducer. A further promoter assay could be performed assessing the three promoter systems employed P_{tcdB}, P_{bgaL} and P_{tet} simultaneously, in order to compare activity levels

between the three promoters. The original paper constructing and characterising the P_{bgaL} promoter system attempted induction using IPTG but reported that no increase in expression was observed, it was suggested that this was either because of non-responsiveness to IPTG or due to IPTG not being able to pass into the cell (Hartman et al., 2011). As can be seen in Figure 5.4, IPTG does appear to induce the system in *C. autoethanogenum* where IPTG is known to be an effective inducer of the P_{fac} promoter, though when at equimolar concentrations with lactose it proves to only induce around one third as effectively. Using an inducible promoter rather than the orthogonal expression system offers the advantage of being able to mutagenise wild type strains without prior chromosomal insertion of *tcdR*. This was recently exemplified in *C. acetobutylicum* using a xylose-inducible promoter system which produced mutant pools of which 91.9% contained no plasmid after three passages (Zhang et al., 2016). The 91.9% figure is analogous to the percentage calculated in this text for the percentage of colonies on mutant selection plates which represent genomic insertions without harbouring plasmid. Every vector assessed in this chapter produced a superior value to this 91.9% except for pMTL-CW22 which produced no transposon mutants. This superiority is unsurprising given the lack of a conditional replicon in the pMTL-YG3 transposon-delivery vector which makes the triple passaging unavoidable and still produces pools with nearly 10% plasmid-harbouring cells. In addition the three passages will inevitably make extremely biased transposon pools which would be inappropriate for TIS. The percentage of the pool which are transposon mutants and not plasmid-harbouring cells was assessed similarly in the referenced text and in this chapter, both used replica plating (for direct

measurement) and CFU counts on Tm compared to erythromycin to calculate the percentage. Both have their advantages, the primary advantage of the replica plating approach is that it is easy to understand. The advantage of using CFU counts to calculate this value is that it is hugely more accurate. If the replica plating (patch plating) method result was quoted in the referenced paper then a value of 100% transposon mutant identity would have been reported since 69 colonies were patch plated and none of them were Erm^R, but instead the 91.9% was reported (Zhang et al., 2016). The percentage transposon mutant value was usually reported with both methods in this chapter, but in this case the reported value was consistently lower when reporting replica plating results.

5.3.1.3 Vectors in other species

The novel transposon-delivery vectors whose construction is detailed in this text have been used to perform transposon mutagenesis in several related organisms including *C. sporogenes*, *C. difficile*, *C. ljungdahlii* and *E. limosum* by several other researchers in this institute. This cross-species applicability is not surprising since the genetic systems on which successful transposon mutagenesis relies have been previously shown to function in a variety of species. As more data emerges it may be interesting and instructive to compare the various values relevant to mutagenesis such as the transposition frequency and plasmid-loss efficiency. One promising result from spreading the novel vectors has been a consistent acknowledgement of an increase in plasmid stability in plasmids which harbour the P_{thl}-Tn-*codA* system. However, by far the most relevant result for assessing plasmid stability was gained from the TraDIS experiment detailed in the following chapter.

5.3.1.4 Plasmid Copy Number, Improved Conditional Replicons.

It was hoped that altering the transposase promoter or conditional plasmid-loss mechanism would create a transposon-delivery vector which would allow almost total elimination of plasmid-harbouring cells on mutant selection plates and be capable of inhibiting transposition until the mutant selection plating stage. With both these attributes a vector would be capable of producing a complex mutant library with a direct relationship between the number of colonies on mutant selection plates and the number of unique insertion events as opposed to the same insertion event representing a large fraction of mutant colonies.

None of the vectors tested provide confidence that early transposition event mutant read-hogging nor plasmid read-contamination problems will be avoided since none offer substantially better results than pMTL-YZ14. However, this analysis does ignore the impact of copy-number, it may be that pCD6 or p19 sustain a lower copy number and hence plasmid read-contamination would not be as detrimental with these vectors. The p19 replicon is of particular interest in this respect since it has been shown to have a significantly lower copy number than pCB102 in *C. beijerinckii* and *E. limosum* and a copy number of just one when measured in the native strain *C. carboxidivorans* (J. Millard, personal communication) (Table 5.11).

Table 5.11: Plasmid copy number of p19 and pCB102.

Plasmid Copy Number (PCN) for p19 and pCB102 in three species. Data provided by J. Millard (unpublished).

Replicon	*C. beijerinckii* PCN	*E. limosum* PCN	*C. carboxidivorans* PCN
p19	5-15	5-10	1
pCB102	60+	15-25	Unknown

If this low copy number is also the case in *C. autoethanogenum* then this replicon could offer a substantial advantage. It is unfortunate that of the three replicons tested, p19 exhibited the least effective forcing of plasmid loss in the presence of aTc. The architecture of the p19 origin of replication is relevant here, it appears to contain a single replication protein which in the conditional system tested is transcribed, though transcription must proceed from P_{tetR} through ~1kb of DNA before reaching the replication protein. Assuming that overexpression of the p19 replication protein is the mechanism of forced plasmid loss it stands to reason that a more proximal positioning of an inducible promoter could increase plasmid-loss efficiency. Optimisation of a conditional replicon based on p19 therefore seems potentially fruitful for future replicative vector transposon mutagenesis with a view to performing TIS experiments.

The lack of ability to remove all plasmid-harbouring cells from the transposon library demonstrated by these vectors is intended to be remedied by the addition of I-SceI sites to the vectors. Without this strategy a more thorough attempt would have been made to optimise plasmid-loss efficiency. There are a number of potential routes to this including experimentation with the placement of the inducible promoter within the origin of replication, and counter-selection of a plasmid-based marker either with the addition of a toxic-precursor compound or using inducible CRISPR targeting of the plasmid backbone. In origins of replication with obvious replication protein encoding genes there is potential for replacing the native promoter with an inducible promoter. This would consequently make the permissive condition the presence of inducer, as has been demonstrated (Hare et al., 2001), where the native RNA II transcript promoter of ColE1 was replaced with P_{lac}. The pCB102 and pCD6 replicons are

less well understood than ColE1 and it is difficult to predict a single useful site to incorporate an inducible plasmid to pursue this strategy. For p19 the replication protein can be predicted with greater confidence, and exchange of the native promoter with an inducible promoter active in clostridial species could be achieved relatively easily. If counter-selection of the plasmid is to be attempted then the vectors created in this study are already appropriate. The incorporation of I-SceI sites into the vectors offers the possibility of counter-selection by the expression of the I-SceI endonuclease encoded by *sceC*. I-SceI counter-selection has been demonstrated in *C. acetobutylicum* and *C. beijerinckii* (Zhang et al., 2015). If *sceC* were under inducible control then mutant selection plates could contain *sceC* inducer and the inducer for transcription into the replicon (IPTG or aTc) adding a secondary selective mechanism in an attempt to remove all plasmid-harbouring cells.

Limitations of the I-SceI Strategy

The benefit of the I-SceI site strategy for removal of plasmid DNA during sequencing library preparation is that a transposon library with a fraction of plasmid-harbouring cells can be tolerated in TraDIS without a domination of plasmid-based reads reducing library saturation. The effect of an I-SceI digest in the sequencing library preparation is shown in chapter 6. While this approach may be effective at increasing the number of unique chromosomal reads obtained it does nothing to prevent the problem of having plasmid-containing cells within the transposon pool. This is particularly problematic in experiments aiming to quantify fitness contribution in various conditions. To do this it is preferable to subject an initial transposon pool to two conditions and track the insertions into each gene comparing the two conditions. In the event of plasmid-

harbouring cells still being present, any of these cells can host a transposition event creating a new transposon mutant that was not present in the initial library. This could cause a newly emerging mutant to seem to have survived the growth condition when in fact the mutant could be non-viable in those conditions. For this reason an improved conditional replicon is still to be desired for use in intractable clostridial species TIS studies.

5.3.1.5 Limitations of I-SceI Digest Strategy, Alternative Strategies.

Elimination of plasmid read-contamination may be overcome at the sequencing library preparation stage, but the domination of early transposition event mutants is a problem which must be overcome during the biological mutant library production. An inducible promoter which is sufficiently tightly supressed until the mutant selection stage could overcome this problem. None of the three promoters tested in this chapter fit this requirement as transposition was still abundant in the uninduced conditions for P_{tcdB} (in an inducible C24 strain), P_{bgaL} and for P_{tet}. Inevitably inducible systems with superior repression in the uninduced state will be developed appropriate to this organism. A sensible step for future work would be to insulate the transposase from transcriptional read-through from upstream promoter elements using a transcriptional terminator upstream of the inducible promoter. Another strategy may be to produce antisense RNA to dampen expression of transposase. Effective repression must also be balanced with sufficiently high expression in the induced state to retain the ease of mutant production.

An alternative approach to using an effective inducible promoter system is offered with the P_{thl}-Tn–*codA* system. This system has proven useful in increasing transposon-delivery vector stability in *E. coli* storage and donor

strains. The functionality may be extended to ameliorate the read-hogging of early transposition event mutants by rendering transposition events lethal to the host cell prior to the mutant selection plate stage.

C. autoethanogenum is sensitive to 5FC at a concentration of 50mg/ml, likely due to the presence of a native cytosine deaminase. A knockout strain of *C. autoethanogenum* would likely exhibit an increased resistance to 5FC, this strain could then serve as the base strain for transposon mutagenesis with a transposon plasmid harbouring the P_{thl}-Tn–*codA* system. This strain and plasmid would allow selection against transposition events during the mating and selection plate stages via the addition of 5FC. Any transposition events would reform P_{thl}-*codA* which should be lethal in the presence of 5FC. Any early transposition events should in this case be lethal. Removal of 5FC for the mutation selection stage would allow transposition to occur unhindered and since no pre-existing transposon mutants would be plated, any colonies appearing would be the result of unique transposition events. This strategy could serve to restore the link between CFU on mutant selection and the number of unique transposon mutants in the library. No Δ*codA* strain was created in this study, however the feasibility of this approach was suggested in the experiment comparing transposition frequency with or without the presence of a sub-lethal concentration of 5FC. In this experiment 5FC was effective at reducing the transposition rate of pMTL-CW20 in the presence of 5FC without affecting that of pMTL-CW14 which lacks the P_{thl}-Tn-*codA* module suggesting that the susceptibility to 5FC was increased in cells harbouring plasmids which had lost their transposon and so placed P_{thl} ahead of *codA*. If this strategy is to

be pursued further, the susceptibility of *C. autoethanogenum* harbouring a plasmid containing P_{thl}-*codA* to 5FC should be compared to wild type.

5.4 Conclusion

5.4.1 Results Summary

pMTL-CW14

- Constructed pMTL-CW14 which has I-SceI sites flanking the transposon to eliminate plasmid reads in TIS.
- Assessed pMTL-CW14 for transposon mutagenesis in *C. autoethanogenum*.
- Created mutant pools rapidly using pMTL-CW14 in liquid culture.

pMTL-CW17

- Measured promoter activity of P_{bgaL} in *C. autoethanogenum* with a CAT assay.
- Constructed pMTL-CW17 which has the transposase under the control of a lactose-inducible promoter as well as I-SceI sites.
- Assessed pMTL-CW17 for transposon mutagenesis in *C. autoethanogenum*.
- Measured the effect of induction on transposition frequency of pMTL-CW17.

pMTL-CW20

- Constructed pMTL-CW20 which has I-SceI sites as well as P_{thl}-Tn-*codA* to increase plasmid stability.

- P_{thl}-Tn-*codA* increases plasmid stability in *E. coli* storage and donor strains.
- Assessed pMTL-CW20 for transposon mutagenesis in *C. autoethanogenum*.
- The addition of P_{thl}-Tn-*codA* in pMTL-CW20 does not significantly affect transposition frequency.

pMTL-CW22, pMTL-CW21, pMTL-CW23

- Constructed pMTL-CW22 which has I-SceI sites, P_{thl}-Tn-*codA* and a divergent tetracycline-inducible promoter system controlling conditional replication and transposase activity.
- Assessed pMTL-CW22 for transposon mutagenesis in *C. autoethanogenum*.
- Assessed pMTL-CW21 for transposon mutagenesis in *C. autoethanogenum*.

6 TraDIS

6.1 Introduction

6.1.1 TraDIS

As discussed in chapter 1, Transposon Insertion Sequencing (TIS) is a high-throughput functional genomics approach. It involves the creation of a transposon mutant library before high-throughput sequencing from the transposon to identify the insertion site. Simultaneous mapping of all insertions within a mutant pool allows essential gene sets to be identified, and repeated rounds of growth and sequencing allows fitness contribution to be quantified for all mutants simultaneously. Transposon Directed Insertion-site Sequencing (TraDIS) is a variant of Transposon Insertion Sequencing (TIS) technique with several minor differences to other TIS techniques (Langridge et al., 2009). TraDIS can in principle use any transposon and the technique is loosely defined by the method of sequencing library preparation. Several adaptations of the method have been made since the first inception, most significantly a follow-up paper by the original authors which includes bioinformatic tools and several sequencing library preparation improvements (Barquist et al., 2016). As with all TIS methodologies a random transposon mutant pool is created and a DNA sequencing library generated to identify the insertion-sites of all transposon mutants in the pool. In this study the TraDIS methodology was adapted to be more suitable for libraries created with replicative vectors by inclusion of an I-SceI digest step to remove plasmid-based reads.

6.1.2 Aims

The aim of this study were to

I) Demonstrate a reduction of plasmid read-contamination using an altered TraDIS technique.

II) Determine the essential gene set for *C. autoethanogenum* on rich medium.

III) Identify genes relevant to isopropanol resistance.

6.1.3 Sequencing Library Preparation and Inclusion of I-SceI Step

TraDIS sequencing library preparation is similar to standard Illumina sequencing for amplicon libraries, i.e. any libraries which involve a PCR enrichment step. A PCR enrichment step is necessary for TraDIS due to the relatively low number of transposon-chromosomal junction-containing fragments in a pool of fragments from an entire genome. The 4Mbp genome of *C. autoethanogenum* will be fragmented to an average size of 400bp, creating 10,000 fragments per genome, of which at most two will contain transposon-chromosomal junctions. In Illumina sequencing of whole genomes the DNA sample is fragmented and the P5 + P7 sequences are added to the ends of all fragments either directly during sample preparation or via PCR. For TraDIS, since we are only interested in transposon-chromosome junctions it is necessary to specifically add both adapter sequences to those junctions and leave other DNA fragments that lack transposon insertions without the adapters required for clustering. The standard Illumina sample preparation protocol most analogous to this study is the TruSeq DNA sample preparation for amplicon sequencing. In this protocol adapters are designed to include hybridisation regions for P5 and P7 primers the PCR products of which will contain both flow-

cell binding P5 and P7 sequences necessary for clustering and sequencing at opposite ends of the DNA fragments. TraDIS uses the same P7 primer, but uses a custom P5 primer (referred to as the library primer) which hybridises to

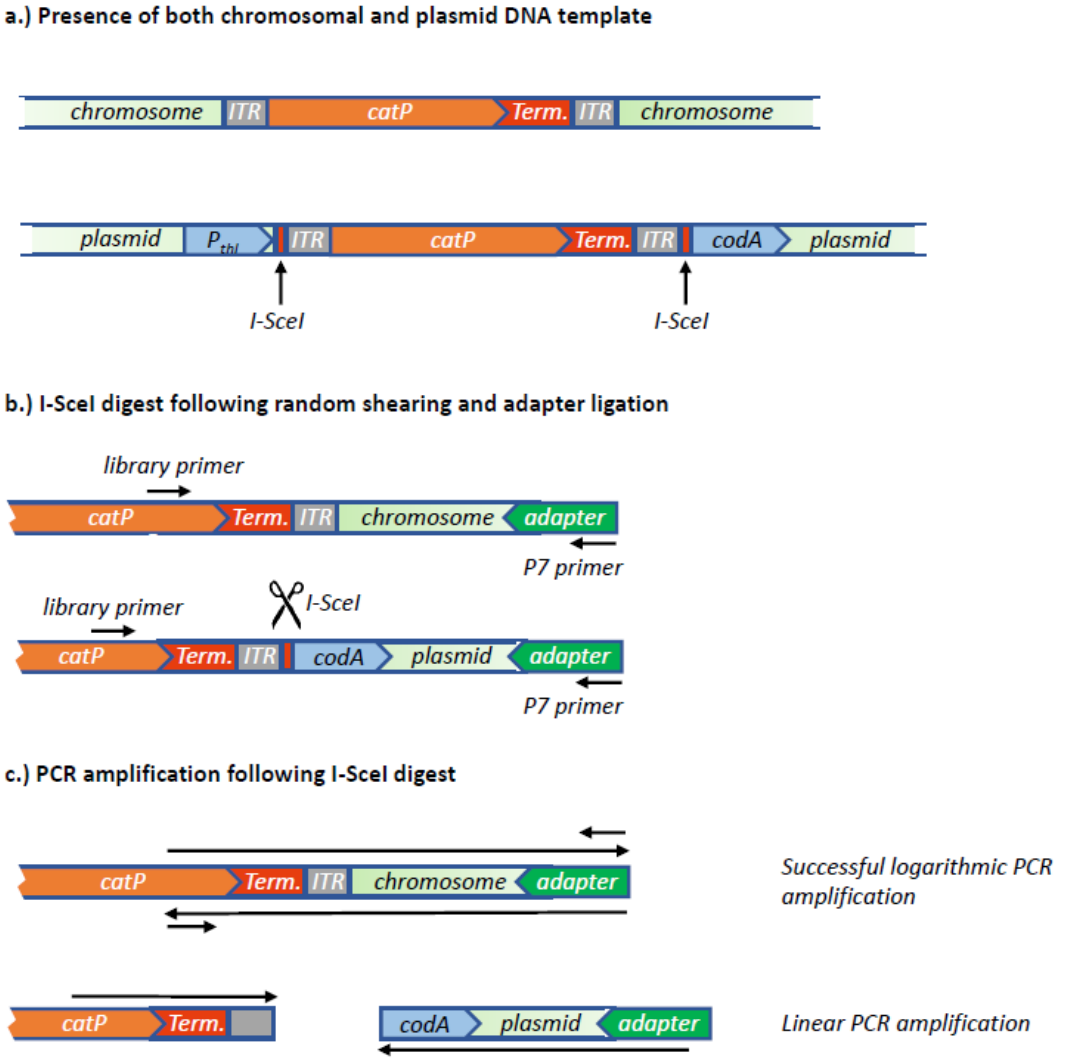


Figure 6.1: I-SceI digestion in sequencing library preparation.

Inclusion of I-SceI recognition sites on the transposon-delivery plasmid allows the removal of plasmid-derived reads at the sequencing library stage. a) Only DNA fragments containing plasmid-chromosomal junction-containing fragments have I-SceI recognition sites. b) Adapter-ligated DNA is exposed to I-SceI which cleaves the region between the library primer hybridization site and adapter where the P7 primer binds in transposon-plasmid junction-containing fragments specifically. c) PCR enrichment of transposon-chromosomal junctions proceeds while transposon-plasmid junctions fail to amplify exponentially.

the transposon such that exponential amplification of products should only proceed using transposon-chromosome junctions. Typical TraDIS sequencing library preparation involves fragmentation, end-repair, ligation of adapters, PCR enrichment and size-selection before sequencing. In this study a restriction digest step was added after the ligation of adapters in order to separate the hybridisation sites for the P7 and library primers in the subsequent PCR step exclusively within DNA fragments which contain the I-SceI recognition site i.e. fragments originating from plasmid DNA. The only fragments which should undergo exponential amplification during the PCR-enrichment stage are therefore fragments originating from transposon-chromosomal junctions. This strategy is illustrated in Figure 6.1.

6.1.4 Essential Genes

Knowledge of the essential gene set of *C. autoethanogenum* under standard laboratory conditions will be valuable for genetic engineering in this organism. Directed knockout mutant creation can be frustrated due to the lethality of the intended knockout. Alternatively a knockout strain could be created which would be lethal but for a compensatory mutation elsewhere in the genome giving a non-isogenic background to strain characterisation. A comprehensive record of essential genes would therefore be useful to avoid wasteful directed knockout efforts. Other transposon mutagenesis studies have revealed essential gene sets including genes that code for such integral processes as DNA replication, RNA synthesis, ribosome biogenesis, metabolism, cell cycle regulation and cell division (Christen et al., 2014). Many of the essential genes identified will be unsurprising and will likely map closely to those identified in related organisms

(Dembek et al., 2015), but of particular interest will be genes in central metabolic pathways toward fermentation products.

6.1.5 Identification of Isopropanol Resistance Genes

C. autoethanogenum does not natively produce isopropanol, but a production strain of isopropanol would be of commercial interest. A 2012 patent filed by Lanzatech reported production of isopropanol via an exogeneous pathway (Koepke et al., 2012). One potential problem for an isopropanol production strain is that of product resistance. *C. ljungdahlii* has also been shown to produce 5mM isopropanol via the incorporation of an artificial acetone synthesis operon with the acetone produced converted to isopropanol via the native NADPH-dependent primary-secondary alcohol dehydrogenase (Bengelsdorf et al., 2016). Tolerance to isopropanol becomes salient when production titres have been optimised to the extent that the product becomes inhibitory. This has occurred in studies on *Cupravidis necator* which was genetically engineered to produce up to 3.44 g L^{-1}, a level at which growth was retarded. Subsequent research overcame isopropanol toxicity via overexpression of GroESL genes allowing a final concentration of 9.8g g L^{-1} to be obtained (Marc et al., 2017). Efforts are ongoing within several research groups and Lanzatech to increase titres of isopropanol in *C. autoethanogenum* which, if successful, will likely render tolerance a desirable trait to incorporate.

The tolerance of wild type *C. autoethanogenum* was established in chapter 4. TIS experiments are capable of tracking the relative abundance of transposon mutants through periods of growth, in so doing relative fitness values can be calculated. The relative fitness values of all non-essential genes in a genome can be calculated simultaneously and compared to those obtained in different

growth conditions. In this case if a gene has a significantly higher fitness contribution value in the presence of a sub-lethal concentration of isopropanol compared to growth with no isopropanol then it must be implicated in isopropanol tolerance. In chapter 4 screening for solvent tolerance was performed by exposing mutant pools to lethal concentrations of ethanol, acetone or isopropanol. Tolerance gene identification using TIS offers a significant advantage over this approach since it is capable of highlighting genes which offer tangible benefit to the cell in the presence of solvent, but no observable change in the minimum inhibitory concentration.

6.1.6 Biological Library Creation Protocol

Despite the ease of liquid selection documented in the previous chapter, it was decided to use mutant selection agar plates. The use of agar plates in library creation does offer significant advantages such as the reliable estimation of library size and the more subtle advantage that it is far less likely that one mutant will come to dominate the mutant pool. On an agar plate each mutant effectively has a separate zone of easily accessible nutrients to deplete forming a colony as opposed to in liquid culture when every mutant is competing for the same pool of nutrients as demonstrated in *B. subtilis* (Fujikawa & Matsushita, 1991). It is possible that inter-colony signalling or even inter-colony killing via secretion of lethal compounds could aid in normalising the numbers of each mutant as reported in the Gram-positive bacterium *Paenibacillus dendritiformis* (Be'er et al., 2009). The result is that the variance in numbers of each mutant is far lower when mutants are selected for on plates as opposed to in liquid selection. There are differences in colony size as a result of differences in fitness, but these fitness effects would produce far greater variance in the

number of cells of each individual mutant if grown in liquid. Furthermore, in liquid selection there is likely to be less correlation between number of mutants at the end of growth and fitness since stochasticity plays a larger role in these conditions- if, at the start of growth a mutant happens to be in maximal cell division conditions then that mutant will proliferate faster than others which may have greater fitness overall and it would proliferate to the detriment of competitors due to nutrient depletion and waste deposition.

6.2 Results

6.2.1 TraDIS First Attempt

6.2.1.1 Mutant Pool Creation

Having established pMTL-CW20 as an effective means of transposon delivery with increased stability in *E. coli* donor and storage strains in chapter 5, it was chosen as the transposon-delivery vector to create a TraDIS library as follows. Twelve separate conjugations were performed from Sexpress. Transconjugants were selected for with clarithromycin and D-cycloserine and harvested from YTF agar plates by flooding with PBS. The cell suspension was then serially diluted in order to assess relevant CFU counts for calculations and 100-200µl of the neat suspension was dispensed to mutant selection plates containing IPTG, thiamphenicol (tm), D-cycloserine (DCS) and 10mM ß-lactose. Around 1.2 million colonies appeared in total on mutant selection plates and were harvested after 72 hours. The plated cell suspension was calculated to produce 262,000 colonies per ml on DCS, Tm, lactose, IPTG and 370,000,000 colonies per ml on DCS, Tm yielding a transposition frequency of 7.08×10^{-4}. Plasmid-loss was calculated by comparing the CFU on Clr to that of Clr yielding close to

100% efficiency with one in 2.46 million plasmid-harbouring cells able to grow in the presence of IPTG. Freezer stocks were created in 10% DMSO and stored at -80°C to serve as inoculum in future outgrowth experiments. Genomic DNA was extracted via phenol-chloroform extraction, detailed in materials and methods. Approximately 7µg of purified genomic DNA was submitted to the next generation sequencing facility DeepSeq.

