Developing Synthetic Biology Tools for Acetogens

Elizabeth R. M. Redfern



Thesis submitted for the degree of Doctor of Philosophy

November 2021

Declaration

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

Elizabeth R.M. Redfern, 2021

Abstract

Acetogenic bacteria are of industrial interest due to their capacity for autotrophic growth on single carbon (C1) feedstocks such as CO or CO_2 and H_2. Both *Eubacterium limosum* and *Clostridium carboxidivorans* can autotrophically produce four carbon (C4) products, with *C. carboxidivorans* also able to produce six carbon (C6) products, such as hexanol and hexanoate. *E. limosum* is also capable of autotrophic growth on methanol. This study outlines efforts to further the industrial potential of these strains through the development of DNA transfer methods, synthetic biology toolkits and a large transposon mutant library elucidating gene essentiality under industrially relevant conditions.

To establish DNA transfer in *C. carboxidivorans,* preliminary genotypic characterisation of the ten RM systems was performed. This substantial barrier was then overcome by the application of multiple circumvention strategies including disruption of key recognition motifs in the shuttle vector sequence, heterologous expression of native methyltransferases in *E. coli* conjugal donors, and isolation of hyper-receptive recipient strains with modifications in the *fabK* gene. Through complementation of these strategies a reliable conjugative DNA transfer method was established. Utilising this method, an inducible exogenous CRISPR-Cas9 system was introduced and used to knock-out *CcaP7ORF17830P*, one of the Type II RM system genes. In implementing this genetic modification system multiple auxillary genetic tools were demonstrated, including a *catP* reporter gene and theophylline inducible riboswitch.

Also reported herein is the generation of a saturated *E. limosum* NG-6849 transposon insertional mutant pool with 179,161 unique insertional mutants. High throughput sequencing of this mutant pool, grown either exclusively on methanol or on glucose, was used to evaluate the respective gene essentiality. This revealed 103 disparate genes required solely for growth on methanol, and not glucose. Due to mutant overgrowth within the methanol conditions, this

process also revealed key candidate genes whose disruption may correlate with improved methanol growth kinetics.

Acknowledgements

My sincere thanks are first extended to my supervisors Professor Nigel Minton and Doctor Ying Zhang for their support and guidance throughout my PhD. Beyond these I am forever indebted to Doctor Chris Humphreys and Doctor James Millard, later additions to my supervision team, from whom I regularly sought advice, expertise, and encouragement. In addition, I would like to acknowledge the instrumental bioinformatic support of Doctor Craig Woods.

My thanks also go to my fellow PhD students and colleagues within the Synthetic Biology Research Centre whose assistance and company over the years has been of great help. Beyond the academic domain I would like to thank my family and friends, particularly my parents and partner, for ensuring my sanity throughout my studies. Finally, I would like to thank the BBSRC and University of Nottingham for supporting and funding this research project.

List of Contents

Developing Synthetic Biology Tools for Acetogens 1
Declaration 2
Abstract 3
Acknowledgements 5
List of Contents 6
Abbreviations 12
List of Figures 14
List of Tables 17
1 Introduction 19
1.1. Biotechnology and sustainable development 19
1.2. Alternative feedstocks 20
1.2.1. Biomass 21
1.2.2. Biomass gasification 23
1.2.3. Syngas 23
1.2.4. Methanol 24
1.3. Microbial catalysis 25
1.4. Acetogenic bacteria 26
1.5.1. The Wood-Ljungdahl Pathway 27
1.5.2. Energetics 31
1.5. Eubacterium limosum 32
1.6. Clostridium carboxidivorans 33
1.7. Strain development 34
1.7.1. Genome 35
1.7.2. DNA transfer 36
1.7.3. Genetic tools in *Clostridium* 37
1.8. Genomic modification in *Clostridium* 41
1.8.1. ClosTron 42
1.8.2. Allelic exchange 43
1.8.3. CRISPR-Cas 44
1.9. Forward genetics studies 46
1.9.1. Transposon mutagenesis 47
1.9.2. Transposon insertion sequencing (TIS) 48
1.10. Project Aims 49
2 Materials and Methods 51
2.1. Chemicals and suppliers 51
2.2. Bacterial strains and routine culture 51
2.1.1. Bacterial strains 51
2.1.2. Routine growth conditions 53

2.1.3. Supplements 54
2.1.4. Strain storage and revival 55
2.3. DNA transfer 55
2.3.1. Preparation of chemically competent *E. coli* cells 55
2.3.2. Transformation of chemically competent *E. coli* 56
2.3.3. Preparation of electrocompetent *E. coli* 56
2.3.4. Transformation of electrocompetent *E. coli* 57
2.3.5. Preparation of electrocompetent *E. limosum* cells 57
2.3.6. Transformation of electrocompetent *E. limosum* 58
2.3.7. Conjugative DNA transfer in *C. sporogenes* 59
2.3.8. Conjugative DNA transfer in *C. autoethanogenum* 60
2.3.9. Conjugative DNA transfer in *C. carboxidivorans* 60
2.4. General molecular techniques 62
2.4.1. Plasmid DNA extraction 62
2.4.2. Genomic DNA extraction 62
2.4.3. Restriction enzyme digestion 62
2.4.4. De-phosphorylation of linear DNA 62
2.4.5. DNA ligation 63
2.4.6. DNA extraction from agarose gels 63
2.4.7. PCR clean up 63
2.4.8. Polymerase chain reaction (PCR) 64
2.4.8.1. High fidelity DNA amplification 64
2.4.8.2. Splicing by overlap extension (SOE) PCR 65
2.4.8.3. Low fidelity DNA amplification 65
2.4.9. HiFi DNA assembly 66
2.4.10. Golden gate assembly 66
2.4.11. DNA visualisation 67
2.4.11.1. Agarose gel electrophoresis 67
2.4.11.2. High sensitivity automated gel electrophoresis 67
2.4.12. DNA quantification 67
2.4.12.1. Nanodrop™ 67
2.4.12.2. Qubit™ 68
2.5. DNA sequencing 68
2.5.1. MicrobesNG sample preparation and sequencing 69
2.5.2. DeepSeq sample preparation and sequencing 70
2.6. Bioinformatics tools 71
2.7. Plasmid Lists 73
2.7.1. Chapter 3 73
2.7.2. Chapter 4 75
2.7.3. Chapter 5 76

2.8. Oligonucleotides and dsDNA sequences 78
2.8.1. Chapter 3 oligonucleotides 78
2.8.2. Chapter 4 oligonucleotides 79
2.8.3. Chapter 5 oligonucleotides 81
2.8.4. Synthetic DNA fragments 82
2.9. Plasmid construction 83
2.9.1. pMTL8315y-v* 83
2.9.2. pRBV1_T* 84
2.9.3. pRBV1_TII-III 85
2.9.4. pRBV2_L2-TIII 86
2.9.5. pMTL8315y-v1-II/III* 87
2.9.6. pMTL8015y- v1-TIII.2R-SC* 87
2.9.7. pMTL8015y-v1-ToxAtox-TIII.2R 88
2.9.8. pMTL8325y-Pxxx-CatP(trunc) 89
2.9.9. pMTL4315_SynCas9_adh1_N20#2 90
2.9.10. pRESynCas(pthI)-v1-TIII.2R-sg* 90
2.9.11. pMTL83254-Rb* 92
2.9.12. pRECas(*)-ORFD 92
2.10. pR702 recombineering 93
2.11. Chloramphenicol acetyltransferase assay 94
2.11.1. CAT assay sample preparation 95
2.11.2. CAT assay 95
2.11.3. Bicinchoninic acid assay 96
2.11.4. CAT/BCA analysis 96
2.12. Bacterial sensitivity assay 97
2.13. TraDIS analysis of *E. limosum* NG-6849 99
2.13.1. Transposon mutagenesis of *E. limosum* 99
2.13.2. TraDIS sample preparation 100
2.13.3. TraDIS analysis 103
3 Establishment of DNA transfer in *Clostridium carboxidivorans* 105
3.1. Introduction 105
3.1.1. Horizontal gene transfer 105
3.1.2. Electroporation and conjugation 106
3.1.3. Barriers to DNA transfer 108
3.1.4. Restriction-modification systems 110
3.1.5. Aims 111
3.2. Results 112
3.2.1. Genomic analysis 112
3.2.1.1. REBASE® Analysis 112
3.2.1.2. PacBio® analysis 115

3.2.2. Evading RM restriction activity 121
3.2.2.1. Thiamphenicol sensitivity 122
3.2.2.2. pMTL8315y-v* construction 123
3.2.2.3. pMTL8315-v1 transfer into *C. carboxidivorans* 125
3.2.2.4. pMTL8315y-v1/2/3 conjugative transfer to *C. sporogenes* 127
3.2.2.5. pMTL8315y-v2/3 in *C. carboxidivorans* 129
3.2.3. Conjugation of *C. carboxidivorans* with *E. coli* CA434 130
3.2.4. Isolation of *C. carboxidivorans* variant C94 131
3.2.5. Further transfer optimisation 136
3.2.5.1. Electroporation 136
3.2.5.2. Alternate *E. coli* donor strains 137
3.2.6. Mimicking native methylation profile 138
3.2.6.1. Overview 138
3.2.6.2. pRBV1_T* vector generation 139
3.2.6.3. Generation of sExpress_TII and sExpress_TIII 142
3.2.6.4. Recombinant sExpress conjugal donor assessment 143
3.2.6.5. Generation of sExpress_TI/TII/III 144
3.2.7. Hyper-receptive conjugative recipients 147
3.2.8. Further RM system elucidation 152
3.2.8.1. pMTL8315y-v1-TII/IIIABC transfer efficacy 152
3.2.8.2. Nanopore methyltransferase analysis 154
3.3. Discussion 157
3.3.1. Overview 157
3.3.2. REBASE® and PacBio® analysis of the RM systems of *C. carboxidivorans* 158
3.3.3. RM system evasion with pMTL8315y-v2/3 159
3.3.4. CA434 conjugation 161
3.3.5. *E. coli* Dam/Dcm methylases 163
3.3.6. *E. coli* sExpress recombinant donors 165
3.3.7. Hyper-receptive conjugative recipient strains characterisation 168
3.3.8. Further RM system characterisation 172
3.4. Chapter summary 175
4 Development of genome modification tools for *C. carboxidivorans* 177
4.1. Introduction 177
4.1.1. Metabolic engineering of *C. carboxidivorans* 177
4.1.2. Allelic exchange genome editing in *Clostridium* 181
4.1.2.1. Genomic selection/counter-selection 181
4.1.2.2. Plasmid selection/counter-selection 183
4.1.3. CRISPR-Cas genome editing systems in *Clostridium* 186
4.1.4. Aims: 189
4.2. Results 189

4.2.1. Establishing key genetic tools in *C. carboxidivorans* 189
4.2.1.1. *ermB* resistance marker 189
4.2.1.2. *catP* reporter assay 191
4.2.1.3. Theophylline growth impact 193
4.2.1.4. RbE*-$P_{fdx(Csp)}$ *catP* reporter assay 194
4.2.2. Genome modification of *C. carboxidivorans* through allelic exchange 197
4.2.2.1. Generation of Δ*ccaP7ORF17830P*::pMTL8015y-v1-TIII.2R-SC* *C. carboxidivorans* single crossover integrants 197
4.2.2.2. Assessing CD2517.1 plasmid counter-selection 201
Targeted disruption of CcaP7ORF17830P with pMTL8015y-v1-ToxATox-TIII.2R 204
4.2.2.3. 204
4.2.3. Exogenous CRISPR-Cas9 genome editing 208
4.2.3.1. Genomic modification of *C. autoethanogenum* with pMTL4315_SynCas9_adh1_N20 #2 209
4.2.3.2. Knockout of *CcaP7ORF17830P* within *C. carboxidivorans* 211
4.3. Discussion 214
4.3.1. Genetic tools 214
4.3.2. Single crossover integration 216
4.3.3. CD2517.1 counter-selection 219
4.3.4. CRISPR-Cas9 genome editing 222
4.3.5. ΔccaP7ORF17830P 225
4.4. Chapter summary 226
5 Application of synthetic biology tools within *E. limosum* NG-6849 and generation of a large-scale mutant library 227
5.1. Introduction 227
5.1.1. Genetic tools within *E. limosum* 228
5.1.2. Other studies within *E. limosum* 229
5.1.3. *mariner* transposon mutagenesis approaches in *Clostridium* 230
5.1.4. pMTL-CW20 transposon mutagenesis vectors 232
5.1.5. Transposon mutagenesis in *E. limosum* ATCC 8486 235
5.2. Results 236
5.2.1. *E. limosum* NG-6849 genetic tractability 236
5.2.1.1. Genome 236
5.2.1.2. *E. limosum* NG-6849 and ATCC 8486 genome comparison 239
5.2.1.3. Antibiotic sensitivity 241
5.2.1.4. pMTL80000 transformation 242
5.2.1.5. Theophylline inducible riboswitch assessment 244
5.2.1.6. pRECas genome editing 246
5.2.2. Transposon mutagenesis of *E. limosum* NG-6849 with pMTL-CW21 249
5.2.2.1. Tet3nO validation 249
5.2.2.2. Plasmid loss and transposition efficiency assessment 250

5.2.2.3. *E. limosum* NG-6849 transposon mutagenesis 252
5.2.3. TraDIS analysis of the essential genes required for the growth of *E. limosum* NG-6849 on methanol 253
5.2.3.1. TraDIS read mapping and gene essentiality 255
5.2.3.2. Genes involved in methanol metabolism via the Wood-Ljungdahl pathway 258
5.2.3.3. Essential genes within the Wood-Ljungdahl pathway 263
5.2.3.4. Integral energy conservation gene clusters 266
5.2.3.5. Key product synthesis genes 269
5.2.3.6. Final methanol subculture composition 271
5.3. Discussion 274
5.3.1. Genome 274
5.3.2. Genetic tools 275
5.3.3. TraDIS coverage and gene essentiality prediction 276
5.3.4. Essential genes for methanol metabolism 280
5.3.5. Core WLP gene essentiality and energetics 283
5.3.6. Mutant overrepresentation 286
5.4. Chapter summary 290
6 Discussion 291
6.1. Clostridium carboxidivorans 291
6.1.1. Original contributions 291
6.1.2. Further studies 292
6.2. Eubacterium limosum 294
6.1.3. Original contributions 294
6.1.4. Further studies 295
Bibliography 298
Appendix 313

Abbreviations

%	percentage
°C	degrees Celsius
5-FOA	5-fluoroorotic acid
ACE	allele-coupled exchange
aTc	anhydrotetracycline
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
BBSRC	Biotechnology and Biological Science Research Council
BLAST	basic local alignment search tool
bp	base pairs
C1/4/6	one/four/six carbon containing compounds
CAT	chloramphenicol acetyltransferase
CFU	colony forming unit
Cla^R	clarithromycin resistant
CO	carbon monoxide
CO_2	carbon dioxide
CoA	conezyme-A
cm	centimeter
Cm^R	chloramphenicol resistant
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation tool
CRISPRi	CRISPR interference tool
crRNA	CRISPR RNA
dCas9	dead *cas9* variant
dH_2O	distilled water
DIN	DNA-integrity number
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSB	double strand DNA breaks
FAST	fluorescent-activating and absorption shifting
g/mg	grams/milligrams
g	gravity
H_2	hydrogen
InDel	insertions and deletions
IPTG	isopropyl ß-D-1-thiogalactopyranoside
kb	kilobase pairs
kV	kilovolts
L/mL/µL	litres/millilitres/microlitres
LB	lysogeny broth
LHA/RHA	left/right homology arm
M/mM	molar/millimolar
mb	megabase pairs
MCS	multiple cloning site
MIC	minimal inhibitory concentration
mS	milliseconds
MSW	municipal solid waste

NAD+	oxidised nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
nCas9	nickase *cas9* variant
NGS	next generation sequencing
NHEJ	non-homologous end joining
nm	nano meters
OD_{600}	optical density at 600nm
ORF	open reading frame
PacBio	Pacific Biosciences
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PCR	polymerase chain reaction
qPCR	quantitative polymerase chain reaction
RM	restriction modification
RNA	ribonucleic acid
RPM	revolutions per minute
SBRC	Synthetic Biology Research Centre
SMRT	single molecule real time
SNP	single nucleotide polymorphism
ssDNA	single-stranded DNA
SSN	single strand nick
Syngas	synthesis gas
TAE	tris base-acetic acid-EDTA buffer
TIS	transposon insertion sequencing
Tm	melting temperature
Tm^R	thiamphenicol resistant
TraDIS	transposon-directed insertion site sequencing
TRD1	target recognition domain
UIS	unique insertion sites
Uro-D	uroporphyrinogen decarboxylase
V	volts
v/v	volume per volume
w/v	weight per volume
WLP	Wood-Ljungdahl pathway
Ω	resistance (ohms)
µF	microfarad

List of Figures

Figure 1 - Wood Ljundahl pathway schematic 29
Figure 2 – Structural diagrams of 5-methylcytosine (5mC) and 6-methyladenine (6mA) 109
Figure 3 – Schematic of the genomic loci of the four RM system Types 110
Figure 4 – The restriction modification (RM) systems of C. carboxidivorans P7. 114
Figure 5 – pMTL80000 shuttle vector structure schematic 121
Figure 6 – Analysis of *C. carboxidivorans* P7 sensitivity to thiamphenicol 123
Figure 7 – pMTL8315y-v1 vector map 124
Figure 8 – Colony PCR of pMTL8315y-v1 within *C. carboxidivorans* P7 126
Figure 9 – Colony PCR of TIII.2 genomic region within pMTL8315y-v1 *C. carboxidivorans* P7 transconjugants 127
Figure 10 – Conjugation efficiency of pMTL8315y-v* vectors as compared to pMTL83151 in *C. sporogenes* 129
Figure 11 - Plasmid maps of pRBV1_TI, TII and TIII 141
Figure 12 – Comparison of the DNA transfer rates into *C. carboxidivoran*s C94 when using sExpress, sExpress_TII and sExpress_TIII as *E. coli* conjugal donors 143
Figure 13 – pRBv1 and pRBV2 plasmid maps 145
Figure 14 — pRBV2_L2-TIII plasmid map 147
Figure 15 – Conjugal transfer efficiency comparison between *C. carboxidivorans* C94, C95, C96 and C97 148
Figure 16 — pMTL8315y-v1-TII/IIIABC plasmid map 152
Figure 17 — Comparison of conjugation efficiency of sExpress, sExpress_TII and sExpress_TIII into *C. carboxidivorans* C94 with pMTL8315y-v1-II/IIIA, pMTL8315y-v1-II/IIIB, pMTL8315y-v1-II/IIIC, and pMTL8315y-v1-II/IIIABC. 153
Figure 18 – Sequence logo representation of MEME identified methylation motifs within nanopore data 156
Figure 19 – Schematic overview of the two RM circumvention approaches undertaken within this study 158
Figure 20 – Allele coupled exchange (ACE) schematic 182
Figure 21 – *C. carboxidivorans* C94 growth in the presence of clarithromycin (0-10 µg/mL). 190
Figure 22 – pMTL8325y-v1 plasmid map 191
Figure 23 – Relative promoter strength of four constituitive promoters in *C. carboxidivorans* C94. 193
Figure 24 – Impacts of 0-10 mM theophylline on *C. carboxidivorans* growth. 194
Figure 25 – pMTL8325y-v1-$P_{fdx(Csp)}$-RbE*-CatP(trunc) plasmid map 195
Figure 26 – $P_{fdx(Csp)}$-RbE* derived expression levels within *C. carboxidivorans* C94 upon induction with 0-10 mM theophylline. 196

Figure 27- pMTL8015y-v1-TIII.2R-SC4 plasmid map 198
Figure 28 – Screening PCR of suicide plasmid transconjugant colonies generated by pMTL8015y-v1-TIII.2R-SC* vector transfer into *C. carboxidivorans* C94. 200
Figure 29 – pMTL8315y-v1-ToxATox-PyrE plasmid map 202
Figure 30 – pMTL8015y-v1-ToxATox-TIII.2R plasmid map 205
Figure 31 – Screening PCR to confirm pMTL8015y-v1-ToxATox-TIII.2R genomic integration. 206
Figure 32 – Plasmid map of pMTL4315_SynCas_adh1_N20#2 209
Figure 33 –pMTL4315_SynCas9_CRISPR_adh1_N20#2 generated mutant sequence. 210
Figure 34 -pRECas(P_{thl})-v1-TIII.2R-sg2 plasmid map 212
Figure 35 – Sanger sequencing and colony PCR of pRESynCas(pthI)-v1-TIII.2-sg2 generated mutants. 213
Figure 37 – pMTL-CW21 vector map 233
Figure 38 – Schematic of C1 metabolism and butyrate/acetate synthesis in *E. limosum* NG-6849, via the WLP with enzymes summarised in Table 28. 238
Figure 39 – Graphical alignment of the two *E. limosum* genomes NG-6849 genome and ATCC 8486. 241
Figure 40 – *E. limosum* NG-6849 24-hour growth OD_{600} values in the presence of clarithromycin (0-10 µg/mL) and thiamphenicol (0-20 µg/mL). 242
Figure 41 – Log10 DNA transfer efficiency of pMTL80000 vectors into *E. limosum* NG-6849 by electroporation. 243
Figure 42 – Impacts of 0-10 mM theophylline on *E. limosum* NG-6849 growth. 244
Figure 43 – $P_{fdx(Csp)}$-Rb* derived CatP expression levels within *E. limosum* NG-6849. 245
Figure 44 – pRECas(araE)-ORFD plasmid map 247
Figure 45 – PCR amplification of R.ORFD gmrSD locus within *E. limosum* NG-6849 colonies 248
Figure 46 – Relative promoter strength of Tet3nO within *E.limosum* NG-6849 250
Figure 47 – Transposase activity and plasmid loss efficiency of pMTL-CW21 in *E. limosum* NG-6849. 251
Figure 48 – Schematic view of the integral TraDIS mutant pool growth and selection steps. 254
Figure 49 – Insertion plots of TraDIS datasets relative to the mapped genome sequence. ... 258
Figure 50 – Insertion plots of candidate methanol metabolism gene clusters mtaABC and metV/F 260
Figure 51 – Schematic of core C1 metabolism within *E. limosum* NG-6849 with associated TraDIS insertion indeces 265
Figure 52 – Insertion plots of candidate V/A-type ATP synthase gene within subculture 3 on methanol and glucose respectively 268
Figure 53 – Graphical representation of *E. limosum* NG-6849 mutant pool diversity in the third and fourth subculture on methanol 272

Figure 54 – Mutant abundance (%) of five key genes within individual replicates of the *E. limosum* NG-6849 TraDIS mutant pool grown on methanol 273
Figure 55 – Rarefaction analysis of TraDIS dataset 289
Figure 56 – Bio-TraDIS pipeline mapping statistics 313

List of Tables

Table 1 – Detailed list of the bacterial strains employed throughout this study 51
Table 2 - YTF medium solution compositions 54
Table 3 – Media supplements for stock and use in Gram-negative and Gram-positive strains 54
Table 4 - Q5® PCR thermocycler program 64
Table 5 - List of plasmids used during the establishment of robust DNA transfer into *C. carboxidivorans* 73
Table 6 - List of plasmids used in the development of tools for the *in vivo* genomic manipulation of *C. carboxidivorans* 75
Table 7 – Lists of plasmids used in the characterisation of *E. limosum* NG-6849, its genetic tractability and essential gene profile for growth on methanol 76
Table 8 – List of primers implemented in the development of a robust DNA transfer protocol for use with *C. carboxidivorans* 78
Table 9 – List of primers utilised in the development of a genetic toolkit for genomic manipulation of *C. carboxidivorans* 79
Table 10 – List of primers employed in the genomic modification of *E. limosum* NG-6849 with both a CRISPR-Cas9 and mariner transposon system. 81
Table 11 – List of longer DNA fragments ordered as gBlocks from IDT 82
Table 12 – Recurrent methylated sequences identified within the *C. carboxidivorans* P7 genome 115
Table 13 – RM systems in the *C. carboxidivorans* P7 genome 117
Table 14 – Relevant alignment scores for the RM systems of *C. carboxidivorans* P7 120
Table 15 – *C. carboxidivorans* P7 derived methylation sequence occurrence in the pMTL80000 vector series modules 122
Table 16 — Conjugal transfer efficiency comparison between *C. carboxidivorans* P7 and *C. carboxidivorans* C94. 132
Table 17 — Summary report of *C. carboxidivorans* C94 Illumina read-mapping 134
Table 18 — Pertinent SNPs called from read mapping of the C94 isolate against the published *C. carboxidivorans* P7 genome 135
Table 19 – Gene inclusion within pRBV1 vectors for expression of *C. carboxidivorans* P7 derived methyltransferase/specificity activity. 140
Table 20 — Conjugation efficacy of sExpress, sExpress_TII and sExpress_TIII with *C. carboxidivorans* P7 144
Table 21 — Summary reports of Illumina read-mapping of *C. carboxidivorans* C95, C96, C97 and plasmid cured C. carboxidivorans P7 149
Table 22 – Sequence variations between key open reading frames of *C. carboxidivorans* hyper-receptive variants and the *C. carboxidivorans* P7 genome. 151

Table 23 – TII/III RM motif frequency in candidate Type II and Type III pRBV1 vectors 154
Table 24 – Read mapping statistics from Deep Seq GridIon run 155
Table 25 – Genotype information for *E. coli* donor strains tested 163
Table 26 – Resultant transconjugant CFUs of TIII.2R integration plasmids. 199
Table 27 – Comparative CFUs of three *C. carboxidivorans* C94 pMTL8015y-v1-ToxATox-TIII.2R chromosomal integrants 208
Table 28 – Enzymes involved in core C1 metabolic pathways, and subsequent acetate and butyrate synthesis within *E. limosum* NG-6849. 239
Table 29 – Summary report of *E. limosum* NG-6849 Illumina read-mapping against the *E. limosum* ATCC 8486 genome and the received NG-6849 draft 240
Table 30 – Overview of variation between the *E. limosum* NG-6849 genome and *E. limosum* ATCC 8486 genome (Song & Cho, 2015) 241
Table 31 – Mapping statistics of TraDIS sequencing data with the *E. limosum* ATCC 8486 annotated genome 257
Table 32 – Candidate methanol metabolism genes with associated locus tag, functional annotation, and insertion index values 262
Table 33 – Core WLP genes with associated locus tag, functional annotation, and insertion index values 263
Table 34 – Gene clusters associated with ATP generation and associated ion gradient with associated locus tag, functional annotation, and insertion index values 267
Table 35 – Gene clusters associated with electron donor regeneration with associated locus tag, functional annotation, and insertion index values 269
Table 36 – Genes and gene clusters associated with the synthesis of industrially relevant products with associated locus tag, functional annotation, and insertion index values 270
Table 37 – Insertional mutants with the greatest increase in prevalence between the third and fourth subcultures on methanol 273
Table 38 – The eight genes with highest increase in mutant abundance during *E. limosum* NG-6849 glucose growth. 288
Table 39 – *E. limosum* NG-6849 gene targets for TraDIS complementary knockouts 296
Table 40 – *E. limosum* NG-6849 genes predicted as essential on methanol but not on glucose 313
Table 41 – progressiveMauve identified Gap regions between the *E. limosum* NG-6849 genome consensus determined within this study, and the published *E. limosum* ATCC 8486 genome 317

1 Introduction

1.1. Biotechnology and sustainable development

The rapid consumption of the earth's finite fossil fuel reserves and the resulting greenhouse gas emissions have been a critical issue at the forefront of the global industrial, political, and public consciences for decades. Whilst every sector embraces innovation as a means of overcoming this issue and mitigating the resultant climate change, growing populations and continued industrialisation are compounding the challenge.

In 2015, 196 parties committed to the Paris Climate Agreement, with the fundamental goal of limiting the overall global temperature rise to 1.5^{o}C (UNFCC, 2015). The United Kingdom was one of these original parties and hosted the United Nations Climate Change Conference of the Parties (COP26) in 2021. To underpin the commitments made by signing the Paris agreement the UK originally outlined a goal of 80% reduction in greenhouse gas emissions by 2030, compared to 1990 levels (UK Government, 2008). This goal was amended in 2019 on the guidance of the non-governmental Committee on Climate Change, extending the target to at least 100% (Climate Change Committee, 2019).

The UK have made strides towards this ambitious goal by several means, including the implementation of the Renewable Transport Fuel Obligation and the CRC Energy Efficiency scheme. While these familiar policies target the energy and transport-fuels sectors as principal producers of greenhouse gas emissions, the Carbon Plan published by the government in 2011 also emphasises the importance of decarbonising integral industrial processes (Department of Energy and Climate Change, 2011). This is further highlighted in the UN's sustainable development goals which focus on establishing a "cleaner and more resilient economy". This development of such an economy is reliant on technological innovation to generate novel, sustainable industrial processes to replace current practices and move us towards a carbon neutral

circular economy (United Nations, n.d.). Biotechnology is at the centre of this innovation, leveraging an understanding of biological systems and technological advances concurrently to provide alternatives to current unsustainable, environmentally detrimental practices. As biotechnological alternatives become more prevalent replacements within current industries the bioeconomy continues to expand rapidly.

This bioeconomy represents the economic potential of bioscience through both the use of renewable biological resources and application of bio-based processes. The Department for Business, Energy and Industrial Strategy and Biotechnology and Biological Science Research Council (BBSRC) report "Evidencing the bioeconomy" estimates the gross value contribution of the bioeconomy to the UK at £220bn (Bauen, et al., 2016). Following this the UK government have released the policy paper "Growing the Bioeconomy" which outlines a strategy to support further development, aiming to double this contribution to £440bn by 2030 and ensure the UK bioeconomy remains a world leader in expertise (Department for Business, Energy & Industrial Strategy, 2018). Beyond governmental funding a change can also be seen in the investment landscape, with the European private sector raising $3.3 billion in 2019, 35% more than in 2018 (Senior, 2020). With expanding foundational support and investment in biotechnological research, great strides are being made in decarbonisation, particularly within the replacement of the petroleum industry.

1.2. Alternative feedstocks

In order to replace fossil fuels as an essential feedstock, alternative renewable feedstocks must be identified. The industry surrounding fossil fuels is extensive. Beyond the integral petroleum derived transport fuels, petrochemical commodities are fundamental resources in diverse industries. These chemicals act as carbon building blocks for the production of pharmaceuticals, solvents, fertilisers, pesticides, synthetic fragrances and

plastics (Vennestrøm *et al.*, 2011). The Chemical and Allied industries quotes use of over 63 million tons of such "carbon building blocks" from petroleum (~83%) and natural gas (~17%) for the production of synthetic fibres and plastics alone in 2008 (Chemical and Allied Industries, 2009).

A suitable replacement feedstock with scope for both fuel and chemical commodity generation must be abundant, cheap and sufficiently rich in accessible carbon so as to produce high product yields. One of the largest hurdles for sustainable manufacturing approaches is that at the commercialisation stage they remain in direct competition with the petrochemical industry. The petrochemical industry benefits from established infrastructure, both physically and legislatively. As such, to ensure economic competitiveness, feedstock choice is even more paramount; not only contributing to the long-term viability and carbon cost of the process, but also constituting up to 70% of the overall production costs (Chemical and Allied Industries, 2009).

1.2.1. Biomass

To date a great deal of focus has been placed on biomass as an abundant and carbonaceous raw material which could replace petroleum. Biomass can be derived from a multitude of sources and consequently varies significantly in composition. Previous commercial success has been achieved using biomass as the feedstock for industrial scale bioethanol production (Haghighi Mood *et al.*, 2013). These cellulosic bioethanol plants represent a boom in first generation biofuels and relied on easily degradable carbohydrate substrates, such as corn starch and sugar (Balan and Venkatesh, 2014). Although such feed was a cheap and relatively abundant biomass feedstock it was not sustainable, as it was grown for purpose and hence competed for land use with growth food crop growth. Food security is another key crisis issue compounded by population growth and the competition with this equally recognised issue is what inevitably led to the decline in these processes. The subsequent shift towards

second generation biofuels, which focus on using biomass derived from waste streams, is a clear example of the shift towards circular economy approaches in sustainable development.

Considering this, biomass derived from agricultural or municipal solid waste (MSW) streams is of particular interest. These feedstocks are established waste streams and are therefore both cheap and abundant feedstocks with good economic margins for valorisation (Vennestrøm *et al.*, 2011; Daniell, Köpke and Simpson, 2012). The accessibility of the carbon within this more lignocellulosic biomass is significantly limited by its heterogeneity and the recalcitrance of the lignin component. Thus far to allow microbes to ferment such bound polysaccharides, the feedstock must undergo multiple expensive and energy intensive pre-treatment steps such as CO_2 explosion, ammonia fibre expansion, ozonolysis and hydrolytic enzyme treatment (Brown and Brown, 2013; Haghighi Mood *et al.*, 2013).

Research into improving accessibility of this carbonaceous biomass is focused on isolation of new organisms with novel lignin degradation pathways, such as bacterial symbionts in wood-feeding insects (Berasategui *et al.*, 2016). Genetic engineering is also being employed to develop a 'consolidated bioprocessing' approach where these lignin degradation pathways are introduced into industrially relevant strains. So far these limitations have prevented commercial competitiveness to be demonstrated in an industrial setting (Brown & Brown, 2013). Consequently, there have been further innovations in biofuel production, and development of new third generation biofuels. Third generation biofuels are extracted from oils within processed algal biomass, generating higher energy density fuels and directly fixing carbon dioxide through photosynthetic means. The resource requirements for algal biofuels are their inherent limiting factor with high water usage, unique sunlight permeation challenges and high nutrient requirements reducing the viability of a commercial scale up at this stage, as well as the necessary development of low-cost, low-environmental impact extraction methods (Saad, et al., 2019).

1.2.2. Biomass gasification

One promising mechanical means of improving carbon availability in lignocellulosic biomass is through conversion to single carbon (C1) compounds by gasification. Gasification uses high temperatures (600-1000^{o}C), high pressure (>30 bar) and an oxidising agent (steam or air) to convert virtually any carbonaceous material into gas. The composition of this gas is dependent on the material type as well as variations in the process, but it is primarily formed of carbon monoxide, carbon dioxide and hydrogen (Spivey et al., 2007).

Although gasification does require an initial energy input, it has been shown to be self-sustaining at steady state (Handler et al., 2015). Within lignocellulosic biomass the proportion of lignin may be as high as 40%. Gasification of biomass cleaves this lignin, allowing utilisation of this previously inaccessible carbon. Gasification is also reported as efficient, with an overall rate of energy conversion of 75-80% (McKendry, 2002). The gaseous mix generated by this process is called synthesis gas (syngas).

1.2.3. Syngas

Syngas is not solely produced by biomass gasification, but also as an off gas in industries including steel mills, refineries, and chemical plants. As a waste product this syngas was initially used for internal heat or energy generation by on site combustion, further contributing to the intensive carbon footprint of the industry (Munasinghe and Khanal, 2011; Heijstra, Leang and Juminaga, 2017).

Alternative valorisation of syngas can occur through the Fischer-Tropsch thermochemical process. This process was developed in 1925 and converts the gases into medium and long chain hydrocarbons. The process is intensive, requiring high temperatures (150-350^{o}C), elevated pressures (30 bar), and heterogeneous catalysts such as cobalt, ruthenium, and iron (Maitlis, 2013). These catalysts are expensive, have minimal selectivity (45%), require high

levels of syngas purity, and only operate at fixed gas ratios. Such restrictions necessitate expensive pre-treatment processes which can account for 60-70% of operating costs and reduce the overall profit margin (Daniell, Köpke and Simpson, 2012; Griffin and Schultz, 2012).

1.2.4. Methanol

Methanol is another alternate C1 feedstock of industrial interest, with an estimated 125 billion litres produced in 2019. Of this methanol approximately 75 billion litres (60%) were used in chemical synthesis of products including formaldehyde, acetic acid and ethylene (IRENA and Methanol Institute, 2021). As a feedstock, methanol offers many advantages over syngas, including a lower occurrence of toxic contaminants, easier and more affordable transport and storage as well as higher solubility, increasing substrate availability within aqueous mixes. Thus far, methanol utilisation has been significantly impeded by cost, with the feedstock in methanol-derived single cell protein generation representing approximately 50% of the total process costs (Nasseri, et al., 2011).

As sustainability is afforded more weight in the dialogue, technological advances in the production of renewable methanol from low value waste sources may provide a solution to this economic challenge. Several companies commercially produce this renewable methanol, otherwise known as biomethanol, including BioMCN, Enerkem and VarmLandsMethanol (Law, et al., 2013). The primary source of biomethanol is biomass, gasified to produced syngas and then converted to methanol through a thermochemical catalytic process. The catalysts and processes used vary, and most are patented or licensed systems with limited publicly available detail, such as the DAVY™ technology pioneered by Johnson Matthey. Alongside biomass, other sources of renewable methanol are also being explored including glycerol, produced as a cheap and abundant by-product of biodiesel manufacture (van Bennekom, et

al., 2012), carbon dioxide derived from atmospheric capture (Kothandaraman, et al., 2016) and industrial waste streams (Zang, et al., 2021).

1.3. Microbial catalysis

Microbial catalysts provide an alternate route for C1 feedstock valorisation. While biotechnology is a relatively modern term, the exploitation of microbes has occurred throughout history, i.e., with yeast fermentation to produce alcohol and bread. Following industrialisation microbial catalysts became integral to the production of key chemical agents such as penicillin (Kocholaty, 1942). Reframed as modern-day biotechnology, the scope of applications has broadened with a range of new targets including biofuels, bulk chemicals, and biological therapeutic agents. The appeal of microbial catalysts lies in their sustainability, product/substrate specificities and relatively low energy demand when juxtaposed with most chemical synthesis approaches (Latif, et al., 2014). Microbial syngas catalysis is a prime example of this, operating at markedly reduced temperature and pressure compared to the Fischer-Tropsch thermochemical process. Microbial syngas catalysts also demonstrate a greater tolerance to gas ratio variability and impurities, allowing greater flexibility in the root biomass source (Brown & Brown, 2013). As a living, growing catalyst, they are also able to regenerate and recover from contamination or equipment failure, eradicating the costly cleaning requirements of rare metal catalysts (Heijstra, et al., 2017).

Microbial syngas fermentation has already proved industrially scalable, with multiple companies operating pre-commercial plants. The most noted company within this field is LanzaTech, founded in 2005 in New Zealand and now based in Illinois, Shanghai and Gurgaon. LanzaTech's technology uses *Clostridium autoethanogenum* to fix carbon monoxide and carbon dioxide from industrial emissions into ethanol. Their first fully commercial waste-gas-to-bioethanol facility, with a 46,000-ton annual capacity, was established in 2018 at Jingtang Steel Mill in Hebei Province, China (Burton, 2018). LanzaTech is also

expanding into a range of new chemical products including butadiene, nylon, aviation fuel and surfactants. This innovation is fostered through numerous partnerships with companies such as INVISTA, SkyNRG and Unilever, and employs their established metabolic toolkit for new pathways introduction and native carbon flux negation within the acetogenic bacteria itself (Lane, 2015).

While this commercial success demonstrates the fundamental feasibility of the economics of microbial syngas fermentation, life cycle assessments (LCA) provide a means of assessing associated environmental impacts. Covering the complete process pipeline LCAs aim to accurately evaluate and contextualise the estimated environmental benefits of sustainable technology adoption. Several LCAs have been published based upon the exemplified LanzaTech process, estimating between 50% and 98% greenhouse gas emission reduction compared with current processes (Ou, et al., 2013; E4tech, 2016).

Thus far, microbial methanol fermentation has only been used at industrial scale for single cell protein production. For example, 'Pruteen', a bacterial protein marketed as animal feed, is harvested from dead *Methylophilus methylotrophus* cell mass produced during large-scale fermentation performed by Imperial Chemical Industries (Schrader, et al., 2009). Unlike syngas fermentation, methanol fermentation is not limited by low volumetric mass-transfer coefficients and therefore has higher substrate availability in the culture media (Behr, 2014).

1.4. Acetogenic bacteria

Acetogens are a diverse group of bacteria with over 100 isolated species in 23 genera, primarily *Clostridium* and *Acetobacterium*. This group represent a novel niche which is metabolically adapted for autotrophic growth on C1 compounds, including methanol, CO, formate, and CO_2(:H_2) (Shin, et al., 2016). This carbon fixation and energy production process generates acetyl coenzyme A (acetyl-CoA) and subsequently acetate via phosphotransacetylase and acetate kinase

enzymes. This results in characteristic high yields of acetate by most acetogenic bacteria. Acetate is itself employed as a petrochemical precursor to a number of industrial products including polyvinyl acetates.

The first pure culture of an acetogenic bacteria reported was *Clostridium aceticum*, isolated in 1936. Unfortunately, this strain was lost in the 1940s and was only successfully isolated again after three decades (Adamse, 1980). As such the characterisation of the enzymatic pathways responsible for this autotrophic growth were elucidated with *Clostridium thermoaceticum* (now *Moorella thermoaceticum*). This research was performed by Harland Wood and Lars Ljungdahl, after whom this pathway is named. This pathway has often been heralded as a key step in the development of microbial life on earth, and its early adoption has been proposed to explain the diversity of ecological niches in which acetogens can now be found (Martin, 2012; Braakman & Smith, 2012).

This diversity also provides a wealth of potential microbial catalysts for C1 feedstock valorisation, each with their own distinct advantages and shortcomings. Some species naturally produce products with commercial potential, alongside the ubiquitous acetate, such as butyrate, butanol, 2,3-butandiol and even ethanol. Together with the commercial potential of these native products, product profile enhancement through targeted strain development and non-native pathway inclusion is an attractive research avenue. A notable impediment to this is the relative immaturity of genetic manipulation within acetogens.

1.5.1. The Wood-Ljungdahl Pathway

Whilst acetogens can be identified by high acetate production and chemolithoautotrophy, these characteristics are not integral to their classification; acetogens are defined by their use of the Wood-Ljunghdahl pathway (WLP). This pathway is a linear mechanism for fixing cellular carbon, from CO_2 or CO, and has been referred to as the “most efficient non-

photosynthetic carbon fixation mechanism" (Fast & Papoutsakis, 2012). The WLP has been reviewed in great detail by a number of groups, with a recent focus on the integration of heterotrophic carbon sources and synthesis of extended product profiles (Bertsch & Müller, 2015; Ragsdale & Pierce, 2008; Bengelsdorf, et al., 2016; Schuchmann & Müller, 2016).

The WLP is formed of two branches, the methyl (eastern) and the carbonyl (western) (**Figure 1**). The methyl branch uses electrons from H_2, CO or heterotrophic sources (via NAD(P)H and ferrodoxin) to reduce CO_2 to a corrinoid iron-sulphur protein (CoFeSP) bound methyl moiety. The first enzymatic step is catalysed by a tungsten and selenocysteine containing formate dehydrogenase which reduces carbon dioxide to formate. A formyl-THF synthase then activates this by condensation with tetrahydrofolate (THF), producing formyl-THF and consuming one molecule of ATP. This is the only ATP dependent step of the WLP. Outside of the WLP formyl-THF is also an important component of amino acid and nucleotide synthesis. Within the WLP this formyl-THF is reduced to methyl-THF through a series of reactions catalysed by a formyl-THF cyclohydrolase, methylene-THF dehydrogenase and a methylene-THF reductase. This five-electron reduction pathway yields methyl-CoFeSP; a methyl moiety bound to a corrinoid iron-sulphur protein (Latif, et al., 2014; Ragsdale & Pierce, 2008).

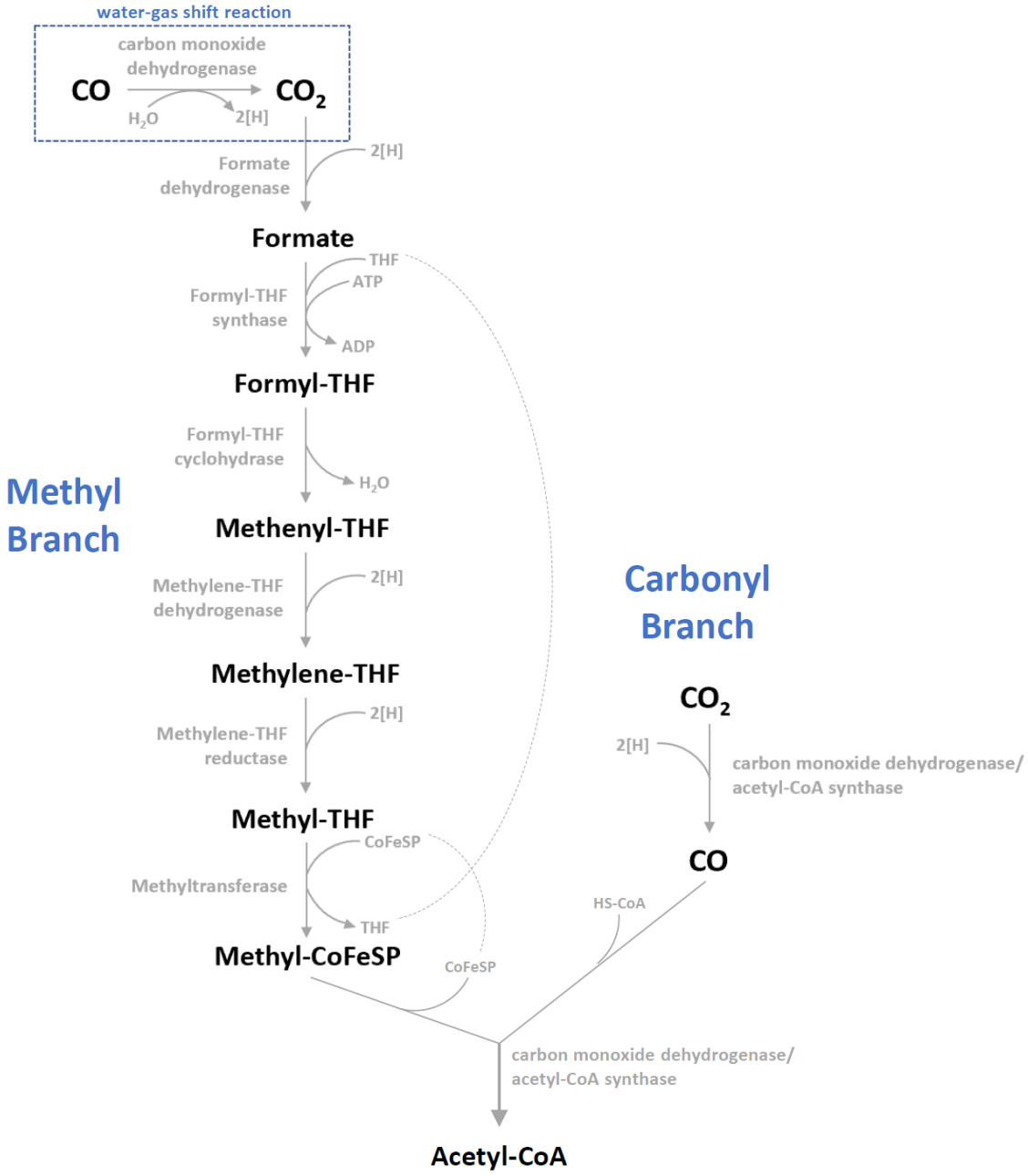


Figure 1 - Wood Ljundahl pathway schematic
Schematic indicating the core enzymatic reactions of both the carbonyl and methyl branches of the WLP, to the point of acetyl-CoA generation. Also included are respective cofactors necessary within each reaction, subsequent cofactor recycling, and CO_2 generation mechanism under exclusive CO conditions.

The carbonyl branch consists of a carbon monoxide dehydrogenase which produces CO from CO_2. This CO is then fixed to the methyl moiety generated by the methyl branch, as well as a molecule of coenzyme A, by the bi-functional carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS) enzyme

complex. This produces acetyl-CoA, a significant carbon flux point from which integral products for biomass accumulation are generated, alongside solvents and organic acids. Exclusive growth on CO is facilitated by the generation of the CO_2 necessary for the methyl branch through a water-gas shift reaction catalysed by the CODH. Hence one molecule of ATP and 8 electrons are used to produce acetyl-CoA from two molecules of CO_2, whereas one molecule of ATP and four electrons are used to produce acetyl-CoA from two molecules of CO (Diender, et al., 2015). To generate acetate, a phosphotransacetylase and acetate kinase enzyme phosphorylate the acetyl moiety of acetyl-CoA, and cleave the CoA which can then be recycled within the pathway. The ATP investment of the two branches is balanced via substrate level phosphorylation during this acetate kinase reaction, which accounts for the high propensity for acetate production in acetogens. The ΔG^0 of acetate production from H_2/CO_2 and CO is -95 kJ/mol and -175 kJ/mol, respectively.

$$4\ H_2 + 2\ CO_2 \rightarrow CH_3COOH + 2\ H_2O$$

Ethanol is another ubiquitous fermentation product generated from acetyl-CoA by an NAD+ dependent bifunctional acetaldehyde. This reaction generates an intermediary acetaldehyde that is then reduced into ethanol by an ethanol dehydrogenase.

Compared to aerobic methods of methanol fixation the WLP is attractive from an industrial perspective as it does not face biochemical limitations which preclude complete carbon fixation. Methanol metabolism via the WLP was originally best characterised in *Moorella thermoacetica* (Pierce, et al., 2008), with a more recent study elucidating the pathway within the model acetogen *Acetobacterium woodii* (Kremp, et al., 2018). To feed methanol into the WLP the initial steps are catalysed by methanol:coenzyme M methyltransferase, a 3-subunit complex denoted as MtaABC. This complex converts methanol to methyl-THF, which then proceeds through the methyl branch as previously described. There is, however, a lower flux alternate method of integration,

occurring with one in every four methanol molecules, which uses a reverse methyl branch process to produce CO_2 and six reducing equivalents [H] from one molecule of methanol. These reducing equivalents are subsequently used within the carbonyl branch for conversion of CO_2 to CO and as such, in the presence of excess CO_2, four methanol molecules yield three molecules of acetate through this process.

$$4CH_3OH + 2CO_2 \rightarrow 3CH_3COOH + 2H_2O$$

1.5.2. Energetics

There is no net yield of ATP within the WLP, even during acetate production wherein the acetate kinase generated ATP only balances that required in the methyl branch for the generation of formyl-THF. Instead, autotrophic energy production in acetogens is dependent on the coupling of redox reactions to the production of a chemiosmotic membrane gradient. This gradient in turn stimulates ATP generation through a membrane bound ATPase. In this regard, the WLP can be regarded as an energy-neutral cytoplasmic system for the oxidation and reduction of electron donor pools. While the electron donors employed within these redox reactions vary between organisms one integral electron donor throughout acetogenic bioenergetics is ferredoxin (Fd). With a relatively low potential, Fd^{2-} is the only donor appropriate for the low redox potential of the CO/CO_2 couple. The subsequent reduction of this oxidised ferredoxin, using NAD+, is performed by membrane bound electron transfer complexes. There are two classes of these complexes described to date, the first of which, termed the Rnf complex, was discovered in *A. woodii* (Müller, et al., 2008). In catalysing this Fd reduction, the ferrodoxin: NAD^+ oxidoreductase (Rnf) complex also translocates either a sodium ion or proton across the membrane. While Rnf complexes have been more widely described within acetogens a further membrane bound ferrodoxin: H^+ oxidoreductase (Ech) system has since been recognised (Hedderich & Forzi, 2005).

Although the basic carbon fixation chemistry of the WLP is well conserved within acetogens, an astonishing repertoire of enzymes catalyse these reactions each with distinct bioenergetic considerations. For example, the reduction of formate within *Moorella thermoacetica* uses NADPH whereas *A. woodii* relies on H_2 (Wang, et al., 2013). To establish the bioenergetics of a novel organism, it is therefore important to consider a broad range of key features: the reductants necessary for enzyme function; whether the methylene-THF reductase uses ferredoxin-based electron bifurcation; what transhydrogenases are present, and; which chemiosmotic enzyme is used to produce ATP (Bertsch & Müller, 2015). The energy conservation system of *A. woodii* is the best characterised thus far with reductants elucidated for each redox step. *A. woodii* also contains a cytoplasmic FeFe-hydrogenase which oxidises molecular hydrogen and produces NADH and Fd^{2-}, a soluble caffeyl-CoA reductase and a F_1F_0-type ATPase (Bertsch, et al., 2013; Schuchmann & Müller, 2012).

A thorough understanding of an organism's energy demands in relation to its metabolism provides an essential framework for determining constraints which dictate the efficacy of non-native pathway introduction (Diender, et al., 2015). Bioenergetic constraints also have a significant effect on carbon flux, such as the predicted negative energy balance of using H_2 as an electron donor leading to higher acetate yields under a CO_2:H_2 headspace. As such the ATP homeostasis and energy balance of the microbe should be considered in the context of their effects on metabolic carbon flux (Valgepea, et al., 2017).

1.5. Eubacterium limosum

E. limosum is one of a limited number of acetogens which naturally produce the four carbon (C4) compound butyrate from C1 substrates. It is a rod-shaped, Gram-positive, obligate anaerobe from the class Clostridiales. It has the capacity for autotrophic growth on CO, CO_2:H_2 and methanol, as well as chemoorganotrophic growth on varied sugars. The type strain ATCC 8486, was

isolated from human faeces (Eggerth, 1935). Other strains have subsequently been isolated from similar anaerobic environments including the gastrointestinal tracts of animals, anaerobic digesters, and sewage sludge.

To date three strain genomes have been published: ATCC 8486 and 81C1 isolated from human faeces (Honda, 2019; Song, et al., 2017), and SA11 isolated from sheep rumen (Kelly, et al., 2016). These genomes measure 4.37 Mb, 4.74 Mb, and 4.15 Mb, respectively. The previously identified 32_A2 strain (Wrighton, 2015), isolated from an underwater shale carbon reservoir, is now regarded as a *Eubacterium maltosivorans* strain, and the previous well studied KIST612 strain has been reclassified as *Eubacterium callanderi*. The primary research strain within this report, NG-6849, was kindly supplied by Prof. Philippe Soucaille, INSA, University of Toulouse, France. No genome has been published for this strain at the time of writing.

1.6. Clostridium carboxidivorans

Only one strain of the species *C. carboxidivorans* has been reported: P7 (ATCC BAA-624, DSM15243). This organism is a Gram-positive, acetogenic, obligate anaerobe, originally isolated from an agricultural settling lagoon in Oklahoma, USA (Liou, et al., 2005). It has the capacity for both autotrophic growth on CO, and CO_2:H_2, as well as chemoorganotrophic growth on variable sugars including xylose, arabinose and cellobiose (Phillips, et al., 2015). *C. carboxidivorans* exhibits characteristic biphasic acetogenic fermentation. This results in an initial accumulation of carboxylic acids occurring concurrently with biomass generation in the exponential phase of growth. The following stationary growth phase, prompted by nutrient limitation and other stress factors, shifts metabolic flux towards solventogenesis and alcohol production (Fernández-Naveira, et al., 2016; Fernández-Naveira, et al., 2017). The complete genome was published in 2015 with a 5.75 Mb chromosome and a 19 kb cryptic plasmid (Li, et al., 2015).

C. carboxidivorans P7 contains a complete WLP gene cluster, but further to this it also produces long-chain acids and alcohols such as butyrate/butanol and hexanoate/hexanol. This product profile has made it an attractive target as a potential industrial chassis to valorise C1 into C4 or C6 compounds. Prior to 2019 no DNA transfer into the organism was reported and hence research focused on extensive empirical exploration of growth parameters with an emphasis on improving alcohol titre. These parameters include pH effects (Fernández-Naveira, et al., 2017), temperature effects (Ramió-Pujol, et al., 2014), product inhibition (Fernández-Naveira, et al., 2017; Fernández-Naveira, et al., 2016; Zhang, et al., 2016) media and trace metal variations (Han, et al., 2020) and the impact of syngas impurities (Rückel, et al., 2021; Ahmed, et al., 2006). A number of different reactors have also been used, including hollow fibre (Shen, et al., 2014), membrane based (Shen, et al., 2014) and packed bed (Shen, et al., 2017). A recent review of *C. carboxidivorans* P7 by Fernández-*Naveira et al* (2017) summarised the key parameters explored to date in terms of the optimum conditions for long-chain alcohol production.

1.7. Strain development

Despite the appeal of acetogens as industrial chassis to utilise C1 feedstocks, genetic engineering technologies within the group remain relatively immature (outside of key model organisms such as *C. autoethanogenum* and *C. ljungdahlii*). This lack of established tools when compared to foundational model strains like *E. coli* has been attributed to factors such as low DNA transfer efficiencies, lower recombination efficacies and complex culturing conditions (Pyne, et al., 2014). The application of promising newly isolated bacterial species as industrial chassis is garnering greater attention as the bioeconomy develops (Heijstra, et al., 2017). This interest has stimulated research into streamlining the approach and methods implemented to rapidly develop novel microbes as amenable chassis for research and potential commercial application. Multiple roadmaps and guides have been published which outline

the key steps within this process for acetogens, primarily those within the *Clostridium* genus. The three fundamental steps outlined are genomic characterisation, establishment of robust genetic transfer, and development of genetic tools (Pyne, et al., 2014; Minton, et al., 2016; Joseph, et al., 2018).

1.7.1. Genome

A complete and annotated genome sequence is foundational in developing the understanding of an organism. Application of this data can inform all research including reverse genetic engineering approaches and metabolic pathway modelling. Sequencing technologies have improved exponentially since the Sanger chain termination method was initially demonstrated (Sanger, et al., 1977). These technological advances have significantly increased the accessibility of whole genome sequencing, reducing previously prohibitive costs and accelerating the overall timeline. These new technologies are coined as next-generation sequencing (NGS) platforms and employ high-throughput parallel processing, allowing sequencing to occur on millions of DNA strands simultaneously. Different NGS technologies have varying error rates and average read length, with Illumina delivering >99.9% read accuracy but an average read length of 50-700 bp (Liu, et al., 2012). As such, Illumina is the standard for most high-fidelity applications. *De novo* genome assembly, as is necessary in novel strains, can however reach only limited success with Illumina data as longer repeated sequence regions prevent the confident overlapping of contig sequences. Other NGS technologies such as PacBio and Nanopore can deliver much higher read lengths, with PacBio's Single Molecule Sequencing Technology (SMRT) quoted with reads in excess of 15000 bp. These technologies employ different approaches, but both show a reduced read accuracy compared to Illumina with SMRT reporting 87% and Nanopore reported at >97-94% (Tyler, et al., 2018; Koren, et al., 2013). *De novo* assembly is hence usually performed with these longer reads, with subsequent error correction through the alignment of Illumina data.

Following assembly, genomes are then annotated. As our databases and understanding of bacterial metabolism expand, automated annotation pipelines continue to improve. A number of freely available software packages exist to execute such pipelines, such as RAST (Aziz, et al., 2008). Such pipelines are invaluable, but still require complementation with manual curation due to prevalence of genes of unknown function and the reattribution of poorly annotated homologues in closely related species.

1.7.2. DNA transfer

There is noted recalcitrance to genetic transfer within the *Clostridium* genus, with no successful DNA transfer or low transfer rates reported for several species. The factors implicated in such recalcitrance are diverse but can broadly be divided into passive and active. Passive impediments to DNA transfer are primarily physical such as cell walls and cytoplasmic membranes. All *Clostridium* are Gram-positive and as such possess extensive peptidoglycan networks, limiting opportunities for DNA-inner cell membrane interaction. Passive barriers must be transiently disrupted, to a sufficient level to facilitate DNA internalisation while not jeopardising the viability of the cell. There are numerous DNA transfer methodologies employed to generate this disruption including electroporation, sonoporation and conjugation. The range of methodologies applied within *Clostridium* is vast, with variability not only in approach but the many specificities therein, such as pre-treatments, electrical pulse parameters or conjugal donors. As such, establishing a robust transfer methodology often requires a dedicated study in which the approach is extensively tailored to optimise yield.

The two best characterised active barriers which impede DNA transfer are restriction-modification (RM) systems and clustered regularly interspaced short palindromic repeats (CRISPR) systems. These both act as rudimentary prokaryotic defence mechanisms by discriminating between self and non-self DNA, and targeting the latter with nucleolytic activity. CRISPR systems target

foreign DNA through the sequence specific base-pairing between guiding RNA molecules and complimentary foreign DNA, exposing this DNA to the associated endonuclease activity of the protein complex. The guiding RNA sequences are stored in a CRISPR loci and provide an archive of previously encountered foreign DNA. This targeting is hence highly specific to acquired target sequences but can be expanded, acting as an immune analogous adaptive memory (Ran, et al., 2013).

RM systems are composed of DNA methyltransferases and restriction enzymes and are found in an estimated 90% of sequenced bacterial genomes (Johnston, et al., 2019). These systems target foreign DNA based on recognition of foreign nucleotide modification patterns within specific recognition sequences. In prokaryotes these modifications are the methylation of adenine and cytosine bases. There are four types of RM system, types I-III cleave DNA based on a lack of methylation, whereas type IV systems cleave inappropriately methylated DNA and do not possess corresponding methyltransferases. RM systems target largely shorter sequences and have less specificity than CRISPR systems so are thought to present the greatest barrier to DNA transfer in intractable organisms (Johnston, et al., 2019; Zhang, et al., 2012).

These challenges are drawing increased attention with several papers describing approaches which mitigate the impacts of these active barriers. These approaches include background methylation selection through conjugal or cloning DNA donor choice (Woods, et al., 2019; Kolek, et al., 2016), native RM system mimicking (Zhang, et al., 2012; Yang, et al., 2016), RM and CRISPR system evasion (Johnston, et al., 2019; Riley, et al., 2019; Purdy, et al., 2002) or CRISPR/RM restriction activity negation through gene disruption (Pyne, et al., 2014; Dong, et al., 2010; Lesiak, et al., 2014).

1.7.3. Genetic tools in *Clostridium*

To select for DNA transfer recipients, exogenous DNA typically includes selectable markers. These markers are genes which confer a selectable

phenotypic change, such as antibiotic resistance. Common examples used within *Clostridium* include the *C. perfringens* derived *catP* and *ermB* genes which each confer resistance to thiamphenicol/chloramphenicol (Tm^R/Cm^R) and clarithromycin/erythromycin (Cla^R/Erm^R), respectively. To ensure growth under selective conditions this exogenous DNA must hence be maintained within the cell, and vertically transferred to each daughter cell. Integrative plasmids, otherwise known as suicide vectors, cannot replicate within the target host and therefore rely on genomic incorporation. To achieve genomic integration intracellular DNA repair machinery is exploited typically through homologous recombination. Integration in some instance can be further supported by the introduction and expression of alternate DNA repair associated genes, such as λ-red recombineering approaches (Yu, et al., 2000). High DNA transfer and homologous recombination efficiencies are necessary for such approaches and as such they are not widely applicable within *Clostridium*.

More prevalent within clostridial studies is the use of self-replicating plasmids, which exploit host machinery to replicate in an independent manner from the host genome. The replication of these plasmids is dependent on the origin of replication and associated genes, or replicon. Replicons also dictate the fundamental stability of the plasmid maintenance, where stable maintenance within a population requires plasmid replication and partition to each daughter cell during cell division. Certain replicons have been implemented relatively broadly within the clostridial species, including pBP1 derived from *C. botulinum*, and pCB102 derived from *C. butyricum*. Despite their prevalence the host range of each replicon is varied, with some origins of replication exhibiting no functionality when implemented in a novel organism. Confirming the efficacy of multiple Gram-positive replicons within a strain is ideal, as features such as copy number and resultant growth/metabolic burden can vary significantly. This also facilitates the use of two-plasmid systems where compatible replicons can simultaneously maintain two vectors within the same host. Where DNA transfer is limited by vector size the use of two-plasmid

systems increases the total size and number of exogenous genes which can be expressed, expanding the complexity of systems which can be implemented *in vivo*. Further research has also led to the development of conditional replicons, with inducible instability, and pseudo-suicide replicons whose low copy number leads to segregational instability (Molitor, et al., 2016). This instability limits bacterial population growth as each daughter cell does not inherit a vector and therefore cannot survive under the selective conditions.

Another consequence of the low DNA transfer rates observed in *Clostridium* is the necessity for standard molecular cloning to first be performed in *E. coli* prior to its application in the target strain. As such, most vectors must contain an additional Gram-negative replicon and ensure that the selection marker is compatible with both hosts. These key requirements have led to the generation of a range of clostridial shuttle vectors with interchangeable modular components for species specific optimisation, such as the pMTL80000 series (Heap, et al., 2009). Within this vector series the Gram-negative replicons also contain an optional origin of transfer (*oriT*) and *traJ* gene to facilitate conjugative transfer of the vector from the *E. coli* donor strain. Besides these fundamental features essential for basic DNA transfer, maintenance and selection, a "toolbox" of well-characterized biological parts is instrumental in further research and strain development. There are various features to consider when expanding the toolbox, including promoters, terminators, ribosomal-binding sites, and reporter genes.

Despite their importance, promoter activities have not been extensively studied in most acetogens. Considered as modular genetic components, promoters are the simplest way to regulate gene expression, therefore it is prudent to have a validated suite available with different constitutive strengths. Most clostridial studies employ highly similar promoters, derived from a few initial *Clostridium* strains. These promoters include the P_{thl} natively driving thiolase gene expression in *C. acetobutylicum* and P_{fdx} driving ferredoxin gene expression in *C. sporogenes*. Efforts to expand the available sequences can be performed directly within the target strain, using rational prediction of

high expression associated genes and subsequent reporter-based validation. A higher throughput identification approach has also been demonstrated using transcriptomic analysis of intracellular RNA levels within *C. ljungdahlii*. This led to the identification of the P_{pta} promoter driving *pta* expression, which was subsequently implemented in the metabolic engineering of increased butyrate production (Ueki, et al., 2014). Such approaches are relatively intensive and time-consuming, and instead, further expansion of the available promoters has been largely based on mutagenesis, through error-prone PCR or other methods. A recent suite of promoters was generated through the rational randomisation and conservation of the P_{thl} promoter sequence, with subsequent synthetic promoter validation and assessment within *C. acetobutylicum* and *C. ljungdahlii* (Yang, et al., 2017).

Inducible promoter systems are also crucial genetic tools, modulating gene expression and repression in a scalable manner and permitting the cloning of otherwise prohibitively toxic genes. The primary hurdle in inducible expression systems is the stringency of repression. Insufficient levels of repression within an organism, as seen with the lactose-inducible *bgaR/*P_{bgaL} system (isolated from *C. perfringens*) in *C. ljungdahlii*, limit its confident application and negate its utility in further tool development (Woolston, et al., 2018). Tetracycline-inducible systems have shown better repression in several acetogens with the P_{2tetO1} promoter, first demonstrated in *C. acetobutylicum*, offering tighter expression in *C. ljungdahlii* (Dong, et al., 2012). Similar results have been observed in *A. woodii* (Back, et al., 2020) and *E. limosum* (Shin, et al., 2019), and as well as with Tet3nO based variants in *C. autoethanogenum* (Walker & Köpke, 2015; Nagaraju, et al., 2016). An interesting inducible system variation was applied for transposase induction within *C. autoethanogenum* by Woods *et al.* (2021). Within this system a chromosomally integrated sigma factor, TcdR, was controlled by the lactose inducible *bgaR*-P_{bgaL} system. The introduced transposon vector relied on P_{tcdB} to drive transposase activity. In the absence of the TcdR sigma factor, such as within the *E. coli* cloning strains, the expression is therefore repressed. Within *C. autoethanogenum* P_{tcdB} expression

was induced by TcdR expression, which was in turn induced by lactose supplementation. The combination of these systems generated a unique indirect orthogonal inducible expression system.

Reporter genes encode various proteins, each yielding a scalable measurable phenotypic change within the organism. This phenotypic change is then measured as a proxy for gene expression levels and therefore corresponding promoter strength. Initial acetogenic reporter genes were typically enzymatic, requiring cell lysis and tailored enzymatic assays, such as the chloramphenicol acetyltransferase (*catP*) assay employed within this study. Fluorescence based systems were not readily implemented within acetogens due to the necessity of oxygen for the fluorophore to function. Recent advances, however have led to the development of three anaerobic fluorescence-based reporter systems. As these systems operate without cell lysis they can be applied for *in vivo* studies, facilitating real-time and high throughput expression analysis. The fluorescent-activating and absorption-shifting tag (FAST) system employs multiple fluorogenic ligands which, when bound to the expressed reporter protein, fluoresce highly within a given range (Steett, et al., 2019). Two further protein-ligand systems, HaloTag and SNAP-tag, have also been demonstrated within *C. ljungdahlii* (Charubin, et al., 2020). The variant fluorogenic ligands and multiple systems afford multiplexing opportunities for simultaneous dual function appraisal within cell cultures.

1.8. Genomic modification in *Clostridium*

Once DNA transfer has been established there is scope to implement *in vivo* genome modification tools. These tools enable the application of reverse genetics approaches to rationally target and directly manipulate the genome. Such tools can be used to further strain understanding by performing basic characterisation, or more crucially they can be used to directly enhance key strain features to improve the commercial prospects of the organism. Though the development of tools was slowed by low transformation and

recombination frequency, a few systems of note have been broadly implemented within *Clostridium*.

1.8.1. ClosTron

One of the earliest fundamental genome modification systems exemplified within *Clostridium* was ClosTron. The ClosTron system employs the targetable group II intron Ll.ltrB from *Lactococcus lactis* to integrate the *ermB* selectable marker from *Enterococcus faecalis* plasmid pAMβ1 at target locations within the genome (Heap, et al., 2007). This *ermB* gene encodes a macrolide–lincosamide–streptogramin B resistance protein, conferring resistance to the macrolide antibiotics clarithromycin and erythromycin.

Group II introns are self-catalytic RNAs which excise themselves from RNA transcripts and insert themselves at new chromosomal sites, to be subsequently integrated through reverse transcriptase and DNA ligase activity. To expedite the process and overcome low transposition frequencies a conditional antibiotic resistance system was developed. This system used an orientation specific *td* group I intron sequence to disrupt the *ermB* gene within the plasmid. This group I intron cannot be excised from the plasmid but can be excised when the juxtaposing orientation RNA transcript is generated within the group II intron insertion process. As such the group II intron system then chromosomally integrates an intact *ermB* gene and stable mutants resistant to erythromycin and clarithromycin can be selected for. This intricate selection approach was coined as a retro-transposition activated marker.

This approach offered a rapid, selectable gene disruption approach without stability concerns, as previously prevalent in the use of single crossover chromosomal integrations. It was also demonstrated to have broad range within *Clostridium* exemplified in *C. acetobutylicum*, *C. sporogenes, C. botulinum* and the now reclassified *Clostridiodes difficile* (Heap, et al., 2007). While the ClosTron and its selection is a remarkable system to expedite mutant isolations and overcome some of the low efficiency hurdles observed within

the genus, it only offers limited scope. The integration of *ermB* cannot be performed iteratively, and despite suggestions to remove the marker and manually screen, implement other markers, or use FRT sites and a flippase system to recycle the marker, these approaches have shown limited applicability and success (Kuehne & Minton, 2012). Further to this as insertional tool it may cause polar effects, it cannot be used to integrate cargo exceeding 1 kb and has shown potential instability in iterative approaches (Jia, et al., 2011).

1.8.2. Allelic exchange

Recent genomic modification systems have focused on exploitation of *in vivo* homologous recombination within the target organism. Homologous recombination is widely used by prokaryotes to repair DNA, replacing chromosomal alleles with near identical sequences. By introducing DNA with homologous sequences framing a target region, the region within the exogenous DNA can replace that at the chromosomal locus. With the non-homologous internal region this approach proceeds via two homologous recombination events, one at each homologous sequence. This approach offers unlimited targeting scope and marker-less mutant generation. Due to the low occurrence of recombination events, however, the challenge lies in selecting for or screening for each recombination event. Of the two crossover events the second is the greatest challenge, with alternate homologous recombination at the initial locus reverting the genome back to the original sequence.

Different approaches employ variable genome and plasmid-based selection/counter-selection markers to select for successful double-crossover genome modification events. The pioneering method to achieve this within *Clostridium* was allele-coupled exchange (ACE) initially demonstrated in *C. acetobutylicum* (Heap, et al., 2012). This uses a defective replicon and selects for disruption of the *pyrE/F* gene. Further methods also exist, with a Triple Cross approach patented for manipulation of *C. autoethanogenum* which uses

a combination of positive and negative selection markers to force a double crossover event, with subsequent optional selection marker recycling to make the change scarless. This was demonstrated through the generation of an in-frame deletion of a 2,3-butanediol dehydrogenase gene (Walker & Köpke, 2015).

A notable alternative rationale for selection is through the generation of double strand DNA breakages (DSBs). Genomic breakages of this nature are lethal in prokaryotes, and as such any cells with a chromosomally integrated targeted sequence are selected against. Cells in which successful recombination has removed this sequence are essentially immunised to this cleavage activity. As such selection enriches the population with mutants as those in which no double-crossover recombination has occurred are gradually killed off. One system employing this rationale was demonstrated within *C. acetobutylicum* and *C. beijerinckii* (Zhang, et al., 2015). This method used I-SceI, an intron-encoded endonuclease derived from the mitochondrial gene of *Saccharomyces cerevisiae* to generate these DSBs. The corresponding 18 bp target sequence was first introduced to the chromosome through the single crossover integration of an introduced allelic exchange plasmid. This initial crossover event relies on an antibiotic marker selection, with a second plasmid introduced expressing the I-SceI to counter-select against cells in which double crossover homologous recombination has not occurred.

1.8.3. CRISPR-Cas

CRISPR-Cas based tools have become a ubiquitous technology within genome engineering. As previously outlined, in their native context these systems act as rudimentary prokaryotic defence mechanisms by accruing short DNA sequences from foreign DNA. The active RNA-protein CRISPR-Cas complex then relies on these guide RNA sequences to target cleavage activity, generating DSBs in any complimentary sequences. It is this high specificity DSB generation which is exploited in genomic manipulation tools, facilitating high specificity

DNA disruption at a specific chromosomal locus. Initial eukaryotic CRISPR-Cas system application relied on DSB repair through error-prone non-homologous end joining (NHEJ) in which insertional or deletional mutations naturally occur. Most prokaryotes do not possess such machinery making these chromosomal DSBs lethal. Instead, prokaryotes rely on a technique referred to as homology directed repair which replaces the targeted sequence preventing persistent chromosomal cleavage (Gasiunas, et al., 2012). As such CRISPR-Cas systems essentially act as a robust counter-selection mechanism for allelic exchange.

The Type II Cas9 system of *Streptococcus pyogenes* has become the most widely implemented CRISPR editing tool and is employed across many taxa. Variations of this system have been employed with good efficiency within a range of acetogens including *C. difficile* (Ingle, et al., 2019), *C. autoethanogenum* (Huang, et al., 2016), *C. acetobutylicum* and *C. beijericnkii* (Li, et al., 2016). Since its discovery this CRISPR-Cas9 system has undergone extensive development to diversify its functionality. This includes the inactivation of the nuclease domain to generate a single strand cleaving nickase-Cas9 (nCas9), a catalytically dead Cas9 (dCas9) applied in gene expression interference assays (CRISPRi) and a dCas9 fused with associated transcriptional effectors to modulate adjacent gene expression (CRISPRa).

The dCas9 utilised in CRISPRi contains two-point mutations (D10A and H840A) to prevent nucleolytic activity. While catalytically inert this system can still target and bind highly specific sequences, and as such is used to downregulate gene expression. The primary targets for downregulation are regions upstream of protein coding sequences, where the binding acts as steric hinderance preventing the interaction of RNA polymerase or transcription factors with the DNA. Dynamic range within this system can be derived through inducible promoter driven control of the sgRNA or dCas9 expression, or through the use of multiple proximal sgRNAs to compound the effect. The ability to knockdown instead of completely negating the activity of genes is also advantageous in metabolic engineering, carbon flux mediation and the perturbation of expression of integral genes which are not feasible knockout targets. Such

approaches have been implemented in C. *ljungdahlii* (Woolston, et al., 2018), *C. autoethanogenum* (Fackler, et al., 2021) and *E. limosum* (Shin, et al., 2019).

Another increasingly prevalent CRISPR-Cas genome modification system applied within acetogens is Cpf1 (**C**RISPR from ***P****revotella* and ***F****rancisella***1**). This employs a Cas12a derived from *Acidaminococcus* sp.BV3L6, a comparatively smaller endonuclease. Cpf1 also requires a shorter CRISPR RNA (crRNA), can be readily used for multiplex editing and has a less GC rich PAM (TTTN) (Liu, et al., 2017). The CRISPR/Cpf1 system has been used for genome editing in several clostridia including *C. difficile* (Hong, et al., 2018) and *C. beijerinckii* (Zhang, et al., 2018). The functional toolkit has also expanded into similar avenues as observed in *cas9* with Cas12a derived CRISPRi applied for flux regulation in *C. ljungdahlii* (Zhao, et al., 2019)

Alongside the implementation of heterologous CRISPR-Cas systems, several examples have arisen of the exploitation of endogenous system for genome self-targeting and manipulation. This is naturally not a universally applicable approach within acetogens, however there is still good scope with an estimated 74% of clostridia harbouring endogenous CRISPR-Cas systems (Pyne, et al., 2016). While each system requires a degree of characterisation prior to exploitation, endogenous system exploitation can be preferable to exogenous system introduction as the required vectors are markedly smaller, containing only the synthetic CRISPR array and regions of homology. As such these systems can be used in situations where transfer rates are prohibitive to larger vector introduction. They also benefit from a decreased likelihood of off-target toxic effects and minimise any associated metabolic burden from high copy number plasmid-based gene expression.

1.9. Forward genetics studies

Forward genetics approaches have been less broadly applied in acetogens than the previously described reverse techniques which rely on genomic

characterisation of a strain before rational genetic manipulation targeting. As the antithesis of this approach, forward genetics techniques rely on the phenotypic characterisation of random genetic mutants, prior to the elucidation of their genetic basis. The first step within such forward genetics approaches is strain mutagenesis. There are several mutagenesis techniques which can be applied, broadly categorised into chemical agents, biological agents or physical agents, such as radiation. Ideal mutagenesis yields a large, unbiased mutant pool with single gene changes in each mutant generating consequently diverse phenotypes. Subsequent phenotypic screening is performed with tailored assays dependent on the characteristic under investigation, or desirable phenotype. Sequencing to determine the mutant genotype responsible for the observed change then follows.

1.9.1. Transposon mutagenesis

An exemplary biological mutagenesis technique is transposon mutagenesis. Transposons are mobile genetic elements found across almost every taxon. Unlike the other genomic modification tools described herein, they are novel in their ability to integrate at non-homologous genomic sites. This integration is performed through the activity of transposase enzymes which recognise and relocate a specific transposon sequence within the chromosome. There are three primary mechanisms of transposition broadly referred to as cut-and-paste, copy-and-paste or copy-out-paste-in. In many instances the transposon sequence itself contains the necessary gene to encode the associated transposase. Two specific recognition sequences called inverted terminal repeats (or alternatively inverted repeats/terminal inverted repeats) typically frame this transposon sequence. The mutant pool generated by transposon mutagenesis is hence made up of insertional mutants, where the transposed sequence disrupts the native loci. Initial approaches to screening and subsequent genomic characterisation of these insertional mutants, included genetic foot printing and tailored DNA hybridisation micro array assays. While

attempts were made to improve the throughput of such techniques, they remained laborious and yielded relatively low precision/resolution.

1.9.2. Transposon insertion sequencing (TIS)

The first application of sequencing to identify transposon insertion sites relied on extensive cloning steps to integrate transposon adjacent sequences into pUC18 vectors (Hutchison, et al., 1999). While this approach exploited Sanger sequencing, capable of processing a single fragment at one time, next generation sequencing technologies such as Illumina processes millions of DNA strands in parallel. As such, simultaneous high throughput sequencing of the entire mutant pool became attainable. Four distinct TIS techniques were published built on this principle, generating amplicon sequencing libraries enriched for transposon-chromosomal DNA junctions from mixed biological samples. These approaches were coined as Insertion Sequencing (INSeq) (Goodman, et al., 2009), transposon sequencing (Tn-seq) (van Opijnen, et al., 2009), High-Throughput Insertion Tracking by Deep Sequencing (HITS) (Gawronski, et al., 2009) and Transposon-Directed Insertion Site Sequencing (TraDIS) (Langridge, et al., 2009).

The scope of this sequencing data greatly exceeds previous assays, allowing not only insertion site identification but also quantitative assessment of the composition of the mutant pool. With single insertional mutant pools each transposon derived sequence is indicative of the presence of this mutant strain. Through selective growth the composition of this mutant pool can change, with mutants whose gene disruption was deleterious not able to proliferate in the library creation process. As such these mutants would be absent within the final dataset, with all remaining transposon associated insertion sequences indicative of mutants which remained viable. With sufficient coverage such data can be used to identify the essential gene set required for growth under a given set of conditions. The implementation of this gene essentiality analysis, or the more tailored perturbation of key characteristics, can significantly

further our genomic understanding of the strain. For example, a TIS study of sporulation within *Bacillus subtilis* identified twenty-four previously disregarded genes essential within the process (Meeske et al., 2016). Gene function allocation is particularly of note as despite the development of high throughput NGS and annotation pipelines a significant number of genes across all genomes cannot be correlated with a specific function by genomic analysis alone. Even within the synthetic minimal genome of Syn 3.0 there are 149 genes (32%) whose functions remain unknown (Hutchinson, et al., 2016).

TIS within *Clostridium* has only been applied to *C. difficile* and *C. autoethanogenum* to date (Dembek, et al., 2015; Woods, et al., 2021). Within *C. difficile* it was used to generate and sequence 70,000 unique mutants, resulting in a 404 essential gene profile for *in vitro* growth. Application of this approach to interrogate sporulation led to the identification of 798 putatively associated genes (Dembek, et al., 2015). Within *C. autoethanogenum* an optimised low transfer and conjugal donor compatible transposon mutagenesis system was exemplified. This system was used to generate mutant profiles in rich media and under CO autotrophy. These conditions yielded 100,065 and 66,524 unique insertion sites respectively, and led to a proposed 439 essential genes for heterotrophic growth on rich YTF medium, and 758 genes for autotrophic growth on CO in minimal media (Woods, et al., 2021)

1.10. Project Aims

Within this project the main goal was to progress research within potentially industrially relevant C4/C6 producing acetogenic bacteria through strain characterisation and the development of amenable synthetic biology tools. The two selected acetogenic organisms posed different challenges; *C. carboxidivorans* as a recalcitrant organism in which heterologous DNA transfer and expression had not been demonstrated at commencement of this project,

and *E. limosum* as a novel uncharacterised strain. As such the key aims of this project were to:

- Establish a robust protocol for DNA transfer into *C. carboxidivorans*
- Develop a genetic toolset for use within *C. carboxidivorans*
- Demonstrate genomic manipulation in *C. carboxidivorans*
- Characterise the synthetic biology prospects and genetic tractability of *E. limosum* NG-6849
- Further understanding of the metabolic processes involved in the growth of *E. limosum* NG-6849 on methanol

2 Materials and Methods

2.1. Chemicals and suppliers

All enzymes and associated reaction mixes and buffers used in this study, including restriction endonucleases, T4 DNA ligase, HiFi 2xMM and Antarctic Phosphatase, were sourced from New England Biolabs, USA (NEB). Three DNA polymerases were used interchangeably within this study: DreamTaq (Thermo Fisher Scientific, USA), Q5® High-Fidelity (NEB), and OneTaq® (NEB). Oligonucleotides were synthesised by Sigma-Aldrich, USA, and Integrated DNA Technologies (IDT), USA. Reference ladders for DNA size analysis in gel electrophoresis included O'Generuler 1 kb/1 kb plus from Thermo Fisher Scientific and 2-Log DNA Ladder from NEB. SYBR Safe was also sourced from Thermo Fisher Scientific. Monarch® DNA Gel/Plasmid extraction kits routinely employed were from NEB, while GenElute™ Bacterial Genomic DNA kits were sourced from Sigma-Aldrich. All chemicals were ordered from Sigma-Aldrich unless otherwise indicated.

