Metabolic Engineering of *Clostridium acetobutylicum* for Gene Essentiality Studies

Cynthia Kelechi Akaluka, MSc

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

August, 2021



Declaration

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

Cynthia Kelechi Akaluka

ABSTRACT

In recent years the need to generate chemicals and fuels by sustainable routes has led to a resurgence of interest in exploiting butanol-producing species *Clostridium acetobutylicum*. It has a longstanding history in the commercial production of solvents in the so-called Acetone-Butanol-Ethanol (ABE) process. However, it has not operated on a commercial scale for several decades. A better understanding of the physiological factors that affect solvent production could have profound effects on the re-emergence of the ABE fermentation process on an industrial scale.

One way to identify those processes of importance in butanol production is through forward genetics using transposon mutagenesis and in particular high-throughput approaches such as Transposon Directed Insertion-site Sequencing (TraDIS). The latter enables the identification of essential genes under niche-specific conditions through the simultaneous sequencing of large transposon mutant libraries. Its application in *C. acetobutylicum* is hindered by poor transformation frequencies. To circumvent this barrier the two endogenous restriction modification systems, *CA_C1502* and *CA_C3535*, were inactivated using allele-coupled exchange (ACE) and a suicide plasmid in a Δ*pyrE* mutant background of *C. acetobutylicum*. Their inactivation led to an almost 6-fold increase in transformation frequency compared to the parent strain.

Though the improvements in transformation efficiency did not allow the suicide delivery, the use of a replicative conditional vector was demonstrated and a preliminary TraDIS library generated from which a list of 537 potentially essential genes was identified. During implementation of TraDIS, a prophage present in the *C. acetobutylicum* genome was shown to be active and capable of excision. As phages can have applications as genetic tools, the phage was further investigated. This study showed that the phage's DNA can be integrated into the chromosome, circularised in the cytoplasm as well as form pseudo-temperate plasmid. The phage was induced with mitomycin C and UV light. Electron microscope characterisation revealed that it has an elongated icosahedral capsid with a long non-contractile tail indicating that it is from the family *Siphoviridae*. Transduction test and neighbour-joining cladogram study suggest that it is a non-transducing phage. The phage integrase was explored as an integrative tool for application in synthetic biology. This is the first published genetic study of a temperate phage in *C. acetobutylicum*.

To my Dad….

Thanks for giving me the best gift ever

Acknowledgements

I want to thank to my supervisors, Nigel P. Minton and Ying Zhang for their advice, support and guidance throughout my research. Most especially thank you for making it comfortable and conduicive to work in the laboratory during my pregnancies. My sincere gratitude goes to the Vice Chancellor's International Scholarship for Research Excellence funded by the University of Nottingham for sponsoring my tuition and the Petroleum Technology Development Fund (PTDF) for providing my stipends.

I want to appreciate Hengzheng for providing bench support during the initial stage of my research. I also want to thank Bunmi Omorotionmwan for the advice and moral support. She never turned down any of my requests, for that I am very grateful. I want to thank our hard-working Lab Technican, Evangelos Karvonis, who was willing to go the extra mile for media preparations and Senior Scientist Technician, Denise McLean, for the beautiful TEM phage images. I also appreciate Matt Abbott for HPLC and GC analysis. I want to thank Craig Woods for his advice and support with my TraDIS work. I cannot forget Bashir Rumah, David Ortega, Raquel Rodrigues, Simona Pentassuglia, Cansu Karyal, Temiloluwa Daike and all my friends and colleagues at Synthetic Biology Research Centre Nottingham, University of Nottingham, UK for their stimulating discussions and for making research fun and interesting.

My sincere thanks goes to my husband, Chiedoziem, for his patience and precious encouragements through out my studies. To my lovely son, Chiazo, your smile is everything, writing this thesis while you stumped, jumped, spinned and rolled had its challenges in the COVID-19 pandemic lockdown but it is the rhythm I needed to bring me down to earth. To my little baby, Ikechi, I deeply cherish the intimate journey of your growth in my womb and birth while writing this thesis. To my Dad and siblings, Uche and Chibueze, I am grateful for your spiritual and emotional support. To the Aganas, without your support, this work would have been more difficult. To my in-laws, Nzeh and Lolo Bon Uwazie Inechi, thank you for all your prayers and support. Ka Chukwu gozie unu nilee!! Ya gazie, nna, nne, nwa, nwanne and enyi m nilee. Onye obula guurum ofoma, ka Chukwu nye unu ihe obula bu ebumnuche unu ugbua na ubochi nilee n'ndu unu. Ese!

Table of Contents

Declaration…….....ii
Abstract…….iii
Acknowledgements…….....v
Table of Contents…….....vi
List of Figures…….……..xiii
List of Tables…….....xvii
List of Abbreviations …….……..xviii
CHAPTER 1 1
Introduction ……...2
1.1 Sources of Energy ……2
1.2 Challenges Created by Depending on Fossil Fuels…….. 3
1.2.1 Environmental Pollution……... 4
1.2.2 Effects of Climate Change…… 4
1.2.3 Fluctuating Crude Oil Prices ……5
1.2.4 Depleting Oil Reserves…… 5
1.2.5 Ever-growing World Energy Demand…… 6
1.3 Biofuels as an Alternative Source of Energy…… 7
1.3.1 Butanol as an Energy Source…… 7
1.3.2 Chemical Synthesis of Butanol……8
1.3.3 History of the Acetone-Butanol-Ethanol Fermentation Pathway……. 8
1.4 The *Clostridium* Genus…… 10
1.4.1 *Clostridium acetobutylicum* ATCC 824…….. 11
1.4.2 Metabolic Pathway for Butanol Production in *C. acetobutylicum*….. 12
1.5 Molecular Biological Tools Developed for Manipulating Genes
in *Clostridia*……... 14
1.5.1 ACE Technology…….. 15
1.5.2 ClosTron …….. 17
1.5.3 CRISPR/Cas Systems …… 19
1.5.4 Transposons …… 21
1.5.5 Phage-mediated Gene Transfer …….. 23
1.6 Background of Project …….. 23
1.7 Aim of Project …….…… 24