6.2.1.2 Sequencing Results

Genomic DNA was processed as detailed in Barquist et al. (2016) except for the inclusion of a restriction digest step prior to PCR amplification and the use of TruSeq adapters rather than those specified in the reference. A MiSeq sequencing platform performed two separate sequencing runs which were identical except for the use of longer sequencing primers in the second run Table 6.1

Table 6.1: Sequencing primers.

Sequencing primers used in the two sequencing runs performed on the first TraDIS library produced in this chapter.

Name	Sequence	Oligoanalyzer Tm (°C)	GC (%)
Run One 5’	aaacaattgtgtcagaccggggacttatcagc	63.1	47
Run One 3’	gcttcttatttttgttagaccggggacttatca	59.9	38
Run Two 5’	cttgccgaaaaacaattgtgtcagaccggggacttatc	65.1	47
Run Two 3’	gcaggcttcttatttttgttagaccggggacttatca	63.3	42

Table 6.2: TraDIS reads summary.

A summary of the sequencing data from the two sequencing runs performed on the first TraDIS library produced in this chapter. The percentage of reads mapping to the PhiX genome is much higher than expected, while the detection of reads containing the transposon tag is very low (<1% of total reads).

Statistic	Run One	Run Two
Total Reads	7,622,373	11,867,744
Reads mapped to PhiX	6,276,237 (81.8%)	9,006,540 (75.5%)
Reads with Transposon Tag	36,857 (0.5%)	77,515 (0.7%)

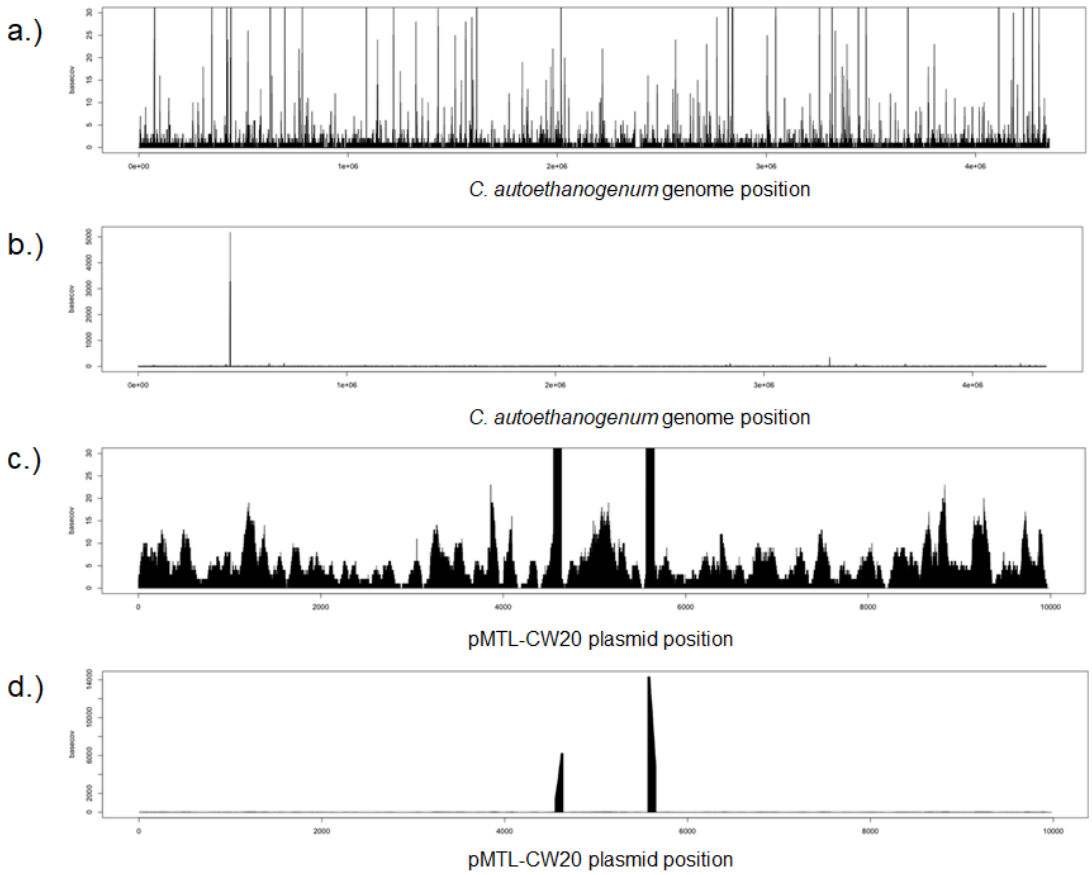


Figure 6.2: Transposon insertions across *C. autoethanogenum* genome and pMTL-CW20.

Transposon-tagged reads were mapped against the *C. autoethanogenum* reference genome and the pMTL-CW20 transposon-delivery plasmid. a.) Mapping to the *C. autoethanogenum* reference genome showing insertions across the genome. b.) Reads mapped to the *C. autoethanogenum* genome at a larger scale allowing the domination of a single gene to be visualised. c.) Reads mapped against the pMTL-CW20 transposon-delivery plasmid at a finer scale showing insertions around the plasmid. d.) Reads mapped against the transposon-delivery plasmid at a larger scale showing that insertions around the plasmid were relatively rare compared to the original transposon location.

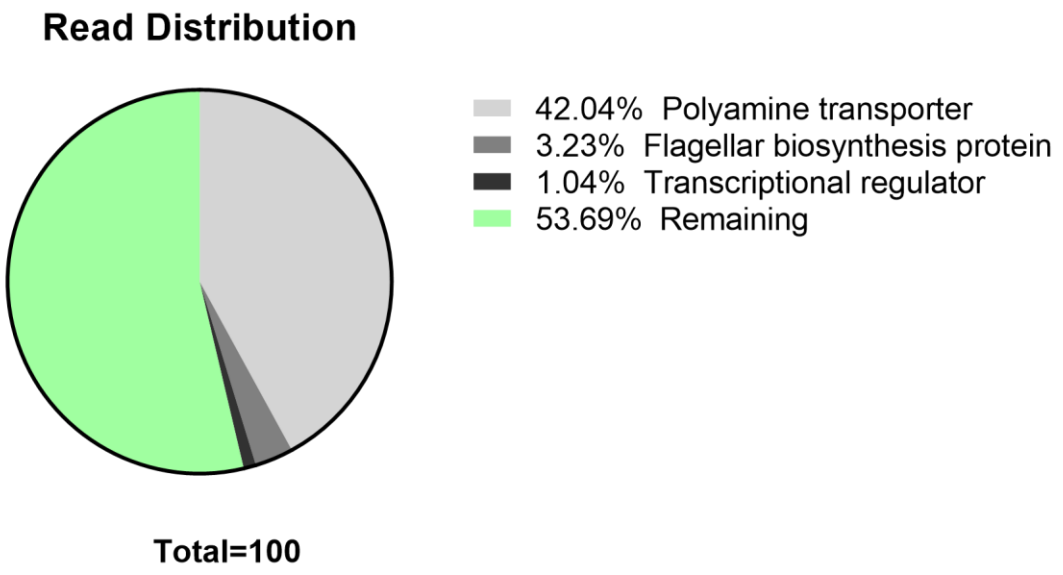


Figure 6.3: Read distribution

Transposon tagged reads with mapped to the chromosome and the bacteria_tradis script from the TraDIS analysis pipeline (Barquist et al., 2016) was used to produce a csv file with the number of mapped for each gene in the provided .embl file. The three most represented genes were plotted to give an impression of the library complexity.

Reads were mapped to the *C. autoethanogenum* reference genome NZ_CP012395. Figure 6.2 shows reads mapping to the genome, in total 18,370 reads contained transposon tags and mapped successfully. Reads appear distributed throughout the genome at one scale, but the domination of a single gene is evidence when the scale is increased. Combining data across both runs revealed that 42% of total mapped chromosomal reads were located in CLAU_0389, a putative amino acid transporter (Figure 6.3). In total 2,275 genes contained insertions leaving 1706 with no insertions.

Reads were also mapped against the pMTL-CW20 plasmid sequence. Figure 6.2(d) shows that insertions elsewhere in the plasmid were extremely rare with the majority of sequencing reads emerging from the original, intended

transposon location within the plasmid. At a finer scale it can be seen that there are other insertion sites around the plasmid. In total 21,855 reads mapped to pMTL-CW20, a similar amount to the genome-mapping reads.

The most germane fact about these results is the extremely low number of useful reads. The reasons for this were professed to be unknown. Therefore an effort was made to troubleshoot the DNA sequencing library process and attempt a repeat.

6.2.2 Troubleshooting

6.2.2.1 Detection of Transposon Tags

When a low number of mapped reads is reported the most pertinent question is: Are there a low number of reads with the transposon tags or do reads with transposon tags fail to map effectively? This is important since it directs troubleshooting to either the *in vitro* sequencing library preparation or the *in silico* read manipulation procedure. If a low number of reads contain the transposon tag (the 8-12bp terminal region of the transposon) this suggests that there was insufficient enrichment of transposon-chromosome junctions *in vitro*. If, however there are ample reads containing the expected transposon tag, but many of them fail to map to the reference genome, this mandates *in silico* analysis to account for the mapping failures. The bioinformatic analysis may for instance reveal a contamination of the mutant library or the presence of a plasmid which was not included in the reference genome.

In this case the problem is a lack of reads containing the transposon tag – 0.5% and 0.7% of total reads for run one and two respectively. The transposon tag in this case was the 11bp terminal region of the transposon 5’-AGCCAACCTGT-

3’. The first step in analysis was to verify that this low value of tagged reads is legitimate and not the result of poor quality sequencing in this region. The detection of mapped reads was repeated this time allowing for 2 mismatches in the sequence, but no significant change in the number of tag-containing reads was reported forcing the conclusion that insufficient enrichment of transposon-chromosome junctions was the overall cause of the low number of mapped reads. To investigate the low number of mapped-reads the sequencing library preparation process was considered.

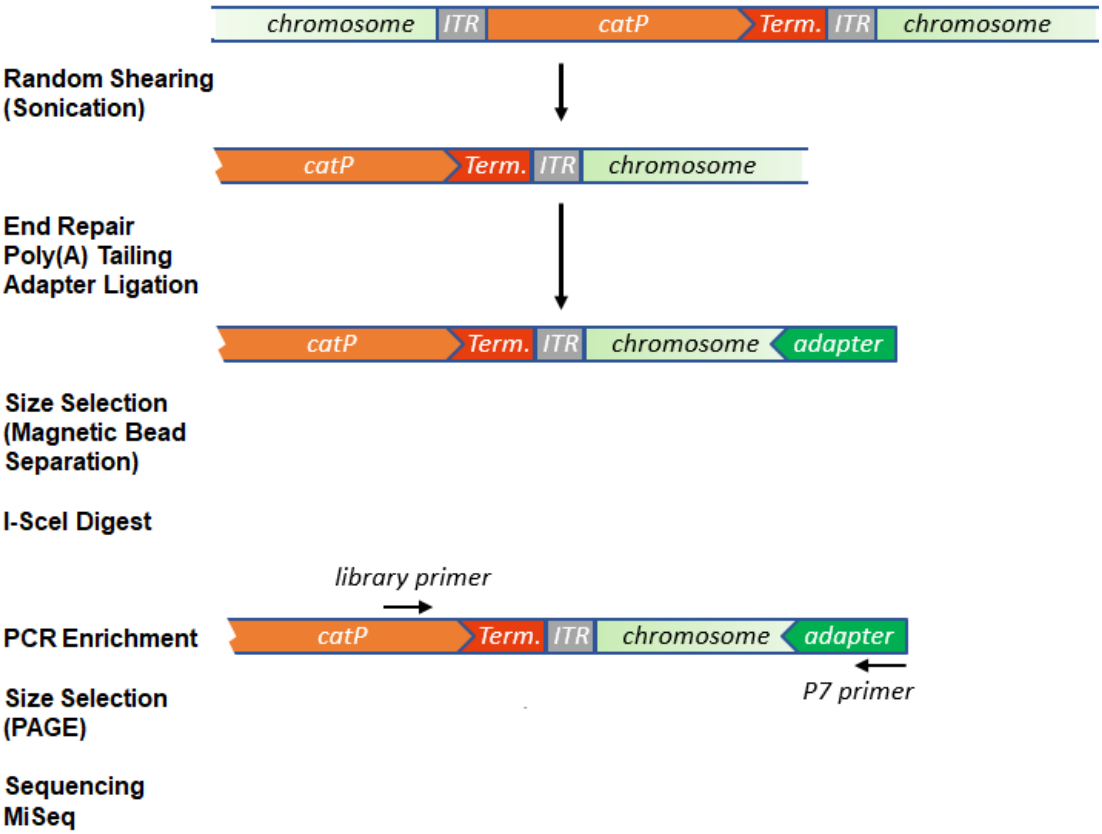


Figure 6.4: Sequencing library preparation summary.

Genomic DNA is sheared via sonication to an average length of 400bp. Sheared fragments were then end repaired and dA tailed using reagents from the NEBNext® Ultra™ II DNA Library Prep kit. Adapters were ligated to fragment ends and the small fragments from the ligation removed with Ampure XP magnetic bead size selection. Adapter-ligated DNA was then used as a template for a test PCR for library primer amplification. Adapter-ligated fragments were treated with 25U of I-SceI overnight at 37°C in a total volume of 70µl. Post-digest samples were PCR amplified to enrich for transposon junctions. PCR products were run on a PAGE gel and a band extracted from a lower limit of 250bp. Gel extracted products were run on a MiSeq sequencing platform.

6.2.2.2 DNA Sequencing Library Preparation

There is a diversity of DNA sequencing procedures employed in high-throughput insertion site sequencing experiments. Initially four similar protocols were published in 2009 each with a different name and each varying at the sequencing library preparation stage (Gawronski et al., 2009; Goodman et al., 2009; Langridge et al., 2009; van Opijnen et al., 2009). Since 2009 there have

been many more modifications to the protocol some of which created yet more novel names such as Rapid transposon liquid enrichment sequencing (TnLE-seq), Cartesian Pooling-Coordinate Sequencing or CP-CSeq and Random Barcode Transposon-Site Sequencing (RB-TnSeq) (Bristow, Butland, & Arkin, 2015; Fels et al., 2013; Vandewalle et al., 2015). A summary of the sequencing library preparation performed at DeepSeq is provided in Figure 6.4.

6.2.2.3 Verification of Transposon-chromosomal Junction Enrichment

It is necessary to enrich for fragments containing transposon-chromosomal junctions since they are present at relatively low concentrations within the total adapter-ligated sheared DNA sample. For this reason amplicon libraries are invariably used for transposon sequencing runs. The amplification step uses one primer which hybridises internally within the transposon (the library primer) and a second primer which binds to the adapter which is ligated to the terminus of all DNA fragments (P7 primer). Optimisation of the PCR step can be crucial to the success of library preparation. During the optimisation the objective is to ensure that reactions which contain both primers (library and P7) give rise to more product than either primer acting individually. Only products arising from the combination of both primers will contain the transposon tag and flanking chromosomal region necessary for read mapping. Higher product concentration in reactions containing the primer pair compared to either primer acting individual acts as an indicator of successful enrichment of transposon-chromosomal junctions. The increase in DNA in the primer pair reaction must be composed of fragments which contain the transposon sequence as well as the adapter sequences needed for flow-cell binding later in the process.

To ensure that the library primers were effective the described test PCRs were performed after adapter-ligation and found to produce stronger bands when visualised on an agarose gel. The next stage was to treat adapter-ligated samples with I-SceI to specifically cleave in-between plasmid-located transposon sequences and the terminal adapter. The addition of I-SceI to adapter-ligated DNA was intended to render transposon-plasmid junctions unable to amplify in the subsequent PCR-enrichment stage.

Since the I-SceI stage effectively removes a portion of the effective template for the PCR-enrichment stage the test PCR performed earlier in the process could be repeated. The primer pair post-I-SceI is required to amplify from a lower concentration of template and as such may require an increased number of cycles in the PCR in order to out-produce either primer acting alone. The adapters used allow amplification of fragments containing the P7 primer acting alone. This sole P7 activity only produces fragments linearly with respect to the number of cycles, but since it can act on every fragment containing an adapter sequence it essentially has a ‘head start’ on transposon-chromosomal junctions which amplify exponentially from a lower initial template concentration. A higher number of cycles in the PCR-enrichment stage therefore favours transposon-chromosomal junctions, but increase cycle number exacerbates ‘amplification jackpots’ resulting in library bias, hence the optimisation. The crossover number of cycles which is sufficient for the primer pair to out-produce P7 acting alone was found in the pre-I-SceI stage adapter-ligated template to be less than the 10 cycles allowed in the test PCR. However, the I-SceI stage removes a portion of the transposon template relative to the plasmid-contamination of the sample, hence increasing the ‘head-start’ of the P7 primer. Therefore an extra test PCR

is warranted post I-SceI treatment to verify that the primer pair still effective enrichments for transposon-chromosomal junctions.

Another method by which enrichment for transposon-chromosomal junctions can be verified is to use the final sample which is ready for sequencing as template in a quantitative PCR (qPCR). qPCR is regularly employed before running a sample on a sequencer, the objective being to ensure that the sample will bind the flow-cell oligonucleotides and to calculate the sample concentration in order to optimise for cluster density. In these routine qPCRs the primers hybridise to the P5 and P7 flow-cell binding regions of DNA fragments necessary for clustering. In order to verify transposon-chromosomal enrichment different primers are used at this stage. To calculate the proportion of the sample which represents transposon-chromosomal junctions the library primer used in the PCR-enrichment stage is used in combination with an adapter-hybridising primer. A second qPCR can also be performed using the custom sequencing primer which hybridises upstream of the transposon tag in combination with the P7 primer, this allows an estimation of the amount of sample likely to produce sequencing reads. In this library preparation process the transposon-chromosomal enrichment qPCRs were not performed. A lack of test PCRs post I-SceI treatment and a lack of qPCRs using library or sequencing primers leave open the possibility that the enrichment of transposon-chromosomal junctions was insufficient.

6.2.2.4 Clustering

Illumina sequencing involves running samples along a flow-cell to which sample DNA will bind via complementary oligos which are fixed to the glass slide. Clusters form by bridge amplification wherever samples successfully bind the

oligos. Inappropriate cluster densities both too high and too low can cause a low output of sequencing data.

Clustering refers to the bridge amplification process occurring on the flow-cell, a cluster forms where a sample containing correct P5 and P7 adapter sequences on both 5' and 3' ends binds to the oligos adhered to the flow-cell, whereupon it undergoes repeated rounds of DNA synthesis. The imager detects the presence of clusters, in order to achieve maximal data output the clusters should be as dense as possible whilst still being far enough apart for the imager to detect each cluster individually.

It is important to quantify the concentration of sample added to the sequencer so as to achieve optimum clustering. Samples are typically quantified using fluorometric assays such as Qubit™. In addition, samples can be visualised using the Aligent Bioanalyzer which gives a detailed breakdown of the fragment sizes in the sample. The Qubit™ gives the total amount of dsDNA and the Bioanalyzer is used to verify the absence of large amounts of primer dimer or adapter artefacts and to ensure the DNA fragments are mostly of an appropriate size.

It is possible to be mistaken in the concentration of sample administered to the sequencer, for instance if only a Qubit™ reading is taken without visualising the fragment size breakdown of the sample. Fluorometic DNA quantification may indicate the presence of a large concentration of DNA and yet a Bioanalyzer image would show that much of this dsDNA is in the form of small primer-dimers or adapter artefacts. In this case under-clustering would be observed since the

primer dimers would not possess the flow-cell binding regions necessary for clustering.

The addition of the phage PhiX genome is frequently included in sequencing runs in order to ensure the sequencing is proceeding as expected. The percentage of reads which map to the PhiX genome can help to determine if the concentration of sample submitted to the sequencer was too high or too low. If a 1% spike of PhiX is added then 1% of total reads and sample concentration was calculated correctly then 1% of reads should map to the PhiX genome, if the percentage is significantly lower than 1% that implies that the concentration of sample was higher than expected. In these cases the density of clusters forming on the flow-cell can be high enough to inhibit the sequencing process. In this case the opposite is true, a 5% spike of PhiX was added and yet between 75.5% and 81.8% of reads mapped to the PhiX genome. This implies that the concentration of appropriate strands of sample DNA containing both flow-cell binding regions was too low, or that the sequencing primer added to sequence out from the transposon was far less efficient than the standard Illumina sequencing primer which sequences the PhiX DNA. The former explanation seems more likely given that low clustering was reported. If the problem were the sequencing primer then clusters should still form, but be unable to produce reads. As discussed previously, the P7 primer generates a band in the control reaction when it is the sole primer. This amplification will only be linear because it never creates more template. However, the number of possible templates in the reaction is very high since (assuming 100% ligation efficiency) every strand of DNA will have a binding site for the P7 sequence. The products of P7

amplification acting alone will not bind flow-cells and form clusters and they could cause inaccurate fluorometric quantification of sample.

6.2.2.5 Size Selection

Another issue that may have contributed to a low number of transposon tagged reads is in the size-selection step. Size-selection is performed after PCR-enrichment to ensure that fragments contain enough flanking chromosomal DNA to identify the insertion site. Figure 6.5 shows a bioanalyzer readout of the sample after shearing and adapter-ligation, at this point the sample appears adequate with more than 70% of the DNA large enough to produce useful reads. The fragmentation also appears to be random producing a characteristic bell shaped distribution with no unexpected peaks.

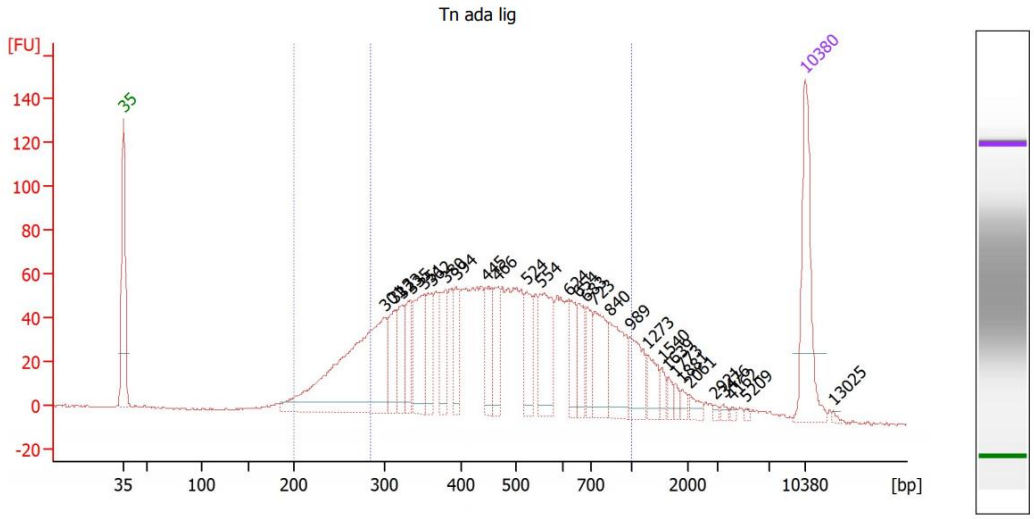


Figure 6.5: Size distribution of DNA fragments after adapter ligation.

DNA fragment sizes were visualised using an Aligent 2100 Bioanalyzer using a high sensitivity dsDNA chip. The sample shows the expected size distribution curve after adapter-ligation.

Following PCR enrichment a size-selection stage is performed to ensure that fragments submitted to the sequencer are of sufficient length to contain both adapter sequences necessary for flow-cell binding, the transposon sequence and a sufficient amount of flanking genomic DNA to identify the insertion site. Normally 12bp of genomic DNA is the minimum length sufficient to allow identification of the insertion site. The minimum length DNA necessary to be useful for each library (both 5' and 3' ends of the transposon are created as separate libraries) is therefore 319bp for the 5' end library and 286bp for the 3' end library as illustrated in Table 6.3. Given these calculations the PAGE size selection performed prior to sequencing should have removed any fragments shorter than 319bp or 286bp. However, size-selection was performed at 250bp retaining fragments which were too short to produce mapped reads whilst occupying valuable flow-cell space. Post size-selection samples were analysed using a Bioanalyzer again as shown in Table 6.3. The Bioanalyzer results show that a significant fraction of the DNA supplied to the sequencer would not produce useful reads.

Table 6.3: Size selection required for libraries.