2.2. Bacterial strains and routine culture

2.1.1. Bacterial strains

Table 1 – Detailed list of the bacterial strains employed throughout this study

Strain	Function and properties	Genotype	Source
E. coli TOP10	Plasmid storage and cloning strain	F- *mcrA, Δ(mrr-hsd*RMS-*mcr*BC*), φ80lacZΔM15 ΔlacX74, nupG, recA1 araD139, Δ(ara-leu)7697 galE15, galK16, rpsL(*Str^R*) endA1, λ-*	Invitrogen
E. coli TopSex	*E. coli* Top10 harbouring pR702	F- *mcrA, Δ(mrr-hsd*RMS-*mcr*BC*), φ80lacZΔM15 ΔlacX74, nupG, recA1 araD139, Δ(ara-leu)7697 galE15, galK16, rpsL(*Str^R*) endA1, λ-*, R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob	This study
NEB® *E. coli* 10-beta	Plasmid storage and cloning strain	F- *Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	NEB

E. coli Stellar	Methylation devoid cloning strain	F-, *ara,Δ(lac-pro*AB*)* [*Φ80d lacZΔ*M15], *rpsL(str), thi, Δ(mrr-hsd*RMS-*mcr*BC), *ΔmcrA, dam, dcm*	CloneTechTM
E. coli InterStellar	*E. coli* Stellar harbouring cryptic plasmid pR702	F-, *ara,Δ(lac-proAB) [Φ80d lacZΔ*M15], *rpsL(str), thi, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, dam, dcm,* R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob	(Mansfield, 2018)
E. coli ReInterStellar	*E. coli* InterStellar harbouring pKD46-cm	F-, *ara,Δ(lac-proAB) [Φ80d lacZΔM15], rpsL(str), thi, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, dam, dcm* R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob, *exo, gam, beta*	(Mansfield, 2018)
E. coli Re2InterStellar	*E. coli* InterStellar harbouring pKD46-cm and recombinant pR702_TII	F-, *ara,Δ(lac-proAB) [Φ80d lacZΔM15], rpsL(str), thi, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, dam, dcm* R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob, ex*o, gam, beta*	This study
E. coli NEBExpress®	Basal strain for receipt of recombinant pR702	*fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^S*)2 [dcm] R(zgb-210::Tn10--*Tet^S*) endA1 Δ(mcrC-mrr)114::IS10*	NEB
E. coli sExpress	*E. coli* Express harbouring cryptic plasmid pR702	*fhuA2 [lon], ompT, gal, sulA11, R(mcr73::miniTn10--*Tet^S*)2, dcm R(zgb210::Tn10--*Tet^S*), endA1, Δ(mcrCmrr)114::IS10* R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob	(Woods, et al., 2019)
E. coli sExpress_TII	*E. coli* Express containing recombinant pR702_TII	*fhuA2 [lon], ompT, gal, sulA11, R(mcr73::miniTn10--*Tet^S*)2, dcm R(zgb210::Tn10--*Tet^S*), endA1, Δ(mcrCmrr)114::IS10* R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob	This study
E. coli sExpress_TIII	*E. coli Express* containing recombinant pR702_TIII	*fhuA2 [lon], ompT, gal, sulA11, R(mcr73::miniTn10--*Tet^S*)2, dcm R(zgb210::Tn10--*Tet^S*), endA1, Δ(mcrCmrr)114::IS10* R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob	This study
E. coli CA434	*E. coli* HB101 harbouring pR702	F–, *thi-1, hsdS20 (rB–, mB–), supE44, recA13, ara-14, leuB6, proA2, lacY1, galK2, rpsL20 (strr), xyl-5, mtl-1.*	M. Young, UCW, Aberystwyth, UK (Purdy, et al., 2002)
E. limosum NG-6849	Research strain	Wild-type	P. Soucaille, INSA, University of Toulouse, France
C. autoethanogenum DSM10061	Research strain	Wild-type	Culture Collection, SBRC, Nottingham
C. sporogenes NCIMB 1069	Research strain	Wild-type	Culture Collection, SBRC, Nottingham
C. carboxidivorans P7	Research strain	Wild-type	DSMZ (Liou, et al., 2005)

2.1.2. Routine growth conditions

E. coli strains were grown in lysogeny broth (LB) media (Sigma-Aldrich) at 37°C in a shaking incubator (225 RPM) when in broth, or static incubator when on solid media. Some exceptions to this temperature are noted where thermally unstable replicons necessitated a 30°C incubation. All Gram-positive strain routine culture (i.e., clostridia and eubacteria) was performed at 37°C in an MG1000 Mark II Anaerobic Workstation (Don Whitley Scientific, UK). Solid media was produced by the addition of 15 g/L agar to the relevant broth media prior to autoclaving. For anaerobic growth broth media was placed into cabinets at least 48 hours prior to use and solid plates at least 16. Where appropriate any media supplementation was performed prior to inoculation.

C. sporogenes NCIMB 1069 was grown using TYG media (tryptone 30 g/L; yeast extract 20 g/L; sodium thioglycollate 1 g/L). *C. autoethanogenum* DSM10061 was cultured with YTF media (tryptone 16 g/L; yeast extract 10 g/L; fructose 10 g/L; NaCl 0.2 g/L; pH 5.8; 1 mL 1000x trace acid solution; 1 mL 1000x trace base solution; 1 mL 1000x vitamin solution; **Table 2**). *C. carboxidivorans* P7 was initially cultured using YTF media, followed by the better buffered YTAFM media (yeast extract 9 g/L; tryptone 14 g/L; L-arginine 1.4 g/L; fructose 10 g/L; MES 10 g/L; pH5.8). *E. limosum* NG6489 was grown on Pacaud media; 5% v/v macromineral solution (KH_2PO_4 6 g/L; $NaCl_2$ 12 g/L; $MgSO_4.7H_2O$ 2.4 g/L; $CaCl_2.2H_2O$ 1.6 g/L), 1% v/v trace element solution ($FeSO_4.7H_2O$ 0.1 g/L; $MnCl_2.4H_2O$ 0.1 g/L; $CaCl_2.H_2O$ 0.17 g/L; $CoCl_2.2H_2O$ 0.1 g/L; $ZnCl_2$ 0.1 g/L; $CuCl_2$ 0.02 g/L; H_3BO_3 0.01 g/L; Na_2MoO_4 0.01 g/L; NaCl 1 g/L; Na_2SeO_3 0.017 g/L; $NiSO_4.6H_2O$ 0.026 g/L; nitrilotriacetic acid 12.8 g/L; pH 6.5), 1% v/v vitamin solution (biotin 2 mg/L; pantothenic acid 5 mg/L; lipoic acid 5 mg/L), NH_4Cl 1 g/L, cysteine-HCl 0.502 g/L and resazurin 0.2 mg/L as a redox potential indicator. For solid media applications 15 g/L of agar was added, as well as 2 g/L yeast extract and 20 mM glucose. Once the media had been left an appropriate amount of time in an anaerobic environment other anaerobic filter-sterilised, de-gassed stock solutions could be added. These included yeast

extract to a final concentration of 2 g/L, glucose to 20 mM and methanol to 62 mM.

Table 2 - YTF medium solution compositions

YTF 1000x vitamin solution	
p-aminobenzoate	114 mg/L
riboflavin	104 mg/L
thiamine	200 mg/L
nicotinate	206 mg/L
pyridoxin	510 mg/L
pantothenate (calcium)	208 mg/L
cyanocobalamin	78 mg/L
d-biotin	22 mg/L
folate	48 mg/L
lipoate/thioctic acid	50 mg/L

YTF 1000x trace acid solution	
HCl	50 mM
H_3BO_3	100 mg/L
$MnCl_2.4H_2O$	230 mg/L
$FeCl_2.4H_2O$	780 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

YTF 1000x trace base solution	
NaOH	10 mM
Na_2SeO_3	58 mg/L
Na_2WO_4	53 mg/L
$Na_{2\ Mb}O_4.2H_2O$	52 mg/L

2.1.3. Supplements

All media supplements were prepared first as stock solutions according to **Table 3** and sterilised by filtration through a 0.2 µM membrane syringe filter (Sartorius Minisart™, Fischer Scientific). Stock solutions were stored according to the manufacturers' recommended temperature.

Table 3 – Media supplements and their respective concentrations for stock and use in Gram-negative and Gram-positive strains.

Supplement	Diluent	Stock	Working concentration	
			E. coli	Gram-positive
Anhydrotetracycline	100% ethanol	2.5 mg/mL	-	0.1 µg/mL
Apramycin	dH_2O	50 mg/mL	50 µg/mL	-
Chloramphenicol	100% ethanol	25 mg/mL	25 µg/mL	-
Clarithromycin	100% DMSO	6 mg/mL	-	6 µg/mL
D-cycloserine	dH_2O	50 mg/mL	250 µg/mL	-
Erythromycin	100% ethanol	50 mg/mL	500 µg/mL	-

Kanamycin	dH_2O	50 mg/mL	50 µg/mL	-
L-arabinose	dH_2O	1M	-	10 mM
Spectinomycin	dH_2O	250 mg/mL	250 µg/mL	-
Theophylline	0.3M NaOH	500 mM	-	0.1-10 mM
Thiamphenicol	100% DMSO	7.5 mg/mL		7.5 µg/mL
Trimethoprim	100% DMSO	10 mg/mL	10 µg/mL	-

2.1.4. Strain storage and revival

E. coli strains were stored using Microbank™ vials (Pro-Lab Diagnostics) at -80°C with exponential phase culture according to manufacturer instruction. Anaerobic strains were stored using 900 µL of exponential phase culture and 100 µL of DMSO to produce a 1 mL 10% (v/v) DMSO stock stored at -80°C in a Cryolock® 2 mL tubes.

2.3. DNA transfer

2.3.1. Preparation of chemically competent *E. coli* cells

Preparation of chemically competent *E. coli* was performed based on the widely used Hanahan method (Chang, et al., 2017). To begin 5 mL of LB broth, supplemented with any necessary antibiotics, was inoculated with the desired *E. coli* strain. After 18 hours of routine growth this culture was used to inoculate similarly supplemented 250 mL LB 1:100 within a sterile 500 mL conical flask, sealed with a foam bung and foil. Placed under appropriate growth conditions once more the optical density at 600 nm (OD_{600}) of this culture was then monitored until it reached 0.4. The 250 mL culture was then decanted into 50 mL falcons and incubated on ice for 15 minutes.

All following steps were performed at 4°C or on ice, including use of pre-chilled solutions. The 50 mL falcons were centrifuged at 3000 x *g* for 10 minutes and subsequently washed by resuspension of the cell pellet in 10 mL of 100 mM $CaCl_2$, 20% v/v glycerol solution. This suspension was then incubated on ice for 15 minutes, followed by centrifugation as above. The final resuspension collated the falcons into one volume of 2.5 mL of 100 mM $CaCl_2$, 20% v/v

glycerol solution. At this stage samples were incubated on ice for up to 15 minutes before decanting into 100 µL aliquots in sterile 1.5 mL Eppendorf tubes (Eppendorf UK Ltd, UK) and stored at -80°C until use.

2.3.2. Transformation of chemically competent *E. coli*

The protocol for DNA transfer into chemically competent *E. coli* cells was adapted from the NEB high efficiency (C2987H/C2987I) heat shock protocol (NEB, n.d.). Aliquots were initially removed from storage and thawed on ice. Upon addition of 50 ng of plasmid DNA, 10 µL HiFi fragment assembly mixture, total volume of purified PCR product, or 10 µL of ligation mix, the aliquot remained on ice for a further 30 minutes. These were then shocked by incubation for 35 seconds at 42^{O}C and immediately returned to the ice for 5 minutes. Recovery was performed by adding room temperature LB to reach a total volume of 1 mL and incubating at 37^{O}C with constant agitation of 225 RPM for 1 hour. Following this recovery period appropriate dilutions were plated to selective LB plates, routinely 30 µL and 970 µL fractions (pelleted first by centrifugation). These plates were then incubated at 37°C for 24 hours before colonies were observed.

2.3.3. Preparation of electrocompetent *E. coli*

For electrocompetent ReInterStellar/Re2InterStellar *E. coli* cell preparation starter cultures were grown in 5 mL LB broth supplemented with 50 µg/mL kanamycin and 250 µg/mL spectinomycin for 18 hours overnight at 30°C with 225 RPM shaking agitation. This culture was subsequently used to inoculate 250 mL of LB in a 500 mL conical flask 1:100, with prior selection maintained. This culture was monitored as it grew at 30°C and 100 mM L-arabinose introduced at 0.1 OD_{600}. Once 0.4 OD_{600} was achieved the culture was decanted

into six pre-chilled 50 mL falcon tubes, and all remaining steps of the preparation were performed on ice or at 4°C, including use of prechilled reagents/buffers. Cells were pelleted by centrifugation at 4000 x *g* for 10 minutes, and the resulting supernatant was discarded. The cell pellets were then washed twice by resuspension in 25 mL of distilled water (dH_2O) and centrifugation repeated. Following the final centrifugation, the cell pellet was resuspended in 1.5 mL of 10% glycerol (v/v in dH_2O), before separation into 50 µL aliquots in sterile 1.5 mL Eppendorf tubes for storage at -80°C.

2.3.4. Transformation of electrocompetent *E. coli*

DNA transfer into electrocompetent *E. coli* cells followed the same thawing and DNA incubation stages outlined in **2.3.2**, with transfer of the cell mix to pre-chilled 2 mm gap electroporation cuvette (BioRad) using a wide-bore pipette during the initial 30-minute incubation. The cuvette was then pulsed using pre-set EC2 conditions of the BioRad Micropulser™ electroporation apparatus (2.5 kV, 1 pulse, 0.2 cm cuvette). Recovery was performed by addition of room temperature LB broth up to a total volume of 1 mL and cell mix transfer to a sterile 1.5 mL Eppendorf tubes prior to 225 RPM agitation and incubation at 30°C. All recovered *E. coli* aliquots were plated onto selective LB agar at appropriate concentrations, routinely 30 µL and 970 µL fractions (the 970 µL pelleted following 30 µL by centrifugation at 1500 x *g*). These plates were then incubated at 30°C for 24 hours prior to observation of growth.

2.3.5. Preparation of electrocompetent *E. limosum* cells

E. limosum electrocompetent cell preparation was performed as previously demonstrated with strain ATCC 8486 (Mansfield, 2018). For *E. limosum* NG-6849 electrocompetent cell preparation all centrifugation and handling steps were performed in an anaerobic tent (Coy Laboratories). An initial starter culture was inoculated from storage into 5 mL of Pacaud media with 20 mM

glucose and 2 g/L yeast extract. Following 16 hours (overnight) growth this initial culture was then used to inoculate 300 mL of similarly supplemented Pacaud media 1:100 and the OD_{600} was monitored. Once the OD_{600nm} of this culture reached 0.4 the 300 mL culture was decanted into prechilled anaerobic 50 mL falcons and transferred on to ice within the anaerobic tent. Following a 15-minute incubation period on ice each falcon was then spun in a 4^{o}C pre-chilled centrifuge at 7000 x *g* for 10 minutes and the resulting supernatant discarded. Each cell pellet was then resuspended in 5 mL chilled Sodium Mono Phosphate (SMP) buffer (270 mM sucrose; 7 mM sodium phosphate; 1 mM $MgCl_2$; pH 7.2) and collated into two falcons, each with a resultant volume of 15 mL. These centrifugation and supernatant removal steps were repeated once more with 10 mL of SMP, before final pellet resuspension in 2 mL of SMP 10% DMSO (v/v). The resuspended cell volumes were aliquoted into 100 µL volumes in chilled 2 mL Cryolock® tubes for immediate use or storage at -80^{o}C.

2.3.6. Transformation of electrocompetent *E. limosum*

All *E. limosum* transformations were performed within the MG1000 Mark II Anaerobic Workstation (Don Whitley Scientific, UK) using Pacaud media with 20 mM glucose and 2 g/L yeast extract. The protocol was adapted from previous work with strain ATCC 8486 (Mansfield, 2018), with competent cell aliquots thawed anaerobically on ice, then incubated with 500 ng of plasmid DNA in 2 mM gap Gene Pulser® electroporation cuvettes (BioRad). Following a 15-minute incubation period a pulse was delivered using a GenePulser Xcell™ electroporator, with both CE and PC modules (BioRad, USA). A time constant pulse of 1700 volts (V) and 6 milliseconds (mS) was applied. These settings typically result in a resistance of 200 Ω and a capacitance of 50 µF.

Following the pulse, media was added to each cuvette to reach a total volume of 2 mL. Samples were then incubated at 37^{o}C for 4 hours. Following this recovery period, the culture was mixed by inversion before volumes of 50 µL and the total 2 mL were spread across two Pacaud media plates supplemented

with appropriate antibiotic selection. To pellet the 2 mL volume cells were transferred to a 2 mL Cryolock® tubes and pelleted at 3000 x *g* for 5 minutes using an aerobic bench top centrifuge before returning to the cabinet. Colonies were observed following 72 hours of appropriately selective plate growth.

2.3.7. Conjugative DNA transfer in *C. sporogenes*

Conjugative DNA transfer into *C. sporogenes* was adapted from Woods et al., (2019), once more with vectors introduced into chemically competent *E. coli* sExpress. This conjugal donor was then cultured in triplicate aerobically in selective LB broth at 37°C for 16 hours (overnight). *C. sporogenes* was also inoculated into 10 mL TYG broth and grown at 37°C overnight in a MG1000 Mark II Anaerobic Workstation (Don Whitley Scientific, UK). A 1 mL aliquot of each donor cell culture was then normalised to the lowest OD_{600} and transferred to a sterile 1.5 mL Eppendorf tube using a wide bore pipette. These were then pelleted by centrifugation at 3000 x *g* for 3 minutes. These cell pellets were then gently resuspended in 1 mL of Phosphate Buffered Saline (PBS) before being pelleted as before with the supernatant removed by careful aspiration.

The cell pellets were then transferred to the anaerobic workstation where they were resuspended in 200 µL of exponential phase *C. sporogenes* culture. The donor and recipient were then incubated together on TYG plates without selection for 24 hours. A bacterial slurry was collected from these plates through flooding with 1 mL of PBS and using a wedge-shaped spreader to gently dislodge and re-suspend the bacterial growth. This slurry was aspirated from the plates and transferred, at appropriate dilution in PBS, to selective TYG agar plates supplemented with 7.5 µg/mL thiamphenicol, 10 µg/mL trimethoprim and 250 µg/mL D-cycloserine. Transconjugant colonies were observed following 24-48 hours of incubation.

2.3.8. Conjugative DNA transfer in *C. autoethanogenum*

Conjugative transfer into *C. autoethanogenum* was performed as detailed in Woods et al., (2019) following transfer of necessary vectors into *E. coli* SExpress. These conjugal donors were then cultured aerobically in selective LB broth at 37°C for 16 hours (overnight). *C. autoethanogenum* was also initially cultivated for 48 hours before being passaged into 10 mL YTF broth and grown at 37°C overnight in a MG1000 Mark II Anaerobic Workstation (Don Whitley Scientific, UK). Donor cultures were subcultured into similarly selective LB broth at 1:100 and grown until an OD_{600} of 0.3 was reached (typically 3-4 hours). 1 mL of each donor cell culture was then normalised to the lowest OD_{600} and transferred to a sterile 1.5 mL Eppendorf tube using a wide bore pipette. These were then pelleted by centrifugation at 3000 x *g* for 3 minutes and the supernatant discarded prior to resuspension in 1 mL of PBS. Centrifugation was repeated as before, and the supernatant removed by careful aspiration.

The cell pellets were then transferred to the anaerobic workstation where they were resuspended in 200 µL of exponential phase *C. autoethanogenum* culture. The donor and recipient were then incubated together on non-selective YTF plates for 24 hours. A bacterial slurry was collected from these plates through flooding with 1 mL of PBS and use of a wedge-shaped spreader to gently dislodge and re-suspend the bacterial growth. This slurry was aspirated from the plates and transferred, at appropriate dilution in PBS for single colony observation, to YTF agar plates supplemented with 7.5 µg/mL thiamphenicol and 250 µg/mL D-cycloserine. Transconjugant colonies were observed following 72 hours of incubation.

2.3.9. Conjugative DNA transfer in *C. carboxidivorans*

To establish DNA transfer *C. carboxidivorans* was inoculated from exponential phase culture into 5-10 mL YTFAM broth. Two further YTAFM broth cultures were also established by passage from this initial culture at dilutions of 10^{-1} and

10^{-2} respectively. All three cultures were grown for 16-20 hours (overnight) anaerobically at 37°C. Prior to conjugation chemically competent *E. coli* sExpress, sExpress_TII and/or sExpress_TIII were transformed with the necessary vectors. These conjugal donors were then cultured aerobically in LB broth at 37°C for 16 hours (overnight) with 50 µg/mL kanamycin selection. Further antibiotic supplementation for plasmid selection was vector dependent, with 15 µg/mL chloramphenicol for CatP module vectors, or 500 µg/mL erythromycin for ErmB module vectors. When recombinant donors, sExpress_TII or sExpress_TIII, were used media was further supplemented with 10 mM L-arabinose. Following this initial growth period each donor culture was subcultured 1:50 into identically supplemented LB broth and grown at 37°C and 225 RPM until an OD_{600} of 0.8-1.1 was reached (3-4 hours). Appropriate culture volume of the *E. coli* donor strain, equivalent to 1 mL of 1 OD_{600}, were then transferred to sterile 1.5 mL Eppendorf tubes using a wide bore pipette. Samples were pelleted by centrifugation at 3000 x g for 3 minutes, and the supernatant discarded. The cell pellet was then gently resuspended in 1 mL sterile PBS before centrifugation and supernatant removal was repeated.

The *E. coli* donor cell pellets were then transferred to the anaerobic cabinet and resuspended in an exponential phase *C. carboxidivorans* culture (0.6-1.1). Volume of culture for resuspension was equivalent to 200 µL of 1 OD_{600}. This cell mix was then spread onto non-selective anaerobic YTFAM mating plates. Recombinant donor mating plates were also supplemented with 10 mM L-arabinose. Following 24 hours growth a bacterial slurry was collected from these plates by flooding with 1 mL of PBS and gentle manual disruption with a sterile wedge-shaped spreader. This bacterial slurry was either applied directly to appropriately selective plates or serially diluted in PBS before plating to attain single colonies. These selective YTAFM plates were all supplemented with 250 µg/mL D-cycloserine to counter-select the *E. coli* donor, as well as the necessary antibiotic to select for the target plasmid (7.5 µg/mL thiamphenicol or 6 µg/mL clarithromycin). Colonies were typically observed after 48 hours.

2.4. General molecular techniques

2.4.1. Plasmid DNA extraction

E. coli cultures of 2-4 mL, grown overnight at 37°C and at 225 RPM in LB with appropriate antibiotic selection for the plasmid, were harvested by 1 minute centrifugation at 17000 x *g*. The Monarch® Plasmid Miniprep Kit (NEB) was then used according to the manufacturer's instructions for plasmid isolation.

2.4.2. Genomic DNA extraction

A 2-10 mL bacterial culture was grown overnight at 37°C and the cells harvested by centrifugation at 16000 x *g*. Genomic DNA extraction and purification was then performed using the GenElute™ Bacterial Genomic DNA Kit (Sigma Aldrich) according to the supplied instructions.

2.4.3. Restriction enzyme digestion

Restriction enzymes and buffers were obtained from New England Biolabs (NEB), and reactions were set up according to the recommended 20 µL reaction protocol: 500 ng DNA, 2 µL CutSmart® Buffer, 5 units of restriction enzyme(s) and nuclease-free water to 20 µL volume. These reaction mixes were then incubated at 37°C for 2 hours prior to downstream manipulation.

2.4.4. De-phosphorylation of linear DNA

De-phosphorylation of the 5' DNA to prevent vector re-circularisation was performed using Antarctic Phosphatase (NEB). Reactions were assembled according to manufacturers' 20 µL volume guidelines with the associated Antarctic Phosphatase Buffer (NEB). This was then incubated in a thermocycler at 37°C for 30 minutes before heat inactivation at 80°C for 2 minutes.

2.4.5. DNA ligation

T4 DNA Ligase and T4 DNA Ligase buffer (NEB) were used to ligate compatible DNA fragments together. Typically, a molar ratio of 3:1 insert:backbone was used, where backbones were linearised plasmids. The ligation mixture contained 2 µL T4 DNA Ligase Buffer (10X), 1 µL of T4 DNA Ligase, 50 ng of backbone DNA and 37.5 ng of insert DNA, made to a total volume of 20 µL with nuclease-free water. These reactions were incubated overnight (18 hours) in thawing ice to provide a temperature gradient.

2.4.6. DNA extraction from agarose gels

SYBR Safe stained agarose gels were analysed on a LiCor Odyssey DNA analyser before transfer to a blue light box for manipulation. Target bands were excised using a scalpel blade and transferred to a 1.5 mL Eppendorf tubes. These fragments were then dissolved and purified using the Monarch® DNA Gel Extraction Kit (NEB), as instructed by manufacturer.

2.4.7. PCR clean up

QIAquick PCR Purification Kits (Qiagen, Germany) were used for purification of PCR products where subsequent agarose gel electrophoresis was not performed. A 125 µL aliquot of Buffer PB was added to the 25 µL PCR reaction mix (5:1) and the solution mixed before loading onto a QIaquick spin column in a 2 mL collection tube. This was centrifuged at 17000 x *g* for 1 minute and the flow-through discarded. A 750 µL volume of Buffer PE was then added to the column followed by a further 1 minute of centrifugation. The flow-through was discarded once more, before a final 2-minute centrifugation. The column was then transferred to a new 1.5 mL Eppendorf tube, and 30-50 µL of nuclease

free water added. This was incubated at room temperature for 5 minutes prior to elution by centrifugation at 17000 x *g* for 1 minute.

2.4.8. Polymerase chain reaction (PCR)

PCR was used to amplify DNA fragments from different templates, using a flanking pair of complimentary primers. Melting temperature (Tm) values and annealing temperature for each primer pair were calculated using the NEB Tm Calculator online tool.

2.4.8.1. High fidelity DNA amplification

Amplification of DNA fragments where fidelity was integral, were performed using Q5® High-Fidelity DNA Polymerase (NEB). Reactions were assembled according to the manufacturers recommended 25 µL reaction mix for each template. The thermocycler program is detailed in **Table 4** below and was adjusted for each reaction according to desired fragment length (30 second elongation per kB) and the final extension also scaled to fragment length.

Table 4 - Q5® PCR thermocycler program

Cycle Step	Temperature (°C)	Duration (s)	
Initial Denaturation	98	60	
Denaturation	98	30	x25-35 cycles
Annealing	50-72	30	
Elongation	72	30/kB	
Final extension	72	90/kB	
Hold	10	-	

Touch-down PCR protocols were used where standard PCR protocols failed. This modified the thermocycler program by decreasing the annealing temperature by 1°C per cycle, starting 5°C above the predicted Tm value, until 5°C below the predicted Tm value, before performing a further 25 cycles and final extension at this final temperature.

2.4.8.2. Splicing by overlap extension (SOE) PCR

Where HiFi assembly approaches struggled to yield CRISPR vectors in a one stage assembly, the two homology arms were joined together by SOE-PCR. This approach relied on complimentary overlapping 5' regions in the internal primer pair (LHA Rv, RHA Fw). When both homology arm fragments are amplified within one PCR mix using the two external primers (LHA Fw, RHA Rv) these overlaps facilitate the PCR-based joining of homology arms. To perform this SOE PCR the purified PCR products of initial homology arm amplification from gDNA were used as the template at a ratio of 1:1 with 50 ng DNA and amplified using the Q5 High Fidelity polymerase protocol (**2.4.8.1**).

2.4.8.3. Low fidelity DNA amplification

Colony PCR was used as a screening method circumventing the need for DNA extraction prior to amplification. In lieu of gDNA, the template for colony PCR was cell mass picked directly from colonies on plates and resuspended in 15 µL dH_2O, or a dilution of 1 µL of turbid liquid culture in 15 µL of dH_2O. An initial denaturation step of 6 minutes at peak temperature (i.e., 98/95/94^oC polymerase dependent) was also added to the thermocycler program to denature cells and expose DNA to the PCR reagents. Unless sequence fidelity was essential lower fidelity polymerases were employed. These were DreamTaq DNA Polymerase (Thermo Fisher Scientific) and OneTaq® DNA Polymerase (NEB). Reaction mixes were composed according to the manufacturers recommended 25 µL mix with the 1 µL template derived from the previously described cell suspensions. Within the DreamTaq standard PCR protocol, denaturation was performed at 95^oC, the annealing temperature was reduced by 5^oC and extension time equivalent to 1 min/kB. Within the OneTaq® standard PCR protocol all denaturation was performed at 94^oC, extension duration was equivalent to 1 min/kB and annealing temperature was dictated by the online NEB Tm Calculator tool.

2.4.9. HiFi DNA assembly

NEBuilder® HiFi DNA Assembly Master Mix was used to assemble plasmids from DNA fragments with terminal overlapping regions of homology. HiFi compatible primers included a non-annealing 5' tail with 15-25 base pairs (Tm >50°C) complimentary to the adjoining fragment. Fragments amplified by PCR using such primers were combined at a ratio of 2:1, insert to linearised plasmid, or 1:1:1:1 if using comparably sized fragments. Reaction mixes consisted of 10 µL 2xHiFi Master Mix, appropriate fragments volumes for desired DNA concentrations and sufficient nuclease-free water to reach a final volume of 20 µL. Following incubation for 1 hour at 50°C, chemically competent *E. coli* were transformed with the reaction mixes.

2.4.10. Golden gate assembly

A BsaI (type IIS restriction enzyme) based assembly approach, often referred to as Golden Gate Assembly (Engler, et al., 2008), was used in the assembly of the pRBV1_T* vectors. This is a 'one pot' method used to assemble multiple DNA fragments using sequential BsaI cleavage and T4 DNA ligase activity. All DNA fragments used, whether PCR amplified, linearised backbone, contained terminally located BsaI recognition sequences. BsaI cleavage occurs 1-5 bp downstream of the site, effectively removing these sequences and generating sticky-ends for subsequent ligation. To perform golden gate assembly purified DNA fragments and EcoRI-HF linearised/de-phosphorylated pRBV1 backbone were combined at a ratio of 2:1. To this was also added 1 µL BsaI (NEB), 2 µL CutSmart® Buffer (NEB), 1 µL T4 DNA Ligase (NEB), 2 µL 10 mM ATP (Sigma-Aldrich) and nuclease-free water to a total volume of 20 µL. These mixes were then transferred to a thermocycler for 5-cycles with 10 minutes at 40°C, 10 minutes at 16°C and 10 minutes at 50°C. Final incubation was then performed

at 80°C for 20 minutes before chemically competent *E. coli* Top10 were transformed with the reaction mixes.

2.4.11. DNA visualisation

2.4.11.1. Agarose gel electrophoresis

For standard DNA fragment size analysis samples underwent agarose gel electrophoresis. The gels were prepared with 1-2% agarose in 1x TAE buffer (40 mM Tris, 1 mM EDTA, 0.1% glacial acetic acid) and 0.001% v/v SYBR Safe DNA gel stain (all Thermo Fisher Scientific) for DNA binding and visualisation. Prior to loading, samples were mixed with 6x Gel Loading Dye (NEB). Gels were run in a Gel Electrophoresis Tank (Cleaver Scientific) at 90-120V for 30-45 minutes. SeaPlaque™ low melt agarose (Lonza Bioscience, USA) was used to purify TraDIS PCR derived DNA prior to Illumina sequencing library preparation. Imaging of agarose gels was achieved using an Odyssey® Fc Imaging System.

2.4.11.2. High sensitivity automated gel electrophoresis

Visualisation of TraDIS library fragment size distribution was performed using an Agilent 2100 Bioanalyzer and High Sensitivity dsDNA Kit (Agilent Technologies, USA). The assay was performed according to manufacturers' instructions with initial gel-dye mix preparation, chip priming, sample loading and analysis using the 2100 Expert Software platform.

2.4.12. DNA quantification

2.4.12.1. Nanodrop™

A SimpliNano™ microvolume spectrophotometer (GE Healthcare Life Sciences) was used for routine quantification of DNA purity and nucleic acid

concentrations. This was used according to manufacturer's instructions, with appropriate blanking measurements performed with 2 µL nuclease-free water (the elution buffer) and subsequent 2 µL sample loading into the port. Nucleic acid concentration was automatically calculated from 260 nm absorbance, and the 260/280 and 230/260 ratios used as a proxy for dsDNA purity measurements.

2.4.12.2. Qubit™

The Qubit™ Fluorometer 3.0 (Thermo Fisher Scientific) and Qubit™ Broad Range dsDNA assay kit was used according to manufacturer's instructions to accurately quantify dsDNA prior to next generation sequencing. To perform the assay all samples and standards were incubated at room temperature for 30 minutes prior to use, and a working solution was produced by 1:200 dilution of assay reagent in dilution buffer. Clear thin-wall 500 µL PCR tubes were then labelled for all samples and two standards. 1 µL of each sample was added to the respective tube and made up to 200 µL total volume using the working solution. Standards were made to the same volume with 10 µL of each initially added. All tubes were then vortexed for 2-3 seconds before incubation for 2 minutes at room temperature. Finally, standards followed by assay tubes were loaded into the Qubit 3.0 Fluorometer to take readings. dsDNA concentrations were calculated according to the recorded fluorescence of each sample in relation to an automatically generated standard curve based on the fluorescence of the standards.

2.5. DNA sequencing

PCR generated DNA fragments and plasmid DNA underwent Sanger sequencing as necessary, with appropriate primers proximal to the target region. Samples and primers were prepared in Eppendorf tubes according to Eurofins Scientific

(Ebersberg, Germany) guidelines. These tubes were then transported at room temperature before undergoing Sanger sequencing.

Illumina sequencing of samples generated within this study was performed by three companies MicrobesNG (University of Birmingham), DBS Genomics (Durham University) and Deep Seq (University of Nottingham). Both Deep Seq and MicrobesNG were used for routine sample analysis using Illumina platforms. Deep Seq samples were supplied as purified DNA extractions, however MicrobesNG samples were supplied as cell mass according to instruction. DBS Genomics was utilised for TraDIS sample analysis using a HiSeq 2500 in Rapid Run mode. Sample preparation and further details of this can be found in Chapter 5.

2.5.1. MicrobesNG sample preparation and sequencing

C. carboxidivorans isolates to be sequenced with Microbes NG were streaked directly from -80^{o}C stocks to YTAFM plates and cultivated for 48 hours. A single colony from each plate was picked and suspended in 100 µL sterile anaerobic PBS buffer. This was subsequently spread onto a fresh YTAFM plate with a goal of attaining 1/3 lawn coverage, alongside streaks to single colonies to confirm sample purity. After 48 hours of growth, cell mass was transferred from the plates to a provided 2 mL cryopreservative sample tubes by manual scraping with sterile 10 µL loops. The same approach was used in *E. limosum* NG-6849 sample generation, except with a PBS volume of 200 µL. The further 100 µL was used to inoculate 50 mL of Pacaud (20 mM glucose) broth. This was cultured for 36 hours before the sample was pelleted by centrifugation. The cell pellet was then resuspended in 500 µL of cryopreservant, removed from the provided bead tube. This cell slurry was then transferred to the tube, inverted, and sealed for transit. All sampled were sent at room temperature and received within 2 days of sample preparation.

MicrobesNG DNA extraction was performed by lysis of 5-50 µL of supplied sample with 120 µL of TE buffer containing 0.1 mg/mL RNase A (ITW Reagents,

Spain). This was incubated for 25 minutes at 37°C before 0.1 mg/mL Proteinase K (VWR Chemicals, USA) and 0.5% v/v SDS (Sigma-Aldrich) were added. This was then incubated at 65°C for 5 minutes before gDNA was extracted through 1:1 SPRI bead binding, and elution in EB buffer (Qiagen, Germany). DNA was then quantified with the Quant-iT dsDNA HS kit (ThermoFisher Scientific) assay in an Eppendorf AF2200 plate reader (Eppendorf).

If DNA quality was sufficient, genomic DNA libraries for Illumina sequencing were then prepared using the Nextera XT Library Prep Kit (Illumina, USA). In use of these kits the manufacturer's protocol was adhered to apart from a 2-fold increase in input DNA and an additional 45 seconds of PCR elongation. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system (Hamilton Bonaduz AG, Switzerland) and pooled libraries quantified using the Kapa Biosystems Library Quantification Kit for Illumina. Libraries were sequenced using either Illumina HiSeq or NovaSeq using a 250 bp paired end protocol. Reads were then adapter trimmed using Trimmomatic 0.30 with a sliding window quality cut off of Q15 (Bolger, et al., 2014).

For long-read genomic DNA libraries sequenced with Oxford Nanopore Technologies (ONT) GridIon, preparation was performed with the Oxford Nanopore SQK-RBK004 kit and/or SQK-LSK109 kit with Native Barcoding EXP-NBD104/114 using 400-500 ng of high molecular weight DNA. Barcoded samples were then pooled into a single library and loaded in a FLO-MIN106 (R.9.4.1.) flow cell.

2.5.2. DeepSeq sample preparation and sequencing

NGS was performed by DeepSeq (University of Nottingham, UK) using the Illumina MISeq platform (Illumina, USA) and GridIon Platform (Oxford Nanopore, UK). Samples were provided as genomic DNA or plasmid DNA isolated as detailed in **2.4.2** and **2.4.1**, but eluted in 10 mM Tris pH 8. Quality assessment was performed by nanodrop (**2.4.12.1**), concentration determined

with the Qubit Fluorometer (**2.4.12.2**), and genomic DNA integrity confirmed by gel electrophoresis (**2.4.11.1**). Genomic DNA libraries were then prepared using TruSeq PCR Free Library Preparation Kits (Illumina, USA) and DNA fragmented by sonication using Carvars S2 sonicator before purification and end-repair with AMPure-XP beads (Beckman Coulter). Y-adaptors were then added and purified before quantification of the resultant DNA libraries by qPCR (Kapa Biosystems). These libraries were sequenced using the Illumina MiSeq 500 bp V2.

For nanopore sequencing of plasmid DNA, libraries were prepared using the Rapid Barcoding Kit (Oxford Nanopore Technologies; SQK-RBK004). Barcoded fragmentation mixes were combined with 400 ng of each plasmid DNA template and reactions were incubated at 30°C for 1 minutes, and then at 80°C for 1 minutes before being briefly cooled on ice. 200 ng of each barcoded sample was then combined in an equimolar ratio to generate a pooled library, and further cleaned up using an equal volume of AMPure XP beads (Beckman Coulter; A63882). This library was then eluted in 10 mM Tris-HCL pH 8.9 with 50 mM NaCl, mixed with 1 µL of Rapid adapter (RAP) and incubated for 10 minutes at room temperature. The completed library was then sequenced on a MinION flow cell (Oxford Nanopore Technologies; FLO-MIN106 R9.4.1) on the GridION X5 mk1.

2.6. Bioinformatics tools

Routine plasmid design, sequence alignments and DNA sequence visualisation was primarily conducted within the browser based Benchling platform (Benchling Ltd). Alongside manual primer designs the Primer3 derived Benchling software was used (Rozen & Skaletsky, 2000). Appropriate *S. pyogenes cas9* compatible protospacer guide RNA sequences were also identified by the built-in tool (Hsu, et al., 2013). The Basic local alignment search tool (BLAST) (Altschul, et al., 1990) was used to compare DNA and protein sequences, as well as conduct searches and alignments to appropriate

databases. Statistical analysis and graphical presentation of results was performed using GraphPad Prism. CLC Genomics Workbench 20 was used to map Illumina data and perform resequencing analysis.

2.7. Plasmid Lists

2.7.1. Chapter 3

Table 5 - List of plasmids used during the establishment of robust DNA transfer into *C. carboxidivorans*

Name	Description	Source
pMTL83151	pCB102, ColE1+traJ, *catP* (Cm^R/Tm^R) containing modular plasmid	SBRC Nottingham Culture Collection
pMTL8315y-v1	Modified pMTL83151 with SNPs and truncations to destroy *C. carboxidivorans* RM motifs.	This study
pMTL8315y-v2	Modified pMTL8315y-v1 with the four final RM motifs destroyed.	This study
pMTL8315y-v3	Modified pMTL8315y-v1 with the four final RM motifs destroyed.	This study
pRBV1	Recombineering base vector for golden gate assembly of methyltransferase gene regions.	(Mansfield, 2018)
pRBV1_TI.1	pRBV1 containing the *M.CcaP7IP, S.CcaP7IP* genes	(Mansfield, 2018)
pRBV1_TI.2	pRBV1 containing the *M.CcaP7IIIP, S.CcaP7IIIP* genes	(Mansfield, 2018)
pRBV1_TI.3	pRBV1 containing *M.CcaP7IIP, S.CcaP7IIP* genes	(Mansfield, 2018)
pRBV1_TII.1	pRBV1 containing the *CcaP7ORF16660P* gene	(Mansfield, 2018)
pRBV1_TII.2	pRBV1 containing the CcaP7IVP gene	(Mansfield, 2018)
pRBV1_TII.3	pRBV1 containing the *CcaP7ORF7120P* gene	(Mansfield, 2018)
pRBV1_TIII.1	pRBV1 containing the *M.CcaP7ORF17830P* gene	(Mansfield, 2018)
pRBV1_TIII.2	pRBV1 containing the *M.CcaP7ORF8705P* gene	(Mansfield, 2018)

pRBV1_TI	pRBV1 containing the *M.CcaP7IP, S.CcaPIP, M.CcaP7IIIP, S.CcaP7IIIP, M.CcaP7IIP* and *S.CcaP7IIP* genes	This study
pRBV1_TII	pRBV1 containing the *CcaP7ORF16660P, CcaP7IVP* and *CcaP7ORF7120P* genes	This study
pRBV1_TIII	pRBV1 containing the *M.CcaP7ORF8705P* and *M.CcaP7ORF17830P* genes	This study
pRBV1_TII/III	pRBV1 containing the *CcaP7ORF16660P, CcaP7IVP, CcaP7ORF7120P, M.CcaP7ORF8705P* and *M.CcaP7ORF17830P* genes	This study
pR702	*Proteus mirabilis* derived conjugation assistance cryptic plasmid.	C. Woods (SBRC, Nottingham)
pR702_TI	pR702 with the TI methyltransferase cassette in lieu of *aadA1*	This study
pR702_TII	pR702 with the TII methyltransferase cassette in lieu of *aadA1*	This study
pR702_TIII	pR702 with the TIII methyltransferase cassette in lieu of *aadA1*	This study
pKD46-cm	Cm^R/Tm^R plasmid with λ red genes under P_{BAD} control	(Mansfield, 2018)
pRBV2	pRBV1 with apramycin resistance gene replaced with the *aadA1* gene.	This study
pRBV2_L2-TIII	pRBV2 with the original LHA replaced with the terminal 500 bp of *CcaP7ORF7120P*, with *M.CcaP7ORF8705P* and *M.CcaP7ORF17830P* introduced within the methyltransferase cassette	This study
pMTL8315y-v1-II/IIIABC	pMTL8315y-v1 with nine TII/III RM motifs introduced, three of each type.	This study
pMTL8315y-v1-II/IIIA	pMTL8315y-v1 with three TII/III-A RM motifs introduced	This study
pMTL8315y-v1-II/IIIB	pMTL8315y-v1 with three TII/III-B RM motifs introduced	This study
pMTL8315y-v1-II/IIIC	pMTL8315y-v1 with three TII/III-C RM motifs introduced	This study

2.7.2. Chapter 4

Table 6 - List of plasmids used in the development of tools for the *in vivo* genomic manipulation of *C. carboxidivorans*

Name	Description	Source
pMTL8325y-v1	pMTL8315y-v1 with CatP marker replaced with ErmB conferring Erm^R/Cla^R	This study
pMTL8325y-v1-$P_{fdx(Csp)}$-CatP(trunc)	pMTL8325y-v1 based CatP reporter vector with *C. sporogenes* derived P_{fdx} promoter	This study
pMTL8325y-v1-P_{araE}-CatP(trunc)	pMTL8325y-v1 based CatP reporter vector with *C. acetobutylicum* derived P_{araE} promoter	This study
pMTL8325y-v1-P_{thl}-CatP(trunc)	pMTL8325y-v1 based CatP reporter vector with *C. acetobutylicum* derived P_{thl} promoter	This study
pMTL8325y-v1-$P_{thl(trunc)}$-CatP(trunc)	pMTL8325y-v1 based CatP reporter vector with modified P_{thl} promoter	This study
pMTL8325y-v1-NP-CatP(trunc)	pMTL8325y-v1 based CatP reporter vector with *catP* expression driven by no promoter sequence	This study
pMTL8325y-v1-$P_{fdx(Csp)}$-RbE*-CatP(trunc)	CatP reporter vector with *catP* expression controlled by $P_{fdx(Csp)}$ and adjacent theophylline responsive riboswitch	This study
pMTL8315y-v1-ToxATox-PyrE	Allelic exchange vector with homology-based targeting of *75and*, and CD2517.1 toxin-antitoxin-based counter-selection marker	C. Humphreys (SBRC, Nottingham)
pMTL8015y-v1-ToxATox-TIII.2R	Integrative allelic exchange vector with homology-based targeting of *CcaP7ORF17830P*, no Gram-positive replicon and CD2517.1 toxin-antitoxin-based counter-selection marker	This study
pMTL4315_pMTL8315_CRISPR_adh1_N20#	RiboCas vector designed for *in vivo* manipulation of the *C. autoethanogenum* genome using *cas9*	C. Humphreys (SBRC, Nottingham)
pMTL4315_SynCas9_adh1_N20#2	Modified pMTL4315_pMTL8315_CRISPR_adh1_N20# with alternate synthetic riboswitch and *cas9* sequence	This study

pRECas(pthI)-v1-TIII.2R	RiboCas vector with pMLT8315y-v1 backbone and modified *cas9* cassette, for generation of a *CcaP7ORF17830P* knock-out	This study
pMTL8425y-v1	pMTL8325y-v1 with pCD6 replicon in lieu of pCB102	This study
pMTL8435y-v1-II/IIIA	pMTL8425y-v1 with three additional TII/III-A motifs introduced	This study
pMTL8425y-v1-II/IIIABC	pMTL8425y-v1 with nine additional TII/III RM motifs, three of each type	This study

2.7.3. Chapter 5

Table 7 – Lists of plasmids used in the characterisation of *E. limosum* NG-6849, its genetic tractability and essential gene profile for growth on methanol

Name	Description	Source
pMTL81151	pAMβ1, ColE1+*traJ*, CatP (Cm^R/Tm^R) containing modular plasmid	SBRC Nottingham Culture Collection
pMTL82151	pBP1, ColE1+*traJ*, CatP (Cm^R/Tm^R) containing modular plasmid	SBRC Nottingham Culture Collection
pMTL83151	pCB102, ColE1+*traJ*, CatP (Cm^R/Tm^R) containing modular plasmid	SBRC Nottingham Culture Collection
pMTL84151	pCD6, ColE1+*traJ*, CatP (Cm^R/Tm^R) containing modular plasmid	SBRC Nottingham Culture Collection
pMTL85151	pIM13, ColE1+*traJ*, CatP (Cm^R/Tm^R) containing modular plasmid	SBRC Nottingham Culture Collection
pMTL86151	pIP404, ColE1+*traJ*, CatP (Cm^R/Tm^R) containing modular plasmid	SBRC Nottingham Culture Collection
pMTL88151	p19r, ColE1+*traJ*, CatP (Cm^R/Tm^R) containing modular plasmid	SBRC Nottingham Culture Collection
pMTL8J15x-4077	CatP reporter vector, with p19 replicon, for determination *C. perfringens* derived P_{fdx} strength	(Mansfield 2018)

pMTL8J15x-999	CatP reporter vector, with p19 replicon, for determination *C. sporogenes* derived P_{fdx} strength	(Mansfield 2018)
pMTL8J15x-araE	CatP reporter vector, with p19 replicon, for determination *C. acetobutylicum* derived P_{araE} strength	(Mansfield 2018)
pMTL8J15x-1673	CatP reporter vector, with p19 replicon, for determination *C. acetobutylicum* derived P_{thl} strength	(Mansfield 2018)
pMTL8J15x-NP	CatP reporter vector, with p19 replicon and *catP* expression driven by no associated promoter sequence	(Mansfield 2018)
pMTL83254-RbD	pMTL-IC101 based reporter vector with *catP* expression controlled by $P_{fdx(Csp)}$ and riboswitch D, with the pBP1 Gram-positive replicon replaced with pCB102	This study
pMTL83254-RbE*	pMTL-IC101 based reporter vector with *catP* expression controlled by $P_{fdx(Csp)}$ and riboswitch E*, with the pBP1 Gram-positive replicon replaced with pCB102	This study
pMTL83254-RbF	pMTL-IC101 based reporter vector with *catP* expression controlled by $P_{fdx(Csp)}$ and riboswitch F, with the pBP1 Gram-positive replicon replaced with pCB102	This study
pMTL83254-RbG	pMTL-IC101 based reporter vector with *catP* expression controlled by $P_{fdx(Csp)}$ and riboswitch G, with the pBP1 Gram-positive replicon replaced with pCB102	This study
pMTL83254-RbH	pMTL-IC101 based reporter vector with *catP* expression controlled by $P_{fdx(Csp)}$ and riboswitch H, with the pBP1 Gram-positive replicon replaced with pCB1042	This study
pMTL83254-RbI	pMTL-IC101 based reporter vector with *catP* expression controlled by $P_{fdx(Csp)}$ and riboswitch I, with pBP1 Gram-positive replicon replaced with pCB102	This study
pMTL83254-RbJ	pMTL-IC101 based reporter vector with *catP* expression controlled by $P_{fdx(Csp)}$ and riboswitch J, with pBP1 Gram-positive replicon replaced with pCB102	This study
pRECas(araE)-ORFD	RiboCas based vector designed to knock-out the *R.EliORFD* gene, with P_{araE} driving sgRNA expression	This study
pRECas(thl)-ORFD	RiboCas based vector designed to knock-out the *R.EliORFD* gene, with P_{thl} driving sgRNA expression	This study
pMTL-JM102	Tetracycline inducible, pMTL-Tet3nO based CatP reporter vector, with pMTL83251 backbone	J. Millard (SBRC, Nottingham)
pMTL-CW21	Tetracycline inducible mariner transposon mutagenesis vector, based on the pMTL-CW20 series	J. Millard (SBRC, Nottingham)

2.8. Oligonucleotides and dsDNA sequences

2.8.1. Chapter 3 oligonucleotides

Table 8 – List of primers implemented in the development of a robust DNA transfer protocol for use with *C. carboxidivorans*

Oligonucleotide	Sequence	Template
Tra SNP Fw	TCAGCCGGGCAGGATAGGTAAAGTAGGCCCACCCGCGAG	pMTL83151
MCS Trunc Rv	CCGGGGATCCTCTAGAGTCGACGGCGCGCCAATAAAAAATAC	pMTL83151
pCB102 SNP1 Fw	TCTAGAGTCGACGGCGCGCCAATAAAAAATACTTGTTATGTATTCC GGCGCGCCAAAAAAAAATACTTGTTATGTATTC	pMTL83151
pCB102 SNP1 Rv	TTTATATCTGTTAAGATAAACTCATAAATTGAATATATATAGTTAAT G	pMTL83151
pCB102 SNP Fw	TATATATTCAATTTATGAGTTTATCTTAACAGATATAAATGTAAATT G	pMTL83151
CatP PmeI Rv	TTTGGTCATGAGGTTTAAACCACCGTATTTCTACGATG	pMTL83151
ColE1 PmeI Fw	AGAAATACGGTGGTTTAAACCTCATGACCAAAATCCCTTAACG	pMTL83151
Tra Rv	CTCGCGGGTGGGCCTACTTTACCTATCCTGCCCGGCTG	pMTL83151
CatP_F1	CGTTTTGTAAACGAATTGCAG	pMTL8X1XX vector
CatP_R1	CAAGTTTATCGCTCTAATGAAC	pMTL8X1XX vector
ColE1_tra-R1	CTTTTCTACGGGGTCTGAC	pMTL80000 vector
p15a_F1	GGGTTCGTGTAGACTTTC	pMTL80000 vector
pCB102_R1	CTGTTATGCCTTTTGACTATC	pMTL83XXX vector
p15a_tra_F2	CAAGAAGAGCGACTTCGC	pMTL80000 vector
pMTL8315y-v2 Frag1 Fw	TGGCGAAGATCGAAGAGAAACAGGACGAGCTTGGCAAGGT	pMTL8315y-v1
pMTL8315y-v2 Frag1 Rv	CACCGCTGGTAGCGGTGGTATTTTTGTTTGCAAGCAGCAG	pMTL8315y-v1
pMTL8315y-v2 Frag2 Fw	TGCTGCTTGCAAACAAAAATACCACCGCTACCAGCGGTGG	pMTL8315y-v1
pMTL8315y-v2 Frag 2 Rv	TATCGTGCGAATAAGGATGGATATACCGAATAAATCGCTATAATG ACCCC	pMTL8315y-v1
pMTL8315y-v2 Frag3 Fw	AGCGATTTATTCGGTATATCCATCCTTATTCGCACGATATACAGG	pMTL8315y-v1
pMTL8315y-v2 Frag3 Rv	CCTTGCCAAGCTCGTCCTGTTTCTCTTCGATCTTCGCCAG	pMTL8315y-v1
pMTL8315y-v3 Frag1 Rv	CCGCTGGTAGCGGTGGTTTATTTGTTTGCAAGCAGCAGATT	pMTL8315y-v1
pMTL8315y-v3 Frag2 Fw	TCTAGAGTCGACGGCGCGCCAATAAAAAATACTTGTTATGTATTC	pMTL8315y-v1
27F	AGAGTTTGATCCTGGCTCAG	*C. carboxidivorans* P7 gDNA
1492R	GGTTACCTTGTTACGACTT	*C. carboxidivorans* P7 gDNA
pMTL8315y Scr Fw	AGCGTGAGCTATGAGAAAGC	pMTL831y-v* vectors
pMTL8315y Scr Rv	ACAGTGAAGAAGGAACACCC	pMTL831y-v* vectors
Ccarb Mtase Fw TI.1	TGTACGCGGTCTCCAATGAAATATGAAAGGAAGAAAACAAATGT	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Rv TI.1	TGTACGCGGTCTCCTCTTAATTAAATAGTTCTCCTTTGAAAGC	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Fw TI.2	TGTACGCGGTCTCGAAGAAAAGATTAAAGGAGATAGAAATGG	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Rv TI.2	TGTACGCGGTCTCGACTAAATTTTTACACCTAAAATCTTCAATT	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Fw TI.3	TGTACGCGGTCTCCTAGTGAAGATAAGGAGATTAGCTAAAGA	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Rv TI.3	TGTACGCGGTCTCCATCGCTATATATCTTTACTCAATATTTCCCTCA	*C. carboxidivorans* P7 gDNA

Ccarb Mtase Fw TII.1	TCCAGCTGGTCTCCAGTTGACGGCTAGCTCAGTC	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Rv TII.1	TGATGGGGGTCTCCACTAAACAAATTTAAGCTGCCTTGTTTC	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Fw TII.2	CAGGTCTCCTAGTATAAAATTAAGAAGGTTTAATATGTATAAAGACG	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Rv TII.2	CAGGTCTCCTTACAGCAGTTTATCTCTATACTTATCAC	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Fw TII.3	CAGGTCTCCGTAATAGCATAAAGAAAGGTAATATTTATGGA	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Rv TII.3	TGATGGGGGTCTCCATCGTTATATCTTAAAAAATAAATCCTTTTTAACTTCT	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Fw TIII.1	TGTACGCGGTCTCCAGTTGACGGCTAGCTCAGTC	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Rv TIII.1	TTCCCAGGGTCTCGTTACACATTCTCAATAAACTTTTCAA	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Fw TIII.2	TTCCCAGGGTCTCGGTAATGTAAATGTTGGAGGTGAAATT	*C. carboxidivorans* P7 gDNA
Ccarb Mtase Rv TIII.2	TTCCCAGGGTCTCGATCGCTACCATGACTTTAGTGCCT	*C. carboxidivorans* P7 gDNA
Recomb LHA Fw	TTCTACGGCACGTTTGAAGG	pRBV1 vectors
Recomb RHA Rv	GCAGCGACTTCCACGATG	pRBV1 vectors
TI Scr Fw5	AAGGATTTAATAGAAACATATGC	pR702_TI
TII Scr Fw5	TGGATTTAAGGAAGAATTAACACC	pR702_TII
TIII Scr Fw3	GGAATTTCTTCTGAGGGAGGAG	pR702_TIII
pR702 Rv	ATTCAGAATGCCGAACACC	pR702
RBV1 Full TIII Fw	ATTTATTTTTTAAGATATAAATGAAATTAAAATTCAACTCAGATTTAG	pRBV1_TIII
RBV1 Full TIII Rv	CGGAATCATATGACTCGAGTCTACCATGACTTTAGTGCCTCC	pRBV1_TIII
TIII Scr Rv1	AGGTTTTTTGATTTATACTCATATGC	pRBV1_TII/III
Aada1 Fw	TCGTGCTATGATCGACTGAGATCCGTGCACAGCACC	pR702
Aada1 Rv	ATTTTTATGCTAGCATCGGCAGCGACTTCCACGATG	pR702
RBV insert Fw	TTTGGGAATGCGAGACCG	pRBV1
RBV insert Rv	TGCTGTGCACGGATCTCAGTCGATCATAGCACGATCAAC	pRBV1
RBV2_TIII_Fw	GGGAATGCGAGACCGAATTCATGAAATTAAAATTCAACTCAG	pRBV1_TIII
RBV2_TIII_Rv	GACTCGAGTATCGTGAGACCCTACCATGACTTTAGTGCCTC	pRBV1_TIII
RBV2_TIII_LHA2_Fw	AGCGCCCAATACGCAGCGATCGCAACTAGAACCAGACATAAGTGAAG	pRBV1_TII
RBV2_TIII_LHA2_Rv	GAATTTTAATTTCATGAATTATCGTTATATCTTAAAAAATAAATC	pRBV1_TII
R2-L2-T3 Fw	CGCAACTAGAACCAGACATAAGTGAAG	pRBV2_L2_TIII

2.8.2. Chapter 4 oligonucleotides

Table 9 – List of primers utilised in the development of a genetic toolkit for genomic manipulation of *C. carboxidivorans*

Oligonucleotide	Sequence	Template
TIII.2R::HC Fw1	AGCCGTAAACTCTTTCTCCC	*C. carboxidivorans* P7 gDNA
TIII.2R::HC Fw2	TCCCATAACTCTTTAAATTCTGGA	*C. carboxidivorans* P7 gDNA
TIII.2R::HC Fw3	GAAGAAAGGTTATTATATCGGGC	*C. carboxidivorans* P7 gDNA
TIII.2R::HC Fw4	TTAATTCCATCAACAATCATTAATC	*C. carboxidivorans* P7 gDNA
TIII.2R::HC Fw5	AAAGTATGACTTTCCACAACACC	*C. carboxidivorans* P7 gDNA

TIII.2R::HC Rv1	GAAGTACGTGGGGAAATAGATG	*C. carboxidivorans* P7 gDNA
TIII.2R::HC Rv2	GAGACAGGAAATAGGAAGAGGAC	*C. carboxidivorans* P7 gDNA
TIII.2R::HC Rv3	GTATGCTCTTTGGTTTGAAGAAG	*C. carboxidivorans* P7 gDNA
TIII.2R::HC Rv4	TCAAATATGTACCTTAAATGAAACC	*C. carboxidivorans* P7 gDNA
TIII.2R::HC Rv5	GAAGTACGTGGGGAAATAGATG	*C. carboxidivorans* P7 gDNA
TIII.2R Scr Fw2	AACTTCAATCCAGTCTACTCCA	*C. carboxidivorans* P7 gDNA
TIII.2R Scr Rv2	GGAAATAAGAAGGCTACAAGAG	*C. carboxidivorans* P7 gDNA
TIII.2R Scr Rv3	CTCAATTAAAGATCCTGATGGT	*C. carboxidivorans* P7 gDNA
CatP Fw	TTAAGGAGGTGTGTTACATATGGTATTTGAAAAAATTGATAAAAA TAG	pMTL8315y-v1
CatP Rv	TTATTGGCGCGCCCGTTTAAACCACCGTATTTCTAC	pMTL8315y-v1
Pfdx(Csp)-RbE* Fw	GGAAACAGCTATGACCGCGGCCGCGTGTAGTAGCCTGTGAAATA AGTAAGG	pMTL8325y-v1-RbE*-CatP(trunc)
Pfdx(Csp)-RbE* Rv	CCTTGTAAAAAATTATCCATATGCTTGTTGTTACCTCCTTAGC	pMTL8325y-v1-RbE*-CatP(trunc)
ToxAtoxFw	TAAGGAGGTAACAACAAGCATATGGATAATTTTTTACAAGGC	ToxAtoxsynthetic DNA fragment
ToxAtoxRv	GTAAAATAGCTACTCCACCGACGTCGAAACTGAAAGAAATCAATG G	ToxAtoxsynthetic DNA fragment
PyrE LHAF1	ATTGATTTCTTTCAGTTTCGACGTCGGTGGAGTAGCTATTTTACTT AAC	*C. carboxidivorans* P7 gDNA
PyrE LHAR1	AGGGTTGTGGGTTGTACGGAAAAAAAATAATTTTCAAGGAGC	*C. carboxidivorans* P7 gDNA
PyrE RHAF1	TTCCGTACAACCCACAACCCTATTCATAATAATTCCCCCTA	*C. carboxidivorans* P7 gDNA
PyrE RHAR1	CACTGGCCGGCCCAGTCGGCGCGCCGAATGTGCAAGAAAAGCTC	*C. carboxidivorans* P7 gDNA
CatP NoRep F1	AGGTGAAAGCTTCACCATCAAAGGGGCGCGCCGACTGGGCCGGC CAGTGGGC	pMTL8315y-v1
CatP NoRep R1	GTTAAGGGATTTTGGTCATG	pMTL8315y-v1
TIII.2R LHA Fw	ATTTCTTTCAGTTTCGACGTCTTTTTATAAAAACTAGGAGC	*C. carboxidivorans* P7 gDNA
TIII.2R LHA Rv	TAGGAGGGATTACTTTATGATGTAATTGTATTTATATTTTATGC	*C. carboxidivorans* P7 gDNA
TIII.2R RHA Fw	AAAATATAAATACAATTACATCATAAAGTAATCCCTCCTA	*C. carboxidivorans* P7 gDNA
TIII.2R RHA Rv	TGGCCGGCCCAGTCGGCGCGCCATATATATCCTATCGCAGCT	*C. carboxidivorans* P7 gDNA
SynCas9-1F1	GCTCACCCAAGTGAATTTGATGGGGAATAGAGC	*cas9* synthetic DNA fragment1
SynCas9-1R1	GGCAAGGCAAGGGACTAAATATAAATCTAGAGTGTAGTAGCCTGT GAAATAAG	*cas9* synthetic DNA fragment1
SynCas9-2F1	TTAAGCTGTTTCATCACCTTATCATCAAAGAGGTG	*cas9* synthetic DNA fragment2
SynCas9-2R1	ATTCCCCATCAAATTCACTTGGGTGAGCTG	*cas9* synthetic DNA fragment2
SynCas9-3F1	CCTTTTTCCACCTTAGCAACCACTAGGACTGAATAAG	*cas9* synthetic DNA fragment3
SynCas9-3R1	CTTTGATGATAAGGTGATGAAACAGCTTAAACGTCG	*cas9* synthetic DNA fragment3
SynCas9-4F1	TTTTTTATCAGGAAACAGCTATGACCGCGGCCGCTGTATCCATATG ACC	*cas9* synthetic DNA fragment4
SynCas9-4R1	TCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATC	*cas9* synthetic DNA fragment4
Cas9scr1R	GAAACTTAATCATATGCHCTAAGG	*cas9*(pMTL4315_SynCas9_adh1_N20#2)
Cas9scr1	ATGGATAAGAAATACTCAATAGGCTTAG	*cas9*(pMTL4315_SynCas9_adh1_N20#2)
Cas9scr2	GCTTTGTCATTGGGTTTGAC	*cas9*(pMTL4315_SynCas9_adh1_N20#2)

Cas9scr3	GTCGATAAAGGTGCTTCAGC	*cas9*(pMTL4315_SynCas9_adh1_N20#2)
Cas9scr4	GAACATATTGCAAATTTAGCTGG	*cas9*(pMTL4315_SynCas9_adh1_N20#2)
Cas9scr5	CTGACTTCCGAAAAGATTTCC	*cas9*(pMTL4315_SynCas9_adh1_N20#2)
Cas9scr6	GAGTTAGAAAACGGTCGTAAACG	*cas9*(pMTL4315_SynCas9_adh1_N20#2)
C.auto Fw	GCTATTTACTTAAATAAGCCTATTCC	*C. autoethangenum* gDNA
C.auto Rv	GGAAGTAGAGTATTCAAATAAGTCG	*C. autoethangenum* gDNA
pthI guide Fw1	ATTTCACAGGCTACTACACTCTAGATTTTTAACAAAATATATTGATAAAAATAATAATAG	pMTL8J15x-1673
pthI guide Rv1	CCTAACTCTTTAATGATATTGTCGACAAATTTTGATACGGGGTAAC	pMTL8J15x-1673
pthI CasGuideSeq	GTTGTTAGAGAAAACGTATAAATTAG	SynCas vectors
TIII.2R LHA Fw	TGCTTTTTTTTAGACGTCCCTTGCCATAAATGAAATGTGC	*C. carboxidivorans* P7 gDNA
TIII.2R LHA Rv	GGAGGGATTACTTTATGAATGTAATTGTATTTATATTTTATGCAAAAACG	*C. carboxidivorans* P7 gDNA
TIII.2R RHA Fw	TGCATAAAATATAAATACAATTACATTCATAAAGTAATCCCTCCTATATACTCATAATATC	*C. carboxidivorans* P7 gDNA
TIII.2R RHA Rv	TTTTTATTGGCGCGCCGTATTAGAAGAGCAGGAGAGAAAATAGTCC	*C. carboxidivorans* P7 gDNA
sgRNA1 TIII.2R Fw	TTTCGTCGACTCGACGAATATAAGAAAATTCAATGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG	pRESynCas9(thI)-v1
sgRNA2 TIII.2R Fw	TTTCGTCGACTCGACTCTGGGTTATAGTTTCCAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG	pRESynCas9(thI)-v1
sgRNA3 TIII.2R Fw	TTTCGTCGACTCGACAAGGAATACGAAGAAGATGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG	pRESynCas9(thI)-v1
sgRNA3 TIII.2R Rv	TTTCATTTATGGCAAGGGACGTCTAAAAAAAAGCACCGAC	pRESynCas9(thI)-v1

2.8.3. Chapter 5 oligonucleotides

Table 10 – List of primers employed in the genomic modification of *E. limosum* NG-6849 with both a CRISPR-Cas9 and *mariner* transposon system.

Oligonucleotide	Sequence	Template
ORFD_sgRNA Fw	TAAGGAGGAGTTTTCGTCGACGGGATGGACAGAAAAAGTGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCC	pRECas vectors
ORFD sgRNA Rv	GATTGCTAATGCTTGCTGGCGATCGCATAAAAATAAGAAGCCTGCAAATG	pRECas vectors
ORFD LHA Fw	TTCTTATTTTTATGCGATCGCCAGCAAGCATTAGCAATCTTTAACTCATGG	*E. limosum* NG-6849 gDNA
ORFD LHA Rv	TGTAATTGTTTTTTCTGATTTTAATCTGGCATAGTTATCACTCCTTTGTCGTTTTTTC	*E. limosum* NG-6849 gDNA
ORFD RHA Fw	AAGGAGTGATAACTATGCCAGATTAAAATCAGAAAAAACAATTACAGTCAGATGAAATATTTAGG	*E. limosum* NG-6849 gDNA
ORFD RHA Rv	AAAAAATAATGGCGGCGCGCCGCCTGCGGACAGATCCCTCC	*E. limosum* NG-6849 gDNA
JM115	TTTTGATACTGTGGCGGTCTGTATTTCC	pRECas vectors
ORFD Scr Fw	TAGATACCACACAGGCTGTTCCTG	*E. limosum* NG-6849 gDNA
ORFD Scr Rv	AATCTGAGCCCCTCAGCATTAAAG	*E. limosum* NG-6849 gDNA
SplA5_top	G*AGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*	-
SplA5_bottom	G*ATCGGAAGAGCGGTTCAGCAGGTTTTTTTTTTCAAAAAAA*A	-
iPCRtagSeq	AAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTC	TraDIS library sample
Library Primer	AATGATACGGCGACCACCGAGATCTACACTTCGACTGCACTCCCGACTTAATAACTTC	TraDIS library sample

5' Tn1	AATGATACGGCGACCACCGAGATCTACACTTCGACTGCACTCCCG ACTTAATAACTTC	TraDIS prepared gDNA
SplAP5.1	C*AAGCAGAAGACGGCATACGAGATAACGTGATGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.2	C*AAGCAGAAGACGGCATACGAGATAAACATCGGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.3	C*AAGCAGAAGACGGCATACGAGATATGCCTAAGAGATCGGTCTC GGCATTC*C	TraDIS prepared gDNA
SplAP5.4	C*AAGCAGAAGACGGCATACGAGATAGTGGTCAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.5	C*AAGCAGAAGACGGCATACGAGATACCACTGTGAGATCGGTCTC GGCATTC*C	TraDIS prepared gDNA
SplAP5.6	C*AAGCAGAAGACGGCATACGAGATACATTGGCGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.7	C*AAGCAGAAGACGGCATACGAGATCAGATCTGGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.9	C*AAGCAGAAGACGGCATACGAGATCGCTGATCGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.10	C*AAGCAGAAGACGGCATACGAGATACAAGCTAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.11	C*AAGCAGAAGACGGCATACGAGATCTGTAGCCGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.12	C*AAGCAGAAGACGGCATACGAGATAGTACAAGGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.13	C*AAGCAGAAGACGGCATACGAGATAACAACCAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.14	C*AAGCAGAAGACGGCATACGAGATAACCGAGAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.15	C*AAGCAGAAGACGGCATACGAGATAACGCTTAGAGATCGGTCTC GGCATTC*C	TraDIS prepared gDNA
SplAP5.16	C*AAGCAGAAGACGGCATACGAGATAAGACGGAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.17	C*AAGCAGAAGACGGCATACGAGATAAGGTACAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.18	C*AAGCAGAAGACGGCATACGAGATACACAGAAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.19	C*AAGCAGAAGACGGCATACGAGATACAGCAGAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.20	C*AAGCAGAAGACGGCATACGAGATACCTCCAAGAGATCGGTCTC GGCATTC*C	TraDIS prepared gDNA
SplAP5.21	C*AAGCAGAAGACGGCATACGAGATACGCTCGAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.22	C*AAGCAGAAGACGGCATACGAGATACGTATCAGAGATCGGTCTC GGCATTC*C	TraDIS prepared gDNA
SplAP5.23	C*AAGCAGAAGACGGCATACGAGATACTATGCAGAGATCGGTCTC GGCATTC*C	TraDIS prepared gDNA
SplAP5.24	C*AAGCAGAAGACGGCATACGAGATAGAGTCAAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA
SplAP5.25	C*AAGCAGAAGACGGCATACGAGATAGATCGCAGAGATCGGTCT CGGCATTC*C	TraDIS prepared gDNA

(*indicating a phosphorothioate group)

2.8.4. Synthetic DNA fragments

Table 11 – List of longer DNA fragments ordered as gBlocks from IDT

Fragment	Length (bp)	Sequence
SynCas1	1494	atttgatggggaatagagccgttgtcaaaggtccgttgcttgcgcagcaaatcttcacgatttagtttcaccaataat tcctcagtaccatccatttttttctaaaattggtttgataaatttataaaattcttcttgtgaagctccccccatcaatata acctgcatatccgttCtttgattgatcaaaGaagatttctttatacttttctggaagttgttgtcgaactaaTgctttt aaaagagtcaagtcttgatgatgttcatcgtagcgtttaatcattgaagctgataggggagccttagttatttcagta tttactcttaggatatctgaaagtaaaatagcatctgataaattcttagctgccaaaaacaaatcagcatattgatct ccaatttgcgccaataaattGtctaaatcatcatcgtaagtatcttttgaaagctgtaatttagcatcttctgccaaat caaaatttgatttaaaattaggggtcaaacccaatgacaaagcaatgagattcccaaataagccatttttcttctca ccggggagctgagcaatgagattttctaatcgtcttgatttactcaatcgtgcagaaagaatcgctttagcatctact ccacttgcgttaatagggttttcttcaaataattgattgtaggtttgtaccaactggataaatagtttgtccacatcac tattGtcaggatttaaatctccctcaatcaaGaaatgaccacgaaacttaatcatgtgcgctaaggccaaatagat

		taagcgcaaatccgctttatcagtagaatctaccaatttttttAcgcagatgatagatagttggatatttctcatgata agcaacttcatctactatatttccaaaaataggatgacgttcatgTttcttgtcttcttccaccaaGaaagactcttc aagtcgatgaaagaaactatcatctactttcgccatctcatttgaaaaaatctcctgtagataacaaatacgattctt ccgacgtgtataccttctacgagctgtccgtttgagacgagtcgcttccgctgtctctccactgtcaaataaaagag cccctataagatttttCttgatactgtggcggtctgtatttcccagaaccttgaacttttttagacggaaccttataGtc atcagtgatcaccgcccatccgacgctatttgtgccgatatctaagcctattgagtatttcttatCCATCTTGTTG TTACCTCCTTAGCAGGGTGCTGCCAAGGGCATCAAGACGATGCTGGTATCACCGGAAC CTATAGTGAGTCGTATTGGTACCGATTTATATTTTACCATGTTTTTATTAATTTTTAAAGT AATTTTTATATAACTGTATAACCTATATTTTCTATGGATTCTACTATCCTATCTACATCTAC AAAATAATCATCATATATAACACTTACTTCTTTTTTTTCCTTACTTATTTCACAGGCTACTA CAC
SynCas2	700	caccttatcatcaaagaggtgagcataagttttaagtctttcctcaatcatctccctatcttcaaataaggtcaatgtt aaaacaatatcctctaagatatcttcattttcttcattGtccaGaaaatctttatctttaataatttttagcaaatcat ggtaggtacctaatgaagcattaaatctatcttcaactcctgaaatttcaacactatcaaaacattctattttCttga aataatcttctttttaattgcttaacggttactttAcgatttgttttgaagagtaaatcaacaatggctttcttctgttca cctgaaagaaatgctggttttcgcattccttcagtaacatatttgacctttgtcaattcgttataaaccgtaaaatact cataaagcaaactatgttttggtagtacttttcatttggaagatttttatcaaagtttgtcatgcgttcaataaatgat tgagctgaagcacctttatcgacaacttcttcaaaattccatggggtaattgtttcttcagacttccgagtcatccatg caaaacgactattgccacgcgccaatggaccaacataataaggaattcgGaaagtcaagatttttttcaatcttctc acgattgtcttttaaaaatggataaaagtcttcttgtcttctcaaaatagcatgcagctcacccaagtga
SynCas3	1500	agcaaccactaggactgaataagctaccgttggactatcaaaaccaccatatttttttAggGtcccagtctttttttac gagcaataagTttgtccgaatttctttttAggtaaaattgactccttggagaatccgcctgtctgtacttctgttttcttt gacaatattgacttggggcatggacaatactttgcgcactgtggcaaaatctcgccctttatcccagacaatttctcc agtttccccattagtttcgattagagggcgtttgcgaatctctccatttgcaagtgtaatttctgttttgaagaagttca tgatattagagtaaaagaaatattttgcggttgctttgcctatttcttgctcagacttagcaatcattttacgaacatc ataaactttataatcaccatagacaaactccgattcaagttttggatatttcttaatcaaagcagttccaacgacgg catttagatacgcatcatgggcatgatggtaattgttaatctcacgtactttatagaattggaaatctttAcggaagt cagaaactaatttagattttaaggtaatcactttaacctctcgaataagtttatcattttcatcgtatttagtattcatg cgactatccaaaatttgtgccacatgcttagtgatttggcgagtttcaaccaattggcgtttgataaaaccagcttta tcaagttcactcaaacctccacgttcagctttcgttaaattGtcaaacttacgttgagtgattaacttggcgtttaga agttgtctccaatagtttttcatcttttCactacttcttcacttggaacgttatccgatttaccacgatttttatcagaa cgcgttaagaccttattgtctattgaatcgtctttaaggaaactttgtggaacaatgtgatcgacatcataatcactt aaacgattaatatctaattcttggtccacatacatgtctcttccattttggagGtaatagagatagagcttttcatttt gcaattgagtattttcaacaggatgctctttaagaatctgacttcctaattctttgatacttcttcgattcgtttcata cgctctcgcgaatttttctggccctttgagttgtctgattttcacgtgccatttcaataacgatattttctggcttatgc cgccccattactttgaccaattcatcaacaacttttacagtctgtaaaatacctttttttaatagcagggctaccagct aaatttgcaatatgttcatgtaaactatcgccttgtccagacacttgtgctttCtgaatgtcttctttaaatgtcaaact atcatcatggatcagctgcataaaattgcgattggcaaaaccatctgatttcaaGaaatctaatattgttttgccag attgcttatccctaataccattaatcaattttcgagacaaacgtccccaaccagtataacggcgacgtttaagctgtt tcat
SynCas4	695	tgtatccatatgaccatgattacgaattctcagtcacctcctagctgactcaaatcaatgcgtgtttcataaagacca gtgatggattgatggataagagtggcatctaaaacttcttttgtagacgtatatcgtttacgatcaattgttgtatcaa aatatttaaaagcagcgggagcaccaagattcgtcaacgtaaataaatgaataatatttttctgcttgttcacgtatt ggtttgtctctatgtttgttatatgcactaagaactttatctaaattggcatctgctaaaataacacgcttagaaaatt cactgatttgctcaataatctcatctaaataatgcttatgctgctccacaaacaattgtttttgttcgttatcttctgga ctacccttcaacttttcataatgactagctaaatataaaaaattcacatatttgcttggcagagccagctcatttccC ttttgtaattctccggcactagccagcatccgtttacgaccgttttctaactcaaaCagactatatttaggtagtttaa tgattaagtctttttttaacttccttatatcctttagcCtctaaaaagtcaatcggatttttttcaaaggaacttctttcca taattgtgatccctagtaactctttaacggattttaacttcttcgatttccctttttccaccttt
II/IIIABC	168	gtacccggggcttctatcgggtcagaagccgcctattagaagctgaaactggatccagaggttttttttgccggcg gtcaaaaagtgtgtgatcgaaaaactccggtcgacgttcgattatcgtggtggtgataatccagtccggataatttc tgggcgcgccaaTa
ToxATox	502	atggataatttttttacaaggcatactagcaagtctatctgctagcttaatagtttatatagctagcaaactatttaga aagcgtaaaaaaccactcaaagcggcaactaagagtggttgggaatttgatttaaaaatcagattccataaaact aagtaatttctaaattatgaactccactctacagcaaaatagattgtagttcttcttgcttttattatatcacaaattg gtacagatattcaaaaataatatttttatgatataataaagtcatagaaattttgcagtgttcgatttttgaagaaat taaggcttaacagttgaaatataaggcattgaggacttgtgataagtgttatcaattgcactactcatggttcactgc aaatttgagagatttgtatatgtgtaggcattggaaatactcaatttatttttggggttttatattaactatatggaatgt aaataaaggtgtccattgatttctttcagtttc

2.9. Plasmid construction

2.9.1. pMTL8315y-v*

pMTL8315y-v1 was assembled through a four-part HiFi reaction, with a ratio of 1:1:1:1 of DNA fragments amplified by PCR from the pMTL83151 using the primer pairs: Tra SNP Fw and MCS Trunc Rv, pCB102 SNP1 Fw and pCB102 SNP1 Rv, pCB102 SNP Fw and CatP PmeI Rv, ColE1 PmeI Fw and Tra Rv (**Table 8**). These fragments were 784 bp, 233 bp, 1999 bp and 999 bp respectively. A 10 µL aliquot of the HiFi reaction mixture was then used to transform chemically competent *E. coli* Top10 by heat shock and recovered on chloramphenicol selective LB plates for 2 days. To screen for correct assembly colonies were picked from these plates into comparatively selective LB broth and grown overnight, prior to plasmid isolation. Successful pMTL8315y-v1 assembly was then confirmed by complete vector sequencing with CatP_R1, ColE1_tra-R1, CatP-F1, pCB102_R1, pCB102-R1 and p15a_tra_F2.

To assemble pMTL8315y-v2 three fragments were amplified by PCR from pMTL8315y-v1 using the respective primer pairs pMTL8315-v2 Frag1 Fw/Rv, pMTL8315-v2 Frag2 Fw/Rv and pMTL8315y-v2 Frag3 Fw/Rv. For pMTL8315y-v3 the PCR amplified 'Frag3' remained the same, but for 'Frag1' the reverse primer was replaced with pMTL8315y-v3 Frag1 Rv. For 'Frag2' the forward primer was replaced with pMTL8315y-v3 Frag2 Fw. Both vectors were then assembled from their respective DNA fragments by a three-part HiFi reaction at a ratio of 1:1:1, and similarly used to transform *E. coli* Top10 before plasmid isolation and verification by sequencing of the modified region using CatP_F1, p15a_F1 and pCB102_R1.

2.9.2. pRBV1_T*

A golden gate assembly method was employed for the transfer of methyltransferase/specificity associated genes from *C. carboxidivorans* into pRBV1. The methyltransferase associated genes of each of the eight RM systems were amplified by PCR from gDNA with the Ccarb Mtase Fw/Rv primer pairs indicated in **Table 8**. Here the RM systems have been denoted as TI.1, TI.2, TI.3, TII.1, TII.2, TII.3, TIII.1 and TIII.2. These DNA fragments were approximately

2.8 kb, 2.9 kb, 2.85 kb, 0.85 kb, 1.7 kb, 3.8 kb, 3 kb and 1.95 kb respectively. pRBV1 was then linearised by digestion with EcoRI-HF and dephosphorylated to prevent plasmid randomly re-circularising (**2.4.4**). Purified vector and DNA fragments were assembled at a ratio of 1:2 in a 20 µL reaction mix and placed in a thermocycler with the program outlined in **2.4.10**, before 10 µL of this mix was used to transform chemically competent *E. coli* Top10. Colonies were screened by PCR with Recomb LHA Fw and Recomb RHA Rv. Those which showed the expected size increase compared to pRBV1 then underwent sequencing to confirm sequence integrity.

The same assembly method was utilised for the assembly of multiple methyltransferase/specificity gene fragments into a singular pRBV1 backbone. Three alternate reactions were undertaken with a variable number of genes to minimise total assembly length: the three TI methyltransferases together, three TII methyltransferases and two TIII methyltransferases. 10 µL of each of these reaction mixes was used to transform chemically competent *E. coli* Top10 and incubated on erythromycin LB selection plates for 24 hours. Initial screening was performed by colony PCR using an internal primer (TI Scr Fw5, TII Scr Fw5 and TIII Scr Fw3) and external primer (Recomb RHA Rv), and positive results were picked into 5 mL selective liquid LB media. The plasmids were then isolated using the Monarch® Miniprep Kit and sequenced.

2.9.3. pRBV1_TII-III

To combine the methyltransferase associated genes of the three Type II and two Type III RM systems, the *CcaP7ORF178030P* and *CcaP7ORF8705P* were amplified together from pRBV1_TIII using pRBV1 Full TIII Fw and RBV1 Full TIII Rv. This fragment was then introduced into pRBV1_TII (linearised by digestion with PspXI) by two-part HiFi. This HiFi mix was then used to transform *E. coli* Top10 and transformants selected on LB agar supplemented with erythromycin. Colonies were screened by PCR with TIII Scr Fw3 and Recomb RHA Rv primers, as well as with TII Scr Fw5 and TIII Scr Rv1 to cover the junction

between the two cassettes. Colonies which generated expected bands for both colony PCRs were picked into selective LB broth and grown overnight before the culture underwent plasmid isolation. Further confirmation was performed by subsequent sequencing of these plasmids.