CHAPTER 2 ……..... 26
2.1 Microbiology ……..……. 27
2.1.1 Bacterial Strains …….. 27
2.1.2 Media, Buffers and Supplements..…….. 28
2.1.2.1 Luria Bertani (LB) Medium …… 29
2.1.2.2 2x Yeast Tryptone Glucose (2xYTG) Broth ……..... 29
2.1.2.3 *Clostridium* Growth Medium (CGM) …… 29
2.1.2.4 *Clostridium* Basal Medium (CBM) …….. 29
2.1.2.5 P2 Minimal Medium ……... 30
2.1.2.6 Electroporation Buffer (EPB) …….. 30
2.1.2.7 TAE Buffer …… 30
2.1.3 Conditions for Culture Storage ……..... 31
2.1.4 Bacterial Growth Quantification ……... 32
2.1.5 Electro-competent Cells ……..……..... 32
2.1.5.1 Preparation of Electro-competent *E.coli* Cells ……. 32
2.1.5.2 Preparation of Electro-competent *C. acetobutylicum* ATCC 824 Δ*pyrE* Mutant Cells…… 32
2.1.6 Electroporation ……... 33
2.1.6.1 Electroporation of Electro-competent *E.coli* Cells ……... 33
2.1.6.2 Electroporation of Electro-competent *C. acetobutylicum* ATCC 824 With Plasmid 33
2.1.7 TraDIS Procedure ……...…….…... 34
2.1.8 Improved Phage Induction Protocol ……...… 35
2.1.9 Plaque Assay ……...…….. 35
2.1.10 Characterisation of Bacteriophages by Transmission Electron Microscope …….....…… 36
2.2 Molecular Biological Techniques ……..……. 37
2.2.1 Primers …… 37
2.2.2 Chromosomal and Plasmid DNA Extraction and Purification …… 39
2.2.3 Polymerase Chain Reaction (PCR) ……..39
2.2.4 Inverse PCR ……. 39
2.2.5 Agarose Gel Electrophoresis ……41
2.2.6 DNA Extraction and Purification From Gel …….....41
2.2.7 DNA Sanger Sequencing ……..... 41

2.2.8 DNA Quantification …….. 42
2.2.9 Cloning Techniques ……... 42
2.2.9.1 Restriction Enzymes ……... 42
2.2.9.2 Ligation Enzyme …….. 42
2.2.9.3 Membrane Dialysis …….. 42
2.2.9.4 Illumina Sequencing ……. 42
2.2.10 Plasmid Constructs and Genomic Mutagenesis of *C. acetobutylicum* …… 43
2.2.10.1 Suicide Knock-out Plasmid …… 44
2.2.10.2 R702 pAN2 Plasmid for Methylation …… 45
2.2.10.3 Allele-Coupled Exchange *pyrE* Repair Plasmid ……... 46
2.2.10.4 CRISPR/Cas9 Gene Deletion Plasmid …….. 46
2.2.10.5 Allele-Coupled Exchange *pyrE* KO Plasmid …… 47
2.2.10.6 Phage Integrase Recombination Plasmid ……... 47
2.3 Solvent Analysis ……. 48
2.3.1 High Performance Liquid Chromatography (HPLC) ……..... 48
2.3.2 Gas Chromatography (GC) ……... 49
2.4 Bioinformatics Tools ……. 49
2.4.1 Primer Design ……. 49
2.4.2 Plasmid Map Construction ……..... 49
2.4.3 Illumina Sequencing Data Analysis …….. 49
2.5.4 Database Search ……. 50
2.5.5 Data Analysis and Visualisation ……..... 50
2.5.6 Library Preparation Process for TraDIS Sequencing 50
CHAPTER 3 ……..……..... 53
Improving the Transformation by Countering Restriction Modification Systems .. 53
3.1 Introduction …….. 54
3.1.1 Restriction Modification Systems in *Clostridium acetobutylicum* … 55
3.1.2 Methylation of Exogenous DNA for Transformation of *C. acetobutylicum* …… 57
3.1.3 Knock-out Restriction Modification Systems …… 58

3.1.4 Previous Studies on the KO of Restriction Modification System in *C. acetobutylicum* …… 58
3.2 Hypothesis and Aim of Chapter ……..… 59
3.3 Results and Discussion……... 60
3.3.1 Creation of Single Crossovers for *CA_C1502* KO in *pyrE* Mutant …… 60
3.3.2 Creation of Double Crossover Mutants for *CA_C1502* KO…….. 62
3.3.3 Creation of a Double Crossovers *CA_C3535* KO in the *CA_C1502* Mutant Strain ……... 66
3.3.4 Creation of a Double Crossovers of *CA_C3535* KO in the Δ*pyrE* Strain …..... 67
3.3.5 Determination of the Transformation Efficiency of all Mutants Created ……... 68
3.3.6 Comparing the Methylation System in R702 With TOP10 pAN2 Methylation System …….. 72
3.3.7 Electrotransformation Efficiency of Methylation Donor Strains Carrying R702::*Φ3T* or pAN2 ……. 73
3.4 Conclusion ……... 75
CHAPTER 4 …..……..…… 77
Investigating Essential and Non-essential Genes Using TraDIS ……. 77
4.1 Introduction …… 78
4.1.1 Transposon Mutagenesis Versus Chemical Mutagenesis …… 78
4.1.2 Transposon Mutagenesis Studies …… 79
4.1.3 Previous Studies on Gene Essentiality and Conditional Studies Using TraDIS ……. 82
4.1.4 Butanol Tolerance Investigations With *C. acetobutylicum* ……. 85
4.1.5 Sodium Ion Tolerance Investigations in *C. acetobutylicum* …… 86
4.1.6 Suicide Versus Conditional Transposon Plasmid ……... 88
4.1.7 Inducible Promoter System in *C. acetobutylicum* 89
4.2 Hypothesis and Aim of Chapter ……... 91
4.3 Results and Discussion …….. 91
4.3.1 Screening for *tcdR* Insert *pyrE* Repair in Single and Double Restriction-less Mutants …….. 91
4.3.2 Design of TraDIS Experiment …… 92

4.3.3 Selecting an Ideal TraDIS Strain which does not have SNPs or Deletions 93
4.3.4 Selection of the Transposon Plasmid that Functions Effectively and Efficiently in *C. acetobutylicum* 95
4.3.4.1 Transposon Suicide Plasmid 95
4.3.4.2 Transposon Conditional Replicative Plasmid, pMTL-CW20 ... 96
4.3.4.3 Transposon Conditional Replicative Plasmid, pMTL-CW21... 98
4.3.5 Determine the Sub-Inhibitory Concentration of Butanol and Sodium Ion for TraDIS 100
4.3.6 Work Flow of the Transposon Library Generation Procedure 104
4.3.7 Library Preparation Process for TraDIS Sequencing 105
4.3.8 First Attempt at TraDIS..... 106
4.3.8.1 Transposon Library Creation-1st Attempt 106
4.3.8.2 Sequencing Data for pMTL-CW20-1st Attempt 106
4.3.8.3 Troubleshooting the Plasmid Loss Challenge (1)..... 109
4.3.9 Second Attempt at TraDIS 111
4.3.9.1 Transposon Library Creation -2nd Attempt 111
4.3.9.2 Sequencing Data for pMTL-CW21- 2nd Attempt..... 111
4.3.9.3 Troubleshooting the Plasmid Read Challenge (2)..... 112
4.3.9.4 Probable Reasons why There Were Plasmid Reads in TraDIS Data 113
4.3.9.4.1 Loss of Activity of Erythromycin 113
4.3.9.4.2 Poor Digestion of I-SceI 114
4.3.9.4.3 Low Transposition Frequency 114
4.3.9.4.4 Plasmid Integration 116
4.3.9.4.5 Problems with Plasmid or Bacteria Strain Used 117
4.3.10 Third Attempt at TraDIS 117
4.3.10.1 Transposon Library Creation - 3rd Attempt 117
4.3.10.2 Analysis of Inverse PCR Products for 3rd Transposon Library 118
4.3.10.3 Growing Transposon Pool in Challenging Conditions 120
4.3.11 Analysis of Sequenced TraDIS Data 121
4.4 Conclusion 126
CHAPTER 5 130