Two libraries were generated in total with each library representing one end of the transposon. Sequencing out of both ends of the transposon allows more confident identification of insertion sites with less ambiguity. The total fragment size required to produce mappable reads was calculated from the composite parts of the DNA fragment after library preparation.

Library End	P5-flow cell binding (bp)	Library primer binding region (bp)	Distance to transposon end. (bp)	Genomic DNA required. (bp)	Adapter. (bp)	Mappable read required length (bp)
5' end	29	36	179	12	63	319
3' end	29	31	151	12	63	286

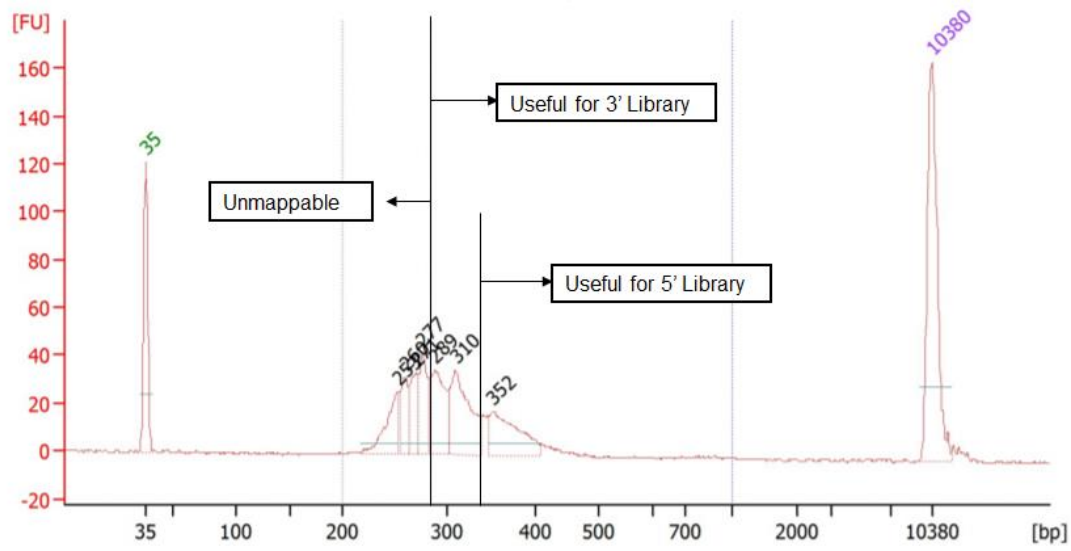


Figure 6.6: Size distribution of DNA fragments after PCR enrichment.

The post size-selection DNA fragments which were run on the MiSeq were analysed on an Aligent 2100 Bioanalyzer using a high sensitivity dsDNA Chip. Fragments smaller than 286bp are unable to produce mappable reads for either library end whilst fragments between 286bp and 319bp were only useful for mapping insertions from the 3' end library. Fragments over 319bp were large enough to generate enough flanking genomic DNA to identify an insertion site, though these represented a minority of the sample.

6.2.2.6 TA Cloning

If much of the DNA post size-selection did not contain the necessary flow-cell binding regions this would explain the under-clustering. To investigate this possibility, post size-selection samples were cloned into TA vectors and sequenced via Sanger sequencing. Samples were ligated into the pCR2.1 vector using a TA cloning kit (ThermoFisher Scientific) following the manufacturer's instructions. Thirty colonies of the resulting transformation were inoculated to overnight cultures from which plasmids were purified. Purified plasmids were subsequently sent for Sanger sequencing using the T7 primer which allows sequencing along the insert region of pCR2.1. These thirty sequences were analysed in order to determine if the flow-cell binding regions were present and correct. In order to form clusters correctly the fragments should have the form indicated in Figure 6.7.

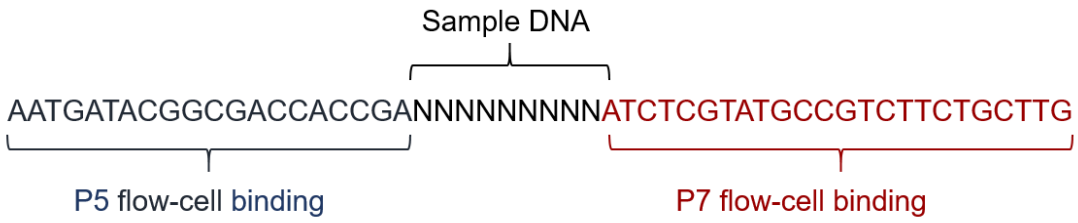


Figure 6.7: Expected adapter-containing fragments for sequencing.

In order for fragments to form clusters and sequence they must contain P5 and P7 flow-cell binding regions flanking the sample DNA.

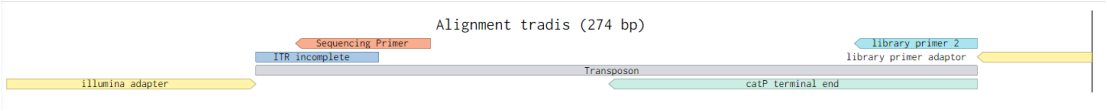


Figure 6.8: Alignment of TA cloning sequencing fragments.

An alignment of the sequencing data which contained the adapter sequence was produced which demonstrated that all fragments sequenced contained terminal P5 and P7 sequences in the correct orientation. A sample of fragments which were sequenced was used to clone into TA cloning vectors. Single colonies were picked and used to prepare plasmid DNA which was then sanger sequenced. Each individual colony resulted from a separate fragment in the library. The sanger sequencing results summarised here show the presence of the two illumina adapters as well as the transposon from the point of amplification with the P5 primer.

6.2.2.7 Summary of Troubleshooting

A low number of mappable reads was obtained from both sequencing runs performed. Less than 1% of reads originated from the transposon. Low clustering was observed which suggests that the concentration of fragments with the necessary flow-cell binding regions was too low. It is possible that a suboptimal sequencing primer design compounded the problem. The percentage of reads mapping to the PhiX genome was far higher than expected given the calculated sample concentration and the amount of PhiX DNA spiked into the sequencer. This suggests that the sample DNA failed to cluster or sequence despite the sequencing chemistry being operational. The failure of size-selection to remove fragments which were too short to map at least contributed to a low number of mappable reads.

6.2.3 TraDIS Second Attempt

6.2.3.1 Library Creation

The pMTL-CW20 transposon-delivery plasmid was chosen as for the previous TraDIS experiment. pMTL-CW20 was transformed into the Sexpress conjugal donor *E. coli* strain and conjugations performed as detailed in materials and methods (section 2.4.7). The remaining donor culture not used in the conjugation was used to extract plasmid DNA for a diagnostic digest to validate plasmid integrity. In total twelve conjugations were performed simultaneously producing cumulatively ~81,000 colonies on plasmid selection plates. Selection plates were flooded with PBS and pooled before being plated to mutant selection plates (1mM IPTG, Tm, 10mM ß-lactose, DCS). Approximately 1.3 million colonies were observed in total over all of the mutant selection plates.

6.2.3.1.1 Quality Control

To determine that the majority of these colonies were indeed transposon mutants and not merely plasmid-harbouring cells quality control was performed involving several different CFU counts from the initial plasmid-selection plate. Resuspension and replica plating to thiamphenicol and clarithromycin was performed on random colonies from mutant selection plates. For each 1ml plasmid-selection plate resuspension mix there were approximately 950,000,000 CFU plasmid-containing cells, 500 of which were able to grow in the presence of IPTG and clarithromycin ($IPTG^R$ plasmid-containing cells), 55,000 cells were able to grow on IPTG, lactose, thiamphenicol plates (putative transposon mutants).

The fact that the CFU on Tm, IPTG is higher than that of Clr, IPTG is an indication of the occurrence of transposition events. The ratio of colonies on mutant selection plate was estimated at 99.1%. Of the 55,000 Tm, IPTG resistant colonies we expect 500 of them to be merely plasmid-harbouring IPTG resistant cells and so the remaining 49,500 cells will likely represent transposon mutants.

$$\frac{55000 - 500}{55000} \times 100 = 99.09\%$$

The transposition frequency was calculated as 5.73×10^{-5} by dividing the estimated number of chromosomal transposon mutants by the total estimated number of total plasmid-harbouring cells.

$$\frac{(55000 - 500)}{950000000} = 5.73 \times 10^{-5}$$

The plasmid loss efficiency was calculated by dividing the CFU on Clr, IPTG by that of Clr alone multiplied by 100 giving 99.99994%, this implies that one in 1.9 million plasmid-containing cells forms a colony despite the presence of IPTG.

Forty-eight colonies were picked at random from mutant selection plates to estimate the percentage of transposon mutants as an alternative to the calculation performed previously. Colonies were streaked sequentially onto agar plates supplemented with clarithromycin followed by thiamphenicol and finally clarithromycin, IPTG.

Colonies on mutant selection plates originated from plasmid-harbouring cells, when performing replica plating or picking colonies to streak onto clarithromycin/thiamphenicol it is possible to pick a part of the colony which despite being a transposon mutant, also holds at least one copy of plasmid. In

these cases there will be growth on clarithromycin due to the lack of IPTG. Two out of forty-eight colonies grew successfully on Clr, IPTG, these are true IPTG resistant plasmids while eight out of forty-eight grew on Clr alone. Growth from clarithromycin plates was re-streaked onto clarithromycin, IPTG and only growth originating from colonies which successfully grew on Clr, IPTG were able to re-streak onto Clr IPTG. These results suggest that the majority (96%) of colonies are transposon mutants. This is 3% lower than that calculated by comparing the CFU on Clr, IPTG to that of Tm IPTG (99%), the latter figure is likely to be more accurate. In either case, a percentage of transposon mutants of 96-99% is sufficient to proceed.

6.2.3.2 Inoculation of Transposon Pool to Challenge Conditions

Transposon mutant colonies were harvested from mutant selection plates by flooding the plates with PBS. The optical density of the cell suspension was measured and used to inoculate YTF or YTF with 2.63%(w/v) isopropanol to OD_{600} 0.2. Each condition was supplemented with thiamphenicol and 1mM IPTG to continue to select for transposon mutants. Cultures were monitored with optical density measurements and at the end of growth freezer stocks were made and the remaining culture was pelleted for later DNA extraction. The two cultures with/without isopropanol grew as expected with the isopropanol causing an observable growth retardation as intended (Figure 6.9). It is this growth retardation which should favour the proliferation of isopropanol tolerant mutants and as such their overrepresentation in the final sequencing library.

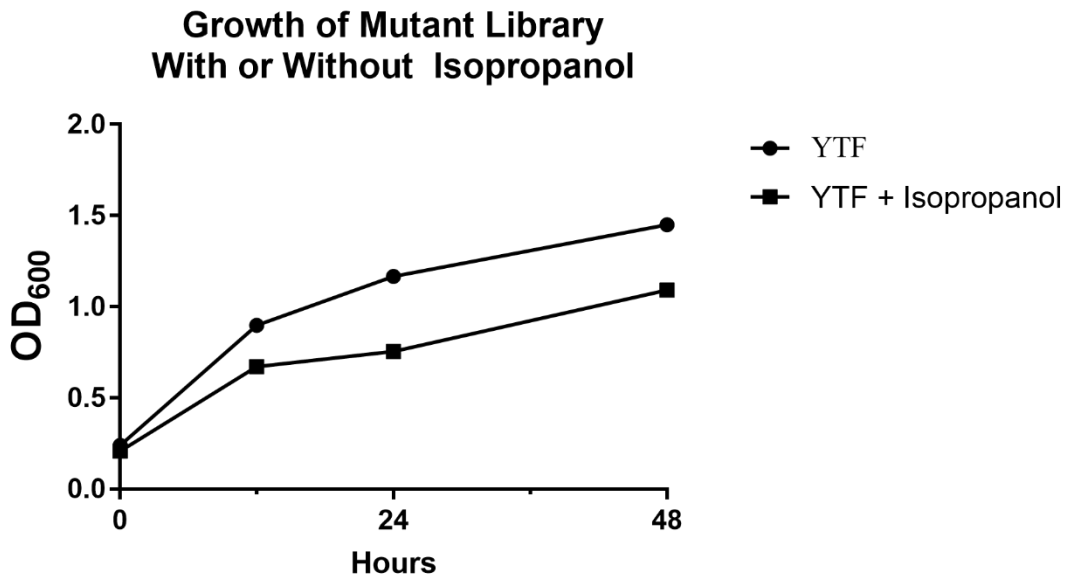


Figure 6.9: Growth curve of transposon mutant libraries.

An initial transposon mutant pool harvested from mutant selection plates was used to inoculate YTF or YTF with 2.63% (w/v) isopropanol, both cultures were supplemented with thiamphenicol and IPTG to continue to select for transposon mutants. The OD_{600} of the cultures was measured after 12, 24 and 48 hours.

6.2.4 Sequencing Library Preparation

6.2.4.1 Redesign of Library Primers

One problem with the initial sequencing library preparation was the lack of fragments large enough to produce useful data as shown in Figure 6.6. A simple way of increasing the proportion of useful fragments is to redesign the library primers to be closer to the ends of the transposon meaning smaller fragments would suffice to give enough genomic DNA to identify an insertion site.

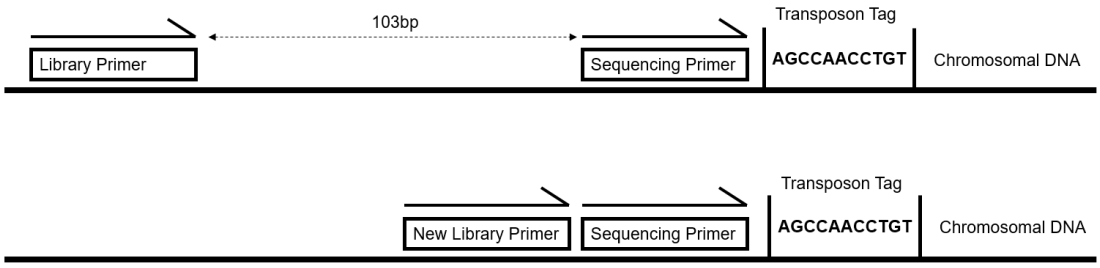


Figure 6.10: Library primer re-design.

The library primers which hybridize to the transposon sequence were re-designed so as to hybridize closer to the end of the transposon. Decreasing the distance between the library primer hybridization sites and the transposon terminus will decrease the size requirement for DNA fragments in order to produce mappable reads.

6.2.4.2 Redesign of Sequencing Primers

Sequencing primers were designed so as to resemble the standard Illumina sequencing primer 'SP1' in terms of GC content and melting temperature (51.5% GC, 65.5°C). The melting temperature of the re-designed primers is slightly increased since increasing the primer length between run 1 and run 2 produces slightly better results.

Table 6.4: Sequencing primer comparison between the two sequencing library preparations.

A comparison of the properties of all six sequencing primers used or supplied to DeepSeq to be used.

Primer Name	Oligoevaluator Tm (°C)	Oligoanalyzer Tm (°C)	GC Content %
New 5' Sequencing Primer	83.8	66.7	46
New 3' Sequencing Primer	81.4	66	44
Run One 5'	76.1	63.1	47
Run One 3'	77.2	59.9	38
Run Two 5'	81	65.1	47
Run Two 3'	76.8	63.3	42

6.2.4.3 Removal of plasmid DNA prior to sequencing library preparation.

Another potential problem in the initial sequencing library was the removal of effective template for the transposon-chromosomal junction enrichment PCR after the I-SceI digest step removed the adapters from the plasmid-located transposons. An attempt was therefore made to remove plasmid DNA prior to the sequencing library preparation. An extra I-SceI digest was performed on genomic DNA samples obtained from transposon mutant pools. The digested DNA was then run on a low-melt agarose gel and the highest band of the gel was excised. The linearised plasmid would run faster than the genomic DNA allowing a large proportion to be removed. The excised gel was then melted and treated with beta-agarase before purification with Ampure XP magnetic beads. Using this method between 30-40% of DNA which was initially exposed to I-SceI was retained.

Two sequencing libraries were prepared to assess if the additional purification step improves the read quantity. The first library was excised from an agarose gel after I-SceI digestion and then purified with beta agarase and Ampure XP magnetic beads. The second library omitted this step and proceeded into the sequencing library preparation as previously. A number of changes were made to the sequencing library process as well as the additional purification step. Library primers were designed to be closer to the transposon ends and size-selection was performed at a level high enough to exclude fragments too small to be useful. In addition, several more test PCRs were performed before and after I-SceI digest to detect if the reduction of effective template significantly inhibited product formation.

6.2.4.4 Sequencing Library Process

The priority of this sequencing library preparation was to establish the method, given the problems associated with the previous run it was decided that only two sequencing libraries would be created and sequenced, one of which had been through the I-SceI digest/gel purification stage and one which had not. Different library primers were used to increase the proportion of useful fragments generated in the PCR. Sonication was performed as previously using a Covaris S-series sonolab single following a protocol optimised to produce 400bp fragments using the following settings: 20W, 200 cycles/burst, intensity 5, 10% duty cycle. A 1µl sample of sonicated product was run on a 2.5% agarose gel to verify fragmentation as expected. Fragments were end repaired and dA-tailed using NEBNext Ultra II DNA Library Prep reagents. For this sequencing run the DNA extracted from a single culture in YTF was used. The treated DNA sample is that which underwent excision from an agarose gel, the untreated did not. Wild type *C. autoethanogenum* DNA was used to provide a control for the test PCRs performed later in the process. Truseq adapters were ligated as previously using indexes 4, 5 and 19 for wild type, untreated and treated samples respectively. After purification using Ampure XP magnetic beads samples were quantified using a Qubit High Sensitivity dsDNA kit. Adapter ligated DNA was treated with 25 units I-SceI (NEB) overnight in a total volume of 70ul. Samples were purified again using Ampure XP beads. Test PCRs were performed using both undigested adapter-ligated DNA and the corresponding samples post-digest. The objective was to verify that the library primers worked as expected by determining that there was a greater amount of DNA in the reactions which had both library primer and P7 primers than with P7

alone. Annealing temperature and the number of cycles were varied to optimise the PCR, using 10ng of template DNA. Having optimised the PCR conditions the final libraries were created using the recipes and settings detailed in Table 6.5 and Table 6.6.

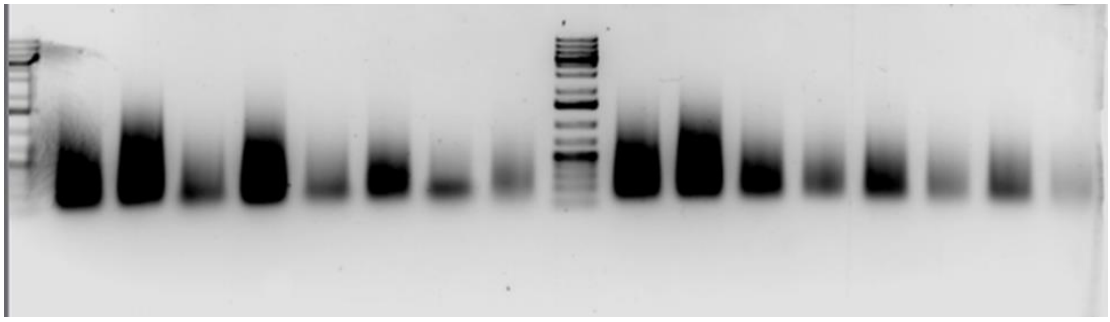


Figure 6.11: Optimisation of transposon-chromosomal enrichment PCR.

Annealing temperature was varied for reactions which either contained both library and P7 primers or the P7 primer alone, the combination of primers reaction always being represented prior to P7 alone. Lanes 1 and 10 1kb+ ladder. Lanes 2-9 used the 3' transposon end library primers while lanes 11-18 used the 5' transposon end library primers. The annealing temperature was varied as follows: lanes 2-3 58°C, lanes 4-5 61°C, lanes 6-7 64°C, lanes 8-9 67°C. Lanes 11-12 58°C, lanes 13-14 61°C, lanes 15-16 64°C, lanes 17-18 67°C. PCR products were also quantified using Qubit readings to determine which reaction condition created more dsDNA product. The annealing temperature chosen for final library creation was 58°C for the 3' end library and 61°C for the 5' end library.

Table 6.5: PCR recipe for transposon-chromosomal junction enrichment.

Component	Volume (µl)
Kapa 2X Master Mix	25
Template. Adapter-ligated, I-SceI treated samples	x (100ng)
P5	5
P7	5
ddH_2O	15 - x

Table 6.6: PCR program for transposon-chromosomal junction enrichment.

Stage	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	95°C	120s	1
Denaturation	95°C	20s	20
Annealing	56° for 3' libraries 61°C for 5' libraries	30s	
Extension	72°C	30s	
Final Extension	72°C	120s	1

PCR products were run on a 2.5% ultra-low range agarose gel which was post-stained and the area above 200bp was excised and purified using NEB monarch gel extraction kit. Samples were quantified using a nanodrop and a Qubit High Sensitivity dsDNA assay before being visualised using an Aligent 2100 Bioanalyzer using a DNA High Sensitivity chip.

Table 6.7: Summary of changes to sequencing library preparation.

Stage	First Library	Second Library
Input DNA	No initial plasmid removal step.	Added initial plasmid-removal step for one library.
Library Primers	Further from transposon end.	Closer to transposon end. 127bp for 5' end and 133bp for 3' end.
Test PCRs	Test PCR performed after adapter ligation.	Test PCR performed after adapter ligation and after digest.
PCR Enrichment of Tn-chromosomal Junctions	10 cycles	20 cycles
Size Selection	250bp lower limit, but 319bp (5') or 286bp (3') needed for detection of insertion site.	250bp lower limit with 192bp (5'), 153bp (3') needed for detection of insertion site.
Sequencing Primers	Lower Tm. Run 1: 63.1°C and 59.9°C Run 2: 65.1°C and 63.3°C	Higher Tm 66.7°C and 66°C

The bioanalyzer results show a much more even distribution of fragment sizes with less exaggerated peaks. This suggests that PCR products are the result of an amplification of a more diverse range of template fragments. The size of the fragments is on average greater which, in combination with the new library primer design closer to the transposon ends, will result in the vast majority of fragments being of an appropriate size to give useful sequencing data. Samples were submitted for sequencing on a Miseq platform and results are pending. Despite promising Bioanalyzer graphs the sequencing results are needed to state whether the changes in library preparation have been effective or not and in addition whether the extra plasmid-removal step was effective.

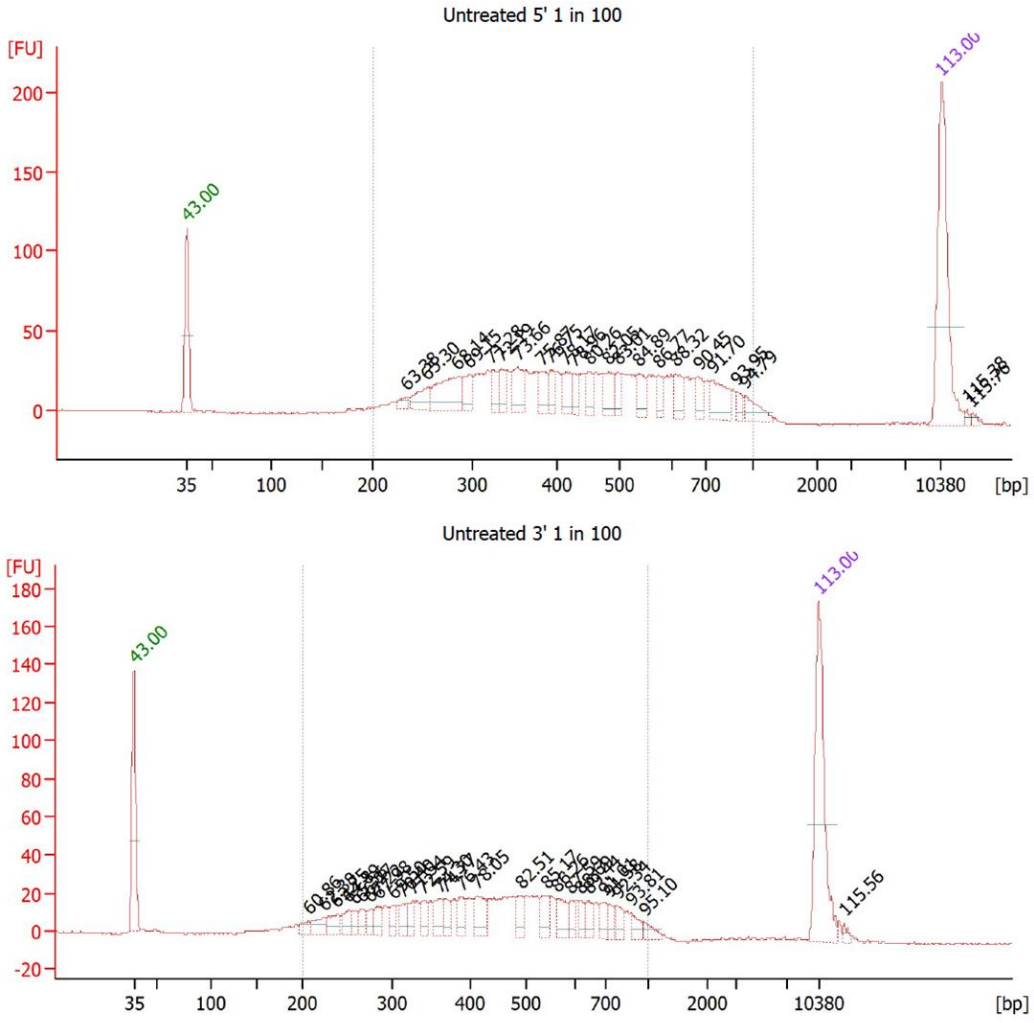


Figure 6.12: Size distribution of DNA fragments prior to sequencing, untreated libraries.