2.9.4. pRBV2_L2-TIII

To initially generate pRBV2 the *aadA1* gene was amplified from pR702 by colony PCR of *E. coli* sExpress using the primers aadA1 Fw and aadA1 Rv. A further PCR of the adjacent insert fragment was performed from pRBV1 with RBV insert Fw and RBV insert Rv. Both fragments were combined with pRBV1, digested with NheI and EcoRI-HF, in a three-part HiFi reaction to generate pRBV2. This HiFi reaction mix was used to transform *E. coli* Top10 and transformants were selected for on erythromycin supplemented LB agar plates. Colonies were picked into comparably selective broth and grown overnight before undergoing plasmid isolation, with pRBV2 assembly screening and final confirmation performed with Sanger sequencing. The two Type III RM system associated methyltransferases within pRBV1_TIII were then amplified by PCR directly from the plasmid, with the primers RBV2_TIII_Fw and RBV2_TIII_Rv. This fragment was assembled into EcoRI-HF linearised pRBV2 by HiFi reaction, with plasmid selection and screening as before. This pRBV2_TIII vector was subsequently digested with AsiSI and EcoRI-HF. A new left homology arm (LHA) for the cassette was produced by PCR amplification of the terminal TII methyltransferase region (CcaP7ORF7120P fragment) from *C. carboxidivorans* genomic DNA with primers RBV2_TIII_LHA2_Fw and RBV2_TIII_LHA2_Rv. HiFi assembly was performed with these two fragments. This HiFi mix was used to transform *E. coli* and subsequently selected as before with pRBV2_L2-TIII assembly confirmation by sequencing.

2.9.5. pMTL8315y-v1-II/III*

pMTL8315y-v1-II/IIIABC was constructed by digestion of both pMTL8315y-v1 and the II/III-ABC synthetic DNA fragment with XmaI and AscI. These two fragments were then ligated as detailed in **2.4.5**, before the resultant ligation mix was used to transform *E. coli* Top10. Transformants were selected on LB and chloramphenicol, and five colonies were picked into liquid media for subsequent plasmid isolation. Successful assembly was confirmed by sequencing with p15a_tra_F2.

The three further vectors pMTL8315y-v1-II/IIIA, pMTL8315y-v1-II/IIIB and pMTL8315y-v1-II/IIIC were constructed in a similar fashion. In lieu of the synthetic DNA fragment pMTL8315y-v1-II/IIIABC was digested, alongside the initial pMTL8315y-v1. For II/III-A XmaI and BamHI-HF were used, for II/III-B BamHI-HF and SalI-HF and for II/III-C SalI-HF and AscI. Ligation, selection and plasmid isolation proceeded as before with p15a_tra_F2 mediated sequencing confirming sequence integration.

2.9.6. pMTL8015y- v1-TIII.2R-SC*

Integration vectors were generated through the removal of the pCB102 Gram-positive replicon cassette of pMTL8315y-v1, by digestion with AscI and FseI. The six homology regions were amplified from *C. carboxidivorans* C94 genomic DNA using the primer pairs: TIII.2R::HC Fw5/Rv3, TIII.2R::HC Fw1/Rv1, TIII.2R::HC Fw2/Rv2, TIII.2R::HC Fw3/Rv4, TIII.2R::HC Fw5/Rv4 and TIII.2R::HC Fw1/Rv1 (**Table 9**). These were then assembled into the linearised backbone using HiFi assembly, and the reaction mix was used to transform chemically competent *E. coli* 10-beta before selection on LB plates with chloramphenicol. Four transformants for each vector were picked into comparably selective LB media and grown for 24 hours before undergoing plasmid isolation and Sanger sequencing with the p15a_tra_F2 primer.

2.9.7. pMTL8015y-v1-ToxAtox-TIII.2R

PCR amplification of the received synthetic CD2517.1-RCd8 region was performed with the primer pair ToxAtoxFw and ToxAtoxRv. Amplification of the $P_{fdx(Csp)}$-RbE* inducible promoter system was performed using Pfdx(Csp)-RbE* Fw and Pfdx(Csp)-RbE* Rv from pMTL8325y-v1-$P_{fdx(Csp)}$-RbE*-CatP(trunc). The two *88and* framing homology arms were similarly amplified by PCR from *C. carboxidivorans* genomic DNA using PyrE LHAF1/R1 and PyrE RHAF1/R1 (**Table 9**). These four fragments were assembled in a HiFi reaction with pMTL8315y-v1 linearised with NotI and AscI. 10 µL of this reaction mix was used to transform chemically competent *E. coli* 10-beta and successful transformants were identified by chloramphenicol selective growth on LB plates. These were picked into comparably selective LB broth and underwent plasmid isolation before successful assembly was confirmed by sequencing with p15a_tra_F2 and pCB102_R1.

Excision of the pCB102 module from this pMTL8315y-v1-ToxATox-PyrE was then performed by restriction digestion with AscI and PmeI. A CatP module without adjacent replicon was then amplified using CatP NoRep F1/R1. This purified 846 bp fragment was once more assembled into the linearised backbone using HiFi, used to transform *E. coli* 10-beta with chloramphenicol selection, before the correct vector was isolated and sequenced using p15a_tra_F2 and CatP_R1. Following this, two homology arms were amplified from *C. carboxidivorans* gDNA using TIII.2R LHA Fw/Rv and TIII.2R RHA Fw/Rv. The pMTL8015y-v1-ToxATox-PyrE vector generated was linearised by digestion with AscI and AatII. These three fragments underwent HiFi assembly and subsequent *E. coli* 10-beta transformation with chloramphenicol selection. Growth in comparably selective broth culture and plasmid isolation preceded preliminary confirmation of successful homology arm integration by restriction digestion with AscI and AatII. Successful integrants yielded a 1950 bp band as

opposed to the 946 bp of the *pyrE* homology arm cassette. Final assembly confirmation was performed using Sanger sequencing.

2.9.8. pMTL8325y-Pxxx-CatP(trunc)

To generate the pMTL8325y-v1 vector, pMLT8315y-v1 and pMTL83251 (SBRC Culture Collection, Nottingham) underwent restriction digestion with the enzymes FseI and PmeI. The purified 906 bp ErmB cassette insert was then ligated into the purified 3256 bp pMTL8015y-v1 backbone. This ligation reaction mix was used to transform into *E. coli* 10-beta and transformants selected with erythromycin. Individual colonies were picked into comparably selective broth and grown overnight before plasmid DNA was extracted. Screening was performed by diagnostic digest with FseI and PmeI to confirm the reduction in antibiotic selection marker from 906 bp to 777 bp, and further confirmed by sequencing with the primer p15a_tra_F2 to ensure the expected truncation of the multiple cloning site (MCS) region present in pMTL8315y-v1 was still present, and hence the vector was not residual pMTL83251.

The $P_{fdx(Csp)}$-RbE*-CatP cassette was then removed from pMTL83254-RbE* by digestion with NotI and AscI. pMTL8325y-v1 was similarly digested and the cassette ligated with the backbone as detailed in **2.4.5**. 10 µL of this reaction mix was then used to transform *E. coli* Top10, with five transformant colonies which arose on erythromycin plates subsequently screened for correct assembly through plasmid isolation and sequencing with p15a_tra_F2 and pCB102_R1. An alternate truncated *catP* reporter module, CatP(trunc), was amplified by PCR from pMLT8315y-v1 using the primers CatP Fw and CatP Rv. This fragment and the pMTL83254_$P_{fdx(Csp)}$-RbE*-CatP vector both underwent digestion with AscI and NdeI, before the backbone and digested amplicon were ligated as before. This was then used to transform *E. coli* 10-beta, with transformants selected on erythromycin. Transformants were picked into LB supplemented with erythromycin and grown overnight before plasmid isolation was performed. These then underwent sequencing with p15a_tra_F2 and pCB102_R1 to confirm the expected truncated CatP module was present.

This resulted in the generation of the initial pMTL8325y-v1-$P_{fdx(Csp)}$-RbE*-CatP(trunc) vector. All further vectors were assembled through the digestion of this backbone with NdeI and NotI, and subsequent ligation with different promoter sequence fragments. These fragments were generated by identical digestion of the vectors pMTL8J15x-999, pMTL8J15x-araE, pMTL8J15x-1673 and pMTL8J15x-NP. All transformations were performed using *E. coli* Top10 as detailed in **2.3.2,** with erythromycin-based selection and sequencing with p15a_tra_F2 to confirm vector assembly.

2.9.9. pMTL4315_SynCas9_adh1_N20#2

pMTL4315_SynCas9_adh1_N20#2 was constructed by HiFi assembly of the four PCR amplified synthetic DNA fragments (**Table 11**) into the linearised backbone of pMTL4315_CRISPR_ah1_N20#2. The linearisation of this vector was performed by restriction digestion using the enzymes SalI-HF and NdeI. The HiFi compatible primers used to amplify the $P_{fdx(Csp)}$-RbE*-Cas9 region are detailed in **Table 9** with the general format of SynCas9-XF1 and SynCas9-XR1 respectively, in which X denotes fragment number. This HiFi reaction mix was used to transform *E. coli* 10-beta and plated onto LB supplemented with chloramphenicol. After 24 hours of growth colonies were picked into comparably selective LB broth. This was cultured overnight before being used for plasmid extraction, and subsequent complete region sequencing with the seven Cas9scr primers (1R, 1, 2, 3, 4, 5 and 6).

2.9.10. pRESynCas(pthI)-v1-TIII.2R-sg*

The full CRISPR-Cas9 functional cassette (*cas9*-sgRNA-LHA-RHA) was initially transferred to pMTL8315y-v1 by digestion of pMTL4315_SynCas9_adh1_N20#2 and pMTL8315y-v1 with SbfI and AscI. The resultant 6584 bp fragment of pMTL4315_SynCas9_adh1_N20#2 and 3887 bp backbone of pMTL8315y-v1 were ligated using T4 DNA ligase. 10 µL of this

reaction mix was then used to transform into *E. coli* 10-beta. Transformants were grown on LB with chloramphenicol selection and transferred to broth following 24 hours. These cultures were grown overnight before undergoing plasmid isolation. The isolated plasmids were then screened by diagnostic digest with SbfI and AscI. Subsequent confirmation of the correct pMTL8315y-v1 backbone was performed by sequencing with p15a_F1.

The P_{araE} promoter driving sgRNA transcription was then replaced. The 138 bp replacement P_{thl} sequences was amplified from pMTL8J15x-1673 using the primers pthI guide Fw1 and pthI guide Rv1. This was then combined with the linearised backbone of the above pRESynCas9-v1 intermediary vector, digested with XbaI and SalI-HF, through a HiFi reaction. 10 µL of this reaction mix was used to transform *E. coli* 10-beta and recovered on LB plates supplemented with chloramphenicol. Colonies were picked into comparably selective LB broth and grown overnight prior to plasmid isolation and sequence confirmation with pthI CasGuideSeq.

Two homology arms of 486 bp and 500 bp respectively, framing *CcaP7ORF17830P* with a 5 bp overlap, were amplified by PCR from *C. carboxidivorans* gDNA using the primer pairs TIII.2R LHA Fw/Rv and TIII.2R RHA Fw/Rv respectively. The resultant 521 bp and 541 bp fragments were assembled into a single 1034 bp homology arm cassette using splicing by overlap extension (SOEing) PCR. This PCR product and the above generated pRESynCas9(thI)-v1 were digested with AscI and AatII, before being combined by ligation using T4 DNA Ligase. 10 µL of this ligation mix was then used to transform *E. coli* 10-beta and plated onto LB agar supplemented with chloramphenicol. Following overnight selective broth growth and plasmid isolation, assembly was confirmed through AscI and AatII diagnostic digestion which yielded a 1682 bp dropout for incorrect vectors and 989 bp for successful assemblies.

20 bp protospacer target sequences, with the appropriate adjoining NGG PAM at the genomic locus, were identified within the *CcaP7ORF17830P* using the online software platform Benchling, with algorithms derived from work by

Doench et al (2016). sgRNAs were chosen to optimise on and off-target scores, with two of the chosen on the forward strand and one on the reverse. Exchange of the sgRNA primer within the vector was performed by amplification of a modified synthetic guide region. Within the primers for this exchange the N20#2 20 bp sequence was replaced with each of the chosen guides sg1-3 respectively. To ensure plasmid annealing these extended length primers contained 34 bp annealing immediately downstream of this non-annealing protospacer sequence (Tm of 63.7^{o}C). Each of these sgRNA* TIII.2R Fw primers (in which * is 1-3) was used in combination with sgRNA TIII.2R Rv to amplify the new sgRNA regions. These were then digested with SalI-HF and AatII, before being ligated into the comparably digested pRESynCas9(thI)-v1-TIII.2R. Ligations were used to transform *E. coli* 10-beta, plated onto LB with chloramphenicol, picked into liquid broth and used for plasmid isolation. Plasmids were then sequenced with pthI CasGuideSeq to confirm new sgRNA inclusion.

2.9.11. pMTL83254-Rb*

pMTL83254-RbE* was initially constructed by the AscI and FseI digestion of pMTL-IC101-RbE* and pMTL83151. The 1466 bp pCB102 Gram-positive replicon module was then isolated and ligated within the pMTL-IC101-RbE* backbone. The ligation mix was used to transform *E. coli* 10-beta and selected on erythromycin. Successful assembly was confirmed by diagnostic digest with AscI and FseI before further confirmation by sequencing. This same assembly approach was repeated for all riboswitch reporter vectors D-J (excluding E).

2.9.12. pRECas(*)-ORFD

Two pRECas vectors were acquired from J. Millard (SBRC, Nottingham), each with a pMTL83151 backbone, but with either a P_{araE} or P_{thI} promoter driving sgRNA expression. To target *R.EliORFD* two homology arms were amplified

from *E. limosum* NG-6849 gDNA with the primer pairs ORFD LHA Fw/Rv and ORFD RHA Fw/Rv (**Table 10**). The purified 805 bp and 779 bp were then assembled into a singular fragment by SOEing PCR **2.4.8.2.** The sgRNA fragment was amplified from the received pRECas(thI) vector using ORFD sgRNA Fw/Rv. Following isolation both fragments were digested, using Sal-HF and AsiSI for the sgRNA fragment, and AscI and AsiSi for the homology cassette. Both pRECas backbones were digested with AscI and Sal-HF before a three-way ligation was performed with the two fragments. 10 µL of each ligation mix was then used to transform *E. coli* Top10 and plated onto LB agar supplemented with chloramphenicol. Following overnight selective broth growth and plasmid isolation, assembly was confirmed through sequencing with JM115 and pCB102_R1.

2.10. pR702 recombineering

Linear DNA was amplified from pRBV1_TI, pRBV1_TII and pRBV1_TIII by PCR using the primer pair Recomb LHA Fw and Recomb RHA Rv. This generated fragments of 10918 bp, 9920 bp and 8569 bp respectively. To ensure sufficient fragment yield the PCR protocol employed a Q5 polymerase, touch-down thermocycler program, a reduced extension temperature (70^{o}C) and an increased extension time. These PCR products were purified using a Qiagen PCR clean up kit and eluted in 44 µL nuclease free water. To this was added 1 µL of Dpn-1 and 5 µL of CutSmart® Buffer before incubation at 37^{o}C for 1 hour. This digested residual plasmid template DNA prior to agarose gel electrophoresis and purification. The total volume of each purified linear fragment was then used to transform electrocompetent ReInterStellar, primed by 100 mM L-arabinose induction of the pKD46 borne λ-red recombineering genes (P_{BAD} promoter) at OD_{600} 0.1 during competent cell preparation. Transformants were recovered at 30°C on LB agar plates supplemented with 50 µg/mL apramycin and 50 µg/mL kanamycin.

Any transformants which arose were re-streaked two further times following 24 hours of growth on comparably selective LB plates at 37^{o}C. These were then replica plated onto LB plates supplemented with 250 µg/mL spectinomycin, and LB plates supplemented with 25 µg/mL chloramphenicol. Colonies exhibiting spectinomycin and chloramphenicol sensitivity were carried forward. These sensitivities were reflective of the excision and replacement of the *aadA1* gene in pR702, and the loss of the pKD46 plasmid due to growth at 37^{o}C destabilising the temperature-sensitive replicon repA101ts, respectively. Further confirmation of methylation cassette integration was performed by colony-PCR. Each PCR used a primer which bound within the introduced cassettes (TI Scr Fw5, TII Scr Fw5, TIII Scr Fw3) and pR702 Rv, an external primer binding outside of the recombination locus. Successful integrants generated DNA fragments of 2837 bp, 2743 bp and 2596 bp respectively.

Transfer of confirmed recombinant pR702s was performed by plasmid isolation from 5 mL of InterStellar pRBV1_T* culture using the Monarch® Miniprep Kit protocol in conjunction with a primed GenElute^{TM} Bacterial Genomic DNA kit column. 10 µL of the 50 µL nuclease-free water elute was then used to transform chemically competent *E. coli* Express. The remaining of this eluate was stored at -20^{o}C prior to next generation sequencing.

2.11. Chloramphenicol acetyltransferase assay

The *catP* gene, initially derived from *Clostridium perfringens*, encodes a chloramphenicol acetyltransferase (CAT) enzyme. This gene has been frequently implemented as a reporter within a CAT assay in which rates of 5-thio-2-nitrobenzoate (TNB) production is used to infer gene expression levels. This TNB is produced by the enzyme catalysed reaction of dithio-bis-[2-nitrobenzoic acid] (DTNB) with the free CoA sulfhydryl group. The rate of TNB production is recorded by measuring absorbance at 412 nm where its molar coefficient is 13600 M^{-1} cm^{-1} (Ellman, 1959).

2.11.1. CAT assay sample preparation

CatP reporter vectors, were introduced into the target organism as outlined in **2.3**. At this stage replicate colonies were used to inoculate each strain's respective media, supplemented with 6 µg/mL clarithromycin, and grown at 37°C in an anaerobic cabinet. The growth of each culture was monitored using OD_{600} and samples taken once they reached mid-late exponential phase (0.8-1.2). Sample volume taken was calculated from OD_{600} to an equivalent 1 OD_{600} of 1 mL culture. Cell pellets were then generated by centrifugation at 17000 x *g* for 10 minutes. Supernatant was discarded and the sample tubes stored at -20°C prior to final assay preparation.

Protein extraction of these cell pellets was performed using 250 µL of 1xBugBuster® supplemented with Calbiochem Protease Inhibitor Cocktail Set VII (Merck) 1:100. Samples were incubated at 37°C with 225 RPM shaking agitation for 45 minutes, before being placed on ice. These were subsequently disrupted by sonication using the Bioruptor® (Diagenode) at 4°C with 10 cycles of high intensity sonication with 30 seconds on, 30 seconds off. Samples were once more pelleted at 17000 x *g* for 2 minutes at 4°C, and 200 µL lysate removed for downstream steps.

2.11.2. CAT assay

An assay master mix was prepared with 0.005% w/v chloramphenicol (Sigma Aldrich); 0.167 mM acetyl-CoA sodium salt (Sigma Aldrich); 0.0833 mM 5,5'-DTNB (ThermoFisher Scientific); 100 mM Tris buffer pH 7.8 (Fisher Scientific). This master mix and the samples were prewarmed to room temperature prior to assay execution. A clear Corning™ polystyrene flat bottomed 96 well plate was loaded with 90 µL of Tris pH7.8 in each assay well, followed by 10ul of each sample. Samples were analysed using a Tecan Infinite® M1000 Pro plate reader (TECAN) maintained at 25°C. The TECAN was prepared for injection by injector

flushing with three 2 mL strokes of dH_2O, followed by three strokes of the prepared master-mix. To commence the assay 100 µL of master-mix was injected into specified wells at a rate of 100 µL/s. This mix was agitated using four seconds of orbital shaking at a magnitude of 1.5 mM and 654 RPM. Recordings were then taken using 10 kinetic cycles at 412 nm (average 25 flashes) absorbance at 10 second intervals as the reaction progressed.

2.11.3. Bicinchoninic acid assay

The overall protein concentration of each sample was quantified using the Pierce™ BCA Protein Assay Kit. A working reagent was prepared by mixing reagents A and B 50:1 at appropriate volume for the assay. A series of nine standards were produced from the supplied sealed albumin stock according to the included protocol with concentrations of 2000 µg/mL, 1500 µg/mL, 1000 µg/mL, 750 µg/mL, 500 µg/mL, 250 µg/mL, 125 µg/mL, 25 µg/mL and 0 µg/mL.

A 25 µL aliquot of each of the standard and sample replicates were pipetted into a clear Corning™ polystyrene flat bottomed 96 well plate followed by 200 µL of the working reagent. Samples were then mixed through 5 seconds of orbital shaking at a magnitude of 1.5 mM and 654 RPM in the TECAN. The plate was then covered and incubated at 37°C for 30 minutes. Once returned to room temperature the average absorbance was recorded at 562 nm across 9 points in each well on the TECAN. The A_{562nm} values of the standard were then used to generate a XY graph with polynomial trendline. The equation of this trendline was subsequently used to extrapolate the mg/mL protein concentrations of each unknown sample.

2.11.4. CAT/BCA analysis

The maximum rate of ΔA_{415nm} was recorded for each sample within the CAT assay, and units of CAT activity (U) were calculated per mL according to the following equation:

$$\frac{(\Delta A_{415min} Test) - (\Delta A_{415min} Blank)(0.2)(df)}{(0.0136)(0.01)}$$

In this the 0.2 integer refers to the volume of each reaction (mL), df the dilution factor, 0.0136 the micromolar extinction coefficient for TNB at 412 nm, and 0.01 the volume of lysed sample used (mL). The blank ΔA_{415nm} value refers to a well undergoing the assay steps containing 100 µL of Tris pH 7.8 but lacking any protein extract supernatant. This U/mL activity was then adjusted according to the total mg protein content calculated within the BCA analysis, to generate the final U/mg value indicating CAT activity normalised by sample protein content.

2.12. Bacterial sensitivity assay

A broth-based method was used to ascertain supplement toxicity and resultant minimum inhibitory concentrations associated with key supplements within the Gram-positive research strains. Four cultures of the assayed strain were inoculated and grown at 37^{o}C in the anaerobic cabinet for 16-48 hours, until OD_{600} readings indicated exponential growth phase. A black walled clear F-bottom/chimney 96 well plate (Greiner Bio-One) was set up with 250 µL of routine growth media supplemented with an appropriate range of supplement concentrations, in each well. This was degassed in the cabinet for 24 hours prior to inoculation. One well of each concentration was inoculated from each initial culture, to a final OD_{600} of 0.05. Three identical plates were inoculated in this manner and incubated for 24, 48 and 72 hours respectively. Each microplate was then transferred to a TECAN Infinite M1000 Pro microplate reader, shaken at a magnitude of 1.5 mM and 654 RPM for five seconds before five OD_{600} measurements were taken across each well. These were then averaged to provide a single output for each well and adjusted according to the average OD_{600} reading of four wells containing uninoculated media.

2.13. TraDIS analysis of *E. limosum* NG-6849

2.13.1. Transposon mutagenesis of *E. limosum*

Competent *E. limosum* NG-6849 was electroporated with 5 µL of pMTL-CW21 as detailed in **2.3.6**. The 2 mL total recovery volume was plated across ten Pacaud (20 mM glucose, 2 g/L yeast extract) plates, with transformant colony selection with 6 µg/mL clarithromycin and 7.5 µg/mL thiamphenicol. Following 96 hours of colony outgrowth, cell mass was harvested through the flooding of plates with 1 mL non-supplemented Pacaud media, manual disruption with a sterile wedge spreader and aspiration of the resultant slurry. Serial 500 µL dilutions were made of this slurry up to 10^{-7} in PBS for preliminary transposon system efficacy assessment, and the remaining slurry was stocked using 10% (v/v) DMSO at -80°C.

To generate the mutant library this bacterial slurry was subsequently plated onto a total of 32 Corning® Gosselin™ 120 x 120 mm Petri Dishes (160 µL each). These Pacaud (20 mM glucose) plates were supplemented with 7.5 µg/mL thiamphenicol and 100 ng/mL anhydrotetracycline. These plates were cultivated for 96 hours before a bacterial slurry was once more collected using nonselective liquid media. 10.8 mL of this bacterial slurry was used to generate twelve 10% DMSO stocks, and the remaining 2 mL was used for subsequent genomic DNA isolation and preliminary TraDIS library assessment.

Three of these stocks were subsequently recovered within 50 mL of Pacaud media, lacking cysteine hydrochloride and supplemented with 20 mM glucose, 7.5 µg/mL thiamphenicol, 100 ng/mL anhydrotetracycline and 2 g/L yeast extract. This was grown within a 250 mL serum bottle with 20% CO_2 and 80% N_2 headspace at 0.2 atm, until the OD_{600} reached 0.8. This culture was then used to inoculate six further 250 mL serum bottles, with identical gas conditions and 50 mL of media, to an initial OD_{600} of 0.1. Three of the serum bottles had Pacaud media (lacking cysteine hydrochloride) with 5% v/v methanol (62 mM) as well as the prior thiamphenicol and anhydrotetracycline concentrations. The other three serum bottles also contained Pacaud media

with 20 mM glucose alongside the thiamphenicol and anhydrotetracycline. These were grown at 37°C within a shaking incubator with 225 RPM agitation, and the OD_{600} tracked, passaged into further identically prepared serum bottles once the OD exceeded 0.6. Following the termination of this assay, a repeat was performed in which all conditions were identical expect for the lack of thiamphenicol in all serum bottles following the initial recovery culture. During this time course 10^{-4} and 10^{-5} dilutions (in PBS) were plated from the methanol assay at each passage. These dilutions were plated on Pacaud agar supplemented with 20 mM glucose, 2 g/L yeast extract and 7.5 µg/mL thiamphenicol, with resultant colonies counted to determine the colony forming units (CFUs). This was also performed without dilution on Pacaud agar supplemented with 20 mM glucose, 2 g/L yeast extract and 6 µg/mL clarithromycin and CFUs also quantified. Cell mass was harvested from the previous subculture following passage. This was performed by centrifugation at 7000 RPM for 10 minutes, followed by the removal of the supernatant and transfer of the cell pellets to -20°C.

2.13.2. TraDIS sample preparation

Following the isolation of all samples for TraDIS library preparation, all cell pellets underwent genomic DNA isolation as detailed in **2.4.2**. dsDNA concentrations were calculated for each sample by QUBIT analysis, and nanopore used to ensure all samples were of sufficient quality (as determined by 260/280 and 260/230 ratios). Repetitions and varied pellet volumes were used as necessary to ensure all samples were of sufficient quality prior to proceeding. 130 µL of each sample was then fragmented by sonication using a covaris S-series, with an intensity of 3, 200 cycles, and duty cycle of 20%.

The DNA was then isolated from this using NEBNext® Ultra™ II DNA Library Prep sample purification beads. The resuspension of the bead mix was ensured by 5 seconds of vortexing, prior to the addition of 190 µL to each sample. This sample-bead mix was vortexed for a further 15 seconds and incubated at room

temperature for 10 minutes before being placed on a magnetic stand. Once the beads had settled the supernatant was gently aspirated off, and the beads flushed with 180 µL of 80% ethanol. This was incubated at room temperature for 30 seconds, before the supernatant was removed by gentle aspiration. The beads were then resuspended in 50 µL of nuclease-free water and vortexed for a minute. These samples were then incubated at 40^{o}C for 10 minutes, before being returned to the magnetic rack. Once the beads had settled once more the eluate was carefully transferred to a new labelled tube, and the beads discarded.

100µM HPLC purified SpIAP5_top and SpIAP5_bottom primers were ordered from IDT. Nuclease-free water diluted working stocks of each were then established at 30µM and assembled at a ratio of 1:1 to the total volume required. This mix was then vortexed for 10 seconds at maximum speed, before being placed in a thermocycler. This then heated the primer mix to 95°C for 2 minutes, before dropping 5^{o}C every 5 minutes for 70 minutes, reaching 25°C. The binding of these oligos generated a splinkerette adapter with P7 flow-cell binding capacity and an index primer compatible region, for ligation with the samples prior to Illumina sequencing.

To ligate this splinkerette with the samples 3 µL of NEBNext® Ultra™ II End Prep Enzyme Mix and 7 µL of NEBNext® Ultra™ II End Prep Reaction buffer was added to each sample. This was then thoroughly mixed through 10x aspiration of 50 µL of the sample, before being incubated in a thermocycler at 20^{o}C for 30 minutes (with a lid temperature of 80^{o}C). The block temperature was then raised to 65^{o}C for 30 minutes before being held at 4^{o}C. Samples then immediately underwent adaptor ligation through addition of 30 µL NEBNext® Ultra™ II Ligation Master Mix, 1 µL NEBNext® Ultra™ II Ligation Enhancer, 2.5 µL of the 15µM SpIAP5 adaptor and identical mixing steps. Samples were incubated once more in the thermocycler at 20^{o}C before sample purification was repeated as described above, with 93.5 µL of magnetic sample purification beads and final elution volume of 44 µL.

To ensure no residual plasmid would be amplified in lieu of insertional mutants, each sample underwent restriction digestion with I-SceI. To each previously eluted 44 µl sample was added 1 µL of I-SceI, and 5 µL of CutSmart® Buffer. This sample was then mixed by 10x aspiration of a 40 µL volume, before being incubated at 37°C for 1 hour. Samples then underwent bead purification once more, with an initial volume of 50 µL bead mix (1:1) and a final elution in 25 µL of nuclease-free water.

100 ng of DNA from each sample was then used as a template for PCR amplification using Q5 polymerase. Each sample was amplified with a unique SpIAP5.* primer, and the 5' Tn1 library primer. This in effect barcoded each sample for demultiplexing sequencing data. To this 100 ng of template was added 25 µL of Q5 2x Master Mix, 5 µL of each 10µM primer stock and sufficient nuclease free water to reach a final volume of 50 µL. The PCR program used within the thermocycler is indicated below:

Cycle Step	Temperature (°C)	Duration (s)
Initial Denaturation	95	120
Denaturation	98	20
Annealing	65	30
Elongation	72	30
Final extension	72	120
Hold	10	-

x20-25 cycles

To ensure optimal fragment size for sequencing coverage (ensuring sufficient transposon adjacent sequence is included for position allocation, alongside the read transposon sequence and terminal adaptor) only PCR amplified fragments between 250 bp and 700 bp were isolated for sequencing. This size exclusion was performed by addition of 10 µL 6x loading dye to each sample, and subsequent loading on a 2% SeaPlaque™ low melt agarose (Lonza Bioscience) gel. Gels were imaged using the Odyssey® Fc Imaging System. Comparison to a o'generuler 1 kb plus (NEB) ladder allowed size discrimination of the resultant smears, and manual excision of the appropriate fragment was performed on a blue light box. These gel fragments then underwent gel purification as

previously described, except for a reduced initial incubation at 35°C to limit denaturing of prevalent AT rich regions within Clostridial genomes.

Assessment of the library quality was performed with DNA concentration quantification by QUBIT analysis (**2.4.12.2**). Validation of the library fragment size distribution was performed using the Agilent 2100 Bioanalyzer as described in **2.4.11.2**. Samples were split into two 10 µL aliquots, and shipped on dry ice to DBS Genomics for further work. 100µM stocks of iPCRtagseq index primer and the library primer were also sent.

iPCRtagSeq	AAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTC
Library Primer	AATGATACGGCGACCACCGAGATCTACACttcgactgcactcccgacttaataacttc

2.13.3. TraDIS analysis

Prior to sequencing, one aliquot of each sample was assessed by qPCR with the iPCRtagSeq and library primer. Samples were then normalised and pooled before being run on a single flow cell within the HiSeq 2500 using v2 chemistry in Rapid Run mode. Raw reads were filtered for the presence of the transposon tag at the start of the read (no mismatches allowed), which was subsequently removed. The resulting fastq files were processed for quality and trimmed accordingly using the trim galore package (https://github.com/FelixKrueger/TrimGalore), before being mapped to the *E. limosum* ATCC 8486 genome (NZCP01662.1) (Song & Cho, 2015).

Various scripts were used from the Bio-TraDIS Toolkit for both mapping and essentiality predictions (Barquist, et al., 2016). Reads were mapped to the reference genome using the bacteria_tradis script and SMALT short read mapper. Reads which mapped to multiple locations on the genome were mapped 103andomly between those locations. To produce data on the number of reads and unique insertions for individual genes the tradis_gene_insert_sites script was used and specified to omit reads in the final 10% of the gene. The NZCP01662.1 genome and annotations were again used at this stage. To produce the predicted essential genes the tradis_essentiality.R

script was used. This script generates separate csv files for essential genes, ambiguous genes, and non-essential genes for each set of reads analysed.

Rarefaction analysis was performed by splitting the fastq samples into appropriate fractions at the command line. These split files were then individually mapped as before, permitting the unique insertions sites to be plotted for four different total read numbers. The declining rate of unique insertions mapped permitted estimation of what further data could be attained through extra sequencing runs.

3 Establishment of DNA transfer in *Clostridium carboxidivorans*

3.1. Introduction

DNA transfer is a ubiquitous tool within microbial research which facilitates various research avenues, including forward genetics studies reliant on vector based genetic engineering tools, or product profile enhancement through exogenous gene introduction. When developing promising novel strains establishing a robust DNA transfer method makes such avenues attainable, underpinning a plethora of strain development approaches.

3.1.1. Horizontal gene transfer

Horizontal gene transfer is the exchange of genetic material between contemporary organisms (as opposed to vertical transfer from parent to progeny). This exchange of genetic material affords many benefits and is an important contributor to evolutionary development. While the exogenous DNA can simply provide an alternate source of carbon, phosphorus, and nitrogen it also has the capacity to increase a host's genetic diversity or provide the means for recombinational repair of maladaptive chromosomal changes within a population.

Some prokaryotic species are naturally competent, possessing systems for active uptake of DNA from their local environment. The first recorded observation of bacterial transformation, in 1928, is of such transfer. This research was performed by Fredrick Griffith with one lethal and one non-lethal Group IV *Streptococccus pneumoniae* strain within a mouse model. His experiments noted that a combination of the heat killed virulent strain and a live culture of the avirulent strain would still confer lethality (Griffith, 1928). Since this discovery, research into mechanisms of natural competency has shown that most systems lack specificity and uptake DNA from the extracellular

environment indiscriminately. Natural competency systems characterised to date are also very complex and appear reliant on large numbers of genes. The expression of these genes is in turn governed by external factors such as nutrient availability, population dynamics (quorum sensing) or starvation (Thomas & Nielsen, 2015). Some common apparatus which appear integral to most characterised systems include Type IV pili, Type II secretion systems and DNA translocase enzymes.

The exploitation of natural competency for DNA transfer is understandably attractive, with this approach used as a staple in a few well documented species. Unfortunately, the presence of active uptake systems is not widespread and has only been documented in 80 bacterial species thus far (Dubnau & Blokesch, 2019).

Another classical form of horizontal gene transfer is transduction, in which DNA transfer is mediated by phages. Bacteriophages infect bacterial cells and exploit the endogenous systems to propagate. Transduction requires an active donor cell from which viral particles are released, these may then bind to and infect other susceptible bacterial species. Specialised transfer of non-phage DNA occurs mostly through phage with a lysogenic lifecycle. Within this lifecycle the phage DNA integrates into the host genome and during the subsequent DNA excision and viral particle formation extra genetic material may be acquired. Genes incorporated in this manner are deemed lysogenic conversion genes (Roytka, et al., 2006). Less frequently occurring phage transduction is generalised transfer, in which lytic lifecycle phage package random host DNA within the viral particle in lieu of the viral vector.

3.1.2. Electroporation and conjugation

Electroporation and conjugation are the two most prevalent DNA transfer approaches used for modern research in Gram-positive prokaryotes, including *Clostridia*. Conjugation is a naturally occurring example of horizontal gene transfer mediated through cell-to-cell contact. Electroporation on the other

hand is an example of an artificial transformation approach, with other such approaches including chemotransformation, sonoporation and spheroplast transformation (Aune & Aachman, 2010)).

Establishing DNA transfer in a novel organism *via* either strategy often requires a dedicated study. Predicting a strains' amenability to each approach is particularly hard and reliant on several factors including cell wall composition, cell fragility, presence of extracellular nucleases, susceptibility to electroporation enhancing molecules and presence of active defence systems (Marraffini & Sontheimer, 2008). In fact, the closely related clostridial species *C. autoethanogenum* and *C. ljungdahlii,* which share 100% 16S sequence identity, differ greatly in amenability to DNA transfer; *C. ljungdahlii* transformed by electroporation routinely whereas *C. autoethanogenum* appearing entirely resistant (Liew, 2016).

Between these approaches' electroporation is often regarded more favourably, largely achieving higher efficiencies and independent from donor species compatibility issues (Pyne, et al., 2013). In electroporation the organism, in the presence of the DNA, is exposed to an electrical field which enhances the membrane permeability and increases DNA uptake from the immediate environment. Although the peptidoglycan wall of Gram-positive species can confer a degree of resistance to this approach, extensive parameters can be explored to improve transfer efficiencies. These include introduction of cell wall weakening agents, alternate electroporation buffers, cell harvest at varying growth phases, cell membrane solubilisation or variations in electrical pulse parameters (Bhattacharjee & Sorg, 2020).

To achieve transfer by conjugation pili and a number of other molecular apparatus are required to mediate cell-to-cell contact. The genes encoding these apparatus are normally contained within a fertility factor (F+) in the donor, which is typically the transferred element in naturally occurring conjugation. Understanding of the fundamentals of the process allows this machinery to be hijacked to transfer additional plasmids, as opposed to the factor itself. To achieve conjugative transfer the plasmid must contain an origin

of transfer (*oriT)* at which a nick occurs to produce the single stranded DNA intermediate which is transferred at the interbacterial junction (Fu, et al., 1991). To test a conjugative protocol a compatible donor strain is needed with an F or R conjugative factor and suitable counter-selection method.

3.1.3. Barriers to DNA transfer

Research into the limitations of DNA transfer have highlighted some key active intracellular systems which impede successful transfer. One such impediment is known as surface exclusion which prevents conjugative transfer into a host maintaining another conjugative plasmid. First described more than 60 years ago (Lederberg, et al., 1952), the mechanisms of action are still relatively poorly understood. The most well studied version in *Escherichia coli* is caused by the F plasmid. This plasmid contains two genes of note: *traS* and *traT*. *traS* is believed to cause entry exclusion by encoding an inner membrane protein which interacts with the *traG* of the donor cell to abort the process once mating pairs have been formed, and potentially inhibit conjugal DNA metabolism. *traT* encodes an outer membrane lipoprotein which, through interaction with the peptidoglycan, reduces F pilus receptivity and reduces the stability of mating aggregates (Jalajakumari, et al., 1987). More systems of exclusion are still being discovered and studied, such as the pLS20 *Bacillus subtilis* plasmid encoded Ses_{pLS20} surface exclusion protein (Gago-Córdoba, et al., 2019).

CRISPR-Cas systems are another active barrier to successful DNA transfer, recognising and degrading foreign DNA upon introduction to the cell. These systems have been likened to a primitive prokaryotic adaptive immune system. The eponymous clustered regularly interspaced short palindromic repeats (CRISPR) frame short (typically in the region of 20 bases) target spacer sequences within a CRISPR loci genomic region. Analysis of these sequences show they are acquired from exogenous sources, and therefore target recurrences of similar foreign DNA incursion into the cell. The transcription and processing of these sequences results in CRISPR RNA (crRNA). By formation of

enzyme complexes with CRISPR associated (Cas) proteins, the crRNA target cleavage activity to complementary DNA downstream of a specific protospacer adjacent motif (PAM). CRISPR systems have been found in approximately 50% of sequenced bacterial genomes and nearly 90% of sequenced archaea (Hille, et al., 2018).

Restriction-modification (RM) systems similarly possess nuclease activity which cleaves foreign DNA upon introduction into the host. These systems are prevalent within prokaryotes with 96% of species on record containing at least one (Roberts, et al., 2015). RM systems recognise "non-self" DNA through specific DNA chemical modification patterns; methylation of the N6 position of adenine residues producing N6-methyladenine (6mA) and methylation of the carbon in position 5 of cytosine residues to produce C5-methylcytosine (5mC) (**Figure 2**). This recognition is less stringent than CRISPR/Cas systems, typically with shorter target sequences of between four and eight bases (Wilson & Murray, 1991). Aside from this identification and cleavage of foreign DNA, RM systems can also methylate the host DNA during replication to ensure a characteristic "self" methylation pattern is present.

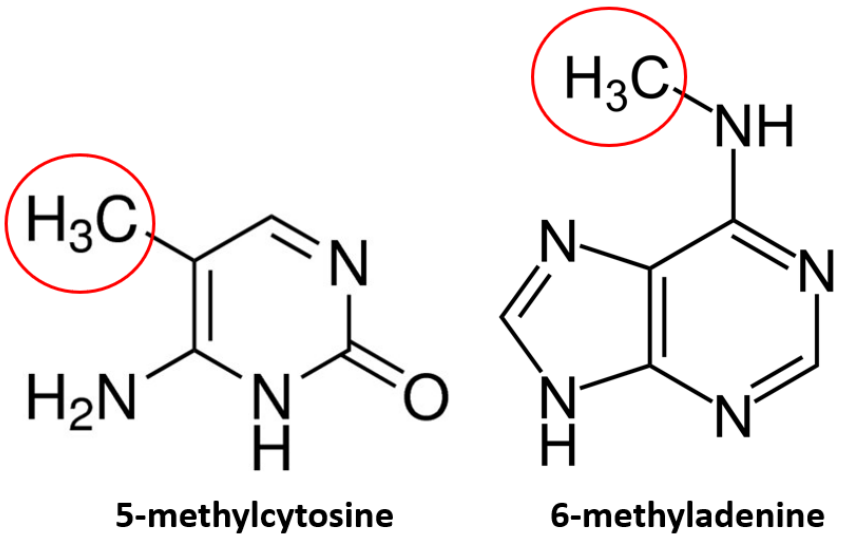


Figure 2 – Structural diagrams of 5-methylcytosine (5mC) and 6-methyladenine (6mA), adapted from Merck *(Merck, n.d.; Merck, n.d.)*. These are two common base modifications created and recognised by RM systems, with the methyl group indicated.

3.1.4. Restriction-modification systems

There are four different types of RM system which have been identified in prokaryotes (**Figure 3**) (Loenen & Raleigh, 2014). The differentiation between these types is based on subunit composition, recognition sequence, modification status of target DNA and proximity of cleavage to this sequence. Broadly Type I, II and III systems target unmethylated DNA, whereas Type IV systems are methylation dependent and target improperly methylated DNA.

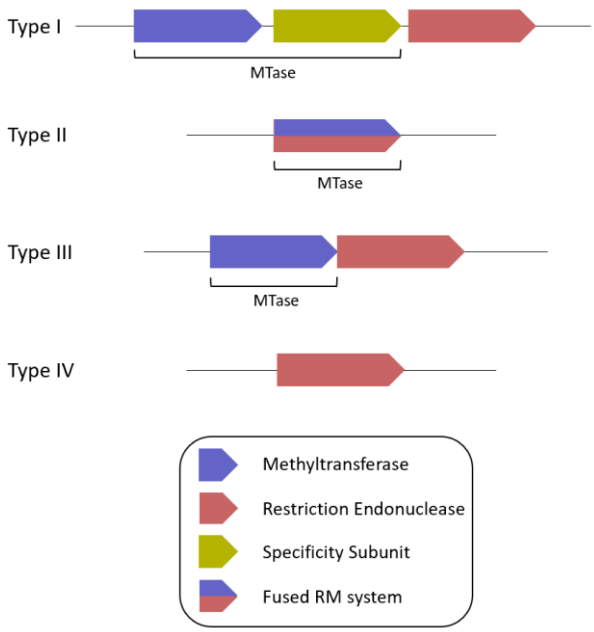


Figure 3 – Schematic of the genomic loci of the four RM system Types
The hetero-oligomeric Type I system consisting of restriction, methyltransferase and specificity subunits; the heterodimeric Type III systems; the monomeric Type II system and Type IV systems. Subunits responsible for methyltransferase (Mtase) activity indicated.

Type I systems form multi-functional, hetero-oligomeric protein complexes where the R, M and S subunits afford distinct functionalities to the protein complex (Wilson & Murray, 1991). The M subunit contains a methyltransferase catalytic domain which selectively methylates newly replicated host DNA. The S subunit dictates sequence specificity, in Tis recognising bipartite target sequences in which two defined ends are separated by several non-specific ('N') nucleotides. The R subunit is a restriction endonuclease comprised of a nucleolytic and ATP hydrolysing translocation domain. This combination of functions cleaves the bound DNA indiscriminately by 'reeling' the downstream

DNA towards the enzyme complex, forming DNA loops, and cleaving as stalls occur due to the presence of other DNA-bound enzymes or secondary structures (Kennaway, et al., 2012).

Type II RM systems are the most prevalent, a broad group classified largely by resultant enzyme activity. There are 11 subgroups within this RM group: A, B, C, E, F, G, H, M, P, S and T (Roberts, et al., 2003). Type II systems are composed of independently acting methyltransferase and restriction endonuclease components. Specificity sequences are largely palindromic or discontinuous and DNA cleavage occurs at a fixed position relative to this sequence.

Type III systems consist of large heterodimeric enzyme complexes. Comparably to Type I systems DNA translocase activity leads to the formation of loops; however, cleavage is not concurrent with stalling. Instead, cleavage of targeted DNA occurs at a defined point, close to the unmethylated recognition sequence (Crampton, et al., 2007). This recognised sequence is characterised by two inversely oriented, non-palindromic sequences.

Type IV systems are in fact the newest distinct class, proposed in 1992 (Janulaitis, et al., 1992). These systems do not possess methyltransferase activity, solely specificity and cleavage. These systems have been recorded as targeting DNA modified by methylation, hydroxymethylation or glucosyl-hydroxymethylation. They are particularly pertinent when using a non-native host for expression as they can exhibit self-targeting behaviour where the host genome is edited and viability reduced (Fukuda, et al., 2008).

3.1.5. Aims

The aim of this study was to establish a robust DNA transfer method for use with the recalcitrant organism *C. carboxidivorans* P7. The proposed method was through identification and characterisation of potential active barriers using genome and methylome analysis. It was hypothesised that the RM systems constitute the most significant barrier, with subsequent validation of

this hypothesis then performed through construction of vectors which evade RM system recognition. Further to this a camouflage analogous system was planned in which recombinant donor strains, expressing the relevant methyltransferase genes of *C. carboxidivorans,* could provide host mimicking methylation coverage which would evade any nucleolytic activity.

3.2. Results

3.2.1. Genomic analysis

At start of this project no instances of DNA transfer into *C. carboxidivorans* P7 had been published. Previous research undertaken by colleagues within the SBRC had corroborated this highlighting a thorough lack of success in traditional approaches to establishing DNA transfer, such as use of plasmid cocktails for electroporation or established *E. coli* conjugal donor strains. As such further elucidation of potential barriers to DNA transfer was pursued through genomic analysis. To identify any CRISPR-Cas associated elements the genome was analysed using CRISPRCasFinder 4.2.20 (Couvin, et al., 2018). The results of this analysis highlighted no CRISPR associated genetic elements within the genome.

3.2.1.1. REBASE® Analysis

For analysis of RM system associated genetic elements the genome was analysed using the Restriction Enzyme DataBASE (REBASE®) by R. Mansfield. This analysis identified ten putative RM systems within the genome: three Type I, three Type II, two Type III and three Type IV (**Figure 4**).

The three Type I systems; CcaP7II (827150-833189), CcaP7I (2727394-2730115 and 2731721-2735029) and CcaP7III (5411396-5416106), all contain restriction endonuclease, methyltransferase and specificity subunits. Of the three Type II systems CcaP7ORF7120P (1646241-1647128) and CcaP7IV (5299217-5300845) both have been identified as fused RM systems with predicted

methyltransferase and restriction activity. The third putative Type II system, *CcaP7ORF16660P* (3757440-3758204), did not have an identified restriction subunit. Both Type III systems, *CcaP7ORF8705P* (2027229-203100) and *CcaP7ORF17830P* (4008355-4003447), contain methyltransferase and restriction endonuclease subunits within the operon. The methyl-directed Type IV systems CcaP7McrB (815415-818568) and CcaP7McrB2 (815415-818568) have two identified subunits, as opposed to the single identified in the CcaP7ORF3495 system (805395-808574).

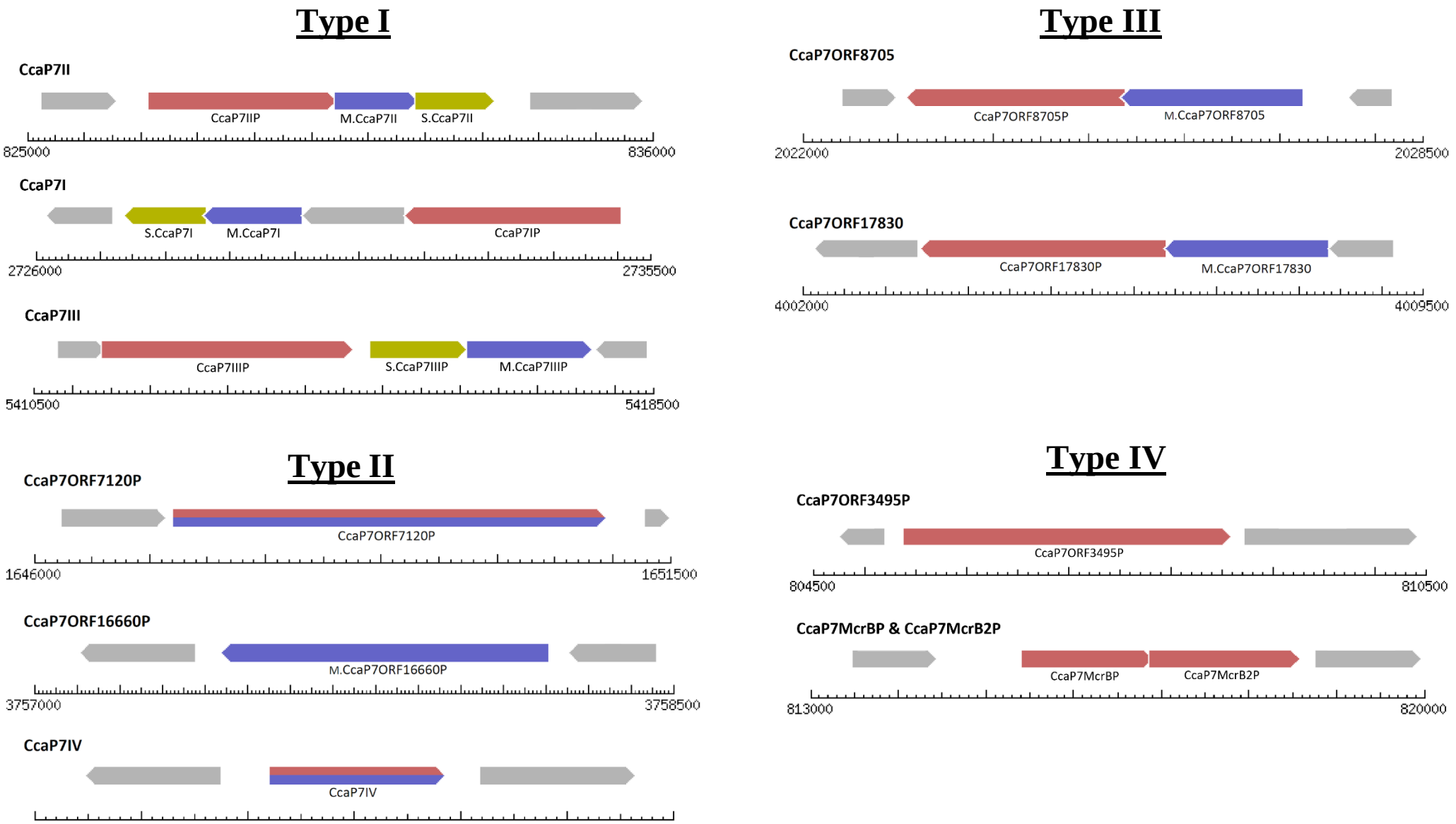


Figure 4 – The restriction modification (RM) systems of *C. carboxidivorans* P7.

A schematic of the genes associated with each of the three Type I, three Type II, two and three Type IV RM systems identified within the *C. carboxidivorans* P7 genome. The red arrows indicate restriction components, blue denote methyltransferases and green for specificity. A red and blue combined arrow is used to depict fused RM systems.

3.2.1.2. PacBio® analysis

Further information was made available through Pacific Biosciences (PacBio®) sequencing data kindly provided by Professor Yang Gu (SIBS, CAS, China). The single-molecule real time (SMRT) sequencing employed within the PacBio® methodology provides epigenetic information pertaining to the modification of genomic DNA bases. Through complete genome analysis recurring patterns of modifications were recognised and the associated sequences and modifications collated in **Table 12**. Six different sequences were confidently identified in which adenosine residues were methylated to N6-methyldeoxyadenosine. No other base modification patterns were recognised.

Table 12 – Recurrent methylated sequences identified within the *C. carboxidivorans* P7 genome.

Modification Sequence (+ strand)	RM System Type	Modification Type	% of occurrences methylated	Coverage	Annotation
C**A**YNNNNNC**T**GC	I	6mA	99.7/99.5	66.4	TI-A
CC**A**NNNNNNNN**T**CG	I	6mA	99.2/98.8	70.9	TI-B
GC**A**NNNNNNN**T**NNCG	I	6mA	99.6/100	69	TI-C
AGA**A**GC	II/III	6mA	97	62.2	TII/III-A
CRAAA**A**R	II/III	6mA	99.7	60.1	TII/III-B
GATA**A**T	II /III	6mA	91.9	60.6	TII/III-C

To allocate these recognition sequences to the corresponding RM systems identified by the REBASE® pipeline, the protein sequences were aligned with the REBASE® Gold Standard Database (BLASTp). This database represents characterized enzymes with experimentally confirmed specificity sequences. Where this yielded no predicted recognition sequence further manual alignments were performed against the complete REBASE® database, permitting further inference but not confident alignment.

All Type I systems were confidently attributed to one of the identified modification sequences. The target recognition domain (TRD1) within the specificity subunit of the RM system CcaP7III has good similarity to that of *S.Spy322I* and *S.Spy8198I*. The recognition sequences of these enzymes are GCANNNNNNRTTG and GCANNNNNNTTAA respectively. This homology distinguishes this RM system as the best candidate for the GCANNNNNNNTNNCG modification sequence. Similar analysis of the TRD1 within *S.CcaP7ORF3570P* shows homology to both *S.BceSVI* (CCANNNNNNNCTTA) and *S.BmeD34I* (CCANNNNNNNCTC), indicating good candidacy for the CCANNNNNNNNTCG modification sequence. The remaining characteristic bipartite Type I recognition sequence CAYNNNNNCTGC was assigned to the Type I system CcaP7I largely through process of elimination. Further manual alignments support this, highlighting homology to *Ces8809IIP* which recognises the sequence CTANNNNNNCTGC.

Of the remaining methylated sequences only one was confidently assigned within the REBASE® pipeline; GATAAT. The allocation of CcaP7IV to this recognition sequence within the REBASE® pipeline is not clearly traceable, however, this may be due to the strong TRD1 homology to *M.CboAf650IV* and *M.CboBDIII* which both recognise GATAAT. The four highest scored protein alignments generated by manual comparison to the Gold Standard database, however, are to *M.CpaYL7II, M.Cdi630V, M.Pdi9689I* and *M.Cdi21062I* all attributed to CAAAAA sequences. This specificity is hypothesised to be conferred by the conserved TaqI-C superfamily structural domain which is not highly conserved in CcaP7IV. Therefore, despite this discrepancy, the strength of TRD1 based alignment scores denotes this as the best candidate for GATAAT.

The best candidate for the AGAAGC modification sequence is CcaP7ORF17830. This TRD1 shows good alignment to *M.XneXORFEP* recognising AGACC as well as *M.XneXORFF* and *M.Sba6781* which recognise AAGACC. The remaining modification sequence, CRAAAAR, could not be confidently assigned. The most similar recognised sequence to CRAAAAR was the CAAAAA sequence recognised by enzymes such as *M.CpaYL7II, M.Cdi630V* and *M.Pdi9689Ie*.

Table 13 – RM systems in the *C. carboxidivorans* P7 genome

RM systems as identified by REBASE® analysis with PACBIO® derived methylation sequences allocated to best candidate systems. *Is used to indicate low confidence assignments.

RM System	Type	ORF Name	Recognition Sequence	Locus Tag (Ccar_)	Size (AA)
CcaP7I	I gamma	*R.CcaP7IP*	CAYNNNNNCTGC	11975	1102
		M.CcaP7IP		11965	498
		S.CcaP7IP		11960	407
CcaP7III	I gamma	*R.CcaP7IIIP*	GCANNNNNNNTNNCG	24300	529
		S.CcaP7IIIP		24305	402
		M.CcaP7IIIP		24310	1069
CcaP7II	I gamma	*R.CcaP7IIP*	CCANNNNNNNNTCG	03565	1090
		M.CcaP7IIP		3570	473
		S.CcaP7IIP		3575	447
CcaP7ORF16660	II	*CcaP7ORF16660P*		16660	254
CcaP7IV	II-G alpha	*CcaP7IVP*	GATAAT	23805	542
CcaP7ORF7120	II-G alpha	*CcaP7ORF7120P*		07120	1241
CcaP7ORF17830	III beta	*CcaP7ORF17830P*	AGAAGC*	17825	980
		M.CcaP7ORF17830P		17830	650
CcaP7ORF8705	III beta	*CcaP7ORF08705P*		08700	752
		M.CcaP7ORF08705P		08705	625
CcaP7McrB2	IV Methyl-directed	*CcaP7McrB2P*		03520	490
CcaP7McrB	IV Methyl-directed	*CcaP7McrBP*		03525	568
CcaP7ORF03495	IV Methyl-directed	*CcaP7ORF03495P*		03495	1059

In manual analysis of the remaining RM systems, the results of BLASTp TRD1 alignments for CcaP7ORF8705 in the REBASE® database yielded low levels matches to *M.SmoLIV, M.Smo310II* and *M.Smo10651ORF176* which recognise the CGWAG sequence. Similar analysis of CcaP7ORF3495 showed no significant homology to any RM system with known sequence specificity. CcaP7ORF7120 has homology to Cch4335ORF15660, and Cch0709ORF9325 which both recognise GNGAAAY sequences. CcaP7ORF16660P shows good homology to CTSAG recognising *M.LlaKF417ORD1059P*, *M.LlaCV56ORF356P* and *M.Csp1100ORF1145P*. Due to a lack of strong homology to similarly specific TRD1 domains of pre-characterised RM systems CRAAAAR remained unassigned.

The three Type IV RM systems could also not be paired with recognition sequences as no strong homology is exhibited by any of the systems to other enzymes with known target sequences. The identification of CcaP7McrB2P and CcaP7McrBP was informed by partial identity with the well-studied McrB cytosine-specific binding component of the McrBC *E. coli* RM system, implying the likelihood of 5-mC recognition. Of note, however, is the third system, CcaP7ORF03495. This has strongest homology within the REBASE® database to *Cau10061ORFEP*. In fact, all three TIV systems have high homology to the five Type IV systems identified in *C. autoethanogenum* 10061, as well as to *C. scatologenes* ATCC 25775 TIV systems.

As such, through comparison to these established databases four of the six identified methylation sequences were successfully assigned to a corresponding RM system (**Table 13**). However, none of the remaining RM systems could be confidently attributed to the CRAAAAAR or AGAAGC sequences. Of the remaining unassigned RM system candidates for these methylation sequences, the highest scored BLASTp alignments to RM systems with confirmed sequence specificity for CcaP7ORF8705 show CGWAG recognition, and for CcaP7ORF7120 show systems which recognise the GNGAAAY sequence. The homology of both systems to these database examples, however, were relatively low and the non-intuitive relationships

between specificity and protein sequence limits the application of this information**.** Analysis of CcaP7ORF16660P highlighted homology to three RM systems which recognise the CTSAG sequence (**Table 14**). Recurrent methylation of this sequence was not detected in the PacBio data. CcaP7ORF16660P may, therefore, be inactive or not constitutively expressed at high levels and hence not at a detectable level at the point of genomic DNA isolation. While BLASTp analysis of the three cytosine-directed Type IV systems, CcaP7ORF03495, CcaP7McrB and CcaP7McrB2, identified comparable systems in *C. autoethanogenum*, none of these enzymes have experimentally confirmed recognition sequences. As such the specificity of these restriction endonucleases, and the barrier they pose is poorly characterised.

Table 14 – Relevant alignment scores for the RM systems of *C. carboxidivorans* P7
TRD1 alignment against REBASE® Gold Standard library (blue), TRD1 alignment against complete REBASE® database (Yellow) and BLASTp alignment of complete enzyme sequence against the Gold Standard library (purple).

RM system	Type	Alignment Target	Score	Identity	E-value
CcaP7I	I gamma	*Ces8809IIP* (CTANNNNNNCTGC)	2155	80	0
CcaP7III	I gamma	*S.Spy8198I* (GCANNNNNNTTAA)	315	36	7.00E-85
		S.Spy360I (GCANNNNNNTTAA)	310	37	2.00E-83
CcaP7II	I gamma	*S.SmaF1I* (CAYNNNNNNTCA)	167	0.99	5.00E-37
CcaP7ORF16660P	II	*M.LlaKF417ORD1059P* (CTSAG)	185	40	5.00E-46
		M.LlaCV56ORF3569P (CTSAG)	179	36	2.00E-44
		M.Csp1100ORF1145P (CTSAG)	178	38	8.00E-44
CcaP7IV	II-G alpha	*M.CboAf650IV (*GATAAT)	413	31	3.00E-114
		M.CboDIII (GATAAT)	413	31	3.00E-114
		M.CpaYL7II (CAAAAA)	179	31	2.00E-49
		M.Cdi360V (CAAAAA)	143	28	2.00E-37
		M.Pdi9689I (CAAAAA)	142	28	2.00E-37
		M.Cdi21062I (CAAAAA)	136	28	3.00E-35
CcaP7ORF7120	II-G alpha	Cch4335ORF15660 (GNGAAAY)	326	43	1.00E-87
		Cch0709ORF9325 (GNGAAAY)	326	42	1.00E-87
CcaP7ORF17830P	III beta	*M.Sba678I*(AAGACC)	631	51	4.00E-138
		M.Cbo1632III (CRACAG)	433	51	2.00E-120
CcaP7ORF8705P	III beta	*M.smoLIV* (CGWAG)	151	16	3.00E-35
		M.Smo310II (CGWAG)	151	16	3.00E-35
		M.Smo10651ORF176 CGWAG)	151	16	3.00E-35
CcaP7McrB2	IV Methyl-directed	*Csc25775McrB3P*	882	45	0
		Cau10061McrB2P	711	50	0
CcaP7McrB	IV Methyl-directed	*Csc25775McrB2P*	1276	94	0
		Cau10061McrBP	1062	83	0
CcaP7ORF03495	IV Methyl-directed	*Cau10061ORFEP*	2435	95	0

3.2.2. Evading RM restriction activity

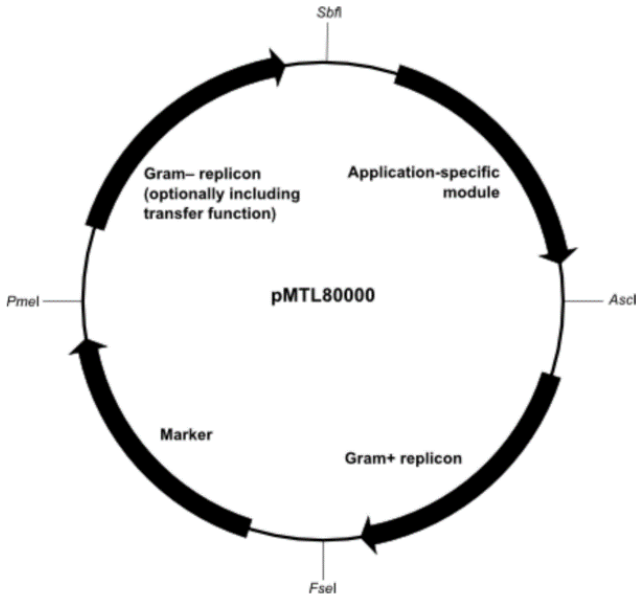


Figure 5 – pMTL80000 shuttle vector structure schematic
Sequencial numbers within pMTL80000 format, following 8, denoting specific Gram-positive replicon modules, selectable markers, Gram-negative replicon modules and application specific modules respectively.

To evaluate the significance of the annotated RM systems as a barrier to DNA transfer in *C. carboxidivorans* P7, a minimal vector was designed in which the recognition sequences identified by PacBio analysis were no longer present. The basis of this vector design was the pMTL80000 *Clostridium* modular shuttle plasmid system (Heap, et al., 2009) (**Figure 5**). These modular vectors contain four distinct interchangeable cassettes: a Gram-negative replicon, Gram-positive replicon, selectable marker, and an application specific module. Each modular sequence was manually analysed identifying the number and type of recognition sequences present (**Table 15**).

Table 15 – *C. carboxidivorans* P7 derived methylation sequence occurrence in the pMTL80000 vector series modules

Gram positive replicon	Count	Selection marker	Count	Gram negative replicon	Count	Application specific module (ASM)	Count
0 — spacer	0	**1 — *catP***	**1**	1 — p15a	2	0 — spacer	0
2- pBP1	8	2 — *ermB*	0	2 — p15a+traJ	6	**1 – multiple cloning site (MCS)**	**5**
3 — pCB102	**3**	3 — *aad9*	8	4 — ColE1	3	2 — pth1 MCS	6
4 — pCD6	13	4 — *tetA(P)*	9	**5 — ColE1+traJ**	**7**	3 — pfdx + MCS	6
5 — pIM13	3					4- CatP reporter	5
6 — pIP404	11						
8 — P19	5						

3.2.2.1. Thiamphenicol sensitivity

To validate *catP* as an appropriate selective marker, *C. carboxidivorans'* sensitivity to thiamphenicol was quantified as detailed in **2.12**. In brief, four exponential *C. carboxidivorans* YTF broth cultures were normalised in regard to OD_{600} values and passaged into 250 µL YTF broth within the wells of a sterile 96 well plate. Each well was supplemented with a range of thiamphenicol concentrations (0-20 µg/mL). OD_{600} readings were taken in technical triplicate at 24, 48 and 72 hours of anaerobic incubation at 37°C. In the absence of a resistance marker, inhibition of *C. carboxidivorans* growth by thiamphenicol occurred at concentrations greater than 2.5 µg/mL (**Figure 6**). This remained consistent across the three timepoints, and as such 7.5 µg/mL was utilised for subsequent plasmid selection within this strain.

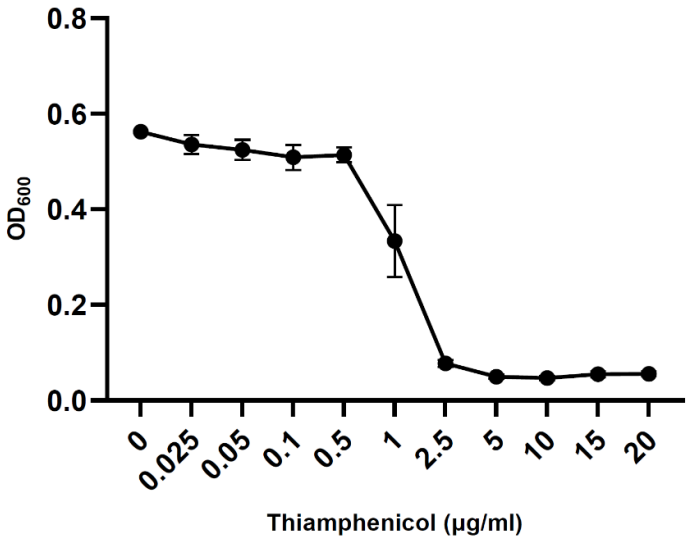


Figure 6 – Analysis of *C. carboxidivorans* P7 sensitivity to thiamphenicol
Sensitivity inferred by growth as indicated by OD600 values achieved following 24 hours of growth in YTF supplemented with 0-20 µg/mL of thiamphenicol. Values represent mean of biological quadruplicate and technical triplicate data points, and error bars reflecting standard deviation.

3.2.2.2. pMTL8315y-v* construction

The modular components suitable for a *C. carboxidivorans* compatible vector were selected to minimise RM system recognised motif inclusion while maintaining desired functionality (**Table 15**). The chosen plasmid pMTL83151 was reliant on a *catP* selection for maintenance and contained 16 RM recognition sequences: two TI-A, two TI-C, three TII/III-A, seven TII/III-B and two TII/III-C.

To remove the RM system recognition sequences within pMTL83151 two single nucleotide changes were made in the Gram-negative replicon (ColE1+*traJ*) and the Gram-positive replicon cassette (pCB102), and two truncations of the selection marker module (CatP) and the multiple cloning site (MCS). These assemblies were generated by PCR amplification of fragments from pMTL83151, and subsequent four-part HiFi assembly. The resultant pMTL8315y-v1 vector, with a total of 12 RM recognition sequences removed, was confirmed by sequencing (**Figure 7**).

A further cloning step was necessary to remove the three TII/III-B and one TII/III-A recognition sequence still present in pMTL8315y-v1. These four sequences were located within the functional coding, and non-coding, structures of the ColE1 + tra module. This module consists of a Gram-negative replicon with associated origin of replication (OriR) and conjugative essential *traJ* gene with associated origin of transfer (OriT). Assembly was performed by PCR amplification of DNA fragments from pMTL8315y-v1 and subsequent HiFi with final confirmation by sequencing as detailed in **2.9.1**. pMTL8315-v2 and pMTL8315y-v3 differ in the single nucleotide change made to disrupt the TII/III-B site within the ColE1 RNA II pre-primer: the 16th base thymine replaced with adenine in v2, or the 14th base adenine replaced with a thymine.

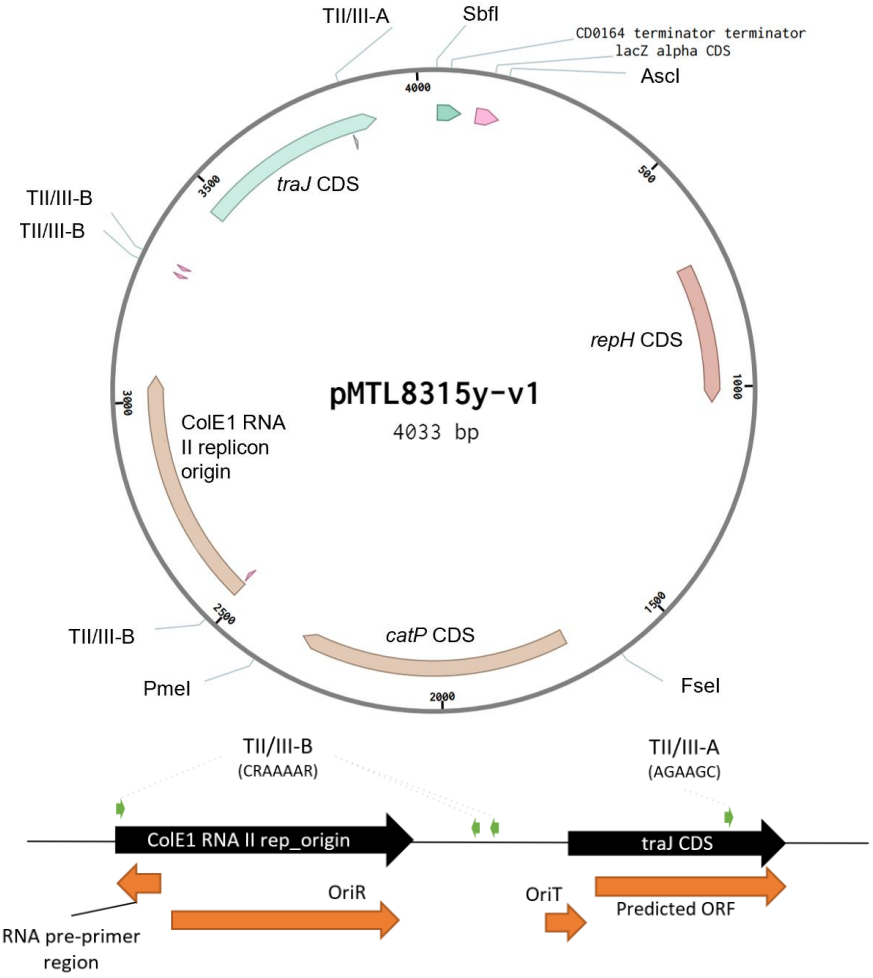


Figure 7 – pMTL8315y-v1 vector map

ColE1 + *traJ* module schematic indicating locations and orientation of remaining recognition sequences in respect to predicted functional elements: TII/III-B-1 10 bp from the end of the RNA pre-primer sequence, TII/III-B-2 and 3 140 bp and 243 bp upstream of the *traJ* region respectively, and TII/III-A 56 bp from the end of *traJ* ORF.

3.2.2.3. pMTL8315-v1 transfer into *C. carboxidivorans*

Prior to confirmed pMTL8315y-v2 and pMTL8315y-v3 assembly a preliminary conjugal DNA transfer attempt was made using pMTL8315y-v1 and *E. coli* sExpress as the conjugal donor. This initial conjugation protocol was largely based on the high efficiency protocol reported for *C. autoethanogenum* 10061 (Woods, et al., 2019). sExpress *E. coli* chemically competent cells were transformed with pMTL83151 and pMTL8315y-v1 and successful transformants selected for on LB broth with 25 µg/mL chloramphenicol and 50 µg/mL kanamycin. These transformants were then grown for 16 hours (overnight) in comparably selective LB broth. Concurrently three *C. carboxidivorans* YTF cultures were grown anaerobically at 37°C for 16 hours (overnight), the first inoculated from an exponential phase culture and the further two additional dilutions of this culture at 10^{-1} and 10^{-2}. The *E. coli* donor cultures were inoculated into fresh selective LB media approximately 3 hours prior to conjugation at a 1:100 dilution. These were grown at 37°C with 225 RPM agitation until an OD_{600} of 0.3 was reached indicative of early exponential phase growth. Both cultures were normalised to an equivalent OD_{600} of 0.3 and pelleted by centrifugation at 3000 x *g*. The pellets were then washed by resuspension in 1 mL PBS before being pelleted once more and transferred to anaerobic conditions to be combined with 200 µL of one of the *C. carboxidivorans* cultures which had reached early exponential phase (OD_{600} of 0.355). This mixed culture was then spread onto non-selective YTF plates and incubated for 24 hours, before bacterial slurry was collected by flooding the plate with 1 mL PBS and manually dislodging cell mass with a sterile wedge spreader. All recovered bacterial slurry from the two samples was plated across three YTF plates each with 7.5 µg/mL thiamphenicol and 250 µg/mL D-cycloserine.

Following 72 hours, 14 *C. carboxidivorans* transconjugant colonies were observed across the pMLT8315y-v1 selective plates, but none across those for pMLT83151. Confirmation of successful transfer of pMLT8315y-v1 was performed by colony PCR, using primer pair pMTL8315y Scr Fw/Rv (**Figure 8**).

This amplified a 580 bp region from the pMTL8315y-v1 vector. pMTL8315y-v1 purified plasmid DNA and *C. carboxidivorans* gDNA also underwent PCR amplification as positive and negative controls. All transconjugants yielded an appropriately sized DNA fragment when analysed by gel electrophoresis.

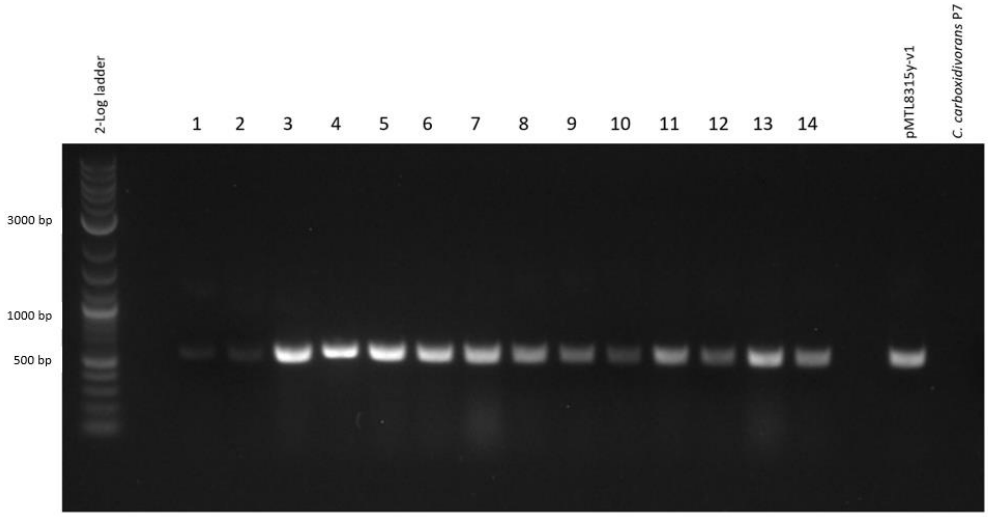


Figure 8 – Colony PCR of pMTL8315y-v1 within *C. carboxidivorans* P7
Colony PCR with pMTL8315y Scr Fw/Rv primers and a DreamTaq based PCR protocol. Expected PCR product size was 580 bp and is consistent with that amplified from 0.5 µL of purified pMTL8315y-v1 plasmid DNA. No band was produced from comparable amplification from *C. carboxidivorans* gDNA.

Strain identity was also confirmed by colony PCR using Ccarb Mtase Fw/Rv TIII.2 primers to amplify a 1878 bp fragment directly from the *C. carboxidivorans* genome (**Figure 9**). The same cell mass template was also used to amplify 16S ribosomal RNA using the primers 27F and 1492R (Miller, et al., 2013). These PCR products were sequenced and aligned to the BLAST database to confirm species identity. The 16s rRNA of *C. carboxidivorans* afforded the highest alignment identity scores for all 14 sequences (>99%). Five of these confirmed transconjugants were also inoculated into 5 mL YTF broth supplemented with 7.5 µg/mL thiamphenicol and 250 µg/mL D-cycloserine and grown for 48 hours prior to cell pelleting at 17,000 x *g* and subsequent plasmid isolation. 10 µL of this purified plasmid DNA was then successfully introduced into chemically competent *E. coli* Top10, and isolated once more prior to complete vector sequencing. This confirmation of successful DNA transfer into *C. carboxidivorans,* using only a minimally evasive vector, was a promising result

which appeared to corroborate the assumption that the RM systems pose the biggest barrier to establishing DNA transfer within this species.

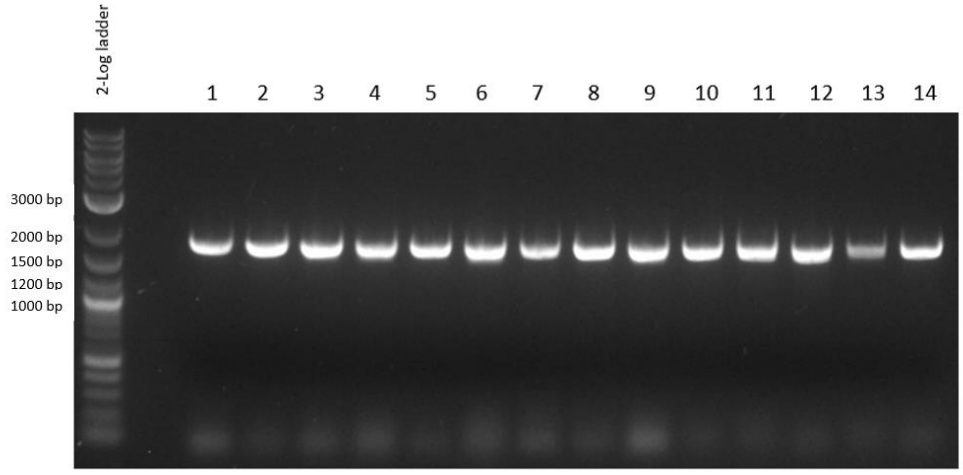


Figure 9 – Colony PCR of TIII.2 genomic region within pMTL8315y-v1 *C. carboxidivorans* P7 transconjugants

Colony PCR with Ccarb Mtase Fw TIII.2 and Ccarb Mtase Rv TIII.2 primers and a DreamTaq based PCR protocol, amplifying a 1878 bp fragment from the genome of *C. carboxidivorans* to confirm the species identity of the 14 transconjugants.

3.2.2.4. pMTL8315y-v1/2/3 conjugative transfer to *C. sporogenes*

To confirm conjugative and replicative functionality was not markedly impeded by the changes made within the Gram-negative replicon of pMLT8315y-v2 and v3, all three generated plasmids were conjugated into *C. sporogenes* NCIMB 1069 as detailed in **2.3.7**. In brief, each vector was used to transform *E. coli* conjugal donor strain sExpress, and then cultured aerobically in chloramphenicol and kanamycin selective LB broth overnight. 1 mL of each donor culture, normalised to the lowest OD_{600}, was then harvested by centrifugation, washed with 1 mL of PBS and combined with 200 µL of exponential phase *C. sporogenes* culture grown anaerobically in TYG broth. The donor and recipient were incubated together on TYG plates without selection for 24 hours, and a bacterial slurry was collected by flooding the plates with 1 mL of PBS and using a wedge-shaped spreader to gently dislodge and re-suspend the bacterial growth. The aspirated slurry was then transferred to selective TYG agar plates supplemented with thiamphenicol, trimethoprim,

and D-cycloserine, at various dilutions in PBS to ensure single CFU observation was possible. Transconjugant colonies were counted within 72 hours of incubation and standardised according to remaining slurry volume (as calculated from weight), to total transconjugant yield of the individual conjugation.

All conjugations were performed in biological triplicate where *E. coli* replicates were biologically independent at the subculture prior to harvesting but were all mated with a single *C. sporogenes* recipient culture. Thiamphenicol resistant (Tm^R) transconjugants were successfully generated by all vectors, confirming the retained functionality of the vectors in conjugative transfer to a comparable Gram-positive strain to the *C. carboxidivorans* target. Notably the efficacy of conjugative transfer for pMTL8315y-v2 is approximately 4-fold higher than that of pMTL8315y-v3, and 3-fold higher than the unmodified vector pMTL83151 (P values of 0.05 and 0.02 respectively) (**Figure 10**). The pMTL8315y-v2 transconjugant *C. sporogenes* colonies also demonstrated rapid growth and were visible following just 18 hours of selective growth. The similar conjugal transfer efficacy of pMTL8315y-v3 and pMTL83151 indicates the silent mutation within the *traJ* operon, and the two SNPs made between the ColE1 RNA II and *traJ* regions, had no notable impact. The 8.5-fold higher transconjugant number observed for conjugal transfer of pMTL8315y-v2, as opposed to pMTL8315y-v3 (0.007), may be attributed to unforeseen impacts of the sequence change in ColE1 on plasmid replication.

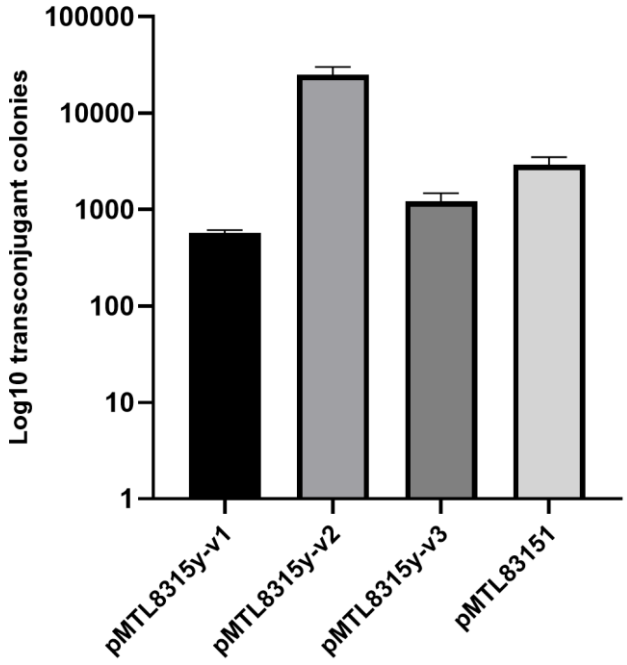


Figure 10 – Conjugation efficiency of pMTL8315y-v* vectors as compared to pMTL83151 in *C. sporogenes*
Conjugation efficiency determined via transconjugant number yielded by *E. coli* sExpress conjugative transfer into *C. sporogenes*. Mean transconjugant number per conjugation was determined from biological triplicate values with error bars reflective of standard deviation.

3.2.2.5. pMTL8315y-v2/3 in *C. carboxidivorans*

Following confirmation of the maintained conjugation functionality of the pMTL8315y-v2/3 vectors, conjugation was then attempted into *C. carboxidivorans*. The sExpress donors maintaining pMLT8315y-v1, pMTL8315y-v2 and pMTL8315y-v3 previously utilised for conjugation of *C. sporogenes* were employed as donors. The conjugation method was identical to that described in **3.2.2.3**, with early exponential phase donor and recipient. Following 72 hours of selective growth on thiamphenicol and D-cycloserine YTF plates no transconjugants were observed for any of the three tested vectors. This conjugation was repeatedly attempted, however no transconjugant CFUs were observed. Various parameters were subsequently explored as routes to improving transfer rates, including varying the growth phase of the donor, the growth phase of the recipient, the ratio of donor to recipient, and changing the donor strain. None of these experiments yielded any confirmed

transconjugants, generating concerns regarding the reproducibility of the initially observed DNA transfer.