Isolation and Characterisation of Lysogenic Phage …… 130
5.1 Introduction …….. 131
5.1.1 Life Cycle of a Phage …… 132
5.1.2 Classification of Bacteriophages According to Their Morphology and Genetic Materials …… 134
5.1.3 Conditions That Affect Phage Infecting Host Cells …… 135
5.1.4 Using Phage Integrases for Site-specific Recombination Technology 138
5.1.5 Phage-host Interactions and Evolutionary Conclusions ……. 140
5.1.6 Exploring Phage Studies for the Development of Industrial Strains .. 142
5.1.7 Previous Studies of Phage in *Clostridium acetobutylicum* and Other Saccharolytic Bacteria ……. 142
5.2 Hypothesis and Aim of This Chapter ……..……. 144
5.3 Results and Discussion …….. 144
5.3.1 How Phage Activity Was Discovered …… 144
5.3.2 Is the Insertion of the *tcdR* Gene Responsible for Inducing Phage ACK1? 146
5.3.3 Gene Map of Phage ACK1 of *Clostridium acetobutylicum* ATCC 824. 149
5.3.3.1 Phage DNA Replication and Transcription Regulators ...… 150
5.3.3.2 Phage Lysis Genes …….. 151
5.3.3.3 Phage Structural Genes …….. 152
5.3.3.4 Phage Packaging Genes …… 152
5.3.3.5 Phage Integrase ……. 153
5.3.4 Site-specific Recombination Region in *C. acetobutylicum* ATCC 824 Phage ACK1 ……..... 154
5.3.5 Phage-host Relationship of *Clostridium* Phage ACK1 in *C. acetobutylicum* ATCC 824 …….. 158
5.3.6 Improving on the Protocol for Phage Induction and Plaque Assay …. 161
5.3.7 Characterisation of Phage ACK1 ……163
5.3.8 Phage ACK1: Transducing or Non-transducing Phage? ……..... 165
5.3.9 Prediction of ACK1 Packaging Using Multiple Sequence Alignment Comparison by Log-extraction (MUSCLE) ……. 167
5.3.10 Development of a Plasmid-based Integration System Utilising the Site-Specific Recombinase and *attP* Locus in Phage ACK1……. 169
5.4 Conclusion ……. 175

CHAPTER 6 ……..……..……..……..……. 179
Conclusion and Further Work…….…….. 180
6.1 Improving the Transformation Efficiency of *C. acetobutylicum* …….. 180
6.2 Investigating the Essential Genes for Survival of *C. acetobutylicum* ……. 181
6.3 Isolation and Characterisation of Lysogenic Phage ……. 181
References …… 183
Appendix …… 202
Appendix A: Supplementary Information for Chapter 4 ……. 202
Appendix B: Supplementary Information for Chapter 5 ……. 214

List of Figures

Figure 1.1: Factors encouraging the search for alternative sources of energy to fossil fuels ……... 3

Figure 1.2: Historical chart by West Texas Intermediate (WTI) of crude oil prices per barrel from 1946 to 2020 ……..…… 5

Figure 1.3: Metabolic pathways of glucose metabolism in *C. acetobutylicum* showing genes encoding each metabolic enzyme ……. 14

Figure 1.4: *pyrE* mutant generation using Allele-Coupled Exchange ……. 16

Figure 1.5: Schematic illustration of ClosTron mutant generation ……. 18

Figure 1.6: CRISPR/Cas9 gene editing …….. 20

Figure 1.7: Illustration of cut-and-paste transposition process …… 22

Figure 2.1: The allelic- exchange suicide restriction modification system knock-out plasmid 45

Figure 2.2: Representation of the methylation cassette in R702 pAN2 plasmid ….. 46

Figure 2.3: The pMTL85141 plasmid backbone for phage integrase vector cloning 48

Figure 2.4: Schematic illustrating workflow for transposon mutant gDNA library preparation 51

Figure 3.1: *Clostridium acetobutylicum* ATCC 824 restriction endonucleases and methylase genes …… 56

Figure 3.2: Schematics of type II restriction modification systems in *C. acetobutylicum* ATCC 824 from Restriction Enzyme Database (REBASE) …… 57

Figure 3.3: PCR screening of 10 single crossover colonies for integration of plasmid pMTL-CAC1502.KO …….. 62

Figure 3.4: PCR analysis revealing reversion of single crossover pMTL-CAC1502.KO plasmid integrants back to wild type ……..…. 64

Figure 3.5: Schematics of right hand single crossover integration and double crossover with deletion of *CA_C1502* gene ……..… 65

Figure 3.6: Isolation of mutants with in-frame deletion of Cac824I……... 66

Figure 3.7: PCR screening for occurrence of *CA_C3535* mutant in 9 thiamphenicol sensitive colonies ……..……... 67

Figure 3.8: PCR screening of putative *CA_C3535* deletions generated in the Δ*pyrE* strain.…….. 68

Figure 3.9: All single and double restriction-less mutants of the *C. acetobutylicum* Δ*pyrE* strain created by in-frame deletion using suicide plasmid ……. 68

Figure 3.10: Transformation efficiency of *C. acetobutylicum* RM mutants as number of transformants/µg plasmid …… 70

Figure 3.11: PCR screening of TOP10 colonies for R702 methylase using primers 338 and 244 ……..... 73

Figure 3.12: Gel image showing SatI (Fnu4HI) treatment of R702::Φ3T protected pMTL85141, pAN2 protected pMTL85141 and unmethylated pMTL85141….....74

Figure 3.13: Electrotransformation efficiency of plasmids prepared in *E. coli* TOP10 methylation donors carrying R702::*Φ3T* or pAN2……75

Figure 4.1: Illustration of the mechanism for cut-and-paste transposition …….. 81

Figure 4.2: Illustration of TraDIS strategy for investigating gene essentiality and conditional gene studies …….. 84

Figure 4.3: Mode of transmission of inducible *Himar1* C9 transposase in pMTL-YZ14……..90

Figure 4.4: PCR screening of *pyrE* repair and *tcdR* insertion in *C. acetobutylicum* Δ*CA_C1502* using primers JH141/F and JH141/R ……. 92

Figure 4.5: Fermentation profile analysis for wild type and restriction-less *tcdR* strains ……..... 94