Post-size selection samples ready for sequencing were visualised using an Aligent 2100 Bioanalyzer using a high sensitivity dsDNA Chip. The expected distribution of fragment sizes is observed with the vast majority of fragments being large enough to produce mappable reads.

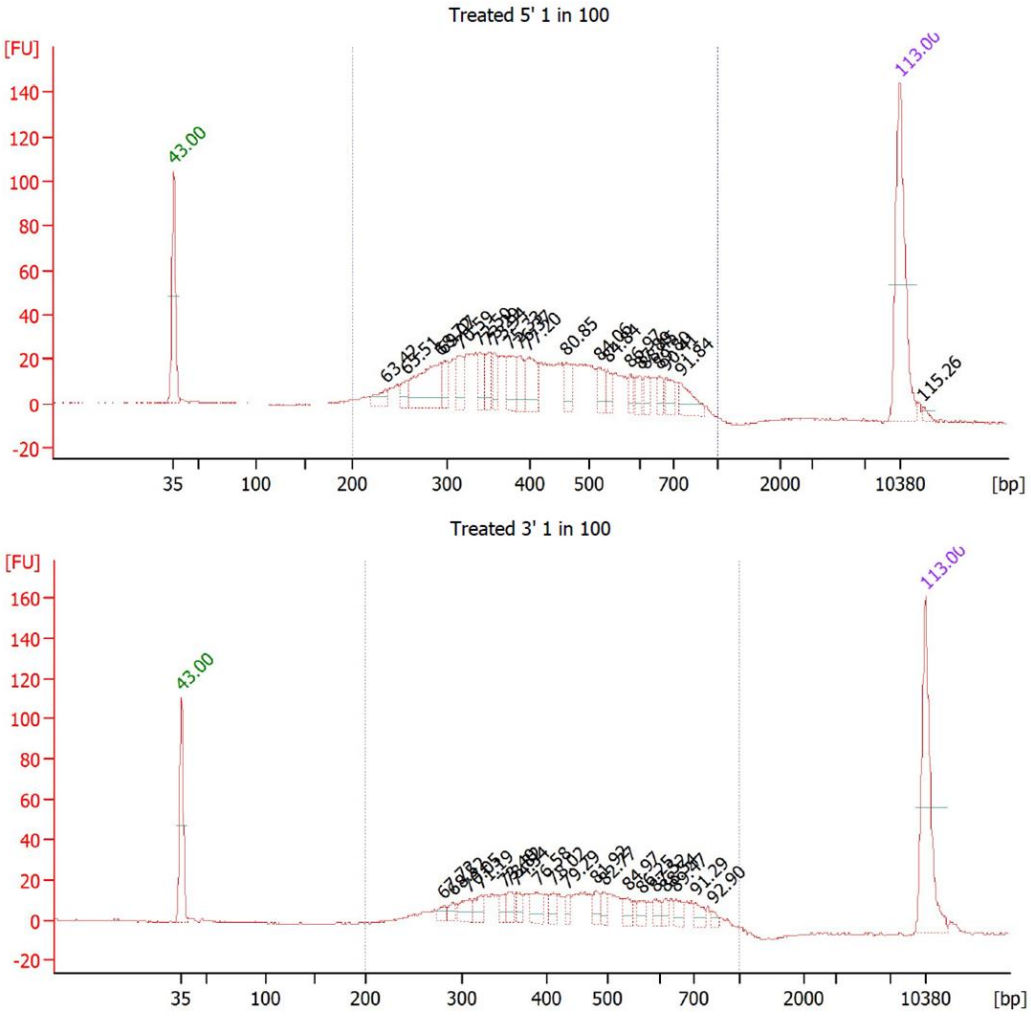


Figure 6.13: Size distribution of DNA fragments prior to sequencing, treated libraries.

Post-size selection samples ready for sequencing were visualised using an Aligent 2100 Bioanalyzer using a high sensitivity dsDNA Chip. The expected distribution of fragment sizes is observed with the vast majority of fragments being large enough to produce mappable reads.

6.3 Discussion

6.3.1 Overview

Two separate biological transposon mutant libraries of sufficient size for high-throughput transposon site sequencing were created a sequencing libraries prepared for each. The first sequencing library was used in two separate sequencing runs using a different primer pair each time. In both of these initial

sequencing runs less than 1% of the total reads contained the transposon tag indicating insufficient enrichment of transposon-chromosomal junctions during library preparation. Troubleshooting of the library preparation revealed several possible explanations and directed adaptations for a repeat of the process. A second biological library was created and used to prepare another sequencing library using an altered protocol. The altered protocol included a novel additional plasmid-removal stage prior to sonication. Bioanalyzer visualisations of the final library demonstrate an increased proportion which is large enough to produce mappable reads as well as a more even distribution of peaks possibly indicating a less biased library or the absence of non-specific adapter artefacts. Sequencing results will reveal if the changes to the library preparation have been effective and if the additional plasmid-removal step increases the sequencing quality and volume. The benefit of performing test PCRs before and after the restriction digest step as well as visualising the sample prior to sequencing is highlighted here and can avoid costly sequencing runs on samples which do not pass quality control. qPCRs using library and sequencing primers with the P7 primer would add another useful quality control measure.

6.3.2 The Importance of Library Complexity

TIS experiments rely on a large number of mutants to increase confidence of calling a given gene to be essential. When plotting number of genes against number of transposon insertions/possible insertion sites a bimodal plot should be observed with essential genes being represented on the left hand side. Essential genes will be underrepresented and will have a very low ratio of insertion sites detected to possible insertion sites. Non-essential genes are expected to have far higher ratios as insertions are non-lethal and thus mutant

cells harbouring them can proliferate. The distance between the left hand peak of essential genes and the plateau of non-essential genes correlates with the statistical certainty with which we can call a gene essential. This distance will be increased by a larger, more complex library since there will be a higher number of insertions in non-essential genes. In the ideal case using a *mariner* transposon all possible insertions in a gene can be calculated due to the requirement of the dinucleotide sequence TA for an insertion. If the biological library is sufficiently large and the sequencing is sufficiently saturated then all possible insertion sites would be mapped.

The first library was sequenced producing 18,370 reads which mapped to the *C. autoethanogenum* genome. This is an insufficient quantity of reads to estimate gene essentiality. In some cases low saturation libraries have been used to predict essentiality purely on the basis of an absence of reads, but even this low-stringency condition is unsuitable since around 2,000 genes lack any insertions. The number of essential genes in *C. autoethanogenum* growing on rich medium is likely to be around 500 based on studied in organisms of similar genome size and relatedness (Dembek et al., 2015). The problem of early transposition event mutants dominating the mutant pool was again exemplified with a single gene occupying 42% of reads. Over-representation of genes can occur due to genuine domination of a mutant in the biological library or via exponential amplification of early or preferred products in the PCR-enrichment stage. It would not be surprising if this were a result of the biological library preparation process since a mutant arising in the conjugation mating plate stage would proliferate exponentially in the plasmid-selection plate stage before many isogenic mutants were plated to mutant-selection plates resulting in multiple

colonies representing the same mutant. It is also possible that this prevalence is the result of an amplification jackpot event in the PCR-enrichment stage. To identify amplification jackpot events bar-codes can be added to adapters allowing amplification products to be attributed to separate template molecules (Zhang et al., 2012).

6.3.3 Removal of Plasmid-based Reads with a Restriction Digest Step

Similar incorporation of a restriction digest step to reduce plasmid reads has been published recently for *Desulfovibrio vulgaris* (Fels et al., 2013). In this study the restriction enzyme BglII was used, which has a 6bp recognition sequence found 150 times in the *D. vulgaris* genome (and 856 times in the *C. autoethanogenum* genome), around which there are inevitable 'dark spots' of sequencing since these reads will be eliminated in the same way of plasmid based reads. The 18bp preferential I-SceI recognition sequence occurs at a frequency of 1 in 6.9×10^9 bases in random DNA and is not found in the *C. autoethanogenum* genome, although it exhibits less than 100% site-specificity stringency (Doyon, Pattanayak, Meyer, & Liu, 2006). The effectiveness of the I-SceI strategy cannot be adequately assessed with the data presented in this chapter. The amount of reads mapping to chromosome and plasmid was approximately equal with 18,370 and 21,855 respectively. The parity of read occupation between chromosome and plasmid is not encouraging, but in this case the failure to effectively amplify transposon-chromosomal junctions has been confirmed and since the I-SceI strategy depends on preferential amplification it is not surprising that we do not see a more promising elimination of plasmid-based reads.

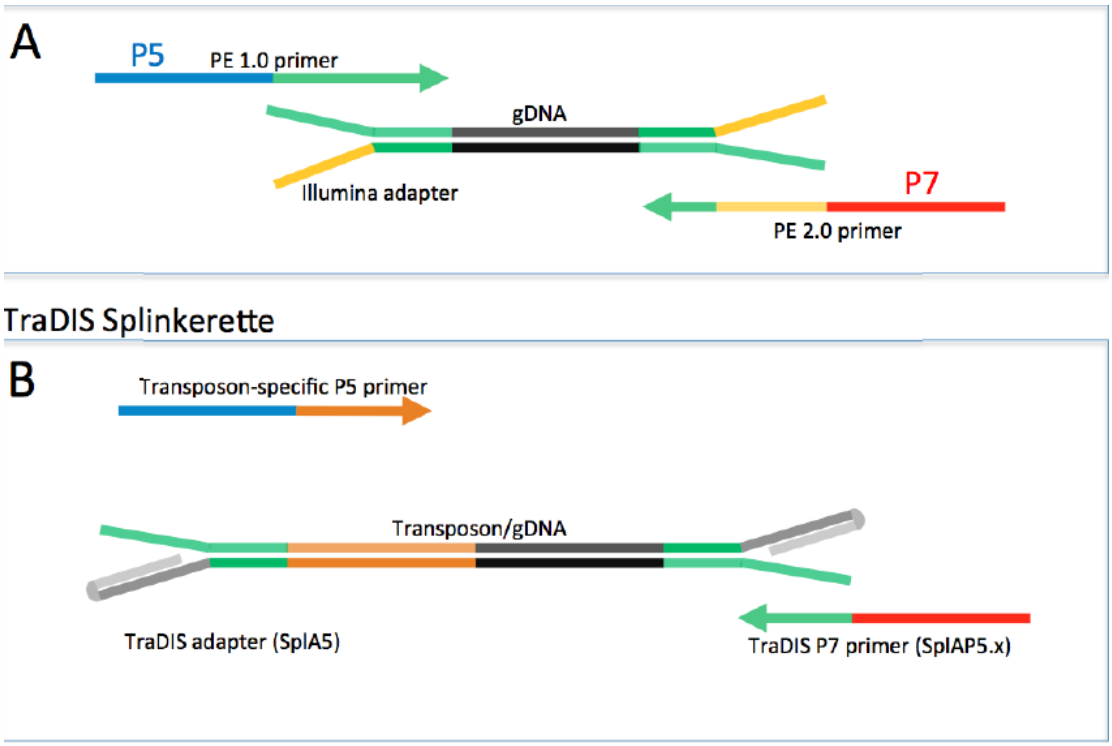


Figure 6.14: Splinkerette adapters.

In standard Illumina library preparation primers PE 1.0 and PE 2.0 are used to amplify fragments. Both these primers have homology to the ligated adapter. The PE 2.0 primer has a longer complementary sequence to the adapter, therefore during PCR samples more easily hybridise with PE 2.0 than PE 1.0. As a result, the first cycle of PCR generates more strands containing the P7 sequence added by PE 2.0 than the P5 sequence added by PE 1.0. The relevance of this to transposon sequencing is that in Tn-seq PE 1.0 is replaced by a transposon-specific binding primer with a region internal to the transposon as well as the P5 sequence. Therefore if the standard Illumina adapter and P7 primer is used in TIS preparations then PCR products this will result in poor enrichment of transposon-chromosomal junctions. The splinkerette adapter and primers detailed in (Barquist et al., 2016) are designed so that the P7 primer can only hybridize to the product of the P5 library primer meaning P7-derived amplification can only proceed on transposon-chromosomal junction-containing fragments.

6.3.4 Splinkerette Adapter

The optimisation of PCR conditions described previously may not have been necessary had a different adapter been used in the library preparation. The biotradis paper (Barquist et al., 2016) describes an adapter and P7 primer set

to be used which in theory should not produce a band without the presence of the transposon-specific P5 primer. The P7 primer will only yield PCR products when operating on template which has been generated by the action of the transposon-specific P5 primer (Figure 6.14). This is important due to the relatively low number of transposon-chromosomal junctions in the PCR template compared to pure genomic DNA.

6.3.5 Future Work

The essential gene set required to utilise CO or CO_2+H_2 as a sole carbon and energy source is an obvious research goal. This would allow a more accurate list of genes for which directed knockouts should not be attempted and could give valuable clues about additional genes implicated in the trait. It would be expected that the Wood-Ljungdahl pathway genes responsible for carbon fixation and energy production would be essential in these conditions, but other genes previously unlinked could also be implicated. The second biological library detailed in this chapter was used to inoculate PETC CO cultures, but poor adaptation to the condition led to concerns of bottlenecking events and no further sequencing analysis was performed.

Several knockout strains of *C. autoethanogenum* have recently been generated including a deletion of *adhE* which led to increased ethanol production (F. Liew, Henstra, Michael, Winzer, & Simpson, 2017). The creation and sequencing of transposon mutant libraries in such knockout strains provides direction in terms of which further genes it is possible to knockout. The same logic to generating a library for wild type applies to every subsequent strain produced since the

different genetic background will make certain genes change their essentiality status when compared to wild type. This phenomenon could be exploited to link genes to their function and reveal genetic interactions. Mutant libraries are produced in the deletion strain and the wild type strain before fitness contribution is assessed for every gene simultaneous as usual. The fitness contribution for each gene is then compared between the two strains and classed as unchanged, aggravating or alleviating depending on whether the mutant was more or less represented in the deletion strain background (van Opijnen et al., 2009).

7 General Discussion

7.1 Context

A sustainable energy supply can be achieved utilising the 120,000 TW of energy which is radiated to Earth. This energy can be directly harvested by photovoltaic cells or via photosynthesis and CO_2 fixation in the production of biomass which can subsequently be converted to liquid fuel. Biofuels have two primary advantages over direct photovoltaic cell energy production – a higher energy density and the abundance of biomass already cultivated for food production. Common lithium-ion batteries have an energy density of approximately 0.3-0.8 MJ kg^{-1} which is considerably lower than that of typical liquid biofuels 27-47 MJ kg^{-1} (Thackeray, Wolverton, & Isaacs, 2012). Liquid fuels are therefore heavily favoured in the transportation sector, especially since infrastructure for liquid fuels is largely already in place. Another concern is the environmental damage caused by the manufacture of batteries, which some life cycle analyses have found to emit more CO_2 than is saved by the avoidance of fossil fuel usage (Ellingsen et al., 2014).

The scientific discussion of the merits first generation biofuels has unfortunately become intrinsically enmeshed with politics in the United States meaning advice from the scientific consensus will be slow to be adopted, much as that of climate change has and continues to be. The EU has been more receptive to data on the limited biofuel yield and associated emissions, with effects on biodiversity and other environmental concerns leading to a limitation of 5% of transport fuels being provided by food-based biofuels (European Academies Science Advisory Council, 2012). This figure was later increased to 7% in 2015 in response to the

challenges of developing biofuels from non-edible food crops and a relaxation of criticism from the food-fuel debate (Phillips et al., 2016). A major obstacle to widespread second generation biofuels is the efficiency of feedstock utilisation. Chemical pre-processing can be expensive and inefficient (Chen, Smith, & Wolcott, 2016). An alternative is to gasify waste resources into syngas and subsequent fuel production via gas fermentation. Traditionally biofuels are defined as fuels produced from biomass, but gas fermentation can utilise non-biomass derived feedstocks in addition, such as industrial waste gases, and as such is not always classed as a biofuels technology. This classification has relevance for government incentives and is likely to change to a broader definition taking account of the presence of biological entities in the production of fuels and not merely in the starting material. Gas fermentation technology has reached maturity since commercial-scale production of ethanol from industrial waste gases is now a reality. One of the pull factors toward first-generation biofuels is the nationalist concern for energy independence. Gas fermentation offers the same benefit due to the relatively low importance of geography in production with this technology. Waste products which can serve as the feedstock for gas fermentation are almost ubiquitous.

It has been suggested that Synthetic Biology will lead to more distributed biological manufacturing (Friedman & Ellington, 2015). This is particularly helpful in healthcare settings where distribution times can increase mortality rates (Harrison, Ruck, Medcalf, & Rafiq, 2017). A programmable biologics-producing microbioreactor system has recently been reported and is expected to play a role in personal biomanufacturing in the future (Perez-Pinera et al., 2016). Gas fermentation could play a role in this decentralisation as energy

production in the future may be relatively decentralised with smart grids encouraging local renewable production.

7.2 DNA Transfer

To fully realise the potential of gas fermentation, strain development via genetic engineering must continue. This will involve improvements to the biological traits of the organism such as robustness, productivity, growth rate and product tolerance as well as an extension of the fermentation products to include non-native compounds such as isopropanol, acetone and 3-hydroxybutyrate. Crucial to strain development is the ability to transfer foreign DNA to the target organism. The use of Sexpress significantly improves frequency of conjugal transfer as detailed in chapter 3. Modification of the donor strain to express *C. autoethanogenum* methyltransferases is likely to offer further improvements. The refinement of conjugation or electroporation protocols is time-consuming with many possible variables potentially offering improvements, but is facilitated by the more reliable novel donor reported in this text since fewer transformation attempts will fail. Genetic alterations of the recipient strain could also be a route to improvements, either by the removal of predicted restriction enzymes or via less intuitive approaches, for instance a recent study observed an increased transformation rate in *Acinetobacter baylyi* ADP1 after removing mobile DNA elements from the genome (Suárez, Renda, Dasgupta, & Barrick, 2017). Improvements in DNA transfer will lead to the adoption of more ambitious genetic tools being applied to this organism as well as facilitating directed strain-modification in general.

7.3 Transposon Mutagenesis

Use of a *mariner* transposon-delivery system with a conditional replicon previously exemplified in related clostridial species was demonstrated in *C. autoethanogenum* and *C. ljungdahlii* (Zhang et al., 2015). The most interesting result obtained from screening was that of the product profile of mutants containing insertions in a transcriptional regulator (CLAU_1148) and a ferredoxin (CLAU_0697). These two mutants were first chosen at random for sequencing and later investigated based on the nature of their insertion site. Verification of these effects must be performed with further genetic studies, but this result demonstrates the potential of random mutant creation and characterisation. Despite not being forward genetics and clearly requiring good fortune it raises the question of how many genes would produce a significant effect on a trait as complex and nebulous as fermentation profile.

Several novel transposon-delivery vectors were constructed and demonstrated in *C. autoethanogenum* and several related species by other researchers. The orthogonal P_{tcdB} promoter system with the alternative sigma factor TcdR expressed from the chromosomal proved to be effective in *C. autoethanogenum* and *C. ljungdahlii*. The relative inactivity of P_{tcdB} in *E. coli* is a useful feature of the promoter, though this could be shared by the two inducible promoters which were also demonstrated P_{bgaL} and P_{tet} (Zhang et al., 2015). The advantages of each of these promoters to *E. coli* activity is not clear at this point and would require further experimentation. However, based on the frequent observation of plasmid re-arrangements in vectors containing all three promoters it seems likely there is still appreciable transposase activity. The P_{thl}-Tn-*codA* strategy detailed in chapter 5 appears to severely reduce the instances of plasmid

rearrangement since no rearrangement event has been reported using either pMTL-CW20, pMTL-CW21, pMTL-CW22 or pMLT-CW23. Further experimentation to more thoroughly quantify the increased stability is necessary, but it seems likely that this system will aid in the practicality of transposon mutagenesis.

The improvements in DNA transfer documented in this text were insufficient to make transposon-delivery via suicide plasmid a feasible option due to practical constraints. TIS studies using replicative transposon-delivery methods are uncommon, but the publication of several successful examples as well as innovations in DNA sequencing library preparations will likely lead to an increased adoption of the technique in less transformable organisms (Charbonneau et al., 2017; Dembek et al., 2015; Fels et al., 2013). Rapid Transposon Liquid Enrichment Sequencing (TnLE-seq) is a technique intended for use in bacteria with under-developed genetic toolkits, as implied it avoids the use of agar plates for mutant selection, instead a conjugation mating stage is allowed followed by a liquid selection stage for transposition events and counter-selection against the *E. coli* donor. The transposon-delivery plasmid is unable to replicate in the non-*pir* host and the DAP- *E. coli* donor is counter-selected by the absence of diaminopimelic acid (Fels et al., 2013). TnLE-seq was demonstrated within *Desulfovibrio vulgaris* Hildenborough, for which the number of transconjugants was sufficient to allow a genome-saturating library to be generated with no replicative plasmid using a single mating plate (Fels et al., 2013). If this constitutes an under-developed bacterial transformation rate then a new label must be applied to *C. autoethanogenum* to signify its recalcitrance, since if an analogous method was adopted, the resultant library

would be composed of approximately fifteen independent mutants as evidenced by the attempt to use the suicide-delivery plasmid pMTL-GL15 in chapter 5. The most significant contribution to TIS in under-developed bacterial chassis in the TnLE-seq paper is perhaps more likely to be the inclusion of the BgIII digest step in DNA sequencing library preparation which is analogous to the I-SceI step employed in this text (Fels et al., 2013). With non-replicative transposon-delivery the chance of obtaining a mutant colony can be broken down into two basic phases, DNA transfer and transposase activity in the cell. Transposase activity can be defined as the chance of a cell which harbours the transposon-transposase DNA fragment to undergo a transposition event. Transposon-delivery on a replicative vector can circumvent the problem of a low DNA transfer rate by outgrowth of cells harbouring plasmid until the point that the number of cells which possess transposon-transposase DNA fragments is similar to that which could be achieved using non-replicative delivery in a highly transformable organism. However, replicative vector transposon-delivery creates two main problems, a biased biological library and plasmid-read contamination, both of which result in a lower number of unique chromosomal reads. Plasmid-read contamination can be addressed with use of a conditional vector such that colonies on mutant selection plates are overwhelmingly not plasmid-harbouring cells. This was exemplified in the TraDIS study on *C. difficile* published in 2015 which pooled approximately 750,000 putative transposon mutant colonies and identified 77,256 unique insertion sites which gave sufficient genome coverage to identify 404 essential genes (Dembek et al., 2015). A similar study used temperature sensitive pWV01 replicon in *Streptococcus pyogenes* to create a genome saturating TIS library (Le Breton

et al., 2015). Plasmid read-contamination was sufficiently problematic that in one T_0 library only 6% of total reads mapped to the chromosome with the “vast majority” aligning to the delivery-plasmid instead (Le Breton et al., 2015). The problem was alleviated in libraries from further passages in the non-permissive condition (Le Breton et al., 2015). An effective conditional replicon is essential to TIS in organisms where non-replicative transposon-delivery is not feasible. Another approach for removal of plasmid read-contamination is at the sequencing library preparation stage as discussed in TnLE-Seq (Fels et al., 2013). A recent study combined these two approaches in a similar way to the work attempted in this thesis. A conditional replicon was used to transform *Streptococcus equi* and a population of cells harbouring plasmid was generated in the permissive condition before plating to the non-permissive condition selecting for the presence of the transposon (Charbonneau et al., 2017b). At the DNA sequencing library stage adapter-ligated DNA was treated with SmaI to remove plasmid-derived reads in an analogous way to the I-SceI strategy employed in this thesis (Charbonneau et al., 2017).