Alternatively, DNA transfer was attempted using electroporation with 5 µg of pMTL8315y-v1. Competent cells were produced in the same manner as for *Eubacterium limosum* (**2.3.5**), using YTF as opposed to Pacaud medium. All centrifugation steps were performed at 4000 x *g*. Electroporation steps and pulse parameters were also consistent with the transformation of *E. limosum* (**2.3.6**) excluding an extended eight-hour non-selective recovery period prior to selective plating on YTF and thiamphenicol. Five voltages were tested: 1, 1.25, 1.5, 1.75 and 2 kV. No transformants were observed for any of the tested parameters following 144 hours of anaerobic incubation at 37^{o}C. This disappointing result led to a renewed focus on optimising the conjugative DNA transfer approach.

3.2.3. Conjugation of *C. carboxidivorans* with *E. coli* CA434

Following confirmation of successful transfer of pMTL8315y-v1 and subsequent troubleshooting, a paper featuring DNA transfer into *C. carboxidivorans* P7 was published (Cheng, et al., 2019). DNA transfer was reportedly achieved by conjugation of *E. coli* CA434 donor cells with pMTL82151. Attempts were made to replicate this; CA434 donors maintaining pMTL83151, pMTL82151 and pMTL8315y-v1 were cultivated aerobically overnight in LB broth with 25 µg/mL chloramphenicol and 50 µg/mL kanamycin. 2 mL of donor cell growth was collected by centrifugation and washed twice with 1 mL PBS. These donor cells were mixed with 0.5 mL *C. carboxidivorans* cells pre-cultured overnight and plated on non-selective YTF agar. Following incubation in an anaerobic chamber at 37°C for 18 hours plates were washed with 1 mL of PBS and all slurry was transferred to YTFAM agar plates with 7.5 µg/mL thiamphenicol and 250 µg/mL D-cycloserine for selection. No transconjugants were observed by any vector following 7 days.

This was an unforeseen result, and hence work pivoted to focus once more on replicating previous successes achieved within this study.

3.2.4. Isolation of *C. carboxidivorans* variant C94

The lack of reproducible conjugative transfer was a significant challenge. Following repeated failure to replicate the initial transfer of pMTL8315-v1 it was hypothesised that a more amenable mutant sub-population was greater represented in the original culture used. It is understood that mutant sub-populations can be maintained in wild-type cultures under routine growth conditions should they not introduce deleterious effects. In order to evaluate this hypothesis, attempts were made to isolate and test this more amenable variant mutant through plasmid loss and re-transformation of previously stocked *C. carboxidivorans* transconjugants.

To isolate potentially advantageous variants three pMTL8315y-v1 *C. carboxidivorans* transconjugant stocks were recovered in YTF broth with thiamphenicol and D-cycloserine selection. Unfortunately, only one of these stocks was successfully recovered. Once this culture had reached mid-exponential phase (>0.8 OD_{600}) it was passaged into non-selective YTF broth and grown for 24 hours before a series of dilutions were made in PBS and plated onto YTF agar. Following a further 48 hours of growth on plates single CFUs were observed on the 10^{-4} dilution plates. 100 colonies from this plate were then re-streaked onto both thiamphenicol containing and non-selective plates. All 100 colonies grew on both plates, demonstrating continued plasmid retention through Tm^R. The 100 colonies from the non-selective plate were transferred to fresh replica plates at 48 hours, and the process repeated. This occurred a total of three times before one colony demonstrated restored thiamphenicol sensitivity. This colony was picked from the non-selective plate into thiamphenicol selective and non-selective broth and grown for 24 hours. Following confirmation of no growth in the thiamphenicol selective YTF broth, the non-selective YTF culture was used to generate six 10% (v/v) DMSO stocks.

A colony PCR was also performed with the primers pMTL8315y Scr Fw and pMTL8315y Scr Rv.

To test the conjugal receptivity of this C94 isolate the four vectors, pMTL8315y-v1, pMTL8315y-v2, pMTL8315y-v3 and pMTL83151 were transferred alongside a vector-less control donor strain. Conjugations were performed as previously described in **3.2.2.3** with *E. coli* sExpress donor cultures and early exponential cultures of both *C. carboxidivorans* P7 and *C. carboxidivorans* C94. Transconjugants were only generated by the conjugation of pMTL8315y-v2 into *C. carboxidivorans* C94. The average number of transconjugants calculated from biological triplicate is 282.3, with a standard deviation of 63 (**Table 16**).

Table 16 — Conjugal transfer efficiency comparison between *C. carboxidivorans* P7 and *C. carboxidivorans* C94.
The total transconjugant number per conjugation was extrapolated from the highest countable dilution plate to the total volume of bacterial slurry recovery from mating plates.

	Total transconjugant colonies					
	***C. carboxidivorans* P7**			**C. *carboxidivorans* C94**		
pMTL8315y-v1	0	0	0	0	0	0
pMTL8315y-v2	0	0	0	254.83	369.54	222.46
pMTL8315y-v3	0	0	0	0	0	0
pMTL83151	0	0	0	0	0	0
No vector	0	0	0	0	0	0

To characterise this more receptive variant, and identify key genetic changes the genome of C94 was subjected to Illumina sequencing using the Illumina MiSeq platform at MicrobesNG (**2.5.1**). The *C. carboxidivorans* P7 genome (Li, et al., 2015) served as the reference sequence for read mapping, with the corresponding statistics of this read mapping is collated in

Table 17. More than 99.9% of reads were successfully mapped with the quality of mapped reads also good with an average read length exceeding 235 bp, median trimmed read size of 621 bp and resultant coverage of 65.58x.

Table 17 — Summary report of *C. carboxidivorans* C94 Illumina read-mapping
C94 isolate Illumina data alignment against reference genome and plasmid of *C. carboxidivorans* P7 (Li, et al., 2015).

Reads	Read Count	Reads (%)	Average length	Bases count	Bases (%)
Mapped	1,806,331	99.87	237.75	429,461,747	99.93
Not mapped	2,375	0.13	126.32	300,013	0.07

These read mappings were used to call Single Nucleotide Polymorphisms (SNPs), insertions and deletions (InDels) and small structural variants. This data is compiled in **Table 18**. Those SNPs with low frequency, insufficient coverage or within an untranslated region were disregarded. Of the remaining mutations, two SNPs were identified as introducing premature stop codons in lieu of glutamic acid residue. These SNPs occur in Ccar_25335, where the translated protein chain length is reduced from 847 residues to 15, and Ccar_04190, which shortens the translated region from 434 amino acids to 228. All other SNPs resulted in single amino acid changes. InDel analysis also highlighted a 52 bp deletion within the Ccar_07510 locus, at 384His. This deletion also generates a frame shift in which a premature stop codon occurs shortening chain length from the full 383 to 208 residues. This gene is annotated as a DEAD/DEAH box helicase. Without further research, this cohort of changes could not be further refined to identify the best candidate genomic changes which resulted in this receptive phenotype.

Table 18 — Pertinent SNPs called from read mapping of the C94 isolate against the published *C. carboxidivorans* P7 genome *(Li, et al., 2015)*.

Locus tag	Position	Base Change		Coverage	Frequency	Average Quality	Predicted Amino Acid Change	Gene Function
Ccar_25335	5629616	G	T	73	100	36.32	Glu15X	calcium-translocating P-type ATPase, PMCA-type
Ccar_04190	970007	G	T	56	100	35.64	Glu229X	DUF2339 domain containing protein, TIGR02679
Ccar_01050	205543	C	A	87	100	35.3	Ala378Asp	MFS transporter
Ccar_10985	2513827	C	A	55	100	34.96	Glu396Cys	cell wall-binding repeat-containing protein
Ccar_17250	3878707	C	A	96	100	36.21	Ser432Ile	murein biosynthesis integral membrane protein MurJ
Ccar_20540	4566888	A	C	87	100	35.41	Cys146Giu, Ile175Arg	enoyl-[acyl-carrier-protein] reductase FabK
Ccar_20565	4582631	C	A	72	100	33.11	Asp305Tyr	phosphoenolpyruvate carboxykinase
Ccar_00825	160189	G	T	71	97.18	34.68	Gly200Val	GTP-sensing pleiotropic transcriptional regulator CodY
Ccar_25825	5726044	G	T	51	98.1	34.8	Ala188Ser	acetylglutamate kinase
Ccar_01705	367599	G	T	96	100	35.44	Ala173Ser	hypothetical protein

3.2.5. Further transfer optimisation

The 4-fold increase in transfer rates afforded by conjugation of isolate C94 with pMTL8315y-v2 proved repeatedly reproducible. In light of this some transfer approaches and conjugative parameters were revisited using the C94 isolate and pMTL8315y-v2 to better quantify any improvements afforded.

3.2.5.1. Electroporation

Although electroporation had previously been tested with *C. carboxidivorans* P7 and pMTL8315y-v1, further attempts were made using *C. carboxidivorans* C94 and pMTL8315y-v2, as well as cell wall weakening agents to improve the receptivity of the Gram-positive bacteria. Electroporation was revisited due to the general preference for this approach due to its broadly higher transfer rates, independence from conjugal donors and the potential for use of a plasmid devoid of conjugation associated features such as origins of transfer. An exponential phase *C. Carboxidivorans* C94 culture was passaged to an OD_{600} of 0.1 into 20 mL of YTF broth containing 1% glycine and 0.2M sucrose. Once an OD_{600} of approximately 0.7 was reached, the culture was incubated on ice and all subsequent steps were performed at 4^{o}C. Cell mass was pelleted by centrifugation at 4000 x *g* and then washed twice with 12.5 mL of SMP before resuspension to a total volume of 800 µL per competent aliquot, corresponding to 20 mL of harvested culture. A comparative competent cell preparation was also performed without glycine and sucrose supplementation in the media.

Each competent cell aliquot was incubated for 15 minutes with 4 µg of each respective plasmid, isolated from the *E. coli* strain NEBExpress. Transferred to either a 2 mM or 4 mM cuvette they were then exposed to an electric pulse of 200 Ω, 50 µF and either 1.5, 1.75 or 2 kV. They were then immediately suspended in 4 mL of YTF (with 0.2M sucrose as continued osmoprotection for those exposed to glycine) and recovered for 12 hours prior to selective plating. No transformants were observed for any condition following 144 hours, once more returning focus to optimising conjugal DNA transfer strategies.

3.2.5.2. Alternate *E. coli* donor strains

Different *E. coli* donor strains display their own inherent methylation backgrounds, due to the expression of native methylase genes. This methyltransferase activity will impact the methylation profile of conjugated DNA, and therefore may also affect the efficacy of this conjugation. The three most common methylases in such strains are encoded by the *dam* gene (Dam methylase) which methylates the adenine residues in the sequence GATC, the *dcm* gene (Dcm methylase) which methylates the second cytosine residues in the sequences CCAGG and CCTGG, and the *M.EcoKI* gene (EcoKI methylase) which modifies the adenine residues in the sequences AAC(N6)GTGC and GCAC(N6)GTT. The higher specificity of the EcoKI sites means they occur much less frequently than the other two, occurring only twice in pMTL83151 as opposed to the nine and ten occurrences of the Dam and Dcm sequences respectively. As such Dcm and Dam methylation should be primarily considered in conjugal donor strain amenability.

To test the impact of these variations in background methylation profile three *E. coli* donor strains were tested: InterStellar, TopSex and sExpress. These represent the alternative background methylation profiles Dcm^-/Dam^-, Dcm^+/Dam^- and Dcm^-/Dam^+, respectively. TopSex was a novel conjugal donor strain generated by transfer of the conjugal plasmid pR702 into chemically competent *E. coli* TOP10. pR702 was isolated from 5 mL of turbid *E. coli* sExpress culture supplemented with kanamycin, using the Monarch® Miniprep kit with a primed GenElute™ column. Successful transfer of the pR702 was confirmed through kanamycin and spectinomycin resistance.

Chemically competent aliquots were generated for each donor, prior to transformation with pMTL8315y-v1, pMTL8315y-v2 and pMTL83151. Triplicate colonies from each vector and donor combination were picked into LB broth with chloramphenicol and kanamycin selection. These were passaged as before and harvested once the OD_{600} had reached 0.3. Conjugation was then attempted into a single *C. carboxidivorans* C94 recipient culture of OD_{600} 0.845

as previously described, with final selection on YTF plates supplemented with thiamphenicol and D-cycloserine. Only the sExpress donor (Dcm- Dam+) yielded transconjugant colonies with pMLT8315y-v2 following 96 hours of growth. Hence sExpress was utilised as the base donor strain in further research steps.

3.2.6. Mimicking native methylation profile

3.2.6.1. Overview

An alternate approach explored in overcoming the barrier presented by the extensive RM systems of *C. carboxidivorans* was to mimic the native methylation fingerprint within the conjugative donor strain. To achieve this strain specific methyltransferase genes and specificity subunits needed to be stably introduced and expressed within the donor prior to conjugation. At conjugation a plasmid transferred by this modified strain would then display the same methylation profile as the recipients DNA and no longer be recognised and targeted as non-self DNA and degraded.

The strategy employed within this study was exemplified within multiple recalcitrant *Clostridium* strains within the SBRC (Mansfield, 2018) and replicates the necessary methyltransferase genes of the target organism within the *P. mirabilis* pKJ439039 derived conjugation assistance plasmid pR702. This strategy eliminates the potential problems associated with conjugative transfer from a donor strain harbouring three plasmids: pR702, a methyltransferase expressing vector and the plasmid of interest. To create these donors, initial assembly of methyltransferase genes and their respective ribosomal binding sites (RBS) was performed within a smaller base vector under inducible promoter control and framed by pR702 homology arms for subsequent cloning. The integration of these methylation cassettes into the pR702 occurred *in vivo* within a specialised *E. coli* strain expressing the λ-red recombineering machinery genes *exo*, *beta* and *gam*. To select for successful recombination the target region for insertion within pR702 was the spectinomycin resistance

gene, whose disruption confers spectinomycin sensitivity to the cell. Further to this a gene conferring resistance to apramycin was introduced as another selectable marker.

3.2.6.2. pRBV1_T* vector generation

Each of the eight methyltransferase genes, their associated RBSs and the specificity subunits, where applicable, were amplified by PCR from the genomic DNA of *C. carboxidivoran*s P7. These were then individually cloned into the recombineering base vector (pRBV1) by golden gate assembly by Georgina Bates as described in **2.9.2**. This generated eight plasmids each reflecting the methylation activity of a singular RM system: TI.1, TI.2, TI.3, TII.1, TII.2, TII.3, TIII.1 and TIII.2. The assembled plasmid names, the introduced genes and their respective resultant vector sizes are listed within **Table 19**. The locus within pRBV1 at which these genes were introduced is downstream of the L-arabinose inducible promoter P_{BAD} to prevent the potential detrimental toxic interactions of non-native methylation with the native RM systems of the *E. coli* Top10 cloning host. The genes were also flanked by an apramycin resistance gene and framed with 500 bp sequences homologous to the adjacent regions of the *aadA1* gene of pR702. This region is termed the methylation cassette of the vector.

Alongside individual methyltransferase expressing pRBV1 vectors, three multi-RM vectors were generated in a similar fashion, as detailed in **2.9.2**. These alternate vectors contain multiple methyltransferase/specificity genes, grouped by RM system type. pRBV1_TI contains the three Type I methyltransferase/specificity genes, pRBV1_TII the three Type II and pRBV1_TIII the two Type III (**Figure 11**). The methylation cassette sequence integrity of all 11 pRBV1_T* vectors was confirmed by sequencing.

Table 19 – Gene inclusion within pRBV1 vectors for expression of *C. carboxidivorans* P7 derived methyltransferase/specificity activity.

Plasmid	Genes	Size(bp)
pRBV1	-	5626
pRBV1_TI.1	*M.CcaP7IP* *S.CcaP7IP*	8348
pRBV1_TI.2	*M.CcaP7IIIP* *S.CcaP7IIIP*	8458
pRBV1_TI.3	*M.CcaP7IIP* *S.CcaP7IIP*	8391
pRBV1_TII.1	*CcaP7ORF16660P*	6391
pRBV1_TII.2	*CcaP7IVP*	7255
pRBV1_TII.3	*CcaP7ORF7120P*	9352
pRBV1_TIII.1	*M.CcaP7ORF17830P*	8569
pRBV1_TIII.2	*M.CcaP7ORF8705P*	7498
pRBV1_TI	*M.CcaP7IP* *S.CcaP7IP* *M.CcaP7IIIP* *S.CcaP7IIIP* *M.CcaP7IIP* *S.CcaP7IIP*	13963
pRBV1_TII	*CcaP7ORF16660P* *CcaP7IVP* *CcaP7ORF7120P*	10673
pRBV1_TIII	*M.CcaP7ORF17830P* *M.CcaP7ORF8705P*	104441

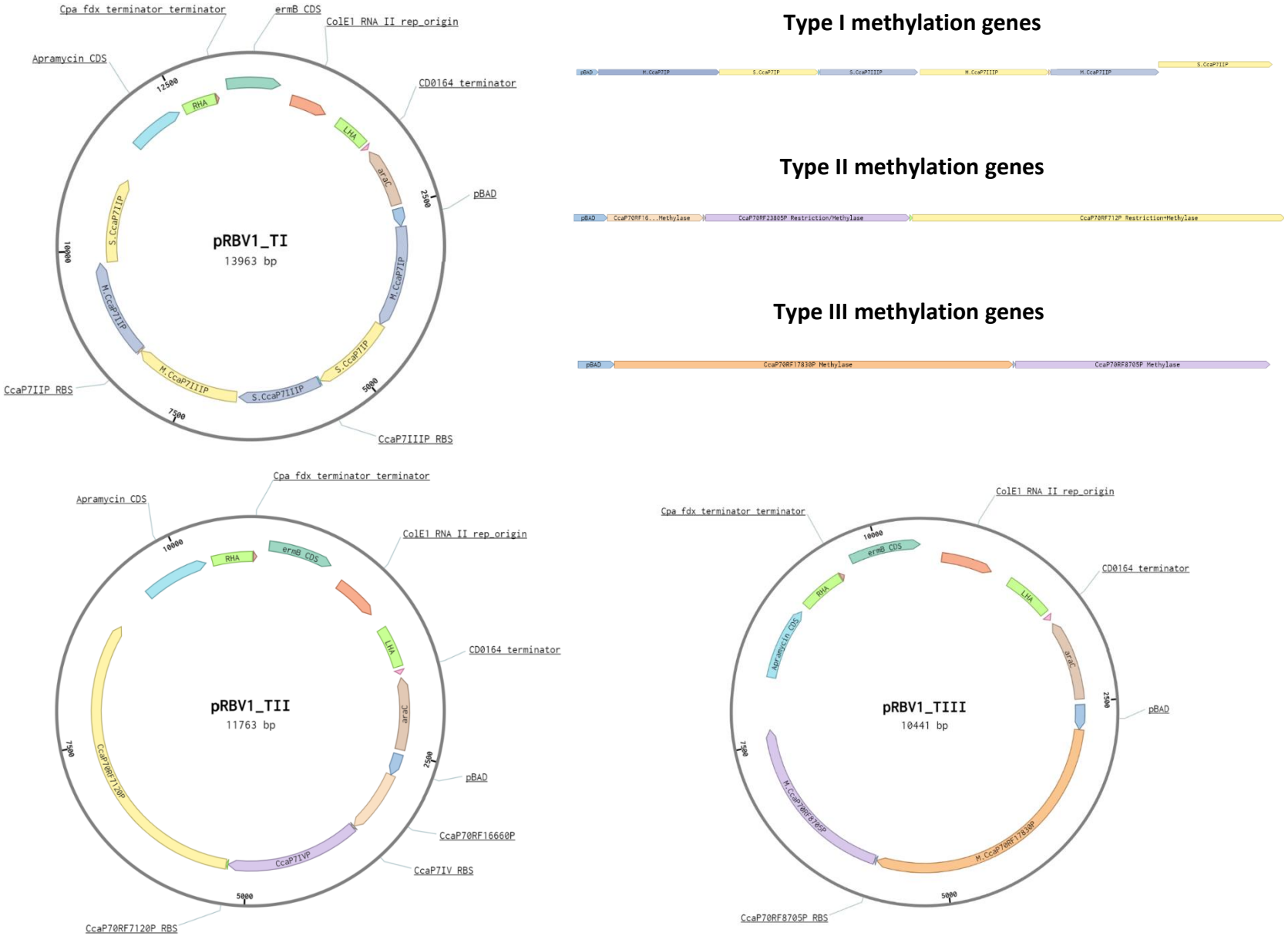


Figure 11 - Plasmid maps of pRBV1_TI, TII and TIII

Initial base vectors assembling multiple methyltransferase genes into a single cassette prior to PCR amplification.

3.2.6.3. Generation of sExpress_TII and sExpress_TIII

Integration of the multi-RM system methylation cassettes into pR702 occurred in InterStellar, a novel strain generated by James Millard (unpublished). InterStellar was created by the conjugal transfer of the R factor from CA434 into a Stellar strain derived from ClonTech™. This commercial strain, and therefore the derived InterStellar, proved an ideal assembly chassis due to the *recA*+, *dam*- and *dcm*- phenotype. The lack of *dam* and *dcm* methyltransferases improves its amenability to RM system cloning, as well as the DNA repair protein RecA enhancing rates of recombination events. Into this strain pKD46cm was also transformed to produce ReInterStellar.

The process of recombinant pR702 generation is detailed in **2.10.** In brief linear DNA fragments were generated from pRBV1_TI, pRBV1_TII and pRBV1_TIII vectors by PCR amplification. These linear fragments were used to transform electrocompetent ReInterStellar where the 500 bp homology arms framing the sequence acted as a template for homologous recombination at the *aadA1* locus of the pR702. Apramycin, and not spectinomycin, was used at this stage to select for inclusion of the *AAC(3)-Iva* encoded apramycin resistance conferred by methylation cassette integration.

Final confirmation of cassette integration was performed by colony PCR with an internal primer specific to the inserts and a reverse primer which bound to the adjacent pR702 region. Three colonies for each recombinant pR702, deemed positive by PCR screening, were immediately isolated and used to transform chemically competent NEB Express *E. coli*. This generated nine novel sExpress donors; sExpress_TIA/B/C, sExpress_TIIA/B/C and sExpress_TIIIA/B/C. After repeated attempts to reach sufficient sequencing coverage to confirm the fidelity of the integrated cassettes using PCR amplification and subsequent Sanger sequencing, all nine donors were instead subjected to Illumina sequencing with DeepSeq. sExpress_TIIA and sExpress_TIIIB exhibited sufficient sequence fidelity and hence were chosen for downstream testing. These are referred to as sExpress_TII and sExpress_TIII henceforth.

3.2.6.4. Recombinant sExpress conjugal donor assessment

The conjugal efficacy of the newly generated sExpress_TII and sExpress_TIII was evaluated in comparison with sExpress using pMTL83151, pTML8315y-v1 and pMTL8315y-v2. Conjugations were then performed in biological triplicate into the C94 isolate with each donor strain and vector as detailed in **2.3.9**. Induction of the methyltransferase cassettes introduced into pR702 was performed with 10 mM L-arabinose at all donor growth stages, including the 24 hours on YTAFM mating plates with *C. carboxidivorans* C94.

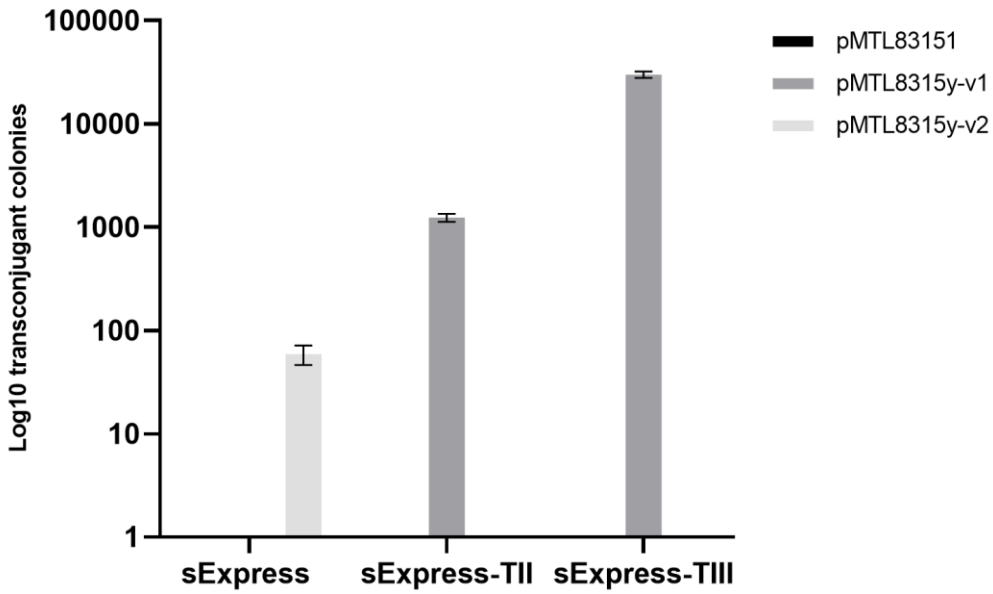


Figure 12 – Comparison of the DNA transfer rates into *C. carboxidivorans* C94 when using sExpress, sExpress_TII and sExpress_TIII as *E. coli* conjugal donors
Data represents average transconjugant number calculated from biological triplicate with the vectors pMTL83151, pMTL8315y-v1 and pMTL8315y-v2. based on the observed selective plate CFU, extrapolated according to dilution and total mating plate bacterial slurry volume. Error bars represent standard deviation.

Conjugation efficiencies were measured by extrapolation of CFUs on dilution plates to the total volume of bacterial slurry collected from the mating plate. Notably while conjugative transfer of pMLT8315y-v1 was not achieved by any sExpress replicate, no transfer was achieved in turn by either recombinant donor with the pMTL8315y-v2 vector (**Figure 12**). Comparing conjugative efficacy between these two vectors, sExpress_TII and sExpress_TIII show a 4-

fold and 9-fold improvement over sExpress respectively (P values of 0.0073 and 0.0052).

Considering the scale of DNA transfer efficiency increase afforded by these recombinant donors into *C. carboxidivorans* C94, conjugal transfer into *C. carboxidivorans* P7 was also attempted with the recombinant donors as described above. The results of this conjugation are presented in **Table 20**. No transconjugants were generated by conjugation with sExpress, but an average transconjugant CFU/conjugation of 11.52 was observed for sExpress_TII and 15.12 for sExpress_TIII. This was an encouraging result, however the lower transconjugant yield compared with C94 using only a 4 kB plasmid may indicate any further drop in transfer rates (i.e., due to larger sequence size or expressed cargo) may once more negate conjugative capacity into the wild-type strain. As such it was important to generate a donor strain with more than three methyltransferase genes expressed, which could afford the combined benefits of both donors used so far.

Table 20 — Conjugation efficacy of sExpress, sExpress_TII and sExpress_TIII with *C. carboxidivorans* P7 using pMTL83151, pMTL8315y-v1 and pMTL8315y-v2, as determined by transconjugant CFU/conjugation.

	Total transconjugant colonies								
	sExpress			sExpress_TII			sExpress_TIII		
pMTL83151	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
pMTL8315y-v1	0.00	0.00	0.00	12.31	8.42	13.83	22.40	9.41	13.56
pMTL8315y-v2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

3.2.6.5. Generation of sExpress_TI/TII/III

For complete, accurate mimicking of the native methylation of *C. carboxidivorans* the aim was to generate a recombinant pR702 in which all eight Type I, Type II and Type III RM system associated methyltransferase/specificity genes could be expressed. Following failure to generate a TI pR702 with sufficient sequence integrity, efforts began with assembly of the five Type II and Type III RM system methyltransferase associated genes into a single cassette: pRBV1_TII/III. The assembly of this

vector is detailed in **2.9.3**. PCR amplification of the methylation cassette from this vector was attempted as previously described using Recomb LHA Fw and Recomb RHA Rv. Repeated attempts to amplify the cassette were entirely unsuccessful. This lack of success was largely attributed to both the extended length of the fragment (~15 kb) and low GC content (29.7%).

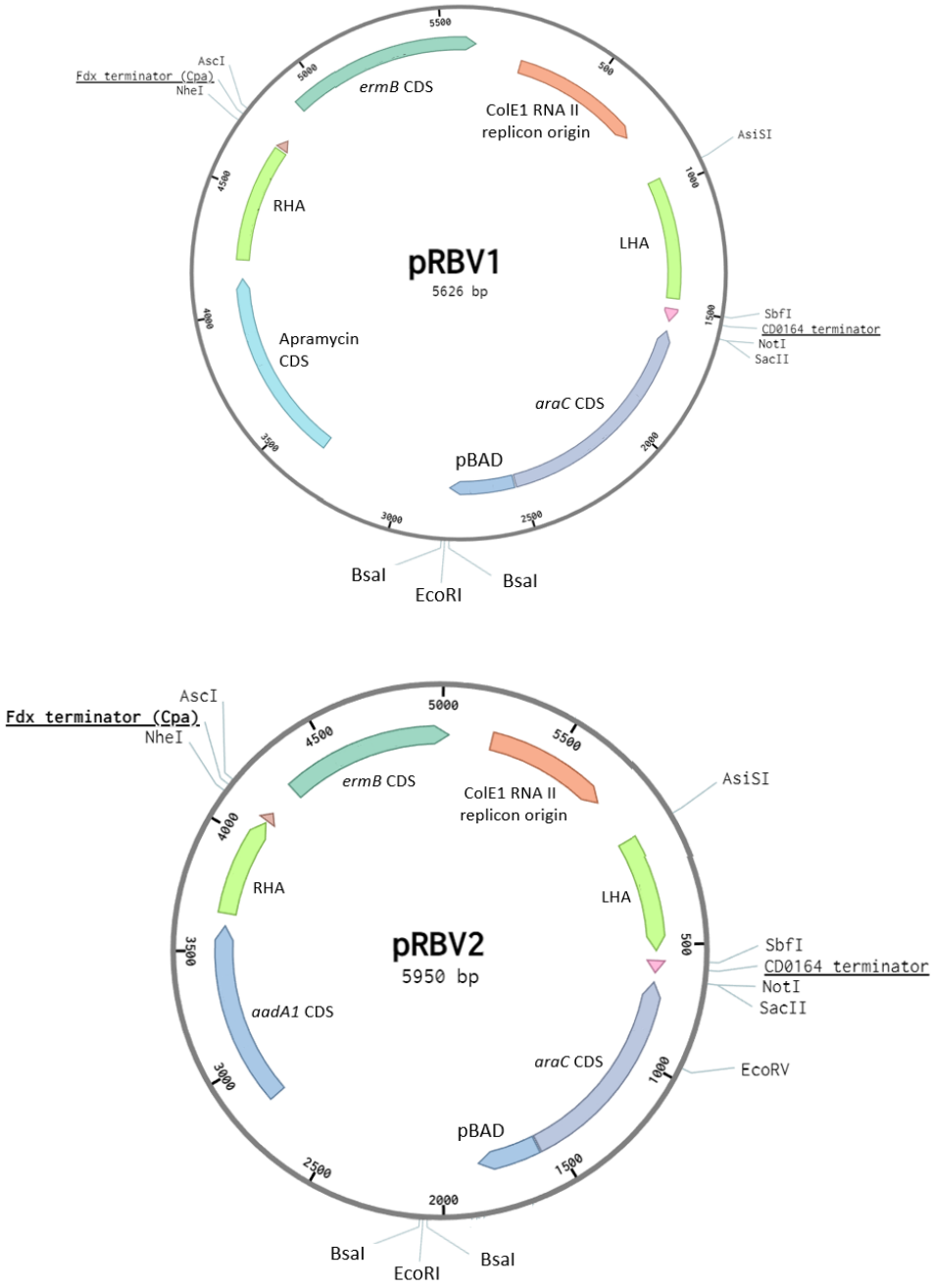


Figure 13 – pRBv1 and pRBV2 plasmid maps

Schematics indicating key components and differences between pRBV1 (kindly provided by Rob Mansfield) and pRBV2 (this study). The key difference indicated in the replacement of the

Apramycin CDS with the *aadA1* CDS to change the antibiotic selection used to select for successful pR702 integrant sequences from apramycin to spectinomycin.

An alternate approach was devised for iterative integration of additional methyltransferase genes directly into the recombinant pR702 *in vivo*. To facilitate this approach the apramycin resistance gene of the methylation cassette in pRBV1 was replaced with *aadA1*. This assembly generated the vector pRBV2 (**Figure 13**). The *aadA1* gene is present in the wild-type pR702 sequence but was knocked-out by initial fragment integration and recombinant pR702 generation. By targeting the sequence downstream of the methyltransferase genes in pR702_TII, the previously introduced apramycin resistance gene would be replaced with this *aadA1* gene once more. Hence this change of selective marker facilitated a similar screening approach to before with new recombinant pR702s demonstrating sensitivity to apramycin, indicative of pR702 target site disruption, and resistance to spectinomycin, indicative of cassette introduction. Next the two TII RM system methyltransferase genes were integrated into this vector, as well as the LHA being replaced with the terminal 944 bp of the final methyltransferase gene of the Type II cassettes (*CcaP7ORF7120P),* as detailed in **2.9.4**, to generate pRBV2_L2-TIII (**Figure 14**).

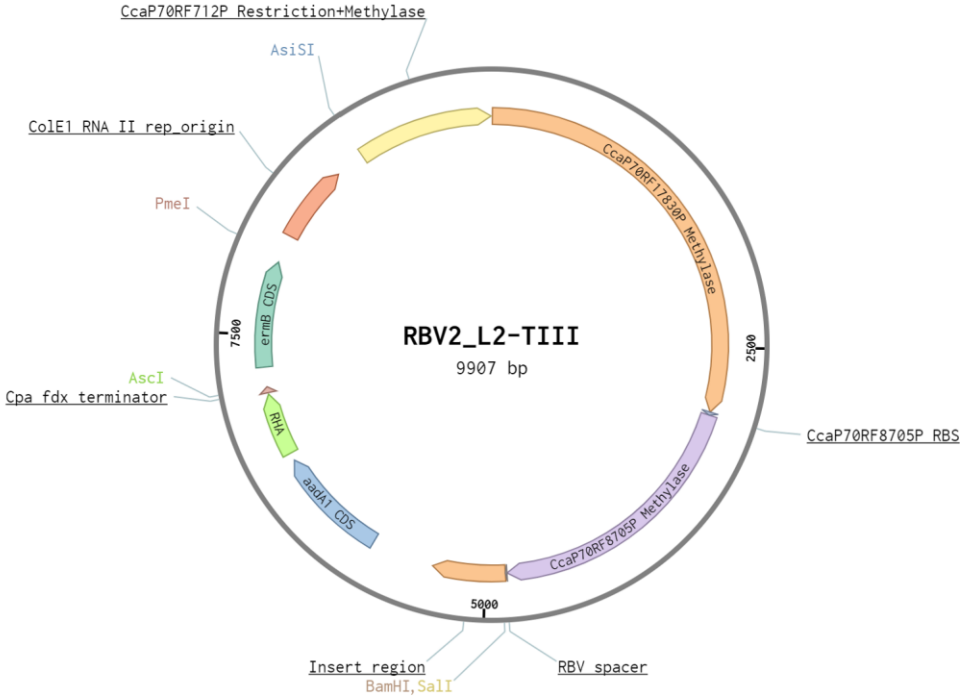


Figure 14 — pRBV2_L2-TIII plasmid map

Vector map generated through assembly of new pR702_TII compatible LHA, *aadA1* selectable marker and *CcaP7ORF17830P/CcaP7ORF8705P* TIII RM system associated methyltransferase genes into the recombineering base vector pRBV1.

For *in vivo* recombination the confirmed pR702_TII was isolated from sExpress_TIIA and used to transform chemically competent Stellar. This new 2InterStellar was then also transformed with the λ-red recombineering machinery encoding pKD46 to generate Re2InterStellar. A linear fragment, generated by PCR amplification from RBV2_L2-TIII with R2-L2-T3 Fw and Recomb RHA Rv, was used to transform electrocompetent Re2InterStellar primed by L-arabinose induction. No transformants were observed following 72 hours of selective growth on spectinomycin and kanamycin. This amplicon transformation process was repeated two further times, however, neither yielded any successful integrant colonies. As such, the recombinant donor rationale was unfortunately not further pursued beyond sExpress_TII and sExpress_TIII within this study.

3.2.7. Hyper-receptive conjugative recipients

Isolation of the *C. carboxidivorans* variant C94 afforded significant improvements in DNA transfer rates, however due to the number of SNPs identified several candidate genes remained. To further elucidate the responsible genotypic change between *C. carboxidivorans* P7 and C94 the isolation of more variants was pursued. Initially this variant isolation was attempted with sExpress and pMTL8315y-v1, pMTL8315y-v2 and pMTL8315y-v3 as the only available tools unfortunately, as previously outlined, no further transconjugants were produced. Following development of the sExpress_TII and sExpress_TIII recombinant *E. coli* donors, transfer into *C. carboxidivorans* P7 was once more attained (**3.2.6.4**). The same variant isolation approach was then applied to three transconjugant colonies derived from pMTL8315y-v1 transfer with sExpress_TIII. Plasmid loss was confirmed through restored thiamphenicol sensitivity and colony PCR, before the three new isolates underwent conjugation with sExpress_TIII and pMTL8315y-v1 to assess the receptivity of the new isolates to conjugal DNA transfer. Transfer was concurrently performed with *C. carboxidivorans* C94 and the same donors as a positive control.

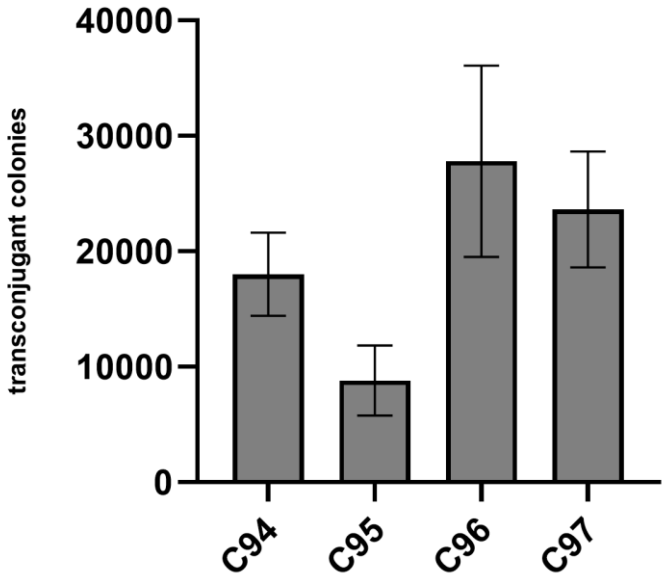


Figure 15 – Conjugal transfer efficiency comparison between *C. carboxidivorans* C94, C95, C96 and C97

Data represents average transconjugant number calculated from biological triplicate through extrapolation of selective plate CFU, extrapolated according to the dilution plated and total

bacterial slurry volume recovered for the replicate from the respective mating plate. Error bars represent standard deviation.

C95 had the lowest rate of transfer with a total average transconjugant CFU per conjugation of 8.8 x 10^3, less than C94 at 1.8 x 10^4. All three isolates demonstrated a comparable increase in DNA transfer efficacy to that observed in C94, with average transconjugant CFUs per conjugation of 2.8 x10^4 for C96 and 2.4 x 10^4 for C97 (**Figure 15**). This was confirmed by statistical analyses using Welch's t-test, which indicated that, compared to C94, the new C95, C96 and C97 isolates did not demonstrate statistically significant changes in resultant transconjugant colonies (at 0.125, 0.365 and 0.420 respectively). These isolates and the initial *C. carboxidivorans* P7 strain from which they were generated underwent Illumina sequencing as described in **2.5.2.** The mapping statistics of each sample's reads against the published genome (Li, et al., 2015) are presented in **Table 21**, all with >99.9% of reads successfully mapped and an average trimmed read length exceeding 285 bp for all samples.

Table 21 — Summary reports of Illumina read-mapping of *C. carboxidivorans* C95, C96, C97 and plasmid cured *C. carboxidivorans* P7 against the wild-type P7 reference genome *(Li, et al., 2015)*.

Reads	Read Count	Reads (%)	Average length	Base count	Bases (%)
			C. carboxidivorans P7 (parent)		
Mapped	3257175	99.94	294.62	959,635,366	99.95
Not mapped	2059	0.06	224.75	462,758	0.05
			C. carboxidivorans C95		
Mapped	2,981,483	99.88	292.33	871,577,932	99.9
Not mapped	3461	0.12	240.44	832,178	0.1
			C. carboxidivorans C96		
Mapped	3,453,195	99.92	287.06	991,286,373	99.94
Not mapped	2865	0.08	223.48	640,283	0.06
			C. carboxidivorans C97		
Mapped	3,991,676	99.93	293.32	1,170,834,415	99.94
Not mapped	2960	0.07	226.85	671,467	0.06

The Illumina data was mapped against the published genome (Li, et al., 2015) and the subsequent mapping used to call single nucleotide polymorphisms (SNPs), insertions, deletions, and small structural variants within CLC Genomic Workbench 20 as before. These SNPs were compared with those of the

parental strain, and all that were consistent with this strain disregarded. The remaining SNPs with a frequency exceeding 90% within the reads were manually aligned and compared. C95, C96 and C97 had seven, eight and nine SNPs above this threshold which were not congruent with the parent. Eight of the nine SNPs identified in C97 were identical to those identified in C96, with C95 showing significantly more divergence. Of these SNPs one, identical across all three samples, introduced a single nucleotide which changed the methionine start codon of a GTP-sensing pleiotropic transcriptional regulator (CodY) to the alternative start codon ATT with a resultant isoleucine residue. One further comparable SNP across all three samples occurred within an AAA family ATPase. While C96 and C97 had an identical change in this region, the SNP in C95 generated a change 10 amino acid residues upstream. Through comparison to those SNPs identified in C94 one further variant region was implicated. This region was within the Ccar_20540 locus. Within C94 this locus contained a SNP, with a comparably located but different resultant SNP identified in C96 and C97. Within C95 a 72 bp region was identified with multiple SNPs, including insertions and deletions. This generated multiple predicted amino acid residue changes, including a premature stop codon introduction at the 159th position, of an otherwise 315 residue proteins. These are summarised in **Table 22**.

Further analysis, however, revealed that two of the SNPs observed in C95-97 also occurred within the parental strain bacterial population. This was not previously highlighted as these SNPs occurred within less than 90% of the reads mapped, with the SNP observed in Ccar_00825 in 55.7% of 158 reads, and the SNP within Ccar_25120 in 50.86% of 116 reads. Analysis of the sequence of Ccar_20450 within the parental data revealed no such mixed population. Further to this a comparable SNP was identified within the Ccar_20450 sequence of the original C94 variant, increasing the likelihood that this is the genotypic change which resulted in the improved DNA transfer receptivity phenotype. Ccar_20450 is denoted as *fabK*, a gene encoding the enoyl-acyl carrier protein (ACP) reductase.

Table 22 – Sequence variations between key open reading frames of *C. carboxidivorans* hyper-receptive variants and the *C. carboxidivorans* P7 genome.

Locus Tag	Function	Position	Isolate	Base Change		Mutation	Coverage	Frequency	Average Quality
Ccar_00825	GTP-sensing pleiotropic transcriptional regulator CodY	159593	C95	G	T	Met1Ile	160	100	36
		159593	C96	G	T	Met1Ile	142	100	35.57
		159593	C97	G	T	Met1Ile	166	98.19	36.31
Ccar_25120	AAA family ATPase	5583825	C95	G	T	Arg194 Ile	82	100	36.16
		5583854	C96	C	T	Pro204Ser	114	100	34.86
		5583854	C97	C	T	Pro204Ser	132	99.24	33.92
Ccar_20540	enoyl-[acyl-carrier-protein] reductase FabK	4566888	C94	C	G	Gly146Arg	87	100	35.41
		4566936>>4567007	C95	ACAACTTGTGGAACTAACGCCATAGTTGTAAGTTCACATACATGGCCTCCTGACTCACATCCTTCTGCTATA		ACACAGCTCTCTCCATCTTTTGCCTAATGCTAAGTGCAACTACATGGCCTCCTGCTGCAATTACTGGTATGCTT			
		4566973	C96	C	A	Gly146Cys	175	100	35.88
		4566973	C97	C	A	Gly146Cys	162	99.38	36.08

3.2.8. Further RM system elucidation

3.2.8.1. pMTL8315y-v1-TII/IIIABC transfer efficacy

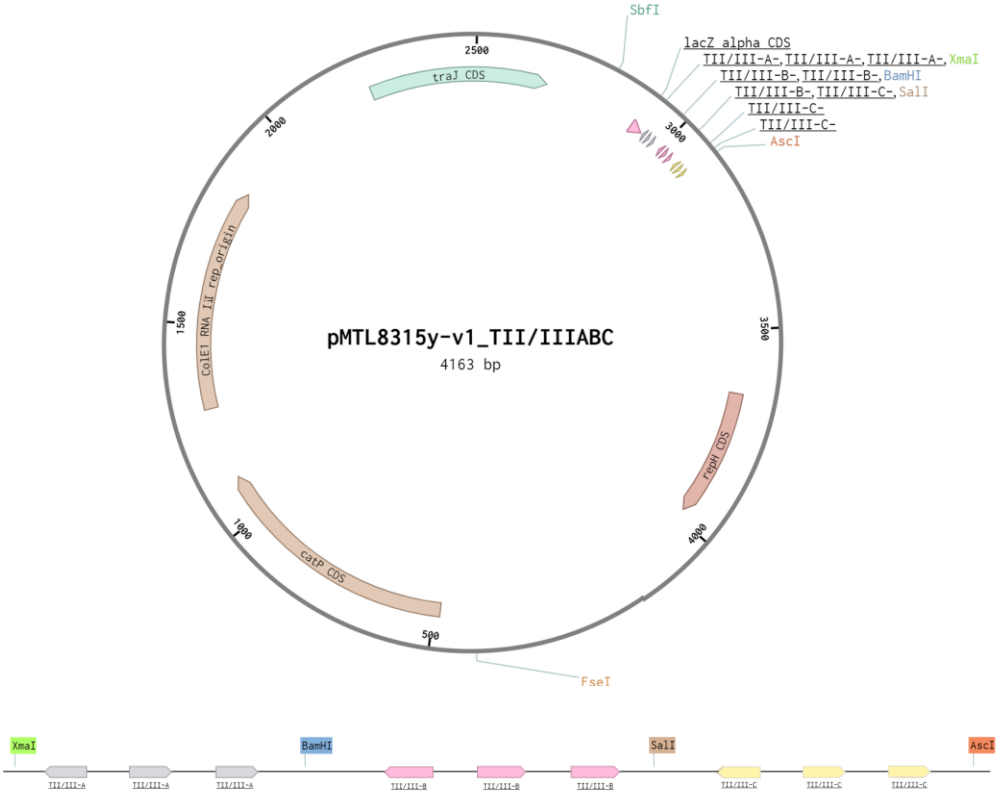


Figure 16 — pMTL8315y-v1-TII/IIIABC plasmid map

Schematic of pMTL8315y-v1-TII/IIIABC with pMTL8315-v1 derived backbone features including ColE1+tra, CatP and pCB102 modules. Also annotated are the additional nine RM motif sequences introduced within the MCS between the XmaI, BamHI, SalI and AscI restriction digestion sites.

To characterise the efficacy of the donors in relation to specific RM motifs a new series of vectors were constructed. A synthetic sequence was designed in which each TII/III RM motif sequence was introduced in triplicate (**Figure 16**). This was integrated into the MCS of pMTL8315y-v1 in four different manners to generate pMTL8315y-v1-TII/IIIABC, pMTL8315y-v1-TII/IIIA, pMTL8315y-v1-TII/IIIB and pMTL8315y-v1-TII/IIIC as described in **2.9.5**.

Conjugation of these vectors was performed in triplicate with sExpress_TII, sExpress_TIII and sExpress as *E. coli* conjugal donors. All conjugations were

performed into a singular exponential phase *C. carboxidivorans* C94 culture. These results are presented in **Figure 17**. The three vectors in which only one type of RM motif was introduced in triplicate (containing a total of seven TII/III sites including the residual ones within the pMTL8315y-v1 backbone), successfully transferred through all donors. pMTL8315y-v1 transfer was consistent with what had previously been observed, achieving transfer only through the recombinant donors. The pMTL8315y-v1-TII/IIIABC vector containing a total of thirteen RM motif sequences was only successfully transferred into *C. carboxidivorans* C94 using the two recombinant donors. While these results were as expected, the successful transfer of the seven RM motif containing vectors *via* the non-recombinant sExpress donor raised further questions regarding the relationship between RM motifs, methylation and conjugative transfer receptivity.

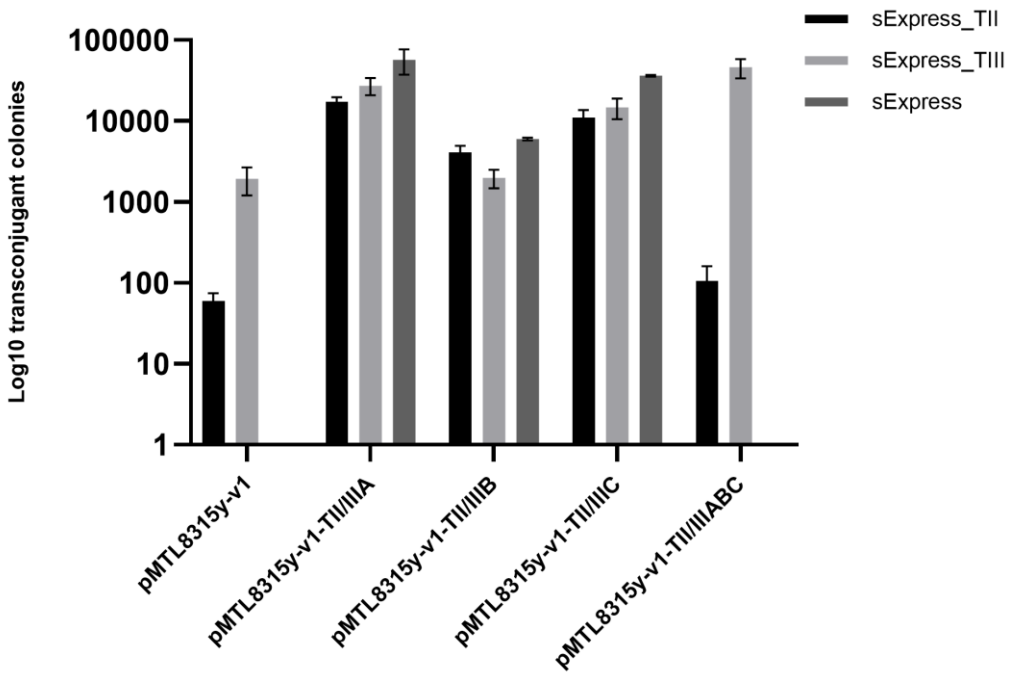


Figure 17 — Comparison of conjugation efficiency of sExpress, sExpress_TII and sExpress_TIII into *C. carboxidivorans* C94 with pMTL8315y-v1-II/IIIA, pMTL8315y-v1-II/IIIB, pMTL8315y-v1-II/IIIC, and pMTL8315y-v1-II/IIIABC.

Data represents average transconjugant number calculated from biological triplicate by extrapolation of selective dilution plate CFUs according to the dilution factor and total mating plate bacterial slurry volume. Error bars represent standard deviation.

3.2.8.2. **Nanopore methyltransferase analysis**

As the cloning burden of complete *C. carboxidivorans'* methyltransferase replication within pR702 was still proving prohibitive, further elucidation of the RM system barrier was also pursued utilising the newly generated pRBV1 vectors. These vectors provide a profile of individual and combined methyltransferase gene cassettes which can be expressed in a non-native background. To assess the impacts of each methyltransferase cassette on the *E. coli* host's methylome using SMRT sequencing with PacBio® would be exceedingly expensive, however recent advances in nanopore technology had expanded into epigenetic analysis. In light of this, a project was devised with Nottingham's in-house sequencing facility, DeepSeq, to employ these new epigenetic modification prediction systems to assess the protections afforded by the recombinant expression of the five Type II and Type III RM system methyltransferase associated genes.

To improve depth of coverage and refine methylation prediction, plasmid DNA and not genomic DNA was sequenced. Due to the confident allocation of TI RM motifs in PacBio®/REBASE® analysis these were not assayed. Instead, all individual Type II and Type III RM system methyltransferase vectors and the combined pRBV1_TII and pRBV1_TIII were chosen, to observe both individual and potential synergistic effects. To ensure the pRBV1 sequence itself was appropriate for this epigenetic analysis the number of known RM motifs within each sequence was determined (**Table 23**). All pRBV1 methyltransferase sequences assessed included four or more of the previously identified TII/III RM motifs, with an average of 7.4, 9.4 and 6.8 of A, B and C respectively in the single methyltransferase expressing vectors.

Table 23 – TII/III RM motif frequency in candidate Type II and Type III pRBV1 vectors

RM motif	TII.1	TII.2	TII.3	TIII.1	TIII.2	TII	TIII
TII/III-A	7	6	8	7	9	8	11
TII/III-B	7	9	10	10	11	12	14
TII/III-C	4	5	7	8	10	10	15

First the pRBV1 vectors (II, II.1, II.2, II.3, III.1, III.2, III) were used to transform chemically competent *E. coli* Stellar to ensure plasmids lacked Dcm and Dam methylation. These transformants were then inoculated in LB broth with erythromycin selection and 50 mM L-arabinose supplementation. These were grown overnight (18-20 hours) before being passaged 1:50 into comparable media and grown for a further 4-6 hours. Plasmid isolation was performed from 5 mL of exponential phase culture (**2.4.1**) and sample quality and concentration measured by nanodrop and concentrations normalised to 100 ng/µL according to Qubit™ derived values. Samples were then prepared by DeepSeq as detailed in **2.5.2** before being sequenced with a GridION x5 mk1 platform. Generated reads which passed quality filters were mapped to the reference sequences using minimap2(2.17) with default parameters (Li, 2018). The read qualities were good with >84% of reads successfully mapped to the reference in all samples excluding pRBV1_TII (**Table 24**). Most aligned reads also have a high alignment identity (>88%). pRBV1_TII analysis highlighted inconsistent reference coverage, with apparent deletion of a region.

Table 24 – Read mapping statistics from Deep Seq GridIon run, demultiplexed and aligned to corresponding pRBV1_T* sequences.

Sample	No. of reads	primary mapped reads	Average coverage	Median aligned read quality and identity
pRBV1_TII	76703	65.90%	20125x	11 / 89.0%
pRBV1_TII.1	315313	87.30%	96745x	10 / 91.5%
pRBV1_TII.2	91902	84.50%	45168x	10 / 91.5%
pRBV1_TII.3	231056	89.10%	36334x	10 / 92.2%
pRBV1_TIII	240587	90.10%	18574x	10 / 92.4%
pRBV1_TIII.1	124491	88.60%	42455x	10 / 92.1%
pRBV1_TIII.2	170654	91.80%	18723x	10 / 92.1%

6mA methylation within this data was identified by Tombo (Stoiber, et al., 2016) with methylation motifs determined and plotted using the MEME Suite (Bailey, et al., 2009). The previously identified RM motifs (AGAAGC, CRAAAAR and GATAAT) were not detected using this analysis, however alternate motifs were predicted with those of sufficient statistical significance detailed in **Figure 18.**

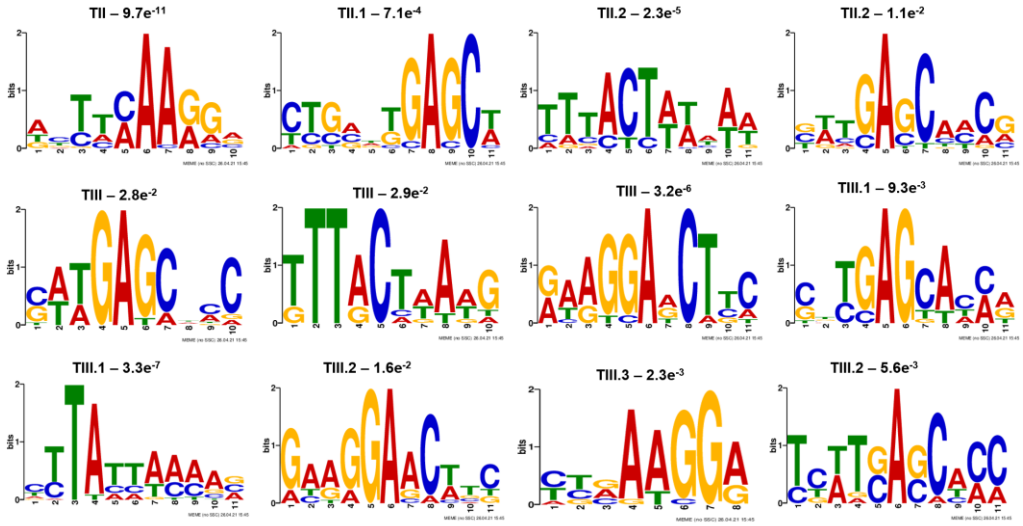


Figure 18 – Sequence logo representation of MEME identified methylation motifs within pRBV_T* reads with E-values ≤ 0.05. Wherein the letter stack height is calculated according to the frequency of the base at a given position within the top 100 Tombo significant regions. Motif width parameters of 8-15, and minimum 2 maximum 100 sites per motif.

Of the motifs identified by MEME analysis, G**A**GC is present within five of the seven samples, and therefore it is probable that this is an *E. coli* Stellar specific motif as opposed to one conferred by the introduced pRBV1_T* plasmids. This does not overlap the known motifs of Dam (G**A**TC), Dcm (C**C**WGG) or EcoKI (**AA**CNNNNNNG**T**GC) the three most prevalent *E. coli* methyltransferases. Another motif identified in three of the seven samples, but once more not correlating with the introduced methyltransferases, was an AAGGAA sequence. Aside from these, the remaining motifs identified were ACTA within pRBV1_TII.2, GAAGC in TIII and TIII.1, as well as TATT and TGAGC in TIII.2. None of these contain the four consecutive adenine residues of the unallocated CRAAAAR TII/III-B sequence, however, the GAAGC of TIII.1 clearly overlaps the AGAAGC TII/III-A motif, further implicating *CcaP7ORF17830P* as a good candidate methyltransferase for recognition of this methylation motif.

3.3. Discussion

3.3.1. Overview

To develop promising bacterial strains for industrial application, a reproducible DNA transfer protocol of reasonable efficacy is a fundamental requirement. Moreover, to better characterise the organism and provide genomic and phenotypic information to inform further development, exogenous plasmid-based tools are invaluable. *C. carboxidivorans* P7 has the natural capacity to produce C4 compounds from C1 feedstocks, making it an attractive means of valorising syngas. As such the lack of receptivity of *C. carboxidivorans* P7 to DNA transfer has been highlighted as a prescient limitation (Shen, et al., 2017). Thus far only one paper has reported successful DNA transfer into *C. carboxidivorans*, and this method did not prove replicable within our research (Cheng, et al., 2019).

This resistance to DNA transfer efforts is a prevalent theme in novel prokaryotes, particularly within clostridia (Pyne, et al., 2014). Gram-positive bacteria are frequently more resistant to transformation efforts due to the thick peptidoglycan layer and protein S-layer acting as a physical barrier, as well as the single cytoplasmic membrane increasing their susceptibility to homeostatic shock. Further to these passive barriers, numerous studies implicate the presence of more impactful, strain specific active systems which must be overcome. The development of a robust DNA transfer method for *C. carboxidivorans* must account for these barriers, applying rational measures to overcome them. This chapter reports the characterisation and circumvention of such barriers within *C. carboxidivorans* and the successful development of a DNA transfer approach (**Figure 19**).

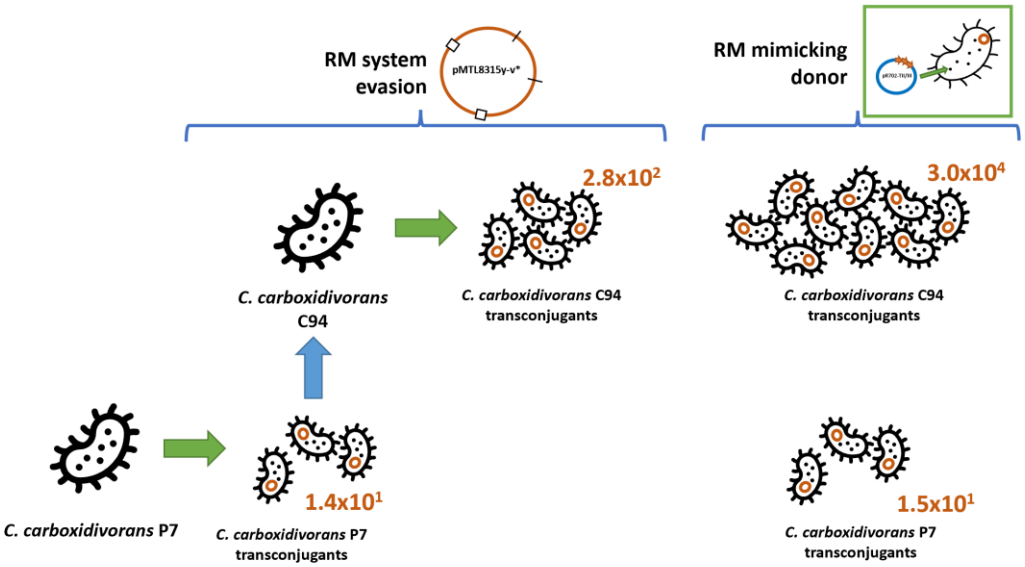


Figure 19 – Schematic overview of the two RM circumvention approaches undertaken within this study, with the wild-type *C. carboxidivorans* P7 strain and *C. carboxidivorans* C94 isolate. Orange writing presenting the optimal average CFU/conjugation yielded using each approach with each strain. Not detailed are the further RM system and hyper-receptive variant characterisation experiments.

3.3.2. REBASE® and PacBio® analysis of the RM systems of *C. carboxidivorans*

The exponential development in next generation sequencing (NGS) technologies has significantly expedited genome acquisition in novel strains. This development is naturally reflected in the exponential expansion of respective genetic databases, including REBASE® (Roberts, et al., 2015) and CRISPRCasdb (Pourcel, et al., 2020). Through application of these databases and associated pipelines, the published genome of *C. carboxidivorans* P7 was analysed. While no CRISPR-Cas associated sequences were identified, ten distinct RM systems were noted. Of these ten, all appear complete, only CcaP7ORF16660 lacking an annotated restriction endonuclease. While the pervasiveness of restriction activity as an active DNA transfer barrier in clostridia has long been acknowledged (Mermelstein & Papoutsakis, 1993), the high number of complete systems identified here presents a particular challenge.

Where previous RM system studies relied on the exposure of plasmids to cell lysate to identify recognition and cleavage sites (Purdy, et al., 2002; Yang, et al., 2016), some NGS technologies can now deliver concurrent epigenetic information which can be utilised. As such methylation, as opposed to restriction, can be readily used as a RM system sequence specificity indicator. Analysis of PacBio sequencing data for *C. carboxidivorans* detected six repeated sequences within the genome at which 6mA methylation occurs with a coverage greater than 99.7% (**Table 12**). No C5-methylcytosine (5mC) methylation patterns were detected. This lack of 5mC identification is likely due to limited sequencing coverage (64x); to recognise the lower kinetic signal of 5mC as opposed to 6mA strand coverage should exceed 250x. Of the 6mA RM motif identified, four were assigned to a corresponding RM system candidate through database alignment. While the CRAAAR and AGAAGC sequences could not be assigned through comparison to these REBASE® Gold Standard Database and REBASE® databases, as they continue to expand further attributions may become possible.

3.3.3. RM system evasion with pMTL8315y-v2/3

RM system evasion through DNA sequence changes is a naturally occurring phenomenon which has been observed in some prokaryotic viruses (Rusinov, et al., 2018). As such, a comparative approach can be employed in DNA transfer studies as a fundamental proof of concept step, with tailored RM system evasion through methylome, lysate or REBASE® informed vector sequence changes (Johnston, et al., 2019). Within *C. difficile*, for example, *Cdi*I and *Sau*96I recognition sites were removed from three vectors, with comparative transfer then performed with the modified and unmodified vectors showing successful transconjugant generation only with the evasive vectors. Further to this the study explored the impacts of various recognition site frequency, with between one and five *Cdi*I recognition sites generating between a five and 16-fold conjugation frequency reduction (Purdy, et al., 2002). This study then

proceeded to development of *in vivo* methylation mimicking, as pursuant with the research reported here.

Through construction and transfer pMTL8315y-v2, devoid of any of the methylated sequences identified by PacBio analysis, the significance of the RM systems as a primary barrier to DNA transfer in *C. carboxidivorans* was similarly confirmed within this study. While the choice of the pMTL83151 vector for this approach was primarily dictated by the number of recognition sequences in the pMTL80000 modules, other factors were considered including the prospective success of the Gram-positive replicon, the necessity of the conjugal *traJ* feature and the maintenance of the MCS for downstream application of the vector.

C. carboxidivorans P7 has a 19 kb plasmid in which the identified replicon, denoted as P19, has been characterised and modularised by James Millard for use with the pMLT80000 vector series (Millard, 2017). This replicon may display incompatibility with other introduced replicons, and due to the toxin-antitoxin (TA) system will outcompete or negate the efficacy of the introduced replicon. With this consideration the pCB102 replicon derived from *Clostridium butyricum* was chosen. Not only has this replicon shown a broad host range in clostridia (Yang, et al., 2016), but previous work by James Millard in p19 replicon characterisation has observed the successful maintenance of both p19 and pCB102 in *E. limosum* ATCC 8486 (unpublished, Millard).

Preliminary vector testing was performed with the sExpress *E. coli* donor strain and *C. sporogenes* NCIMB 1069 as a robust conjugal proxy (Woods, et al., 2019). This+ revealed an elevated transfer rate for pMTL8315y-v2 compared to all other vectors assayed. pMTL831y-v2 and pMTL8315y-v3 are nearly identical sequences, with a slight variation in the placement of changes to disrupt the TII/III-B site in the ColE1 RNA II region. Replication of ColE1-based plasmids is initiated by an RNA complex known as the RNA II pre-primer. This replication complex is destabilised and repressed by a 108 nucleotide-long antisense RNA known as RNA I. The higher RNA I antisense expression results in a 100-fold excess of inhibitor relative to pre-primer. This negative feedback loop dictates plasmid copy number in each cell and can be influenced by metabolic

fluctuations and the presence of additional plasmids in the host (Camps, 2010). If the affinity of the RNA I antisense regulatory system was reduced by the sequence change made in pMTL8315y-v2, the increase in RNA II pre-primer stability would increase plasmid replication and could lead to a higher copy number. With higher plasmid transcript availability in the donor cell the overall conjugal transfer efficacy may also be improved. Alternatively, a reduction in stability and resultant reduction in copy number could reduce the metabolic burden of the plasmid within the cell, increasing the likelihood of conjugation machinery expression and increasing the number of transfer competent cells within the conjugating population. More extreme reduction in RNA I antisense efficacy has resulted in runaway plasmid replication, which can be lethal for the host (Lorenzo-Díaz, et al., 2017). This does not appear as a concern as no variation in growth rate was observed between the sExpress donor strains maintaining the different plasmids, however this was not specifically assayed within this study.

3.3.4. CA434 conjugation

The conjugal DNA transfer method reported by Cheng et al. (2019) utilising CA434 as the donor strain was attempted twice within this study, with both pMTL8315y-v1 and pMTL82151 in triplicate. pMTL8315y-v1 was additionally used in conjugal efforts as it possesses only four of the methylated sequences identified by PacBio analysis of *C. carboxidivorans*, compared to pMTL82151 which possesses 21. No transconjugants were observed from either vector or repetition following 168h incubation under selective conditions.

CA434 is a conjugal *E. coli* strain created by the transfer of the R702 R factor into HB101. It is based predominantly on the K12 strain (96.9%) with the remaining portion derived from the *E. coli* B strain. Despite its prevalence as a conjugal strain for clostridial studies (Jeong, et al., 2017), conjugations with the CA434 strain have historically shown irreproducibility. In 2019 only six of a reported ten replicates of CA434 mediated conjugal transfer into *C.*

autoethanogenum yielded transconjugants (Woods, et al.). Failure to replicate this DNA transfer method within this study may be further illustrative of the low efficacy of this donor. As such, if the overall efficacy is low, the lack of success reported here may be rectified by increasing replicates to detect lower transfer rates.

Within the published study it was hypothesised that successful DNA transfer into *C. carboxidivorans* was attained due to evasion of RM system nucleolytic activity. This evasion was afforded by conjugation which delivers plasmids as single-stranded DNA (ssDNA) which would not be readily recognised by the RM systems which primarily target unmethylated double-stranded DNA (dsDNA). As this research was focused on metabolic engineering efforts there was no information regarding the efficacy or reproducibility of the conjugal transfer. This rationale of conjugative evasion of RM systems is echoed in other studies, but despite its prevalence contradictions were observed as early as 1965, when the varied conjugational transfer efficacy of episomes into *E. coli* was attributed to host specificity (RM systems) (Arber & Morse, 1965). This is corroborated by further studies which exemplify the benefits and sometimes necessity of RM system evasion even when transferring DNA by conjugation, whether by disruption of restriction endonuclease genes or methylation mimicry. Some clostridial examples include: *C. difficile* CD3 and CD6 (Purdy, et al., 2002), *C. autoethanogenum* DSM10061 and *C. sporogenes* NCIMB 10696 (Woods, et al., 2019) or *C. thermocellum* ATCC 27405 (Riley, et al., 2019). This is likely due to the *in vivo* processing of introduced ssDNA into replicative plasmid, in which the DNA proceeds through a hemi-methylated dsDNA stage sensitive to restriction activity which may be degraded in lieu of maintained.

An alternative hypothesis as to the lack of success reported herein replicating transfer with this method, is the presence of more amenable mutant sub-populations within the original culture. Bacterial populations, depending on mutation rate, can acquire and maintain mutant sub-populations if the changes do not have deleterious effects under standard growth conditions. The bacterial culture used within the published study may hence be benefitting

from an elevated presence of mutants, underrepresented within our stocks, with increased receptivity to DNA transfer.

3.3.5. *E. coli* Dam/Dcm methylases

When considering the impact of RM systems on DNA transfer a necessary concern is the methylome of the donor strain. Two methyltransferases found in *E. coli* strains are Dam and Dcm. Dam is responsible for methylation of the adenine residue in the sequence G**A**TC, whereas Dcm methylates the second cytosine residue in the sequence C**C**WGG. These sites are relatively prevalent within the vectors of this study, each appearing eight times within the pMTL8315y vectors, and ten and nine times respectively in pMTL83151.

Three *E. coli* donor strains, Interstellar, TopSex and sExpress were all employed as conjugal donors for the transfer of pTML8315y-v2 into *C. carboxidivorans* C94 (**Table 25**). Of the three strains tested transconjugants were only observed with the sExpress (*dam-, dcm+*) donor. While there are other phenotypic differences between the tested donors the impacts of *dam* and *dcm* methylation on DNA transfer have been noted as significant in many studies; *C. pasteurianum* NRRL B-598 (Kolek, et al., 2016), *C. thermocellum* DSM 1313 (Guss, et al., 2012), *C. sporogenes* NCIMB 10696, *C. autoethanogenum* DSM 10061 and *C. difficile* R20291 (Woods, et al., 2019). Within *C. thermocellum* in fact the two RM system genes *Cthe_2470* and *Cthe_1511* were identified as Dam methyltransferases, meaning Dam+ donor strains were clearly required (Riley, et al., 2019). This correlation is less evident in other strains, where overlap of the associated Dam methylation sequence with native RM motifs is not always observed.

Table 25 – Genotype information for *E. coli* donor strains tested including *dcm* and *dam* gene integrity and strain origins.

***E. coli* strain**	**Derivation**	***dcm***	***dam***	**Genotype information**

InterStellar	Clontech™ Stellar, HST08 derived	-	-	F-, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ(lacZYA-argF) U169, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, λ-, R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob
sExpress	NEBExpress®, BL21 derived	-	+	*fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10*—Tet^S*)2 [dcm] R(zgb-210::Tn10*—Tet^S*) endA1 Δ(mcrC-mrr)114::IS10*, R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob
Top10	Invitrogen MG1655 derived	+	+	F- *mcrA, Δ(mrr-hsdRMS-mcrBC), φ80lacZΔM15 ΔlacX74, nupG, recA1 araD139, Δ(ara-leu)7697 galE15, galK16, rpsL(StrR) endA1, λ-*, R702-Tc^R, Sm^R, Su^R, Hg^R Tra+, Mob

A lack of Dcm methylation may improve DNA transfer efficacy in *C. carboxidivorans*, as is observed in the successful transfer of pMTL8315y-v2 by sExpress but not by TopSex. It is hypothesised that the lack of cytosine methylation reduces the nucleolytic activity of the three Type IV systems on the transferred DNA. BLASTp analysis of these Type IV systems shows high homology between the TRD1 regions of CcaP7McrB, CcaP7McrB2P and CcaP7ORF03495 and three respective Type IV systems of *C. autoethanogenum* DSM10061: Cau10061McrB2P, Cau10061McrBP and Cau10061ORFEP. sExpress is reported to be a highly effective donor for transfer to *C. autoethanogenum* DSM10061 with 600 times the transconjugants observed for sExpress over CA434 (Woods, et al., 2019). CA434 also exhibits Dcm and Dam methylation. The recognition sequences of the six Type IV systems of both *C. autoethanogenum* and *C. carboxidivorans*, may hence partially overlap the Dcm methylated C**C**WGG sequence.

Lack of successful transfer from the *dcm*+ *dam*- donor Interstellar, as compared to sExpress indicates *dam* methylation may be advantageous for DNA transfer into *C. carboxidivorans.* This advantage is unlikely to be afforded by evasion of the Type I, II and III RM systems as there is no overlap of the Dam methylated GATC sequence with any of the 6mA methylated sequences identified by PacBio analysis (**Table 12**). Dam methylation lowers the thermodynamic stability of DNA, alters its curvature, and introduces steric hindrance. These

changes have an impact on DNA-protein interactions, improving affinity in instances where they are expected. As such an alternate rationale is that Dam methylation affords advantages to DNA transfer efficacy because methylated DNA, as opposed to unmethylated, presents a better substrate for processing within *C. carboxidivorans* following initial ssDNA introduction. This could be tested by observing whether expression of the *C. carboxidivorans* derived methyltransferases in InterStellar restores conjugative functionality.

3.3.6. *E. coli* sExpress recombinant donors

The application of *in vivo* methylation mimicking within donor strains is an increasingly prevalent approach to attaining DNA transfer in recalcitrant clostridial strains. Unlike RM evasion this approach affords RM system circumvention irrespective of the vector sequence, allowing the transfer of larger molecular tool vectors in which the destruction of RM motifs would be burdensome and may impact system efficacy. This approach was applied within the *Clostridium* genus as early as 1993 when the *B. subtilis* phage phi 3T I methyltransferase was expressed in *E. coli* to protect the Cac824I recognition/cleavage sites within vectors used to transform *C. acetobutylicum* ATCC 824 (Mermelstein & Papoutsakis, 1993). More recent studies include the previously mentioned *C. difficile* CD3 and CD6 RM system circumvention, initially with CdiI site removal followed by M.*Sau*96I methylase gene expression within the conjugal donor CA434 (Purdy, et al., 2002).

Similar studies have been performed with *Clostridium saccharobutylicum* NCP 262 (Lesiak, et al., 2014), *Clostridium cellulovorans* 743B (Yang, et al., 2016) and *Clostridium thermocellum* ATCC 27405 (Riley, et al., 2019). RM systems were first identified within each strain, with two in *C. saccharobutylicum*, eight within *C. thermocellum* and twelve putative operons within *C. cellulovorans*. These were refined further where possible, with those without associated restriction endonucleases disregarded in *C. cellulovorans* narrowing the number to six. Further analysis of the methylome of *C. thermocellum* also

highlighted six recurrent methylation motifs which were attributed to RM systems where possible. Despite this distinct variation in RM system number and type between each strain, each study concluded in the heterologous expression of only two methyltransferase genes within the recombinant *E. coli* strain. Within the *C. saccharobutylicum* these were the two native methyltransferases *hsdM1* and *hsdM2,* within *C. cellulovorans* the M. *Cce*743I and M. *Cce*743II, and within the *C. thermocellum E. coli* donor Cthe_0519 and *phi3TI* were chromosomally integrated. To narrow the field in *C. cellulovorans* genomic analysis and BLAST alignments were performed which highlighted the presence of one Type I RM system with EcoKI analogous specificity, a site which did not occur within the vectors used in this study. Further to this four Type II RM genes were devised as overlapping in site specificity, as such only targeting two motifs; CCWGG (*Lla*I sites) and GMSKT (*Lla*JI sites). This was further confirmed through lysate restriction of appropriate vectors. Within *C. thermocellum* the use of a Dam+ strain provided protection for one predicted RM motif, with *phi3TI* also providing protection from both the Cthe_1749 and 2321 sites and the further CNCANNNNNNTTC site not occurring within the chosen vector.

With limited sequence specificities identified, no apparent RM system redundancy and the prevalence of all RM motifs within the vectors of this study the initial *C. carboxidivorans* methyltransferase focus could not be so specifically refined. Instead, the five Type II RM system methyltransferases associated with the three most prevalent RM motifs were focused on, with the further goal of complete native methyltransferase cohort assembly in a singular donor strain to completely divorce all future efforts from any sequence restrictions (still present in two of the aforementioned studies). Initial assessment of the recombinant *E. coli* donors sExpress_TII and sExpress_TIII was performed by conjugal transfer of pMTL8315y-v1, pMTL8315y-v2 and pMTL83151 into *C. carboxidivorans* C94. No transconjugants were attained through any donor with pMTL83151, as was expected due to the presence of four Type I RM motif sequences (corresponding to all three native Type I

systems). Plasmid pMTL8315y-v1, however, underwent conjugal transfer through both recombinant donors into *C. carboxidivorans* P7 and C94. pMTL8315y-v1 contains only two types of TII/III RM motif, at a frequency of one and three respectively. As such, this is not the most ideal vector for recombinant donor validation. It was chosen for preliminary assessment due to the lack of Type I recognition sequences, a feature not readily available in the other plasmids within this study.

While the successful transfer of pMTL8315y-v1 into *C. carboxidivorans* through the recombinant donors further corroborates the significance of the RM systems as a primary barrier to DNA transfer within the strain, the extremely low rates of transfer (0.8-1.4 x 10^1 CFU/conjugation) are unlikely to permit the introduction of extensive vector-based molecular tools for *in vivo* genomic manipulation. The exception to this, as observed within the development of similarly limited *C. tyrobutyricum* is endogenous CRISPR system exploitation (Zhang, et al., 2018). As this is not an option within *C. carboxidivorans* it would be advisable for downstream tool development to proceed within the C94 variant strain. To confirm its suitability for downstream applications C94 was grown autotrophically in serum bottles with 50% CO, 40% N_2, 5% H_2, and 5% CO_2 at 1 bar, by C. Humphreys (unpublished), to ensure its amenability for application as an industrial chassis for syngas fermentation. No variation in autotrophic growth capabilities between C94 and P7 were observed.

Aside from the significant improvement in transconjugant numbers produced by both sExpress_TII and sExpress_TIII over sExpress, the difference between the two recombinant donors is also statistically significant (0.0002). The greater success of the two-methyltransferase gene expressing sExpress_TIII was also consistent with pMTL8315y-v1-II/IIIABC, a vector containing twelve TII/III RM motif sequence of all types. The higher transfer rate observed in sExpress_TIII over sExpress_TII could be attributed to the number of sites the donor protects, indicating *CcaP7ORF8705* or *CcaP7ORF17830* are likely candidates for recognition and methylation of the three TII/III-B (CRAAAAR) sequences which occur in the pMTL8315y-v1 backbone. However, this hypothesis is reliant on

number of assumptions including that only the PacBio identified sequences are relevant, that all methylated sequences have a comparable impact on transfer efficacy and that a single methyltransferase is responsible for each sequence (there is no redundancy). Therefore, no such conclusions can confidently be made without further methyltransferase characterisation.

The notable lack of successful pMTL8315y-v2 conjugation was not further explored within this study, but may be due to adverse effects of methyltransferase activity on plasmid integrity within the conjugal strain. In routine cloning pMTL8315y-v2 demonstrated a reduced receptivity to digestion and Sanger sequencing which may further corroborate this abnormal behaviour. Alternatively, if pMTL8315y-v2 does have an elevated plasmid copy number compared to pMTL8315y-v1 due to RNA II pre-primer stability variance, the concurrent L-arabinose induction of methyltransferase gene expression may significantly compound the metabolic burden experienced by the cell. This stress may prevent the epigenetic alleviation of conjugation machinery repression, resulting in a reduced number of transfer competent cells within the conjugating population.

3.3.7. Hyper-receptive conjugative recipient strains characterisation

The isolation of hyper-receptive variants as a means of increasing the frequency of transformation has previously been demonstrated in *C. pasteurianum* ATCC 6013 (Grosse-Honebrink, et al., 2017). Prior to this study DNA transfer efforts within *C. pasteurianum* had similarly focused on elucidation of the native RM systems, perceiving them to be the greatest hurdle within the organism. REBASE® analysis of the genome of *C. pasteurianum* ATCC 6013 (DSM 525) predicted a total of seven RM systems, two Type I and five Type II. Of these four have associated restriction endonuclease activity, with one orphan Type I specificity gene, and two orphan Type II methyltransferase genes. The REBASE® predicted recognition sequences were further

corroborated with methylome data revealing four distinct m6A methylation motifs. While this RM system profile is slightly less extensive than that discerned within *C. carboxidivorans, C. pasteurianum* also contains two putative CRISPR arrays corresponding to a native Type I-B CRISPR-Cas system (Pyne, et al., 2016).

Analogous to the preliminary C94 isolation described herein, the work of Grosse-Honebrink *et al.* initially identified a single plasmid-cured variant with significantly improved DNA transfer receptivity, designated 525-H1. This isolate afforded an increase in the frequency of DNA transfer of three orders of magnitude, reaching 10^5 per µg DNA with *in vivo* M.*Bep*I-methylated pMTL85151. This leap in efficiency is also directly comparable to that observed within this study, at 1.5 x 10^1 CFU/conjugation observed when conjugating *C. carboxidivorans* P7 with pMTL8315y-v1 and sExpress_TIII, to 3 x 10^4 for comparable transfer into *C. carboxidivorans* C94. Genomic analysis of 525-H1 implicated two genes, CLPA_c13710 (*ablB*) and CLPA_c33080 (*resE9*), predicted to encode a β-lysine N6-acetyltransferase and a histidine kinase, respectively. Subsequent repetition of variant isolation generated two further hyper-transformable variant strains with comparable DNA transfer frequencies to 525-H1 denoted 525-H3 and 525-H4. These were also sequenced with genomic analysis identifying mutations in the structural maintenance of chromosome (SMC) proteins. None of the genes implicated in genomic analysis of C94-97 were SMC proteins.