Figure 4.6: Map of conditional replicative vector pMTL-CW20……..... 97

Figure 4.7: Map of conditional replicative vector pMTL-CW21 ……..... 99

Figure 4.8: Growth of *C. acetobutylicum* ATCC 824 in the presence of various concentrations of butanol …….....……..... 101

Figure 4.9: Growth of *C. acetobutylicum* ATCC 824 at various concentrations of sodium chloride 103

Figure 4.10: Schematic illustration of TraDIS library construction, indicating areas modified to improve library creation ……..... 104

Figure 4.11: Transposon-tagged maps and unique insertion gene frequency for first attempt pMTL-CW20 TraDIS library sequencing ……. 108

Figure 4.12: Plasmid loss efficiency of pMTL-CW20 in *C. acetobutylicum* using varying concentrations of IPTG …….....……..... 110

Figure 4.13: Trimmed sequencing plasmid reads mapping to the pMTL-CW21 transposon delivery plasmid …… 113

Figure 4.14: Inverse PCR fragments for 24 randomly selected transposon mutant colonies …….....115

Figure 4.15: PCR screening reveals no insertion at acetyl-CoA transferase gene locus …..... 116

Figure 4.16: Inverse PCR screening obtained for 25 transposon mutant colonies with random insertions …….....…….....….. 118

Figure 4.17: Frequency distribution plot of gene insertion index for the summary sets of TraDIS pool ……. 122

Figure 5.1: Phage attacks in the ABE industries in various countries ……. 132

Figure 5.2: Classification of phages according to their genetic material and morphology 135

Figure 5.3: Sequencing coverage of phage-related region in different *C. acetobutylicum* mutants …….....…….....…… 145

Figure 5.4: Position and location of primers used for phage screening.....…….. 146

Figure 5.5: Twenty-five *tcdR*Δ*CAC1502* mutants from 5 Independent transformations screened for the presence of lysogenic phage …….....…… 147

Figure 5.6: Screening of *C. acetobutylicum* Δ*phe*Δ*arg*Δ*pyrE* strains revealed the presence of the lysogenic phage ACK1 in all independent isolates .....…… 148

Figure 5.7: Gene map of *Clostridium* phage ACK1 phage drawn to scale …..... 150

Figure 5.8: Conserved protein domains identified in *CA_C1951* recombinase by NCBI BLASTx search .....…….....……. 154

Figure 5.9: Alignment of trimmed mutant reads covering the phage integration region in a *tcdR*Δ*CAC1502* phage-free strain …….....…… 155

Figure 5.10: Alignment of trimmed *pyrE* mutant reads covering the phage integration region in a lysogen strain .....…… 156

Figure 5.11: Screening the attachment sites in phage-free and phage-containing *C. acetobutylicum* strains with Cac1866F and Cac1951R2……...….. 156

Figure 5.12: Putative site-specific recombination model for *Clostridium* phage ACK1 recombinase ……. 157

Figure 5.13: PCR screening for circularised phage ACK1 genomes in different *C. acetobutylicum* host …….....……. 159

Figure 5.14: The different lifestyles of phage ACK1 in *C. acetobutylicum* ATCC 824 …….....…….. 160

Figure 5.15: Phage ACK1 plaques ……... 163

Figure 5.16: Transmission electron microscope images of phage ACK1 ……. 164

Figure 5.17: Demonstration of transduction test technique …… 166

Figure 5.18: Neighbour-joining cladogram tree of the large terminase subunit of ACK1 Phage of *C. acetobutylicum* compared with 54 phages with known DNA termini. ……..... 169

Figure 5.19: Primer locations for phage integrase cloning …...…… 170

Figure 5.20: Plasmid map of integrase suicide plasmid pMTL85141-51-50 …….. 171

Figure 5.21: Plasmid map for integrase suicide plasmid pMTL85141-67-50 ……. 172

Figure 5.22: Plasmid map of integrase suicide plasmid pMTL85141-68-50 …….. 173

Figure 5.23: PCR screening of *C. acetobutylicum* transformants for recombination of pMTL85141-67-50 and pMTL85141-68-50 …….....…….. 174

List of Tables

Table 1.1: Comparison of butanol, ethanol, gasoline and diesel …….. 8

Table 2.1: List of bacterial strains, their description and sources …… 27

Table 2.2: List of supplements and antibiotics added to media ……31

Table 2.3: List of primers …… 37

Table 2.4: Inverse PCR touchdown details for screening transposon mutants 41

Table 2.5: List of plasmids ……. 43

Table 3.1: Expected DNA fragment sizes for right and left single crossover integrants……61

Table 3.2: Transformation efficiencies for *C. acetobutylicum* RM mutants in cfu/ µg plasmid DNA (10^3).…….. 70

Table 4.1: TraDIS pMTL-CW20 library sequencing result for 1st attempt ……..... 106

Table 4.2: TraDIS pMTL-CW21 library sequencing result for 2nd attempt…… 112

Table 4.3: Transposon pMTL-CW21 library creation result for 3rd attempt. …… 117

Table 4.4: Insertion positions of transposons in the *C. acetobutylicum* wild type chromosome ……..……...…….....……..... 119

Table 4.5: Summary of the mapping statistics of TraDIS library MiSeq sequencing data …….... 121

Table 4.6: KEGG functional categories for candidate essential genes of *C. acetobutylicum* ATCC 824 ……...……. 124

Table 7.1: Canditates for *C. acetobutylicum* ATCC 824 essential gene list …… 193

Table 7.2: *C. acetobutylicum* ATCC 824 bacteriophage ACK1 gene list …..... 205

List of Abbreviations

2xYTG	Two-times Yeast-Tryptone-Glucose Medium
5-FC	5-Fluorocytosine
5-FU	5-Fluorouracil
5-FOA	5-Fluorooratic acid
ABE	Acetone Butanol and Ethanol
ACE	Allele-Coupled Exchange
ATCC	American Type Culture Colection
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
Cas	CRISPR-associated
cfu/ml	Colony forming units per millilitre
CGM	Clostridial Growth Medium
CRISPR	Clustered Regularly Interspaced Palindromic Repeats
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetraacetic acid
EPB	Electropora\tion buffer
Erm	Erythromycin
g	Gram
kb	Kilobase pair
KO	Knock-out
l	Litre
LB	Luria-Bertani medium
LHA	Left homology arm
µl	Microlitre
mg	Milligram
ml	Millilitre
mM	Millimolar
NADH	Nicotinamide adenine dinuceotide, reduced
NEB	New England Biolabs
NGS	Next Generation Sequencing
OD	Opitical Density
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
Pfu/ml	Plaque forming unit per millilitre
RHA	Right homology arm
rpm	Revolutions per minute
SNP	Single Nucleotide Polymorphism
SOE	Splicing by Overlap Extension
Tm	Thiamphenicol
Ur	Uracil
UV	Ultraviolet
V	Volt
w/v	Weight per volume
WT	Wild type
µg	Microgram
%	Percentage

°C	Degree Celsius

Chapter 1

General Introduction

Chapter 1

1.0 Introduction

1.1 Sources of Energy

According to the International Energy Agency (2016), billions of people around the world lack access to basic energy services. The importance of energy in contemporary living cannot be overemphasized as several domestic actions such as cooking, illuminating rooms, boiling water, washing clothes, charging gadgets and engaging with electronic media all require energy. Energy is also required for adequate provision of healthcare, sanitation, clean water, telecommunication and transportation (International Energy Agency, 2016). Without access to energy, many contemporary jobs would not exist. Modern humanity has come to depend so much on the abundance, convenience and potential of energy (Coyle and Simmons, 2014).