The application of TIS in less transformable organisms using a conditionally-replicative transposon-delivery plasmid is likely to become increasingly common, though the quality of these studies will potentially be limited due to the continual presence of plasmid in mutant libraries. The problem is essentially a lack of ability to select for genomic transposon mutants without simultaneously selecting for the presence of the delivery plasmid since the antibiotic marker is active in either case. Future work will perhaps employ a system in which the transposon marker is inactive in the presence of another element of the plasmid, for instance the production of antisense RNA to the marker. Alternatively, a

classical counter-selection marker could be incorporated to the plasmid backbone or the endogenous CRISPR system could be inducibly activated to target a backbone region. These suggestions could serve as a secondary measure to eliminate plasmid, though with a sufficiently efficient conditional replicon they make not be necessary, especially with the inclusion of additional passages.

In the introductory chapter the creation of random mutant libraries was described as an attractive alternative to defined single-mutant collections. However, a recent study demonstrated that transposon insertion-site sequencing methods can be applied to facilitating the creation of a defined single-mutant library (Baym, Shaket, Anzai, Adesina, & Barstow, 2016). In this approach combinatorial pooling is followed by sequencing from the transposon insertion-site using barcoded adapters to locate the sequence read to the pool it originated from. The mapped reads and attached bar codes are used to deduce the location of each mutant via a probabilistic algorithm (Baym et al., 2016). Using this approach a single-gene knockout collection in *C. autoethanogenum* is an achievable aim, despite the problems of sequencing a mutant containing plasmid-harbouring cells and the biased mutant pool likely to result from use of a replicative vector. The genetic techniques which have been available to model organisms are therefore becoming increasingly accessible to genetically recalcitrant species such as *C. autoethanogenum*. This is an encouraging trend for the discovery and understanding of the diversity of traits within biology.

TIS studies are producing increasingly complex and saturated libraries allowing the importance of small non-coding RNAs, domain level functional

characterisation and intergenic chromosomal regions to be understood. Sufficiently saturated libraries have now demonstrated the limitations of *mariner* transposition in terms of intrinsic bias of *Himar1* transposase in several prokaryotes (Long et al., 2017). The importance of protein-DNA interactions has also recently been further elucidated with the discovery of a histone-like nucleoid structuring protein H-NS reducing the chance of transposition in *Vibrio cholerae* resulting in a misidentification of essential genes (Akerley, 2016; Kimura, Hubbard, Davis, & Waldor, 2016). Important future discoveries from TIS studies will be allowed by clever experimental design to discriminate between groups of mutants based on a phenotypic trait. Previously the ratios of transposon mutants in a pool has always been linked to their relative fitness values, however a recent paper made use of fluorescence activated cell sorting (FACS) to sort mutants on the basis of physical characteristics. This approach broadens the applications of TIS studies by allowing high-throughput screening of a more diverse array of biological traits (Paulsen, Cain, & Hassan, 2017).

8 Appendix

8.1 Oligonucleotide Primers Used

Primer Number	Primer Name	Sequence 5' – 3'
24	TcdR_seq1	GAGCTTATGCAATTCAAGTAGGTACTGCAAAC
25	ACE pyrE FSP	GAGCTTATGCAATTCAAGTAGGTACTGCAAAC
26	ACE pyrE RSP	CATCAAAGCTATACTATTTTCCGTATTTACATTTGGG
27	CMH1	TAAATCATTTTTAGCAGATTATGAAAGTGATACGCAACGGTATGG
28	catP_INV_F1	TATTGTATAGCTTGGTATCATCTCATCATATATCCCCAATTCACC
29	catP_INV_R1	TATTTGTGTGATATCCACTTTAACGGTCATGCTGTAGGTACAAGG
30	catP-INV-R2	GGCAAGTGTTCAAGAAGTTATTAAGTCGGGAGTGCAGTCGAAGTGG
31	catP-F1	TGAAGTTAACTATTTATCAATTCCTGCAATTCGTTTACAAAACGG
32	catP-R1	CTAACAGGTTGGCTGATAAGTC
46	Himar1C9_Seq_F	TTTTTTTATCAGGAAACAGCTATGACCG
47	PtcdB_Seq_R	TTAACAGCTGCAAACACCGC
54	pyrE_longhomology_R	AGTAAAGGAGAAAATCATATGATGGAAAAAAAGG
82	82_pBAD-araC_fwd	CATGTTTGACAGCTTATCATCGATAAGCTTATGACAACTTGACGGC
83	83_pBAD-araC_rev	GCTAGTCTATTTTCCCAAAAAAACGGGTATG
84	84_CLAU_0514_fwd	CCGTTTTTTTGGGAAAATAGACTAGCACATTTATTGG
85	85_CLAU_0514_rev	TCTCCTTAAATTATCATATTTTGCACAGCATC
86	86_CLAU_1608_fwd	GTGCAAAATATGATAATTTAAGGAGAATACTATATGTCTTATG
87	87_CLAU_1608_rev	TAACTGTGATAAACTACCGCATTAAAGCTTCTATATACTGTCATTTTCAACTAAATTTTTTAG
109	109_CW14_backbone_F	AAGGAGTGATTACATGAACAAAAATATAAAATATTCTCAAAACTTTTTTAAC
110	110_CW14_backbone_R	ATTACCCTGTTATCCCTATAGTTTCTCGGTACCCGG
111	111_CW14_TN_F	TACCGAGAAACTATAGGGATAACAGGGTAATACAGGTTGGCTGATAAGTC
112	112_CW14_TN_R	ATTACCCTGTTATCCCTAACAGGTTGGCTGATAAGTC
113	113_CW14_1_F	TCAGCCAACCTGTTAGGGATAACAGGGTAATTATACCTCGAATTCGTAATCATG
114	114_CW14_1_R	AGAGGTGCAACATATGAAACCAGTAACGTTATACG
115	115_CW14_2_F	TTACTGGTTTCATATGTTGCACCTCTACTTTAATAATTTTTAAC
116	116_CW14_2_R	TATTTTTGTTCATGTAATCACTCCTTCTTAATTACAAATTTTTAG
129	129_Tn_I-SceI_F	GCGGCCGCTAGGGATAACAGGGTAATACAGGTTGGCTGATAAGTCCCCGGTCTAACAAAAAATAAGAAGCCTGCA
130	130_Tn_I-SceI_R	GCTAGCATTACCCTGTTATCCCTAACAGGTTGGCTGATAAGTCCCCGGTCTGACACAATTGTTTTTCGGCAAGTG
131	131_CW17_Backbone_F	GAAGCAAACTTAAGAGTGTG
132	132_CW17_Backbone_R	TAGTTTCTCGGTACCCGG
133	133_CW17_Lactose_F	AGGATCCCCGGGTACCGAGAAACTATTATATACTTGGTTTATTTACTTGATTATTTC

134	134_CW17_Lactose_R	CCTTTTTTTCCATCATTTGATATGCCTCCTAAATGG
135	135_CW17_Transposase_F	GGAGGCATATCAAATGATGGAAAAAAAGGAATTTCG
136	136_CW17_Transposase_R	TGACCGCGGCCGCTTATTCAACATAGTTCCCTTCAAG
137	137_CW17_GramPosRep_F	CTATGTTGAATAAGCGGCCGCGGTCATAGCT
138	138_CW17_GramPosRep_R	CCCTAGCGGCCGCGGCCGGCCGCTTATAATCCATAAC
139	139_CW17_Tn_F	TAAGCGGCCGGCCGCGGCCGCTAGGGATAAC
140	140_CW17_Tn_R	ATCAACACACTCTTAAGTTTGCTTCGCTAGCATTACCCTGTTATCCC
162	162_CW20_1F	GGCAAGACCGATCGGGCCCCCTGCATTAATGAATTTAAAGAAATATTTACAATAGAAATC
163	163_CW20_1R	ATTTTGTTAAAAATTATTCAACATAGTTCCCTTC
164	164_CW20_2F	CTATGTTGAATAATTTTTAACAAAATATATTGATAAAAATAATAATAGTG
165	165_CW20_2R	ATAACAGGGTAATAACTAACCTCCTAAATTTTGATAC
166	166_CW20_3F	TAGGAGGTTAGTTATTACCCTGTTATCCCTAACAGG
167	167_CW20_3R	GCAGGCTTCTTATTTTTATCACTTCGACTGCACTCCC
168	168_CW20_4F	TGCAGTCGAAGTGATAAAAATAAGAAGCCTGCATTTGCAGGCTTCTTATTTTTATGGCAAGTTGAAAAATTCAC
169	169_CW20_4R	ACACCTCCTTAAATAGGGATAACAGGGTAATAC
170	170_CW20_5F	CCTGTTATCCCTATTTAAGGAGGTGTATTTCATAATG
171	171_CW20_5R	TTATCTACAATTTTTTTATCCTGCATCAACGTTTGTAATCGATG
172	172-CW22-1F	GGGCGGGCGCGCCTTAAGACCCACTTTCACATTTAAG
173	173-CW22-1R	CCTTTTTTTCCATATGTATTTCCTCCTCTTCAATATATTTAAG
174	174-CW22-2F	GGAGGAAATACATATGGAAAAAAAGGAATTTCG
175	175-CW22-2R	TTAAGTTTGCTTCTTATTCAACATAGTTCCCTTC
176	176-CW22-3F	CTATGTTGAATAAGAAGCAAACTTAAGAGTGTGTTG
177	177-CW22-3R	ATTTTGTTAAAAAGATCGGTCTTGCCTTGCTC
178	178-CW22-4F	GGCAAGACCGATCTTTTTAACAAAATATATTGATAAAAATAATAATAGTG
179	179-CW22-4R	ATTTCGGCCGGCCTCAACGTTTGTAATCGATG
180	180-CW22-5F	TTACAAACGTTGAGGCCGGCCGAAATGCAAGTTTC
181	181-CW22-5R	AAGTGGGTCTTAAGGCGCGCCCGCCCTTAAG
194	194-cw24-backbone-reverse	GGGCCCTGCGTATTGGGC
195	195-cw24-LHA-fwd	GAAGAGCGCCCAATACGCAGGGCCCTTCTACGGCACGTTTGAAG
196	196-cw24-LHA-rev	CAAGTTGTCATAACTGGCTTCAGGAGATCGG
197	197-cw24-meth-fwd	CTCCTGAAGCCAGTTATGACAACTTGACGGC
198	198-cw24-meth-rev	CCGGAATCATATGCTATATACTGTCATTTTCAACTAAATTTTTTAG
199	199-cw24-RHA-fwd	TGACAGTATATAGCATATGATTCCGGGGATCC
200	200-cw24-RHA-rev	TTGTGAATTTTTCAACTTGCCCACTGCAGCGACTTCCACGATG
201	201-cw24-backbone-forward	AGTGGGCAAGTTGAAAAATTC
266	RecoV2-LHM-Fw	TTCTACGGCACGTTTGAAGG

267	RecoV2-RHM-Rv	GCAGCGACTTCCACGATG
268	Run One 5’	AAACAATTGTGTCAGACCGGGGACTTATCAGC
269	Run One 3’	GCTTCTTATTTTTTGTTAGACCGGGGACTTATCA
270	Run Two 5’	CTTGCCGAAAAACAATTGTGTCAGACCGGGGACTTATC
271	Run Two 3’	GCAGGCTTCTTATTTTTTGTTAGACCGGGGACTTATCA
272	New 5’ Sequencing Primer	CTTGAACACTTGCCGAAAAACAATTGTGTCAGACCGGGGACTTATC
273	New 3’ Sequencing Primer	GCAAATGCAGGCTTCTTATTTTTTGTTAGACCGGGGACTTATCAGC
274	Library Primer 5end	AATGATACGGCGACCACCGAGATCTACACTTCGACTGCACTCCCGACTTAATAACTTC
275	Library Primer 3end	AATGATACGGCGACCACCGAGATCTACACTGCAGGCTTCTTATTTTTTGTTAGACC

8.2 TraDIS Results Summary

The following are the results of the sequencing performed on the final library preparation created in chapter 6. A superior number of transposon-tagged reads which mapped to the genome were detected. The I-SceI plasmid read removal strategy appears to have been successful. The biological mutant library appears to be sufficiently complex for essentiality determination. Predicted essential genes are presented in table 8.3.

Summary of TraDIS in *C. autoethanogenum*

Unique insertion sites = 96,302

Average of 1 insertion per 45bp.

Comparison to first *C. autoethanogenum* library preparation.

Table 8.1 A comparison of the two library preparations performed in this work.

Two separate biological libraries were created in this text, the first of which was used to create a sequencing library from, this was performed by the sequencing company Deep Seq. A second biological library was created and subsequently used to create a sequencing library. Both libraries were sequenced with the second library giving superior results in terms of mapped insertion-sites.

	First library	Second library
Transposon-tagged reads	36,857	6,822,472
Successfully mapped to genome	18,370	6,570,667

Reduction of plasmid-based reads.

A novel method of reducing plasmid-based reads in addition to the I-SceI step. Genomic DNA was run on a low-melt agarose gel and the top band extracted, treated with agarase and purified. These samples are referred to as 'gel extracted' on the following table. Both gel extracted and untreated *C.*

autoethanogenum samples were digested with I-SceI. The *C. beijerinckii* library had no I-SceI treatment.

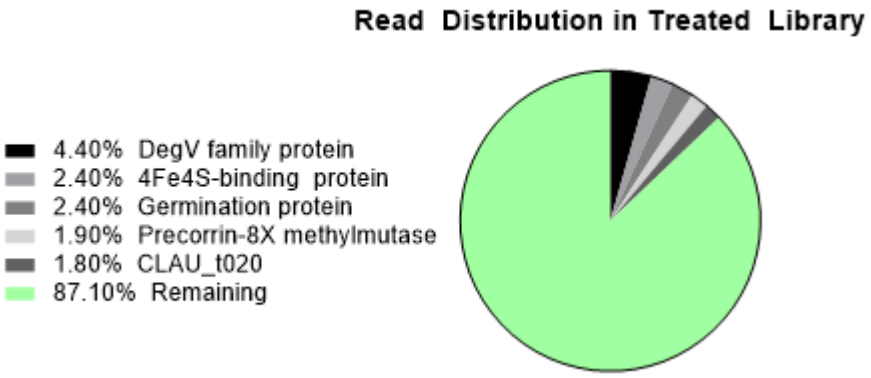
Table 8.2 Effect of methods to remove plasmid based reads

A comparison of three sequencing library methods with various treatments for the removal of plasmid-based reads. The *C. beijerinckii* library was created and sequenced in 2014 by Deep Seq and used no technique to attempt to remove plasmid-based reads. The two *C. autoethanogenum* examples are both from the same final biological library created in this text and subjected to identical sequencing library preparations but for the inclusion of a gel extraction step for the example in column four.

Organism	***C. beijerinckii***	***C. autoethanogenum***	***C. autoethanogenum***
Treatment	**No I-SceI treatment**	**I-SceI treated**	**Gel extracted and I-SceI treated**
Reads mapping to plasmid	54%	0.48%	0.06%

Improved library diversity

Two sequencing libraries were generated on the same run using different barcoded adapters. The libraries were the product of separate PCRs and yet have strong correlations in mutant abundance implying that over-represented mutants are not merely sequencing artefacts.



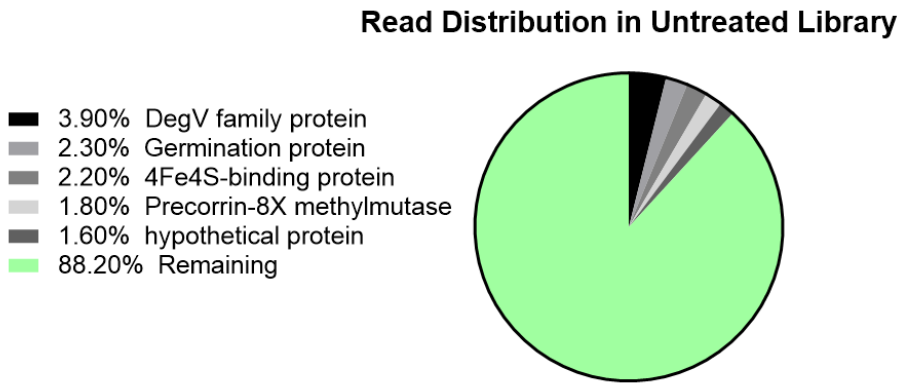


Figure 8.1 Read Distribution

The five genes with the highest number of insertion-sites are shown as a percentage of the total reads alongside the percentage representing every other gene in order to show a lack of domination by a small selection of mutants. This is shown for two sequencing libraries which were subject to separate PCRs in order to demonstrate that a high percentage of read occupation is a consequence of a high prevalence in the biological library and not merely as a result of a jackpot amplification event at the PCR stage of library preparation.

C. autoethanogenum **Essential Genes**

The bioinformatic pipeline outlined in (Barquist et al., 2016) was followed which generates a list of predicted essential genes. The transposon tag ‘agccaacctgt’ was used in the final essentiality analysis with a 2bp mismatch being allowed. This allowed the prediction of 419 essential genes which are listed in Table 8.3. Another 155 genes were assigned as ‘ambiguous’ meaning they could not be placed as essential or non-essential with a good degree of certainty. These may represent mutant which are viable but with significantly lower fitness values, or they may represent essential genes which have had experienced transposition events later on in the growth period from a plasmid-harbouring cell (Table 8.4).

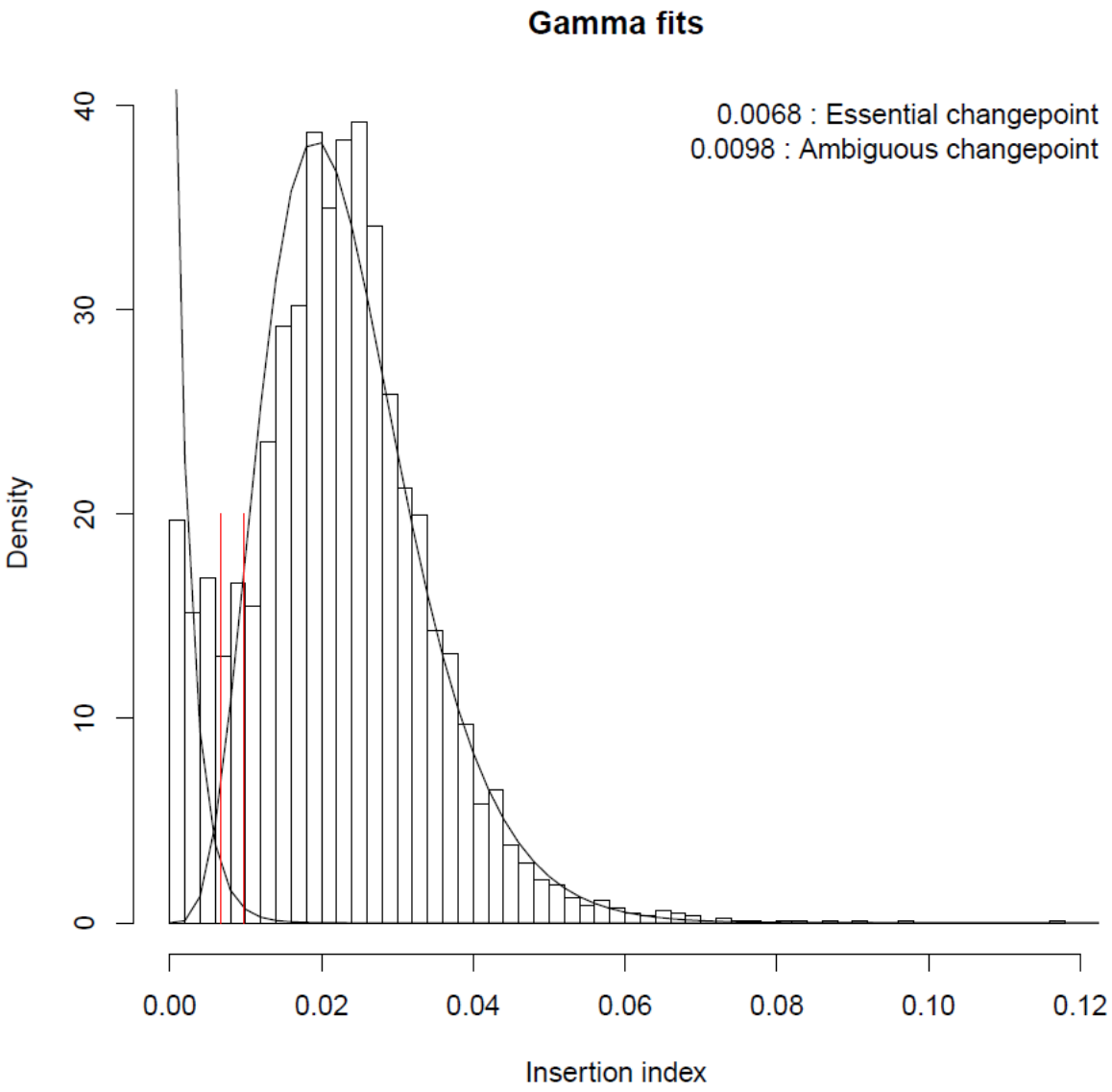


Figure 8.2 Histogram of Genes and Insertion Index

Essential genes are represented on the left of the histogram. The leftmost red line represents the cut-off point for assigning essentiality. The second red line separates ambiguous genes from non-essential genes.

Cut-off points were calculated using the tradis_essentiality.R script included in the TraDIS toolkit (Barquist et al., 2016). The script takes account of TraDIS data generated in *Salmonella typhi* and *S. typhimurium* which when compared to the Keio essentiality score (Baba et al., 2006) in related *E. coli* genes was accurate with a false positive rate of ~2.7%.

Table 8.3 List of candidate essential genes of *C. autoethangenum* for growth on YTF medium

The script TraDIS_essentiality.R produced predictions of gene essentiality based on the bimodal distribution of insertion sites over genes as described in (Barquist et al., 2016).