This hyper-transformable variant isolation approach has since been repeated in *C. pasteurianum* in one other study. This study only isolated one hyper-transformable mutant, referred to as R525 (Schmitz, et al., 2019). Genomic sequencing and analysis identified a total of eleven SNPs, two insertions and two deletions within R525. These were refined by comparison to the concurrently acquired parental genome sequence, with synonymous mutations disregarded as well as those within non-coding regions. This refined the results to two mutations of note, the first of which was a 3 bp insertion in the DNA binding domain a transposase, encoded by F502_06873. The second

was an SNP resulting in a single amino acid change in a non-transmembrane domain of a hypothetical predicted integral transmembrane protein, encoded by F502_05127. As within the Grosse-Honebrink study, the mechanism of this improved DNA transfer frequency was only hypothesised and not experimentally determined. Isolation of R525 however, facilitated subsequent metabolic engineering efforts including the transfer of a nCas9 based CRISPR vector-based system for deletion of the glycerol dehydratase large subunit *dhaB*. The SNPs identified throughout the two *C. pasteurianum* studies are diverse, with a β-lysine N6-acetyltransferase, a histidine kinase, SMC proteins, a transposase and hypothetical transmembrane protein all implicated. Of the three SNPs identified by analysis of C94-97 none are congruent with any observed within *C. pasteurianum.*

The best gene candidate for the increased receptivity phenotype identified within the *C. carboxidivorans* variants was identified as Ccar_20450/*fabK*. tBLASTn alignment of the *C. difficile* FabK protein sequence against the *C. carboxidivorans* genome identified no alternate candidate *fabK* genes within the genome or plasmid, with further analysis highlighting no alternate enoyl-ACP reductase genes which would duplicate the function of FabK. The role of FabK within the cell is in the bacterial fatty acid type II biosynthesis pathway, converting trans-2-enoyl-[acp]s in cyclical chain elongation to butyryl-[acp], hexanoyl-[acp], decanoyl-[acp], tetradecanoyl-[acp], hexa-decanoyl-[acp] and octadecanoyl-[acp]. This fatty acid synthesis is integral to various biological processes, including phospholipid production. This may, therefore, impact cell membrane formation, changing both fluidity and permeability. FabK has in fact recently drawn attention as a high specificity antibacterial targeting for the eradication of *C. difficile* in gut microbiomes (Jones, et al., 2019). The amino acid altered within FabK, in three of the four variant strains, is Gly146. In structural analysis and characterisation of the *Streptococcus pneumoniae* FabK, His144 is proposed as catalytic, with the two preceding Glycine residues integral to substrate binding stability through solvent hydrogen bond bridging (Saito, et al., 2008). While not directly implicated in substrate binding, BLAST

alignment of this FabK with that encoded by the wild-type *fabK* gene of *C. carboxidivorans* (311 score, 59% identity) does demonstrate the conservation of Gly146 between both sequences. The proximity of this change to the catalytic residue and substrate binding site, and resultant transition from the smallest amino acid residue to the much larger arginine and sulphur containing cysteine in the mutant sequences, may impact the substrate binding affinity and stability at these adjacent residues.

The significant improvements afforded by isolation of *C. carboxidivorans* and *C. pasteurianum* variants highlights the promise of forward genetics approaches within recalcitrant organisms, where high efficiency DNA transfer cannot be attained solely through rational approaches. With perfect information on an organism's genotype-to-phenotype relationship such approaches would not be necessary, and desired traits could simply be engineered in rather than evolved. Unfortunately, the complexity of biological systems renders our current knowledgebase insufficient to inform such rational design, and as such forward genetics approaches are indispensable. These approaches are typically reliant on the generation of a large diverse mutant pool, however, the protocol described within this study is reliant on the natural occurrence of mutant subpopulations. As such these mutants typically generate changes which are not disadvantageous or advantageous under standard growth conditions, permitting them to establish stably within the population despite clonal interference. The high occurrence Ccar_00825 and Ccar_25120, therefore, is likely to indicate these mutations have little to no impact on cell viability under the selected conditions. Other mutations, such as that observed in Ccar_20450, appear to occur more rarely within the population, perhaps due to deleterious effects of minor severity. As such their occurrence is dictated in part by the strain's specific mutation rate, recently estimated for *E. coli* at 10^{10} per generation (Wielgoss, et al., 2011).

This approach was only successfully applied within *C. carboxidivorans* through the initial circumvention of the RM barrier which facilitated low levels of DNA transfer. As such it is not broadly applicable to recalcitrant organisms in which

no DNA transfer has been achieved. Further mutant pool enrichment prior to the selective screening by transformation would likely increase the chance of relevant mutant isolation, potentially increasing both the number and the type of hyper-receptive mutants yielded by this approach. While this would likely implicate a wider profile of genes not previously considered in DNA transfer studies, importantly mutagenesis can be performed with extracellular tools. As such this approach could be applied to species where there is both no basal transfer level and low mutation rate.

Thus far the mechanisms responsible for the increased plasmid transfer have not been elucidated, with different genes implicated between *C. pasteurianum* and *C. carboxidivorans,* and even within various strains of the former. Determination of the exact *in vivo* impacts of these minimal genome variations would likely require a dedicated study. Repetitions within this study were limited by time constraints, but to further confirm the significance of FabK on DNA transfer frequency it would be prudent to perform more hyper-receptive variant isolations through independent conjugations into *C. carboxidivorans* P7. The significance could be further corroborated by employing genetic modification tools to restore the wild-type gene sequence within variant strains. Subsequent assessment of the DNA transfer frequency using these restored strains compared to the non-modified version should highlight any restored recalcitrance.

3.3.8. Further RM system characterisation

Unfortunately, the application of nanopore sequencing to characterise the recombinantly expressed methyltransferase genes of the *C. carboxidivorans* TII and TIII RM systems did not facilitate the confident allocation of the remaining TII/III RM motifs identified by initial PacBio® SMRT sequencing. This recombinant expression approach to methyltransferase characterisation has been previously effectively demonstrated using SMRT sequencing, with a recent study within *E. coli* ER2796 not only identifying previously unknown

methylation motifs, but also successfully allocating a known motif with only two occurrences within the sequenced plasmid sequence (Clark, et al., 2012). Thus far, however, this has not yet been replicated with nanopore technologies, with the majority of nanopore DNA-modification studies focused on whole genome and transcriptome applications (Xu & Seki, 2020).

Interestingly, the expression of *CcaP7IVP* within pRBV1_TII.2, the methyltransferase confidently attributed to the GATAAT methylation motif, did not generate an identifiable GATAAT 6mA methylation motif via this approach, in fact identifying an ACTA motif instead. To ensure good methylation coverage all samples were induced with 50 mM L-arabinose for 24 hour, five times the induction levels which generated *C. carboxidivorans* C94 transconjugants with the 13 RM motif containing pMTL8315y-v1-TII/IIIABC vector. Prior expression characterisation of an L-arabinose induced P_{BAD} promoter during development of the pRBV1-R702 recombination system by R. Mansfield demonstrated significant expression of a P_{BAD} driven RFP reporter gene following 15.5 hours at 10 mM and 100 mM L-arabinose, even when the RFP was integrated into the single copy pR702 plasmid (Mansfield, 2018). Further unpresented data also indicated that an uninduced recombinant donor managed to increase DNA transfer frequency within *C. acetobutylicum* 824, indicating that the level of expression required to afford protective methylation of the plasmid could be markedly low. As such it appears unlikely that the induction protocol was not sufficient to induce methyltransferase expression. Instead, the lack of identification of the GATAAT sequence may be due to a combination of limitations in the sensitivity of the nanopore approach and the existence of variable methylation across the extracted multi-copy plasmid DNA. Nanopore DNA modification identification is a rapidly developing field which is only preferentially applied over the gold standard SMRT sequencing for 5mC identification, a shortfall of the SMRT technique. Otherwise, there are still key hurdles within the technology including the identification of rare motifs and accurate assessment of partial methylation. As such SMRT and nanopore

sequencing are largely viewed as complementary DNA modification tools at this stage (Tourancheau, et al., 2021).

This lack of accuracy is also apparent in the identification of the G**A**GC sequence, hypothesised to be derived from the background *E. coli* phenotype, in only five of the seven samples analysed. Further REBASE® PacBio analysis was performed to highlight prevalent GAGC containing motifs within *E. coli.* This highlighted a GAG**C**C methylated motif, however, solely with 5mC methylation of the first cytosine residue. The *E. coli* HST08 (Stellar) RM profile had not been submitted to the REBASE® database at time of writing, however, the HST04 strain had. Unfortunately, only one recognition sequence had been predicted within this strain, GATC, with the corresponding base modification type unknown. Similar analysis of the AAGGAA sequence, identified in three of the seven samples, could not attribute it to a corresponding background methyltransferase within the *E. coli* donor.

Efforts to further elucidate the RM systems through conjugative transfer of pMTL8315y-v1-II/IIIABC, and respective variant vectors, also did not permit the confident allocation of the TII/III-A and TII/III-B RM motifs to respective RM systems due to the lack of significant variance between the sExpress_TII and TIII donor transconjugant number generated with each vector. Successful transfer of the vectors through a non-recombinant donor strain did, however, highlight the reduced receptivity of the C94 strain to RM motif inclusion. While the genomic characterisation of *C. carboxidivorans* C94, C95, C96 and C97 implicated *fabK* within this more receptive phenotype, the mechanism of this increase in DNA receptivity is far from being understood. Intriguingly all vectors containing seven RM motifs demonstrated higher DNA transfer frequency than the four-site containing pMTL8315y-v1 base vector. This increase in DNA transfer frequency could be due to the timeframe in which hemi-methylated conjugal DNA is likely to be degraded within the cell; through the inclusion of further RM motifs, a greater number of hemi-methylated 6mA sites may increase the affinity of replicating enzymes within *C. carboxidivorans,* where 6mA methylation at these sites is expected, accelerating the transition into

double-stranded DNA, and minimising time present in the vulnerable hemi-methylated stage.

This hypothesis, however, contradicts what was observed for pMTL8315y-v2, with no successful transfer through recombinant donors, but minimal transfer using a wild-type sExpress *E. coli* strain. It is difficult to disentangle the potential repercussions of the original RM evasion sequence modifications on vector functionality from the significance of the reduced RM motif inclusion, particularly as they relate to the inability to generate successful transconjugants. As such the lack of transfer of pMTL8315y-v2 by recombinant donor could be the result of the sequence changes reducing respective functionality, or due to the lack of *C. carboxidivorans* RM motifs reducing the affinity of DNA-processing enzymes within the recipient cell following dsDNA synthesis from conjugated ssDNA. This hypothesis could easily be tested through the integration of the synthetic TII/III-A, B or C fragments, to identify whether RM motif inclusion restores conjugative transfer prospects for this vector within the recombinant donors.

3.4. Chapter summary

- Preliminary characterisation of the native RM systems within *C. carboxidivorans* was performed by REBASE® alignments and methylome data analysis
- The sensitivity of *C. carboxidivorans* to thiamphenicol, and applicability of the *catP* selection marker was confirmed.
- Successful DNA transfer into *C. carboxidivorans* was initially attained through RM system evasion with sequence modified minimal plasmids
- Four hyper-receptive conjugative *C. carboxidivorans* recipient variants were isolated and genetically characterised, all with mutations of an enoyl-acyl carrier protein *fabK*
- Two recombinant *E. coli* strains were generated expressing *C. carboxidivorans* methyltransferase genes affording partial RM system

circumvention, and DNA transfer of a minimal vector reaching 3 x 10^4 CFU/conjugation into *C. carboxidivorans* C94

- Further RM system characterisation was attempted through nanopore sequencing and conjugal transfer of new pMTL8315y-v1 vectors with additional TII/III RM motif inclusion

4 Development of genome modification tools for *C. carboxidivorans*

4.1. Introduction

4.1.1. Metabolic engineering of *C. carboxidivorans*

Establishing genetic modification tools within an organism facilitates the application of reverse genetics to investigate key genotypic/phenotypic relationships and allows the informed manipulation of a strain to improve desired traits. *C. carboxidivorans* is the only bacterial species reported to produce hexanol directly from syngas, alongside the industrially relevant compounds ethanol, hexanoate, butyrate and *n*-butanol. Both ethanol and butanol can be used as directly compatible alternatives in current petroleum-based energy infrastructure. As such the principal goal pursued to date within *C. carboxidivorans* studies has been improved long-chain alcohol yields under autotrophic conditions. While comparative analysis between studies is limited by the diverse fermentation approaches taken, the highest reported hexanol titre thus far was 2.4 g/L (Kottenhahn, et al., 2021). This yield was achieved by application of *in situ* product extraction to mitigate the impacts of hexanol toxicity, but other studies have also noted significant yield improvements through changes in other key fermentation parameters, such as pressure, reactor type and trace metal supplementation (Han, et al., 2020; Shen, et al., 2017; Shen, et al., 2014). This extensive exploration of fermentation parameters is not, however, complimented with extensive metabolic engineering efforts, with only one study published to date implementing *in vivo* approaches to this challenge and none applying genomic modification tools.

The study in question, published by Cheng et al. in 2019, utilises the pMTL82151 plasmid vector as a means of overexpressing a heterologous aldehyde/alcohol dehydrogenase (ALD) and ferredoxin-NAD^+ reductase (FNR), as well as the native aldehyde::ferredoxin oxidoreductase (AOR) within *C. carboxidivorans* P7 (Cheng, et al., 2019). Both heterologous genes were derived

from *C. acetobutylicum*, with the bifunctional aldehyde/alcohol dehydrogenase encoded by *adhE2* utilised for similar studies in a plethora of *Clostridium*. The results of these studies have demonstrated the significance of *adhE2* within ethanol and butanol biosynthesis, however the significance of AOR within alcohol biosynthesis is not well understood. AOR has been noted as important in ATP generation under energy-limited conditions in acetogens, and FNR is integral to electron donor balance and regeneration through the transfer of electrons from reduced ferredoxin to NAD^+. Hence overexpression of these genes has energetic repercussions which in turn impact carbon flux within the organism (Müller, 2003).

Four overexpression strains and the wild-type *C. carboxidivorans* P7 were analysed by batch growth on either glucose or a syngas mix. Each overexpression strain had a differing gene combination with ALD and FNR concurrently expressed, as well as AOR and FNR. ALD and AOR were also expressed individually. When grown on glucose all strains showed higher terminal alcohol yields compared to the wild-type. The primary changes were in ethanol, with butanol/butyrate production fluctuations negligible. Within the ALD overexpressing strain, a 31% higher ethanol yield was observed, and within the AOR strain a 70% higher ethanol terminal yield. This AOR overexpressing strain also exhibited higher cumulative total product yields. It was hypothesised this was due to the re-assimilation of CO_2, which stoichiometrically constitutes one third of the carbon flux from glycolysis. Through solventogenic tracking it was apparent that no clear acidogenic to solventogenic shift occurred, with acid yields exceeding alcohol throughout.

When grown autotrophically on syngas ALD overexpression produced the highest ethanol yield observed within the study at 3.00 ± 0.061 g/L, an increase of ~50% over the wild-type strain assayed. The strain overexpressing both ALD and FNR produced ~22% more ethanol and the highest butanol yield of the study at 0.351 ± 0.010 g/L, ~18% greater than wild-type. Both strains also showed obvious acid re-assimilation, producing ethanol and butanol in later growth stages and reducing terminal acid yield to <0.15 g/L. Compared to the

fermentations on glucose the impact of AOR overexpression was minimal within the syngas fermentations, which was hypothesised to be due to the already elevated gene expression stimulated by autotrophic growth. ALD activity requires NADH as a cofactor, which can be produced by FNR. The concurrent expression of these, however, lowered alcohol yields, likely because FNR does not couple NADH production to the generation of a chemiosmotic gradient as an Rnf complex does. This gradient is integral to autotrophic ATP generation, and as such would have significant repercussions in bioenergetic balancing.

This study is a promising start for metabolic engineering efforts in *C. carboxidivorans*, demonstrating the significant improvements in medium-chain alcohol yields which can be readily achieved through *in vivo* metabolic manipulation. It also examines some energetic considerations and the roles of redox donors to further the fundamental understanding of the strain. The alcohol titre reported, however, is still sub-par with limited prospects for industrial applications. Some alternate targets are proposed within the study to further address this, including increasing gaseous pressures to improve aqueous substrate availability (previously demonstrated by Fernández-Naveira *et al* (2016)), and the development of genomic manipulation tools to disrupt the acid synthesis pathways and shift carbon flux.

The balance of C2 and C4 products within the study was also disappointing, with butanol yields the least impacted by the metabolic changes made (Cheng, et al., 2019). A pointed focus on chain elongation routes in future metabolic engineering endeavours may hence be key. While no further *in vivo* studies have been published within *C. carboxidivorans*, a more recent study has taken strides in elucidating its integral genes for C6 product formation through heterologous expression within the more amenable strains *A. woodii* and *Clostridium saccharoperbutylacetonicum* (Wirth & Dürre, 2021). Two putative long-chain alcohol gene clusters were identified and expressed in *A. woodii* denoted as *hcs* cluster 1 (Ccar_22805_Ccar22810), and *hcs* cluster 2 (Ccar_01430-Ccar_01436). Expression of *hcs* cluster 1 within *C.*

saccharoperbutylacetonicum facilitated hexanol production, a product which had never been previously produced by the organism. The genes within these clusters include *thl*, *hbd*, *crt*, and *bcd*, all of which are hypothesised to be downregulated under autotrophic conditions and as such would be interesting targets for genomic engineering (Bruant, et al., 2010).

While plasmid-based metabolic engineering studies are useful and provide key insights, the ideal in *in vivo* engineering is genomic modifications. The use of plasmid-based expression, while potentially conferring overexpression through high copy number, is inherently unstable. As such plasmid maintenance typically relies on antibiotic selection markers, whose corresponding supplementation within the media would be prohibitively expensive at scale and generate significant logistical considerations. Alternatively, plasmid addiction systems could be implemented, for example engineered prototrophy as demonstrated in *C. autoethanogenum* (Annan, et al., 2019). While interesting, such systems have not yet been widely demonstrated within the genus and require considerable dedicated research efforts. Plasmid based expression also applies significant metabolic burden and was noted as generating an observable reduction in growth rate within the *C. carboxidivorans* study. This would in turn limit the overall process efficacy of fermentation when employing this strain as an industrial chassis.

Considering this, and the alternative approaches afforded by genomic modification, the Cheng *et al C. carboxidivorans* metabolic engineering study itself emphasises the pertinence of genomic manipulation tool development within the strain. Key considerations when adapting and developing such tools include the diversity and efficiency of the system, scope for marker-less changes of significant size, and ability to multiplex or iteratively apply the system. The best candidate genomic manipulation systems used broadly within *Clostridium* for such applications are allelic exchange and CRISPR-Cas systems.

4.1.2. Allelic exchange genome editing in *Clostridium*

The most widely applied modern strategy for genomic modification in prokaryotes is the use of allelic exchange. Within allelic exchange homologous recombination is exploited to replace wild-type genomic DNA with a heterologous sequence. The sequence which is being introduced is framed by sequences identical to those flanking the native allele. It is between these identical sequences that homologous recombination occurs, replacing the native allele with the heterologous sequence. Various methods are used to select for or pressure these recombination events, isolating successful double crossover mutants whether they be in-frame deletions or heterologous gene introductions.

4.1.2.1. Genomic selection/counter-selection

The targeted disruption of specific genes within the genome can confer phenotypic changes which form the basis of mutant selection. One early prominent example of this used within clostridia is the Allele-coupled exchange (ACE) approach first exemplified in *C. acetobutylicum* (Heap, et al., 2012). ACE is a genome modification approach which targets the *pyrE* or *pyrF* gene. These genes encode for the enzymes orotate phosphoribosyltransferase and orotidine 5'-phosphate decarboxylase, essential catalysts within the pyrimidine biosynthesis pathway. They also convert 5-fluoroorotic acid (5-FOA) to 5-fluoroorotidine monophosphate and then the toxic compound 5-fluorouridine monophosphate. As such, 5-FOA is toxic for bacteria with actively expressed *pyrE* and *pyrF* genes.

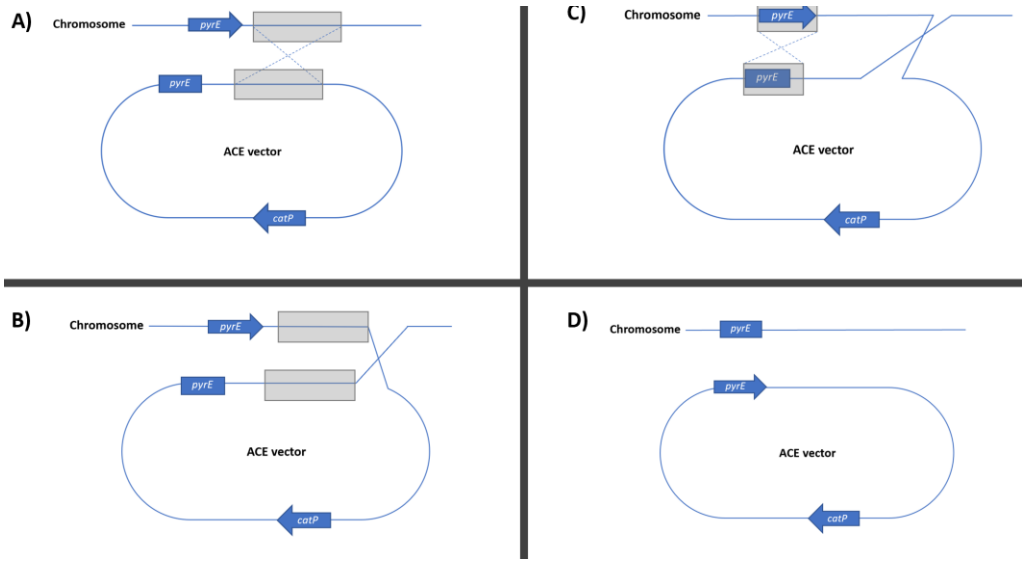


Figure 20 – Allele coupled exchange (ACE) schematic
Schematic outlining the *in vivo* progression of ACE through plasmid introduction (A), single crossover (B), double crossover (C) and resultant *pyrE* truncation (D).

The ACE process proceeds through two sequential crossover events between the regions of homology on an introduced ACE vector and the chromosome (**Figure 20**). To preferentially stimulate the first crossover event at the locus immediately downstream of *pyrE/F* a large (~1.2 kb) homologous sequence is used. To further encourage initial crossovers to occur, and visibly differentiate colonies in which this has occurred, a replication-deficient replicon is employed. These replicons, deemed pseudo-suicide, replicate less effectively than the chromosome. As such cell division can occur with a daughter cell unable to inherit an antibiotic-resistance providing plasmid, and therefore not viable when cultivated in the presence of this antibiotic. This reduces the overall growth rate of these plasmid maintaining bacteria under appropriate selection. Cells within which a crossover event has occurred will present as single-integrant clones whose growth is no longer limited by the plasmid replication allowing them to grow visibly quicker than those in which this recombination has not occurred.

A secondary crossover event is then achieved at the *pyrE/F* locus where a shorter homology arm within the ACE vector contains a 3' terminally truncated (~200 bp) version of the *pyrE/F* sequence. The dual functionality of *pyrE/F* results in distinct selectable phenotypes conferred by the functional and non-

functional version of this gene. The functional wild-type phenotype is sensitive to the presence of 5-FOA as well as growing in the absence of uracil within the media. The truncated inactive *pyrE* generated by a double crossover event no longer exhibits 5-FOA sensitivity and is a uracil auxotroph which cannot grow in the absence of supplementation. This selection approach can be reversed for repair of the gene, with cargo for chromosomal integration situated between the homology arms.

ACE is particularly prevalent due to its dual functionality affording a robust selection process and has been implemented in a number of acetogenic species including *C. difficile* (Ng, et al., 2013)*, C. autoethanogenum* (Annan, et al., 2019) and *E. limosum* ATCC 8486 (Millard, 2017). Similar targets which only confer singular phenotypic changes are also pursued, such as integral enzymes within amino acid biosynthesis including *hisI/*histidine, *pheA*/phenylalanine and *argH*/L-arginine. Other targets focus on toxicity in lieu of auxotrophy, metabolising otherwise innocuous precursors into toxic metabolites *in vivo.* The *codA* gene is one such target encoding for a cytosine deaminase, hydrolysing 5-fluorocytosine (5-FC) into toxic 5-fluorouracil (Cartman, et al., 2012).

4.1.2.2. Plasmid selection/counter-selection

While genes which encode distinct selectable phenotypes are targeted as sites of chromosomal disruption, they have also been implemented as plasmid-based selection tools. The two-stage allelic exchange approach relies on an initial recombination event integrating the allelic exchange vector into the chromosome. The selection marker for this stage is typically an antibiotic resistance gene within the vector backbone, such as *ermB* or *catP.* In some systems these markers can be within the introduced DNA sequence itself, acting as a single selection pressure for double crossover recombination. This approach was used to knock-out of multiple genes in *C. ljungdahlii* and replace them with the *ermC* gene conferring clarithromycin resistance (Leang, et al.,

2013). Marker-less knockouts are, however, broadly preferred as multiple genes can be targeted sequentially, and the cell's phenotype is otherwise not altered. There are two main mechanisms which, paired with these selection markers, stimulate chromosomal integration of the vector. The first of these methods, as outlined in ACE, is the use of a pseudo-suicide replicon. The limiting nature of these replicons confers observable reductions in growth rate and therefore affords the capacity for manual selection of faster growing integrants. The second of these approaches is the use of vectors which are completely incapable of replicating in the target organism, termed suicide vectors. As these rely on both DNA transfer and homologous recombination into the genome to confer resistance to the selection condition there is significant reduction in transformant frequency observed with these vectors.

The second recombination event which excises the integrated plasmid DNA can occur without plasmid counter-selection (Faulds-Pain & Wren, 2013), however issues of recombination frequency remain a significant hurdle. As such, a range of counter-selectable marker genes have been demonstrated within various *Clostridium* spp., including *pyrE*, *pyrF*, *codA*, *mazF* and *upp*. Counter selectable markers must demonstrate inducible toxicity by only stimulating cell death under specific conditions. One such marker demonstrated in *C. perfringens* is based on the galactose/GalK counter-selection system initially exemplified in *Myxococcus xanthus* (Ueki, et al., 1996). The *galK* gene employed was derived from *C. acetobutylicum* ATCC 824 and encodes a galactokinase, catalysing galactose-1-phosphate production from galactose. Galactose-1-phosphate accumulation is believed to incur toxicity through intracellular stress, which inhibits cell growth (Lee, et al., 2009). This system was initially used to generate an in-frame deletion of the *galKT* operon, with counter-selection for the vector and the wild-type operon concurrent. This mutant was then used as the appropriate background for further knockouts. *C. perfingens* contains a native galactose utilisation operon, leading to the assumption it contains passive/active uptake systems. Further to this, the *C. acetobutylicum* derived homologue was chosen due to both its phylogenetic proximity, thought to

increase the likelihood of good expression and expected functionality within *C. perfringens*, and sequence divergence to prevent homologous recombination (61.2%) (Nariya, et al., 2011).

The predictions of counter-selection system compatibility are aided by genomic and metabolic understanding, for example an organism which can grow on a particular substrate must contain the necessary passive or active uptake mechanism. This knowledge however is limited in novel strains. Further to this as seen within the application of *galKT* in *C. perfringens,* chromosomal inactivation of homologous genes is commonly required before a marker can be reapplied within an exogenous sequence. Other examples include ACE mediated disruption of *pyrE* in *C. difficile* prior to the implementation of *pyrE* as a 5-FOA induced plasmid counter-selection system (Ehsaan, et al., 2016). As such, many markers are not readily applicable within novel strains where genetic manipulation tools have not yet been developed.

Toxin genes, derived from toxin-antitoxin (TA) systems, are an alternate type of counter-selectable marker. As expression of such genes is inherently lethal to the cell under any condition, their use in allelic exchange systems is reliant on the availability of suitably strict inducible control systems. One such example employed within *Clostridium* is the *E. coli* derived type II TA system toxin gene *mazF,* as previously demonstrated within *B. subtilis* (Zhang, et al., 2006). This encodes an mRNA interferase which targets and cleaves mRNA at the ACA sequence, inhibiting protein synthesis and inducing bacterial stasis. Allelic exchange was performed in *C. perfringens* using a codon optimised *mazF* gene, under the control of the lactose inducible promoter system *bgaR*-P_{bgaL} (Al-Hinai, et al., 2012). This system has since been successfully used in *C. pasteurianum* (Sandoval, et al., 2015) and under arabinose inducible control in *C. cellulolyticum* (Zhang, et al., 2015). More recently the *C. difficile* 630 derived CD2517.1 toxin gene was similarly employed as an inducible counter-selection marker for the genetic modification of multiple *C. difficile* strains (Peltier, et al., 2020). Within this system the toxin expression was controlled by a tetracycline inducible promoter, with further mitigation of any leaky expression through

the concurrent integration of the RCd8 antitoxin gene and native constitutive promoter into the vector.

4.1.3. CRISPR-Cas genome editing systems in *Clostridium*

The development of CRISPR-Cas from initial identification in *E. coli* in 1987 (Ishino, et al., 1987) to widespread heterologous implementation is reminiscent of the development restriction endonucleases as ubiquitous modern molecular biology tools. Very similar to allelic exchange systems, CRISPR-Cas mediated genome editing supplies a comparable homologous recombination cassette and essentially employs the CRISPR-Cas complex as a counter-selection tool against the unmodified genome sequence. As an *in vivo* genome manipulation tool for use in prokaryotes, vector based CRISPR-Cas systems rely on three key features: a single guide RNA (sgRNA), a Cas endonuclease and a donor region for homologous recombination. The size of the homology region of the vector is dependent on both the homologous recombination efficiency of the strain and the desired genomic change. Most clostridial studies employ two homology arms totalling typically 1-2 kb. Recently a modified system has been demonstrated in *C. beijerinckii* which does not include a homologous region, instead using a fused cytidine deaminase and uracil DMA glycosylase inhibitor to repair the DNA through specific base pair substitutions. This is reminiscent of non-homologous end joining exploited in eukaryotes to generate potential missense or null mutations (Li, et al., 2019).

The sgRNA region within CRISPR-Cas vectors is a simplification of the crRNA system associated with the given enzyme, in which fusion of the crRNA and the tracrRNA produces a single RNA molecule. This sgRNA forms a complex with the Cas endonuclease for direction of nucleolytic activity. The specificities of the sgRNA and Cas region are dependent on the CRISPR-Cas9 system employed, with the most prevalent *S. pyogenes* derived system requiring one *cas9* endonuclease with specificity for the PAM sequence 5’-NGG-3’. This genomic PAM sequence is also an integral feature for functional CRISPR-Cas

operation, necessary directly adjacent to the protospacer sequence targeted by the sgRNA to enable the binding of the complex. 5'-NGG-3' is a relatively prevalent sequence within genomes, even within the *Clostridium* genus where low GC percentages are characteristic. This ensures there are appropriate protospacer targets across all regions, with 135 identified within the 2943 bp *CcaP717830P* gene targeted within this study (**4.2.3.2**).

The first report of CRISPR-Cas9 mediated genome editing within an acetogen, occurred within *C. ljungdahlii* (Huang, et al., 2016). Subsequent development of the system for application within *C. difficile* highlighted a frameshift mutation which prevented expression of the RuvCI nucleolytic domain (Ingle, et al., 2019). This nucleolytic domain is one of two present in *S. pyogenes* Cas9 and as such the resultant enzyme would be expected to only demonstrate single strand cleavage. These single strand cleaving Cas systems are widely called nickases (nCas9) with the single strand nicks (SSN) typically less toxic in prokaryotes than the DSBs generated by the wild-type Cas9. This lowers the inherent toxicity of exogenous Cas system expression, and as such nCas9 usage is one prominent tactic used to ensure successful system transfer in less amenable strains. nCas9 based exogenous systems have hence been employed in a number of clostridial strains, including *C. difficile* (Ingle, et al., 2019), *C. ljungdahlii* (Huang, et al., 2016), *C. acetobutylicum* and *C. beijerinckii* (Li, et al., 2016; Li, et al., 2019). While SSN as opposed to DSB generation does have its benefits, there is also a resultant drop in genome editing efficacy observed with nCas9 systems. This leads to more time-consuming screening with longer periods of selective growth required to identify and isolate mutants under the weaker selective pressure.

Another prominent approach to mitigation of Cas9 toxicity in clostridial studies is the use of strict inducible expression. The exact inducible expression system employed for *cas9* control varies between strain, based largely on the efficacy of system repression. Within *C. autoethanogenum* a series of tetracycline-inducible promoters were tested and implemented when constitutive expression was not viable for CRISPR-Cas9 mediated genome editing (Nagaraju,

et al., 2016). A similar tetracycline-inducible system was applied for the CRISPR-Cas9-mediated deletion of three genes in *E. limosum* (Shin, et al., 2019) and a lactose inducible CRISPR-Cpf1 system was employed within *C. beijerinckii* (Zhang, et al., 2018). Some systems further employ a two-plasmid based approach, such as the anhydrotetracycline-inducible promoter-controlled genome editing of *C. acetobutylicum* and *C. beijerinckii* (Wasels, et al., 2017; Diallo, et al., 2020). Recent work by Ines Cañadas has exemplified a new inducible expression system for use within *Clostridium*. This system places *cas9* expression under stringent control using a high strength constitutive promoter paired with RBS sequestration within a theophylline inducible riboswitch. This system was demonstrated in *C. sporogenes*, *C. pasteurianum*, *C. difficile* and *C. botulinum,* four non-acetogenic Clostridial species.

A third approach to overcoming CRISPR-Cas toxicity concerns in tool development, is the exploitation of a strains' endogenous system in lieu of exogenous system introduction. Endogenous systems are reliant on the native expression levels of the CRISPR-Cas system within the cell, hence causing only limited metabolic burden upon vector introduction. Also, in their native state endogenous systems do not exhibit any inherent toxicity, only proving lethal when self-targeting sgRNA sequences are introduced. Following system characterisation and PAM identification, the chosen sgRNAs are introduced on vectors containing a synthetic array sequence and appropriate homologous donor region. As these vectors do not also need associated Cas genes, they are typically much smaller than exogenous alternatives, and may be more suited to strains in which only limited DNA transfer frequency has been established.

One such example was the exploitation of the endogenous Type I-B CRISPR-Cas system in *Clostridium tyrobutyricum* (Zhang, et al., 2018). The inducible CRISPR array vectors developed in this study were introduced into this strain at a rate of 1.7-18.2 CFU/mL per donor (dependent on the 38 bp spacer sequence used), where extensive previous exogenous system approaches had yielded zero transconjugants. A comparable increase in transfer frequency has been observed in other clostridia, such as *C. pasteurianum* ATCC 6013 in which the

transformant numbers for endogenous vectors were quadruple those of the heterologous CRISPR-Cas9 vectors (Pyne, et al., 2016). Interestingly the development of the endogenous CRISPR-Cas system for genome editing in *C. tyrobutyricum* facilitated subsequent Type I RM system and native plasmid disruption. These changes improved the DNA transfer rates 3.7-fold and 15.3-fold respectively, enabling subsequent implementation of exogenous CRISPR-Cas9/Cpf1 based genome editing (Zhang, et al., 2020).

4.1.4. Aims:

The aim of this study was to develop genomic manipulation tools for use within the *C. carboxidivorans* C94 strain generated in Chapter 3. This began with fundamental genetic tool validation, including an inducible promoter system and reporter gene. Further to this an allelic exchange approach was implemented, and tailored CRISPR-Cas9 system developed. The CRISPR-Cas9 system developed in this study was applied to demonstrate a proof-of-concept gene knockout of *CcaP7ORF17830P*.

4.2. Results

4.2.1. Establishing key genetic tools in *C. carboxidivorans*

4.2.1.1. *ermB* resistance marker

To expand the available genetic toolset and facilitate the execution of a chloramphenicol acetyltransferase (CAT) reporter assay, the viability of the *ermB* plasmid selection marker was assessed. This was done by confirming the sensitivity of *C. carboxidivorans* C94 to the antibiotic clarithromycin, for which *ermB* confers resistance, before attempting to transform *C. carboxidivorans* with pMTL8325y-v1 (**Figure 22**). No significant hurdle to DNA transfer was expected due to the lack of known RM motifs identified in initial sequence analysis of the ErmB cassette.

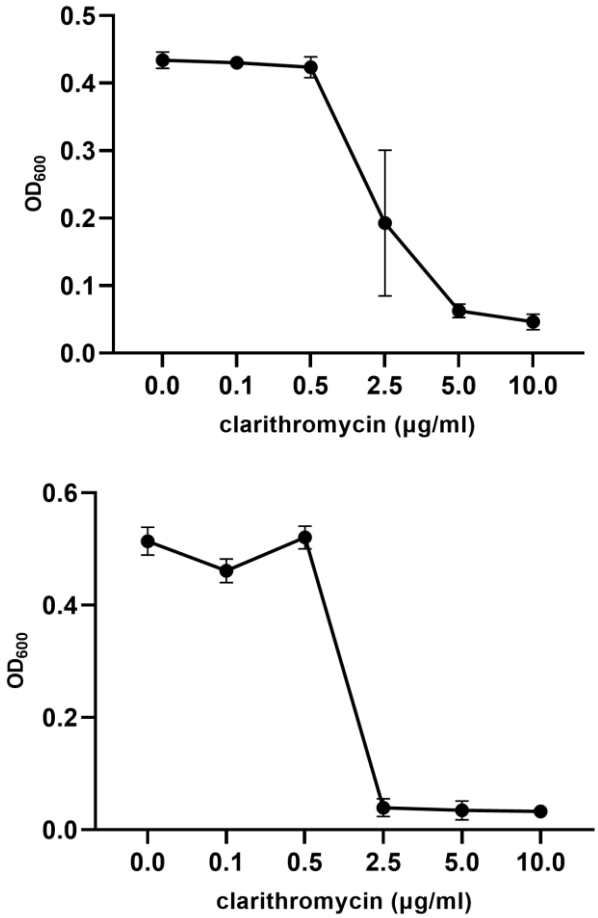


Figure 21 – *C. carboxidivorans* C94 growth in the presence of clarithromycin (0-10 µg/mL)
OD_{600} of biological quadruplicate *C. carboxidivorans* C94 YTAFM broth cultures supplemented with 0-10 µg/mL clarithromycin, recorded at 24 and 72 hours of growth. Error bars represent standard deviation.

A toxicity assay was performed to quantify the sensitivity of *C. carboxidivorans* C94 to the antibiotic clarithromycin, as detailed in **2.12**. In brief four exponential YTFAM broth cultures were normalised to an OD_{600} of 0.05, into 250 µL YTFAM broth supplemented with a range of clarithromycin concentrations (0-10 µg/mL). OD_{600} readings were then taken following 24, 48 and 72 hours of anaerobic incubation at 37°C. In the absence of the *ermB* resistance marker, inhibition of *C. carboxidivorans* growth by clarithromycin occurred at concentrations greater than 2.5 µg/mL, however, following sustained growth some reduced sensitivity was observed with 50% of replicates at 2.5 µg/mL demonstrating growth at 72 hours (**Figure 21Error! R eference source not found.**). As such a clarithromycin concentration exceeding

5 µg/mL (6 µg/mL) was used to select for *ermB* containing plasmids within this study.

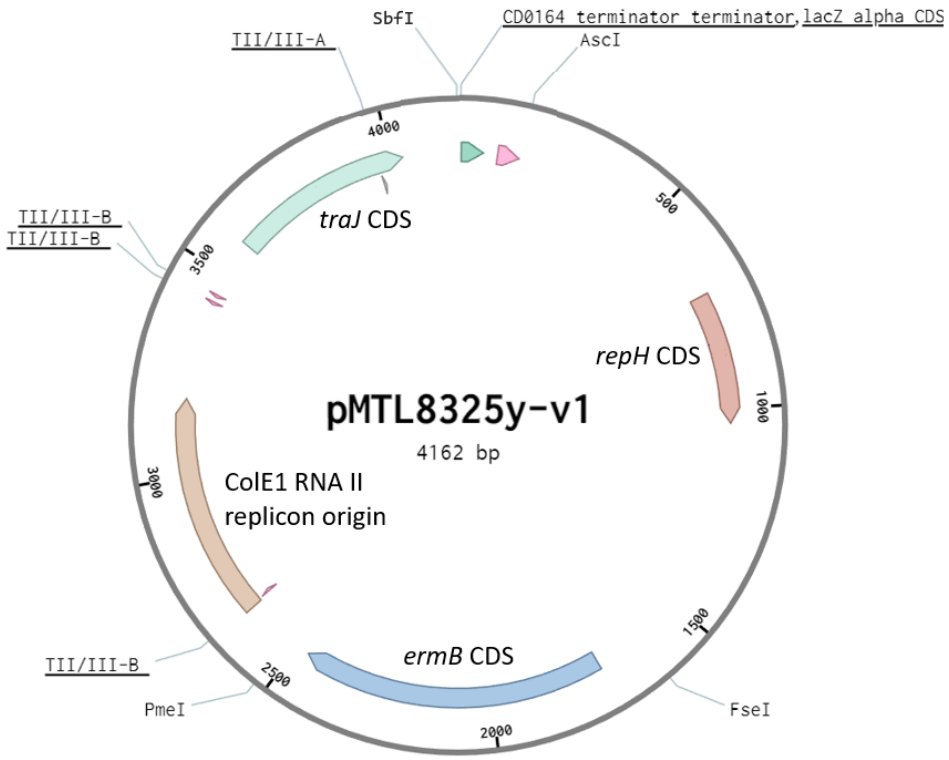


Figure 22 – pMTL8325y-v1 plasmid map

Schematic of pMTL8325y-v1, derived from pMLT8315y-v1 with the ErmB cassette between the FseI and PmeI cut sites, in lieu of the CatP cassette. Also annotated are the four residual RM motif sequences within the vector.

pMTL8325y-v1 was created as detailed in **2.9.8**, with the ErmB resistance marker cassette of the pMTL80000 vector series replacing the modified CatP module in pMTL8315y-v1. This vector was then used to transform chemically competent sExpress_TIII, before transfer was attempted into *C. carboxidivorans* C94 by conjugation, as detailed in **2.3.9**. Transconjugants were observed following 48 hours of incubation on clarithromycin and D-cycloserine selective YTAFM plates. No observable variance in conjugation efficacy was observed between pMTL8315y-v1 and pMTL8325y-v1.

4.2.1.2. *catP* reporter assay

A *catP* reporter assay was conducted to determine the expression levels derived from key promoter sequences within *C. carboxidivorans* C94. To

maximise transfer efficiency, the truncated CatP module employed within pMTL8315y-v1 was used as a template for a 195 bp shorter *catP* reporter gene cassette. This was amplified by PCR and assembled with various promoter sequences of note into pMTL8325y-v1. The generation of these reporter vectors is detailed in **2.9.8**. Following transfer of these reporter vectors (pMTL8325y-v1-P****-CatP(trunc) in which **** is replaced with the appropriate promoter's name) into *C. carboxidivorans* using *E. coli* sExpress_TIII (**3.2.6.4**), four transconjugant colonies of each were picked into YTAFM broth supplemented with clarithromycin and D-cycloserine. A CAT assay was then performed with these samples as detailed in **2.11**. In brief cell mass was harvested in mid-exponential phase and cell pellets lysed before the CatP activity of the resultant supernatant was determined spectrophotometrically using a Tecan M1000 Pro plate reader. Five readings were taken across each well and averaged, further to the biological quadruplicates. All values were also adjusted according to blank controls and the BCA assay determined protein content of each sample.

Three constitutively expressed promoters were tested: P_{araE} and P_{thl} derived from *C. acetobutylicum* ATCC 824 (Heap, et al., 2009; Ingle, et al., 2019) and P_{fdx} derived from *C. sporogenes* NCIMB 1069 (Heap, et al., 2009). Both P_{thl} and P_{fdx} have previously been implemented in a number of clostridial species, with P_{thl} and P_{araE} used within the pMTL40000 vector CRISPR-Cas9 system exemplified within *C. difficile* (Ingle, et al., 2019). A further modified P_{thl} promoter was also tested, in which a 5' truncation on sequence reduced the initial 150 bp to 72 bp to determine the essentiality of the full sequence. The full P_{thl} was the strongest promoter tested, with a CAT assay (**2.11**) activity of 103.07 U per mg of isolated protein, markedly reduced in the truncated version to 32.05 U/mg (**Figure 23**) (P value <0.0001). P_{fdx} was the second strongest tested at 79.19 U/mg. P_{araE} had an expression level of 38.26 U/mg. A negative control denoted as pMTL8325y-v1-NoProm-CatP(trunc) was used to determine basal chloramphenicol acetyltransferase activity in a strain in which the *catP* gene within pMTL8325y-v1 was not preceded by an RBS or associated promoter,

instead flanked by a *C. difficile* derived CD0164 terminator sequence. This basal level was recorded as 1.64 U/mg indicating extremely low levels of chloramphenicol acetyltransferase activity in the sample. As such all promoters stimulated statistically significant increases in CatP activity compared to this, with P values of <0.0001, 0.0075, 0.0004, and <0.0001 when using Welch's t-test.

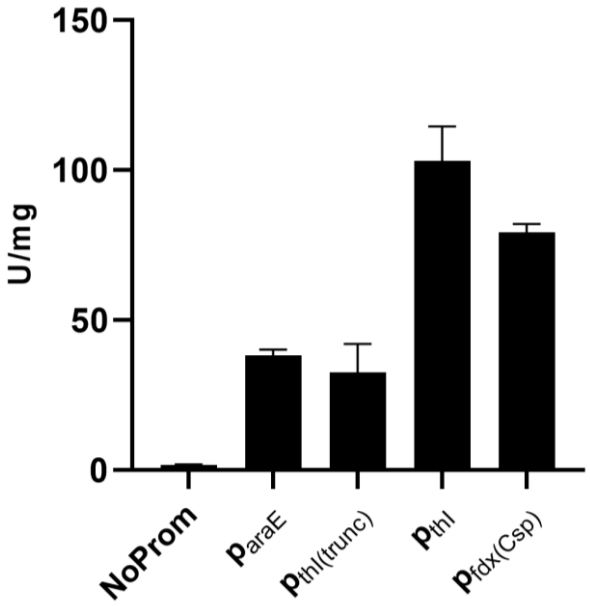


Figure 23 – Relative promoter strength of four constituitive promoters in mid-exponential *C. carboxidivorans* C94.
CAT assay derived CatP protein activity levels of *C. carboxidivorans* C94 maintaining five pMTL8325y-v1 based reporter vectors, with catP expression under the control of four different promoter sequences. Resultant CatP values determined according to enzymatic activity per mg protein, determined by cell lysis and 5-thio-2-nitrobenzoate production of the lysate measured according to OD_{412nm}. Final values adjusted according to BCA assay derived total protein contents. Performed in biological triplicate with error bars indicative of standard deviation.

4.2.1.3. Theophylline growth impact

Further to this constitutive promoter validation, the stringency of the P_{fdx} based theophylline inducible riboswitch system published by Ines Canadas (2019) was also tested. The repression of gene expression within this system functions through the sequestering of a synthetic ribosomal binding site within the mRNA stem-loop structure of the riboswitch. This riboswitch directly follows a constitutive promoter, but due to this sequestration gene sequence translation is inhibited. This repression is relieved through changes in the secondary structure confirmation induced by binding of theophylline to the aptamer. The

structural changes expose the RBS allowing higher rates of translation to occur. As such the impacts of theophylline supplementation on *C. carboxidivorans* C94 growth was assessed as detailed in **2.12**. In brief four independent exponential *C. carboxidivorans* C94 cultures were standardised and inoculated into YTAFM with theophylline concentrations ranging from 0 mM to 10 mM. These cultures were grown anaerobically for 24, 48 and 72 hours before the OD_{600} of each replicate recorded (**Figure 24**). This assay highlighted minimal reduction in the growth of *C. carboxidivorans* C94, with the greatest theophylline induced negative growth impacts observed at 10 mM. This was consistent across all three timepoints. Notably the growth appeared greater with the presence of theophylline concentrations than in its absence, wherein the OD_{600} only reached 0.357. This was an unexpected result, but due to the lack of negative impacts of theophylline within this inductive range, validation of the riboswitch proceeded utilising these assayed concentrations.

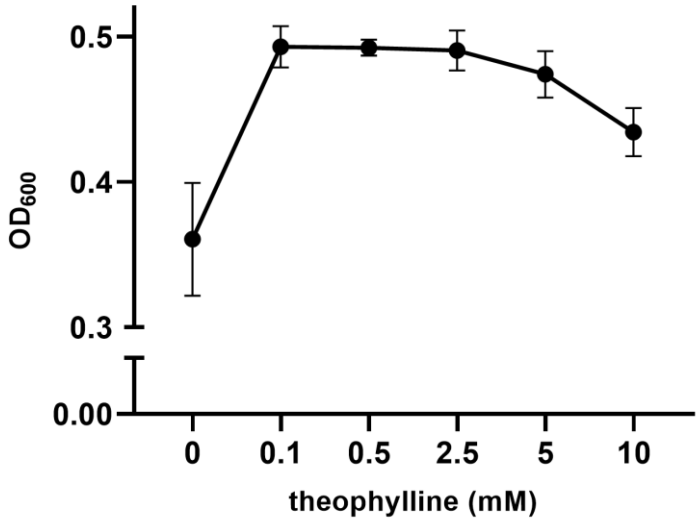


Figure 24 – Impacts of 0-10 mM theophylline on *C. carboxidivorans* growth.
Average OD_{600} values and standard deviation of triplicate *C. carboxidivorans* C94 cultures grown for 24 hours in YTAFM broth supplemented with 0-10 mM theophylline.

4.2.1.4. RbE*-$P_{fdx(Csp)}$ *catP* reporter assay

The riboswitch sequence denoted as E* is widely regarded as the best of the theophylline-dependent riboregulators studied within Gram-positive bacteria,

due to high repression, low leakage, and good dynamic range (Rudolph, et al., 2015). As such, the vector pMTL8325y-v1-$P_{fdx(Csp)}$-RbE*-CatP(trunc) was constructed to assess the stringency and viability of the inducible expression system within *C. carboxidivorans* C94 (**Figure 25**). This vector contains one single nucleotide change within the promoter sequence from cytosine to adenosine at the 80th position. This change was made to disrupt a TII/III-C RM motif sequence.

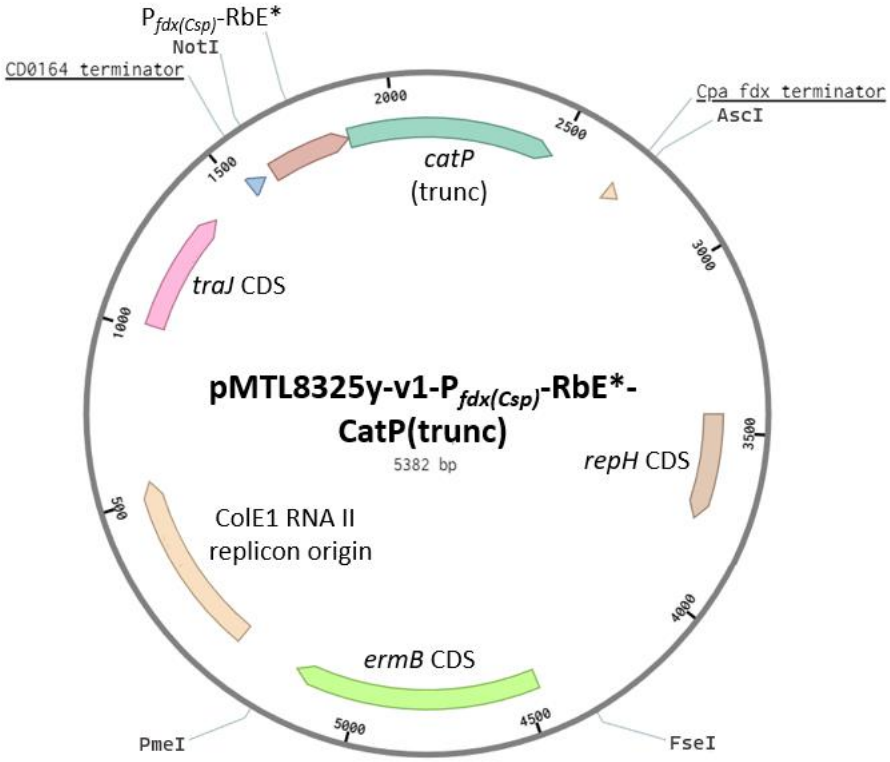


Figure 25 – pMTL8325y-v1-$P_{fdx(Csp)}$-RbE*-CatP(trunc) plasmid map
Schematic detailing key sequence features of the pMTL8325y-v1-$P_{fdx(Csp)}$-RbE*-CatP(trunc) reporter vector, including truncated CatP module, theophylline responsive riboswitch E* and P_{fdx} promoter sequence originally derived from *C. sporogenes*.

Once assembled this vector was introduced to *C. carboxidivorans* C94 by conjugation with sExpress_TIII. Four of the resultant transconjugants colonies were picked independently into YTAFM supplemented with clarithromycin and D-cycloserine. Following 24 hours of growth these were each passaged into YTAFM broth with clarithromycin and theophylline concentrations of 0, 1.25, 2.5, 5 and 10 mM. A CAT assay was performed using these cultures as previously described. The resultant CAT activity at each concentration shows

expected scalability of induction, with highest activity per mg of protein recorded as 89.25 U/mg at 10 mM and comparable induction at 2.5 mM and 5 mM of 67.50 and 67.68 U/mg respectively (**Figure 26**). The incremental differences between 0 and 2.5 mM are statistically significant (0.0052 and 0.0205), but not between 2.5 mM and 10mM (0.9796 and 0.0529). All were significantly elevated when compared to the uninduced samples (0.0052, 0.0009, 0.0011 and 0.0014 respectively). In an uninduced state the $P_{fdx(Csp)}$-RbE* promoter system still generated an elevated CAT activity at 8.92 U/mg. This activity is more than five times higher than the previously quantified no promoter control, indicating a very low level of expression even in a repressed state. This is expected as the reporter gene is no longer preceded solely by a terminator sequence, but a promoter sequence where even with riboswitch sequestration some minimal binding and expression may occur.

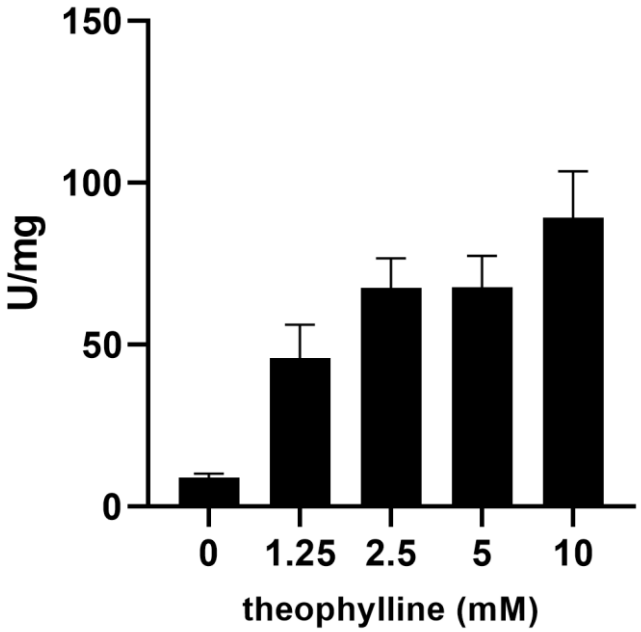


Figure 26 – $P_{fdx(Csp)}$-RbE* derived expression levels within *C. carboxidivorans* C94 upon induction with 0-10 mM theophylline.

CAT assay derived CatP protein levels of *C. carboxidivorans* C94 maintaining pMTL8325y-v1-$P_{fdx(Csp)}$-RbE*-CatP when supplemented with 0-10 mM theophylline. Resultant CatP values determined according to enzymatic activity per mg protein, determined by cell lysis and 5-thio-2-nitrobenzoate production of the lysate measured according to OD_{412nm}. Final values adjusted according to BCA assay derived total protein contents. Performed in biological triplicate with error bars indicative of standard deviation.

4.2.2. Genome modification of *C. carboxidivorans* through allelic exchange

The first genomic modification approach attempted within *C. carboxidivorans* C94 was allelic exchange. The viability of suicide integrant vector transfer was first assessed, with subsequent validation of a *C. difficile* derived inducible toxin system as a plasmid counter-selection method. This was followed by targeting of the *CcaP7ORF17830P* gene (otherwise denoted as TIII.2R) for attempted knockout. This gene was previously identified as the restriction endonuclease responsible for the nucleolytic activity of the TIII.2 RM system. An alternative pseudo-suicide replicon-based approach was explored concurrently by Chris Humphreys (SBRC, Nottingham).

4.2.2.1. Generation of Δ*ccaP7ORF17830P*::pMTL8015y-v1-TIII.2R-SC* *C. carboxidivorans* single crossover integrants

Low rates of DNA transfer and homologous recombination are prohibitive to the successful implementation of non-replicative vector chromosomal integration. While basal levels of DNA transfer of 3 x 10^4 (CFU/conjugation) had been achieved with the evasive vector pMTL8315y-v1, recombinant sExpress donors and *C. carboxidivorans* C94, the barrier posed by efficient chromosomal integration was uncharacterised. To test this three pMTL8315y-v1 derived vectors were constructed in which DNA fragments of approximately 500 bp, homologous to the *CcaP7ORF17830P* gene, were introduced in lieu of the pCB102 Gram-positive replicon. These vectors were constructed by HiFi with PCR amplified DNA fragments from *C. carboxidivorans* gDNA and linearised pMTL8315y-v1, as detailed in **2.9.6**. pMLTL8015y-v1-TIII.2R-SC1 included an additional TII/III-C RM motif, pMLTL8015y-v1-TIII.2R-SC2 one TII/III-C and one TII/III-B, and pMLTL8015y-v1-TIII.2R-SC3 none, as indicated in **Table 26**. Following sequence confirmation each vector was then used to transform the chemically competent *E. coli* sExpress_TIII donor cells.

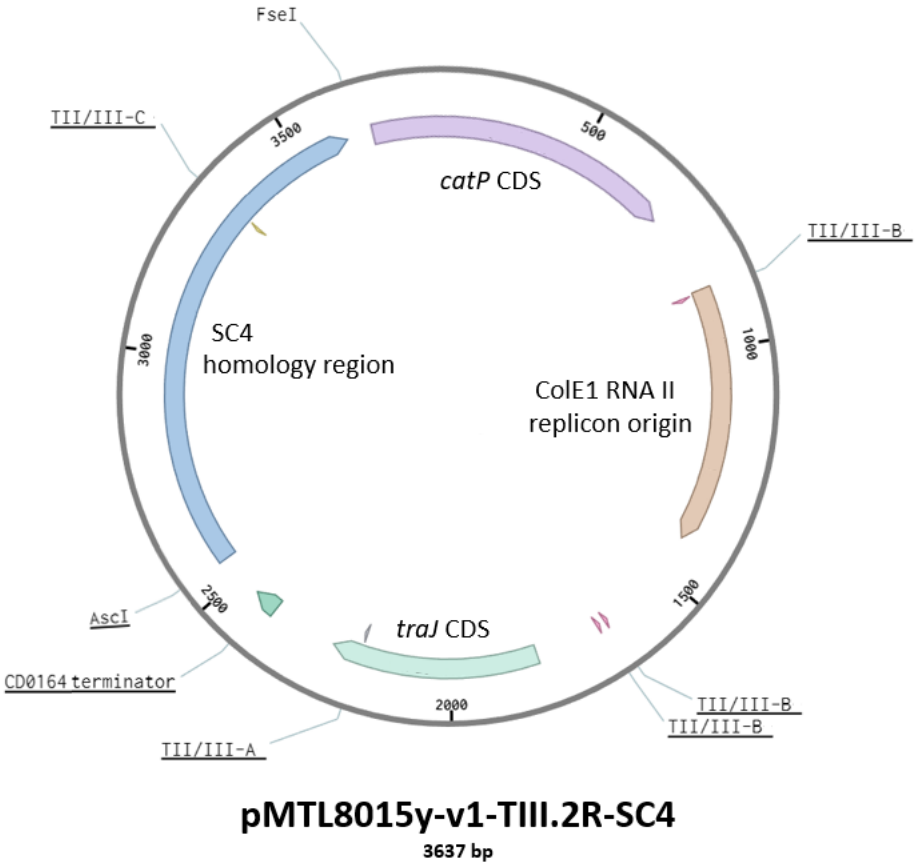


Figure 27- pMTL8015y-v1-TIII.2R-SC4 plasmid map
Schematic of the integrative vector pMTL8015y-v1-TIII.2R-SC4, in which a 1058 bp homologous sequence from the *CcaP7ORF17830P* gene of *C. carboxidivorans* was introduced between the AscI and FseI cut sites in lieu of the Gram-positive replicon.

Conjugation of a single *C. carboxidivorans* C94 culture was performed in triplicate with each vector as well as pMTL8315y-v1, with independent sExpress_TIII donor cultures at the point of initial inoculation. The conjugation was performed as detailed in **2.3.9** with total bacterial slurry from the mating plates spread across four YTAFM plates with thiamphenicol and D-cycloserine selection. Following 120 hours no transconjugants were observed for any of the three suicide vectors. Three further vectors were subsequently constructed with increased regions of homology to improve rates of *in vivo* homologous recombination. pMTL8015y-v1-TIII.2R-SC4, pMTL8015y-v1-TIII.2R-SC5 and pMTL8015y-v1-TIII.2R-SC6 were constructed in an identical manner to the previous vectors, and contained homologous regions of 1058 bp, 922 bp and 704 bp respectively (**Figure 27**). The introduction of these vectors into the recombinant *E. coli* donor sExpress_TIII and triplicate conjugation into *C.*

carboxidivorans C94 was then repeated. A total of 16 transconjugants were observed following 72 hours of selective growth. Nine transconjugants were generated by pMTL8015y-v1-TIII.2R-SC4 which contained the longest homologous region.

Table 26 – Resultant transconjugant CFUs of TIII.2R integration plasmids.
Integration plasmids with total *CcaP7ORF17830P* homologous sequence lengths integrated in lieu of a Gram-positive replicon. Also indicated are the further RM motif sequences introduced to each vector through the homologous region, and the resultant transconjugant CFUs observed through triplicate conjugations into *C. carboxidivorans* C94.

Plasmid	Homologous length (bp)	RM motifs	CFUs		
pMTL8015y-v1-TIII.2R-SC1	519	TII/III-C x1	0	0	0
pMTL8015y-v1-TIII.2R-SC2	583	TII/III-C x1 TII/III-B x1	0	0	0
pMTL8015y-v1-TIII.2R-SC3	495		0	0	0
pMTL8015y-v1-TIII.2R-SC4	1058	TII/III-C x1	5	2	2
pMTL8015y-v1-TIII.2R-SC5	922	TII/III-A x1	3	2	1
pMTL8015y-v1-TIII.2R-SC6	704	TII/III-C x1 TII/III-B x1	0	0	1

To confirm integration of the plasmid sequence at the desired locus, a PCR was performed with primers framing *CcaP7ORF17830P*. Two pMTL8015y-v1-TIII.2R-SC4 transconjugant colonies, and one of each of the other vector derived transconjugants, were picked into YTAFM media with thiamphenicol and D-cycloserine selection. These were grown at 37^{o}C for 48 hours before 10% (v/v) DMSO stocks were prepared, and the cultures removed from the cabinet. The residual 4.1 mL of each culture was then pelleted by centrifugation prior to genomic DNA extraction (**2.4.2**). 1 µL of each 100 µL genomic DNA elute,

alongside 1 µL of previously isolated *C. carboxidivorans* C94 gDNA, was used as a template for PCR amplification with the primers TIII.2R Scr Fw2 and TIII.2R Scr Rv2. These primers bind 1213 bp downstream and 796 bp upstream of *CcaP7ORF17830P* respectively. A modified Q5 High-Fidelity polymerase protocol was utilised to confidently amplify the longer regions, with an initial denaturation step of five minutes at 98°C, a reduced extension temperature of 70°C and increased extension time. The cycle number was also reduced from 35 to 25 cycles, allowing for observable band intensity variation between the wild-type band at 4952 bp, still present due to the inherent instability of the integrants, and the integrant band (**Figure *28***). All screened transconjugants exhibited the expected ~8600 bp band indicative of vector backbone integration. This indicated that levels of homologous recombination within *C. carboxidivorans* were not so low as to prove prohibitive to suicide vector based genetic systems, provided homologous regions exceeded 700bp, with a greater transconjugant yield when the sequence was approximately 1 kB.

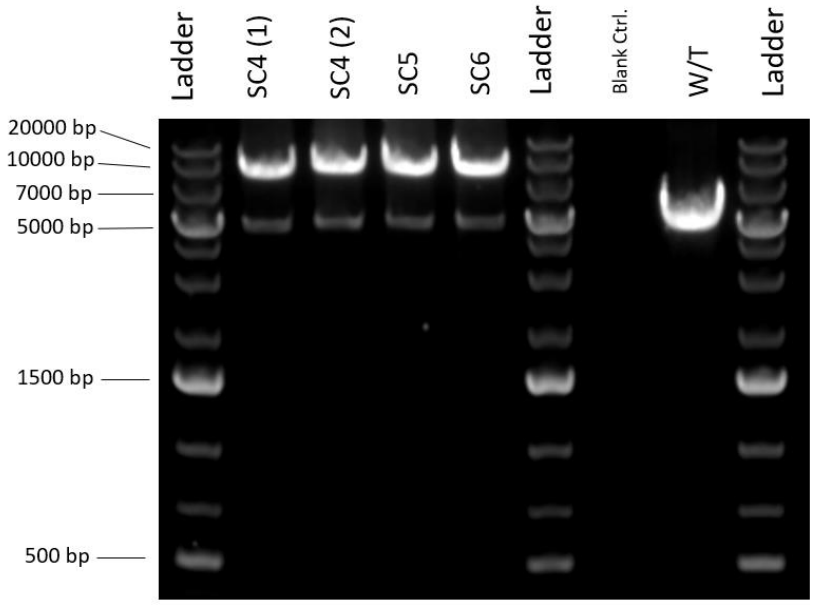


Figure 28 – Screening PCR of suicide plasmid transconjugant colonies generated by pMTL8015y-v1-TIII.2R-SC* vector transfer into *C. carboxidivorans* C94.

Gel electrophoresis analysis of colony PCR products generated by amplification of the *CcaP7ORF17830P* locus using the primers TIII.2R Scr Fw2 and TIII.2R Scr Rv2. 4952 bp band generated from wild-type locus amplification, and 3600/3700 bp increase to approximately 8600 bp for plasmid integration. Ladder: O'generuler 1 kb plus. SC4-6: transconjugant colonies. Blank Ctrl.: identical PCR mix with dH_2O template. W/T: identical PCR mix with *C. carboxidivorans* C94 gDNA template.

4.2.2.2. Assessing CD2517.1 plasmid counter-selection

Following confirmation of successful single crossover plasmid integration at *CcaP7ORF17830P*, an allelic exchange vector was assembled as detailed in **2.9.7**. This vector employed inducible expression of the CD2517.1 toxin gene derived from *C. difficile* 630 as a counter-selection marker. This CD2517.1 toxin was initially identified as part of a type I toxin-antitoxin (TA) system. This protein exhibits characteristic features associated with a membrane disruption toxicity strategy with a hydrophobic *N*-terminal region, a positively charged tail and relatively small size. This TA system has recently been exemplified as a counter-selectable marker for the genetic modification of not only *C. difficile* 630, but also other non-native *C. difficile* ribotype 027 strains (Peltier, et al., 2020).

As the sole inducible promoter system validated in *C. carboxidivorans* within this study, the CD2517.1 gene was placed under the control of $P_{fdx(Csp)}$-RbE*. Antitoxin activity in type I TA systems relies on an antisense RNA approach, in which the transcribed RNA of the antitoxin binds the toxin mRNA preventing expression through steric hinderance and stimulating transcript degradation. The associated antitoxin sequence of CD2517.1 (RCd8) is also present within the plasmid, under control of the native promoter of the original chromosomal locus. The presence of this constitutive basal level of antitoxin transcription was regarded as essential in non-native bacterial application due to the potential lethality of any “leakiness” of the inducible toxin expression system (Peltier, et al., 2020). The elevated basal expression level of the $P_{fdx(Csp)}$-RbE* compared to the no promoter control observed in the CAT assay within *C. carboxidivorans* C94 indicates the necessity of this within our system. As such a $P_{fdx(Csp)}$-RbE*-CD2517.1-RCd8 synthetic DNA region was commissioned. This region introduces one TII/III-C motif and one TII/III-B motif, within the toxin and antitoxin sequences respectively.

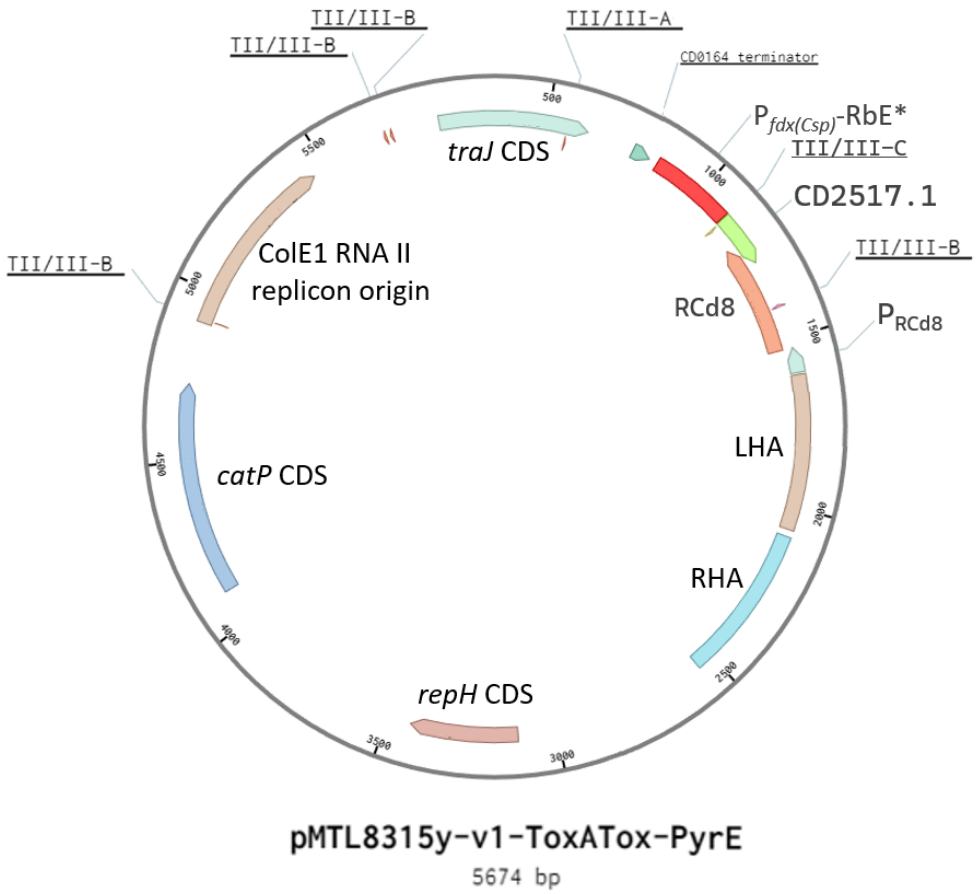


Figure 29 – pMTL8315y-v1-ToxATox-PyrE plasmid map
Schematic of the initial vector implementing the CD2517.1 toxin gene, alongside the RCd8 antitoxin and native promoter (P_{RCd8}). This module assembled in the pMTL8315y-v1 backbone, which utilised pCB102 (*repH*) as the Gram-positive replicon.

Initial assembly of this inducible system within pMTL8315y-v1 was performed by Chris Humphreys, concurrently pursuing a *C. carboxidivorans* amenable allelic exchange system based on use of a pseudo-suicide replicon and disruption of the *pyrE* gene. As such this initial assembly generated the pMTL8315y-v1-ToxATox-PyrE vector which incorporated not only the TA system, but also two homology arms (totalling 939 bp) framing the *pyrE* gene within the *C. carboxidivorans* chromosome (**Figure 29**). During repeated cloning attempts the $P_{fdx(Csp)}$-RbE*-CD2517.1-RCd8 region was never successfully integrated with 100% sequence integrity; all *E. coli* 10-beta transformants screened had acquired SNPs within the toxin gene. This resistance to toxin gene integrity was attributed to deleterious levels of toxicity of CD2517.1 within the *E. coli* cloning strain, despite the riboswitch repressed basal expression. Therefore, the CD2517.1 sequence carried forward for

subsequent testing had a single nucleotide deletion of the first base, adenine. This excision resulted in the generation of a TTG codon with the now adjacent thymine residue, an alternate start codon within *E. coli* and a number of prokaryotes (Hecht, et al., 2017; Belinky, et al., 2017). This vector was transferred into *C. carboxidivorans* C94 by conjugation using the sExpress_TIII recombinant *E. coli* donor strain, and subsequently stocked at -80^{o}C prior to downstream assessment.

To assess the efficacy of the toxin system as a counter-selectable marker the rates of Tm^{R} colonies following theophylline induction were tested. This was initially performed by inoculation of an exponential phase *C. carboxidivorans* C94 pMTL8315y-v1-ToxATox-PyrE culture (with thiamphenicol and D-cycloserine selection), into three YTAFM cultures supplemented solely with 2.5 mM theophylline. These were grown anaerobically at 37^{o}C for 24 hours before 100 µL of the cultures were spread onto YTAFM plates with no selection, and YTAFM plates with 7.5 µg/mL thiamphenicol. This replica plating was also performed with 100 µL of samples diluted at 10^{-1}, 10^{-2}, and 10^{-3} in PBS. Following 48 hours of growth, CFUs were counted across these plates. Comparison of CFUs between these two conditions indicated a plasmid loss of 99.8% (± 0.04%) with only 0.02% of viable colonies exhibiting Tm^{R} after 24 hours of 2.5 mM theophylline induction.

Another method was also used to assess the efficacy of plate-based induction, with the initial selective pMTL8315y-v1-ToxATox-PyrE *C. carboxidivorans* C94 culture directly plated on 2.5 mM theophylline inductive YTAFM plates. To ensure observation of distinct CFUs was possible this was also performed with dilutions of culture in PBS reaching 10^{-4}. These dilutions were concurrently plated on non-selective YTAFM plates, permitting comparative CFU counts between the presence and absence of theophylline. This analysis indicated that 99.17% of cells were not viable on the theophylline supplemented plates. 100 colonies from these induced theophylline plates were then replica plated to thiamphenicol selective and non-selective plates. Following 96 hours no growth was evident from any of the screened colonies on the thiamphenicol

selective plates. This result appeared to further support the application of the inducible toxin system as a robust counter-selection method to stimulate plasmid loss, wherein plasmid loss in this assay was deduced from restored thiamphenicol sensitivity through presumed loss of the *catP* sequence.

4.2.2.3. Targeted disruption of CcaP7ORF17830P with pMTL8015y-v1-ToxATox-TIII.2R

Following validation of the efficacy of the CD2517.1 TA system as a counter-selectable marker, a non-replicative integration vector was generated through excision of the Gram-positive replicon. The vector was also modified to target *CcaP7ORF17830P* and as such the homology arms were replaced with two new, approximately 1000 bp homology arms framing the sequence ATGAATGTAA. This central sequence between the homology arms is the first five bases of the *CcaP7ORF17830P*, and terminal five. The integration of these homology arm sequences introduced one TII/III-A motif sequence, five TII/III-B and four TII/III-C (**Figure 30**). The targeted *CcaP7ORF17830P* gene was identified by REBASE® analysis as the restriction endonuclease of the second Type III system, putatively assigned to the AGAAGC sequence (TII/II-A). The assembly of this pMTL8015y-ToxATox-TIII.2R vector is detailed in **2.9.7**.

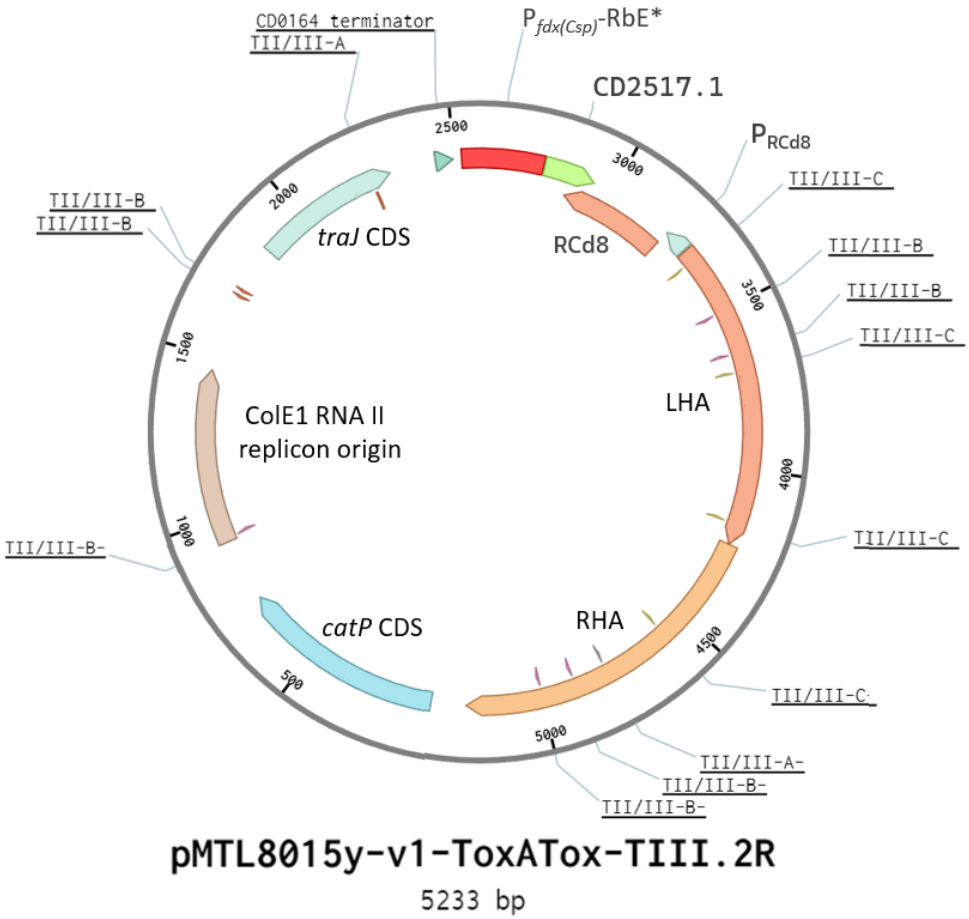


Figure 30 – pMTL8015y-v1-ToxATox-TIII.2R plasmid map
Schematic of the pMTL8015y-v1-ToxATox-TIII.2R plasmid sequence with key features annotated, including the $P_{fdx(Csp)}$-RbE* riboswitch and promoter region, CD2517.1 toxin and corresponding RCd8 antitoxin sequence, alongside the native antitoxin promoter P_{RCc8}.

Following sequence confirmation pMTL8015y-v1-ToxATox-TIII.2R was used to transform chemically competent *E. coli* sExpress_TIII. This donor was then used to transform *C. carboxidivorans* C94 by conjugation (**2.3.9**). A total of five CFUs were observed following 72 hours growth on YTAFM plates with thiamphenicol and D-cycloserine selection. All five colonies were picked into 5 mL of comparably selective YTAFM broth and grown anaerobically at 37°C until turbid (36 hours). A 900 µL aliquot of each culture was used to generate 10% (v/v) DMSO stocks of the transconjugants, and the remaining 4.1 mL used in genomic DNA extraction. A 0.5 µL volume of each 100 µL eluate was then used as a template for PCR amplification of the *CcaP7ORF17830P* locus to confirm plasmid integration. This PCR was performed with the primers TIII.2R Scr Fw2 and TIII.2R Scr Rv3. These primers bind 1213 bp downstream and 1019 bp upstream of *CcaP7ORF17830P* respectively, outside of the recombination locus

and are hence only complimentary to the chromosome and not to the plasmid sequence. The previously used modified Q5 High-Fidelity polymerase protocol was utilised to ensure amplification of the longer regions, with a reduced extension temperature of 70°C and increased extension time. The cycle number was also reduced from 35 to 25 cycles, however only integrant bands of 10390 bp were observed. These were the expected size for successful genomic integration of the plasmid, although the previously observed wild-type bands-present due to integrant instability-were unexpectedly absent. A 0.5 µL aliquot of *C. carboxidivorans* C94 gDNA was also used as a positive control, from which a 5175 bp band was generated congruent with the wild-type locus size (**Figure 31**). These five amplicons were further analysed by sequencing using p15a_tra_F2 to confirm toxin and antitoxin sequence integrity. The generation of transconjugants and amplification of the integration locus confirmed the success of the initial first crossover event.

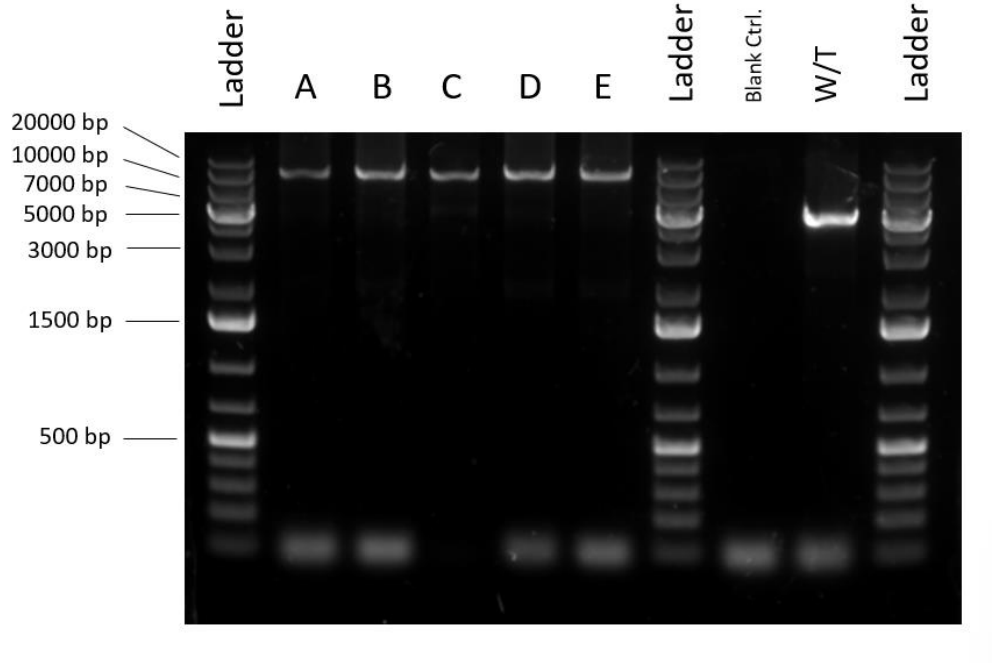


Figure 31 – Screening PCR to confirm pMTL8015y-v1-ToxATox-TIII.2R genomic integration.
Gel electrophoresis analysis of colony PCR products generated by amplification of the *CcaP7ORF17830P* locus using the primers TIII.2R Scr Fw2 and TIII.2R Scr Rv3. 5157 bp band generated from wild-type locus amplification, and 10390 bp from integrant. Ladder: O'generuler 1 kb plus. A-D: transconjugant colonies. Blank Ctrl.: identical PCR mix with dH_2O template. W/T: identical PCR mix with *C. carboxidivorans* C94 gDNA template.

For counterselection to stimulate the second crossover, toxin induction was performed in liquid YTAFM culture as previously described, with 2.5 mM theophylline and no other selection for 24 hours. This was then plated on thiamphenicol selective and non-selective YTAFM plates at dilutions up to 10^{-5}. Following 48 hours of growth the CFUs were counted, showing a 18.58% drop in colonies on the selective plate. To isolate any double-crossover events, 100 colonies were then replica plated from these non-selective plates once more onto thiamphenicol selective and non-selective YTAFM plates. Following 48 hours of growth all 100 colonies had grown under both conditions, and hence no thiamphenicol sensitive colonies could be isolated. This was perplexing as previous results had indicated good efficacy of the counter selectable marker upon induction.

Hence other counter selective attempts were pursued in an effort to stimulate plasmid excision. This counter selective attempt was made utilising inductive plates. To maximise rates of double crossover events three integrant cultures were grown for 24, 72 and 196 hours within non-selective YTAFM broth prior to plating dilutions on YTAFM agar with 10 mM theophylline to maximise CD2517.1 expression. All dilutions were also plated on YTAFM plates with no selection and the CFUs at each stage counted. Following 48 hours of growth 35 colonies from each theophylline inductive plate were further replica plated onto thiamphenicol selective and non-selective plates. Of the 105 colonies replica plated across the timepoints none exhibited thiamphenicol sensitivity (**Table *27***). Notably prior to the 196 hour timepoint all comparative CFUs between inductive and non-inductive plates showed a higher colony number in the presence of theophylline. To determine whether this loss of counterselective functionality was due to sequences changes which may have occurred, mitigating the effectiveness of the induced cassette the colony PCRs of the triplicate integrants assayed were extracted and sequenced with p15a_tra_F2. There were no sequence changes identified in any of these Rb-toxin-antitoxin loci which explained this failed inductive counterselection. With no evidence of functional counter-selection at the single-crossover integrant

level, research efforts were shifted to alternative genomic manipulation strategies.