For a long time, conventional energy sources, such as coal, petroleum, natural gas, nuclear reactors, wood and biomass have been used worldwide (Hina *et al*., 2014). The main sources of energy frequently used in the world today are fossil fuels such as crude oil, coal, and gas. Fossil fuel is the combustible organic matter obtained from the remains (or fossils) of former life forms like plants, animals and microorganisms (Salameh, 2014). While these forms of energy are important, there are many reasons why these sources of energy are not efficient and sufficient to meet the increasing demand per individual in our population.

1.2 Challenges Created by Depending on Fossil Fuels

There is increasing interest in alternative sources of energy other than fossil fuels for the following reasons: environmental pollution, effects of climatic change, fluctuation of crude oil prices, depleting oil reserves and ever-growing world energy demand (as illustrated in Fig 1.1 and explained in sections 1.2.1 to 1.2.5)

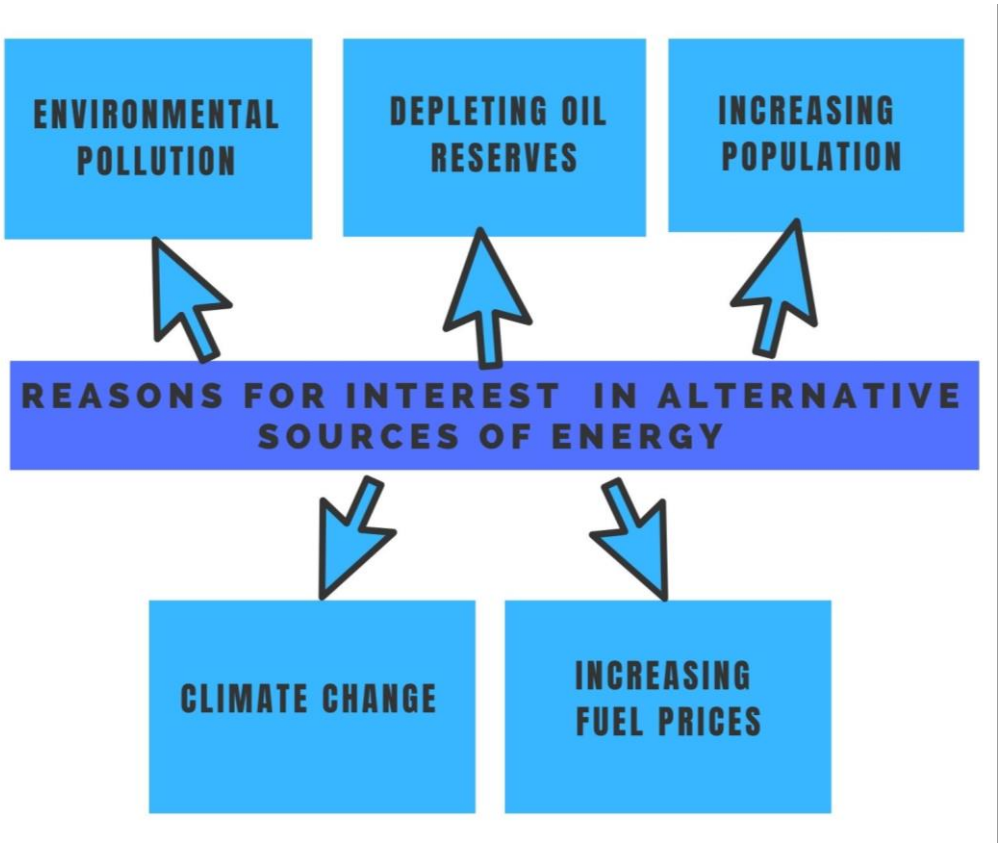

Figure 1.1: Factors encouraging the search for alternative sources of energy to fossil fuels

1.2.1 Environmental Pollution

Conventional energy sources pollute the environment with solid, liquid and gaseous waste which can lead to adverse health problems on living things. In 2010, household air pollution was responsible for 3.9 million premature deaths and approximately 4.8% of lost healthy life years, thus ranking it the highest among environmental risk factors investigated (Smith *et al*., 2014). According to the Centre for Research on Energy and Clean Air (CREA) report in 2020, high air pollution is an important risk factor that increases the probability of one contracting SARS-CoV-2, falling ill, needing intensive care, requiring mechanical ventilator and dying from it (Myllyvirta and Thieriot, 2020). During the COVID-19 quarantine in the first quarter of 2020, there was an unprecedented decline in CO_2 emissions to almost twice as large as the combination of all reductions since the end of World War II (International Energy Agency, 2020). Although there were reports of 30% reduction in air pollution in Italy, USA, Wuhan, Brazil and Spain during the COVID-19 pandemic lockdown (Urrutia-Pereira *et al*.,

2020), there is a high probability that it will get back up as soon as the restrictions and curtailment on mobility are lifted.

Nuclear energy is an alternative source of energy to fossil fuels. One disadvantage of using radioactive substances as a source of energy is that it causes cancer in humans (Hina *et al*., 2014). Improper operation, waste management and maintenance of nuclear power plants can lead to accidents which have detriment consequences on people and the environment. Examples of such radioactive disasters are those that happened in Fukushima Daiichi in 2011 and Chernobyl in 1986 (Hina *et al*., 2014).

During the procurement of oil, spills can happen and cause extensive damage to the environment. In April, 2014, one of the most disastrous oil spills in US history happened on the Deep Horizon rig, the underground well blew up which led to the discharge of 4.9 million barrels of oil into the ocean (Blackburn *et al*., 2014). This is still having an adverse effect on marine plants and animals. Additionally, ground-level ozone pollution mainly caused by motor vehicles, emits 52% reactive hydrocarbon and 72% NO_x (Salameh, 2014).

1.2.2 Effects of Climate Change

Climate change caused by increasing concentrations of greenhouse gases (GHGs) in the atmosphere can contribute to natural disasters that lead to the destruction of valuable assets: prevalence of diseases and pests problems due to droughts, floods and heat shock; poor yield in crops as a result of reduced rainfalls; and rises in prices of food affecting food availability. All of these impede poverty reduction (Hallegatte *et al.*, 2016).