Number	Gene Name	Description
1	CLAU_0033	Flavodoxin
2	CLAU_0137	PTS system fructose subfamily IIC subunit
3	CLAU_0138	1-phosphofructokinase
4	CLAU_0139	Transcriptional regulator DeoR family
5	CLAU_0172	hypothetical protein
6	CLAU_0174	hypothetical protein
7	CLAU_0178	Integrase family protein
8	lexA	SOS-response transcriptional repressor lexA
9	CLAU_0194	Hypothetical protein
10	CLAU_0198	Methionine gamma-lyase
11	miaA	TRNA dimethylallyltransferase
12	CLAU_0215	Glycosyl transferase
13	CLAU_0243	FeoA family protein
14	argS	Arginyl-tRNA synthetase
15	CLAU_0265	Small acid-soluble spore protein alpha/beta type
16	CLAU_0345	CRISPR/Cas system-associated protein Cas8a1
17	hypD	Hydrogenase expression/formation protein HypD
18	mnmA	TRNA-specific 2-thiouridylase MnmA
19	CLAU_0385	Cysteine desulfurase
20	CLAU_0488	Dihydrofolate reductase
21	CLAU_0491	Hypothetical protein
22	CLAU_0522	NADPH-dependent FMN reductase
23	CLAU_0535	Copper ion binding protein
24	moaC	Molybdenum cofactor biosynthesis C
25	CLAU_0553	putative membrane protein
26	CLAU_0554	Lipoteichoic acid synthase 2
27	citG	2-(5''-triphosphoribosyl)-3'-dephosphocoenzyme-A synthase
28	CLAU_0601	Hypothetical protein
29	CLAU_0603	Hypothetical protein
30	CLAU_0671	Transcriptional regulator AsnC family
31	CLAU_0723	Hypothetical protein
32	CLAU_0729	Glutaminyl-tRNA synthetase
33	CLAU_0733	DegT/DnrJ/EryC1/StrS aminotransferase
34	CLAU_0734	Glutamine-scyllo-inositol transaminase
35	CLAU_0745	hypothetical protein
36	CLAU_0774	Response regulator
37	CLAU_r001	
38	CLAU_t001	
39	CLAU_t002	

40	CLAU_0866	Fructose-1
41	CLAU_0883	Short-chain dehydrogenase/reductase SDR
42	CLAU_0913	Hypothetical protein
43	CLAU_1019	GntR family transcriptional regulator
44	polC	DNA polymerase III polC-type
45	CLAU_1055	Hypothetical protein
46	CLAU_t004	
47	CLAU_1073	Cobalt-precorrin-6A reductase
48	CLAU_1095	GntR family transcriptional regulator
49	ddl	D-alanine-D-alanine ligase
50	CLAU_1101	Alanine racemase
51	CLAU_1103	Metal sensitive transcriptional repressor
52	CLAU_1109	Hypothetical protein
53	CLAU_1122	Primosome DnaD subunit
54	CLAU_t005	
55	CLAU_1171	Transcriptional regulator MarR family
56	addA	ATP-dependent helicase
57	cspA	Cold-shock DNA-binding domain protein
58	CLAU_t006	
59	CLAU_1219	Lytic transglycosylase
60	coaE	Dephospho-CoA kinase
61	polA	DNA polymerase I
62	aspS	Aspartyl-tRNA synthetase
63	hisS	Histidyl-tRNA synthetase
64	CLAU_1237	Hypothetical protein
65	ruvA	Holliday junction DNA helicase ruvA
66	CLAU_1261	Glycosyl transferase WecB/TagA/CpsF family
67	CLAU_1262	Undecaprenyl-phosphate galactosephosphotransferase
68	CLAU_1263	Acyltransferase 3
69	CLAU_1266	Hypothetical protein
70	CLAU_1267	UDP-N-acetylglucosamine 2-epimerase
71	CLAU_1268	Transferase hexapeptide repeat-containing protein
72	CLAU_1269	Oxidoreductase
73	CLAU_1270	UDP-N-acetyl-D-glucosamine dehydrogenase
74	CLAU_1271	Maltose-binding protein
75	CLAU_1273	Acyltransferase 3
76	CLAU_1274	Hypothetical protein
77	CLAU_1275	Glycosyl transferase family 2
78	CLAU_1276	Glycosyl transferase
79	CLAU_1277	Glycosyl transferase group 1
80	pheT	Phenylalanyl-tRNA synthetase subunit beta
81	pheS	Phenylalanyl-tRNA synthetase alpha subunit
82	rplT	50S ribosomal protein L20
83	infC2	Translation initiation factor IF-3
84	thrS	Threonyl-tRNA synthetase

85	CLAU_t007	
86	CLAU_t008	
87	CLAU_t009	
88	CLAU_t010	
89	dapH	2
90	CLAU_1323	Hypothetical protein
91	CLAU_1325	Valyl-tRNA synthetase
92	CLAU_1412	YbaK/prolyl-tRNA synthetase associated region
93	engB	GTP-binding protein engB
94	clpX	ATP-dependent Clp protease ATP-binding subunit clpX
95	CLAU_1446	Methionine aminopeptidase type I
96	CLAU_1481	Hypothetical protein
97	CLAU_1486	Chromate transporter
98	CLAU_1502	Hypothetical protein
99	CLAU_t012	
100	CLAU_t013	
101	CLAU_t014	
102	CLAU_t015	
103	CLAU_t017	
104	CLAU_t018	
105	CLAU_t019	
106	CLAU_t021	
107	CLAU_t022	
108	CLAU_t023	
109	CLAU_t025	
110	CLAU_1516	Hypothetical protein
111	CLAU_1517	DNA ligase
112	CLAU_1531	Chaperonin
113	CLAU_1532	Chaperonin
114	CLAU_1539	Glutamate synthase (NADPH) homotetrameric
115	CLAU_1554	O-sialoglycoprotein endopeptidase
116	CLAU_1555	Endodeoxyribonuclease RusA
117	CLAU_1570	Cobyrinic acid a
118	CLAU_1571	Dihydrolipoamide dehydrogenase
119	CLAU_1576	Formate-tetrahydrofolate ligase
120	CLAU_1579	Carbon-monoxide dehydrogenase (acceptor)
121	CLAU_1609	hypothetical protein
122	CLAU_t026	
123	CLAU_1632	ATPase putative protein family UPF0079
124	CLAU_1644	Tryptophan-tRNA ligase bacterial-type
125	CLAU_1651	Prolyl-tRNA synthetase
126	smpB	SsrA-binding protein
127	CLAU_1709	Preprotein translocase SecG subunit
128	eno	Enolase
129	gpmI	2

130	pgk	Phosphoglycerate kinase
131	CLAU_1715	Transcriptional regulator DeoR family
132	CLAU_t027	
133	CLAU_1726	Phospholipid/glycerol acyltransferase
134	glmS1	Glucosamine-fructose-6-phosphate aminotransferase (isomerizing)
135	CLAU_1848	Phosphoenolpyruvate-protein phosphotransferase
136	CLAU_1851	Citrate lyase alpha subunit
137	CLAU_1852	Citryl-CoA lyase
138	CLAU_1853	Citrate lyase acyl carrier protein
139	CLAU_1854	Hydro-lyase Fe-S type tartrate/fumarate subfamily beta subunit
140	CLAU_1856	Methylaspartate ammonia-lyase
141	glmS2	Methylaspartate mutase S chain
142	rpsI	Ribosomal protein S9
143	rplM	Ribosomal protein L13
144	rplQ	50S ribosomal protein L17
145	rpoA	DNA-directed RNA polymerase subunit alpha
146	rpsD	Ribosomal protein S4
147	rpmJ	50S ribosomal protein L36
148	adK	Adenylate kinase
149	rplO	Ribosomal protein L15
150	rpmD	Ribosomal protein L30
151	rpsE	Ribosomal protein S5
152	rplF	Ribosomal protein L6
153	rplE	Ribosomal protein L5
154	rplX	Ribosomal protein L24
155	rplN	Ribosomal protein L14
156	rpsQ	Ribosomal protein S17
157	rplP	50S ribosomal protein L16
158	rplV	Ribosomal protein L22
159	rpsS	Ribosomal protein S19
160	rplD	Ribosomal protein L4/L1e
161	rplC	Ribosomal protein L3
162	rpsJ	30S ribosomal protein S10
163	fusA	Translation elongation factor G
164	rpsG	Ribosomal protein S7
165	rpsL	Ribosomal protein S12
166	rpoC	DNA-directed RNA polymerase subunit beta
167	rpoB	DNA-directed RNA polymerase subunit beta
168	rplA	Ribosomal protein L1
169	secE	Preprotein translocase SecE subunit
170	rpmG	50S ribosomal protein L33
171	CLAU_t029	
172	cysS2	Cysteinyl-tRNA synthetase
173	ispD	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
174	CLAU_t030	

175	CLAU_t031	
176	CLAU_t032	
177	murD	UDP-N-acetylmuramoylalanine-D-glutamate ligase
178	glyQS	Glycine-tRNA ligase
179	CLAU_t033	
180	lysS	Lysyl-tRNA synthetase
181	CLAU_t034	
182	CLAU_t037	
183	CLAU_t038	
184	CLAU_1944	Septum formation initiator
185	prs	Ribose-phosphate pyrophosphokinase
186	glmU	Bifunctional protein glmU
187	murC	UDP-N-acetylmuramate-L-alanine ligase
188	CLAU_t039	
189	murI	Glutamate racemase
190	murA1	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
191	asnS	Asparaginyl-tRNA synthetase
192	CLAU_1993	Acetyl-CoA/biotin carboxylase
193	CLAU_1995	Acetyl-CoA carboxylase/biotin carboxyl carrier protein
194	CLAU_1996	3-oxoacyl-(acyl-carrier-protein) synthase 2
195	fabD	ACP S-malonyltransferase
196	fabH	3-oxoacyl-(acyl-carrier-protein) synthase 3
197	CLAU_2001	Transcriptional regulator MarR family
198	CLAU_2002	2-nitropropane dioxygenase NPD
199	CLAU_2003	Primosome DnaD subunit
200	CLAU_2004	IstB domain protein ATP-binding protein
201	CLAU_t040	
202	CLAU_t041	
203	CLAU_t042	
204	CLAU_t043	
205	CLAU_2029	Hypothetical protein
206	CLAU_t044	
207	CLAU_2045	Replicative DNA helicase
208	rpsR	30S ribosomal protein S18
209	rpsF	30S ribosomal protein S6
210	CLAU_2054	hypothetical protein
211	rpmH	50S ribosomal protein L34
212	dnaA	Chromosomal replication initiator protein dnaA
213	recF	DNA replication and repair protein recF
214	CLAU_2076	Hypothetical protein
215	gyrB	DNA gyrase B subunit
216	gyrA	DNA gyrase subunit A
217	CLAU_r009	
218	serS	Seryl-tRNA synthetase
219	CLAU_t046	

220	CLAU_t047	
221	CLAU_t048	
222	psd	Phosphatidylserine decarboxylase proenzyme
223	CLAU_t049	
224	CLAU_t050	
225	CLAU_2146	DNA polymerase III subunits gamma and tau
226	CLAU_2147	Nucleoid-associated protein
227	recR	Recombination protein recR
228	CLAU_2155	Hypothetical protein
229	CLAU_r012	
230	CLAU_2172	PpiC-type peptidyl-prolyl cis-trans isomerase
231	CLAU_2173	Aldose 1-epimerase
232	CLAU_2174	L-arabinose isomerase
233	CLAU_2175	L-ribulose-5-phosphate 4-epimerase
234	CLAU_2176	Carbohydrate kinase FGGY
235	CLAU_2177	Transcriptional regulator GntR family
236	CLAU_2178	Alpha-L-arabinofuranosidase domain protein
237	CLAU_2179	Sugar (Glycoside-Pentoside-Hexuronide) transporter
238	CLAU_2180	Periplasmic binding protein domain-containing protein
239	rbsA	Ribose import ATP-binding protein rbsA
240	CLAU_2182	ABC-type transporter integral membrane subunit
241	CLAU_r018	
242	CLAU_2183	Hypothetical protein
243	CLAU_t054	
244	CLAU_2190	Thymidylate kinase
245	CLAU_2192	DNA polymerase III gamma subunit domain III
246	CLAU_2195	4Fe-4S ferredoxin iron-sulpur binding domain-containing protein
247	CLAU_2202	Transcriptional regulator AbrB family
248	ispF	2-C-methyl-D-erythritol 2
249	metG	Methionyl-tRNA synthetase
250	ileS	Isoleucyl-tRNA synthetase
251	ispE	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
252	pyrG	CTP synthase
253	rpmE	50S ribosomal protein L31
254	CLAU_2273	Peptide chain release factor 1
255	CLAU_2278	Sugar-phosphate isomerase RpiB/LacA/LacB family
256	CLAU_2281	CMP/dCMP deaminase zinc-binding protein
257	CLAU_2282	Glycosyl transferase family 4
258	CLAU_2283	UDP-N-acetylglucosamine 2-epimerase
259	atpB	ATP synthase subunit a
260	CLAU_2287	ATP synthase subunit b
261	atpG	ATP synthase gamma chain
262	CLAU_2298	Cell shape determining protein MreB/Mrl family
263	CLAU_t055	
264	CLAU_t056	

265	CLAU_t057	
266	CLAU_t058	
267	CLAU_t059	
268	metK	S-adenosylmethionine synthase
269	secA	Protein translocase subunit secA
270	CLAU_2306	Peptide chain release factor 2
271	CLAU_2313	Type III secretion exporter
272	CLAU_2314	Calcineurin-like phosphoesterase superfamily domain-containing protein
273	CLAU_t060	
274	CLAU_2322	Fructose-1
275	acpS	Holo-(acyl-carrier-protein) synthase
276	CLAU_2359	1-deoxy-D-xylulose-5-phosphate synthase
277	murB	UDP-N-acetylenolpyruvoylglucosamine reductase
278	CLAU_2373	putative protein family UPF0052
279	dnaE	DNA polymerase III alpha subunit
280	pyk	Pyruvate kinase
281	CLAU_2433	Transcriptional regulator AraC family
282	CLAU_2440	Transcriptional regulator MarR family
283	CLAU_2490	Cell wall binding repeat 2-containing protein
284	CLAU_2492	Cell wall binding repeat 2-containing protein
285	CLAU_2529	hypothetical protein
286	CLAU_2631	Ferric uptake regulator Fur family
287	CLAU_2643	CobB/CobQ domain protein glutamine amidotransferase
288	CLAU_2644	UDP-N-acetlymuramyl peptide synthase
289	CLAU_t061	
290	CLAU_t062	
291	CLAU_r021	
292	CLAU_t063	
293	CLAU_t064	
294	CLAU_t065	
295	CLAU_t066	
296	CLAU_2673	1-acyl-sn-glycerol-3-phosphate acyltransferase
297	CLAU_2686	Transcriptional regulator AsnC family
298	CLAU_2705	NAD(+) synthetase
299	CLAU_2711	ATP-dependent helicase/deoxyribonuclease subunit B
300	CLAU_2717	NADH-quinone oxidoreductase E subunit
301	CLAU_2734	Septum formation protein Maf
302	CLAU_2736	Cell shape determining protein MreB/Mrl family
303	CLAU_2737	Rod shape-determining protein MreC
304	CLAU_2739	Penicillin-binding Protein dimerization domain
305	CLAU_2742	Cell division topological specificity factor MinE
306	CLAU_2743	putative rod shape-determining protein RodA
307	CLAU_2751	hypothetical protein
308	CLAU_2752	50S ribosomal protein L27

309	CLAU_2761	L-seryl-tRNA(Sec) selenium transferase
310	CLAU_2780	Hypothetical protein
311	CLAU_2803	DNA polymerase III delta subunit
312	CLAU_2804	30S ribosomal protein S20
313	hrcA	Heat-inducible transcription repressor HrcA
314	CLAU_2811	Heat shock protein grpE
315	dnaK	Molecular chaperone DnaK
316	rpsU	30S ribosomal protein S21
317	CLAU_2823	Heat-shock protein
318	CLAU_2825	GTPase Era
319	recO	DNA repair protein recO
320	CLAU_2836	DNA primase
321	rpoD	RNA polymerase sigma factor
322	CLAU_2867	N5-carboxyaminoimidazole ribonucleotide mutase
323	CLAU_2947	Pyruvate ferredoxin/flavodoxin oxidoreductase
324	CLAU_2949	HAD family hydrolase
325	flgM	Flagellar biosynthesis anti-sigma factor protein FlgM
326	CLAU_2973	Flagellin domain protein
327	CLAU_3050	hypothetical protein
328	CLAU_3055	putative signal transduction protein with CBS domain-containing protein
329	ftsL	Cell division protein FtsL family protein
330	murE	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2
331	murF	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase
332	mraY	Phospho-N-acetylmuramoyl-pentapeptide-transferase
333	CLAU_3068	Stage V sporulation protein E
334	CLAU_3069	Polypeptide-transport-associated domain protein FtsQ-type
335	CLAU_3073	hypothetical protein
336	sepF	Cell division protein sepF
337	CLAU_3078	5'-methylthioadenosine nucleosidase / S-adenosylhomocysteine nucleosidase
338	dapF	Diaminopimelate epimerase
339	efp	Elongation factor P
340	dxs	1-deoxy-D-xylulose-5-phosphate synthase
341	nadK	NAD kinase
342	CLAU_3155	Metal sensitive transcriptional repressor
343	CLAU_3167	6-phosphogluconate dehydrogenase
344	CLAU_3201	Ethanolamine utilization protein EutN/carboxysome structural protein Ccml
345	CLAU_3207	Microcompartments protein
346	CLAU_3210	Cysteine desulfurase
347	alaS	Alanine-tRNA ligase
348	CLAU_3215	UPF0297 protein
349	CLAU_3216	Holliday junction resolvase
350	CLAU_3219	Ribocuclease J
351	CLAU_3226	Twitching motility protein
352	ftsA	Cell division protein FtsA

353	ftsZ	Cell division protein FtsZ
354	CLAU_3243	putative phosphate transport regulator
355	CLAU_3244	Phosphate transporter
356	CLAU_3250	hypothetical protein
357	gmk	Guanylate kinase
358	rpoZ	DNA-directed RNA polymerase subunit omega
359	CLAU_3253	Phosphopantothenoylcysteine decarboxylase/phosphopantothenate/cysteine ligase
360	CLAU_3254	Primosomal protein N
361	fmt	Methionyl-tRNA formyltransferase
362	CLAU_3261	Serine/threonine protein kinase with PASTA sensor(s)
363	rpmB	50S ribosomal protein L28
364	coaD	Phosphopantetheine adenylyltransferase
365	CLAU_3271	hypothetical protein
366	CLAU_3274	Phosphate acetyltransferase
367	ackA	Acetate kinase
368	rpmF	50S ribosomal protein L32
369	acpP	Acyl carrier protein
370	ftsY	Cell division transporter substrate-binding protein FtsY
371	CLAU_3285	hypothetical protein
372	ffh	Signal recognition protein
373	rpsP	30S ribosomal protein S16
374	CLAU_3288	hypothetical protein
375	rimM	Ribosome maturation factor rimM
376	trmD	TRNA (guanine-N(1)-)-methyltransferase
377	rplS	50S ribosomal protein L19
378	codY	GTP-sensing pleiotropic transcriptional repressor CodY
379	rpsB	Ribosomal protein S2
380	tsf	Elongation factor Ts
381	pyrH	Uridylate kinase
382	frr	Ribosome-recycling factor
383	CLAU_3305	Undecaprenyl pyrophosphate synthase
384	CLAU_3306	Phosphatidate cytidylyltransferase
385	dxr	1-deoxy-D-xylulose 5-phosphate reductoisomerase
386	ispG	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
387	nusA	Transcription elongation factor NusA
388	CLAU_3312	Hypothetical protein
389	infB	Translation initiation factor IF-2
390	ribF	Bifunctional riboflavin kinase/FMN adenylyltransferase
391	pnpA	Polynucleotide phosphorylase
392	recA	Protein recA
393	ubiA	Prenyltransferase
394	CLAU_3334	Phosphotransferase system phosphocarrier protein HPr
395	purB	Adenylosuccinate lyase
396	gltX	Glutamate-tRNA ligase

397	CLAU_3342	Nicotinate phosphoribosyltransferase
398	CLAU_3369	hypothetical protein
399	CLAU_3390	Hypothetical protein
400	CLAU_3400	hypothetical protein
401	CLAU_3409	putative transcriptional regulator
402	CLAU_3416	Cupin 2 conserved barrel domain protein
403	CLAU_3455	AAA-ATPase-like protein
404	CLAU_3462	hypothetical protein
405	CLAU_3464	Transcriptional regulator XRE family
406	CLAU_3494	Transcriptional regulator LclR family
407	CLAU_t067	
408	CLAU_r027	
409	CLAU_3516	putative NUDIX hydrolase
410	CLAU_3571	C_GCAxxG_C_C family protein
411	CLAU_3669	putative transcription regulator
412	CLAU_3682	Hypothetical protein
413	CLAU_3709	Hypothetical protein
414	CLAU_3755	Amidohydrolase
415	CLAU_3781	Hypothetical protein
416	CLAU_3867	Hypothetical protein
417	CLAU_3880	Hypothetical protein
418	CLAU_3903	Cupin 2 conserved barrel domain protein
419	CLAU_3956	Cys/Met metabolism pyridoxal-phosphate-dependent protein

Table 8.4 List of ambiguous genes

The script tradis_essentiality.R produces predictions of gene essentiality based on a bimodal distribution of insertion sites over genes, but some genes appear in the 'ambiguous' region between the two peaks, these genes are listed below.

Number	Gene Name	Description
1	CLAU_0014	Transcriptional regulator XRE family
2	CLAU_0034	Cold-shock DNA-binding domain protein
3	CLAU_0049	Hypothetical protein
4	CLAU_0179	Hypothetical protein
5	miaB	(Dimethylallyl)adenosine tRNA methylthiotransferase miaB
6	speH	S-adenosylmethionine decarboxylase proenzyme
7	CLAU_0278	Sporulation stage 0 Spo0E-like regulatory phosphatase
8	CLAU_0330	Succinate dehydogenase/fumarate reductase
9	hypC	Hydrogenase assembly chaperone hypC/hupF
10	nifH2	Nitrogenase component 2 NifH ATPase
11	CLAU_0416	Pyridoxamine 5-phosphate oxidase-related FMN-binding protein
12	cysK	Cysteine synthase A
13	CLAU_0520	Small acid-soluble spore protein alpha/beta type

14	CLAU_0524	Transcriptional regulator XRE family
15	CLAU_0552	Molybdenum cofactor synthesis domain-containing protein
16	CLAU_0569	Ornithine carbamoyltransferase
17	CLAU_0636	putative zinc-finger-containing protein
18	CLAU_0639	putative secreted protein containing pentapeptide repeat
19	CLAU_0647	MotA/TolQ/ExbB proton channel
20	CLAU_0775	Sensory transduction histidine kinase
21	CLAU_0801	Hypothetical protein
22	CLAU_0813	Sporulation stage 0 Spo0E-like regulatory phosphatase
23	CLAU_0897	hypothetical protein
24	CLAU_1038	Hypothetical protein
25	CLAU_1046	Hypothetical protein
26	CLAU_1054	Hypothetical protein
27	CLAU_1077	Precorrin-4 C(11)-methyltransferase
28	CLAU_1093	Cobalt permease
29	CLAU_1130	Hypothetical protein
30	CLAU_1141	Transcriptional regulator
31	CLAU_1181	Cobalamin (vitamin B12) biosynthesis CbiM protein
32	CLAU_1193	Hypothetical protein
33	CLAU_1201	Coat F domain protein
34	CLAU_1231	(p)ppGpp synthetase I
35	ruvB	Holliday junction DNA helicase ruvB
36	CLAU_1249	Transcriptional regulator GntR family
37	CLAU_1265	Glycosyl transferase
38	capD	Polysaccharide biosynthesis protein CapD
39	CLAU_1295	Iron (metal) dependent repressor DtxR family
40	rpmI	50S ribosomal protein L35
41	CLAU_1415	Di-trans
42	CLAU_1417	hypothetical protein
43	CLAU_1418	hypothetical protein
44	CLAU_1448	putative zinc-binding domain-containing protein
45	CLAU_1449	Hypothetical protein
46	CLAU_1503	Hypothetical protein
47	CLAU_1526	Transcriptional regulator HxlR family
48	CLAU_1529	GMP synthase (glutamine-hydrolyzing)
49	CLAU_1534	Iron-dependent hydrogenase
50	CLAU_1563	RNA-binding protein
51	CLAU_1564	Ferredoxin
52	gcvH	Glycine cleavage system H protein
53	CLAU_1567	Dihydropteroate synthase DHPS
54	CLAU_1568	CO dehydrogenase/acetyl-CoA synthase delta subunit TIM barrel
55	CLAU_1569	CO dehydrogenase/acetyl-CoA synthase delta subunit TIM barrel
56	CLAU_1584	Hypothetical protein
57	CLAU_1626	Hypothetical protein
58	CLAU_1628	Transcriptional regulator XRE family

59	CLAU_1638	putative zinc-binding domain-containing protein
60	CLAU_1648	(FeFe)-hydrogenase maturation hydE
61	CLAU_1656	Hypothetical protein
62	CLAU_1706	Hypothetical protein
63	tpiA	Triosephosphate isomerase
64	CLAU_1714	Glyceraldehyde-3-phosphate dehydrogenase type I
65	CLAU_1728	Serine O-acetyltransferase
66	CLAU_1765	C_GCAxxG_C_C family protein
67	glmM	Phosphoglucosamine mutase
68	CLAU_1850	Citrate lyase ligase
69	CLAU_1855	Fumarate hydratase
70	glmL	GlmL
71	rpsK	30S ribosomal protein S11
72	rpsM	Ribosomal protein S13
73	infA	Translation initiation factor IF-1
74	secY	Protein translocase subunit SecY
75	rplR	Ribosomal protein L18
76	rpsH	Ribosomal protein S8
77	rpsZ	Ribosomal protein S14 type Z
78	rpmC	50S ribosomal protein L29
79	rpsC	Ribosomal protein S3
80	rplB	50S Ribosomal protein L2
81	rplK	Ribosomal protein L11
82	CLAU_1916	RNA polymerase sigma-24 subunit SigH ECF subfamily
83	tils	TRNA(Ile)-lysidine synthase
84	pth	Peptidyl-tRNA hydrolase
85	CLAU_1956	Integral membrane sensor signal transduction histidine kinase
86	accA	Acetyl-CoA carboxylase carboxyl transferase subunit alpha
87	accD	Acetyl-CoA carboxylase carboxyl transferase subunit beta
88	CLAU_2018	Cupin 2 conserved barrel domain protein
89	CLAU_2052	Single-strand binding protein
90	CLAU_2073	DNA polymerase III beta subunit
91	CLAU_2142	ABC-type transporter related protein
92	CLAU_2193	PSP1 domain protein
93	CLAU_2209	Hypothetical protein
94	CLAU_2215	Hypothetical protein
95	CLAU_2230	Hypothetical protein
96	CLAU_2284	Hypothetical protein
97	atpH	ATP synthase subunit delta
98	atpA	ATP synthase subunit alpha
99	atpD	ATP synthase subunit beta
100	atpC	ATP synthase epsilon chain
101	leuS	Leucine-tRNA ligase
102	CLAU_2326	Transposase IS111A/IS1328/IS1533
103	CLAU_2335	Transcriptional regulator IclR family