Table 27 – Comparative CFUs of three *C. carboxidivorans* C94 pMTL8015y-v1-ToxATox-TIII.2R chromosomal integrants upon theophylline inductive and non-inductive plates following 24-, 72- or 196-hours growth in non-selective YTAFM broth.

Non-selective growth	replicate	CFU (10^{-4}) 10 mM theophylline	CFU (10^{-4}) 0 mM theophylline
24 hours	A	154	88
	B	111	62
	C	124	45
72 hours	A	345	307
	B	123	37
	C	247	89
196 hours	A	150	119
	B	142	113
	C	199	60

4.2.3. Exogenous CRISPR-Cas9 genome editing

CRISPR-Cas systems are an invaluable modern molecular biology tool offering highly adaptable, rapid and marker less genome editing *in vivo*. While the low DNA transfer rates observed within the wild-type *C. carboxidivorans* P7 strain are more suited to smaller, less toxic vector delivery more characteristic in endogenous system exploitation, the lack of any endogenous CRISPR-Cas system within the strain necessitated the development of an optimised exogenous CRISPR-Cas9 system for use within strain C94. The effectiveness of an exogenous system is impacted by many factors including expression levels of the guide RNA, expression levels of the Cas9, transfer frequency and homologous recombination efficiency. As such the viability of a sequence modified CRISPR-Cas9 system with *S. pyogenes* wild-type *cas9* under the control of an inducible promoter was tested first within the more amenable *C. auoethanogenum* strain, before being applied within *C. carboxidivorans*.

4.2.3.1. Genomic modification of *C. autoethanogenum* with pMTL4315_SynCas9_adh1_N20 #2

With the efficacy of the inducible $P_{fdx(Csp)}$-RbE* promoter system confirmed within *C. carboxidivorans,* the pRXCasN inducible CRISPR-Cas9 vector series concurrently exemplified by I.C. Canadas (2019) was an ideal candidate for testing within *C. carboxidivorans*. The wild-type *S. pyogenes cas9* gene and associated P_{fdx}-RbE* sequence of the functional module contained a total of twenty-five TII/III motifs: five TII/III-C, eighteen TII/IIII-B and two TII/III-A motifs. To mitigate the impact of these on conjugation efficacy, twenty-five single nucleotide changes were made to the sequence *in silica.* Each change destroyed the RM recognition sequence while conserving the translated amino acid sequence, resulting in a 4358 bp fragment which was synthesised as four synthetic DNA gBlocks by IDT.

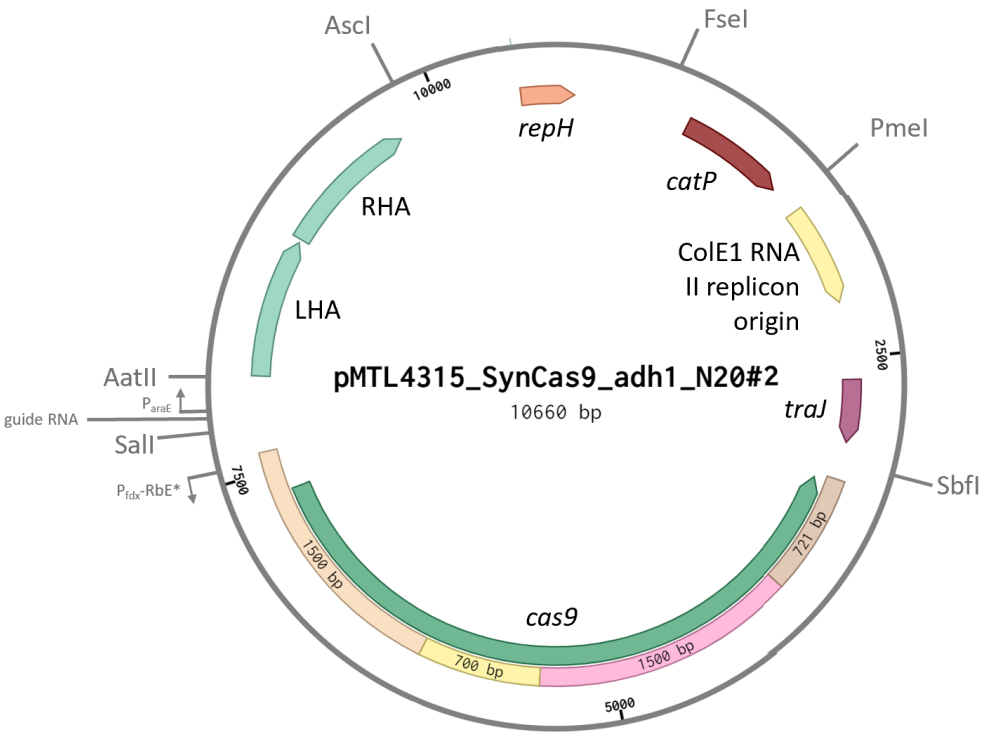


Figure 32 – Plasmid map of pMTL4315_SynCas_adh1_N20#2
Vector map generated through assembly of modified $P_{fdx(Csp)}$-RbE*-*cas9* cassette into pMTL4315_CRISPR_adh1_N20#. Pfdx_Rb3: $P_{fdx(Csp)}$-RbE* driving *cas9* expression. 1500 bp/700 bp/ 1500 bp/721 bp: PCR amplified synthetic fragments.

Initial assembly of the modified $P_{fdx(Csp)}$-RbE*-*cas9* cassette was performed into a previously demonstrated pRECas vector, to facilitate immediate validation of the maintained efficacy of the modified sequence. The chosen vector, pMTL4315_CRISPR_adh1_N20#, had been successfully applied in the genomic modification of *C. autoethanogenum* DSM10061 (**Figure 32**). The genomic region targeted by this vector was 339 bp upstream of CLAU_1435 and 93 bp downstream of the *pyrE* gene. The vector itself was designed to introduce the sequence CCTCGAGTATATAGACGT between two predicted terminators at this locus (C. Humphreys, personal communication). To introduce the modified inducible *cas9* cassette, the four fragments were amplified by PCR for subsequent HiFi assembly into linearised pMTL4315_CRISPR_adh1_N20# as detailed in **2.9.9**.

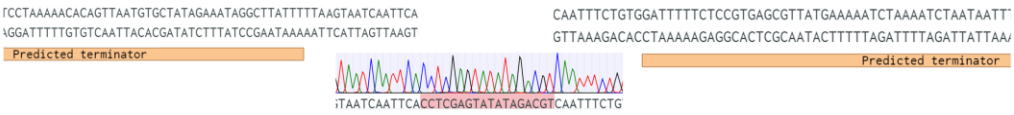


Figure 33 –pMTL4315_SynCas9_CRISPR_adh1_N20#2 generated mutant sequence. Sanger sequencing trace of PCR amplified genomic region of *C. autoethanogenum* (bottom), aligned to the native wild-type locus represented by the above sequence top strands. This highlights the non-native CCTCGAGTATATAGACGT sequence introduction at the expected location.

This resultant pMTL4315_SynCas9_CRISPR_adh1_N20# vector was first used to transform chemically competent *E. coli* sExpress, before being further transferred into *C. autoethanogenum* by conjugation as detailed in **2.3.8**. No notable change in conjugation efficiency was observed comparatively to the original vector. Following 72 hours of transconjugant colony growth on D-cycloserine and thiamphenicol selective plates, five colonies were streaked onto YTF plates supplemented with 2.5 mM of theophylline. Following a further 72 hours of growth these colonies were analysed by colony PCR using the primer pair C.auto Scr Fw and C.auto Scr Rv. This amplified a 196 bp region framing the targeted locus, which was isolated using gel electrophoresis before undergoing sequencing with the above primers (**Error! Reference source not found.**). 100% of the colonies showed successful integration of the

CCTCGAGTATATAGACGT sequence at the CRISPR-Cas9 targeted locus, indicating the retained functionality of the *cas9* sequence. As such the inducible CRISPR-Cas9 system was then tested within *C. carboxidivorans* C94.

4.2.3.2. Knockout of *CcaP7ORF17830P* within *C. carboxidivorans*

Before transfer to *C. carboxidivorans* C94 the CRISPR-Cas9 system was transferred to pMTL8315y-v1. The P_{araE} promoter controlling transcription of the synthetic guide RNA was also replaced with the previously tested P_{thl}. As demonstrated by the CAT reporter assay P_{thl} is more than 2.5x stronger than P_{araE} therefore affording the highest transcription rates and bioavailability of guide RNA for CRISPR-Cas9 complex assembly *in vivo*. The initial target of this system was knockout of *CcaP7ORF17830P*, the restriction endonuclease of the TIII.2 RM system (otherwise denoted as Ccar_17825 or TIII.2R). The assembly of these two pRESynCas(pthI)-v1-TIII.2-sg2 and pRESynCas(pthI)-v1-TIII.2-sg3 vectors, including sgRNA selection and homology arm integration is detailed in **2.9.10** (**Figure 34**). The concurrent assembly of a third vector with an alternate guide sequence (sg1) was unsuccessful, perhaps due to unexpected off-target binding of the guide sequence within the *E. coli* 10-beta cloning strain genome proving lethal even at repressed *cas9* expression levels. The resultant sequence confirmed vectors were used to transform chemically competent sExpress_TIII before being transferred by conjugation into *C. carboxidivorans* C94. Following 72 hours of selective growth a total of 16 and 18 transconjugant colonies were observed for the sg2 and sg3 vectors, respectively.

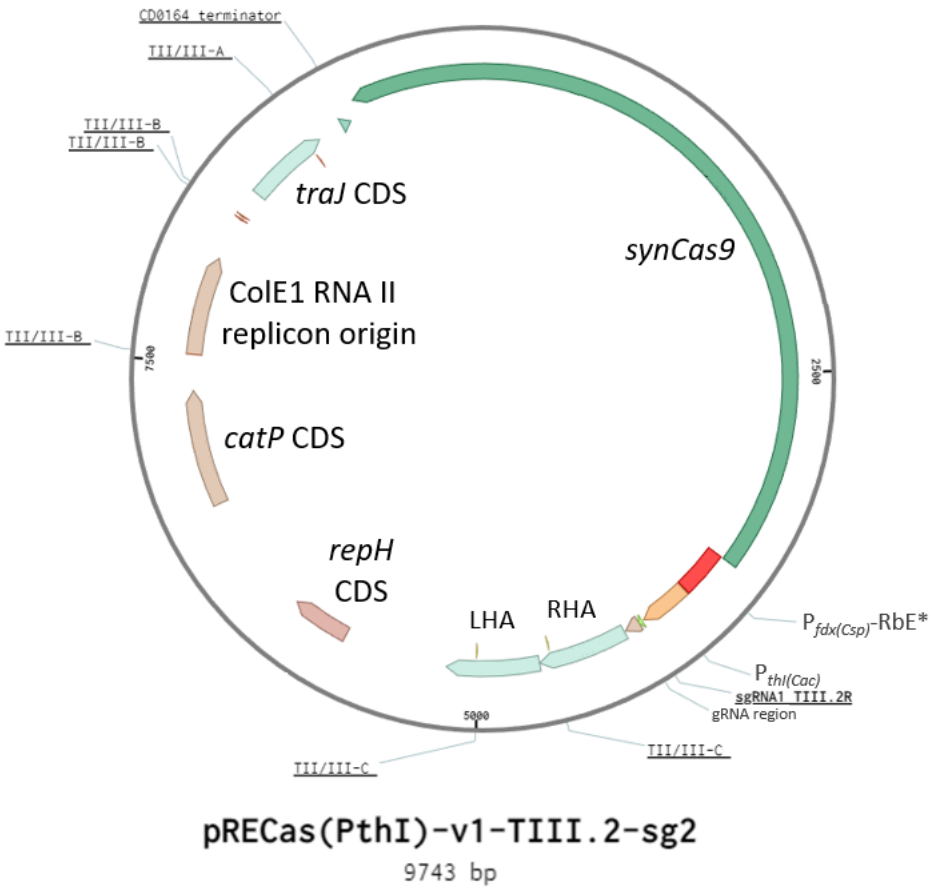


Figure 34 -pRECas(P_{thl})-v1-TIII.2R-sg2 plasmid map
Schematic detailing the pRECas(PthI)-v1-TIII.2R-sg2 vector used to knockout *CcaP7ORF17830P* within *C. carboxidivorans*. Assembled with the modified *cas9* gene, P_{thl} promoter, two homology arms of approximately 500 bp and pMTL8315y-v1 backbone.

Four transconjugants of each vector were then directly re-streaked to 2.5 mM theophylline supplemented YTAFM plates and grown for 48 hours. These were screened, alongside 0.5 µL of *C. carboxidivorans* C94 gDNA and 0.5 µL of water, by colony PCR using primers TIII.2R Scr Fw2 and TIII.2 Scr Rv2. Seven of the eight colonies' screens produced bands when analysed using gel electrophoresis (**Figure 35**). Of these, two had truncated bands not congruent with the 4517 bp fragment expected from the wild-type locus, but comparable to the 1584 bp band indicative of successful *CcaP7ORF17830P* knock-out. These two DNA fragments were extracted and sequenced with TIII.2R Scr Fw2/Rv3 to confirm the expected region had been excised.

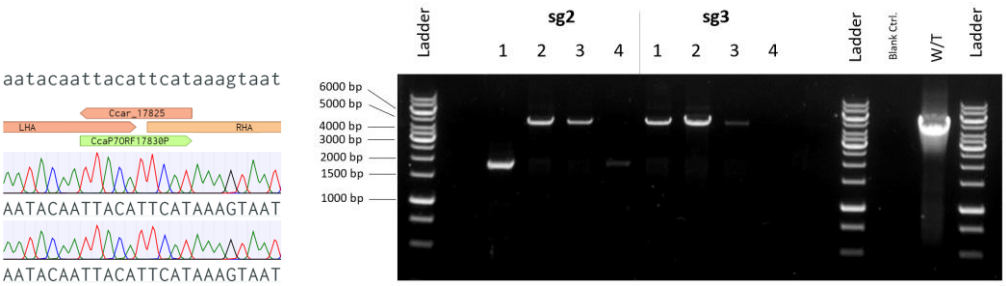


Figure 35 – Sanger sequencing and colony PCR of pRESynCas(pthI)-v1-TIII.2-sg2 generated mutants. W/T band showing evidence of overloading and resultant reduced band size. Ladder: O'generuler 1 kb. Sg2(1-4): *C. carboxidivorans* pRESynCas(pthI)-v1-TIII.2-sg2 induced colonies, with lanes 1 and 4 showing expected band size reduction due to genetic knock-out. Sg3(1-4): *C. carboxidivorans* pRESynCas(pthI)-v1-TIII.2-sg3 induced colonies, all displaying wild-type unmodified band sizes. Blank Ctrl.: identical PCR mix with dH_2O template. W/T: identical PCR mix with *C. carboxidivorans* C94 gDNA template.

Following confirmed knockout of the target gene, the two replicates were picked into thiamphenicol selective YTAFM broth and grown for 48 hours prior to generation of stocks. At 48 hours this culture was also passaged into non-selective YTAFM broth. This non-selective serial passage continued at 24-hour intervals for 196 hours. 10^{-4} and 10^{-5} dilutions were made in PBS and plated on YTAFM following 24 and 120 hours of non-selective growth. These plates were incubated for 48 hours before 100 colonies were replica plated onto thiamphenicol selective and non-selective plates to screen for plasmid loss. Both following 72 and 196 total hours of non-selective growth all screened colonies exhibited maintained thiamphenicol resistance. This lack of degradation in thiamphenicol resistance was unexpected, and comparative plasmid loss through serial non-selective growth was performed with a previous stock of *C. carboxidivorans* C94 maintaining pMTL8315y-v1. Following triplicate non-selective YTAFM broth growth for 72 and 196 hours 10^{-3}, 10^{-4} and 10^{-5} dilutions of the culture were plated directly to thiamphenicol selective and non-selective plates. pMTL8315y-v1 did not show the unexpected plasmid retention observed with the pRESynCas(pthI)-v1-TIII.2R vectors, with plasmid loss of 45.13% (± 4.72%) observed at 72 hours and 98.83% (± 0.44%) at 196 hours, as calculated from through comparative CFU counts. This was a perplexing result, implicating a sequence change between pMTL8315y-v1 and the CRISPR-Cas9 vector as preventing comparable plasmid loss following the successful knock-out of *CcaP7ORF17830P.*

4.3. Discussion

This chapter outlines the development and implementation of two genetic modification systems within *C. carboxidivorans* C94, and generation of a proof-of-concept knockout of the *CcaP7ORF17830P* gene. The ability to perform marker-less genomic modifications efficiently and at any genomic locus is an invaluable tool which underpins many strain development research avenues, as well as facilitating advanced genetic investigations. Where lack of transfer has precluded previous efforts to engineer this organism, the advances within this study have attained a transfer rate of 3 x 10^4 CFU/conjugation. While this observed transfer efficacy with pMTL8315y-v1 is promising for the successful introduction of plasmid-based tools, transfer efficacy remains a key consideration in vector design. In particular the highest efficacy conjugal donor available (sExpress_TIII) is still limited in its scope, and not expected to afford any protection to RM system targeted degradation at TII/III-C RM motifs, or the three Type I RM motifs.

4.3.1. Genetic tools

Outside of this study the only genetic tools demonstrated within *C. carboxidivorans* are those within the pMTL82151 vector sequence. These include the *catP* selectable marker, pBP1 Gram-positive replicon (derived from *C. botulinum* NCTC2916) and the P_{thl} promoter derived from the thiolase gene of *C. tyrobutyricum* ATCC 25755 (Genbank accession no. HM989902) (Cheng, et al., 2019). With such limited scope of previously validated tools, initial efforts within this study focused on assessing the compatibility of widespread tools used in *Clostridium* before subsequently exploiting them in the development of genetic engineering systems. The first genetic tool tested was the *ermB* selection marker, a readily available resistance gene within the broadly used clostridial pMTL80000 shuttle vector series. Initial testing confirmed the sensitivity of *C. carboxidivorans* to clarithromycin at concentrations exceeding 5 µg/mL. Clarithromycin was used in lieu of erythromycin due to the pH

sensitivity of the former, and cultures cultivated for up to 72 hours in accordance with procedural requirements and observed growth rates. Following the confirmation of this sensitivity a pMTL8325y-v1 vector was successfully transferred by conjugation, conferring resistance to 6 µg/mL clarithromycin and ensuring plasmid maintenance. Validation of *ermB* as an amenable selection marker within *C. carboxidivorans* facilitated the subsequent use of the *catP* gene as a non-maintenance integral reporter gene. The availability of two selectable markers and the three replicons used (pCB102, pCD6 and pBP1) could also permit the use of dual plasmid approaches within the strain. This is particularly useful in species with low transfer rates and has typically been applied in the development of high-throughput rapid CRISPR-Cas systems within a number of Clostridia (Diallo, et al., 2020; Wasels, et al., 2017).

The strengths of four constitutive promoters were assayed using the *catP* reporter gene on an *ermB* plasmid backbone. Validation of these promoters provides a good core genetic toolset from which to further expand the genetic toolbox. Alongside constitutive promoters an inducible expression system was viewed as integral to both prospective genomic engineering systems being considered within this strain. Some of the most common inducible expression systems employed within clostridial studies include the Tet3nO system (Walker & Köpke, 2015) and the *bgaR*-P_{bgaL} lactose inducible system derived from *C. perfringens* (Hartman, et al., 2011). These systems are 825 bp and 1406 bp respectively, with Tet3nO containing a TI-C, TI-A, TII/III-A and TII/III-B RM motif, and *bgaR*-P_{bgaL} containing one of each TII/III RM motif sequences. Due to the lack of Type I protection afforded by the high transfer frequency sExpress_TII and sExpress_TIII *E. coli* conjugal donors, the prospects of transferring this system were unfavourable. The *bgaR*-P_{bgaL} inducible system, while containing sites previously successfully transferred in concert via sExpress_TIII and sExpress_TII, further increased the vector length considerably. The theophylline responsive riboswitch system exemplified within the RiboCas genome editing system (Cañadas, et al., 2019) are only 253 bp total length. Of

this 253 bp, 166 bp have already been transferred into *C. carboxidivorans* through validation of P_{fdx} as a constitutive promoter. As such the efficacy of the riboswitch E* based expression system was also tested using a *catP* reporter assay, to confirm sufficient stringency of repression and theophylline responsiveness within the strain. This inducible system facilitated the transfer of both a highly toxic *S. pyogenes cas9* gene and CD25171 *C. difficile* toxin and could be further implemented in the development of other genomic tools reliant on scalable expression such as CRISPRi.

The only other promoter assessed within *C. carboxidivorans* within the literature is $P_{thl(Ctb)}$. Confirmation of heterologous gene expression driven by this constitutive promoter was initially performed through RT-PCR amplification of the two non-native genes within the transformed strains. Activity was further assayed through butanol/ethanol dehydrogenase enzymatic assays, in which the activity (U/mg) within the ALD strain was approximately double that of the wild-type. While this successfully confirmed the overexpression of these genes within the study through comparison to the untransformed strain, the divergence in enzymatic assays unfortunately prevents the direct comparison of the $P_{thl(Ctb)}$ promoter strength with those assayed within this study.

4.3.2. Single crossover integration

Allelic exchange is a common genomic engineering tool employed within the *Clostridium* genus, with the most prominent example being the *pyrE* focused ACE approach published in 2012 (Heap, et al., 2012). While allelic exchange can be mediated through genome-based selection, with different phenotypes conferred by the disruption of a target gene, this significantly limits the scope only allowing the knockout of key genes, and subsequent use of these as an integration locus. More widely applicable are allelic exchange systems reliant on plasmid-based selection and counter-selection markers. These are not limited in their targeting scope, instead reliant on the validation of appropriate

selection markers and systems for isolating initial single crossover integrants, as well as their more elusive double crossover counterparts.

Within this study the initial single integrant crossover selection in *C. carboxidivorans* was performed with non-replicative suicide vectors containing the *catP* gene to confer thiamphenicol resistance. Preliminary evaluation of the viability of such an approach was performed with replication-deficient variants of the original pMTL8315y-v1 vector. Non-replicative genomic tool vectors are significantly less prevalent in clostridial studies than replicative approaches due to the hurdle posed by combined low homologous recombination efficiencies and transfer efficiencies within this single step. To overcome this in some studies a selectable marker is placed within the introduced sequence, meaning mutants maintain a selectable phenotype following double crossover events. This is a powerful selection tool, as can be seen in the knockouts of the *fliA*, *adhE1* and *adhE2* of *C. ljungdahlii* with *ermC* in 2013, where 30% of the initial transconjugants screened were double crossover mutants (Leang, et al., 2013). While FLP-FRT recombination has been used in some instances to remove the selectable marker, broadly this is not a favoured approach. This is in part due to the integration of a selectable marker introducing further phenotypic alterations complicating forward genetics studies and being legislatively more problematic in the application of a strain within industry. It is also due to the efficiency afforded by marker integration being somewhat negated by these extra regeneration steps.

Other marker-less suicide vector allelic exchange systems have been demonstrated in *Clostridium*. For example, a restriction-deficient *C. saccharobutylicum* strain (Ch2) was generated using a series of vectors with approximately 1 kb homology arms and *codAB* counter-selection marker to iteratively target the *hsdR1, hsdR2* and *hsdR3* genes (Huang, et al., 2018). This restriction-deficient mutant was then used once more to demonstrate an alternate suicide allelic exchange vector using *upp* based counterselection. To apply this counter-selection approach the native *upp* locus was the first target for knockout. Further genomic manipulations were then performed within this

Δ*upp*Δ*hsdR1*Δ*hsdR2*Δ*hsdR3* strain, including the introduction of *spo0A* from *C. acetobutylicum* to generate a novel sporulating strain (Foulquier, et al., 2019). This system was concurrently applied within a similarly restriction-deficient *C. acetobutylicum* strain with DNA transfer frequency previously reported as 6 x 10^4/μg DNA.

With the dual considerations of homologous recombination efficiency and transfer efficiency within suicide vector approaches there is notable trade-off associated with reaching sufficient homologous region length to stimulate homologous recombination and minimising overall plasmid length to ensure the highest transfer frequency possible is attained. This was somewhat addressed by Foulquier and co-workers (2019) where a pointed focus on plasmid size reduction was incorporated into the system design. This successfully generated minimal vectors of 2582 bp and 2892 bp respectively for use *C. saccharobutylicum* and *C. acetobutylicum*. For targeted genome manipulation two ~1 kb homology arms were then incorporated into these minimal vectors, further increasing the size. Although the total recombination cassette size is greater than what was tested within this initial assay in *C. carboxidivorans*, homologous recombination of these vectors into the chromosome still occurs at only one of these 1 kb regions at a time. 1 kb is the most prevalent homology arm size within published clostridial studies employing such suicide allelic exchange systems (McAllister & Sorg, 2019). This size is also congruent with what was observed within this study, where robust transfer was only observed with a homologous region exceeding 900 bp.

Concurrent with this study, research was also performed by C. Humphreys into the efficacy of a different allelic exchange system also employing inducible CD2517.1-based counter-selection. This approach differed in the mechanism used to isolate initial single crossover events; instead of DNA integration through use of a suicide vector, this approach employed pCD6 as a pseudo-suicide replicon to retard cell growth, as exemplified within ACE. Following initial vector transfer, colonies visibly exhibiting improved growth rates were selected as probable single crossover mutants whose growth was no longer

restricted by this segregationally instability. This approach benefited from much higher transfer efficacies, however, it was also unsuccessful in the isolation of double crossover mutants. The lack of success of this pseudo-suicide approach was attributed to struggles in ensuring single crossover integrant purity, a common issue with replicative allelic exchange approaches. This instability could be observed through PCR amplification of the target genomic locus of all prospective colonies which yielded both wild-type and integrant bands even following repeated purification attempts. As previously demonstrated, toxin induction with a chromosomally independent plasmid is an extremely effective counter-selection method, with 99.8% of induced single-crossover colonies no longer exhibiting Tm^R (**4.2.2.2**). If the unstable single-crossover colonies retained a subpopulation of cells harbouring residual extrachromosomal plasmid these would readily lose the plasmid, as demonstrated with pMTL8315y-ToxATox-PyrE. As such 5 mM theophylline induction of 25 of these impure single-crossover isolates yielded 23 colonies exhibiting restored thiamphenicol sensitivity acquisition (92%), however, all with revertant wild-type sequences upon screening. This was a negative result, indicating that the instability of the integration led to an overwhelming majority of wild-type genotypes when the plasmid was counterselected.

4.3.3. CD2517.1 counter-selection

Following generation of single crossover mutants, the allelic exchange protocol then induced expression of the CD2517.1 toxin gene, derived from *C. difficile* 630, as a counter-selectable marker. Assessment of the efficacy of toxin induction was performed through the determination of thiamphenicol resistant colony ratios, with successful plasmid loss excising the plasmid-based *catP* gene responsible for thiamphenicol resistance. Initial assessment of this inducible counter-selection system using pMTL8315y-v1-ToxATox-PyrE was promising, indicating 99.8% of induced colonies had regained thiamphenicol sensitivity. Induction of the counter-selection system, however, did not

successfully yield plasmid loss when applied to the single crossover *C. carboxidivorans* integrants isolated through transfer of pMTL8015y-v1-ToxATox-TIII.2R.

The failure to isolate double crossover mutants through this integrative vector approach did not occur as observed within the pseudo-suicide approach employed by C. Humphreys (due to single-crossover integrant instability and residual independent plasmid retention). In contrast PCR amplification of the single crossover chromosomal plasmid integration at the *CcaP7ORF17830P* locus (**Figure 31**) yielded no bands congruent with the wild-type locus size, suggesting good integrant stability. Theophylline induction of three transconjugant colonies, however, did not yield any discernible increase in thiamphenicol sensitivity. It is difficult to compare the CD2517.1 derived allelic exchange vectors employed within this study with the original vector applied within *C. difficile* as the paper does not report efficacy characterisation. It is does, however, note a reduction in false-positive change-of-function mutations within the toxin system, compared to the previously employed *codA* counter-selection system. When the three integrants' colony PCRs were sequenced with p15a_tra_F2, no such loss-of function mutations were identified.

The initial characterisation of the CD2157.1 toxin induction was not performed in a directly comparable system to that of the subsequent allelic exchange vector, relying on a pCB102 replicative vector as opposed to single chromosomal integrant. Comparative data on the qualities of Gram-positive replicons within clostridia is limited, with significant variance in compatibility between species attributed to partial dependence on host factors. Some previous early studies employed gel electrophoresis to assess levels of plasmid within *C. perfringens* (Allen & Blaschek, 1990) and *C. acetobutylicum* (Lee, et al., 1993). More recently qPCR quantification of the pCB102-derived plasmid copy number in *C. beijericnkii* showed 63 (± 6) in mid-exponential phase (Millard, 2017). This is a relatively high copy number, and although plasmid copy number was not determined for any vector within this study, at any copy number exceeding one exclusive chromosomal integration of the vector would reduce

the overall copy number of the toxin gene within each cell. As such the toxin load of the cell, despite high induction levels, may be markedly lower in the single-crossover integrant than during the characterisation of the system. This would mitigate the efficacy of the toxin, particularly due to the mechanism of toxicity being membrane disruption.

This effect may be further compounded by the issues observed in *C. carboxidivorans* theophylline sensitivity determination (**Figure 24**) where theophylline appears beneficial to cell growth at low concentrations. In fact, this appears further reflected in the pMTL8015y-v1-ToxATox-TIII.2R counter-selection CFUs (**Table 27**) where all timepoints and replicates demonstrate a higher occurrence of colonies from identical dilutions on theophylline plates, as opposed to their non-selective counterparts. This response was unexpected, and the mechanism of this theophylline conferred advantage is not immediately apparent. As a xanthine (1,3-dimethylxanthine), theophylline has frequently been studied in plants and fungi alongside caffeine (1,3,7-trimethylxanthine), which is structurally similar. Bacterial metabolism of these compounds is not common but has been reasonably well characterised in *Pseudomonas putida,* which can use caffeine as a sole carbon and nitrogen source (Summers, et al., 2015). BLAST alignments and analysis of the genome of *C. carboxidivorans* reveal no analogous metabolic pathway components, and as such an alternate mechanism must be implicated. While weak mutagenic effects, morphological and sporulation changes, as well as β-Galactosidase inhibition have been reported in bacteria exposed to theophylline, including *C. perfringens* and *B. subtilis*, research on the antibacterial properties and mechanism is extremely limited (Horne, et al., 2020; Giri, et al., 1999; Sacks & Thompson, 1977; AL-Janabi, 2011).

A recent publication employing allelic exchange in *C. difficile* and *Streptococcus sius* demonstrated the isolation of double crossover mutants without counter-selection (Faulds-Pain & Wren, 2013). As such, once it became apparent that CD2517.1 induction may be ineffectual, a non-selective growth phase was introduced into the protocol analogous to that used within this study. Within

the *C. Difficile* allelic exchange protocol, isolation of double crossover mutants was performed by replica plating *C. difficile* integrants on thiamphenicol and non-selective plates following three or four days of subcultured growth in nonselective media. With the vector containing 1200 bp homology arms the homologous recombination efficiency was such that the fifth nonselective subcultures/days yielded one successful double crossover mutant in 50 replica-plated colonies, increasing the next day to eight in 75. This rate of double crossover frequency within the bacterial population was not observed within *C. carboxidivorans,* with the replica plating of 105 colonies following 196 hours of non-selective growth yielding no thiamphenicol sensitive colonies or double crossover mutants. This is likely attributed to the variances in homologous recombination efficiencies between the strains which are notoriously low across the Clostridial genus, or perhaps to differences in the stability of the integrated plasmid.

4.3.4. CRISPR-Cas9 genome editing

CRISPR-Cas9 systems are the most prevalent genomic engineering tool applied within *Clostridium* today. This prevalence is due to the plethora of advantages afforded comparative with previous technologies such as ClosTron and ACE. One such advantage is the high specificity of targeting, paired with prevalence of appropriate target sites. The generation of mutants is also rapid, with scope for both knock-ins and knockouts without the necessity of selectable marker integration or gene disruption induced phenotypic changes. These systems function by leveraging the homologous recombination activity of the cell alongside powerful counter-selection against the wild-type sequence. This counter-selection is conferred by the CRISPR-Cas9 complex which generates DSBs at a targeted sequence within the genome. Without repair these genomic cleavages prove lethal to the cell, and the CRISPR-Cas9 complex is hence toxic to cells with a wild-type locus, due to the persistent targeted genomic cleavage of the un-modified target sequence. This enriches the bacterial population with

mutants in which the targeted sequence is no longer present in the chromosome.

While CRISPR-Cas9 genome editing is becoming ubiquitous, there are still notable drawbacks to the application of these plasmid-based exogenous genome editing tools. These plasmids are typically large due to the incorporation of the *cas9* (4.1 kb) and the homology region (1-2 kb). The homology region can in fact be significantly larger if delivering cargo in lieu of removing sequences. For functional expression of the CRISPR-Cas system these features must be integrated into an appropriate vector with necessary functional modules, compounding overall vector size so it is typically in excess of 9.5 kb. The largest vector transferred into *C. carboxidivorans* within this study prior to this was pMTL8315y-v1-ToxATox-PyrE at 5.7 kb, at which point the transfer frequency was not assessed. Plasmid length is a key concern as increased plasmid length has been observably deleterious to DNA transfer frequency throughout clostridial studies (Heap, et al., 2009). This has been noted as even more apparent in some CRISPR studies where homology arm sizes were accordingly reduced to ensure the final vector size was less than 9 kb (Zhang, et al., 2018).

Another drawback to exogenous CRISPR-Cas9 application in genome editing is the inherent lethality of the active complex. Many previous clostridial studies have employed a modified nickase *cas9* variant which confers only single strand breaks, and therefore has a lower lethality. This initial preference for nCas9 systems within *Clostridium* has been largely attributed to the increased numbers of primary transformants/transconjugant cells obtained in contrast to the native *cas9* gene (Xu, et al., 2015; Li, et al., 2016). Even with this reduced lethality, the rates of transfer are still impacted, with a drop of 2-3 orders of magnitude in DNA transfer frequency observed when transforming *E. limosum* ATCC 8486 with a pMTL40000 vector targeting *R.Eli8486ORFD* (Millard, 2017). The inherent lethality of the native *cas9* during cloning stages has even been reported as generating SNPs and other cloning artifacts which lower the toxicity, like that observed with the CD2517.1 cloning stages (Ingle, et al.,

2019). An alternate approach to mitigating CRISPR-Cas9 lethality is placing the *cas9* gene under inducible expression, with sufficient repression to negate the associated lethality. The wild-type *cas9* sequence is a more powerful selection tool compared to nCas9 which has more burdensome and time-consuming mutant isolation. The alternate application of endogenous system exploitation, while particularly attractive in low transfer strains, was unfortunately not applicable within *C. carboxidivorans.*

These considerations were particularly prominent due to the extensive work undertaken to reach the conjugation efficiency observed within *C. carboxidivorans* C94. If a comparable drop in transfer was observed with the CRISPR-Cas vector in *C. carboxidivorans* as with *E. limosum* then the DNA transfer frequency could be reduced to just 30 transconjugants per conjugation. If this were to be further compounded by RM activity or CRISPR-Cas toxicity the resultant lack of transconjugants would completely preclude the application of this genomic engineering approach. As such the P_{fdx}-RbE* inducible expression system was employed to repress *cas9* expression in *E. coli* and *C. carboxidivorans.* Alongside this all RM motifs within the sequence were disrupted to evade any residual RM system restriction activity while retaining the original amino acid sequence, hence no significant loss of function was expected. The success of the system within *C. autoethanogenum* further validated the functionality of this modified system prior to application within *C. carboxidivorans*.

The final CRISPR-Cas9 vector produced within this study for knockout of *CcaP7ORF17830P* totalled 9743 bp. Fortunately, homology arm inclusion only added two TII/III-C RM motifs and successful transfer was attained using the sExpress_TIII *E. coli* donor. The transconjugant numbers reflect a drop in transfer frequency of more than three orders of magnitude, slightly greater than that observed with *E. limosum*. The efficiency of the induced system also appears relatively good, with 20% of screened colonies exhibiting the mutant genotype following only 48h hours growth on inductive plates. While this is

likely to vary based on the guide used and target change, this is sufficient for rapid mutant isolation in future application of the approach.

Following confirmation of *CcaP7ORF17830P* excision and isolation of these knockout variants, it unfortunately proved difficult to lose the thiamphenicol resistance conferred by the CRISPR-Cas9 vector within the course of this study. This was not expected, with previous work demonstrating plasmid loss of 98.83% (± 0.44%) with pMTL8315-v1 after only 72 hours of non-selective growth. This may indicate an integrative event occurred outside of the screened locus; however, it is not clear where such homologous recombination could occur due to the lack of notable genomic homologs of any of the backbone sequence. To confirm this, it would be pertinent to perform next generation sequencing of the entire genome to locate such an integrant. Prior to this, however, a longer duration non-selective growth phase should be performed to ensure the stability of the maintained resistance is sufficient as to preclude any isolation of thiamphenicol sensitive cells. If this characterisation does highlight the persistence of this concern, it may be necessary to explore other counter-selection marker options, such as *upp.*

4.3.5. ΔccaP7ORF17830P

The *CcaP7ORF17830P* gene targeted within this study encodes the restriction enzyme of one of the Type III RM systems, putatively assigned to the AGAAGC RM motif. This was targeted in an effort to apply a third approach to improving DNA transfer frequency in *C. carboxidivorans* through the disruption of the RM barrier. This approach has frequently occurred within DNA transfer studies in *Clostridium*, with RM system restriction activity negation applied to demonstrate the significance of the barrier and confirm the efficacy of the transfer approach exemplified by the study. For example, within two ClosTron generated *C. saccharobutyricum* Type I restriction endonuclease mutants, DNA transfer rates with non-cognately protected DNA increased 8-fold and 3-fold respectively, but only 1-fold when the cognate methyltransferase was

expressed within the donor *E. coli* strain (Lesiak, et al., 2014). These mutants were also able to receive unmethylated DNA, unlike the wild-type control. Other examples of targeted RM system disruption generating restriction-deficient mutants receptive of unmethylated DNA have also occurred in *C. pasteurianum* and *C. acetobutylicum* (Pyne, et al., 2014; Dong, et al., 2010). Following successful plasmid loss with the mutants isolated within this study, it would be interesting to characterise the DNA transfer frequency of this novel mutant compared to the original *C. carboxidivorans* C94 isolate.

4.4. Chapter summary

- The sensitivity of *C. carboxidivorans* to clarithromycin, and applicability of the *ermB* selection marker was confirmed.
- Four constitutive promoter strengths within *C. carboxidivorans* were assessed with a *catP* reporter assay.
- The strength and stringency of a theophylline responsive riboswitch and $P_{fdx(Csp)}$ inducible promoter system was assessed.
- The validity of a suicide vector based single crossover integration approach was proven.
- An allelic exchange vector with CD2517.1-based counter-selection was tested, without success when combined with a suicide vector approach.
- A modified CRISPR-Cas9 system based on the RiboCas system (Cañadas, et al., 2019) was successfully used to generate a knockout of *CcaP7ORF17830P* gene in *C. carboxidivorans* C94, however with residual undesirable plasmid retention observed.

5 Application of synthetic biology tools within *E. limosum* NG-6849 and generation of a large-scale mutant library

5.1. Introduction

Eubacterium limosum is another attractive acetogenic bacterial chassis for the industrial valorisation of syngas (**1.5**). As a prevalent species within the human gut microbiome, it is reported as having a diverse heterotrophic metabolism with extensive substrate utilisation (Kelly, et al., 2016). It is also an obligate anaerobe capable of butyrate and butanol production during autotrophic growth on C1 feedstocks. ATCC 8486, the type strain, is the most studied within the species to date. Until recently the most studied strain was KIST612, which has since been reclassified as *Eubacterium callanderi.* The original draft genome of the ATCC 8486 strain was published in 2015 (Song & Cho, 2015) with further elaboration and completion in 2017 (Song, et al., 2017). The KIST612 genome sequence was published in 2011 (Roh, et al., 2011). The strain utilised within this study is NG-6849, for which there is no published genome, and only one published study thus far (Flaiz, et al., 2021).

Prior to receipt of *E. limosum* NG-6849, previous research within the SBRC had utilised other *E. limosum* strains including DSM 2593, DSM 2594 and ATCC 8486 (the type strain). Most research had been focused on ATCC 8486, but of note across all species was a phenotype characterised by the production of significant amounts of extracellular polysaccharide. This made handling of the strains problematic and complicated several essential research procedures including cell pelleting and DNA extractions. *E. limosum* NG-6849 does not exhibit this phenotype, making it a more desirable research organism. It is also more suitable for large scale fermentations where extracellular polysaccharide interactions could complicate features such as agitation/foaming.

5.1.1. Genetic tools within *E. limosum*

DNA transfer into *E. limosum* ATCC 8486 was first reported in 2018 (Song, et al.). Within this report a β-galactosidase reporter system was introduced to assess the impact of the 5' untranslated regions of key Wood-Ljungdahl pathway (WLP) carbonyl branch genes, as identified by transcriptomic and translatomic analysis. In 2018, R. Mansfield (2018) also demonstrated an electroporation protocol for DNA transfer into *E. limosum* ATCC 8486, achieving $1x10^4$ colony forming units (CFU) per µg of plasmid DNA using pMTL80000 series vectors (Heap et al., 2009). Mansfield and Millard (2017) then both employed the recently published pMTL40000 CRISPR-Cas9 vector series (Ingle, et al., 2019) to generate proof-of-concept genetic knockouts. In the development and application of this system several genetic tools were demonstrated, including vectors based on the pCB102 and p19 Gram-positive replicons, and a library of inducible and constitutive promoters using a *catP* reporter gene.

More recently a somewhat analogous study was published in which genetic tools were sequentially validated in *E. limosum* ATCC 8486 (Shin, et al., 2019). This began with the assessment of the minimal inhibitory concentrations (MICs) of clarithromycin and thiamphenicol, with and without plasmid-based *ermB* and *catP* selection markers. This was followed by the assessment of the transformation frequencies of pMTL80000 series vectors containing various Gram-positive replicons. The CRISPR-Cas9 system reported herein initially utilised a constitutively expressed wild-type *S. pyogenes cas9* gene, under the control of the strong native ELIM_c2885 (pyruvate ferredoxin oxidoreductase) promoter. This approach encountered similar hurdles to those outlined by Ingle *et al.* (2019) with all isolated *cas9* sequences containing frameshift mutations. As such an anhydrotetracycline inducible promoter, Pcm-tetO1 (Dong, et al., 2012), was subsequently used in the generation of two knockout strains: Δ*folD* and Δ*acsC*, both WLP genes. Five CRISPRi "knock-downs" were also generated, four within the WLP gene clusters (*fhs1, folD, acsC, acsD*), and one of the fructose-PTS system gene (*ptsF*).

Interestingly, the same wild-type *cas9* gene which could not be constitutively implemented in *E. limosum* ATCC 8486 was successfully expressed within *E. callanderi* KIST 612 using a strong constitutive P_{Rbo} promoter (Jeong, et al., 2020). This constitutive *cas9* expression caused no notable impact on the growth rate or other indicator of inherent lethality when introduced in a minimal CRISPR-Cas9 vector (without an associated sgRNA targeting sequence or homology cassette). This system was subsequently used to knockout the *pyrF* gene with an efficiency of 97% and, unlike within the *E. limosum* ATCC 8486 study, this knockout did not simultaneously integrate a selectable marker within the genome.

These studies each implement a different reporter system to assess promoters; β-galactosidase, β-glucuronidase and CreiLOV, respectively. Within a recent study using *E. limosum* NG-6849 an alternate FAST (fluorescence-activating and absorption shifting tag) reporter system was demonstrated (Flaiz, et al., 2021). When this *feg* reporter gene was placed under the control of the lactose-inducible *bgaR*-P_{bgaL} promoter system a 5-fold fluorescence increase was observed upon induction. The fluorescence reported was also 13.3-fold better when cultivated on glucose as opposed to methanol, with codon optimisation of the reporter gene improving this a further 1.5-fold. Concurrent *feg* reporter and exogenous gene expression was also demonstrated, initially with the a*dhE2* gene of *C. acetobutylicum.* The two genes were fused via a flexible glycine linker at either the *N* or *C*-terminal, with *adhE2* activity assessed by solvent profiling, and *feg* by fluorescence microscopy. Multiple assays demonstrated no apparent negation of the functionality of the alcohol dehydrogenase or reporter gene in any condition, although *C*-terminal fusion resulted in both the highest fluorescence and alcohol yields.

5.1.2. Other studies within *E. limosum*

Aside from synthetic tool development further studies employing *E. limosum* ATCC 8486 have been performed to elucidate the intricacies of autotrophic

metabolism, energy conservation and regulation. This includes the previously mentioned transcriptomic and translatomic analysis performed by Song *et al* (2018) in which RNA-Seq and Ribo-Seq were applied to samples grown on 5 g/L glucose and CO_2:H_2, respectively. The RNA-seq dataset generated within the course of the study was thorough, with read mappings equating to 146.1-fold genomic coverage. This analysis revealed strict translational regulation within the WLP, ensuring conservation of equal molar ratios irrespective of transcript levels. Recently a forwards genetics approach was also taken within *E. limosum* ATCC 8486 in which CO tolerance was improved through 150 generations of adaptive laboratory evolution (Kang, et al., 2020). Of the five common mutations identified across all replicates of elevated growth rate, the key mutation was hypothesised as being a SNP within the *acsA* gene. This gene was proposed as encoding the catalytic subunit of the CODH enzyme, with the mutation substituting a small non-polar amino acid for a large polar amino acid at a distal site from the active pocket. Similar studies using analysis of energy conservation in autotrophic growth conditions, and the implications this has on the resultant solvent profiles, have also been pursued within *E. callanderi* KIST612 (Litty & Müller, 2021), including some enzymatic studies (Jeong, et al., 2015).

5.1.3. *mariner* transposon mutagenesis approaches in *Clostridium*

Transposon mutagenesis is a forwards genetics tool for generating diverse insertional mutant pools. It has been implemented within several acetogenic clostridial species, with two high throughput transposon insertion sequencing (TIS) studies within the genus to date (**1.9**). Early transposon mutagenesis studies within clostridia relied on an *in vitro* approach, whereby a transposon and transposase enzyme are combined to form a transposome. This transposome is then transformed into the cell by electroporation. While this approach, using the commercially available EZ-Tn5 transposome system, generated erythromycin resistant *C. perfringens* insertional mutants at a rate

of ~11,200 transformants/µg transposon DNA (Vidal, et al., 2009), it unfortunately relies on high electroporation efficiencies to ensure a suitably saturated mutant pool is generated. This limits the amenable target strains, and hence focus has since shifted towards *in vivo* systems. These systems generate mutants through intracellular expression of the transposase gene from a plasmid sequence and can hence be applied within any organism in which an appropriately efficient DNA transfer method has been established.

These *in vivo* systems, however, are subject to other key considerations. One concern being prolonged exposure to transposase activity, which has the potential to generate multiple insertions within the genome of a single cell. This significantly complicates mutant characterisation and downstream application of the mutant pool. Suicide vector-based transposon mutagenesis, while sufficiently transient to ensure single insertional mutant generation, would require unfeasible throughput to generate a suitably saturated mutant pool in the majority of clostridial species. Instead, replicative plasmid-based approaches with efficient plasmid loss mechanisms have been developed. For plasmid loss within *C. perfringens galK*, encoding a galactokinase enzyme, was employed as a counter-selectable marker as previously described in **4.1.2.2**. This stimulated plasmid loss in the presence of 3% galactose, however, was reliant on the initial generation of a Δ*galKT* host strain (Liu, et al., 2013). Alternatively, the transposon mutagenesis vector used within *C. acetobutylicum* and *C. sporogenes* used an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible promoter to stimulate plasmid loss (Zhang, et al., 2015). This promoter was placed upstream of the Gram-positive replicon, in this instance pCB102, causing high transcriptional read through of the replicon sequence upon induction. This read through interfered with the key plasmid replication and maintenance functions of the replicon, destabilising the plasmid. This plasmid loss mechanism was effectively demonstrated in both species.

Another important consideration in plasmid-based transposon mutagenesis systems is the repression of transposase activity. General molecular laboratory

practices rely on *E. coli* cloning hosts for plasmid proliferation and isolation. Transposase activity within these strains would not only mutagenise the *E. coli* genome, but also potentially negate the efficacy of the vector through the removal of the transposon sequence (or sometimes movement of the sequence within the vector itself as a relatively abundant DNA target within the cell). One means of preventing such deleterious transposase activity is the use of orthogonal expression systems, as exemplified in *C. difficile* R20291 (Cartman & Minton, 2010). Within this study the transposase expression was driven by P_{tcdB}, derived from the toxin B gene of *C. difficile.* This is expressed in the presence of the *tcdR* encoded sigma factor which was naturally present within the *C. difficile* target strain but absent from the *E. coli* conjugal donor strains.

This orthogonal expression approach was subsequently demonstrated in recombinant *C. acetobutylicum* and *C. sporogenes* strains, in which the *tcdR* gene was integrated into the genome at the *pyrE* locus using ACE (Zhang, et al., 2015). A modified inducible *tcdR* system was integrated via ACE within *C. autoethanogenum* (Woods, et al., 2021). This *bgaR*-P_{bgaL}::*tcdR* sequence placed *tcdR* under the control of the lactose regulated P_{bgaL} promoter. This expression of this promoter is afforded by a lactose bound BgaR transcriptional regulator. As such, the introduced *Himar1C9* transposase gene, controlled by the P_{tcdB} promoter, could only be expressed following the 10 mM lactose induced expression of the *tcdR* gene within the genome of the recombinant strain.

5.1.4. pMTL-CW20 transposon mutagenesis vectors

The transposon mutagenesis vector applied within this study, pMTL-CW21, is based on the recently published pMTL-CW20 vector used within *C. autoethanogenum* (Woods, et al., 2021). Both vectors employ the *Himar1C9* transposase and *catP* mini-transposon previously exemplified within *C. sporogenes* and *C. acetobutylicum* (Zhang, et al., 2015). The *Himar1C9* transposase was originally isolated as *Himar1* from the mariner system of

Haemotobia irritans (Lampe, et al., 1996). Error-prone PCR was then used to generate diverse mutant sequences analysed by papillae-based screening in *E. coli*. This facilitated the identification of the hyperactive mutant *Himar1C9*, in which two amino acid changes, Q131R and E137K, synergistically conferred a 50-fold increase in *in vivo* transposition rates (Lampe, et al., 1999). The sequence requirement for this *mariner* transposase mediated genomic insertion is a thymine-adenine dinucleotide (TA). As an exceedingly prevalent sequence across any genome this transposase is unlikely to generate any insertional bias which may limit complete genomic coverage. Additionally, the transposition activity of *Himar1C9* does not rely on any host cell systems, ensuring broad species applicability.

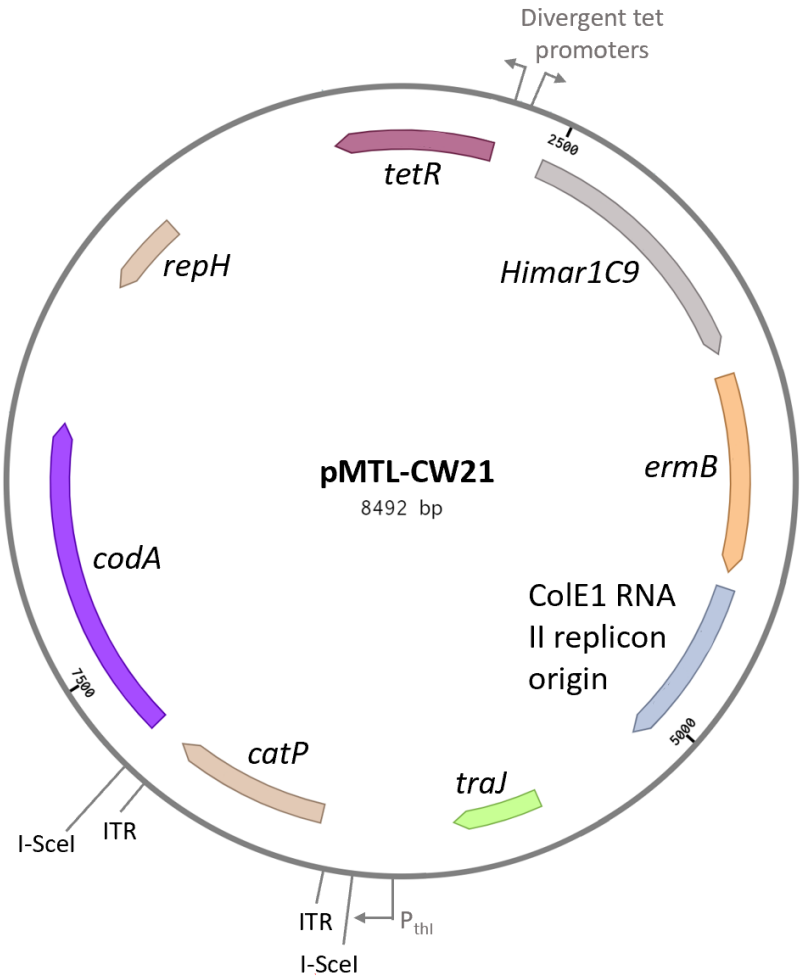


Figure 36 – pMTL-CW21 vector map
Transposon mutagenesis vector employed within this study, implementing the *Himar1C9* transposase with expression controlled by the Tet3nO derived divergent tetracycline inducible promoter system. The other divergent orientation promoter controlling expression of the *tetR* gene with read-through into the pCB102 replicon module, containing the *repH* gene. Selection

afforded by the *ermB* erythromycin resistance conferring gene, and Gram-negative replication and conjugation functionality afforded by the ColE1 replicon and *traJ* gene. Also, of note the *catP* mini transposon framed by the characteristic inverted terminal repeats (ITR) as well as two I-SceI recognition sites. This, in turn, preceded by the P_{thl} promoter system, conferring elevated expression of the cytosine deaminase encoded by *codA* upon transposition.

Another concern within *in vivo* transposon systems is the prevalence of plasmid-based reads at the sequencing stage. To combat this the pMTL-CW20 and pMTL-CW21 (**Figure 36**) vectors both contain a plasmid sequence removal system. This system relies on the inclusion of two I-SceI restriction enzyme recognition sequences framing the ITR sites of the transposon sequence. Their placement, outside of the transposed sequence, ensures that I-SceI sites will only be present within plasmid-derived DNA and not within the successful chromosomal integrants. Hence, during the I-SceI restriction digestion step of the transposon directed insertion sequencing (TraDIS) library preparation method, only plasmid-derived DNA undergoes cleavage. This cleavage occurs between the two primer binding sites (the adapter and transposon sequence), preventing any amplification of plasmid-derived DNA during the subsequent transposon-chromosome junction PCR enrichment step. This approach appeared effective within the published pMTL-CW20 study, with only 0.2% of rich medium library reads mapped to the transposon-delivery plasmid.

Both pMTL-CW20 and pMTL-CW21 also contain the *codA* gene, derived from *E. coli* and encoding a cytosine deaminase. Cytosine deaminase catalyses the conversion of supplemented 5-fluorocytosine to 5-fluorouracil, a lethal chemical causing occlusion of DNA and protein synthesis within the cell. As a part of a conditional lethality system this *codA* gene, without associated RBS and promoter sequence, is located downstream of the transposon. Upon transposition and excision of this region this *codA* gene is then placed under the control of an upstream constitutive P_{thl} promoter, causing elevated expression of cytosine deaminase. Hence transposition, in the presence 5-fluorocytosine, would be a lethal event.

Aside from prevention of premature transposition events, this system has also been noted to improve the stability of the vector within the *E. coli* cloning host, wherein assembly of an intact constitutive *codA* expression cassette appeared

lethal irrespective of 5-fluorocytosine supplementation (C. Woods, personal communication). The efficacy of this system was confirmed through the comparison of the stabilities of pMTL-CW20 and pMTL-YZ14 vectors (Zhang, et al., 2015) within *E. coli* TOP10. Analytic digestion of five plasmid isolates of each vector revealed three pMTL-YZ14 replicates with transposition conferred plasmid rearrangements, compared to zero in six replicates with pMTL-CW20. This counter-selection also proved effective in increasing the reproducibility of intact plasmid transfer to *C. autoethanogenum.*

5.1.5. Transposon mutagenesis in *E. limosum* ATCC 8486

The inducible orthogonal expression system exemplified within *C. autoethanogenum* for transpose regulation required initial generation and validation of a recombinant target strain. As such this expression system is not appropriate for use within undeveloped novel strains such as *E. limosum* NG-6849. In lieu of this pMTL-CW21 employs an anhydrotetracycline inducible promoter system, Tet3nO (Walker & Köpke, 2015). This system is comparable to that applied within *C. difficile* (Dembek, et al., 2015) with two divergent promoters which overlap a *tet* operator sequence. One of the promoters, P_{tetR}, drives the expression of a *tetR* gene. This gene encodes a transcriptional regulator which binds the *tet* operator region, effectively repressing both promoters. The second promoter, P_{tet}, can then be used to drive expression of the transposase gene, *Himar1* in the *C. difficile* study and *Himar1C9* within pMTL-CW21. Tetracycline induces a conformational change in TetR that prevents it binding to the operator region, relieving the repression and permitting the transcription of the divergent sequences. TetR can also efficiently bind anhydrotetracycline, which is more typically used within research applications due to its lower toxicity and similar passive membrane diffusion capacity (Gossen & Bujard, 1993).

This inducible promoter system not only tightly regulates transposition through transposase gene repression, but due to the divergent promoters can be used

to simultaneously induce plasmid loss. The plasmid instability mechanism of pMTL-CW21 functions in a comparable way to that previously described in *C. acetobutylicum* and *C. sporogenes* (Zhang, et al., 2015). In this instance, however, the destabilising pCB102 replicon read through is generated by the induced P_{tetR} promoter. Preliminary assessment and application of pMTL-CW21 has been previously performed within *E. limosum* ATCC 8486 (Millard, 2017). This assessment began with characterisation of the sensitivity of the strain to anhydrotetracycline, followed by Tet3nO validation with a *catP* reporter assay and preliminary pML-CW21 plasmid loss and transposition frequency assessment. This study concluded in the successful generation of single insertional mutants within the strain, of which 61 unique insertional sites were identified and mapped to the genome. This study effectively demonstrated the functionality of the core features of pMTL-CW21 within *E. limosum*, but due to the low genomic coverage left in question the potential insertional bias of the system.

5.2. Results

5.2.1. *E. limosum* NG-6849 genetic tractability

5.2.1.1. Genome

The MicrobesNG Enhanced Genome sequencing service was employed in an effort to generate a closed genome sequence for *E. limosum* NG-6849, a strain kindly provided by P. Soucaille (University of Toulouse, France). This is detailed in **2.5.1**. In brief the cell mass of an exponential phase culture was harvested and sent to MicrobesNG at ambient temperature, before undergoing DNA extraction and sequencing. The Illumina data generated was of high quality, with a mean coverage of 94.6x and median insert size of 650 bp. Following initial lysis optimisation an initial Oxford Nanopore sequencing run was performed with 25 µL of a 16 ng/µL sample with DNA integrity number (DIN) of 9.8. This unfortunately yielded only 0.408 Mb of sequencing, with 35.1% of

this corresponding to reads longer than 7 kb. A repeated attempt with the sample only generated a further 2.17 Mb of sequencing (with ~30% of this from reads longer than 7 kb). Further DNA extraction optimisation yielded 1.1 µg of DNA, with 25 µL of the 22 ng/µL sample (DIN 9.5) run once more, generating 15 Mb of sequencing reads. Of these reads only 5% exceeded 7 kb, so a further sequencing run was performed with the remaining sample which obtained 25 Mb of sequencing, but again with only 5% of reads exceeding 7 kb. Despite these repeated attempts the final coverage with these crucial longer reads remained very low (<1x), and without sufficient long read sequencing data the genome was not successfully closed within this study, instead yielding 27 contigs. The ATCC 8486 published genome was not deemed a suitable template for alignment based genomic closure due to previously observed significant structural rearrangements between *E. limosum* species (Mansfield, 2018)

Initial analysis was performed of the core WLP genes by BLASTn alignment of the respective carbonyl and methyl gene loci of *E. limosum* ATCC 8486 to each contig. All WLP genes were identified, similarly within the two distinct operons and with 100% sequence identity of the coding regions. All further alignments performed to elucidate the C1 metabolism, acetate and butyrate production pathways also demonstrated 100% identity. The corresponding genes, according to their ATCC 8486 loci are indicated in **Figure 37** and **Table 28**. The pathway for butyrate synthesis is somewhat obtuse, with two candidate phosphate butyryltransferases yet no apparent butyrate kinase (the only gene identified by tBLASTn analysis being a poor alignment with acetate kinase). This is not uncommon in butyrate-producing bacteria within the human colon, which typically contain a butyryl-CoA: acetate CoA-transferase which transfers the acetyl moiety of butyryl-CoA to acetate (Louis, et al., 2004).

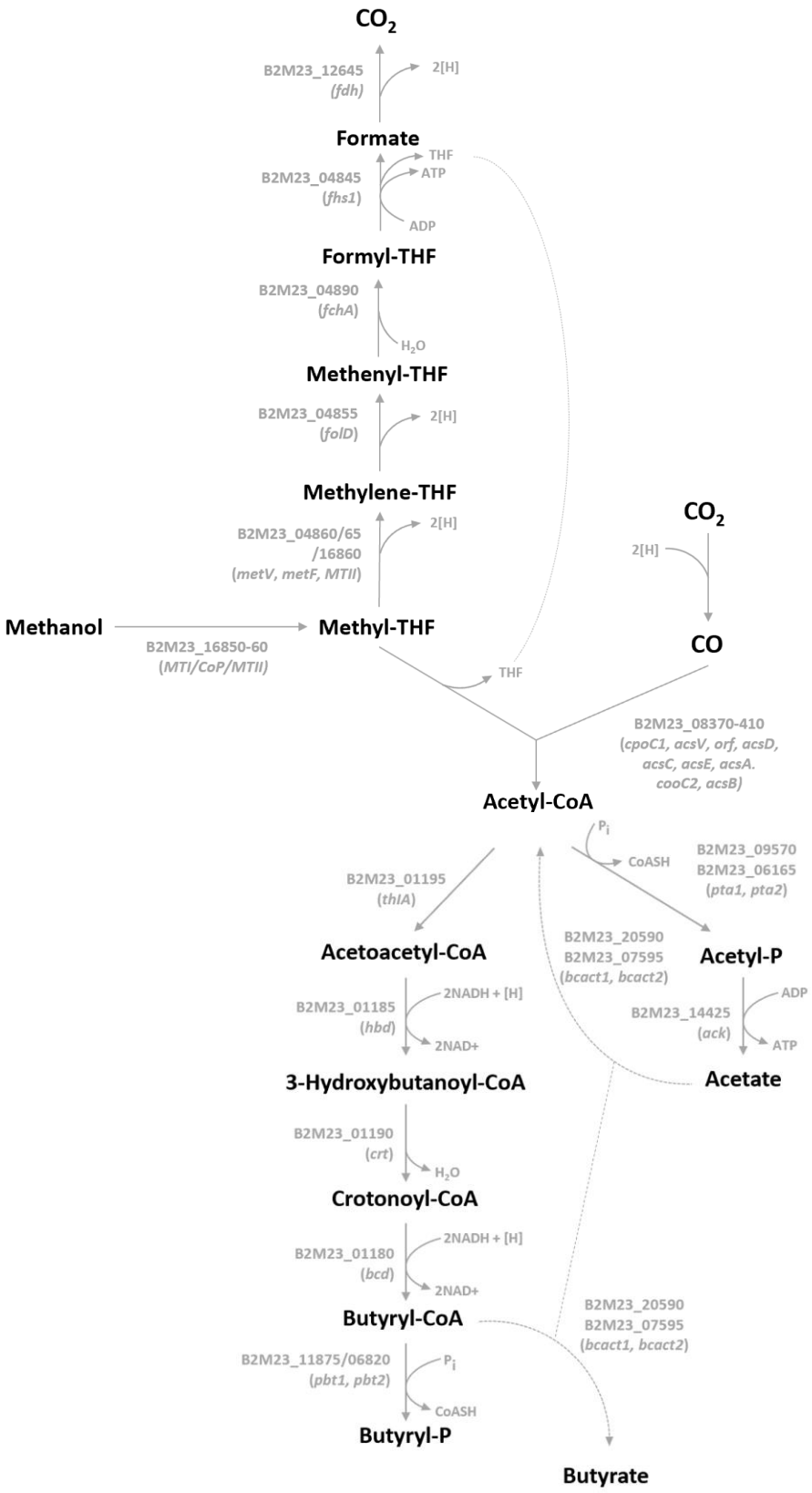


Figure 37 – Schematic of C1 metabolism and butyrate/acetate synthesis in *E. limosum* NG-6849, via the WLP with enzymes summarised in Table 28.

Locus tags derived from *E. limosum* ATCC 8486 annotations, subsequently aligned to the *E. limosum* NG-6849 genomic sequences with BLAST. The WLP genes are assembled into two

distinct loci corresponding to the methyl and carbonyl branches. Formate dehydrogenase is the only gene located independently.

Table 28 – Enzymes involved in core C1 metabolic pathways, and subsequent acetate and butyrate synthesis within *E. limosum* NG-6849.

Enzyme	Abbreviation	locus tag (B2M23_)
Formate dehydrogenase	*fdh*	12645
Formate-THF synthetase	*fhs1*	04845
Formyl-THF cyclohydrolase	*fchA*	04890
Methylene-THF dehydrogenase	*folD*	04855
Methylene-THF reductase	*metV, metF, MTII*	04860, 04860, 16850
CO dehydrogenase/acetyl-CoA synthase	*cpoC1, AscV, orf, acsD, acsC, acsE, acsA, cooC2, acsB*	08370, 08375, 08380, 08385, 08390, 08395, 08400, 08405, 08410
methanol: coenzyme M methyltransferase	*MTI*	16850
Corrinoid/iron-sulphur protein	*CoFeSP*	16855
Thiolase	*thlA*	01195
3-hydroxybutyryl-CoA dehydrogenase	*hbd*	01185
Crotonase	*crt*	01190
Butyryl-CoA dehydrogenase	*bcd*	01180
Phosphate butyryltransferase	*pbt1/2*	11875, 06820
Phosphate acetyltransferase	*pta1, pta2*	06165, 09570
Acetate kinase	*ack*	14425
Butyryl-CoA: acetate CoA – transferase	*bcact1, bcact2*	07595, 20590

5.2.1.2. *E. limosum* NG-6849 and ATCC 8486 genome comparison

Although initial genomic assembly methods were curtailed by a lack of long-read sequencing, towards the end of this study P. Soucaille (concurrently pursuing research with *E. limosum* NG-6849) provided a preliminary closed genome sequence for this strain. This was used a reference for Illumina read mapping for the previously attained sequencing data (mapping statistics summaries in **Table 29**), and a consensus sequence attained using the CLC Genomics workbench. This process yielded a highly similar sequence to that supplied by Soucaille, with 17 SNPs identified, of which only two exceeded a

frequency of 90%, generating a S24F mutation in a prepilin peptidase gene (B2M23_05875) and a X303Y mutation in a pseudogene (B2M23_17485).

Table 29 – Summary report of *E. limosum* NG-6849 Illumina read-mapping against both the published *E. limosum* ATCC 8486 genome *(Song & Cho, 2015)*, and the received NG-6849 draft kindly supplied by P. Soucaille.

Reads	Read Count	Reads (%)	Average length	Base count	Bases (%)
E. limosum ATCC 8486 genome alignment					
Mapped	1,796,422	99.83	234.28	420,860,162	99.94
Not mapped	3130	0.17	83.51	261,382	0.06
E. limosum NG-6849 draft closed genome alignment					
Mapped	1,796,462	99.83	234.28	420,869,884	99.94
Not mapped	3090	0.17	81.44	251,660	0.06

To further asses the relatedness of NG-6849 to the type strain ATCC 8486 a genome comparison was performed using progressiveMauve alignment (Darling, et al., 2010). This identified four local collinear blocks (minimal weight 11590) of 1286651 bp, 2374389 bp, 73242 bp and 11589 bp respectively (**Figure 38**). Also identified were 421 SNPs of which 53 were gaps, twelve of a single nucleotide polymorphism, eight dinucleotide polymorphisms, sixteen between 3-7 nucleotide changes, four between 29-62 bp and the remaining thirteen exceeding 100 nucleotides (see Appendix **Table 41**). Further gap analysis highlighted four key regions of variation in which divergence led to multiple calls across the region, and six further significant gaps (>30 bp) (**Table 30**). The gap at 4395297 bp appears to correspond to the end of the third locally colinear block, and 4406928 the fourth.

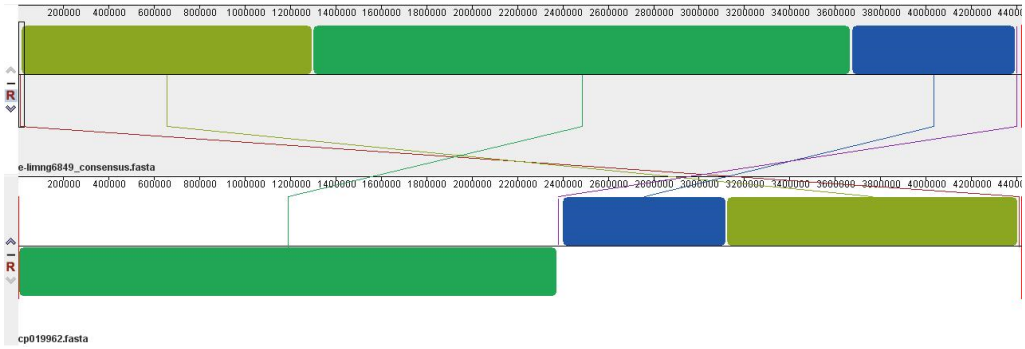


Figure 38 – Graphical alignment of the two *E. limosum* genomes NG-6849 genome and ATCC 8486.
Alignment was performed using progressiveMauve genome comparison software (Darling, et al., 2010), with a seed weight of 15 and locally colinear blocks exceeding a weight of 11590. Both genomes are displayed linearly, with the upper horizontal sequence the NG-6849 chromosome, and the lower the ATCC 8486 chromosome. Each coloured region represents a homologous region between the two genomes, otherwise denoted as a locally colinear block.

Table 30 – Overview of variation between the *E. limosum* NG-6849 genome and *E. limosum* ATCC 8486 genome (Song & Cho, 2015)**.** Table of inter-genome variation collated into four primary regions of difference (A-D) and a fifth group containing all remaining significant gaps identified.

Region	NG-6849 Start position	NG-6849 End position	Region Length (bp)
A	864675	864690	15
	864800	864920	120
	865015	865030	15
B	1701570	1701825	255
	1702190	1702380	190
	1702805	1703065	260
C	4406820	4406830	10
	4406880	4406890	10
	4406910	4406928	18
D	71035	71065	30

Position	Gap Length
1727594	1662
4395297	62
4406928	10427
1442434	31
1298056	1241
74022	2900

5.2.1.3. Antibiotic sensitivity

To ensure the applicability of *ermB* and *catP* selection markers for plasmid maintenance within the strain, preliminary assessment of the antibiotic sensitivity of *E. limosum* NG-6849 to clarithromycin and thiamphenicol was performed as detailed in **2.12**. In brief three exponential phase *E. limosum*

cultures were standardised and passaged into comparable growth conditions, supplemented with a range of antibiotic concentrations. This was then cultivated for 24 hours before OD_{600} readings were taken in technical triplicate and blanked against the average OD_{600} value of non-inoculated media. Both clarithromycin and thiamphenicol prevented *E. limosum* NG-6849 growth for 72 hours at concentrations greater than 2.5 µg/mL (**Figure 39**). Hence, for robust selection in further research concentrations of 7.5 µg/mL thiamphenicol and 6 µg/mL clarithromycin were used respectively.

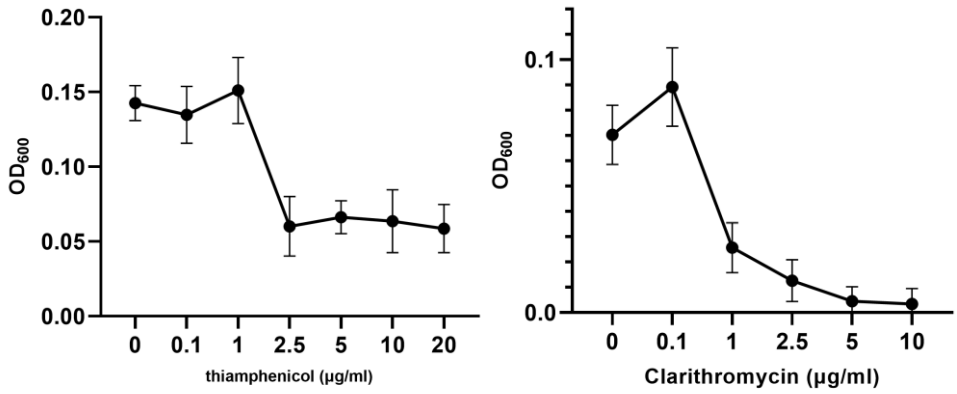


Figure 39 – *E. limosum* NG-6849 24-hour growth OD_{600} values in the presence of clarithromycin (0-10 µg/mL) and thiamphenicol (0-20 µg/mL).
Average OD_{600} values of biological quadruplicate *E. limosum* NG-6849 broth cultures supplemented with 0-20 µg/mL thiamphenicol and 0-10 µg/mL clarithromycin, recorded at 24 hours of growth. Error bars represent standard deviation.

5.2.1.4. pMTL80000 transformation

Transformation of *E. limosum* NG-6849 was attempted using electroporation as previously implemented successfully by J. Millard with strain ATCC 8486, with electrocompetent cell preparation as detailed in **2.3.5**. In brief, an early exponential *E. limosum* culture was placed on ice for 15 minutes and the cells subsequently centrifuged and washed twice by resuspension in chilled anaerobic SMP buffer. The final resuspension was performed using 10% (v/v) DMSO SMP buffer and separated into 100 µL aliquots in prechilled, de-gassed 2 mL Cryolock tubes sealed for later use. Triplicate electroporation was then performed as detailed in **2.3.6**, with 5 µg of various pMTL8*151 vectors, each

with an alternate Gram-positive replicon (for pMTL80000 schematic see **Figure 5**). All replicons, excluding pBP1 derived from *C. botulinum* NCTC2916, were successfully transferred, with the highest transfer efficacy at 2 x 10^6 CFU/µg DNA observed when using p19 (P value of 0.0046 compared to second highest pMTL83151), the *C. carboxidivorans* P7 plasmid derived replicon modularised by Millard (2017) (**Figure 40**). These transfer rates, using a base vector, were promising for further application of an exogenous genetic tools, including genetic modification machinery whose presence may reduce overall transfer efficacy due to vector size or deleterious activity. Efforts proceeded with pCB102, due to the ongoing characterisation and modularisation efforts in p19 and ready availability of 8315* sequences. The transfer rates observed for pMTL83151 were also deemed sufficient for preliminary GM tool application, with p19 an alternate option should pCB102 transfer rates prove limiting, particularly in achieving sufficient genome saturation with transposon mutagenesis.

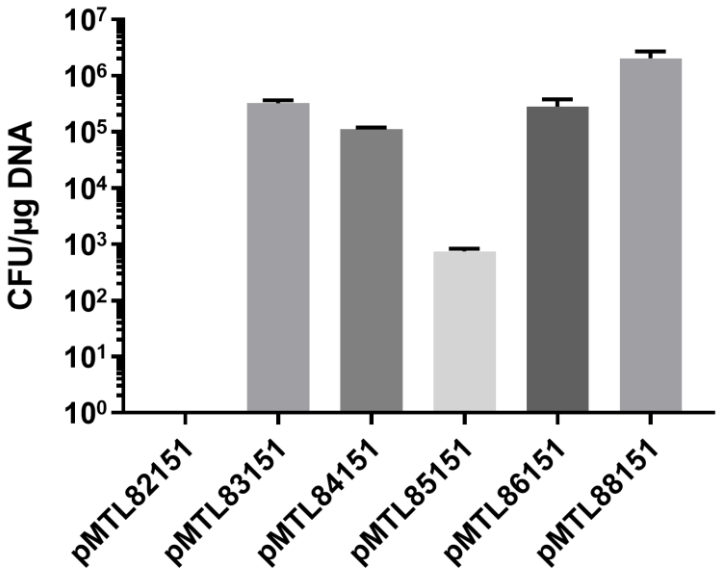


Figure 40 – Log10 DNA transfer efficiency of pMTL80000 vectors into *E. limosum* NG-6849 by electroporation.
Frequency determined *via* transformant CFU number yielded by 5 µg of DNA of pMTL8*151 vectors with *catP* selection marker, ColE1 Gram-negative replicon and varied Gram-positive replicons. 2: pBP1. 3: pCB102. 4: pCD6. 5: pIM13. 6: pIP404. 8: p19. Values presented represent mean transformant CFU per µg of DNA, extrapolated from appropriate dilution plates, with error bars reflective of the standard deviation of each biological triplicate.

5.2.1.5. Theophylline inducible riboswitch assessment

As previously described, a theophylline inducible expression system consisting of a theophylline responsive riboswitch, a sequestered synthetic RBS and constitutive promoter was exemplified by I. Cañadas for application within clostridial species (Cañadas, et al., 2019). To assess the applicability of this system for inducible expression within *E. limosum* NG-6849 the sensitivity of the strain to an appropriate range of theophylline concentrations was first assessed as detailed in **4.2.1.3**, with Pacaud media used in lieu of YTAFM. Once more the initial bacterial growth appeared marginally decreased, or entirely absent, at theophylline concentrations lower than 0.5 mM. Also, no significant growth impact of theophylline concentrations of up to 10 mM on *E. limosum* NG-6849 (**Figure 41**). As such theophylline was deemed an appropriate supplement for use the strain at this threshold, and further characterisation of the system was pursued.

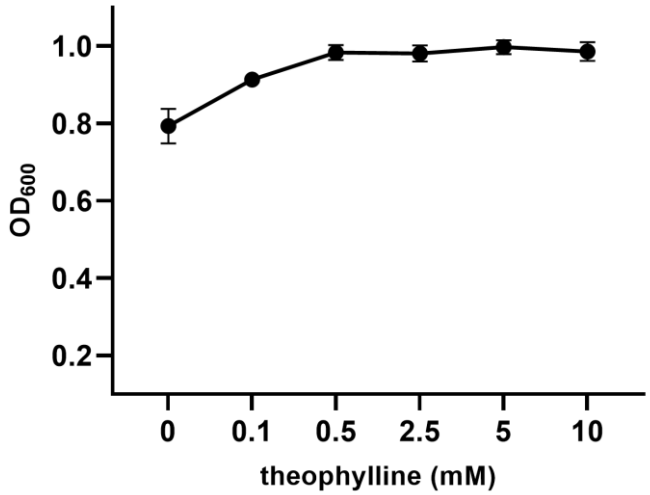


Figure 41 – Impacts of 0-10 mM theophylline on *E. limosum* NG-6849 growth.
Average OD_{600} values and standard deviation of triplicate *E. limosum* NG-6849 cultures grown for 24 hours in Pacaud broth supplemented with yeast extract, glucose and 0-10 mM theophylline.

Eight riboswitches are detailed within the aforementioned study: RbD, RbE, RbE*, RbF, RbG, RbH, RbI and RbJ. The *catP* reporter vectors of each of these (pMTL-IC101-Rb*) were kindly provided by I. Cañadas (excluding RbE) and the Gram-positive replicon pBP1 was exchanged for pCB102 as detailed in **2.9.11**. These seven vectors were then used to transform *E. limosum* NG-6849

alongside a no promoter *catP* reporter vector control and the unmodified $P_{fdx(Csp)}$ promoter, before a *catP* reporter assay was performed as detailed in **2.11.** This included growth on Pacaud broth supplemented with glucose, yeast extract, clarithromycin and 5 mM theophylline for riboswitch induction. The resultant CatP activity indicated slightly reduced levels of expression across all induced riboswitches compared to the $P_{fdx(Csp)}$ vector assayed (**Figure 42**). The greatest induced activity was observed with the pMTL83254-RbE* vector, at an average activity of 23.05 U/mg. Of all riboswitches assayed only RbJ, RbE* and RbF demonstrated significant difference in the induced and uninduced samples (confidence interval 99%). There was no statistical significance between the induced strength of RbE*, RbF and RbJ. RbE* also demonstrated relatively good repression, with an uninduced activity of 6.58 U/mg, compared to 5.60 U/mg for the no promoter control.

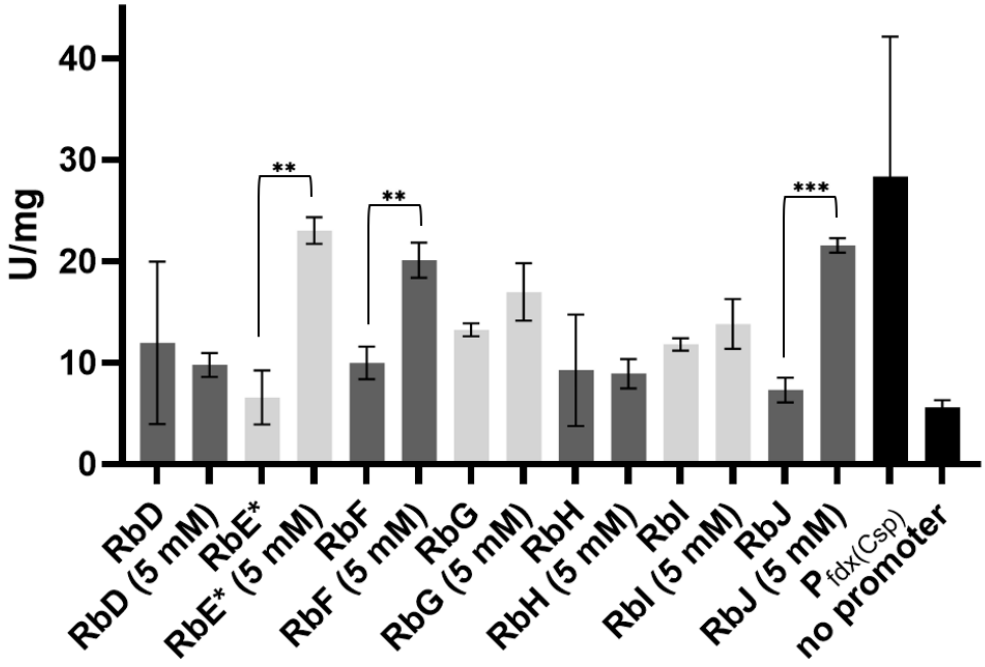


Figure 42 – $P_{fdx(Csp)}$-Rb* derived CatP expression levels within *E. limosum* NG-6849 upon induction with 5 mM theophylline, alongisde $P_{fdx(Csp)}$ and a no promoter control.
catP reporter assay derived enzymatic activity levels of CatP in *E. limosum* NG-6849 maintaining pMTL83254-Rb* reporter vectors in which *catP* expression is controlled by the $P_{fdx(Csp)}$ promoter and adjacent riboswitch structure. Induction was performed with 5 mM theophylline and P values calculated using Welch's t-test. Resultant U/mg values were determined according to the enzymatic activity identified through 5-thio-2-nitrobenzoate production of the cell lysate measured with OD_{412nm} readings. Final values adjusted according to BCA assay derived total protein contents. Performed in biological triplicate with error bars indicative of standard deviation.

5.2.1.6. pRECas genome editing

Previous efforts by J. Millard (2017) demonstrated the knockout the *R.ORFD* and *gmrSD* genes within *E. limosum* ATCC 8486 through the application of a pMTL40000 vector series CRISPR-Cas9 system (Ingle, et al., 2019). These genes represent the restriction subunit of a Type I RM system Eli8486ORFDP, and a putative adjacent Type IV RM system identified within *E. limosum* ATCC 8486 by REABASE® analysis. Alignment of this genomic region with the genome sequences acquired for NG-6849 revealed the presence of the same RM system genes within NG-6849, with 100% sequence identity for both genes and 1.5 kb of the downstream sequence. There was however a divergent ~700 bp region between the two genomes, beginning 75 bp upstream of *R.ORFD*. This is an intergenic region downstream of the terminal annotated specificity subunit of the Type I RM system.