1.2.3 Fluctuating Crude Oil Prices

Coal was the most prominent source of energy in the world until the mid-1950s, when crude oil took over from coal. Approximately 90% of motor vehicles in the world today consume products of crude oil (Salameh, 2014). Since the inception of consuming oil and its derivatives, the prices of oil have continued to fluctuate. Generally, the prices do not low for a long time, they keep going up and up (Salameh, 2014). Figure 1.2 shows the 1946-2020 historical chart by West Texas Intermediate (WTI) of crude oil prices per barrel. Crude oil prices dropped to an historical time low of $18.93 per barrel during the COVID-19 pandemic in April 2020, but prices continued rising sharply to

$41.22 in August 2020 (Macrotrends, 2020). The high/fluctuating price of crude oil is one of the driving forces behind the quest for alternative sources of energy (Bankar *et al*., 2013).

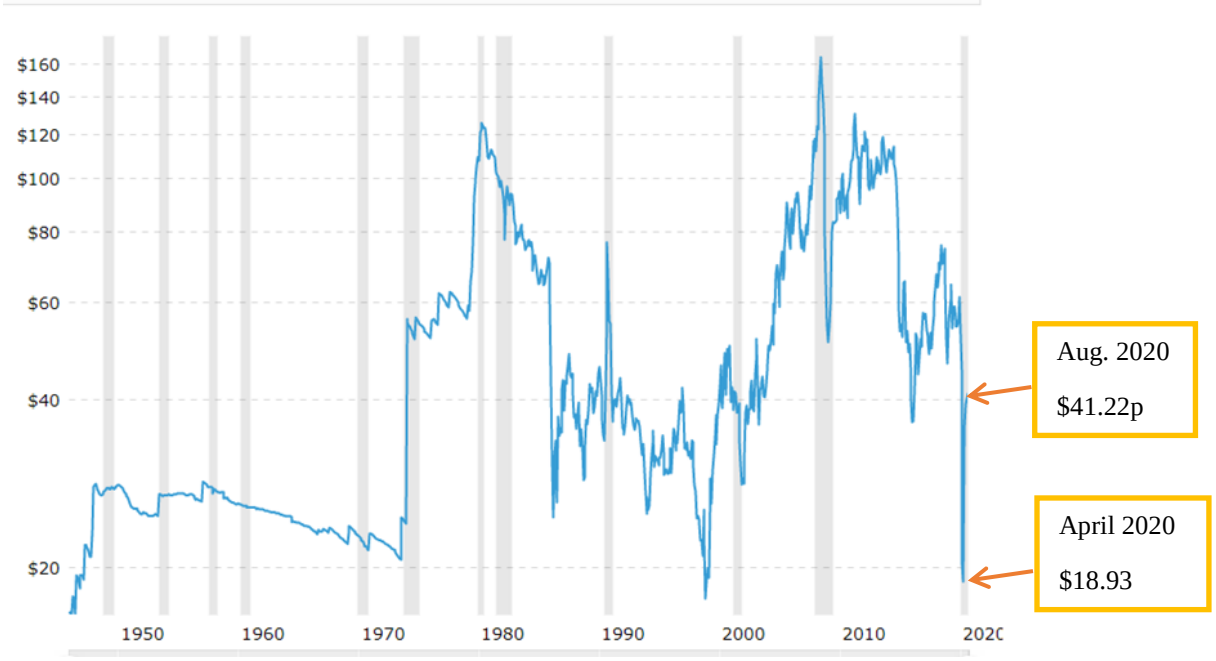

Figure 1.2: Historical chart by West Texas Intermediate (WTI) of crude oil prices per barrel from 1946 to 2020 (Sources: Macrotrends, 2020).

1.2.4 Depleting Oil Reserves

Energy sources derived from fossil fuels are important in meeting the increasing world demand, nevertheless, the realisation that these sources are limited has led to interests in the search for alternative sources of energy (Hossain *et al*., 2017; Hina *et al*., 2014). Most conventional sources of energy are non-renewable as their reserves are rapidly depleting due to prevalent consumption (Hina *et al*., 2014). For this reason, in the past 20 years, there has been much interest in renewable energy sources in developing and developed countries (Hossain *et al*., 2017). The Middle East has about 80% of the world's accessible crude oil reserves with about 62.5% situated in Saudi Arabia, Kuwait, Iraq, United Arab Emirates and Qatar.

1.2.5 Ever-growing World Energy Demand

The world consumption of energy has been increasing steadily each year since 1970's. According to International Energy Agency (IEA) (2016), the world's energy needs continues increasing while millions of people lack access to energy. In 2016, the IEA

further predicted a 30% rise in global energy demand by 2040, translating into an increase in consumption of all modern fuels. On the otherhand, there was 5% drop in oil demand in the first quarter of 2020 which was due to retrictions on mobility and aviation during COVID-19 lockdown pandemic while the demand for gas was relatively moderately affected with 2% drop in demand (IEA, 2020). Notwithstanding this drop in energy demand, the world is still in an energy crisis. The current energy crisis could become worse if sufficient efforts are not put into exploring alternative sources of energy. The global energy resources have diminished over the past years while the world's population has kept increasing (Ohimain and Izah, 2017). This cry for energy security has attracted a lot of international attention.

It is obvious that we should not wait for fossil reserves to be totally depleted or for climatic changes to get worse before we develop renewable, sustainable, reliable and more environmental friendly sources of energy. Research in alternative sources of energy such as bio-butanol is necessary if the world is going to beat the global energy crisis and tackle environmental pollution. There is an urgent need to research renewable liquid fuels as a source of energy, since liquid fuel (especially gasoline and diesel) continue to fulfil most transportation energy demand from vehicles to aviation (International Energy Outlook, 2021).

1.3 Biofuels as an Alternative Source of Energy

Renewable energy is energy that can be replenished. For more than twenty years, several alternative liquid fuels have been researched that could either replace or be combined with crude oil-derived fuels (Bankar, *et al* 2013). In some countries, biofuel, biogas and some other form of biomass have been commercialised as source of energy (Ohimain and Izah, 2017). Liquid fuels have been the major source of energy for transportation as they are used to run ships, aircrafts and automobiles, thus ensuring that society progresses in an effective and efficient manner (Balan, 2014). Biofuels are produced by microorganisms through fermentation of feedstock. Microorganisms can use a wide range of materials as growth substrate thereby converting it to a desired biofuel product. The potential to convert industrial, agricultural, domestic wastes and biomasses to biofuel has further promoted the idea of using microbes for production of biofuels as a renewable source of energy. Furthermore, the combustion of biofuels compared to fossil fuels is more environmental friendly (Elshahed, 2010) due to

reduced particulate matter (PM), carbon monoxide, sulphur oxides, nitrogen oxides and polycyclic aromatic hydrocarbons (PAH) released into the environment (Jin *et al*., 2011; Joshi *et al*., 2015). This is because biofuel gives greater complete combustion than fossil fuels. According to the Carbon Disclosure Project (CDP) Climate Change Report (2016), "renewable energy will need to play a major role in any global shift to a low carbon energy."