104	pssA	CDP-diacylglycerol/serine O-phosphatidyltransferase
105	CLAU_2358	Transketolase
106	CLAU_2392	hypothetical protein
107	CLAU_2405	putative RNA methylase family protein
108	CLAU_2477	hypothetical protein
109	CLAU_2611	hypothetical protein
110	CLAU_2639	HTH domain-containing protein
111	CLAU_2690	Transcriptional regulator GntR family with UTRA sensor domain-containing protein
112	CLAU_2692	Penicillinase repressor
113	CLAU_2731	Acetyltransferase (GNAT) domain
114	CLAU_2738	Rod shape-determining protein MreD
115	CLAU_2750	50S ribosomal L21 protein
116	CLAU_2753	GTPase obg
117	CLAU_2754	putative nicotinate-nucleotide adenylyltransferase
118	CLAU_2755	Metal dependent phosphohydrolase
119	CLAU_2762	Translation elongation factor
120	CLAU_2765	putative cobalamin-binding protein
121	CLAU_2770	putative cobalamin-binding protein
122	CLAU_2828	Transcriptional regulator containing CBS sensor domains
123	CLAU_2848	Hypothetical protein
124	CLAU_2923	Nitrate reductase subunit beta
125	CLAU_2946	UDP-N-acetylglucosamine transferase
126	CLAU_3012	putative motility protein
127	ychF	Ribosome-binding ATPase ychF
128	CLAU_3091	putative SAM-dependent methyltransferase
129	CLAU_3096	hypothetical protein
130	CLAU_3112	Stage III sporulation protein AC
131	der	GTP-binding protein engA
132	CLAU_3249	hypothetical protein
133	rsgA	Ribosome biogenesis GTPase RsgA
134	CLAU_3263	Ribulose-phosphate 3-epimerase
135	CLAU_3264	Thiamine pyrophosphokinase
136	nrnA	3'(2')
137	pgsA	CDP-diacylglycerol/glycerol-3-phosphate 3-phosphatidyltransferase
138	rny	Ribonuclease Y
139	CLAU_3333	Aspartate transaminase
140	CLAU_3362	Hypothetical protein
141	CLAU_3365	putative transcriptional regulator
142	kdpD	Histidine kinase
143	CLAU_3393	Coenzyme PQQ synthesis protein D
144	CLAU_3435	hypothetical protein
145	CLAU_3441	hypothetical protein
146	CLAU_3481	NADH dehydrogenase
147	CLAU_3568	Transcriptional regulator MarR family

148	CLAU_3572	C_GCAxxG_C_C family protein
149	CLAU_3718	Hypothetical protein
150	CLAU_3777	Hypothetical protein
151	CLAU_3785	2-dehydropantoate 2-reductase
152	CLAU_3794	HAD-superfamily hydrolase subfamily IA variant 1
153	CLAU_3843	Sugar kinase ROK family
154	CLAU_3895	Hypothetical protein
155	CLAU_3966	Hypothetical protein

9 References

Abrini, J., Naveau, H., & Nyns, E.-J. (1994). *Clostridium autoethanogenum*, sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide. Archives of Microbiology, 161(4), 345–351. https://doi.org/10.1007/bf00303591

Abubackar, H. N., Veiga, M. C., & Kennes, C. (2012). Biological conversion of carbon monoxide to ethanol: Effect of pH, gas pressure, reducing agent and yeast extract. Bioresource Technology, 114, 518–522. https://doi.org/10.1016/j.biortech.2012.03.027

Abubackar, H. N., Veiga, M. C., & Kennes, C. (2015). Carbon monoxide fermentation to ethanol by *Clostridium autoethanogenum* in a bioreactor with no accumulation of acetic acid. Bioresource Technology, 186, 122–127. https://doi.org/10.1016/j.biortech.2015.02.113

Akerley, B. J. (2016). No-Go Zones for Mariner Transposition Brian. mBio, 7(5), 5–7. https://doi.org/10.1128/mBio.01351-16

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. Applied and Environmental Microbiology, 78(22), 8112–8121. https://doi.org/10.1128/AEM.02214-12

Allen, T. D., Caldwell, M. E., Lawson, P. A., Huhnke, R. L., & Tanner, R. S. (2010). *Alkalibaculum bacchi* gen. nov., sp. nov., a CO-oxidizing, ethanol-producing acetogen isolated from livestock-impacted soil. International Journal of Systematic and Evolutionary Microbiology, 60(10), 2483–2489. https://doi.org/10.1099/ijs.0.018507-0

Aloui, A., Tagourti, J., El May, A., Joseleau Petit, D., & Landoulsi, A. (2011). The effect of methylation on some biological parameters in *Salmonella enterica* serovar Typhimurium. Pathol Biol (Paris), 59(4), 192–198. https://doi.org/10.1016/j.patbio.2009.03.011

Aranda, J., Poza, M., Pardo, B. G., Rumbo, S., Rumbo, C., Parreira, J. R., … Bou, G. (2010). A rapid and simple method for constructing stable mutants of *Acinetobacter baumannii*. BMC Microbiol, 10(1), 279. https://doi.org/10.1186/1471-2180-10-279

Awad, M. M., Bryant, A. E., Stevens, D. L., & Rood, J. I. (1995). Virulence studies on chromosomal alpha-toxin and theta-toxin mutants constructed by allelic exchange provide genetic evidence for the essential role of alpha-toxin in *Clostridium perfringens*-mediated gas gangrene. Molecular Microbiology, 15(2), 191–202. https://doi.org/10.1111/j.1365-2958.1995.tb02234.x

Ayarpadikannan, S., & Kim, H.-S. (2014). The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. Genomics & Informatics, 12(3), 98–104. https://doi.org/10.5808/GI.2014.12.3.98

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2006;2 2006.0008.

Banerjee, A., Leang, C., Ueki, T., Nevin, K. P., & Lovley, D. R. (2014). Lactose-inducible system for metabolic engineering of *clostridium ljungdahlii*. Applied and Environmental Microbiology, 80(8), 2410–2416. https://doi.org/10.1128/AEM.03666-13

Barquist, L., Boinett, C. J., & Cain, A. K. (2013). Approaches to querying bacterial genomes with transposon-insertion sequencing. RNA Biology, 10(7), 1161–1169. https://doi.org/10.4161/rna.24765

Barquist, L., Mayho, M., Cummins, C., Cain, A. K., Boinett, C. J., Page, A. J., … Parkhill, J. (2016). The TraDIS toolkit: Sequencing and analysis for dense transposon mutant libraries. Bioinformatics, 32(7), 1109–1111. https://doi.org/10.1093/bioinformatics/btw022

Barry, A., Wolfe, A., English, C., Ruddick, C., & Lambert, D. (2016). National Algal Biofuels Technology Review. Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office., (June), 212.

Baym, M., Shaket, L., Anzai, I. A., Adesina, O., & Barstow, B. (2016). Rapid construction of a whole-genome transposon insertion collection for *Shewanella*

oneidensis by Knockout Sudoku. Nature Communications, 7, 13270. https://doi.org/10.1038/ncomms13270

Behera, S., Singh, R., Arora, R., Sharma, N. K., Shukla, M., & Kumar, S. (2015). Scope of Algae as Third Generation Biofuels. Frontiers in Bioengineering and Biotechnology, 2(February), 1–13. https://doi.org/10.3389/fbioe.2014.00090

Bengelsdorf, F. R., Poehlein, A., Linder, S., Erz, C., Hummel, T., Hoffmeister, S., Dürre, P. (2016). Industrial acetogenic biocatalysts: A comparative metabolic and genomic analysis. Frontiers in Microbiology, 7(JUL), 1–15. https://doi.org/10.3389/fmicb.2016.01036

Bengelsdorf, F. R., Straub, M., & Dürre, P. (2013). Bacterial synthesis gas (syngas) fermentation. Environmental Technology, 34(13–14), 1639–1651. https://doi.org/10.1080/09593330.2013.827747

Bertram, J., & Dürre, P. (1989). Conjugal transfer and expression of streptococcal transposons in *Clostridium acetobutylicum*. Archives of Microbiology, 151(6), 551–557. https://doi.org/10.1007/BF00454874

Bestor, T. H. (1990). DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 326(1235), 179–87. https://doi.org/10.1098/rstb.1990.0002

Braakman, R., & Smith, E. (2012). The emergence and early evolution of biological carbon-fixation. PLoS Computational Biology, 8(4). https://doi.org/10.1371/journal.pcbi.1002455

Breault, R. W. (2010). Gasification processes old and new: A basic review of the major technologies. Energies, 3(2), 216–240. https://doi.org/10.3390/en3020216

Bristow, J., Butland, G., & Arkin, A. P. (2015). Sequencing Randomly Bar-Coded Transposons, 6(6), 1–15. https://doi.org/10.1128/mBio.00306-15.Editor

Brown, S. D., Nagaraju, S., Utturkar, S., De Tissera, S., Segovia, S., Mitchell, W., Land, L.M., Dassanayake, S. & Köpke, M. (2014). Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. Biotechnology for Biofuels, 7, 40. https://doi.org/10.1186/1754-6834-7-40

Burby, P. E., Nye, T. M., Schroeder, J. W., & Simmons, L. A. (2017). Implementation and Data Analysis of Sequencing for Bacterial Genetics, 199(1), 1–11. https://doi.org/10.1128/JB.00560-16

Burk, M. J., & Van Dien, S. (2016). Biotechnology for Chemical Production: Challenges and Opportunities. Trends in Biotechnology, 34(3), 187–190. https://doi.org/10.1016/j.tibtech.2015.10.007

Byrne, R. T., Chen, S. H., Wood, E. A., Cabot, E. L., & Cox, M. M. (2014). *Escherichia coli* genes and pathways involved in surviving extreme exposure to ionizing radiation. Journal of Bacteriology, 196(20), 3534–3545. https://doi.org/10.1128/JB.01589-14

Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T., & Minton, N. P. (2012). Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype and toxin production. Applied and

Environmental Microbiology, 78(13), 4683–4690. https://doi.org/10.1128/AEM.00249-12

Cartman, S. T., & Minton, N. P. (2010). A mariner-Based transposon system for in vivo random mutagenesis of *clostridium difficile*. Applied and Environmental Microbiology, 76(4), 1103–1109. https://doi.org/10.1128/AEM.02525-09

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 Biochemistry, 37(4), 411–421. https://doi.org/10.1016/S0032-9592(01)00227-8

Charbonneau, A. R. L., Forman, O. P., Cain, A. K., Newland, G., Robinson, C., Boursnell, M., … Waller, A. S. (2017). Defining the ABC of gene essentiality in streptococci. BMC Genomics, 18(1), 426. https://doi.org/10.1186/s12864-017-3794-3

Chaudhuri, R. R., Allen, A. G., Owen, P. J., Shalom, G., Stone, K., Harrison, M., … Charles, I. G. (2009). Comprehensive identification of essential Staphylococcus aureus genes using Transposon-Mediated Differential Hybridisation (TMDH). BMC Genomics, 10(1), 291. https://doi.org/10.1186/1471-2164-10-291

Chen, C. K., Boucle, C. M., & Blaschek, H. P. (1996). Factors involved in the transformation of previously non-transformable *Clostridium perfringens* type B. FEMS Microbiology Letters, 140(2–3), 185–191. https://doi.org/10.1016/0378-1097(96)00177-2

Chen, M., Smith, P. M., & Wolcott, M. P. (2016). U.S. Biofuels Industry: A Critical Review of Opportunities and Challenges. Bioproducts Business, 1(4), 42–59.

Clark, T. A., Murray, I. A., Morgan, R. D., Kislyuk, A. O., Spittle, K. E., Boitano, M., … Korlach, J. (2012). Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Research, 40(4). https://doi.org/10.1093/nar/gkr1146

Cotton, C. A., Edlich-Muth, C., & Bar-Even, A. (2017). Reinforcing carbon fixation: CO 2 reduction replacing and supporting carboxylation. Current Opinion in Biotechnology, 49, 49–56. https://doi.org/10.1016/j.copbio.2017.07.014

Craig, N. L. (1991). Tn7: a target site-specific transposon. Molecular Microbiology, 5(11), 2569–2573. https://doi.org/10.1111/j.1365-2958.1991.tb01964.x

Crimmins, G. T., Mohammadi, S., Green, E. R., Bergman, M. A., Isberg, R. R., & Mecsas, J. (2012). Identification of MrtAB, an ABC Transporter Specifically Required for *Yersinia pseudotuberculosis* to Colonize the Mesenteric Lymph Nodes. PLoS Pathogens, 8(8). https://doi.org/10.1371/journal.ppat.1002828

Croux, C., Nguyen, N.-P.-T., Lee, J., Raynaud, C., Saint-Prix, F., Gonzalez-Pajuelo, M., … Soucaille, P. (2016). Construction of a restriction-less, marker-less mutant useful for functional genomic and metabolic engineering of the biofuel producer *Clostridium acetobutylicum*. Biotechnology for Biofuels, 9(1), 23. https://doi.org/10.1186/s13068-016-0432-2

Cui, G. Z., Hong, W., Zhang, J., Li, W. L., Feng, Y., Liu, Y. J., & Cui, Q. (2012). Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation. Journal of Microbiological Methods, 89(3), 201–208. https://doi.org/10.1016/j.mimet.2012.02.015

Curcio, M. J., & Derbyshire, K. M. (2003). The outs and ins of transposition: from mu to kangaroo. Nature Reviews Molecular Cell Biology, 4(11), 865–877. https://doi.org/10.1038/nrm1241

Daas, M. J. A., van de Weijer, A. H. P., de Vos, W. M., van der Oost, J., & van Kranenburg, R. (2016). Isolation of a genetically accessible thermophilic xylan degrading bacterium from compost. Biotechnology for Biofuels, 9(1), 210. https://doi.org/10.1186/s13068-016-0618-7

DeJesus, M.A., Gerrick, E.R., Xu, W.Z., Park, S.W., Long, J.E., Boutte, C.C., Rubin, E.J., Schnappinger, D., Ehrt, S., Fortune, S.M., et al. (2017). Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. Mbio 8.

Dembek, M., Barquist, L., Boinett, C., Cain, A., Mayho, M., Lawley, T., … Fagan, R. (2015). High-throughput analysis of gene essentiality and sporulation in Clostridium difficile. mBio, 6(2), 1–13. https://doi.org/10.1128/mBio.02383-14

Demirbas, A. (2007). Progress and recent trends in biofuels. Progress in Energy and Combustion Science, 33(1), 1–18. https://doi.org/10.1016/j.pecs.2006.06.001

Desai, R. P., & Papoutsakis, E. T. (1999). Antisense RNA strategies for metabolic engineering of *Clostridium acetobutylicum*. Applied and Environmental Microbiology, 65(3), 936–945.

Doyon, J. B., Pattanayak, V., Meyer, C. B., & Liu, D. R. (2006). Directed evolution and substrate specificity profile of homing endonuclease I-SceI. Journal of the American Chemical Society, 128(7), 2477–2484. https://doi.org/10.1021/ja057519l

Ellingsen, L. A. W., Majeau-Bettez, G., Singh, B., Srivastava, A. K., Valøen, L. O., & Strømman, A. H. (2014). Life Cycle Assessment of a Lithium-Ion Battery Vehicle Pack. Journal of Industrial Ecology, 18(1), 113–124. https://doi.org/10.1111/jiec.12072

Tanner,R. S., Miller, L. M., & Yang, D. (1993). Clostridial rRNA Homology Group I, 43(2), 232–236.

Erik, T., Aune, V., & Aachmann, F. L. (2010). Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed, 1301–1313. https://doi.org/10.1007/s00253-009-2349-1

Errington, J. (2003). Regulation of endospore formation in *Bacillus subtilis*. Nature Reviews Microbiology, 1(2), 117–126. https://doi.org/10.1038/nrmicro750

European Academies Science Advisory Council. (2012). The current status of biofuels in the European Union, their environmental impacts and future prospects; EASAC policy report 19. https://doi.org/ISBN: 978-3-8047-3118-9

Fargione, J. E., Cooper, T. R., Flaspohler, D. J., Hill, J., Lehman, C., Tilman, D., Coy, McCoy, T., McLeod, S, Nelson, E. J, Oberhauser, K. S., (2009). Bioenergy and Wildlife : Threats and Opportunities for Grassland Conservation, 59(9), 767–777. https://doi.org/10.1525/bio.2009.59.9.8

Fels, S. R., Zane, G. M., Blake, S. M., & Wall, J. D. (2013). Rapid transposon liquid enrichment sequencing (TnLE-seq) for gene fitness evaluation in underdeveloped bacterial systems. Applied and Environmental Microbiology, 79(23), 7510–7517. https://doi.org/10.1128/AEM.02051-13

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA 3rd, Venter, J. C. (1995). The Minimal Gene Complement of *Mycoplasma genitalium*. Science, 270(5235), 397–403. https://doi.org/10.1126/science.270.5235.397

Friedman, D. C., & Ellington, A. D. (2015). Industrialization of Biology. ACS Synthetic Biology (Vol. 4). https://doi.org/10.1021/acssynbio.5b00190

Gallagher, L. A., Shendure, J., & Manoil, C. (2011). Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. mBio, 2(1), 1–8. https://doi.org/10.1128/mBio.00315-10

Gawronski, J. D., Wong, S. M. S., 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. Proceedings of the National Academy of Sciences, 106(38), 16422–16427. https://doi.org/10.1073/pnas.0906627106

Geddes, C. C., Nieves, I. U., & Ingram, L. O. (2011). Advances in ethanol production. Current Opinion in Biotechnology, 22(3), 312–319. https://doi.org/10.1016/j.copbio.2011.04.012

Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., Clyde, A., Hutchison III, Hamilton O. Smith & Venter, J. C. (2006). Essential genes of a minimal bacterium. Proceedings of the National Academy of Sciences of the United States of America, 103(2), 425–430. https://doi.org/10.1073/pnas.0510013103

González-Cerón, G., Miranda-Olivares, O. J., & Servín-González, L. (2009). Characterization of the methyl-specific restriction system of Streptomyces coelicolor A3(2) and of the role played by laterally acquired nucleases. FEMS Microbiology Letters, 301(1), 35–43. https://doi.org/10.1111/j.1574-6968.2009.01790.x

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, 6(3), 279–289. https://doi.org/10.1016/j.chom.2009.08.003.Identifying

Goryshin, I. Y., Jendrisak, J., Hoffman, L. M., Meis, R., & Reznikoff, W. S. (2000). Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat Biotechnol, 18(1), 97–100. https://doi.org/10.1038/72017

Griffin, D. W., Schultz, M. A., Irving, E., & Road, P. (2012). Fuel and Chemical Products from Biomass Syngas : A Comparison of Gas Fermentation to Thermochemical Conversion Routes, 31(2), 219–224. https://doi.org/10.1002/ep

Griffin, J. E., Gawronski, J. D., DeJesus, M. A., Ioerger, T. R., Akerley, B. J., & Sassetti, C. M. (2011). High-resolution phenotypic profiling defines genes

essential for mycobacterial growth and cholesterol catabolism. PLoS Pathogens, 7(9), 1–9. https://doi.org/10.1371/journal.ppat.1002251

Guo, Y., Xu, J., Zhang, Y., Xu, H., Yuan, Z., & Li, D. (2010). Medium optimization for ethanol production with *Clostridium autoethanogenum* with carbon monoxide as sole carbon source. Bioresource Technology, 101(22), 8784–8789. https://doi.org/10.1016/j.biortech.2010.06.072

Guss, A.M., Olson, D.G., Caiazza, N.C., and Lynd, L.R. (2012). Dcm methylation is detrimental to plasmid transformation in Clostridium thermocellum. Biotechnology for Biofuels 5.

Handler, R. M., Shonnard, D. R., Griffing, E. M., Lai, A., & Palou-Rivera, I. (2016). Life Cycle Assessments of Ethanol Production via Gas Fermentation: Anticipated Greenhouse Gas Emissions for Cellulosic and Waste Gas Feedstocks. Industrial and Engineering Chemistry Research, 55(12), 3253–3261. https://doi.org/10.1021/acs.iecr.5b03215

Hare, R.S., Walker, S.S., Dorman, T.E., Greene, J.R., Guzman, L.M., Kenney, T.J., Sulavik, M.C., Baradaran, K., Houseweart, C., Yu, H.Y., et al. (2001). Genetic footprinting in bacteria. Journal of Bacteriology 183, 1694-1706.

Harris, L. M., Welker, N. E., & Papoutsakis, E. T. (2002a). Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. Journal of Bacteriology, 184(13), 3586–3597. https://doi.org/10.1128/JB.184.13.3586-3597.2002

Harris, L. M., Welker, N. E., & Papoutsakis, E. T. (2002b). Northern, Morphological, and Fermentation Analysis of spo0A Inactivation and

Overexpression in *Clostridium acetobutylicum* ATCC 824, 184(13), 3586–3597. https://doi.org/10.1128/JB.184.13.3586

Harrison, R. P., Ruck, S., Medcalf, N., & Rafiq, Q. A. (2017). Decentralized manufacturing of cell and gene therapies: Overcoming challenges and identifying opportunities. Cytotherapy, 19(10), 1140–1151. https://doi.org/10.1016/j.jcyt.2017.07.005

Hartman, A. H., Liu, H., & Melville, S. B. (2011). Construction and characterization of a lactose-inducible promoter system for controlled gene expression in *Clostridium perfringens*. Applied and Environmental Microbiology, 77(2), 471–8. https://doi.org/10.1128/AEM.01536-10

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 Research, 40(8). https://doi.org/10.1093/nar/gkr1321

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. Journal of Microbiological Methods, 70(3), 452–464. https://doi.org/10.1016/j.mimet.2007.05.021

Heap, J. T., Pennington, O. J., Cartman, S. T., & Minton, N. P. (2009). A modular system for Clostridium shuttle plasmids. Journal of Microbiological Methods, 78(1), 79–85. https://doi.org/10.1016/j.mimet.2009.05.004

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E., & Holden, D. W. (1995). Simultaneous Identification of Bacterial Virulence Genes by Negative Selection Published by : American Association for the Advancement of Science

Stable URL : http://www.jstor.org/stable/2888285. Advancement Of Science, 269(5222), 400–403.

Hickman, A. B., & Dyda, F. (2016). DNA Transposition at Work. Chemical Reviews,116(20), 12758–12784. https://doi.org/10.1021/acs.chemrev.6b00003

Higgins, B. P., Popkowski, A. C., Caruana, P. R., & Karls, A. C. (2009). Site-specific insertion of IS492 in *Pseudoalteromonas atlantica*. Journal of Bacteriology, 191(20), 6408–6414. https://doi.org/10.1128/JB.00771-09

Hiom, K., Melek, M., & Gellert, M. (1998). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell, 94(4), 463–470. https://doi.org/10.1016/S0092-8674(00)81587-1

Ho, D. P., Ngo, H. H., & Guo, W. (2014). A mini review on renewable sources for biofuel. Bioresource Technology, 169(July), 742–749. https://doi.org/10.1016/j.biortech.2014.07.022

Hochman, G., Rajagopal, D., & Zilberman, D. (2010). Are biofuels the culprit? OPEC, food, and fuel. American Economic Review, 100(2), 183–187. https://doi.org/10.1257/aer.100.2.183

Humphreys, C.M., McLean, S., Schatschneider, S., Millat, T., Henstra, A.M., Annan, F.J., Breitkopf, R., Pander, B., Piatek, P., Rowe, P., *et al.* (2015). Whole genome sequence and manual annotation of *Clostridium autoethanogenum*, an industrially relevant bacterium. Bmc Genomics 16.

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M., & Darzins, A. (2008). Microalgal triacylglycerols as feedstocks for biofuel

production: Perspectives and advances. Plant Journal, 54(4), 621–639. https://doi.org/10.1111/j.1365-313X.2008.03492.x

Hutchison, C.A., Chuang, R.Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., et al. (2016). Design and synthesis of a minimal bacterial genome. Science 351.

Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O., and Venter, J.C. (1999). Global transposon mutagenesis and a minimal mycoplasma genome. Science 286, 2165-2169.

IPCC. (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report (AR5)(Cambridge Univ Press, New York), 1535. https://doi.org/10.1029/2000JD000115

Itaya, M., Fujita, K., Kuroki, A., & Tsuge, K. (2008). Bottom-up genome assembly using the *Bacillus subtilis* genome vector. Nature Methods, 5(1), 41–3. https://doi.org/10.1038/nmeth1143

Jeltsch, A. (2003). Maintenance of species identity and controlling speciation of bacteria: A new function for restriction/modification systems? Gene, 317(1–2), 13–16. https://doi.org/10.1016/S0378-1119(03)00652-8

John, R. P., Anisha, G. S., Nampoothiri, K. M., & Pandey, A. (2011). Micro and macroalgal biomass: A renewable source for bioethanol. Bioresource Technology, 102(1), 186–193. https://doi.org/10.1016/j.biortech.2010.06.139

Karlsson, S., Burman, L. G., & Åkerlund, T. (2008). Induction of toxins in *Clostridium difficile* is associated with dramatic changes of its metabolism. Microbiology, 154(11), 3430–3436. https://doi.org/10.1099/mic.0.2008/019778-0

Kashiwagi, K., Shibuya, S., Kuraishi, A., & Igarashi, K. (1997). Molecular Genetics : Excretion and Uptake of Putrescine by the PotE Protein in *Escherichia coli*, 272(10), 6318–6323.