As a previously validated target within ATCC 8486 this locus was chosen as a preliminary knockout target within NG-6849 to assess the amenability of CRISPR-Cas9 genome editing using the pRECas RiboCas plasmid system. Two vectors were constructed, as detailed in **2.9.12**. These vectors were based on the pMTL83151 modular vector series backbone, with corresponding *catP* selection marker and pCB102 Gram-positive replicon, with the riboswitch E*and P_{fdx} promoter driving the *cas9* gene expression as previously described. The two vectors differ solely in the promoter driving sgRNA transcription, with P_{araE} in pRECas(araE)-ORFD, and P_{thl} within pRECAS(thI)-ORFD (**Figure 43**). Both vectors contained a recombination cassette with two homology arms of 805 bp and a 779 bp, respectively. The homology arms primers and sgRNA sequence used within NG-6849 were identical to those previously applied within ATCC 8486.

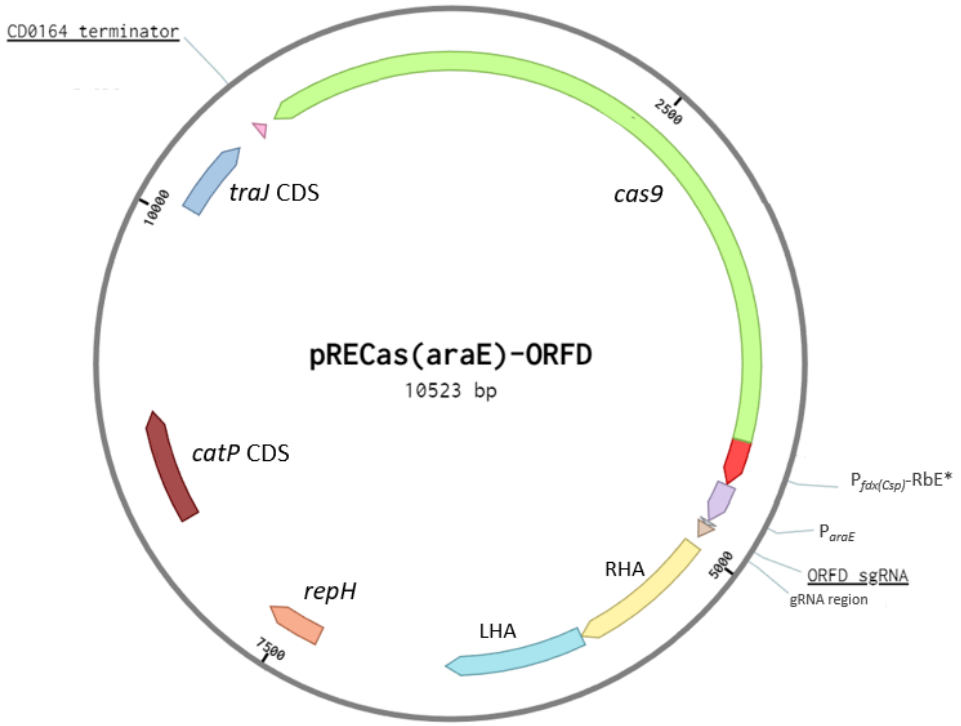


Figure 43 – pRECas(araE)-ORFD plasmid map
Schematic detailing the sequence features of the pRECas(araE)-ORFD, one of two CRISPR-Cas9 vector systems used to generate a *gmrSD* and *R.ORFD* knockout in *E. limosum* NG-6849. This vector differs from the alternate solely in the promoter used to express the guide RNA region, in this instance P_{araE}, and alternatively P_{thl}.

The two vectors were used to transform *E. limosum* NG-6849 as described in **2.3.6** and ten transformants of each vector were then re-streaked onto media supplemented with 5 mM theophylline, and thiamphenicol for plasmid selection. Following 48 hours of growth each of these colonies were screened by PCR with the primers ORFD Scr Fw and ORFD Scr Rv (**Table 10**). Both primers bind to regions of genomic DNA only, outside of the recombination locus, resulting in amplification of a 7043 bp PCR fragment from the wild-type locus. Of the colonies screened, four of the ten pRECas(thI)-ORFD transformants generated a truncated band size indicative of the excision of the 5.2 kb genomic fragment through double crossover recombination with the vector backbone (**Figure 44**). Of these four mutant bands, two also generated a band congruent with the wild-type fragment size and are hence mixed colonies from which pure mutant isolates could be attained through further re-streaking. With the pRECas(araE)-ORFD transformants only two colonies generated a knockout band, both alongside a wild-type band. Two knockout bands were extracted for

each vector and Sanger sequenced to confirm the gene deletion. The higher number of knockout genotypes generated by pRECas(thI)-ORFD could indicate the P_{thI} has a higher expression level than P_{araE}, providing a greater availability of the sgRNA sequence for CRISPR-Cas9 assembly and genomic targeting *in vivo*. However due to the low number of colonies screened at this stage, this cannot be ascertained without further research, such as a reporter assay.

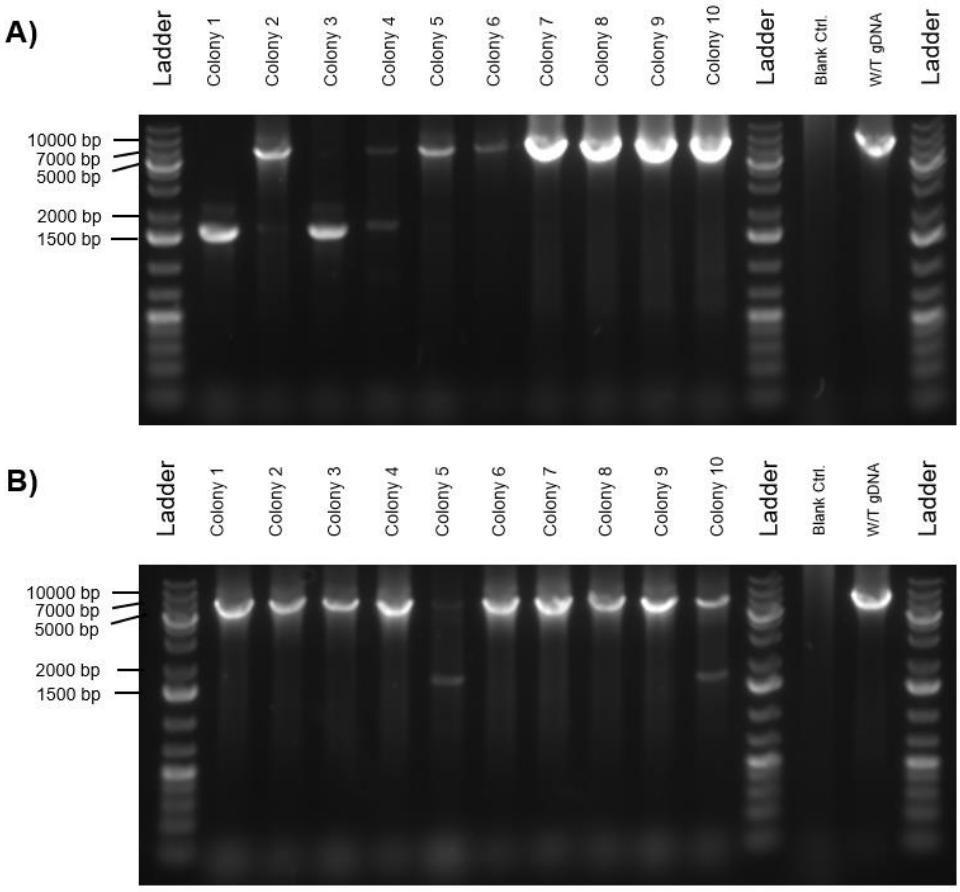


Figure 44 – PCR amplification of *R.ORFD gmrSD* locus within *E. limosum* NG-6849 colonies with pRECas(*) vectors following 5 mM theophylline induction.
Colony PCR with ORFD Scr Fw/Rv primers and a DreamTaq based PCR protocol. Expected PCR product size was 7043 bp for unmodified genome, and 1831 bp following gene excision. Blank controls used the same PCR/primer master mix but with no template, and positive controls utilised *E. limosum* NG-6849 genomic DNA.

Colony 1 and 3 of the induced pRECAS(thI)-ORFD transformants, the confirmed *R.ORFD* and *gmrSD* knock-out colonies, were then subjected to plasmid loss attempts as previously described for *C. carboxidivorans* (**4.2.3.2**). This included serial passage through non-selective Pacaud media (supplemented with 20 mM glucose, 2 g/L yeast extract) for 168 hours before the plating of 10^{-4} and 10^{-5}

dilutions. A total of 100 colonies from these dilutions were then replica plated on thiamphenicol and non-selective Pacaud agar, however, all screened colonies still exhibited Tm^R. As such no Δ*R.ORFD-gmrSD* strain without residual plasmid was successfully isolated during the course of this study.

5.2.2. Transposon mutagenesis of *E. limosum* NG-6849 with pMTL-CW21

5.2.2.1. Tet3nO validation

To determine the applicability of pMTL-CW21 for transposon mutagenesis in *E. limosum* NG-6849 the stringency and response of the Tet3nO promoter was first assessed. This was performed using a *catP* reporter assay, as detailed in **2.11**. In brief pMTL-JM102 was used to transform electrocompetent *E. limosum* NG-6849 and three resultant transformant colonies picked into liquid culture with clarithromycin selection. These were then grown until mid-exponential phase and cell mass harvested. Three further transformants were picked into clarithromycin selective liquid cultures, further supplemented with 100 ng/mL anhydrotetracycline. Cell mass was once again harvested in mid-exponential phase before final sample preparation, CatP activity assessment and BCA assay-based protein quantification. *E. limosum* NG-6849 maintaining pMTL8324_NP-CatP, a no promoter control, and a $P_{fdx(Csp)}$ as a positive control were also assayed in a similar fashion. The expression level observed upon induction was comparable to that previously observed in ATCC 8486 at 81.08 U/mg (**Figure 45**). This confirms the anhydrotetracycline responsiveness of the promoter within NG-6849 ($P < 0.01$, Welch's t-test), and the functionality of a key component of the pMTL-CW21 vector system within this strain.

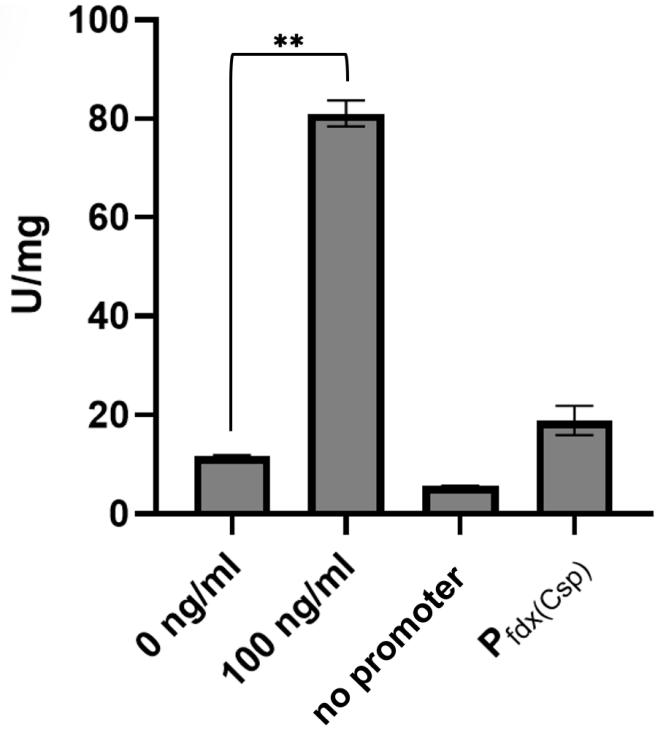


Figure 45 – Relative promoter strength of Tet3nO within *E.limosum* NG-6849 when incubated with 0 ng/mL and 100 ng/mL anhydrotetracycline respectively.
catP reporter assay derived chloramphenicol acetyltransferase activity levels within *E. limosum* NG-6849 maintaining pMTL-JM102, a Tet3nO reporter vector, in mid-exponential phase, with and without 100 ng/mL anhydrotetracycline induction (P 0.001275). Assayed alongside a comparable pMTL83254 backbone with no promoter or $P_{fdx(Csp)}$ driving *catP* expression. Resultant CatP values were calculated according to enzymatic activity of cell lysate through rate of 5-thio-2-nitrobenzoate production, measured according to OD_{412nm} readings. Final values adjusted according to BCA assay derived total protein content. Performed in biological triplicate with error bars indicative of standard deviation. The 100 ng/mL value is statistically significantly elevated compared to that of 0 ng/mL (P values of 0.0013 according to Welch's t-test).

5.2.2.2. Plasmid loss and transposition efficiency assessment

Following validation of the efficacy and stringency of the Tet3nO inducible promoter system, further assessment of the resultant transposition frequency and conditional plasmid loss frequency of pMTL-CW21 within *E. limosum* NG-6849 was performed. This began with the transformation of electrocompetent *E. limosum* NG-6849 with pMTL-CW21 as detailed in **2.3.6**, with transformant selection performed with both clarithromycin and thiamphenicol. CFUs were observed following 72 hours and total cell mass harvested at 96 hours. The cumulative bacterial slurry recovered from these plates was used to generate

a dilution series in PBS reaching 10^{-7}. Each of these dilutions were then plated onto three Pacaud plates (supplemented with 20 mM glucose and 2 g/L yeast extract) with different selective pressures: clarithromycin (Cla), clarithromycin and anhydrotetracycline (Cla+aTc), and thiamphenicol and anhydrotetracycline (Tm+aTc) respectively. The remainder of this bacterial slurry was stocked at -80°C following addition of 10% (v/v) DMSO.

The CFUs observed under each selective pressure were used to isolate and quantify a specific subset of the transformant population (**Figure 46**). Selection solely with Cla is permissive to any plasmid-bearing cells, representing the entire population as selected thus far. Selection with Tm+aTc should be permissive only to pure insertional mutants which now chromosomally bear the *catP* selective marker within the mini-transposon sequence but have lost the pMTL-CW21 vector due to replicon destabilisation. Selection with Cla+aTc permits cells which have retained the plasmid despite the induction of transposition and the associated plasmid instability system. This is of note as it quantifies the limitations or possible ineffectiveness of the plasmid loss mechanism.

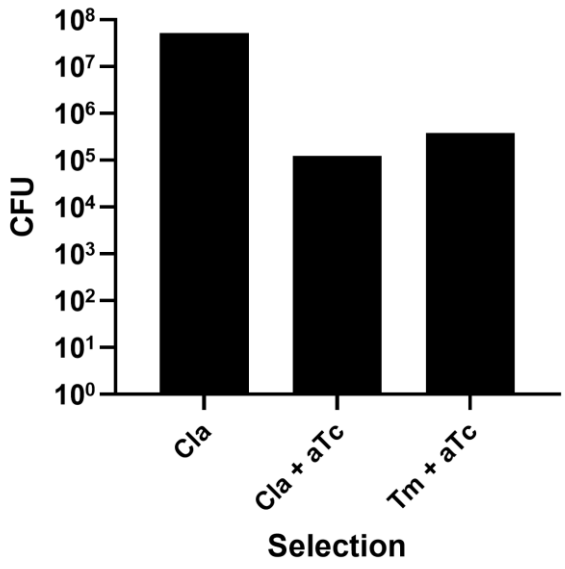


Figure 46 – Transposase activity and plasmid loss efficiency of pMTL-CW21 in *E. limosum* NG-6849.

An indicative result determined through assessment of the pMTL-CW21 transformant cell population capable of growth under different selection, with clarithromycin (Cla) selection representing the total transformant pool, thiamphenicol and anhydrotetracycline (Tm+aTc)

selecting for the *catP* transposon with induced transposase and plasmid instability, and the clarithromycin and anhydrotetracycline (Cla+aTc) selecting for cells which maintain the plasmid backbone despite plasmid instability and transposase induction.

The percentage of colonies arising on Tm+aTc plates relative to the total CFU on Cla is indicative of transposition frequency, with the missing CFU in the Tm+aTc condition largely representing those cells which failed to transpose successfully prior to plasmid loss. Within this assay the transposition frequency was, therefore, calculated as 9.54 x 10^{-9} %. The plasmid loss efficiency within this assay was calculated as 99.8%, according to the percentage of Cla^R+aTc colonies relative to the total population. This result appears to confirm that the plasmid destabilisation, through anhydrotetracycline-induced transcriptional readthrough of the Gram-positive replicon, is active within this strain. It is of note that while this may be an illustrative assay there are other factors which can also impact these CFU counts, such as single crossover integration of the vector within the chromosome generating a Cla+aTc^R phenotype.

5.2.2.3. *E. limosum* NG-6849 transposon mutagenesis

Following the confirmation of the suitability of pMTL-CW21 for transposon mutagenesis within *E. limosum* NG-6849, a transposon mutant library was generated as detailed in **2.13.1**. In brief the total bacterial slurry recovered following transformation of pMTL-CW21 was plated onto 32 large Pacaud agar plates supplemented with glucose, yeast extract, thiamphenicol and anhydrotetracycline. These plates were then grown anaerobically for 96 hours before a further bacterial slurry was attained by flooding each plate with 1 mL of Pacaud media, followed by agitation with a sterile wedge spreader and slurry aspiration. The final OD_{600} of this recovery was 13.6 and volume 12.8 mL. This was used to produce twelve 10% (v/v) DMSO stocks stored at -80°C to serve as inoculum in future outgrowth experiments, with the remaining 2 mL diluted and used to produce an initial transposon directed insertional sequencing (TraDIS) library for further assessment, as detailed in **2.13.2**. This was then sequenced by Deep Seq on the MiSeq platform, and resultant data mapped to

the *E. limosum* ATCC 8486 published genome (Song & Cho, 2015). This resulted in 48,887 mapped reads with a total of 22,241 unique insertions, approximately equivalent to an insertion every 200 bp and with no evidence of insertional bias.

5.2.3. TraDIS analysis of the essential genes required for the growth of *E. limosum* NG-6849 on methanol

To select for the essential gene profile of *E. limosum* NG-6849 corresponding to the use of glucose or methanol the transposon mutant library was grown within defined media, under 80% N_2 and 20% CO_2 headspace on each substrate as detailed in **2.13.1**. In brief, three of the library stocks were combined and used to inoculate a single serum bottle with Pacaud media supplemented with glucose, yeast extract, thiamphenicol and anhydrotetracycline. This was then used to inoculate triplicate serum bottles with defined media supplemented with either 20 mM glucose or 62 mM methanol, as well as thiamphenicol and anhydrotetracycline. This was serially passaged in exponential phase, with samples taken from each subculture prior to inoculation of the next. Unfortunately, the first iteration of this experiment was curtailed, with the second passage on methanol not exceeding an OD_{600} of 0.350 despite continuous culturing for more than 336 hours. This was attributed to an increased sensitivity observed in *E. limosum* NG-6849 to thiamphenicol in the presence of methanol, also observed within wild-type growth assays (M. Flaiz, personal communication). As such the assay was repeated with three new TraDIS library stocks as before, but with thiamphenicol selection removed following initial recovery.

To observe any significant variation in viable CFU on thiamphenicol, or residual clarithromycin resistance, dilutions of methanol replicate A were plated following each subculture. These did not show any notable drop in Tm^R colony number and, following the initial subculture, no cells exhibited Cla^R. When library preparation was performed using the samples of this second assay, a

reduction in yield from transposon-chromosomal junction PCR amplification was investigated with the glucose samples of the final subculture. This prevented acquisition of sufficient DNA yields for sequencing without further amplification which would encourage PCR associated bias. This was also a concern as it may imply a large wild-type population within the sample, which unfortunately did not undergo similar CFU assaying throughout the time course. As such, subsequent analysis focused on the glucose samples of the first run, and methanol samples of the second run (**Figure 47**).

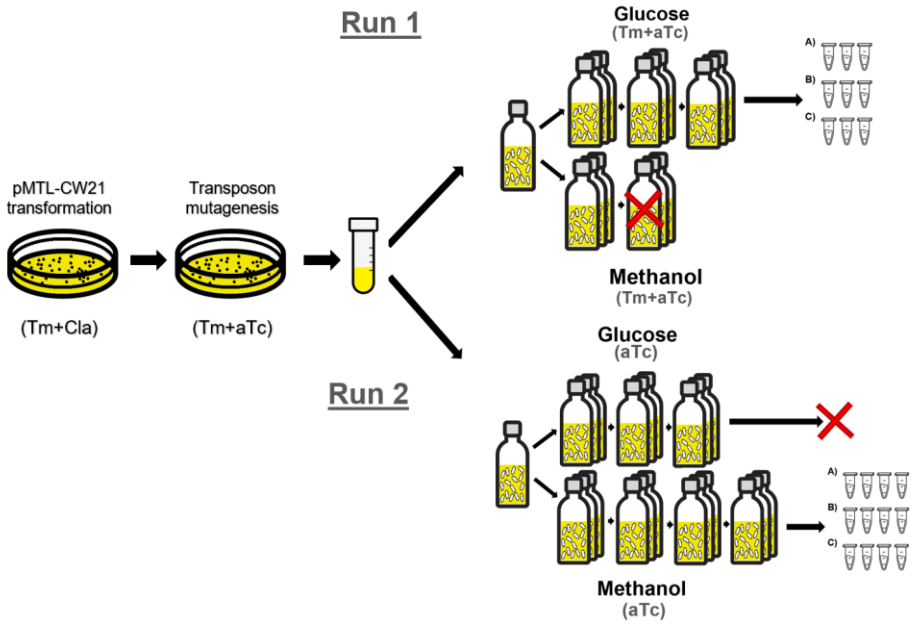


Figure 47 – Schematic view of the integral TraDIS mutant pool growth and selection steps.

TraDIS samples were prepared for each subculture as detailed in **2.13.2,** at timepoints equating to 35.5, 48.5 and 76 hours on defined glucose medium, and 95.5, 334.5, 430.5, and 478.5 hours on methanol. It was evident in growth tracking by OD_{600} measurements that the methanol cultures shifted carbon source only following depletion of the residual glucose within the media, reaching an OD_{600} of 0.6 within 100 hours in the first subculture, but with considerable lag observed in subculture two which only reached a comparable OD following 239 hours of growth. This lag was not apparent within the subsequent two subcultures as the strain had adapted to the new growth conditions, and the mutant population likely shifted towards those mutants

without detrimental genotypic changes causing delayed or no growth in the assay conditions.

TraDIS sequencing library preparations began with genomic DNA extraction followed by fragmentation of the DNA by sonication. The latter step was performed by DeepSeq due to COVID-19 restrictions, with conditions optimised according to preliminary parameter variations on a single replicate sample and micro-gel electrophoresis visualisation. Then all samples underwent adaptor ligation, plasmid backbone I-SceI digestion and genome-transposon junction amplification by PCR. These amplicons were size selected to an appropriate range for meaningful sequencing by gel electrophoresis, and quality ensured with the Qubit 3.0 fluorometer and Agilent 2100 Bioanalyzer. Following receipt of the samples on dry ice DBS genomics performed qPCR to standardise samples, preparing, pooling, and running the samples on a third party HiSeq platform as detailed in **2.13.3**. Sequencing library preparation was attempted for every sample, however, due to unforeseen issues sequencing data for the first subculture within glucose, and of replicate A in the fourth subculture of methanol was not successfully attained within the course of this study.

5.2.3.1. TraDIS read mapping and gene essentiality

Following receipt of the sequencing data, the Bio-TraDIS pipeline was used to map the reads to the *E. limosum* ATCC 8486 genome (Song & Cho, 2015) (**2.13.3**). This genome was used due to the lack of a closed and annotated NG-6849 genome at this stage, and the similarity observed between the genomes of the strains thus far (**5.2.1.1**). This genomic similarity between the strains was observed not only in the 100% sequence identities observed between the key WLP genes, but also the high read mapping statistics of the NG-6849 Illumina data against the published genome. As TraDIS data is also composed of Illumina reads, these high mapping statistics were particularly encouraging and the *E. limosum* ATCC 8486 genome was deemed an appropriate candidate for initial TraDIS analysis.

The mapped TraDIS Illumina data was then used to plot transposon insertion sites throughout the genome and extrapolate the predicted essentiality of each gene within the genome annotation according to insertion number and distribution (see Appendix **Figure 55**, to be deposited at ncbi.nlm.nih.gov/sra). Notably this pipeline disregards insertions within the terminal 10% of each gene, as these are deemed less likely to significantly impact the associated function (Barquist, et al., 2016). Each triplicate dataset was combined for this analysis, with resultant mapping statistics as indicated in **Table 31**.

By comparing the number of unique insertions to the total genome length a broad metric for coverage was generated approximate to the average frequency of insertions across the entire genome sequence. For the initial glucose dataset this is 25 bp, increasing to 49 bp in the third subculture as the mutant pool complexity decreases with non-viable mutants no longer represented within the data. Within the initial methanol sample this coverage is also reasonably high at 25 bp, falling to 43 bp, 60 bp and then 253 bp by the fourth subculture. The number of unique insertions relative to the number of insertions mapped also falls across the subcultures, illustrating this expected decline in pool diversity. To further assess the insertion site coverage the insertion site frequency of the earliest datapoints within each condition were plotted against the relative gene position, as can be seen in **Figure 48.** These plots have been capped at 10,000 read counts to provide better resolution of the lower frequency insertion sites, which appear prevalent throughout the entire sequence. These plots do not highlight any broad regions of exceedingly low coverage, appearing to demonstrate a lack of bias in the system and a diverse mutant pool.

Essential gene profiles were determined for each condition from the third subculture samples of each. In total 607 of the 4167 annotated genes were predicted as essential for growth in defined media supplemented with methanol (in excess CO_2), with a further 190 ambiguous at this stage. For glucose this essential gene profile was reduced to 547 genes, with only 74 ambiguous. Comparison of the combined essential and ambiguous gene

profiles of the third subculture under each condition revealed 103 non-synonymous essential genes (see Appendix **Table 40**) which were not predicted as essential for the growth of *E. limosum* NG-6849 on defined media with glucose but were essential when grown on methanol. Interestingly this subset of essential genes contains 37 hypothetical proteins with no associated predicted function.

Table 31 – Mapping statistics of TraDIS sequencing data with the *E. limosum* ATCC 8486 annotated genome *(Song & Cho, 2015)*.
Sample names indicative of glucose (G) or methanol (M) conditions, and respective subculture (S1-4). Analysis performed of biological triplicate samples with essential and ambiguous genes as predicted by the Bio-TraDIS pipeline analysis.

Sample	Reads			Total Unique Insertions	Gene Number		
	Total	Mapped	Percent-age (%)		Essential	Ambig-uous	Total
GS2	28,175,366	26,868,489	95	174,554	560	115	675
GS3	27,839,004	20,280,373	96	90,407	547	74	621
MS1	26,135,723	23,933,249	92	179,161	524	125	649
MS2	33,948,623	30,800,748	91	101,735	713	255	968
MS3	27,860,653	26,889,592	97	73,968	607	190	797
MS4	20,109,594	19,501,962	97	17,483	-	-	-

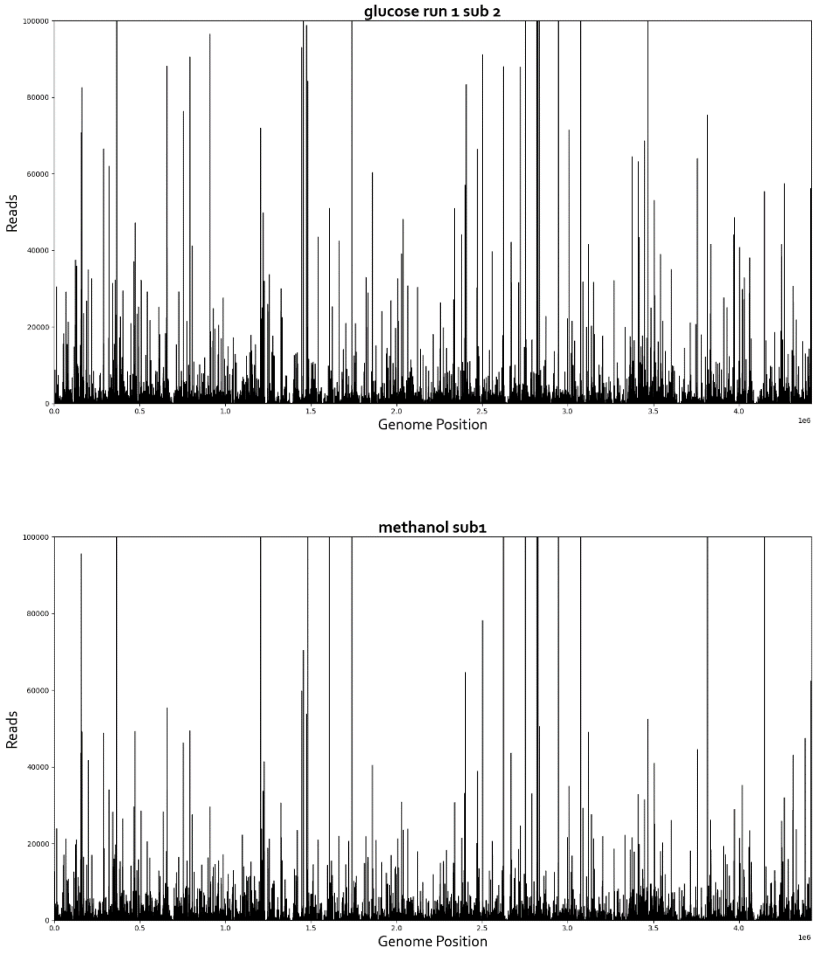


Figure 48 – Insertion plots of TraDIS datasets relative to the mapped genome sequence. Data plotted for the earliest available samples under each condition i.e., MS1 and GS2. Reads indicative of the number of sequencing reads mapped at that position, and therefore the number of transposon insertion sequences at each locus amplified from the original sample within the sample preparation pipeline. Reads have been capped at 100,000 to ensure sufficient resolution of lower frequency insertions.

5.2.3.2. Genes involved in methanol metabolism via the Wood-Ljungdahl pathway

As previously outlined, the best characterisation of methanol metabolism using the Wood-Ljungdahl pathway (WLP) has been achieved within *Moorella thermoacetica* and *Acetobacterium woodii* (Pierce, et al., 2008; Kremp, et al., 2018). The combination of three genes *mtaA/MTII, mtaB/MTI* and *mtaC/CoP*, forms the methanol:coenzyme M methyltransferase complex, in which MtaB mediates the transfer of the methyl group of methanol to the MtaC corrinoid prosthetic group. MtaA then catalyses the transfer of the methyl group from

MtaC to tetrahydrofolate to form methyl-THF (Ragsdale & Pierce, 2008). This proceeds according to the traditional WLP with the methyl group of methyl-THF transferred to a corrinoid iron sulphur protein before being combined with CO produced by the carbonyl branch via the ACS/CODH complex to form acetyl-CoA. The other key pathway for methanol metabolism, occurring concurrently with this route, proceeds by a reverse methyl branch process, generating CO_2 and six reducing equivalents from one molecule of methanol. This pathway begins with a steeply endergonic methylene-THF reductase catalysed reaction. As such four methanol molecules yield three acetyl-CoA and 0.83 ATP/acetate within *A. woodii* (Kremp, et al., 2018).

Initial genomic analysis of the NG-6849 strain through BLAST alignment to the *E. limosum* ATCC 8486 genome sequence identified six 100% homologous gene candidates for these integral methanol metabolising enzymes. These included B2M23_04860 and B2M23_04865 (otherwise denoted as *metV* and *metF*); prospective methylene-THF reductases within the methyl WLP gene cluster. Additionally, one independent methylene-THF reductase candidate, B2M23_18520, was identified and denoted as *orfMetF*. The candidate MtaABC complex genes were also clustered with locus tags of B2M23_16850 to B2M23_16860 predicted as encoding a methanol-corrinoid methyltransferase (*mtaB/MTI*), corrinoid protein (*mtaC2/CoP*) and methyltetrahydrofolate cobalamin methyltransferase (*mtaA/MTII*), respectively. These are also partially homologous to a comparable cluster in *A. woodii* of *mtaB* (Awo_c22760, 71% identity), *mtaC2* (Awo_c22750, 69% identity) and *mtaA* (Awo_c22740, 77% identity) (Kremp, et al., 2018).

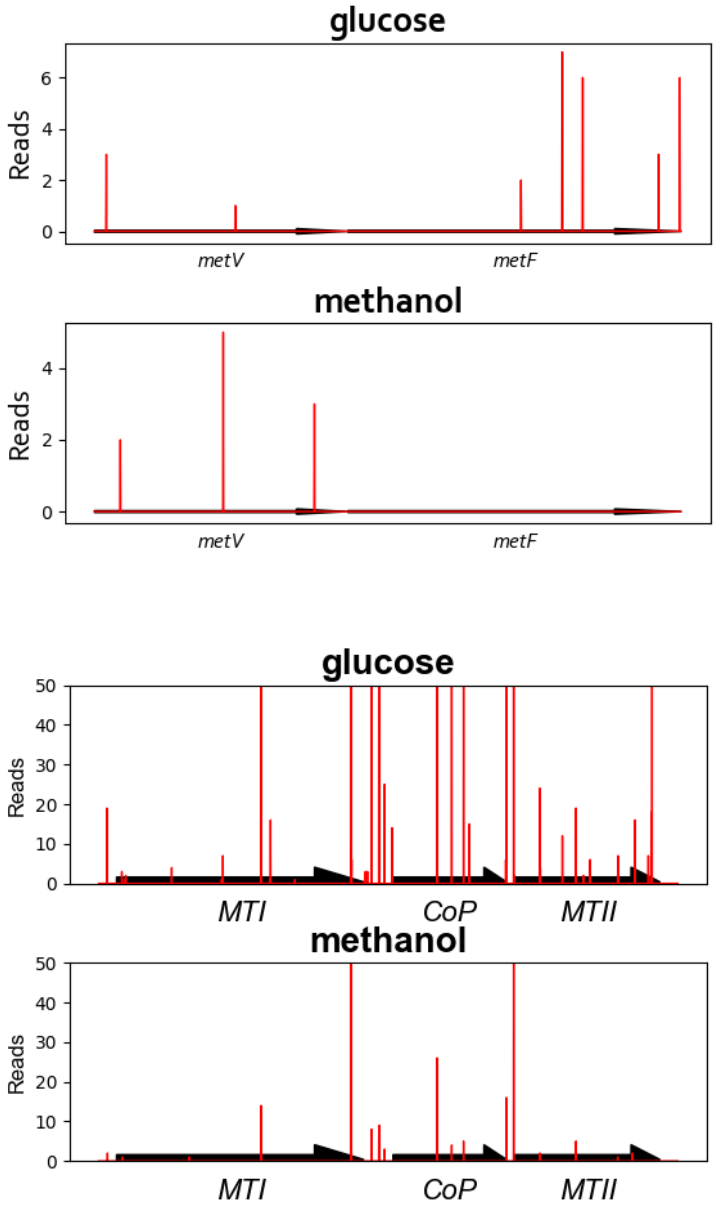


Figure 49 – Insertion plots of candidate methanol metabolism gene clusters within subculture 3 on methanol and glucose respectively. Specifically, the putative MtaABC cluster and two methylene-THF reductase genes *metF* and *metV*. Plots have been capped at 50 reads to ensure good resolution of lower frequency insertions.

Of these six candidates only two were predicted as essential within the methanol defined medium growth replicates, these being the *MTI/mtaB* (**Figure 49**) and *metV.* To identify further candidate *mtaA* and *mtaC* genes the corresponding enzymes within *M. thermoacetica* were aligned using tBLASTn to the NG-6849 genome sequence. When performed with the MtaA encoding gene (Moth_1208) 20 partial alignments were identified, all of which were annotated as methyltransferases within the *E. limosum* ATCC 8486 genome,

and only one of which was predicted as essential for methanol growth; B2M23_08240 annotated as a homocysteine methyltransferase (see Appendix **Table 40**). This was however, also predicted as essential within the glucose conditions. The alignment of the MtaC (Moth_1209) (Das, et al., 2007) only yielded one candidate, the previously identified *CoP* gene (B2M23_16850) which was not predicted as essential to growth on methanol (**Table 32**). Alignment of the final MtaB candidate (Moth_2346) generated three low homology regions, the corresponding genes of which were annotated as uroporphyrinogen decarboxylases (Uro-D), but once more none of which were predicted as essential for *E. limosum* growth on methanol.

To further the search both the *E. limosum* SA11 and *E. callanderi* KIST 612 genomes were also analysed, with any candidate genes within the published genome annotations aligned by BLASTn to the NG-6849 genome to identify any homologues. While none of the SA11 MtaA family gene homologues within the NG-6849 genome were predicted to be essential within the methanol dataset, one Uro-D gene within the KIST-612 had a homologue within NG-6849 which was predicted as essential. Uro-D enzymes catalyse the sequential decarboxylation of the four acetyl side chains of uroporphyrinogen to produce coporphyrinogen, part of the porphyrin synthesis pathway. These enzymes, however, have an interesting structural overlap with methyltransferase genes involved in C1 metabolism (MtaA/CmuA) which leads to some overlap in automated annotation.

Concurrently with this study, work was performed by A. Agius in the generation and characterisation of various gene knockouts within *E. limosum* NG-6849. This included *ΔMTI, ΔCoP* and *ΔMTII*, all of which showed significant perturbation of growth on methanol, with no evidence of growth within 140 hours (A. Agius, personal communication). Contrary to this the *ΔorfmetF* strain demonstrated no perturbation of growth on methanol. An attempt was also made to knockout *metVF* (both genes concurrently) which was not successful, perhaps indicating the essentiality of these under the conditions of mutant generation.

Table 32 – Candidate methanol metabolism genes with associated locus tag, functional annotation, and insertion index values. All methanol insertion index values below 0.0029 predicted as essential (white), and 0.0029-0.0046 ambiguous (light grey). All glucose insertion index values below 0.002 predicted as essential (white), and 0.002 to 0.003 ambiguous (light grey). Those genes which were not predicted as essential are indicated in dark grey.

	Locus tag	Gene	Enzyme	ATCC 8486 Locus tag (B2M23_)	Insertion index	
					Methanol	Glucose
E. limosum ATCC 8486	16850	*MTI mtaB*	methanol–corrinoid methyltransferase	16850	0.0024	0.0088
	16855	*CoP mtaC2*	corrinoid protein	16855	0.0052	0.0104
	16860	*MTII mtaA*	methyltetrahydrofolate cobalamin methyltransferase	16860	0.0054	0.0136
	04860	*metV*	methylene-THF reductase	04860	0.0050	0.0033
	04865	*metF*	methylene-THF reductase	04865	0.0000	0.0037
	18520	*orfmetF*	methylene-THF reductase	18520	0.0206	0.0232
E. limosum SA11 (ACH52_)	2399	*mtaA*	methyltransferase MtaA/CmuA family	15120	0.0096	0.0144
	2967	*mtaA*	methyltransferase MtaA/CmuA family	08135	0.0353	0.0084
	3803	*mtaA*	methyltransferase MtaA/CmuA family	02650	0.0202	0.0181
	3266		hydroxymethylbilane synthase/uroporphyrinogen-III methyltransferase	11550	0.0110	0.0110
E. callanderi KIST612 (ELI_)	0771		uroporphyrinogen III synthase / methyltransferase	11550	0.0110	0.0110
	2005	*MTII*	methyltetrahydrofolate: corrinoid/iron-sulfur protein methyltransferase	16860	0.0054	0.0136
	3604	*acsE*	methyltetrahydrofolate: corrinoid/iron-sulfur methyltransferase protein	08395	0.0042	0.0056
	4468		Uroporphyrinogen decarboxylase (URO-D)	01960	0.0027	0.0108

While the *metF* and the *MTI/mtaB* (B2M23_16850) candidate genes were confidently predicted as essential under methanol growth conditions, the corresponding corrinoid protein gene (*CoP/mtaC/*B2M23_16855) and methyltransferase gene (*MTII/MtaA*/B2M23_16860) of the cluster were not predicted as essential in either condition. Through BLASTp alignment of the corresponding gene products involved in the elucidated methanol metabolism pathways of *M. thermoacetica* and *A. woodii*, combined with the manual

screening of available *E. limosum* genomes, one *mtaA/MTII* candidate was identified exhibiting essentiality only in methanol and not glucose-based growth: B2M23_01960. Unfortunately, the BLASTp analysis revealed no significant sequence similarity to Moth_2346 or Awo_c22740, the candidate MtaA/MTII-encoding genes of *M. thermoacetica* and *A. woodii* (Pierce, et al., 2008; Kremp, et al., 2018). As such to confirm any indicated significance it would be prudent to perform further research and characterisation.

5.2.3.3. Essential genes within the Wood-Ljungdahl pathway

The TraDIS determined gene essentiality predictions for well conserved WLP gene clusters within ATCC 8486 were collated (**Table 33**; **Figure *50***), highlighting the integral nature of *cpoC1, fhs1* and *metF.* Interestingly *fhs1* and *folD* were also predicted as essential or ambiguously essential within the *E. limosum* replicates grown on glucose. Several genes were ambiguously predicted within the methanol samples including *acsD, acsC, acsE, cooC2, acsB, fchA* and *folD.* None of the remaining WLP genes within the identified clusters were predicted as essential for *E. limosum* NG-6849 growth on glucose.

Table 33 – Core WLP genes with associated locus tag, functional annotation, and insertion index values. All methanol insertion index values below 0.0029 predicted as essential (white), and 0.0029-0.0046 ambiguous (light grey). All glucose insertion index values below 0.002 predicted as essential (white), and 0.002 to 0.003 ambiguous (light grey). Those genes which were not predicted as essential are indicated in dark grey.

	ATCC 8486 locus tag (B2M23_)	Gene	Enzyme	Insertion index	
				Methanol	Glucose
carbonyl cluster	08370	*cpoC1*	corrinoid activation/regeneration protein	0.0028	0.0086
	08375	*acsV*	Putative metal-binding protein	0.0051	0.0173
	08380	*orf*	Hypothetical protein	0.0119	0.0325
	08385	*acsD*	CO dehydrogenase/acetyl-CoA synthase delta subunit	0.0035	0.0083
	08390	*acsC*	CO dehydrogenase/acetyl-CoA synthase gamma subunit	0.0032	0.0164

	08395	*acsE*	Putative methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase	0.0042	0.0057
	08400	*acsA*	CO dehydrogenase/acetyl-CoA synthase delta subunit	0.0070	0.0240
	08405	*cooC2*	CODH nickel insertion accessory protein	0.0043	0.0099
	08410	*acsB*	Bifunctional acetyl-CoA decarbonylase/synthase complex subunit alpha/beta	0.0046	0.0098
FDH cluster	12645	*fdh*	formate dehydrogenase (FDH)	0.0123	0.0153
	12650	*mob*	molybdopterin-guanine dinucleotide biosynthesis protein	0.0316	0.035
	12655	*fdhD*	FDH-accessory protein	0.0187	0.0159
methyl cluster	04840	*MTIII*	methyltetrahydrofolate cobalamin methyltransferase	0.0217	0.039
	04845	*fhs1*	formate-THF ligase	0.0000	0.0000
	04850	*fchA*	methenyl-THF cyclohydrolase	0.0035	0.0035
	04855	*folD*	methylene-THF dehydrogenase	0.0037	0.0024
	04860	*metV*	methylene-THF reductase	0.0050	0.0033
	04865	*metF*	methylene-THF reductase	0.0000	0.0038

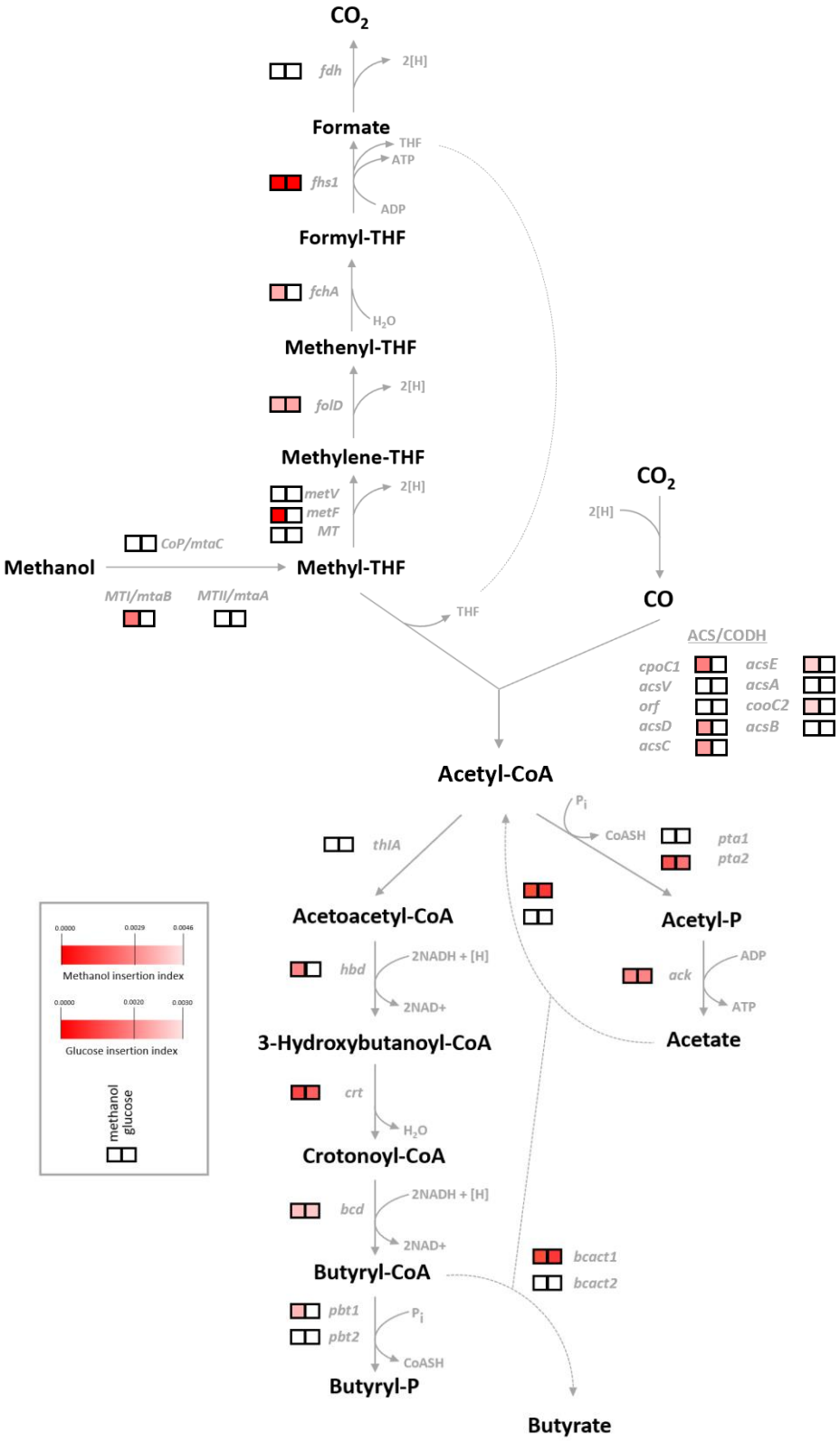


Figure 50 – Schematic of core C1 metabolism within *E. limosum* NG-6849 with associated TraDIS insertion indeces below the ambiguous threshold in both conditions indicated (i.e., 0-0.0046 methanol, 0-0.003 glucose).

5.2.3.4. Integral energy conservation gene clusters

Two ATP synthase gene clusters have previously been identified within *E. limosum* ATCC 8486 (Song, et al., 2017). BLASTn alignment of the genes involved in each cluster with the NG-6849 genome highlighted 100% identity homologues for the c3666-673 V/A-type ATPase cluster 1, and c3453-60 v/A-type ATPase cluster 2. Within the first cluster seven of the eight subunit genes were predicted as essential within the methanol-based library, with three of eight essential and three of eight ambiguous within the glucose-based population. Only one, non-synonymous subunit was predicted as essential for each condition assayed within the V/A-type ATPase cluster 2, and three further subunits predicted as ambiguous solely on methanol (**Table 34**; **Figure 51**).

The respective genes of the Rnf complex cluster (c3879-84) identified within ATCC 8486 were also well conserved within the *E. limosum* NG-6849 genome, all with 100% identities. Of the six subunits only two were predicted as essential on methanol *rsxC* (B2M23_19790), which was comparably predicted as essential for glucose conditions, and r*nfD* (B2M23_19785). The previous insertion index values for r*sxC* show a decline in both conditions, from 0.0132 to 0.0058 on methanol, and from 0.0082 in the previous glucose subculture, indicating a decline in viable unique insertion mutants over time and not a lack of insertion mutants in earlier samples. The same decline is apparent for *rnfD* with an earlier insertion index value on glucose of 0.0238, and earlier subcultures on methanol of 0.0057 (subculture 2) and 0.0249 (subculture 1).

Table 34 – Gene clusters associated with ATP generation and associated ion gradient with associated locus tag, functional annotation, and insertion index values. All methanol insertion index values below 0.0029 predicted as essential (white), and 0.0029-0.0046 ambiguous (light grey). All glucose insertion index values below 0.002 predicted as essential (white), and 0.002 to 0.003 ambiguous (light grey). Those genes which were not predicted as essential are indicated in dark grey.

	ATCC 8486 Locus tag (B2M23_)	Gene	Enzyme	Essentiality	
				Methanol	Glucose
Rnf cluster	19790	*rsxC*	ion-translocating oxidoreductase complex subunit C	0.0016	0.0016
	19785	*rnfD*	ion-translocating oxidoreductase complex subunit D	0.0011	0.0034
	19780	*rnfG*	ion-translocating oxidoreductase complex subunit G	0.0071	0.0176
	19775	*rnfE*	ion-translocating oxidoreductase complex subunit E	0.0089	0.0195
	19770	*rnfA*	ion-translocating oxidoreductase complex subunit A	0.0096	0.0115
	19765	*rnfB*	ion-translocating oxidoreductase complex subunit B	0.0075	0.0138
V/A-type ATPase cluster 1	18720	*vatD*	V-type ATP synthase subunit D	0.0000	0.0035
	18725	*vatB*	V-type ATP synthase subunit B	0.0016	0.0008
	18730	*vatA*	V-type ATP synthase subunit A	0.0037	0.0019
	18735	*vatE*	V-type ATPase subunit E	0.0054	0.0018
	18740	*vatF*	V type ATPase subunit F	0.0000	0.0036
	18745	*vatC1*	ATP synthase subunit C	0.0000	0.0025
	18750	*vatI*	V-type sodium ATPase subunit I	0.0034	0.0045
	18755	*vatC2*	V-type ATPase subunit C (H)	0.0021	0.0021
V/A-type ATPase cluster 2	17645	*vasI*	V-type ATP synthase subunit I	0.0071	0.0124
	17650	*vasK*	V-type ATP synthase subunit K	0.0211	0.0211
	17655	*vasE*	V-type ATP synthase subunit E	0.0000	0.0037
	17660	*vas*	V-type ATP synthase subunit	0.0034	0.0080
	17665	*vasF*	V-type ATP synthase subunit F	0.0035	0.0105
	17670	*vasA*	V-type ATP synthase subunit A	0.0082	0.0131
	17675	*vasB*	V-type ATP synthase subunit B	0.0071	0.0126
	17680	*vasD*	V-type ATP synthase subunit D	0.0052	0.0052

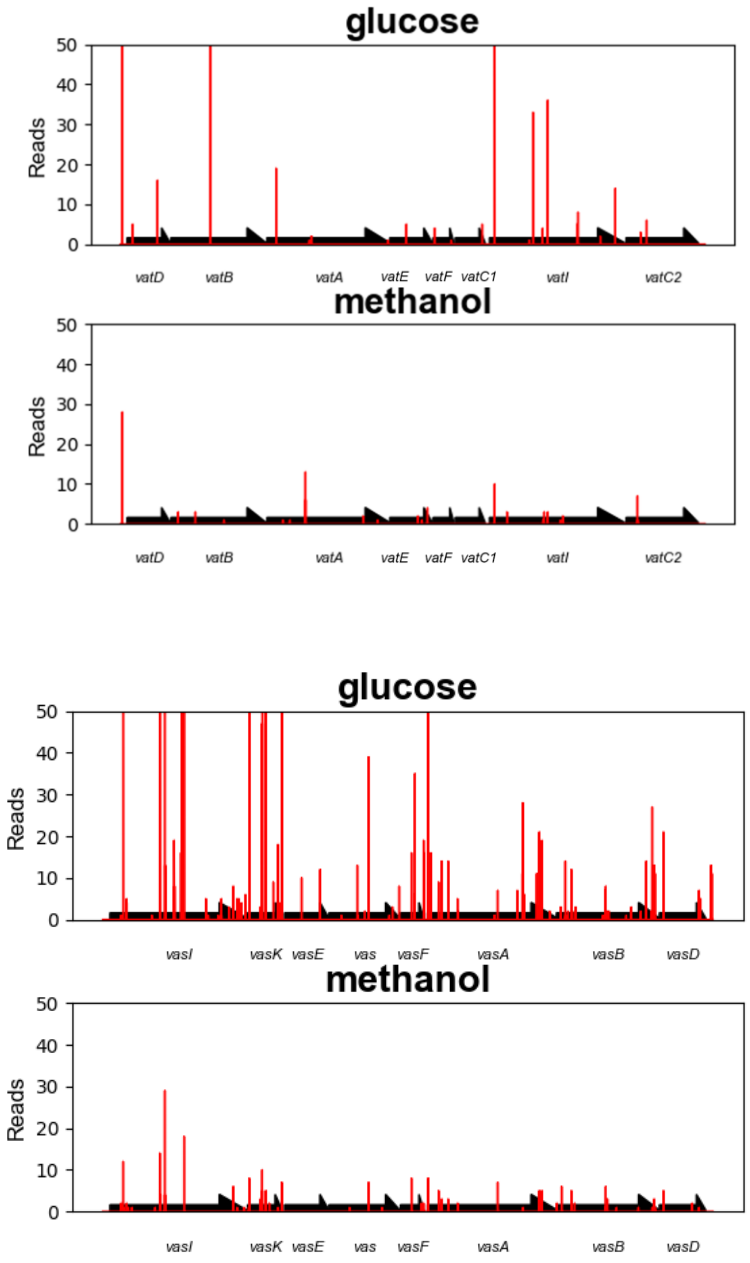


Figure 51 – Insertion plots of candidate V/A-type ATP synthase gene within subculture 3 on methanol and glucose respectively. Specifically, cluster 1 containing *vatD-vatC2* and cluster 2 containing *vasI-vasD* genes. Plots have been capped at 50 reads to ensure good resolution of lower frequency insertions.

Similarly, within the same study two hydrogenase gene clusters and an independently located hydrogenase gene were identified; hydrogenase cluster 1 (c3457-51), hydrogenase cluster 2 (c3127-32) and a single iron hydrogenase (c3789) (Song, et al., 2017). BLASTn alignment of these with the NG-6849 found comparable gene clusters with 100% identity. None of the genes within hydrogenase cluster 1 were predicted essential under any condition, but *hydA1*

(B2M23_11935) was ambiguously essential on methanol (**Table 35**). Similarly, none of the hydrogenase cluster 2 genes were predicted as essential in either condition. The independent gene B2M23_19340 was, however, predicted as essential within the *E. limosum* insertion mutant library grown on methanol and predicted as ambiguously essential on glucose.

Table 35 – Gene clusters associated with electron donor regeneration with associated locus tag, functional annotation, and insertion index values. All methanol insertion index values below 0.0029 predicted as essential (white), and 0.0029-0.0046 ambiguous (light grey). All glucose insertion index values below 0.002 predicted as essential (white), and 0.002 to 0.003 ambiguous (light grey). Those genes which were not predicted as essential are indicated in dark grey.

	ATCC 8486 Locus tag (B2M23_)	Gene	Enzyme	Essentiality	
				Methanol	Glucose
hydrogenase cluster 1	11915	*hydC*	NADH dehydrogenase (Quinone)	0.0088	0.0088
	11920	*hydE*	ATP-binding protein	0.0141	0.0202
	11925	*hydD1*	(2Fe-2S) ferredoxin domain-containing protein	0.0055	0.0083
	11930	*hydB1*	4Fe-4S binding protein/NADH dehydrogenase	0.0055	0.0080
	11935	*hydA1*	2Fe-2S iron-sulfur cluster binding domain-containing protein/hydrogenase	0.0044	0.0114
hydrogenase cluster 2	16020	*c3127*	4Fe-4S ferredoxin	0.0094	0.0142
	16025	*aor*	aldehyde ferredoxin oxidoreductase	0.0111	0.0136
	16030	*moaD*	MoaD family protein	0.0161	0.0120
	16035	*hydC2*	NAD(P)H-dependent oxidoreductase subunit E	0.0249	0.0317
	16040	*hydB2*	4Fe-4S dicluster domain-containing protein	0.0166	0.0191
	16045	*hydA2*	4Fe-4S dicluster domain-containing protein	0.0178	0.0238
hydrogenase	19315	*hydE2*	NAD(P)H-dependent oxidoreductase subunit E	0.0043	0.0213
	19320	*hyd*	4Fe-4S dicluster domain-containing protein	0.0022	0.0029

5.2.3.5. Key product synthesis genes

As the most studied products within *E. limosum* the candidate genes for butyrate, acetate, ethanol, and butanol synthesis were also assessed. As previously outlined, there are two candidate phosphate acetyltransferases,

two candidate phosphate butyryltransferases and two candidate butyryl-CoA:acetate CoA-transferases. Only one phosphate acetyltransferase (B2M23_09570) and one candidate butyryl-CoA:acetate CoA-transferase (B2M23_20590) were predicted as essential for growth in both conditions, indicating these are likely the preferred, if not the only functional route to corresponding product synthesis (**Table 36**). This was corroborated by work by A. Agius (personal communication) who demonstrated the maintained butyrate production of a *Δbcact2 E. limosum* NG-6849 strain. Only one phosphate butyryltransferase (B2M23_06820) was predicted as ambiguously essential in methanol conditions. Also, two alcohol dehydrogenase genes were predicted as essential in both growth conditions: B2M23_12405 and B2M23_15170.

Table 36 – Genes and gene clusters associated with the synthesis of industrially relevant products with associated locus tag, functional annotation, and insertion index values. Specifically, acetate, butyrate, ethanol and butanol synthesis. All methanol insertion index values below 0.0029 predicted as essential (white), and 0.0029-0.0046 ambiguous (light grey). All glucose insertion index values below 0.002 predicted as essential (white), and 0.002 to 0.003 ambiguous (light grey). Those genes which were not predicted as essential are indicated in dark grey.

	ATCC 8486 locus tag (B2M23_)	Gene	Enzyme	Insertion index	
				Methanol	Glucose
butyrate synthesis	01170	*etfA*	electron transfer flavoprotein subunit alpha	0.0010	0.0065
	01175	*etfB*	electron transfer flavoprotein subunit beta	0.0057	0.0057
	01180	*bcd*	acyl/butyryl-CoA dehydrogenase	0.0039	0.0029
	01185	*hbd*	3-hydroxybutyryl-CoA dehydrogenase	0.0027	0.004
	01190	*crt*	crotonase	0.0014	0.0014
	01195	*thlA*	thiolase	0.0056	0.0066
	06820	*pbt2*	phosphate butyryltransferase	0.0037	0.0061
	11875	*pbt1*	phosphate butyryltransferase	0.0194	0.0109
acetate synthesis	14425	*ack*	acetate kinase	0.0028	0.0019
	06165	*pta1*	phosphate propanoyltransferase/acetyltransferase	0.0111	0.0129
	09570	*pta2*	phosphate propanoyltransferase/acetyltransferase	0.0017	0.0017
Both	20590	*bcact1*	butyryl-CoA:acetate CoA-transferase	0.0016	0.0008

	07595	*bcact2*	3-oxoacid CoA-transferase/butyryl-CoA:acetate CoA-transferase	0.0242	0.0271
alcohol synthesis	16025	*aor1*	aldehyde ferredoxin oxidoreductase	0.0111	0.0136
	12405	*adh1*	alcohol dehydrogenase	0.0015	0.0005
	15170	*adh2*	alcohol dehydrogenase	0.0000	0.0000
	20225	*adh3*	hypothetical protein	0.0179	0.0279
	02350	*adh4*	butanol dehdrogenase	0.0182	0.0192
	09455	*adh6*	Zn-dependent alcohol dehydrogenase	0.0106	0.0138
	04655	*aor2*	aldehyde:ferredoxin oxidoreductase	0.0188	0.0111

5.2.3.6. Final methanol subculture composition

To identify any overrepresented mutants within the final subculture on methanol preliminary assessment was performed by ranking the read counts. This identified a few significantly overrepresented genes with highly elevated read counts in the final two subcultures. This included a trimethylamine methyltransferase (B2M23_02925), a hypothetical gene of 345 bp (B2M23_08625), two genes of unassigned function of 2361 bp (B2M23_14320) and 735 bp (B2M23_09980) and a restriction endonuclease subunit (B2M23_05065). Collectively in the third subculture these constituted 51.58% of the total insertion reads successfully mapped, further increasing to 79.59% by the final subculture and reducing the mutant pool diversity such that resolution of essential and non-essential gene peaks and subsequent prediction could no longer be performed (**Figure 52**).

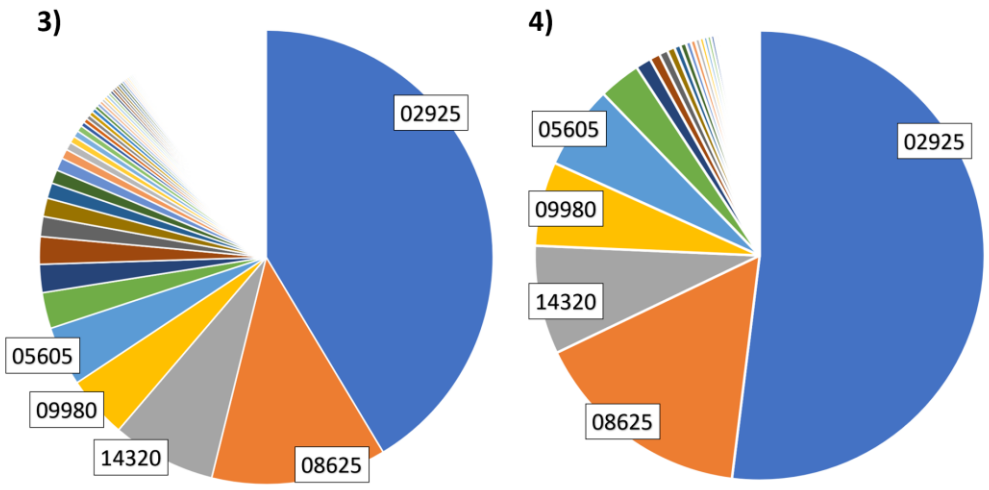


Figure 52 – Graphical representation of *E. limosum* NG-6849 mutant pool diversity in the third and fourth subculture on methanol, with each chart reflective of total reads with individual wedges indicating those genes with >0.2% occurrence within the dataset.

While this could be due to an early transposition event saturating the sample with insertions at these positions, examination of the mutant pool in the earliest samples (methanol subculture 1) negated this assumption, with much lower read counts observed for each gene, corresponding to a mutant abundance within the total reads mapped of 0.23%, 0.03%, 1.19%, 0.01% and 0.04%, respectively. To assess the likelihood of the apparent *mttb* (B2M23_02925) mutant abundance being the result of a PCR artifact, the abundance of mutants in each of the five genes was assessed across the individual replicates of methanol subculture 3 (**Figure 53**). The mutant abundance was determined by the read counts for each gene locus relative to the total mapped reads of each replicate. This highlighted an abundance of B2M23_02925 mutants within replicates A and C reaching 65.84% and 94.43% respectively (**Table 37**). To better quantify the increase in abundance of these mutants the third and fourth subcultures were directly compared, and insertion reads/total reads normalised according to gene length. This highlighted the over representation of B2M23_02925 in two of three replicates, and a mixture of genes overrepresented in replicate B.

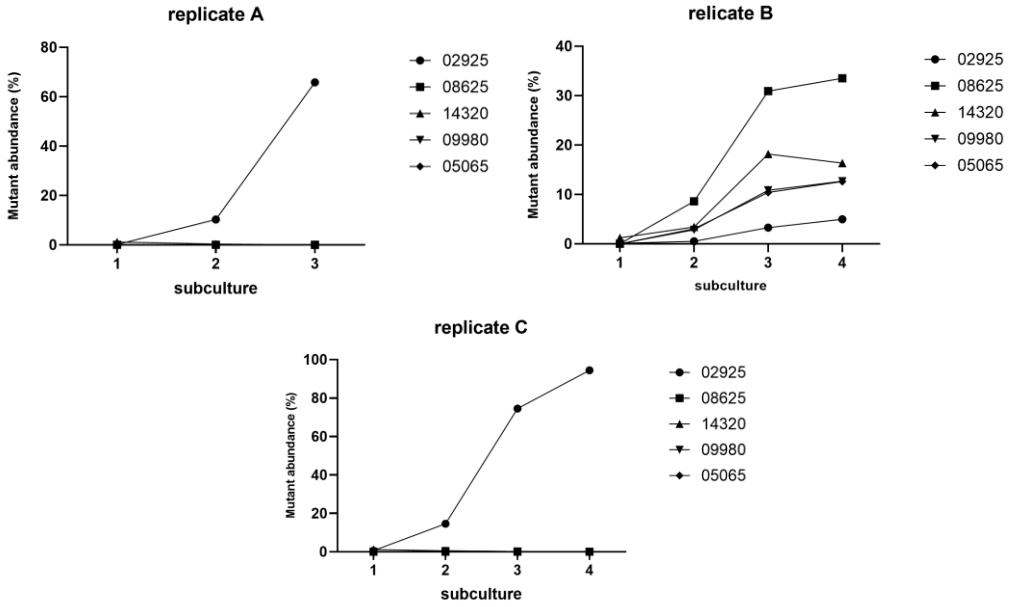


Figure 53 – Mutant abundance (%) of five key genes within individual replicates of the *E. limosum* NG-6849 TraDIS mutant pool grown on methanol

Mutant abundance calculated by mapped read count for the respective gene to total mapped read count per replicate. Analysed gene locus tags of B2M23_02925, B2M23_08625, B2M23_14320, B2M23_09980 and B2M23_05065.

Table 37 – Insertional mutants with the greatest increase in prevalence between the third and fourth subcultures on methanol. Unique insertion sites indicative of independent integration events into the genome.

ATCC 8486 locus tag (B2M23_)	Percentage increase %	Normalised percentage increase	Gene length	Unique insertion sites	Annotated function
08625	5.23	15.16	345	14	hypothetical protein
02925	16.61	11.78	1410	63	trimethylamine methyltransferase
05065	2.33	3.81	612	19	restriction endonuclease subunit S
09980	2.22	3.02	735	17	hypothetical protein
02875	1.25	1.40	894	19	ABC transporter permease
14320	1.62	0.69	2361	70	hypothetical protein
19560	0.26	0.39	660	9	DNA-binding response regulator
10505	0.16	0.36	444	17	hypothetical protein
07440	0.16	0.18	873	28	hypothetical protein

5.3. Discussion

5.3.1. Genome

While attempts were made within this study to acquire a closed genome for *E. limosum* NG-6849, they were unfortunately not successful within the timeframe. Although attempts to map the Illumina data derived from *E. limosum* NG-6849 sequencing to this published genome (Song & Cho, 2015) were highly effective, with 99.83% of genes successfully mapped, this was not deemed a suitable closure approach due to previous observation of significant structural rearrangements apparent between *E. limosum* species (Mansfield, 2018). As such, initial research progressed with several contigs as primary genomic reference. The Bio-TraDIS pipeline analysis relies on a closed and annotated genome which permits the elucidation of gene specific insertion indices. As such the analysis of TraDIS data was pursued with the published ATCC 8486 genome.

Following receipt of a closed draft *E. limosum* NG-6849 genome, this was used as a template for the mapping of the previously acquired NG-6849 genome Illumina data. The mapping statistics reveal a comparable efficacy in this mapping to the previous attempts made with the ATCC 8486 genome as a reference (**Table 29**). This Illumina corrected NG-6849 consensus genome was then compared to the type-strain genome using progressiveMauve (Darling, et al., 2010). This highlighted high conservation between the two strains, with four homologous locally colinear blocks of 1,286,651 bp, 2,374,389 bp, 73,242 bp and 11,589 bp, respectively. These long homologous regions effectively demonstrate 100% identity in 84.37% of the genome. The presence of these four locally colinear blocks also appears to indicate at least two significant genomic rearrangements have occurred since the strains diverged from a common ancestor.

5.3.2. Genetic tools

Within the course of this study both the antibiotic sensitivity of *E. limosum* NG-6849 and the compatibility of various Gram-positive replicons from the pMTL80000 series were assessed (Heap, et al., 2009). Similar assessment has previously been performed within *E. limosum* ATCC 8486 by Shin et al. (2019). To determine the minimal inhibitory concentrations (MICs) of each antibiotic in ATCC 8486, Shin and co-workers used a plate-based approach in which CFUs were compared at a range of antibiotic concentrations (0-20 μg/mL of thiamphenicol and 0-4 μg/mL of clarithromycin) (Shin, et al., 2019). This was thought to occur due to *E. limosum* catalysing the reduction of the nitro group of chloramphenicol, effectively inactivating the antibiotic. Such reduction has previously been observed in many clostridial species and hence use of chloramphenicol was not used with NG-6849 (O'Brien & Morris, 1971).

Sufficient growth inhibition of *E. limosum* ATCC 8486 was observed at 15 μg/mL thiamphenicol and 1 μg/mL clarithromycin. This conclusion regarding the MIC of clarithromycin for *E. limosum* is identical to that determined by the broth-based inhibition approach utilised within this study. Thiamphenicol sensitivity, however, was observed at a lower threshold for *E. limosum* NG-6849, leading to the conclusion of a 2.5 μg/mL MIC as opposed to 10 μg/mL. This could be attributed to variation in the assay approach, with a differing exposure method and more extended growth period assayed by Shin et al. (2019). A similar discrepancy has previously also been noted by R. Mansfield (2018), who observed a 2.5 μg/mL thiamphenicol MIC for *E. limosum* ATCC 8486 at three days of growth, which increased to 5 μg/mL following six days of extended exposure.

Discrepancies between NG-6849 and ATCC 8486 were also observed within the Gram-positive replicon assessment, wherein transfer of pMTL8*151 vectors was attempted with five replicons consistent in both the work of Shin et al. (2019), and this study: pBP1 (2), pCB102 (3), pCD6 (4), pIM13 (5), and pIP404 (6). This study also reports the assessment of a further new modular replicon,

p19 (8), derived from *C. carboxidivorans* P7 (Millard, 2017). Extremely low transformation efficiencies were observed for nearly all vectors in *E. limosum* ATCC 8486 by Shin *et al*, with a range of only 0-1.8 transformants observed per µg of DNA for most replicons. The highest transfer efficiency reported was for pIP404, which generated an average transformant CFU of 104.3 per µg of DNA. In contrast, within this study a high transformation frequency was observed for all but one replicon, with most vectors yielding transformant CFUs exceeding 6.8×10^2 per µg of DNA, and four of the five achieving more than 1.0×10^5. This significant discrepancy between NG-6849 and *E. limosum* ATCC is likely attributable to differences in the competent cell generation and transformation protocol.

The lack of transformant colonies generated by transfer of pMTL82151 within this study was thought to be indicative of pBP1 lacking functionality within *E. limosum* NG-6849. However, the successful transformation of this vector into the highly similar *E. limosum* ATCC 8486 strain, alongside lack of pIM13-based plasmid transfer, calls this into question. This variance could be due strain divergence but following preliminary genomic analysis this appears unlikely. Instead, this discrepancy may be attributable to the low transfer frequency observed by Shin et al. (2019) somewhat reducing the dynamic range and corresponding sensitivity of the assay within *E. limosum* ATCC 8486. R. Mansfield (2018), using a near identical competent cell protocol to that employed within this study, reported only three of the eight replicons tested as demonstrating compatibility with *E. limosum* ATCC 8486; pCB102, pIP404 and p19.

5.3.3. TraDIS coverage and gene essentiality prediction

Thorough TIS studies have only been published for two clostridial species to date; *C. difficile* and *C. autoethanogenum* (Dembek, et al., 2015; Woods, et al., 2021). While transposon mutagenesis was performed within *C. perfringens,* this study did not apply NGS technologies to quantify the total mutant pool,

instead initially screening for motility mutants and using Sanger sequencing to identify the insertional loci of only 24 selected mutants (Liu, et al., 2013). Within the *C. difficile* and *C. autoethanogenum,* however, a high throughput NGS approach was applied for quantification of the diversity of the mutant pool and corresponding essentiality predictions.

Within *C. difficile* a reported 77,256 unique insertional mutants were sequenced following preliminary growth, approximately equivalent to one insertion every 54 bp within the genome. This was even greater in *C. autoethanogenum* reaching 100,065 unique insertional mutants, broadly equivalent to one insertion every 43 bp. In both instances this sequencing was performed following growth in rich media in order to determine the core essential gene profile of each strain. Further to this, the study in *C. autoethanogenum* also grew the mutant pool in autotrophic conditions on CO_2 and H_2 to provide a comparative dataset of insertional mutants. This reduced the total unique insertion sites (UIS) recorded to 66,254 broadly equivalent to one insertion every 65 bp, with a corresponding increase from the original 439 core genes predicted as essential for growth in enriched conditions to 758 genes predicted as essential for autotrophic growth. This constitutes approximately 11% and 19% of the total 4059 genes annotated within the *C. autoethanogenum* genome, respectively (Humphreys, et al., 2015). Only the initial core gene essentiality prediction was performed for *C. difficile,* highlighting 404 confident essential assignments and 33 ambiguous.

Within this study similar essentiality profiling was performed within *E. limosum* NG-6849. Across the earliest most saturated mutant pool samples (MS1) 179,161 UIS were mapped, with a range of 112,010-135,470 within each individual replicate. This is comparable to that published to date, with near equivalent total UIS at the individual replicate level generated by both the pMTL-CW20 and derivative pMTL-CW21 vectors when applied in *C. autoethanogenum* and *E. limosum*. The elevated saturation provided by the combination of the biological repeats led to an approximate insertional frequency throughout the genome across all samples of one every 15 bp

(288,171 UIS). For prediction of the minimal gene set the final subculture on glucose was used (GS3), in which a total of 90,407 UIS were mapped with a corresponding insertional frequency of approximately one every 49 bp. This insertional frequency is within the range of those previously used to predict core gene essentiality in the two clostridial strains, 43-54 bp. From this read mapping 547 genes were predicted as essential, and 74 ambiguously predicted, cumulatively equivalent to 12.47% of the annotated genes. For growth on methanol the third subculture was also used (MS3) in which 73,968 unique insertional mutants were identified, broadly equivalent to one insertion every 60 bp. This is comparable to the mutant diversity observed in autotrophic growth of *C. autoethanogenum* at an approximate frequency of one insertion every 65 bp. From this sequencing 607 genes were predicted as essential for growth of *E. limosum* on methanol, with a further 190 genes ambiguously essential. Cumulatively this represents 16.00% of the total annotated genes within the ATCC 8486 genome used as reference (Song & Cho, 2015)

This core essential gene set within *E. limosum* NG-6849, totalling 621 genes, is higher than both the core gene set predictions for *C. difficile* and *C. autoethanogenum.* This significant increase is due to the divergence in culturing conditions: for *E. limosum* NG-6849 growth a fully defined medium was used which supplied no amino acids and only biotin, pantothenic acid, and lipoic acid as vitamins. In contrast, *C. autoethanogenum* and *C. difficile* were both grown in medium enriched with both yeast extract and tryptone (YTF and TY broth, respectively). These are undefined components providing the organism with an external source of amino acids, peptides, water soluble vitamins and carbohydrates. As such the approximately 180 additional genes implicated within *E. limosum* NG-6849 for growth on glucose in Pacaud media are likely to correspond to the biosynthetic pathways for these essential components, such as B2M23_13930 encoding a threonine synthase and B2M23_15475 encoding tryptophan synthase, both of which were predicted as essential within this assay. This defined media provides a more complete picture of the organism in an industrially relevant context, wherein defined and

minimal media is preferred due to concerns including cumulative costs and contamination.

The remaining genes predicted as essential in both conditions are also involved in fundamental cellular processes such as DNA replication, protein synthesis and cell division. These include the tRNA ligase genes for the twenty key amino acids, of which 19 are clearly annotated within the genome. For phenylalanine two subunit genes are annotated, B2M23_17785 and B2M23_17790, both predicted as essential for each condition indicating their integral nature and lack of functional redundancy. B2M23_04525 encodes a methionine-tRNA ligase which was not predicted as essential, indicating it likely a non-active duplicate with the B2M23_03035 encoded methionine-tRNA ligase the functional version employed by the cell. While a tRNA ligase specific to asparagine was not present within the genome this is not uncommon in prokaryotes, with multiple bacteria relying on a non-discriminatory aspartic-acid tRNA-ligase and subsequent amidotransferase activity. Three amidotransferase subunit genes were identified within the genome, B2M23_12425/30/35, all of which were predicted as essential for *E. limosum* NG-6849 growth in both conditions, indicating this is the likely route within this strain. This pathway may also be employed within *C. autoethanogenum,* in which neither of the two asparagine synthase genes (CLAU_2687, CLAU_1984) were predicted as essential (Woods, et al., 2021). Alternatively, this could indicate functional redundancy, in which instance both insertions would remain viable within the sequenced mutant pool. This is an unfortunate shortcoming of the single insertional characterisation process.

Due to issues encountered in the growth of *E. limosum* NG-6849 on methanol in the presence of thiamphenicol, this selection was absent in the methanol growth conditions. Limitations in the use of thiamphenicol with *E. limosum,* despite the presence of an appropriately expressed *catP* marker, has also been noted in other studies within the wild-type ATCC 8486 strain. For example, when developing the CRISPRi system for knock-down of uninduced cells, autotrophic characterisation efforts were curtailed when all vectors showed

significant growth defects (OD_{600} < 0.1) under CO_2:H_2 conditions even when CRISPR-Cas9 activity was repressed (Shin, et al., 2019). As such this appears as an inherent feature of *E. limosum* when growth conditions dictate the use of the WLP.

5.3.4. Essential genes for methanol metabolism

The *mtaB/MTI* gene, encoding a methanol-dependent methyltransferase, was confirmed as integral to the growth of *E. limosum* NG-6849 on methanol through TraDIS predicted gene essentiality, and not integral when the strain was comparably grown on glucose (**Table 32**). The lack of essentiality predictions for the other key genes hypothesised as forming the methanol metabolising MtaABC complex (the corrinoid protein gene; *CoP/mtaC/*B2M23_16855, and methyltransferase gene; *MTII/MtaA*/B2M23_16860) was particularly perplexing. To further complicate interpretation, the *cpoC1* gene was essential solely on methanol, appearing to indicate the integral nature of corrinoid protein regeneration in methanol metabolism.

Alongside *MTI* however, a single methylene-THF reductase candidate gene, *metF* (B2M23_04865), was predicted as essential solely within the TraDIS library grown on methanol. This supports what was previously hypothesised regarding the dual flux of methanol within the WLP including reverse methyl branch oxidation reactions which generate both reducing equivalents and ATP (**Figure 50Error! Reference source not found.**). The lack of predicted essentiality of the adjacent *metV* gene is not entirely surprising; *metV* (methylene-THF reductase) genes are noted as conserved dispensable genes in acetogenic bacteria, in part due to their enzymatic diversity (Song, et al., 2018). For example, *A. woodii* is reported as containing a NADH-oxidising trimeric enzyme-complex, but *C. ljungdahlii* a heterodimeric reduced ferredoxin dependent enzyme complex (Bertsch, et al., 2015; Yi, et al., 2021).

This reverse methyl flux was similarly confirmed with *A. woodii* through transcriptomic and enzymatic characterisation of the methylene-THF reductase (Kremp, et al., 2018). Interestingly, to stimulate methylene-THF reductase catalysation of the reverse methanol oxidation reaction, and demonstrate associated NAD+ reduction, additional cytoplasm had to be added to the assay conditions. This led to the conclusion that due to the highly endergonic nature of this reaction the subsequent metabolites/reactions were necessary to "pull" this essential initial methylene-THF reaction. It was proposed in studying the energy conservation of *E. callanderi* KIST 612, that the methylene-THF reductase itself may be bifurcating, providing reduced ferredoxin for the FDH reaction (Jeong, et al., 2015). The expression of the corresponding electron-transferring flavoproteins (B2M23_01170-75) within *E. limosum* ATCC 8486 was also found to be upregulated in autotrophic conditions (Song, et al., 2018). Within this study only *etfA* was predicted as essential for growth on methanol. This appears to indicate that although this is a prospective pathway in *E. limosum* NG-6849, it is not essential to redox cofactor balancing with considerable functional redundancy potentially present in the form of multiple hydrogenases (**Table 36**).

M. thermoacetica and *A. woodii* both contain a high number of methyltransferase genes, a common acetogenic trait thought to be a characteristic feature of a broad substrate profile (Pierce, et al., 2008; Kremp, et al., 2018). *A. woodii* is estimated to have 30 methyltransferase encoding genes, with transcriptomic analysis during growth on methanol highlighting three specific methyltransferase gene clusters which were significantly upregulated. This included Awo_c22760-40, the proposed *mtaABC* cluster, alongside Awo_c7520-60 and Awo_c23710-70. While the Awo_7520-60 cluster was concluded to be a glycine betaine metabolism pathway, upregulated due to the presence of this substrate within the undefined yeast extract within the medium, the second could not be attributed to a specific substrate. While the precise number of methyltransferase genes within *M. thermoacetica* has not been quantified, an abundance was noted of genes with high homology to

uroporphyrinogen decarboxylase sequences during genomic analysis (Pierce, et al., 2008). These included the *mtaABC* gene cluster (Moth_1208-Moth_1209, Moth_2346) and the *mtvABC* gene cluster (Moth_1316-1318 and Moth_385- Moth_387) implicated in the catabolism of methoxylated aromatic compounds. This methyltransferase abundance led to the further proposal of Moth_1446, Moth_2100 and Moth_2102 as potential MtaA analogous enzymes involved in the transfer of methyl groups to tetrahydrofolate.

Within the *E limosum* ATCC 8486 genome the number of methyltransferase genes has also not been quantified, however, 42 *mttb* genes have been reported as encoding putative MttB-family trimethylamine methyltransferase proteins (Picking, et al., 2019). These methyltransferases facilitate growth on various methyl group containing compounds, and three *mttb* genes have been characterised thus far which enable the demethylation of proline betaine (Picking, et al., 2019), carnitine (Kountz, et al., 2020) and γ-butyrobetaine (Ellenborgen, et al., 2021), respectively. Despite these distinct substrate specificities, all three characterised Mttb have been shown to methylate the same corrinoid protein encoded by *mtqC*, with subsequent methyl group transfer to the WLP also via the same *mtqA* encoded methyltransferase. These both have 100% identity homologs within NG-6849, with the previously analysed *MTII* and an alternate corrinoid protein encoded by B2M23_05770-5. With insertion indexes within the methanol growth conditions of 0.0139 and 0.333, respectively, neither of these were predicted as essential. This lack of confident *mtaA/MTII* and *mtaC/CoP* gene essentiality within the TraDIS data may indicate the necessity for further genomic analysis to determine alternate candidates, such as B2M23_01090. Alternatively, the challenge encountered in this endeavour thus far may indicate functional redundancy occurs within these WLP integrating enzymes, with the capacity for crosstalk between different methyltransferase substrate utilisation systems

5.3.5. Core WLP gene essentiality and energetics

Four of the six WLP genes within the methyl cluster were predicted, or ambiguously predicted, as essential during growth on methanol. As methanol flux in acetogens has been proposed to proceed in part via a reverse methyl branch, this high number of genes provides further corroboration. Of the three genes in the formate dehydrogenase cluster (B2M23_12645-12655), however, none were predicted as essential for either condition (**Table 33**). This is likely due to the growth conditions, in which the headspace contained an excess of CO_2 and hence the carbonyl branch was not reliant on CO_2 generation via FDH.

On the other hand, only the *fhs1* and *folD* genes of the methyl branch were predicted as essential within the glucose conditions. This is in direct agreement with previous transcriptomic and translatomic analysis reported for *E. limosum* ATCC 8486 comparing autotrophic (CO_2:H_2) and heterotrophic growth (glucose) (Song, et al., 2018). Within this study the transcript levels of *fhs*1 (ELIM_c0957) and *folD* (ELIM_c0959) were relatively high in both growth conditions, and therefore did not show a significant differential expression between the two conditions (DESeq $P > 9.93 \times 10^{-2}$). This appears to indicate that *folD* and *fhs1* are expressed at a reasonably high level on glucose, so the enzymes are relatively abundant and not upregulated when they become necessary for the methyl branch of the WLP.