Ethanol is the most widely used biofuel for internal combustion of engines produced, in fact some countries such as United States of America, Brazil and China use ethanol as a blend (Jin *et al*., 2011). However, during combustion, ethanol does not yield as much energy as gasoline (Jin *et al*., 2011; Joshi *et al*., 2015). In recent years, a search for a biofuel that has as much energy value as gasoline has led to an increase in research in butanol.

1.3.1 Butanol as an Energy Source

Butanol is a four carbon alcohol that can be used as transportation fuel. Like ethanol, butanol can be produced through both chemical and biological processes. Butanol is produced by fermentation of renewable biological materials by *Clostridium* species (Jin *et al*., 2011) while ethanol is produced by fermentation by yeast, clostridia and other organisms. Despite both being alcohols, butanol has been shown to be a more competitive biofuel than ethanol (Jin *et al*., 2011). Butanol has been agreed to be a superior fuel to ethanol for the following reasons. Firstly, it is gives a better performance with gasoline engines and requires no modification of the engine (Jin *et al*., 2011). This eliminates the need to redesign existing engines to suit it, thus saving cost. Secondly, butanol is less corrosive to pipeline and storage structures (Bankar *et al*., 2013: Jin *et al*., 2011). In addition, butanol is safer to handle than gasoline, due to its higher flash point and lower vapor pressure as shown in Table 1.1. Furthermore, butanol is miscible with gasoline in any ratio. Regulations proposed by the US Environmental Protection Agency (EPA) allow addition of butanol in concentrations up to 11.5% (v/v) to gasoline. Another advantage over ethanol and other alcohols are its non-hygroscopic characteristic (see Table 1.1) (Bankar *et al*., 2013).

Table 1.1 Comparsion of butanol, ethanol, gasoline and diesel

Fuel	n-Butanol	Ethanol	Gasoline	Diesel
Energy density (MJ l-1)	29	19.6	32	39
Vapour pressure (kPa) at 20°C	0.53	7.58	0.7-207	less than 0.07
Vapour pressure of mixture with gasoline (kPa)	44.1	138	53.8-103.4	none
Air: fuel ratio	11.2	9	14.6	none
Motor octane number	78	102	81-89	none
Freezing temperature (°C)	-89.5	-114.5	less than -60	from -30 to -9.9
Hygroscopicity	low	high	low	very low
Compactibility with current infrastructure	yes	no	yes	yes

(Source: Bankar *et al*., 2013)

1.3.2 Chemical Synthesis of Butanol

There are three important processes for commercial production of butanol. The first and most important chemical process is oxo synthesis (Lee *et al*., 2008). Oxo synthesis, also known as hydroxlformation, is the most widely used chemical synthesis process today (Lee *et al*., 2008: Atmadjaja, *et al*., 2016). It is carried out in two steps. The first step is catalysed by Rh, Co or Ru which involves the addition of a C=C double bond to carbon monoxide and hydrogen to form aldehyde (Lee *et al*., 2008). The second step is dehydrogenation of the aldehyde to form different isomers of butanol. Reppe synthesis is the second chemical process which entails direct method of producing butanol under low temperature and pressure. The third chemical process is crotonaldehyde hydrogenation which converts acetaldehyde to butanol through aldol condensation, dehydration and hydrogenation (Ho *et al*., 2016). While Reppe synthesis has never been industrially applied, currently, crotonaldehyde hydrogenation is rarely used (Lee *et al*., 2008).

1.3.3 History of the Acetone-Butanol-Ethanol Fermentation Pathway

Production of butanol by fermentation is one of the oldest microbial industrial processes (Sauer, 2016). Louis Pasteur was the first to report the production of butanol by microbes (Pasteur, 1862). He observed that butanol is a product of microbial alcohol fermentation - although it was a mixed culture of several microorganisms with at least a strain of *Clostridium* (Sauer, 2016; Dürre, 2007*).* Chaim Weizmann, a chemist who

worked at the University of Manchester, triggered the development of acetone-butanol-ethanol (ABE) production for commercial purposes in the early part of the 20th century. He stumbled upon this microbial fermentation when making synthetic rubber containing isoprene or butadiene from butanol and isoamyl alcohol (Patakova *et al*., 2011). Weizmann ended up studying the ABE production process and the microbial culture involved in the process of which *Clostridium acetobutylicum* was dominant (Sauer, 2016).

This revolution into the development of acetone and butanol microbial process was further boosted by the first World War when acetone demand increased for the production of cordite, a smokeless gun powder (Patakova *et al*., 2011). *C. acetobutylicum* was used for the production of butanol from corn or molasses. In 1935 as prices of molasses dropped, the first plant to use new solventogenic strain feeding on molasses was built in Bromborough, England. In 1936, after Chaim Weizmann's patent expired, there was great expansion of fermentation-based industries in countries like USA, South Africa, Japan, Australia and India. Further expansion and development of the ABE fermentation process was boosted by the second World War in which acetone was the major product of interest (Patakova *et al*., 2011).

In the early 1960's, fermentation-based production of acetone and butanol shrunk owing to fierce competition with the growing petrochemical industry (Jones and Woods, 1986). The development of the fermentation industry was also affected negatively by the rise in prices of molasses due to the increase in its demand for cattle rearing. These two factors are responsible for the decline in the ABE fermentation industries in most parts of the world (Sreekumar *et al*., 2015). Nevertheless the 1960's was when fermentation-based production of acetone and butanol started in China and great expansion of the industrial process took place in 1980. The majority of these industries in China shut down their operations in the late 20th century (Chiao and Sun, 2007).

Currently the looming energy crises, environmental impact of fossil fuels and fluctuation in prices of crude oil are leading the new drive toward revitalising the biobutanol industry. Despite the remarkable advantages of butanol as a fuel, some challenges hinder the re-commercialisation of its fermentative-based production in the world today. These include poor butanol production yield, and low final product

concentration as compared with petrochemical synthesis, as well as, the loss of spore formation and solventogenic capabilities in strains of solvent-producing clostridia, also known as degeneration (Chen and Blaschek, 1999). To compete favourably with chemical synthesis, these challenges will have to be dealt with. Metabolic engineering of clostridia strains based on the availability of complete genome sequences and new genomic manipulation tools is one of the major keys to overcoming these barriers (Bankar *et al*., 2013).