Kieser, T., & Hopwood, D. A. (1991). Genetic manipulation of Streptomyces: integrating vectors and gene replacement. Methods in Enzymology, 204, 430–58. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1943784

Kimura, S., Hubbard, T. P., Davis, B. M., & Waldor, M. K. (2016). The nucleoid binding protein H-NS biases genome-wide transposon insertion landscapes. mBio, 7(4), 5–7. https://doi.org/10.1128/mBio.01351-16

Klapatch, T. R., Demain, A. L., & Lynd, L. R. (1996). Restriction endonuclease activity in *Clostridium thermocellum* and *Clostridium thermosaccharolyticum*. Applied Microbiology and Biotechnology, 45(1–2), 127–131. https://doi.org/10.1007/s002530050659

Koepke, M., Simpson, S., Liew, F., & Chen, W. (2012). fermentation process for producing isopropanol using a recombinant microorganism.

Kolek, J., Sedlar, K., Provaznik, I., & Patakova, P. (2016). Biotechnology for Biofuels Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B- 598 : development of methods for electrotransformation , conjugation, and sonoporation. Biotechnology for Biofuels, 1–14. https://doi.org/10.1186/s13068-016-0436-y

Köpke, M., Gerth, M. L., Maddock, D. J., Mueller, A. P., Liew, F., Simpson, S. D., & Patrick, W. M. (2014). Reconstruction of an acetogenic 2,3-butanediol pathway involving a novel NADPH-dependent primary-secondary alcohol dehydrogenase. Applied and Environmental Microbiology, 80(11), 3394–3303. https://doi.org/10.1128/AEM.00301-14

Köpke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., … Dürre, P. (2010). *Clostridium ljungdahlii* represents a microbial production platform based on syngas. Proceedings of the National Academy of Sciences of the United States of America, 107(29), 13087–92. https://doi.org/10.1073/pnas.1004716107

Köpke, M., & Liew, F. (2011). Recombinant Microorganism and Methods of Production Thereof. Washington, DC.

Köpke, M., Noack, S., & Dürre, P. (2011). The Past, Present, and Future of Biofuels – Biobutanol as Promising Alternative. Biofuel Production-Recent Developments and Prospects, 451–486. https://doi.org/10.5772/20113

Köpke, M., Straub, M., & Dürre, P. (2013). *Clostridium difficile* Is an Autotrophic Bacterial Pathogen. PLoS ONE, 8(4), 1–7. https://doi.org/10.1371/journal.pone.0062157

Kuehne, S. A., Heap, J. T., Cooksley, C. M., Cartman, S. T., & Minton, N. P. (2011). ClosTron-mediated engineering of clostridium. Methods in Molecular Biology, 765(August), 389–407. https://doi.org/10.1007/978-1-61779-197-0_23

Kundiyana, D. K., Huhnke, R. L., Maddipati, P., Atiyeh, H. K., & Wilkins, M. R. (2010). Feasibility of incorporating cotton seed extract in Clostridium strain P11 fermentation medium during synthesis gas fermentation. Bioresource

Technology, 101(24), 9673–9680. https://doi.org/10.1016/j.biortech.2010.07.054

Lamberg, A., Nieminen, S., Qiao, M., & Savilahti, H. (2002). Efficient insertion mutagenesis strategy for bacterial genomes involving electroporation of in vitro-assembled DNA transposition complexes of bacteriophage Mu. Applied and Environmental Microbiology, 68(2), 705–712. https://doi.org/10.1128/AEM.68.2.705-712.2002

Lamichhane, G., Tyagi, S., & Bishai, W. R. (2005). Designer Arrays for De ned Mutant Analysis To Detect Genes Essential for Survival of. Society, 73(4), 2533–2540. https://doi.org/10.1128/IAI.73.4.2533

Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J., & Robertson, H. M. (1999). Hyperactive transposase mutants of the Himar1 mariner transposon. Proceedings of the National Academy of Sciences of the United States of America, 96(20), 11428–33. https://doi.org/10.1073/pnas.96.20.11428

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., et al. (2009). Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. Genome Research 19, 2308-2316.

Latif, H., Zeidan, A. A., Nielsen, A. T., & Zengler, K. (2014). Trash to treasure: Production of biofuels and commodity chemicals via syngas fermenting microorganisms. Current Opinion in Biotechnology, 27, 79–87. https://doi.org/10.1016/j.copbio.2013.12.001

Le Breton, Y., Belew, A. T., Valdes, K. M., Islam, E., Curry, P., Tettelin, H., … McIver, K. S. (2015). Essential Genes in the Core Genome of the Human

Pathogen *Streptococcus pyogenes*. Scientific Reports, 5(1), 9838. https://doi.org/10.1038/srep09838

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. Applied and Environmental Microbiology (Vol. 79). https://doi.org/10.1128/AEM.02891-12

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. Journal of Biotechnology, 188, 97–99. https://doi.org/10.1016/j.jbiotec.2014.07.005

Liao, J. C., Mi, L., Pontrelli, S., & Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. Nature Reviews. Microbiology, 14(5), 288–304. https://doi.org/10.1038/nrmicro.2016.32

Liew, F. (2015). Metabolic Engineering of Clostridium autoethanogenum. University of Nottingham.

Liew, F., Henstra, A.M., Winzer, K., Kopke, M., Simpson, S.D., and Minton, N.P. (2016). Insights into CO2 Fixation Pathway of *Clostridium autoethanogenum* by Targeted Mutagenesis. Mbio 7.

Liew, F., Henstra, A. M., Michael, K., Winzer, K., & Simpson, S. D. (2017). Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production,40(January),104–114. https://doi.org/10.1016/j.ymben.2017.01.007

Liew, F. M., Martin, M. E., Tappel, R. C., Heijstra, B. D., Mihalcea, C., & Köpke, M. (2016). Gas Fermentation-A flexible platform for commercial scale

production of low-carbon-fuels and chemicals from waste and renewable feedstocks. Frontiers in Microbiology, 7(MAY). https://doi.org/10.3389/fmicb.2016.00694

Liou, J. S. C., 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. International Journal of Systematic and Evolutionary Microbiology, 55(5), 2085–2091. https://doi.org/10.1099/ijs.0.63482-0

Liu, G., Ou, H. Y., Wang, T., Li, L., Tan, H., Zhou, X., … He, X. (2010). Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. PLoS Genetics, 6(12), 1–13. https://doi.org/10.1371/journal.pgen.1001253

Ljungdahl, L. G. (1986). the Autotrophic Pathway of. Annual Review of Microbiology, (40), 415–450. https://doi.org/10.1146/annurev.mi.40.100186.002215

Loenen, W. A. M., & Raleigh, E. A. (2014). The other face of restriction: Modification-dependent enzymes. Nucleic Acids Research, 42(1), 56–69. https://doi.org/10.1093/nar/gkt747

Low, D. A., & Casadesús, J. (2008). Clocks and switches: bacterial gene regulation by DNA adenine methylation. Current Opinion in Microbiology, 11(2), 106–112. https://doi.org/10.1016/j.mib.2008.02.012

Marc, J., Grousseau, E., Lombard, E., Sinskey, A. J., Gorret, N., & Guillouet, S. E. (2017). Over expression of GroESL in *Cupriavidus necator* for heterotrophic

and autotrophic isopropanol production. Metabolic Engineering, 42(June), 74–84. https://doi.org/10.1016/j.ymben.2017.05.007

Marinus, M. G. (2010). DNA methylation and mutator genes in *Escherichia coli* K-12. Mutation Research/Reviews in Mutation Research, 705(2), 71–76. https://doi.org/10.1016/j.mrrev.2010.05.001

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. Science, 322(5909), 1843–1845. https://doi.org/10.1126/science.1165771

Martin, M. E., Richter, H., Saha, S., & Angenent, L. T. (2016a). Traits of selected Clostridium strains for syngas fermentation to ethanol. Biotechnology and Bioengineering, 113(3), 531–539. https://doi.org/10.1002/bit.25827

Martin, M. E., Richter, H., Saha, S., & Angenent, L. T. (2016b). Traits of Selected Clostridium Strains for Syngas Fermentation to Ethanol, 113(3), 531–539. https://doi.org/10.1002/bit.25827

McKown, R. L., Orle, K. a., Chen, T., & Craig, N. L. (1988). Sequence requirements of *Escherichia coli* attTn7, a specific site of transposon Tn7 insertion. J. Bacteriol., 170(1), 352–358.

Meeske, A. J., Rodrigues, C. D. A., Brady, J., Lim, H. C., Bernhardt, T. G., & Rudner, D. Z. (2016). High-Throughput Genetic Screens Identify a Large and Diverse Collection of New Sporulation Genes in Bacillus subtilis. PLoS Biology, 14(1), 1–33. https://doi.org/10.1371/journal.pbio.1002341

Mermelstein, L. D., & Papoutsakis, E. T. (1993). *In vivo* methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to

protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. Appl Environ Microbiol, 59(4), 1077–1081. https://doi.org/10.1016/j.bbrc.2011.01.095

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., *et al.* (2016). A roadmap for gene system development in Clostridium. Anaerobe 41, 104-112.

Mock, J., Zheng, Y.N., Mueller, A.P., Ly, S., Tran, L., Segovia, S., Nagaraju, S., Kopke, M., Durre, P., and Thauer, R.K. (2015). Energy Conservation Associated with Ethanol Formation from H-2 and CO2 in *Clostridium autoethanogenum* Involving Electron Bifurcation. Journal of Bacteriology 197, 2965-2980.

Mohammadi, M., Najafpour, G. D., Younesi, H., Lahijani, P., Uzir, M. H., & Mohamed, A. R. (2011). Bioconversion of synthesis gas to second generation biofuels: A review. Renewable and Sustainable Energy Reviews, 15(9), 4255–4273. https://doi.org/10.1016/j.rser.2011.07.124

Munasinghe, P. C., & Khanal, S. K. (2011). Biomass-derived syngas fermentation into biofuels. Biofuels, 101(13), 79–98. https://doi.org/10.1016/B978-0-12-385099-7.00004-8

Nagaraju, S., Davies, N. K., Jeffrey, D., Walker, F., Köpke, M., & Simpson, S. D. (2016). Biotechnology for Biofuels Genome editing of *Clostridium autoethanogenum* using CRISPR / Cas9. Biotechnology for Biofuels, 1–8. https://doi.org/10.1186/s13068-016-0638-3

Nakamura, T. M., & Cech, T. R. (1998). Reversing time: Origin of telomerase. Cell, 92(5), 587–590. https://doi.org/10.1016/S0092-8674(00)81123-X

Neill, B. C. O., Oppenheimer, M., Neill, B. C. O., & Oppenheimer, M. (2016). Dangerous Climate Impacts and the Kyoto Protocol Published by : American Association for the Advancement of Science Linked references are available on JSTOR for this article : Dangerous Climate Impacts and the Kyoto Protocol, 296(5575), 1971–1972.

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

OECD, & IEA. (2014). Renewable Energy. Medium-Term Market Report 2014. Market Analysis and Forecasts to 2020. Executive Summary, 20. https://doi.org/http://dx.doi.org/10.1787/renewmar-2014-en

Opijnen, T. Van, Camilli, A., Opijnen, T. Van, & Camilli, A. (2012). A fine scale phenotype − genotype virulence map of a bacterial pathogen, 2541–2551. https://doi.org/10.1101/gr.137430.112

Ou, X., Zhang, X., Zhang, Q., & Zhang, X. (2013). Life-cycle analysis of energy use and greenhouse gas emissions of gas-to-liquid fuel pathway from steel mill off-gas in China by the LanzaTech process. Frontiers in Energy, 7(3), 263–270. https://doi.org/10.1007/s11708-013-0263-9

Pacala, S. (2004). Stabilization Wedges: Solving the Climate Problem for the Next 50 Years with Current Technologies. Science, 305(5686), 968–972. https://doi.org/10.1126/science.1100103

Pajunen, M. I., Pulliainen, A. T., Finne, J., & Savilahti, H. (2005). Generation of transposon insertion mutant libraries for Gram-positive bacteria by

electroporation of phage Mu DNA transposition complexes. Microbiology, 151(4), 1209–1218. https://doi.org/10.1099/mic.0.27807-0

Paulsen, I. T., Cain, A. K., & Hassan, K. A. (2017). Physical enrichment of transposon mutants from saturation mutant libraries using the TraDISort approach. Mobile Genetic Elements, 7(3), 1–7. https://doi.org/10.1080/2159256X.2017.1313805

Perez-Pinera, P., Han, N., Cleto, S., Cao, J., Purcell, O., Shah, K. A., … Lu, T. K. (2016). Synthetic biology and microbioreactor platforms for programmable production of biologics at the point-of-care. Nature Communications, 7, 12211. https://doi.org/10.1038/ncomms12211

Peters, J. E., & Craig, N. L. (2001). Tn7 : smarter than we thought, 2(November), 806–814.

Phillips, S., Flach, B., Lieberz, S., Rondon, M., Williams, B., & Wilson, C. (2016). EU Biofuels Annual 2016.

Pimentel, D., Marklein, A., Toth, M.A., Karpoff, M.N., Paul, G.S., McCormack, R., Kyriazis, J., and Krueger, T. (2009). Food Versus Biofuels: Environmental and Economic Costs. Human Ecology 37, 1-12.

Pyne, M. E., Bruder, M., Moo-Young, M., Chung, D. A., & Chou, C. P. (2014). Technical guide for genetic advancement of underdeveloped and intractable Clostridium. Biotechnology Advances, 32(3), 623–641. https://doi.org/10.1016/j.biotechadv.2014.04.003

Raftery, J. P., & Karim, M. N. (2017). Economic viability of consolidated bioprocessing utilizing multiple biomass substrates for commercial-scale

cellulosic bioethanol production. Biomass and Bioenergy, 103, 35–46. https://doi.org/10.1016/j.biombioe.2017.05.012

Rajagopalan, S., P. Datar, R., & Lewis, R. S. (2002). Formation of ethanol from carbon monoxide via a new microbial catalyst. Biomass and Bioenergy, 23(6), 487–493. https://doi.org/10.1016/S0961-9534(02)00071-5

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2010). REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. Nucleic Acids Research, 43(D1), D298–D299. https://doi.org/10.1093/nar/gku1046

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2015). REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. Nucleic Acids Research, 43(D1), D298–D299. https://doi.org/10.1093/nar/gku1046

Russell, D. W., & Zinder, N. D. (1987). Hemimethylation prevents DNA replication in *E. coli.* Cell, 50(7), 1071–1079. https://doi.org/10.1016/0092-8674(87)90173-5

Salama, N. R., Shepherd, B., & Falkow, S. (2004). Global Transposon Mutagenesis and Essential Gene Analysis of *Helicobacter pylori* †. Society, 186(23), 7926–7935. https://doi.org/10.1128/JB.186.23.7926

Sasakawa, C., Carle, G. F., & Berg, D. E. (1983). Sequences essential for transposition at the termini of IS50. Proc Natl Acad Sci U S A, 80(23), 7293–7297.

Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2001). Comprehensive identification of conditionally essential genes in mycobacteria. Proceedings of the National

Academy of Sciences, 98(22), 12712–12717. https://doi.org/10.1073/pnas.231275498

Satomura, T., Shimura, D., Asai, K., Sadaie, Y., & Hirooka, K. (2005). Enhancement of Glutamine Utilization in Bacillus subtilis through the GlnK-GlnL Two-Component Regulatory System Enhancement of Glutamine Utilization in *Bacillus subtilis* through the GlnK-GlnL Two-Component Regulatory System. Society, 187(14), 4813–4821. https://doi.org/10.1128/JB.187.14.4813

Saxena, J., & Tanner, R. S. (2011). Effect of trace metals on ethanol production from synthesis gas by the ethanologenic acetogen, *Clostridium ragsdalei*. Journal of Industrial Microbiology and Biotechnology, 38(4), 513–521. https://doi.org/10.1007/s10295-010-0794-6

Saxena, J., & Tanner, R. S. (2012). Optimization of a corn steep medium for production of ethanol from synthesis gas fermentation by *Clostridium ragsdalei*. World Journal of Microbiology and Biotechnology, 28(4), 1553–1561. https://doi.org/10.1007/s11274-011-0959-0

Schulz, H. (1999). Short history and present trends of Fischer–Tropsch synthesis. Applied Catalysis A: General, 186(1–2), 3–12. https://doi.org/http://dx.doi.org/10.1016/S0926-860X(99)00160-X

Shaw, W. V. (1967). Chloramphenicol Acetyltransferase from Chloramphenicol-Resistant Bacteria. Antibiotic Inactivation and Modification, 687(Cm), 737–755.

Shen, B. W., Heiter, D. F., Chan, S. H., Wang, H., Xu, S. Y., Morgan, R. D., … Stoddard, B. L. (2010). Unusual Target Site Disruption by the Rare-Cutting HNH Restriction Endonuclease PacI. Structure, 18(6), 734–743. https://doi.org/10.1016/j.str.2010.03.009

Shen, G. J., Shieh, J. S., Grethlein, A. J., Jain, M. K., & Zeikus, J. G. (1999). Biochemical basis for carbon monoxide tolerance and butanol production by *Butyribacterium methylotrophicum*. Applied Microbiology and Biotechnology, 51(6), 827–832. https://doi.org/10.1007/s002530051469

Sitaraman, R., & Leppla, S. H. (2012). Methylation-dependent DNA restriction in *Bacillus anthracis*. Gene, 494(1), 44–50. https://doi.org/10.1016/j.gene.2011.11.061

Smith, V., Botstein, D., & Brown, P. O. (1995). Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. Proceedings of the National Academy of Sciences of the United States of America, 92(14), 6479–6483. https://doi.org/10.1073/pnas.92.14.6479

Smith, V., Botstein, D., & Browntt, P. (1995). Genetic footprinting : A genomic strategy for determining a genes function given its sequence, 92(July), 6479–6483.

Spath, K., Heinl, S., & Grabherr, R. (2012). "Direct cloning in *Lactobacillus plantarum*: electroporation with non-methylated plasmid DNA enhances transformation efficiency and makes shuttle vectors obsolete". Microbial Cell Factories, 11(1), 141. https://doi.org/10.1186/1475-2859-11-141

Suárez, G. A., Renda, B. A., Dasgupta, A., & Barrick, J. E. (2017). Reduced mutation rate and increased transformability of transposon-free *Acinetobacter baylyi* ADP1-ISx. Applied and Environmental Microbiology, 83(17). https://doi.org/10.1128/AEM.01025-17

Tamura, G. S., Bratt, D. S., Yim, H. H., & Nittayajarn, A. (2005). Use of glnQ as a counterselectable marker for creation of allelic exchange mutations in group

B streptococci. Applied and Environmental Microbiology, 71(1), 587–590. https://doi.org/10.1128/AEM.71.1.587-590.2005

Thackeray, M. M., Wolverton, C., & Isaacs, E. D. (2012). Electrical energy storage for transportation—approaching the limits of, and going beyond, lithium-ion batteries. Energy & Environmental Science, 5(7), 7854. https://doi.org/10.1039/c2ee21892e

Tol, R. S. J. (2014). Quantifying the consensus on anthropogenic global warming in the literature: Rejoinder. Energy Policy, 73, 709. https://doi.org/10.1016/j.enpol.2014.06.003

Tummala, S. B., Welker, N. E., & Papoutsakis, E. T. (1999). Development and characterization of a gene expression reporter system for *Clostridium acetobutylicum* ATCC 824. Applied and Environmental Microbiology, 65(9), 3793–3799.

Tummala, S. B., Welker, N. E., & Papoutsakis, E. T. (2003). Design of antisense RNA constructs for downregulation of the acetone formation pathway of *Clostridium acetobutylicum*. Journal of Bacteriology, 185(6), 1923–1934. https://doi.org/10.1128/JB.185.6.1923-1934.2003

Tyurin, M. V, Desai, S. G., & Lynd, L. R. (2004). Electrotransformation of *Clostridium thermocellum*. Appl Environ Microbiol, 70(2), 883–890. https://doi.org/10.1128/AEM.70.2.883

Ungerman, A. J., & Heindel, T. J. (2007). Carbon monoxide mass transfer for syngas fermentation in a stirred tank reactor with dual impeller configurations. Biotechnology Progress, 23(3), 613–620. https://doi.org/10.1021/bp060311z

UNICEF/WHO/World Bank. (2016). Levels and trends in child malnutrition. UNICEF-WHO-World Bank Joint Child Malnutrition Estimates., 2016(accessed 12 October 2016), 1–8. https://doi.org/10.1016/S0266-6138(96)90067-4

Utturkar, S.M., Klingeman, D.M., Bruno-Barcena, J.M., Chinn, M.S., Grunden, A.M., Kopke, M., and Brown, S.D. (2015). Sequence data for *Clostridium autoethanogenum* using three generations of sequencing technologies. Scientific Data 2.

Whitaker, R.D., Dorner, L.F., and Schildkraut, I. (1999). A mutant of BamHI restriction endonuclease which requires N6-methyladenine for cleavage. Journal of Molecular Biology 285, 1525-1536.

van der Sluijs, J. P., & Turkenburg, W. C. (2006). Climate Change and the Precautionary Principle. Implementing The Precautionary Principle, Perspectives and Prospects, 245–269.

van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nature Methods, 6(10), 767–72. https://doi.org/10.1038/nmeth.1377

Vandewalle, K., Festjens, N., Plets, E., Vuylsteke, M., Saeys, Y., & Callewaert, N. (2015). by Cartesian Pooling-Coordinate Sequencing. Nature Communications, 6(May), 1–7. https://doi.org/10.1038/ncomms8106

Vasu, K., & Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. Microbiology and Molecular Biology Reviews : MMBR, 77(1), 53–72. https://doi.org/10.1128/MMBR.00044-12

Vega, J. L., Antorrena, G. M., Clausen, E. C., & Gaddy, J. L. (1989). Study of gaseous substrate fermentations: Carbon monoxide conversion to acetate. 2. Continuous culture. Biotechnology and Bioengineering, 34(6), 785–793. https://doi.org/10.1002/bit.260340608

Walker, D. J. ., & Köpke, M. (2015). Method of Producing a Recombinant Microorganism. Washington,DC.: U.S.PatentandTrademark Office.

Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K., Lu, T., Jin, Y. S., & Blaschek, H. P. (2015). Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. Journal of Biotechnology, 200, 1–5. https://doi.org/10.1016/j.jbiotec.2015.02.005

Wilkins, M. R., & Atiyeh, H. K. (2011). Microbial production of ethanol from carbon monoxide. Current Opinion in Biotechnology, 22(3), 326–330. https://doi.org/10.1016/j.copbio.2011.03.005

Williams, D. R., Young, D. I., & Young, M. (1990). Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. Journal of General Microbiology, 136(1 990), 819–826. https://doi.org/10.1099/00221287-136-5-819

Whitaker, R.D., Dorner, L.F., and Schildkraut, I. (1999). A mutant of BamHI restriction endonuclease which requires N6-methyladenine for cleavage. Journal of Molecular Biology 285, 1525-1536.

Wion, D., & Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nature Reviews. Microbiology, 4(3), 183–192. https://doi.org/10.1038/nrmicro1350

Wise, T. A. (2012). The Cost to Mexico of U.S. Corn Ethanol Expansion. Global Development and Environment Institute, (May), 14. Retrieved from http://ase.tufts.edu/gdae/Pubs/wp/12-01WiseBiofuels.pdf

Woolley, R. C., Pennock, A., Ashton, R. J., Davies, A., & Young, M. (1989). Transfer of Tn1545 and Tn916 to *Clostridium acetobutylicum*. Plasmid, 22(2), 169–174. https://doi.org/10.1016/0147-619X(89)90027-9

Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., … Zhou, J. (2015). Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. Applied and Environmental Microbiology, 81(13), 4423–4431. https://doi.org/10.1128/AEM.00873-15

Zhang, N., Shao, L., Jiang, Y., Gu, Y., Li, Q., Liu, J., … Yang, S. (2015). I-SceI-mediated scarless gene modification via allelic exchange in Clostridium. Journal of Microbiological Methods, 108, 49–60. https://doi.org/10.1016/j.mimet.2014.11.004

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(4), 1–21. https://doi.org/10.1371/journal.pone.0122411

Zhang, Y. J., Ioerger, T. R., Huttenhower, C., Long, J. E., Sassetti, C. M., Sacchettini, J. C., & Rubin, E. J. (2012). Global Assessment of Genomic Regions Required for Growth in Mycobacterium tuberculosis. PLoS Pathogens, 8(9). https://doi.org/10.1371/journal.ppat.1002946

Zhang, Y., Xu, S., Chai, C., Yang, S., Jiang, W., Minton, N. P., & Gu, Y. (2016). Development of an inducible transposon system for efficient random

mutagenesis in *Clostridium acetobutylicum*. FEMS Microbiology Letters, 363(8), 1–7. https://doi.org/10.1093/femsle/fnw065