Within autotrophic characterisation of a *C. autoethanogenum* transposon insertional mutant pool only the *acsA* subunit of the CODH cluster was predicted as essential (Woods, et al., 2021). This was corroborated by another study in which individual ACS/CODH subunit genes were disrupted using ClosTron; for all subunits, excluding *acsA,* this disruption did not prevent autotrophic growth in the resultant mutant (Brown, et al., 2014). Interestingly, pan-genomic analysis of the core WLP genes within autotrophic bacteria identified *C. autoethanogenum* as one of only three species (of fifteen analysed) with below average conservation of this subunit (Shin, et al., 2016). The most conserved ACS/CODH genes identified within this analysis were *acsA,*

acsV, and *acsC.* Within the TraDIS analysis of the ACS/CODH gene cluster of *E. limosum* NG-6849 *acsA* (B2M23_08400) was not predicted as essential, but the genes *cpoC1* (B2M23_08370), *acsD* (B2M23_08385), *acsC* (B2M23_08390), *acsE* (B2M23_08395), *acsB* (B2M23_08410) and *cooC2* (B2M23_08405) were ambiguously predicted (**Table 33**; **Figure 50**). It is highly unlikely that the CODH activity is not essential under methanol conditions, given that the WLP is essential for methanol fixation.

cpoC1 and *cooC2* were originally annotated as encoding AAA family ATPase, with further manual analysis assigning functions as a putative CODH nickel-insertion accessory protein and corrinoid activation/regeneration protein (Song, et al., 2018). Hence these may be involved in assembling and regenerating the corrinoid protein involved in methyl group transfer from methanol to THF. This study did not, however, note any upregulation of *cpoC1* in transcriptomic and translatomic analysis of *E. limosum* ATCC 8486 during autotrophic growth. The core subunits, *acsA, acsB,* and *acsD,* however, all had notable upregulation of expression. *acsE* is annotated as a dihydropteroate synthase and thought to transfer the methyl group of the methyl-THF to the *acsCD* subunits of CODH/ACS. With all three subunits implicated in the methanol-based growth of *E. limosum* NG-6849 growth it is likely that this reaction is well conserved, with the remaining CODH subunits potentially exhibiting functional redundancy, allowing the proliferation of single insertional mutants of each independently.

Energy conservation during methanol metabolism is more complicated to disentangle within the dataset for several reasons including the unconfirmed/varied electron donor essentiality of specific WLP reactions, the potential functional redundancy of the hydrogenases and the two identified V/A-type ATPase gene clusters (**Table 34**). Within the Rnf gene cluster only r*sxC* and *rnfD* were essential on methanol. The C subunit is the most conserved Rnf complex subunit across acetogenic bacterial species (Shin, et al., 2016) and recently confirmed, through total *A. woodii* Rnf complex isolation, to be a NADH-binding and catalysing subunit (Kuhns, et al., 2020). During this analysis

RnfD was also confirmed as flavin-binding, and hence presumed to be integral to the unique electron bifurcation process within acetogens. This unfortunately could not be further confirmed within the study due to the lack of complete enzymatic catalysis being demonstrated by the isolated complex.

During autotrophic growth on CO_2 and H_2, transcriptomic and translatomic profiling of *E. limosum* ATCC 8486 highlighted the significant upregulation of only one V/A-type ATPase cluster: c3452-3469 (Song, et al., 2018). Intriguingly within the comparable cluster of *E. limosum* NG-6849 (B2M23_17640-17680) during growth on methanol only four of these nine genes are predicted (or ambiguously predicted) as essential. On the other hand, within gene cluster 1 all but one of the eight genes (B2M23_18720-18755) were confidently predicted as essential for methanol growth (**Table 34**). These divergent results appear to indicate that both ATPase systems within *E. limosum* are functional, and despite its lack of upregulation in autotrophic growth conditions V/A-type ATPase 2 remains integral to energy conservation when grown on methanol.

Also, within the transcriptomic and translatomic analysis of *E. limosum* ATCC 8486 a hydrogenase cluster, comparable to the electron bifurcating *A. woodii* gene cluster *hydCEDBA*, was identified and found to be upregulated during autotrophic growth. The corresponding hydrogenase cluster of *E. limosum* NG-6849 (B2M23_11915-11935) had no genes confidently predicted as essential, with only *hydA1* (B2M23_11935) ambiguous in methanol growth (**Table 35**). *hydA1* was identified as coding for one of the two most conserved subunits within this hydrogenase cluster during previous pan-genomic analysis (Shin, et al., 2016). No genes were predicted as essential within the second hydrogenase cluster (c3127-3132) as well, despite pan-genomic analysis previously highlighting *hydB2* (B2M23_16040) as highly conserved across acetogenic species. This is to be expected given the lack of hydrogen provided in the culturing conditions, leading to the necessity for electron donors to be generated directly from methanol via the reverse WLP pathway steps. Only the independent hydrogenase gene c3789 (B2M23_19320) was predicted as essential for both methanol and glucose growth.

5.3.6. Mutant overrepresentation

With replicative plasmid-based transposon mutagenesis systems, typically plasmid-based sequencing reads can be an issue, reducing the diversity of the library through read saturation. In fact, when applying TIS in *C. difficile* grown in rich medium, 48% of the resultant reads were mapped to the transposon-delivery plasmid (Dembek, et al., 2015; Woods, et al., 2021). While this was not a hurdle within this study due to the pMTL-CW20 based vector design (Woods, et al., 2021) a significant decline in library diversity was observed. This similarly occurred due to overrepresented reads, in this instance from a small cohort of genes. These genes had elevated read counts, with insertional mutants within each gene significantly more prevalent within the sequenced population of later *E. limosum* NG-6849 growth stages. This correlation with the time course of the experiment, and the occurrence of the B2M23_02925 mutant bias in two of three replicates confirmed this is not likely to be solely the result of a PCR acquired artifact or bias. The resultant decline in diversity prevented originally planned gene essentiality prediction within the fourth subculture of methanol growth.

Although this mutant overrepresentation was originally noted within methanol sequencing analysis due to the statistical implications preventing essentiality prediction, comparative analysis of the glucose conditions also highlighted a rapid decline in diversity between the second and third subcultures (

Table *38*). Within the second subculture the gene with the highest number of transposon insertional reads represented only 2.8% of the total sequenced mutant population. In comparison, within the third subculture the most occurrent mutant represented 25.40% of the total reads. These genes were also completely distinct, with the top fourteen most abundant mutants within the second subculture on glucose declining in read counts by the third subculture.

Table 38 – The eight genes with highest increase in mutant abundance, as determined by percentage of the total read count, between the second and third subcultures of *E. limosum* NG-6849 glucose growth.

Locus Tag (B2M23_)	Gene length	Function	Glucose subculture mutant abundance (%)	
			2nd	3rd
07305	777	hypothetical protein	0.0054	25.3956
18130	1275	hypothetical protein	0.0198	20.6829
17470	1566	sodium:proton antiporter	0.0198	19.5071
00625	2316	bifunctional glutamate–cysteine ligase/glutathione synthetase	0.3869	8.4431
13265	2427	hypothetical protein	2.8674	1.1819
11395	261	hypothetical protein	0.4964	0.5037
12945	8634	hypothetical protein	1.5914	0.4152

Essential gene prediction relies on the assumption that mutants exhibiting a disadvantageous or unviable phenotype within a given condition cannot survive, and hence decline in prevalence and become underrepresented in the mutant population. In which instance the increase in reads within some genes in relation to the total mutant pool may be indicative of those which proliferate more rapidly within the given conditions, out competing those mutants which, while still viable, have somewhat impaired growth rates. This is somewhat analogous to clonal interference observed during adaptive laboratory evolution, but with a much greater number of mutants.

During this analysis, it was noted that several genes had zero reads mapped to them in subculture three, but an increase to a positive integer in the fourth subculture. The lack of residual transposon mutagenesis vector, demonstrated by the lack of Cla^R colonies identified during plating (within 100 µl of each culture), completely prevents new transposition events occurring in these later subcultures. With new mutant generation hence impossible, this threw into question the robustness of the dataset, and the extent to which the sequencing run provided complete coverage of the amplified mutant pool. To assess this rarefaction analysis was performed (**2.13.3**). Within rarefaction analysis the total reads of the sequencing run were combined, before being artificially split into four subsets, each mapped individually and sequentially. The new UIS identified by the mapping of each new set of reads was then plotted against

the total reads mapped (**Figure 54**). If the level of sequencing provided sufficient coverage for the library diversity, we would expect diminishing returns with less new unique insertion sites identified as more reads are mapped. If the sequencing was insufficient, however, a more linear relationship will be observed between reads mapped and UIS identified.

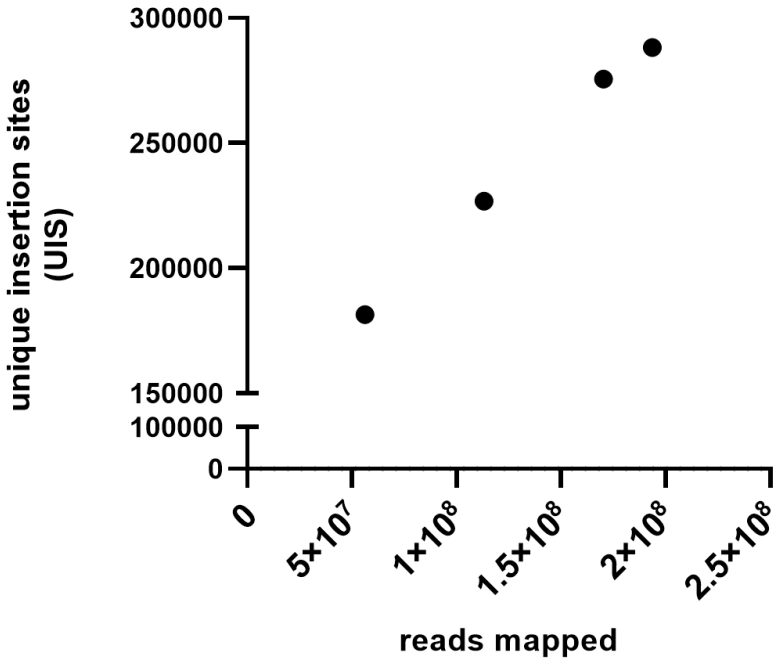


Figure 54 – Rarefaction analysis of TraDIS dataset
Total unique insertion sites mapped for four subsets of the total TraDIS data representing 6 x 10^7, 12 x 10^7, 18 x 10^7 and 20.46 x10^7 of the received reads each, of which 93.74%, 93.29%, 94.60% and 94.59% were successfully mapped, respectively.

The equation of the graph (y=0.0008x+137289) was used to perform a preliminary prediction of the further UIS attainable with a second HiSeq run. This predicts a potential further 158,683 UIS could be identified. This is a large extrapolation of the data, and hence the confidence level is not high. The most cogent conclusion from this analysis is that the biological library has greater diversity than the sequencing has identified. This is the likely explanation for the apparently inexplicable appearance and disappearance of certain mutants when tracking them throughout the time course of the experiment.

5.4. Chapter summary

- Confirmed the functional compatibility of *ermB, catP*, pCB102, pCD6, pIM13, pIP404 and p19 for pMLT8000 clostridial shuttle vector implementation within *E. limosum* NG-6849
- Demonstrated the pRECas CRISPR-Cas9 system within *E. limosum* NG-6849 through a proof-of-concept knockout of the *R.ORFD* and *gmrSD* genes
- Implemented the transposon mutagenesis vector pMTL-CW21 to generate 179,161 unique insertional mutants
- Sequenced triplicate transposon libraries to generate the 797 essential gene profile of *E. limosum* NG-6849 for growth methanol
- Identified 103 genes implicated in growth of *E. limosum* NG-6849 on methanol, but not within comparable growth on glucose
- Confirmed the integral nature of *MTI* (B2M23_16850) and *metF* (B2M23_04865) for methanol utilisation
- Identified five genes with elevated mutant abundance in later stages of *E. limosum* NG-6849 growth on methanol, collectively constituting 79.59% of the total mapped reads in subculture 4

6 Discussion

This study describes the development and application of synthetic biology tools in the C6 and C4 producing acetogens *C. carboxidivorans* and *E. limosum*. These are both promising bacterial chassis for the valorisation of syngas, subject to further strain development and process optimisation. While the application and success of the bacterial catalysts within industry is dictated by an expansive number of factors, the acquisition of synthetic biology tools increases the potential routes to process optimisation. As carbon dioxide and other greenhouse gases attract increasing tariffs, and biotechnology is pursued as a route to industrial decarbonisation such strain development is particularly prescient. Both *E. limosum* and *C. carboxidivorans* have aligning strengths which make them promising candidates for such efforts, i.e., the use of the reductive acetyl-CoA pathway to efficiently fix C1 feedstocks, native production of high value targets, and expanding strain specific knowledge bases. They differ, however in their key challenges; *C. carboxidivorans* in its recalcitrance to DNA transfer, and *E. limosum* NG-6849 in its lack of characterisation.

6.1. *Clostridium carboxidivorans*

6.1.1. Original contributions

Efficient conjugal DNA transfer into *C. carboxidivorans* via recombinant *E. coli* donors was demonstrated for the first time within this study. Preliminary characterisation of the barrier posed by the ten identified RM systems was performed using PacBio® data, kindly provided by Professor Yang Gu (SIBS, CAS, China), and analysed both by the REBASE® pipeline and manually. While this allowed the confident allocation of four (of six) recurrent methylation motifs identified within *C. carboxidivorans* to corresponding RM systems, one further was confidently allocated within this study through nanopore sequence analysis; AGAAGC to the Type II system encoded by *CcaP7ORF17830P.* Different strategies were taken to mitigate the impact of these RM systems on

the frequency of DNA transfer, the first being RM system evasion. This method facilitated transfer only inconsistently as well as attaching significant cloning burden and plasmid design concerns. This was followed by specific methylation mimicry, via the expression of *C. carboxidivorans'* methyltransferase genes within *E. coli* sExpress conjugal donors. This facilitated DNA transfer at a rate of 1.4-1.5 x 10^1 CFU/conjugation. The most significant increases in DNA transfer frequency were attained through the isolation of more amenable *C. carboxidivorans* variant strains. This resulted in an increase in DNA transfer efficiency of over three orders of magnitude, reaching 3 x 10^4 CFU/conjugation and permitting transfer from non-recombinant *E. coli* donor strains. Characterisation of four hyper-receptive conjugative *C. carboxidivorans* recipient variants implicated changes in the *fabK* gene, responsible for chain elongation during fatty acid synthesis.

One final RM mitigation strategy was pursued within this study. This was the targeted knockout of *C. carboxidivorans'* RM system restriction endonuclease activity. To apply this strategy a genomic engineering system first had to be established within *C. carboxidivorans*. While allelic exchange using the *C. difficile* derived CD2517.1 toxin proved ineffective, a modified RiboCas system (Cañadas, et al., 2019) was used to generate a Δ*ccaP7ORF17830P* knockout strain. This represents the first report of targeted genetic modification of any kind within *C. carboxidivorans*. Unfortunately, during the course of this study no successful plasmid loss of the CRISPR-Cas9 vector could be identified. This tool has great scope for future applications within *C. carboxidivorans,* essentially opening a new untapped avenue of strain development and research. Thus far, this system has already been applied by C. Humphreys (personal communication) to generate three further knockouts within *C. carboxidivorans* C94.

6.1.2. Further studies

Despite the strides made within this study in DNA transfer into *C. carboxidivorans,* there are still further gains to be attained. In particular, Type

I RM system mitigation would allow the sequence itself to be completely disregarded (Type I RM motif inclusion within the plasmid sequence still a concern at this stage). While this study demonstrates the efficacy of the methyltransferase expression strategy, it fell short of delivering any heterologous Type I methyltransferase containing vectors of sufficient sequence fidelity. As such, the alternative restriction endonuclease disruption approach may pose a better candidate in this instance. This approach has already begun to be applied by C. Humphreys (personal communication), who has used the reported RiboCas system to generate a knockout of *R.CcaP7IP* (Ccar_11975). A double knockout has also been attained within the original Δ*CcaP7ORF17830P* strain, generating a *ΔR.CcaP7IPΔCcaP7ORF17830P* strain whose conjugal receptivity is yet to be assessed. An interesting target could also be restoring the non-receptive phenotype to C94 by reversing the observed amino acid change in *fabK*, or alternatively determining the receptiveness of a complete knockout of *fabK*. Beyond methylation there are many variables within the conjugative protocol which could also be explored. Even seemingly minor alterations, such as variations in media component concentrations, have been reported to yield 4250-fold improvement in transfer in *Geobacillus thermodenitrifican* (Daas, et al., 2016). Also, while no successful transfer was generated by electroporation in the course of this study, these attempts were made prior to the demonstration of the methyltransferase expression. As such these efforts could be repeated with more appropriately methylated DNA isolated from sExpress_TII or sExpress_TIII.

Interestingly conjugal DNA transfer in *C. tyrobutyricum* was improved 15.3-fold both through the knockout of a Type I restriction endonuclease and the elimination of the 16 kb native plasmid pCTK01 (Zhang, et al., 2020). This increase in conjugation efficiency was partially attributed to the removal of the genes CTK_P00290-410 putatively encoding conjugal machinery potentially incurring surface exclusion, and partially to alleviating metabolic burden. With these hypotheses in mind this does not appear directly applicable to the p19 plasmid in *C. carboxidivorans* which lacks any conjugation associated genes and

has a copy number of approximately one (Millard, 2017), however, an increase in DNA transfer frequency between 20 and 2000-fold was observed for *C. beijerinckii* DSM 6423 upon curing of the native pNF2 plasmid (Diallo, et al., 2020). This plasmid was only 4282 bp and contained four genes, of which none were associated with conjugal machinery. As such the p19 native plasmid of *C. carboxidivorans* may still be an interesting alternative target for further elevating DNA transfer frequency. This would need to begin with knockout of the RelE/ParE toxin gene (Ccar_25955) proposed as stabilising the plasmid, followed by CRISPR-Cas9 mediated cleavage of the backbone (without homologous region complementation).

Another pertinent assay which should likely be performed is the characterisation of the sensitivity of *C. carboxidivorans* C94 to hexanol. Hexanol toxicity is thought to occur through impacts on membrane fluidity (Ingram & Vreeland, 1980). As the *fabK* mutation is also thought to impact the membrane structure, this may generate a change in the hexanol sensitivity of the strain. Product toxicity is a key concern observed in the development of solventogenic clostridia as industrial chassis, with the use of *in situ* hexanol in *C. carboxidivorans* fermentation achieving the highest terminal hexanol yield thus far (Kottenhahn, et al., 2021). As hexanol production is one of the most attractive features of *C. carboxidivorans* from an industrial perspective, negative implications on hexanol tolerance may merit the application of similar technology to this strain in future endeavours.

6.2. *Eubacterium limosum*

6.1.3. Original contributions

Within this study genetic tool validation was performed in the novel *E. limosum* strain NG-6849. This culminated in the application of the transposon mutagenesis vector pMTL-CW21 to generate an insertional mutant pool containing 288,171 unique insertion loci, approximately equivalent to one

insertion every 15 bp. This is the highest genome saturation observed within an acetogenic transposon mutagenesis study to date. Preliminary assessment was performed of pertinent discoveries within this dataset, which was also used as a reference for the reverse genetics' studies of A. Agius and J. Millard, wherein the essentiality of certain genes dictated the most appropriate knockout targets within the strain.

6.1.4. Further studies

The high throughput nature of TIS generates a large dataset, which can act as a reference for future studies, for example elucidating the roles of the 37 functionally unknown genes implicated solely during methanol growth of *E. limosum* NG-6849. The genes which were notably overrepresented by the fourth subculture in methanol, B2M23_02925 and B2M23_08625, may also be interesting candidates for knockout with the CRISPR-Cas9 system used within this study. The hypothesis that the increase in mutant prevalence corresponds to an improved growth rate on methanol could then be directly tested. This could also generate a more desirable strain for industrial biotechnological application, which demonstrates improved growth kinetics and product yields on methanol feedstocks. Prior to such studies, it would be pertinent to map the results to an annotated NG-6849 genome when this is successfully attained. Alternatively, more extensive comparison of the ATCC 8486 and NG-6849 genomes could be performed to identify non-synonymous regions which may result in low coverage within the current dataset.

To further assess the validity of the Bio-TraDIS predictions several potential knockout targets have been identified within the dataset, specifically genes predicted as essential solely for growth on methanol (**Table 39**). To corroborate the predictions of the Bio-TraDIS analysis these knockout strains should lack the capacity for growth on methanol. The B2M23_01960 gene identified as a putative methyltransferase (*mtaB*) is one such target of particular interest. Additionally, a further six stocks of the initial TraDIS library remain which could

also prove instrumental in future research, perhaps being applied to examine the essential gene profile of *E. limosum* NG-6849 when grown autotrophically on CO or CO_2 and H_2.

The decrease in library saturation and elevated levels of specific mutants observed within this assay was an unexpected result. To assess the hypothesis regarding growth rates in *mttb* insertional mutants, a knockout strain should be generated, and comparative growth rate assessed between methanol and glucose. This could similarly be performed with the hypothetical protein encoded by B2M23_08625, which represented 33.51% of the total mutant pool in subculture 3 of replicate B on methanol. Within this analysis the length of the characteristic lag phase when shifting carbon source to methanol may also prove pertinent. To ensure the accuracy of the TraDIS essentiality predictions, rarefaction analysis highlighted the capacity for a repeated sequencing run to attain more thorough sequencing sample saturation.

Table 39 – *E. limosum* NG-6849 gene targets for TraDIS complementary knockouts

Locus Tag (B2M23_)	Read Count	Insertion Index	Gene Length	Unique Insertions	Function
08235	0	0	714	0	hypothetical protein
07750	0	0	708	0	sugar fermentation stimulation protein SfsA
13560	0	0	708	0	phosphoribosylaminoimidazolesuccinocarboxamide synthase
19750	0	0	678	0	hypothetical protein
03220	0	0	660	0	ribulose-phosphate 3-epimerase
19175	0	0	660	0	cytidylate kinase
18720	0	0	627	0	V-type ATP synthase subunit D
17655	0	0	606	0	hypothetical protein
02895	0	0	558	0	flavoprotein
18170	0	0	468	0	6,7-dimethyl-8-ribityllumazine synthase
05950	0	0	450	0	ribose 5-phosphate isomerase B
05955	18	0.0008	1323	1	nucleotide sugar dehydrogenase
01170	1	0.0011	1011	1	electron transfer flavoprotein subunit alpha
08330	1	0.0011	996	1	HPr(Ser) kinase/phosphatase
01260	2	0.0012	924	1	two-component sensor histidine kinase
15645	1	0.0013	846	1	PHP domain-containing protein
19505	2	0.0015	768	1	ABC transporter ATP-binding protein
13065	3	0.0015	753	1	hypothetical protein

05935	26	0.0016	1434	2	nicotinate phosphoribosyltransferase
01960	1	0.0027	411	1	hypothetical protein (Uro-D)

Bibliography

2021. *Plasmid Vectors: General Cloning (Clostridium).* [Online] Available at: https://plasmidvectors.com/general-cloning/ [Accessed 16 09 2021].

Adamse, A., 1980. New isolation of *Clostridium aceticum* (Wieringa). *Antonie van Leeuwenhoek,* Volume 46, pp. 523-531.

Ahmed, A., Cateni, B., Huhnke, R. & Lewis, R., 2006. Effects of biomass-generated producer gas constituents on cell growth, product distribution and hydrogenase activity of *Clostridium carboxidivorans* P7T. *Biomass and Bioenergy,* Volume 30, pp. 665-672.

Al-Hinai, M., 2012. Novel system for efficient isolation of Clostridium double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration. *Applied and Environmenal Microbiology,* 78(22), pp. 8112-8121.

Al-Hinai, M., Fast, A. & Papoutsakis, E., 2012. Novel system for efficient isolation of Clostridium double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration. *Applied and Environment Microbiology,* Nov, 78(22), pp. 8112-8121.

AL-Janabi, A., 2011. Potential Activity of the Purine Compounds Caffeine and Aminophylline on Bacteria. *Journal of Global infectious Diseases,* 3(2), pp. 133-137.

Allen, S. & Blaschek, H., 1990. Plasmid copy number and stability determination in *Clostridium perfringens* transformants. *FEMS Microbiology Letters,* 60(3), pp. 323-327.

Altschul, S. et al., 1990. Basic local alignment search tool. *Journal of Molecular Biology,* Volume 215, pp. 403-410.

Annan, F. et al., 2019. Engineering of vitamin prototrophy in *Clostridium ljungdahlii* and *Clostridium autoethanogenum*. *Applied Microbiology and Biotechnology,* Volume 103, pp. 4633-4648.

Arber, W. & Morse, M., 1965. Host Specificity of DNA Produced by *Escherichia coli*. VI. Effects on Bacterial Conjugation. *Genetics,* Jan, 51(1), pp. 137-148.

Aune, T. & Aachman, F., 2010. Methodologies to increase transformation efficiencies and the range of bacteria that can be transformed. *Applied Microbiology and Biotechnology,* Feb, 85(5), pp. 1301-1313.

Aziz, R. et al., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics,* 9(7), p. 75.

Back, M., Flaiz, M., Bengelsdorf, F. & Dürre, 2020. Induced heterologous expression of the arginine deiminase pathway promotes growth advantages in the strict anaerobe *Acetobacterium woodii*. *Applied Microbiology and Biotechnology,* 104(2), pp. 687-699.

Bailey, T. et al., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research,* Volume 37, pp. W202-208.

Barquist, L. et al., 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics,* 32(7), pp. 1109-1111.

Bauen, A. et al., 2016. *Evidencing the Bioeconomy,* s.l.: s.n.

Behr, A., 2014. *Methanol: The Basic Chemical and Energy Feedstock of the Future.* Berlin: Springer.

Belinky, F., Rogozin, I. & Koonin, E., 2017. Selection on start codons in prokaryotes and potential compensatory nucleotide substitutions. *Scientific Reports,* Volume 7, p. 12422.

Bengelsdorf, F. et al., 2016. Industrial Acetogenic Biocatalysts: A Comparative Metabolic and Genomic Analysis. *Frontiers in Microbiology,* Volume 7, p. 1036.

Bertsch, J. & Müller, V., 2015. Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. *Biotechnology for Biofuels,* Volume 8, p. 210.

Bertsch, J. et al., 2015. Heterotrimeric NADH-Oxidizing Methylenetetrahydrofolate Reductase from the Acetogenic Bacterium *Acetobacterium woodii*. *Journal of Bacteriology,* 197(9), pp. 1681-1689.

Bertsch, J., Parthasarathy, A. & Müller, V., 2013. An Electron-bifurcating Caffeyl-CoA Reductase. *The Jounral of Biological Chemistry,* 288(16), p. 11304 –11311.

Bhattacharjee, D. & Sorg, J., 2020. Factors and conditions That Impact Electroporation of *Clostridioides difficile* Strains. *mSphere- American Society of Microbiology,* Mar, 5(2), pp. e00941-19.

Bolger, A., Lohse, M. & Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Biotechnology,* 30(15), pp. 2114-2120.

Braakman, R. & Smith, E., 2012. The Emergence and Early Evolution of Biological Carbon-Fixation. *PLOS Computational Biology,* 8(4), p. e1002455.

Brown, S. et al., 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,* Volume 70.

Brown, T. & Brown, R., 2013. A review of cellulosic bioduel commercial-scale projects in the United States. *Biofuels, Bioproducts and Biorefining,* 7(3), pp. 235-245.

Bruant, G. et al., 2010. Genomic analysis of carbon monoxide utilization and butanol production by *Clostridium carboxidivorans* strain P7. *PLoS ONE,* 5(9).

Burton, F., 2018. *World's First Commercial Waste Gas to Ethanol Plant Starts Up,* s.l.: LanzaTech.

Camps, M., 2010. Modulation of ColE1-like Plasmid Replication for Recombinant Gene Expression. *Recent Patents on DNA and Gene Sequences,* Jan, 4(1), pp. 58-73.

Cañadas, I. et al., 2019. RiboCas: A Universal CRISPR-Based Editing Tool for Clostridium. *ACS Synthetic Biology,* Volume 8, pp. 1379-1390.

Cartman, S. et al., 2012. Precise Manipulation of the *Clostridium difficile* Chromosome Reveals a Lack of Association between the *tcdC* Genotype and Toxin Production. *Applied and Environemntal Microbiology,* 78(13), pp. 4683-4690.

Cartman, S. & Minton, N., 2010. A *mariner*-Based Transposon System for In Vivo Random Mutagenesis of *Clostridium difficile*. *Applied and Environmental Microbiology,* 76(4), pp. 1103-1109.

Chang, A., Chau, V., Landas, J. & Pang, Y., 2017. Preparation of calcium competent *Escherichia coli* and. *JEMI Methods,* Volume 1, pp. 22-25.

Charubin, K., Streett, H. & Papoutsakis, E., 2020. Development of Strong Anaerobic Fluorescent Reporters for *Clostridium acetobutylicum* and *Clostridium ljungdahlii* Using HaloTag and SNAP-tag Proteins. *Applied and Environmental Microbiology,* 86(20), pp. e01271-20.

Chemical and Allied Industries, 2009. *Sustainable U.S. Manufacturing: Alternative Feedstocks,* s.l.: American Chemical Society.

Cheng, C., Li, W., Lin, M. & Yang, S., 2019. Metabolic engineering of *Clostridium carboxidivorans* for enhanced ethanol and butanol production from syngas and glucose. *Bioresource Technology,* Jul, Volume 284, pp. 415-423.

Clark, T. et al., 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Research,* 40(4).

Climate Change Committee, 2019. *NEt Zero - The UK's contribution to stopping global warming,* s.l.: s.n.

Cotton, C. A., Edlich-Muth, C. & Bar-Even, A., 2018. Reinforcing carbon fixation: CO_2 reduction replacing and supporting carboxylation. *Current opinion in biotechnology,* Volume 49, pp. 49-56.

Couvin, D. et al., 2018. CRSIPRCasFinder, an update of the CRISPRFinder includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research,* Jul, 2(46), pp. W246-W251.

Crampton, N. et al., 2007. DNA looping and translocation providean optimal cleavage mechanism for the type III restriction enzymes. *The EMBO Journal,* Aug, 26(16), pp. 3815-3825.

Daas, M. et al., 2016. Isolation of a genetically accessible thermophilic xylan degrading bacterium from compost. *Biotechnology for Biofuels,* 9(210).

Darling, A., Mau, B. & Perna, N., 2010. progressiveMauve: Multiple Genome Alignment with Gain, Loss and Rearrangement. *PLoS ONE,* 5(6).

Das, A. et al., 2007. Characterization of a corrinoid protein involved in the C1 metabolism of strict anaerobic bacterium *Moorella thermoacetica*. *Proteins: Structure, Function, and Bioinformatics,* 67(1), pp. 167-176.

Dembek, M. et al., 2015. High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *mBio,* 6(2), p. e02383.

Department for Business, Energy & Industrial Strategy, 2018. *Growing the bioeconomy: a national bioeconomy strategy to 2030,* s.l.: s.n.

Department of Energy and Climate Change, 2011. *The Carbon Plan - reducing greenhouse gas emissions,* s.l.: s.n.

Diallo, M. et al., 2020. Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*. *Methods,* Volume 172, pp. 51-60.

Diender, M., Stams, A. & Sousa, D., 2015. Pathways and Bioenergetics of Anaerobic Carbon Monoxide Fermentation.. *Frontiers in Microbiology,* Volume 6.

Doench, J. et al., 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology,* Volume 34, pp. 184-191.

Dong, H., Tao, W., Zhang, Y. & Li, Y., 2012. Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering. *Metabolic Engineering,* 14(1), pp. 59-67.

Dong, H., Zhang, Y., Dai, Z. & Li, Y., 2010. Engineering Clostridium Strain to Accept Unmethylated DNA. *PLoS One,* 5(2).

Dubnau, D. & Blokesch, M., 2019. Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annual Review of Genetics,* 21 August, Volume 53, pp. 217-237.

E4tech, 2016. *The commercial and strategic opportunity offered by gas fermentation in the UK.,* s.l.: E4tech (UK) Ltd.

Eggerth, A., 1935. The Gram-positive non-spore-bearing anaerobic bacilli of human feces. *Journal of Bacteriology,* 30(3), pp. 277-299.

Ehsaan, M., Keuhne, S. & Minton, N., 2016. *Clostridium difficile* genome editing using pyrE alleles. *Methods in Molecular Biology,* Issue 1476, pp. 35-52.

Ellenborgen, J. et al., 2021. The MttB superfamily member MtyB from the human gut symbiont *Eubacterium limosum* is a cobalamin-dependent γ-butyrobetaine methyltransferase. *Journal of Biological Chemsitry,* 297(5).

Ellman, G., 1959. Tissue sulfhydryl groups. *Archives of Biochemistry and Biophysics,* 82(1), pp. 70-77.

Engler, C., Kandzia, R. & Marillonnet, S., 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE,* 3(11), p. e3647.

Fackler, N. et al., 2021. Transcriptional control of *Clostridium autoethanogenum* using CRISPRi. *Synthetic Biology,* 6(1).

Fast, A. & Papoutsakis, E., 2012. Stoichiometric and energetic analyses of non-photosynthetic CO_2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Current Opinion in Chemical Engineering,* 1(4), pp. 380-395.

Faulds-Pain, A. & Wren, B., 2013. Improved Bacterial Mutagenesis by High-Frequency Allele Exchange, Demonstrated in *Clostridium difficile* and *Streptococcus suis*. *Applied and Environmental Microbiology,* 79(15), pp. 4768-4771.

Fernández-Naveira, Á., Abubackar, H., Veiga, M. & Kennes, C., 2016. Carbon monoxide bioconversion to butanol-ethanol by *Clostridium carboxidivorans*: kinetics and toxicity of alcohols. *Applied Microbiology and Biotenchnology,* 100(9), pp. 4231-4240.

Fernández-Naveira, Á., Abubackar, H., Veiga, M. & Kennes, C., 2016. Efficient butanol-ethanol (B-E) production from carbon monoxide fermentation by *Clostridium carboxidivorans*. *Applied Microbiology and Biotechnology,* 100(7), pp. 3361-3370.

Fernández-Naveira, A., Abubackar, H., Veiga, M. & Kennes, C., 2017. Production of chemicals from C1 gases (CO, CO 2) by *Clostridium carboxidivorans*. *World Journal of Microbiology and Biotechnology,* 33(3), p. 43.

Fernández-Naveira, Á., Veiga, M. & Kennes, C., 2017. Glucose bioconversion profile in the syngas-metabolizing species *Clostridium carboxidivorans*. *Bioresource Technology,* 244(Pt 1), pp. 552-559.

Flaiz, M., Ludwig, G., Bengelsdorf, F. & Dürre, P., 2021. Production of the biocommodities butanol and acetone from methanol with fluorescent FAST-tagged proteins using metabolically engineered strains of *Eubacterium limosum*. *Biotechnology for Biofuels,* 14(1).

Foulquier, C. et al., 2019. An efficient method for markerless mutant generation by allelic exchange in *Closridium acetobutylicum* and *Clostridium saccharobutylicum* using suicide vectors. *Biotechnology and Biofuels,* Volume 12.

Fremp, F., Poehlein, A., Daniel, R. & Müller, V., 2018. Methanol metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environmental Microbiology,* 20(12), pp. 4369-4384.

Fukuda, E., Kaminska, K., Bujnicki, J. & Kobayashi, I., 2008. Cell death upon epigenetic genome methylation: a novel function of methyl-specific deoxyribonucleases. *Genome Biology,* Nov, 9(11), p. R163.

Fu, Y., Tsai, M., Luo, Y. & Deonier, R., 1991. Deletion anaylsis of the F plasmid *oriT* locus. *Journal of Bacteriology,* 108(11), pp. 1012-1020.

Gago-Córdoba, C. et al., 2019. Surface Exclusion Revisitied: Function Related to Differential Expression of the Surface Exclusion System of *Bacillus subtilis* Plasmid pLS20. *Frontiers in Microbiology,* Jul, 10(10), p. 1502.

Gasiunas, G., Barrangou, R., Hovarth, P. & Siksnys, V., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS,* 109(39), pp. E2579-E2586.

Gawronski, J. et al., 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *PNAS,* 106(38), pp. 16422-16427.

Giri, A., Das, V., Reddy, V. & Pal, A., 1999. Mutagenic and genotoxic effects of theophylline and theobromine in Salmonella assay and in vivo sister chromatid exchanges in bone marrow cells of mice. *Mutation Research,* 444(1), pp. 17-23.

Gomaa, A. et al., 2017. An enhanced vector-free allele exchange (VFAE) mutagenesis protocol for genome editing in a wide range of bacterial species. *AMB Express,* Volume 7, p. 125.

Goodman, A. et al., 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host and Microbe,* 6(3), pp. 279-289.

Gossen, M. & Bujard, H., 1993. Anhydrotetracycline, a novel effector for tetracycline controlled gene expression systems in eukaryotic cells. *Nucleic Acids Research,* 21(18), pp. 4411-4412.

Gouil, Q. & Keniry, A., 2019. *Essays in Biochemistry,* 63(6), pp. 639-648.

Green, E. & Bennett, G., 1996. Inactivation of an aldehyde/alcohol dehydrogenase gene from *Clostridium acetobutylicum* ATCC 824. *Applied Biochemistry and Biotechnology,* Volume 57-58, pp. 213-221.

Griffith, F., 1928. The Significance of Pneumococcal Types. *Journal of Hygeine,* 27(2), pp. 113-159.

Grosse-Honebrink, A. et al., 2017. Improving gene transfer in *Clostridium pasteurianum* through the isolation of rare hypertransformable variants. *Anaerobe,* Volume 48, pp. 203-205.

Guo, C.-J.et al., 2019. Depletion of microbiome-derived molecules in the host using Clostridium genetics. *Science,* 366(6471).

Guss, A., Olson, D., Caiazza, N. & Lynd, L., 2012. Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. *Biotechnology for Biofuels,* Volume 5.

Han, Y.-F.et al., 2020. Combination of Trace Metal to Improve Solventogenesis of *Clostridium carboxidivorans* P7 in Syngas Fermentation. *Frontiers in Microbiology,* Volume 11, p. 577266.

Hartman, A., Liu, H. & Melville, S., 2011. Construction and Characterization of a Lactose-Inducible Promoter System for Controlled Gene Expression in *Clostridium perfringens*. *Applied and Environmental Microbiology,* 77(2), pp. 471-478.

Heap, J. et al., 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research,* 40(8), p. e59.

Heap, J. et al., 2007. The ClosTron: A universal gene knock-out system for the genus Clostridium. *Journal of Micrbiological Methods,* 70(3), pp. 452-464.

Heap, J., Pennington, O., Cartman, S. & Minton, N., 2009. A modular system for Clostridium shuttle plasmids. *Jounal of Microbiological Methods,* Jul, 78(1), pp. 79-85.

Hecht, A. et al., 2017. Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Research,* 45(7), pp. 3615-3626.

Hedderich, R. & Forzi, L., 2005. Energy-converting [NiFe] hydrogenases: more than just H_2 activation. *Journal of Molecular Microbiology and Biotechnology,* 10(2), pp. 92-104.

Heijstra, B., Leang, C. & Juminaga, A., 2017. Gas fermentation: cellular engineering possibilities and scale up. *Microbial Cell Factories,* 16(1), p. 60.

Heijstra, B., Leang, C. & Juminaga, A., 2017. Gas fermentation: cellular engineering possibilities and scale up. *Microbial Cell Factories,* Volume 16.

Hille, F. et al., 2018. The Biology of CRISPR-Cas: Backward and Forward. *Cell,* Mar, 172(6), pp. 1239-1259.

Hillen, W. & Berens, C., 1994. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Anual Review of Microbiology,* Volume 48, pp. 345-367.

Honda, K., 2019. *Eubacterium Isolate 81C1: chromosome 1: NZ_LR215983,* s.l.: NCBI.

Hong, W. et al., 2018. Multiplexed CRISPR-Cpf1-Mediated Genome Editing in *Clostridium difficile* toward the Understanding of Pathogenesis of C. difficile Infection. *ACS Synthetic Biology,* 7(6), pp. 1588-1600.

Horne, J. et al., 2020. Caffeine and Theophylline Inhibit β-Galactosidase Activity and Reduce Expression in *Escherichia coli*. *ACS Omega,* 5(50), pp. 32250-32255.

Hsu, P. et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology,* Volume 31, pp. 827-832.

Huang, C.-N., Liebl, W. & Ehrenreich, A., 2018. Restriction-defcient mutants and marker-less genomic modification for metabolic engineering of the solvent produceer *Clostridium saccharobutylicum*. *Biotechnology for Biofuels,* Volume 11.

Huang, H. et al., 2016. CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. *ACS Synthetic Biology,* 5(12), pp. 1355-1361.

Humphreys, C. et al., 2015. Whole genome sequence and manual annotation of *Clostridium autoethanogenum*, an industrially relevant bacterium. *BMC Genomics,* Volume 16.

Hutchinson, C. et al., 2016. Design and synthesis of a minimal bacterial genome. *Science,* 351(6280).

Hutchison, C. et al., 1999. Global transposon mutagenesis and a minimal Mycoplasma genome. *Science,* 286(5447), pp. 2165-2169.

Ingle, P. et al., 2019. Generation of a fully erythromycin-sensitive strain of *Clostridioles difficile* using a novel CRISPR-Cas9 genome editing system. *Scientific Reports,* Volume 9, p. 8123.

Ingram, L. & Vreeland, N., 1980. Differential effects of ethanol and hexanol on the *Escherichia coli* cell envelope. *Journal of Bacteriology,* 144(2), pp. 481-488.

IRENA and Methanol Institute, 2021. *Innovation Outlook: Renewable Methanol,* Abu Dhabi: International Renewable Energy Agency.

Ishino, M. et al., 1987. Nucleotide sequence of the adenovirus type 40 inverted terminal repeat: Close relation to that of adenovirus 5. *Virology,* 156(2), pp. 414-416.

Jalajakumari, M. et al., 1987. Surface exclusion genes traS and traT of the sex factor of *Escherichia coli* K-12: Determination of the nucleotide sequence and promoter and terminator activities. *Journal of Molecular Biology,* Nov, 198(1), pp. 1-11.

Janulaitis, A. et al., 1992. Purification and properties of the Eco57I restriction endonuclease and methylase--prototypes of a new class (type IV). *Nucleic Acids Research,* Nov, 20(22), pp. 6043-6049.

Jeong, H., Sim, Y., Kim, H. & Lee, S., 2017. Unveiling the Hybrid Genome Structure of *Eschirichia coli* RR1 (HB101 RecA+). *Frontiers in Microbiology,* Apr, 4(8), p. 585.

Jeong, J. et al., 2015. Energy Conservation Model Based on Genomic and Experimental Analyses of a Carbon Monoxide-Utilizing, Butyrate-Forming Acetogen, *Eubacterium limosum* KIST612. *Applied and Environmental Microbiology,* Volume 81, pp. 4782-4790.

Jeong, J. et al., 2020. Genetic engineering system for syngas-utilizing acetogen, *Eubacterium limosum* KIST612. *Bioresource Technology Reports,* Volume 11.

Jia, K., Zhu, Y., Zhang, Y. & Li, Y., 2011. Group II intron-anchored gene deletion in Clostridium. *PLOS ONE,* 6(1), p. e16693.

Johnston, C. et al., 2019. Systematic evasion of the restriction-modification barrier in bacteria. *Proceedings of the National Academy of Sciences of the USA,* 116(23), pp. 11454-11459.

Jones, J. et al., 2019. Small-Molecule Inhibition of the *C. difficile* FAS-II Enzyme, FabK, Results in Selective Activity. *ACS Chemical Biology,* Volume 14, pp. 1528-1535.

Joseph, R., Kim, N. & Sandoval, N., 2018. Recent Developments of the Synthetic Biology Toolkit for Clostridium. *Frontiers in Microbiology,* Volume 9, p. 154.

Kang, S. et al., 2020. Adaptive Laboratory Evolution of *Eubacterium limosum* ATCC 8486 on Carbon Monoxide. *Frontiers in Microbiology,* Volume 11.

Kelly, W. et al., 2016. The complete genome sequence of *Eubacterium limosum* SA11, a metabolically versatile rumen acetogen. *Standards in Genomic Science,* 11(26), pp. -.

Kennaway, C. et al., 2012. Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Genes & Development,* Jan, 26(1), pp. 92-104.

Kocholaty, W., 1942. Cultural characteristics of *Penicillium notatum* in relation to the production of antibacterial substance. Indication of the dual nature of the antibacterial substance. *Journal of Bacteriology,* Volume 44, pp. 469-477.

Kolek, J., Sedlar, K., Provaznik, I. & Patakova, P., 2016. Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnology for Biofuels,* Jan.9(14).

Koren, S. et al., 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biology,* Volume 14, p. R101.

Kothandaraman, J. et al., 2016. Conversion of CO_2 from Air into Methanol Using a Polyamine and a Homogeneous Ruthenium Catalyst. *Journal of the American Chemical Society,* 138(3), pp. 778-781.

Kottenhahn, P., Philipps, G. & Jennewein, S., 2021. Hexanol biosynthesis from syngas by *Clostridium carboxidivorans* P7 – product toxicity, temperature dependence and in situ extraction. *Heliyon,* 7(8).

Kountz, D., Berham, E., Zhang, L. & Krzycki, J., 2020. MtcB, a member of the MttB superfamily from the human gut acetogen *Eubacterium limosum*, is a cobalamin-dependent carnitine demethylase. *Journal of Biological Chemistry,* 295(34), pp. 11971-11981.

Kremp, F., Poehlein, A., Daniel, R. & Müller, V., 2018. Methanol metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environmental Microbiology,* 20(12), pp. 4369-4384.

Kuehne, S. & Minton, N., 2012. ClosTron-mediated engineering of Clostridium. *Bioengineered,* 3(4), pp. 247-254.

Kuhns, M. et al., 2020. The Rnf complex from the acetogenic bacterium *Acetobacterium woodii*: Purification and characterization of RnfC and RnfB. *Biochimica et biophysica acta. Bioenergetcis,* 1861(11).

Lampe, D. et al., 1999. Hyperactive transposase mutants of the *Himar1 mariner* transposon. *PNAS,* 96(20), pp. 11428-11433.

Lampe, D., Churchill, M. & Robertson, H., 1996. A purified *mariner* transposase is sufficient to mediate transposition in vitro. *EMBO Journal,* 15(19), pp. 5470-5479.

Lanckriet, A. et al., 2009. Generation of single-copy transposon insertions in *Clostridium perfringens* by electroporation of phage mu DNA transposition complexes. *Applied and Environmental Microbiology,* 75(9), pp. 2638-2642.

Lane, J., 2015. *LanzaTech, INVISTA find direct pathway to bio-based butadiene,* s.l.: BioFuels Digest.

Langridge, G. et al., 2009. Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants. *Genome Research,* 19(12), pp. 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,* Volume 27, pp. 79-87.

Law, K., Rosenfeld, J. & Jackson, M., 2013. *Methanol as a Renewable Energy Resource,* s.l.: TIAX LLC.

Leang, C., Ueki, T., Nevin, K. & Lovley, D., 2013. A Genetic System for *Clostridium ljungdahlii*: a Chassis for Autotrophic Production of Biocommodities and a Model Homoacetogen. *Applied and Environmental Microbiology,* 79(4), pp. 1102-1109.

Leang, C., Ueki, T., Nevin, K. & Lovley, D., 2013. A Genetic System for *Clostridium ljungdahlii*: a Chassis for Autotrophic Production of Biocommodities and a Model Homoacetogen. *Applied and Environmental Microbiology,* 79(4), pp. 1102-1109.

Lederberg, J., Cavalli, L. & Lederberg, E., 1952. Sex Compatability in *Escherichia Coli*. *Genetics,* Nov, 37(6), pp. 720-730.

Lee, S., Mermelstein, L. & Papoutsakis, E., 1993. Determination of plasmid copy number and stability in *Clostridium acetobutylicum* ATCC 824.. *FEMS Micrbiology Letters,* 108(3), pp. 319-323.

Lee, S. et al., 2009. Cellular stress created by intermediary metabolite imbalances. *Proceedings of the National Academy of Sciences, USA,* Volume 106, pp. 19515-19520.

Lesiak, J., 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,* Volume 188, pp. 97-99.

Liew, F., 2016. *Metabolic engineering of Clostridium autoethanogenum,* s.l.: University of Nottingham.

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics,* 34(18), pp. 3094-3100.

Li, N. et al., 2015. Complete genome sequence of *Clostridium carboxidivorans* P7(T), a syngas-fermenting bacterium capable of producing long-chain alcohols. *Journal of Biotechnology,* Oct, Volume 211, pp. 44-45.

Liou, J. S.-C., Balkwill, D., Drake, G. & Tanner, R., 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 Biology,* 55(Pt 5), pp. 2085-2091.

Li, Q. et al., 2016. CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnology Journal,* Volume 11, pp. 961-972.

Li, Q. et al., 2019. CRISPR-Cas9(D10A) nickase-assisted base editing in solvent producer *Clostridium beijerinckii*. *Biotechnology and Bioengineering,* Volume 116, pp. 1475-1483.

Litty, D. & Müller, V., 2021. Butyrate production in the acetogen *Eubacterium limosum* is dependent on the carbon and energy source. *Microbial Biotechnology,* 0(0), pp. 1-7.

Liu, H., Bouillaut, L., Sonenshein, A. & Melville, S., 2013. Use of a Mariner-Based Transposon Mutagenesis System To Isolate *Clostridium perfringens* Mutants Deficient in Gliding Motility. *Journal of Bacteriology,* 195(3), pp. 629-636.

Liu, L. et al., 2012. Comparison of Next-Generation Sequencing Systems. *BioMed Research International,* p. 251364.

Liu, Y. et al., 2017. Engineering cell signaling using tunable CRISPR–Cpf1-based transcription factors. *Nature Communications,* Volume 8, p. 2095.

Loenen, W. & Raleigh, E., 2014. The other face of restriction: modification-dependent enzymes. *Nucleic Acids Research,* 42(1), pp. 56-59.

Lorenzo-Díaz, F. et al., 2017. Crosstalk between vertical and horizontal gene transfer: plasmid replication control by a conjugative relaxase. *Nucleic Acids Research,* Jul, 45(13), pp. 7774-7785.

Louis, P. et al., 2004. Restricted Distribution of the Butyrate Kinase Pathway among Butyrate-Producing Bacteria from the Human Colon. *Journal of Bacteriology,* 186(7), pp. 2099-2106.

Mansfield, R., 2018. *Developments in genetic engineering of novel acetogens,* s.l.: University of Nottingham.

Marraffini, L. & Sontheimer, E., 2008. CRISPR interference limits horizontal gene transfer in stphylococci by targeting DNA. *Science,* Dec, 322(5909), pp. 1843-1845.

Martin, F., 2012. Hydrogen, metals, bifurcating electrons, and proton gradients: The early evolution of biological energy conservation. *FEBs Letters,* 586(5), pp. 485-493.

McAllister, K. & Sorg, J., 2019. CRISPR Genome Editing Systems in the Genus Clostridium: a Timely Advancement. *Journal of Bacteriology,* 201(16), pp. e00219-19.

Merck, n.d. *65700 N6-Methyladenine.* [Online] Available at: https://www.sigmaaldrich.com/GB/en/product/aldrich/65700?gclid=Cj0KCQjworiXBhDJARIsAMuzAuyGHERV0ayi8mAZnoWtDUgPXG21XajJlE00gNelkqn_sLuTogXPv30aAsdlEALw_wcB [Accessed 2018].

Merck, n.d. *M6751 5-Methylcytosine hydrochloride.* [Online] Available at: https://www.sigmaaldrich.com/GB/en/product/sigma/m6751?gclid=Cj0KCQjworiXBhDJARIsAMuzAuyAcLvmHu59FTproJiP7XCpR2VtlZJsRYUlbQTTouLTzOXkck0ZzCIaAq2YEALw_wcB [Accessed 2018].

Mermelstein, L. & Papoutsakis, E., 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.. *Applied Environmental Microbiology,* Apr, 59(4), pp. 1077-1081.

Merritt, J. et al., 2007. Construction of a counterselection-based in-frame deletion system for genetic studies of *Streptococcus mutans*. *Oral Microbiology and Immunology,* 22(2), pp. 95-102.

Millard, C., 2017. *Development of Forward and Reverse Genetic Tools for Eubacterium limosum,* s.l.: University of Nottingham.

Miller, C. et al., 2013. Short-read assembly of full-length 16S amplicons reveals bacterial diversity in subsurface sediments. *PLoS One,* 8(2), p. e56018.

Minton, N. et al., 2016. A roadmap for gene system development in Clostridium. *Anaerobe,* Volume 41, pp. 104-112.

Mock, J. et al., 2014. Evidence for a Hexaheteromeric Methylenetetrahydrofolate Reductase in *Moorella thermoacetica*. *Journal of Bacteriology,* 196(18), pp. 3303-3314.

Molitor, B. et al., 2016. Expanding the molecular toolkit for the homoacetogen *Clostridium ljungdahlii*. *Scientific reports,* Volume 6.

Müller, V., 2003. Energy Conservation in Acetogenic Bacteria. *Applied and Environmental Microbiology,* 69(11), pp. 6345-6353.

Müller, V. et al., 2008. Discovery of a ferredoxin:NAD+-oxidoreductase (Rnf) in *Acetobacterium woodii*: a novel potential coupling site in acetogens. *Annals of the New York Academy of Sciences,* Volume 1125, pp. 137-146.

Nagaraju, S. et al., 2016. Genome editing of *Clostridium autoethanogenum* using CRISPR/Cas9. *Methodology,* Volume 9.

Nariya, H. et al., 2011. Development and Application of a Method for Counterselectable In-Frame Deletion in *Clostridium perfringens*. *Applied and Environmental Microbiology,* 77(4), pp. 1375-1382.

Nasseri, A., Rasoul-Amini, S., Morowvat, M. & Ghasemi, Y., 2011. Single Cell Protein: Production and Process. *American Journal of Food Technology,* 6(2), pp. 103-116.

NEB, n.d. *New England Biolabs- High Efficiency Transformation Protocol (C2987H/C2987I).* [Online]
Available at: https://international.neb.com/protocols/0001/01/01/high-efficiency-transformation-protocol-c2987
[Accessed 2017].

Ng, Y. et al., 2013. Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using pyrE alleles. *PLOS ONE,* 8(2), p. e56051.

O'Brien, R. & Morris, J., 1971. The Ferredoxin-dependent Reduction of Chloramphenicol by *Clostridium acetobutylicum*. *Microbiology,* 67(3).

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,* Volume 7, pp. 263-270.

Peltier, J. et al., 2020. Type I toxin-antitoxin systems contribute to mobile genetic elements maintenance in *Clostridioides difficile*. *Communications Biology,* Volume 3, p. 718.

Phillips, J. et al., 2015. Butanol and hexanol production in *Clostridium carboxidivorans* syngas fermentation: Medium development and culture techniques. *Bioresource Technology,* Volume 190, pp. 114-121.

Picking, J., Behrman, E., Zhang, L. & Krzycki, J., 2019. MtpB, a member of the MttB superfamily from the human intestinal acetogen *Eubacterium limosum*, catalyzes proline betaine demethylation. *Journal of Biological Chemistry,* 294(37), pp. 13697-13707.

Pierce, E. et al., 2008. The Complete Genome Sequence of Moorella thermoacetica (f. *Clostridium thermoaceticum*). *Environmental Microbiology,* 10(10), pp. 2250-2273.

Pourcel, C. et al., 2020. CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. *Nucleic Acids Research,* Jan, 48(D1), pp. D535-D544.

Purdy, D. et al., 2002. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular Microbiology,* Oct, 46(2), pp. 439-452.

Pyne, M. et al., 2014. Technical guide for genetic advancement of underdeveloped and intractable Clostridium. *Biotechnology Advances,* May, 32(3), pp. 623-641.

Pyne, M. et al., 2014. Technical guide for genetic advancement of underdeveloped and intractable Clostridium. *Biotechnology Advances,* 32(3), pp. 623-641.

Pyne, M. et al., 2016. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Scientific Reports,* Volume 6, p. 25666.

Pyne, M. et al., 2016. Genome-directed analysis of prophage excision, host defence systems, and central fermentative metabolism in *Clostridium pasteurianum*. *Scientific Reports,* Volume 6.

Pyne, M., Moo-Young, M., Chung, D. & Chou, C., 2013. Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. *Biotechnology for Biofuels,* Apr, Volume 6, p. 50.

Pyne, M., Moo-Young, M., Chung, D. & Chou, C., 2014. Expansion of the genetic toolkit for metabolic engineering of *Clostridium pasteurianum*: chromosomal gene disruption of the endogenous CpaAI restriction enzyme. *Biotchnology for Biofuels,* Volume 7.

Ragsdale, S. & Pierce, E., 2008. Acetogenesis and the Wood-Ljungdahl Pathway of CO2 Fixation. *Biochimica et Biophysica Acta,* 1784(12), pp. 1873-1898.

Ramió-Pujol, S., Ganigué, R., Bañeras, L. & Colprim, J., 2014. Impact of formate on the growth and productivity of *Clostridium ljungdahlii* PETC and *Clostridium carboxidivorans* P7 grown on syngas. *International Microbiology: the official journal of the Spanish Society for Microbiology,* 17(4), pp. 195-204.

Ran, F. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols,* 8(11), pp. 2281-2308.

Riley, L. et al., 2019. Rational development of transformation in *Clostridium thermocellum* ATCC 27405 via complete methylome analysis and evasion of native restriction–modification systems. *Journal of Industrial Microbiology & Biotechnology,* Jul, Volume 46, pp. 1435-1443.

Roberts, R. et al., 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acid Research,* Apr, 31(7), pp. 1805-1812.

Roberts, R., Vincze, T., Posfai, J. & Macelis, D., 2015. REBASE -- a database for DNA restriction and modification enzymes, genes and genomes. *Nucleic Acids Research,* Jan, 43(Database issue), pp. D298-D299.

Roh, H. et al., 2011. Complete genome sequence of a carbon monoxide-utilizing acetogen, *Eubacterium limosum* KIST612. *Journal of Bacteriology,* 193(1), pp. 307-8.

Roytka, D., Nicze, T., Posfai, J. & Macelis, D., 2006. Horizontal Gene Transfer and Evolution of Microvirid Coliphage Genomes. *Journal of Bacteriology,* 188(3), pp. 1134-1142.

Rozen, S. & Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology,* Volume 132, pp. 365-386.

Rückel, A. et al., 2021. Studies on Syngas Fermentation With *Clostridium carboxidivorans* in Stirred-Tank Reactors With Defined Gas Impurities. *Frontiers in Microbiology,* Volume 12, p. 655390.

Rudolph, M., Vockenhuber, M.-P. & Suess, B., 2015. Conditional control of gene expression by synthetic riboswitches in *Streptomyces coelicolor*. *Methods in Enzymology,* Volume 550, pp. 283-299.

Rusinov, I. et al., 2018. Avoidance of recognition sites of restriction-modification systems is a widespread but not universal anti-restriction strategy of prokaryotic viruses. *BMC Genomics,* Dec, 19(1), p. 885.

Saad, M., Dosoky, N., Zoromba, M. & Shafik, H., 2019. Algal Biofuels: Current Status and Key Challenges. *Energies,* 12(10).

Sacks, L. & Thompson, P., 1977. Increased spore yields of *Clostridium perfringens* in the presence of methylxanthines. *Applied and Environmental Microbiology,* 34(2), pp. 189-193.

Saito, J. et al., 2008. Crystal structure of enoyl–acyl carrier protein reductase (FabK) from *Streptococcus pneumoniae* reveals the binding mode of an inhibitor. *Protein Science,* 17(4), pp. 691-699.

Sandoval, N., Venkataramanan, K., Groth, T. & Papoutsakis, E., 2015. Whole-genome sequence of an evolved *Clostridium pasteurianum* strain reveals Spo0A deficiency responsible for increased butanol production and superior growth. *Biotechnology and Biofuels,* Volume 8.

Sanger, F., Nicklen, S. & Coulson, A., 1977. DNA sequencing with chain-terminating inhibitors. *PNAS,* 74(12), pp. 5463-5467.

Schmitz, R. et al., 2019. Improved electrocompetence and metabolic engineering of *Clostridium pasteurianum* reveals a new regulation pattern of glycerol fermentation. *Engineering in Life Sciences,* 19(6), pp. 412-422.

Schrader, J. et al., 2009. Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. *Trends in biotechnology,* 27(7), pp. 107-115.

Schuchmann, K. & Müller, V., 2012. A Bacterial Electron-bifurcating Hydrogenase. *Journal of Biological Chemistry,* 287(37), pp. 31165-31171.

Schuchmann, K. & Müller, V., 2016. Energetics and Application of Heterotrophy in Acetogenic Bacteria. *Applied and Environmental Microbiology,* 82(14), pp. 4056-4069.

Senior, M., 2020. Europe's biotech renaissance. *Nature Biotechnology,* Volume 38, pp. 408-415.

Shen, S. et al., 2017. Enhanced alcohol titre and ratio in carbon monoxide-rich off-gas fermentation of *Clostridium carboxidivorans* through combination of trace metals optimization with variable-temperature cultivation. *Bioresource Technology,* Volume 239, pp. 236-243.

Shen, Y., Brown, R. & Wen, Z., 2014. Enhancing mass transfer and ethanol production in syngas fermentation of *Clostridium carboxidivorans* P7 through a monolithic biofilm reactor. *Applied Energy,* 136(C), pp. 68-76.

Shen, Y., Brown, R. & Wen, Z., 2014. Syngas fermentation of *Clostridium carboxidivoran* P7 in a hollow fiber membrane biofilm reactor: Evaluating the mass transfer coefficient and ethanol production performance. *Biochemical Engineering Journal,* Volume 85.

Shen, Y., Brown, R. & Wen, Z., 2017. Syngas fermentation by Clostridium carboxidivorans P7 in a horizontal rotating packed bed biofilm reactor with enhanced ethanol production. *Applied Energy,* Feb, Volume 187, pp. 585-594.

Shin, J. et al., 2019. Genome Engineering of *Eubacterium limosum* Using Expanded Genetic Tools and the CRISPR-Cas9 system. *ACS Synthetic Biology,* 8(9), pp. 2059-2068.

Shin, J., Song, Y., JEong, Y. & Cho, B.-K., 2016. Analysis of the Core Genome and Pan-Genome of Autotrophic Acetogenic Bacteria. *Frontiers in Microbiology,* Volume 7, p. 1531.

Song, Y. & Cho, B.-K., 2015. Draft Genome Sequence of Chemolithoautotrophic Acetogenic Butanol-Producing *Eubacterium limosum* ATCC 8486. *Genome Announcements,* 3(1).

Song, Y. et al., 2017. Determination of the Genome and Primary Transcriptome of Syngas Fermenting *Eubacterium limosum* ATCC 8486. *Scientific Reports,* Volume 7.

Song, Y. et al., 2018. Genome-scale analysis of syngas fermenting acetogenic bacteria reveals the translational regulation for its autotrophic growth. *BMC Genomics,* Volume 19.

Steett, H., Kalis, K. & Papoutsakis, E., 2019. A strongly fluorescing anaerobic reporter and protein-tagging system for Clostridium organisms based on the fluorescence-activating and absorption-shifting tag protein (FAST). *Applied and Environmental Microbiology,* 85(14), pp. e00622-19.

Stoiber, M. et al., 2016. De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. *bioRxiv.*

Streett, H., Charubin, K. & Papoutsakis, E., 2021. Anaerobic fluorescent reporters for cell identification, microbial cell biology and high-throughput screening of microbiota and genomic libraries. *Current Opinion in Biotechnology,* Volume 71, pp. 151-163.

Summers, R., Mohanty, S., Gopishett, S. & Subramanian, M., 2015. Genetic characterization of caffeine degradation by bacteria and its potential applications. *Microbial Biotechnology,* 8(3), pp. 369-378.

Thomas, C. & Nielsen, K., 2015. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Review Microbiology,* 3(9), pp. 711-721.

Tourancheau, A., Mead, E., Zhang, X.-S. & Fang, G., 2021. Discovering multiple types of DNA methylation from bacteria and microbiome using nanopore sequencing. *Nature Methods,* Volume 18, pp. 491-498.

Tyler, A. et al., 2018. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Scientific Reports,* Volume 8, p. 10931.

Ueki, T., Inouye, S. & Inouye, M., 1996. Positive-negative KG cassettes for construction of multi-gene deletions using a single drug marker. *Gene,* 183(1-2), pp. 153-157.

Ueki, T., Nevin, K., Woodard, T. & Lovley, D., 2014. Converting Carbon Dioxide to Butyrate with an Engineered Strain of *Clostridium ljungdahlii*. *mBio,* 5(5), pp. e01636-14.

UK Government, 2008. *Climate Change Act,* s.l.: s.n.

UNFCC, 2015. *Paris Agreement,* s.l.: United Nations.

United Nations, n.d. *Sustainable Development Goals.* [Online] Available at: https://sdgs.un.org/goals [Accessed 2021].

Valgepea, K. et al., 2017. Maintenance of ATP Homeostasis Triggers Metabolic Shifts in Gas-Fermenting Acetogens. *Cell Systems,* 4(5), pp. 505-515.

van Bennekom, J., Venderbosch, R. & Heeres, H., 2012. Biomethanol from Glycerol. In: *Biodiesel - Feedstocks, Production and Applications.* s.l.:s.n.

van Opijnen, T., Bodi, K. & Camilli, A., 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods,* 6(10), pp. 767-772.

Vidal, J., Chen, J., Li, J. & McClane, B., 2009. Use of an EZ-Tn5-Based Random Mutagenesis System to Identify a Novel Toxin Regulatory Locus in *Clostridium perfringens* Strain 13. *PLOS ONE.*

Walker, D. & Köpke, M., 2015. *Method of Producing a Recombinant Microorganism.* Washington, DC, Patent No. 9315830.

Wang, S. et al., 2013. NADP-Specific Electron-Bifurcating [FeFe]-Hydrogenase in a Functional Complex with Formate Dehydrogenase in *Clostridium autoethanogenum* Grown on CO. *Journal of Bacteriology,* 195(19), pp. 4373-4386.

Wang, T. & Smith, K., 1986. Inviability of dam recA and dam recB cells of *Escherichia coli* is correlated with their inability to repair DNA double-strand breaks produced by mismatch repair. *Journal of Bacteriology,* Mar, 165(3), pp. 1023-1025.

Wasels, F. et al., 2017. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *Journal of Microbiological Methods,* Volume 140, pp. 5-11.

Wielgoss, S. et al., 2011. Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment with *Escherichia coli*. *G3 (Bethesda),* 1(3), pp. 183-186.

Wilding-Steele, T., Ramette, Q., Jacottin, P. & Souicaille, P., 2021. Improved CRISPR/Cas9 Tools for the Rapid Metabolic Engineering of *Clostridium acetobutylicum*. *Inernational Journal of Molecular Sciences,* 22(7), p. 3704.

Wilson, G. & Murray, N., 1991. Restriction and modification systems. *Annual Review of Genetics,* Volume 25, pp. 585-627.

Wirth, S. & Dürre, P., 2021. Investigation of putative genes for the production of medium-chained acids and alcohols in autotrophic acetogenic bacteria. *Metabolic Engineering,* Volume 66, pp. 296-307.

Woods, C. et al., 2019. A novel conjugal donor strain for improved transfer into *Clostridium* spp.. *Anaerobe,* Oct, Volume 59, pp. 184-191.

Woods, C. et al., 2021. Application of transposon-insertion sequencing to determine gene essentiality in the acetogen *Clostridium autoethanogenum*. *bioRxiv.*

Woolston, B., Emerson, D., Currie, D. & Stephanopoulos, G., 2018. Rediverting carbon flux in *Clostridium ljungdahlii* using CRISPR interference (CRISPRi). *Metabolic engineering,* Volume 48, pp. 243-253.

Wrighton, K., 2015. *Microbial controls on biogeochemical cycling in deep subsurface shale carbon reservoirs,* s.l.: JGI Genome Portal.

Xu, L. & Seki, M., 2020. Recent advances in the detection of base modifications using the Nanopore sequencer. *Journal of Human Genetics,* Volume 65, pp. 25-33.

Xu, T. et al., 2015. Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase. *American Society for Microbiology: Applied and Environmental Microbiology,* 81(13), pp. 4423-4431.

Yang, G. et al., 2017. Rapid Generation of Universal Synthetic Promoters for Controlled Gene Expression in Both Gas-Fermenting and Saccharolytic Clostridium Species. *ACS Synthetic Biology,* 6(9), pp. 1672-1678.

Yang, X., Xu, M. & Yang, S., 2016. Restriction modification system analysis and development of in vivo methylation for the transformation of *Clostridium cellulovorans*. *Applied Genetics and Molecular Biotechnology,* Nov, Volume 100, pp. 2289-2299.

Yi, J. et al., 2021. A Heterodimeric Reduced-Ferredoxin-Dependent Methylenetetrahydrofolate Reductase from Syngas-Fermenting *Clostridium ljungdahlii*. *Microbiology Spectrum,* 9(2), pp. e00958-21.

Yu, D. et al., 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the USA,* 97(11), pp. 5978-5983.

Zang, G. et al., 2021. Synthetic Methanol/Fischer-Tropsch Fuel Production Capacity, Cost, and Carbon Intensity Utilizing CO_2 from Industrial and Power Plants in the United States. *Environmental Science and Technology,* 55(11), pp. 7595-7604.

Zhang, G. et al., 2012. A mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria. *PLoS Genetics,* 8(9), p. e1002987.

Zhang, J. et al., 2020. Enhancing plasmid transformation efficiency and enabling CRISPR-Cas9/Cpf1-based genome editing in *Clostridium tyrobutyricum*. *Biotechnology and Bioengineering,* 117(9), pp. 2911-2917.

Zhang, J. et al., 2020. Enhancing plasmid transformation efficiency and enabling CRISPR-Cas9/Cpf1-based genome editing in *Clostridium tyrobutyricum*. *Biotechnology and Bioengineering,* 117(9), pp. 2911-2917.

Zhang, J. et al., 2018. Markerless genome editing in *Clostridium beijerinckii* using the CRISPR-Cpf1 system. *Journal of Biotechnology,* Volume 284, pp. 27-30.

Zhang, J., Liu, Y.-J., Cui, G.-Z. & Cui, Q., 2015. A novel arabinose-inducible genetic operation system developed for *Clostridium cellulolyticum*. *Biotechnology and Biofuels,* Volume 8.

Zhang, J., Taylor, S. & Wang, Y., 2016. Effects of end products on fermentation profiles in *Clostridium carboxidivorans* P7 for syngas fermentation. *Bioresource Technology,* Volume 218, pp. 1055-1063.

Zhang, J. et al., 2018. Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metabolic Engineering,* Volume 47, pp. 49-59.

Zhang, N. et al., 2015. I-SceI-mediated scarless gene modification via allelic exchange in Clostridium. *Journal of Microbiological Methods,* Volume 108, pp. 49-60.

Zhang, X. et al., 2006. *mazF*, a novel counterselectable marker for unmarked chromosomal manipulation in *Bacillus subtilus*. *Nucleic Acids Research,* 34(9).

Zhang, Y., Grosse-Honebrink, A. & Minton, N., 2015. A Universal Mariner Transposon System for Forward Genetic Studies in the Genus *Clostridium*. *PloS One,* 10(4), p. e0122411.

Zhao, R. et al., 2019. CRISPR-Cas12a-Mediated Gene Deletion and Regulation in *Clostridium ljungdahlii* and Its Application in Carbon Flux Redirection in Synthesis Gas Fermentation. *ACS Synthetic Biology,* 8(10), pp. 2270-2279.

Appendix

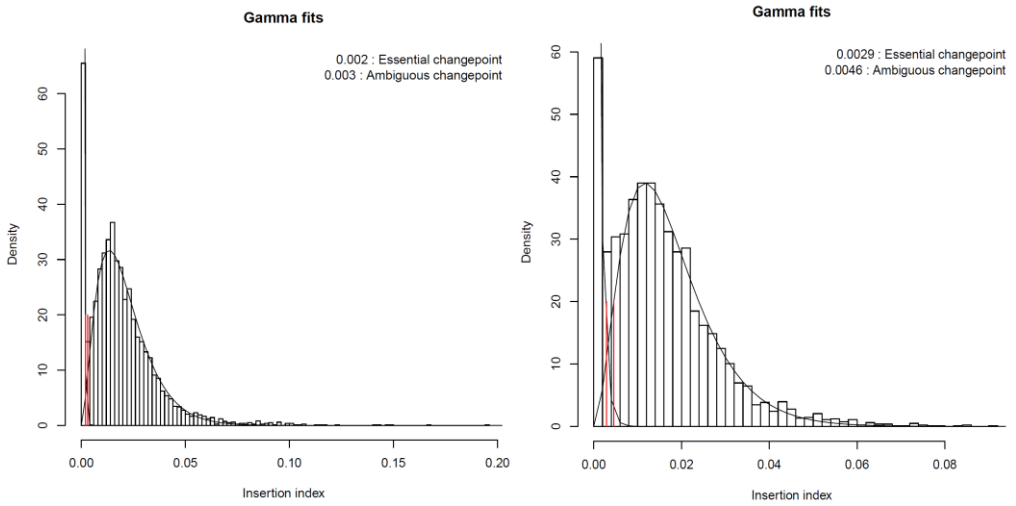


Figure 55 – Bio-TraDIS pipeline mapping statistics, graphing insertion index distribution for essentiality threshold prediction within glucose subculture 3 and methanol subculture 3, respectively.

Table 40 – *E. limosum* NG-6849 genes predicted as essential via the Bio-TraDIS pipeline, through alignment of transposon insertion sequences to the *E. limosum* ATCC 8486 genome *(Song & Cho, 2015)*, on methanol but not on glucose.

Locus Tag (B2M23_)	Read Count	Insertion index	Gene Length	Gene Function
13375	23	0.001005	2211	conjugal transfer protein TraA
19165	5	0.001013	2193	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
08820	9	0.00204	2178	(p)ppGpp synthetase
08725	7	0.002568	2163	penicillin-binding protein
17100	7	0.001603	2079	polyribonucleotide nucleotidyltransferase
13210	264	0.002365	1878	hypothetical protein
19320	6	0.002191	1521	transcriptional regulator MraZ (cell wall/division)
05935	26	0.001549	1434	nicotinate phosphoribosyltransferase
16850	16	0.00241	1383	methanol--corrinoid methyltransferase
03155	15	0.001613	1377	tRNA modification GTPase
04870	7	0.001617	1374	dihydrolipoyl dehydrogenase
17115	3	0.002486	1341	ribosomal protein S12 methylthiotransferase RimO
05955	18	0.00084	1323	nucleotide sugar dehydrogenase
15300	10	0.002542	1311	peptide-binding protein
05590	15	0.002595	1284	phosphomethylpyrimidine synthase

04800	1	0.000886	1254	ATP phosphoribosyltransferase regulatory subunit
14150	6	0.001771	1254	hypothetical protein
00605	5	0.001789	1242	hypothetical protein
14410	9	0.00271	1230	aminopeptidase
15470	5	0.002812	1185	tryptophan synthase subunit beta
15560	42	0.002045	1086	GTP-binding protein YchF
18755	8	0.002116	1050	ATPase
11465	3	0.001073	1035	iron ABC transporter
13570	3	0.002157	1029	phosphoribosylformylglycinamidine cyclo-ligase
07945	328	0.002191	1014	hypothetical protein
01170	1	0.001099	1011	electron transfer flavoprotein subunit alpha
15455	2	0.001099	1011	anthranilate phosphoribosyltransferase
11425	4	0.002222	999	GGGtGRT protein
13205	0	0	996	alpha-ketoacid dehydrogenase subunit beta
08330	1	0.001115	996	HPr(Ser) kinase/phosphatase
19030	0	0	975	recombinase RecJ
11460	2	0.001168	951	ABC transporter substrate-binding protein
07735	18	0.00235	945	aldo/keto reductase
12235	17	0.00235	945	ribosomal protein L11 methyltransferase
19855	3	0.001186	936	tRNA dimethylallyltransferase
01260	2	0.001202	924	two-component sensor histidine kinase
03250	6	0.002484	894	hypothetical protein
04865	0	0	879	5
09995	1	0.001263	879	diaminopimelate epimerase
06300	21	0.002608	852	23S rRNA (guanosine(2251)-2'-O)-methyltransferase RlmB
15645	1	0.001312	846	PHP domain-containing protein
01185	5	0.002646	840	3-hydroxybutyryl-CoA dehydrogenase
16870	8	0.002681	828	RNA polymerase subunit sigma-24
01620	11	0.002721	816	TIGR00268 family protein
15370	10	0.002721	816	sortase
00950	12	0.002837	783	SDR family oxidoreductase
08370	7	0.002882	771	carbon monoxide dehydrogenase
11010	16	0.002882	771	nitroreductase
15110	0	0	768	DeoR family transcriptional regulator
19505	2	0.001445	768	ABC transporter ATP-binding protein
14765	0	0	765	hypothetical protein
01335	8	0.001475	753	hypothetical protein
13065	3	0.001475	753	hypothetical protein
00535	1	0.001499	741	triose-phosphate isomerase
08235	0	0	714	hypothetical protein
01025	1	0.001563	711	ABC transporter

07750	0	0	708	sugar fermentation stimulation protein SfsA
13560	0	0	708	phosphoribosylaminoimidazolesuccinocarboxamide synthase
09290	70	0.001582	702	hypothetical protein
19750	0	0	678	hypothetical protein
03220	0	0	660	ribulose-phosphate 3-epimerase
19175	0	0	660	cytidylate kinase
18720	0	0	627	V-type ATP synthase subunit D
03160	3	0.00177	627	protein jag
17655	0	0	606	hypothetical protein
16835	1	0.001832	606	cobalamin biosynthesis protein CobW
01565	2	0.001898	585	maltose acetyltransferase
14600	25	0.001969	564	peptidase M15
02895	0	0	558	flavoprotein
20670	36	0.002	555	glycerol-3-phosphate responsive antiterminator GlpP
20530	1	0.002045	543	hypothetical protein
01425	3	0.002058	540	hypothetical protein
07940	3	0.002058	540	hypothetical protein
20630	5	0.002066	537	hypothetical protein
08205	1	0.002079	534	ribosomal subunit interface protein
19135	24	0.002079	534	ATP-dependent protease subunit HslV
19225	0	0	528	NUDIX domain-containing protein
18395	3	0.002151	516	trans-hexaprenyltranstransferase
11050	0	0	501	hypothetical protein
10995	33	0.002227	498	hypothetical protein
20185	0	0	492	3-isopropylmalate dehydratase small subunit
19595	10	0.002268	489	N-acetyltransferase
00915	5	0.002283	486	hypothetical protein
11735	2	0.002283	486	hypothetical protein
00695	1	0.002315	480	23S rRNA (pseudouridine(1915)-N(3))-methyltransferase RlmH
13230	44	0.002342	474	transcriptional regulator
05745	2	0.002358	471	30S ribosomal protein S7
09590	1	0.002358	471	hypothetical protein
10970	9	0.002358	471	hypothetical protein
20555	1	0.002358	471	rRNA maturation RNase YbeY
18170	0	0	468	6,7-dimethyl-8-ribityllumazine synthase
08275	6	0.002387	465	anaerobic ribonucleoside-triphosphate reductase activating protein
19055	1	0.002387	465	ribosome maturation factor
13160	1	0.002415	459	tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex ATPase subunit type 1 TsaE
15400	12	0.002433	456	hypothetical protein
19635	3	0.002451	453	hypothetical protein

05950	0	0	450	ribose 5-phosphate isomerase B
15190	0	0	450	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit
18745	0	0	435	ATPase
08485	0	0	426	50S ribosomal protein L11
19300	0	0	423	division/cell wall cluster transcriptional repressor MraZ
00765	1	0.002625	423	hypothetical protein
07645	0	0	417	transcriptional regulator
09860	0	0	408	hypothetical protein
15925	42	0.002778	399	hypothetical protein
12750	5	0.002865	387	MerR family transcriptional regulator
09220	0	0	384	molybdopterin oxidoreductase
11740	0	0	384	hypothetical protein
05215	0	0	375	transcriptional regulator
18290	0	0	375	reactive intermediate/imine deaminase
03805	0	0	360	transcriptional regulator
10070	0	0	348	alkaline-shock protein
08960	0	0	345	hypothetical protein
14520	0	0	342	mobilization protein
00155	0	0	315	hypothetical protein
07745	0	0	303	hypothetical protein
08415	0	0	303	hypothetical protein
18740	0	0	303	ATPase
06465	0	0	297	hypothetical protein
06360	0	0	291	50S ribosomal protein L23
18060	0	0	270	hypothetical protein
20770	0	0	261	hypothetical protein
16945	0	0	258	hypothetical protein
03575	0	0	243	hypothetical protein
09400	0	0	240	hypothetical protein
17065	0	0	234	hypothetical protein
01455	0	0	231	hypothetical protein
14480	0	0	231	RNA-binding protein
03655	0	0	228	hypothetical protein
08430	0	0	219	hypothetical protein
18630	0	0	210	hypothetical protein
05225	0	0	207	hypothetical protein
00730	0	0	204	hypothetical protein
10310	0	0	204	transcriptional regulator
08420	0	0	201	RNA-binding protein
14700	0	0	201	hypothetical protein
06415	0	0	186	30S ribosomal protein S14
09195	0	0	186	pseudogene
14580	0	0	186	hypothetical protein
19810	0	0	156	pseudogene
14375	0	0	117	

08445	0	0	77
10395	0	0	77
10390	0	0	76
10405	0	0	75
12665	0	0	75
05245	0	0	66

Table 41 – progressiveMauve identified Gap regions between the *E. limosum* NG-6849 genome consensus determined within this study, and the published *E. limosum* ATCC 8486 genome *(Song & Cho, 2015)*.

Genome	Position	Length
E.limNG6849_consensus	74022	2900
E.limNG6849_consensus	694245	1
E.limNG6849_consensus	864807	103
E.limNG6849_consensus	864821	1
E.limNG6849_consensus	865023	121
E.limNG6849_consensus	1298055	2
E.limNG6849_consensus	1442402	4
E.limNG6849_consensus	1701638	6
E.limNG6849_consensus	1701675	6
E.limNG6849_consensus	1701747	6
E.limNG6849_consensus	1701783	4
E.limNG6849_consensus	1701825	430
E.limNG6849_consensus	1702240	7
E.limNG6849_consensus	1702311	2
E.limNG6849_consensus	1702380	372
E.limNG6849_consensus	1702908	2
E.limNG6849_consensus	1702942	2
E.limNG6849_consensus	1702953	1
E.limNG6849_consensus	1703037	1
E.limNG6849_consensus	1703051	7
E.limNG6849_consensus	1703062	3
E.limNG6849_consensus	1727594	1662
E.limNG6849_consensus	3264446	1
E.limNG6849_consensus	3670813	2
E.limNG6849_consensus	4395297	62
E.limNG6849_consensus	4406928	10427
E.limNG6849_consensus	1727599	4
E.limNG6849_consensus	1703015	6
E.limNG6849_consensus	1702972	6
E.limNG6849_consensus	1702899	6
E.limNG6849_consensus	1702857	4
E.limNG6849_consensus	1702812	431
E.limNG6849_consensus	1702337	7
E.limNG6849_consensus	1702261	2

E.limNG6849_consensus	1702197	371
E.limNG6849_consensus	1701738	2
E.limNG6849_consensus	1701702	2
E.limNG6849_consensus	1701689	1
E.limNG6849_consensus	1701616	1
E.limNG6849_consensus	1701601	7
E.limNG6849_consensus	1701583	3
E.limNG6849_consensus	1442434	31
E.limNG6849_consensus	1298056	1241
E.limNG6849_consensus	4406826	1
E.limNG6849_consensus	4406888	1
E.limNG6849_consensus	4406918	1
E.limNG6849_consensus	4395297	38
E.limNG6849_consensus	71052	1
E.limNG6849_consensus	71063	2959
E.limNG6849_consensus	864686	121
E.limNG6849_consensus	864905	1
E.limNG6849_consensus	864920	103
E.limNG6849_consensus	6250	29