1.4 The Genus *Clostridium*

The genus *Clostridium* belongs to the Phylum Firmicutes, which has 190 species of which some are clinically important while others are biotechnologically important (Bergey's Manual of Systematic Bacteriology, 2009). *Clostridium* species are phylogenetically diverse as they consist of spore formers and non-spore formers; Gram-positive but can easily lose the primary dye to appear Gram-negative or give variable Gram reaction; cocci and rod-shaped; and anaerobic and aerotolerant to varying degrees, which some species will grow but not sporulate when exposed to air (Onderdonk and Garrett, 2020). Generally, species of *Clostridium* have peritrichous flagellae for motility with spherical endospores (Bergey's Manual of Systematic Bacteriology, 2009).

In nature, *Clostridium* species can break down a wide range of organic matter and can be found in the soil as well as in the intestine of humans and animals as normal flora (Wells and Wilkins, 1996). A few of them produce endotoxins and are pathogenic. The major pathogenic species of medical importance are *Clostridium botulinum, C. perfringens, Clostridium difficile, Clostridium septicum, Clostridium novyi, Clostridium tetani and Clostridium sordelli* (The UK Standards for Microbiological Investigations, 2015).

Clostridium species are capable of breaking down proteins into amino acids or polypeptides (proteolytic) or have the ability of breaking down sugars with the production of energy (saccharolytic) and are catalase negative (Bergey's Manual of Systematic Bacteriology, 2009). The end products of sugar fermentation include acetic acid, butyric acid, acetone, butanol and large amounts of carbon dioxide and hydrogen gases. Some species of *Clostridium* also produce different varieties of extracellular enzyme for the breakdown of collagen, cellulose, lipids, proteins and other large

organic molecules into fermentable compounds. For example, *Clostridium cellulolyticum* and *Clostridium thermocellum* can hydrolyse cell walls of plants containing cellulose, hemicellulose and lignocellulose, thereby playing important roles in carbon cycling and biodegradation (Desvaux *et al*., 2000; Van der Lelie *et al*., 2012). On the other hand, cellulose-hydrolysing clostridia are very sensitive to oxygen (O_2), thus have low level of tolerance to air (Bankar *et al*., 2013).

Clostridium saccharolyticum, Clostridium saccharoperbutylacetonicum, Clostridium beijerinckii and *C. acetobutylicum* are the four most important saccharolytic species linked to the Acetone-Butanol-Ethanol fermentation process (Jones and Woods, 1986).

1.4.1 *Clostridium acetobutylicum* ATCC 824

Clostridium acetobutylicum is a Gram-positive, rod-shaped, anaerobic and mesophilic bacterium. Since *C. acetobutylicum* is the first choice *Clostridium* species for industrial biobutanol production, most genetic and biochemical studies focus on it as a representative species (Bankar *et al*., 2013). *C. acetobutylicum* ATCC 824 is an important ABE-producing bacterium with great superiority in stress tolerance, biosynthetic capability and feedstock flexibility and diversity (Wu *et al*., 2015).

In general, the accessibility of annotated genome sequences of clostridia has presented the following opportunities: (1) a better understanding of cellular processes through their biological and biochemical functions (Koonin and Galperin, 2003); (2) identification of target genes availability; (3) circumventing barriers posed by the restriction modification for transformation of the organism (Minton *et al*., 2016). Nevertheless, the rise in research on *C. acetobutylicum* in the last two decades has been as a result of the availability of the complete genome sequence of the organism (Nölling *et al*., 2001) and metabolic engineering tools.

The ATCC 824 strain was first isolated from soil in Connecticut in 1924 and was shown to be similar to the strain used by Chaim Weizmann (Nölling *et al*., 2001). In 2001, the shotgun sequencing approach was used to determine the whole genome sequence of the solventogenic bacterium *C. acetobutylicum* ATCC 824 (Nölling *et al*., 2001). The model strain's genome consists of a 3,940,880 base pair chromosome including 11 ribosomal operons and a megaplasmid of 192,000 base pairs. pSOL1, the megaplasmid, has 178 genes for the polypeptides responsible for the solventogenic properties of the

organism. Thus, if ATCC 824 loses its megaplasmid which holds all the major genes for ABE production, it is said to be "degenerated" as it cannot carry out its solventogenic roles (Nölling *et al.*, 2001). The macromolecular complex for degradation of cellulose materials have not been genomically characterised in *C. acetobutylicum* as it can only degrade carboxyl methyl form of cellulose and not the crystalline or amorphous form (Shoham *et al*., 1999).

1.4.2 Metabolic Pathway for Butanol Production in *C. acetobutylicum*

The metabolism of ABE-producing *C. acetobutylicum* is biphasic, divided into acidogenesis and solventogenesis occurring at the exponential and stationary stages of growth respectively (Jones and Woods, 1986). It starts with the metabolism of 6-carbon sugars like glucose through the Emben-Meyerhof-Parnas pathway. When one molecule of hexose is converted to 2 molecules of pyruvate, 2 molecules of ATP and NADPH are produced (Tashiro and Sonomoto, 2010). On the other hand, the metabolism of pentose sugars through pentose phosphate pathway (PPP) yields pentose 5-phosphate which is subsequently converted to fructose 6-phosphate and glyceraldehyde 3-phosphate via the transketolase–transaldolase process. The two intermediates of PPP, fructose 6-phosphate and glyceraldehyde 3-phosphate, are then channeled into the glycolysis pathway. Three (3) molecules of pentose yield 5 molecules pyruvate, 5 molecules of ATP and 5 molecules of NADH (Tashiro and Sonomoto, 2010).

Then the pyruvate produced from the glycolytic pathway is cleaved in the presence of coenzyme (CoA) to produce carbon dioxide (CO_2) and acetyl-CoA (Tashiro and Sonomoto, 2010). The three main groups of fermentation products derived from pyruvate are (a) solvents: acetone, butanol and ethanol; (b) organic acids: acetic acids, butyric acids and lactic acids; and (c) gases: hydrogen and carbon dioxide (Jin *et al*., 2011).

Acetyl-CoA is the main intermediate converted to acetate and ethanol. After acetyl-CoA is generated, it is either reduced or condensed to give acetaldehyde and acetoacetyl-CoA respectively. Acetyl-CoA is also phosphorylated to generate acetyl-P. They are all intermediates leading to the production of acids, acetone (Jones and Woods, 1986), except acetaldehyde, which is reduced to ethanol by alcohol and butanol dehydrogenase (AdhE2/BhdABC). During the process of acidogenesis, acetate is

produced from acetyl-CoA by the sequential action of phosphotransacetylase (Pta) and acetate kinase (Ack) while butyrate is produced from butyrl-CoA through phosphotransbutyrylase (Ptb) and butyrate kinase (Buk). Energy is conserved as ATP in both reactions, however, the pH of the culture drops due to accumulation of organic acids. During solventogenesis, the organic acids are partially reassimilated thereby leading to a rise in the pH of the culture. In this phase, acetate is reassimilated through the acetoacetyl-CoA:acetate (CoAT) pathway. Acetate is reassimilated by CoA transferase (CtfAB) conversion to acetyl-CoA, which leads to ethanol production while butyrate is reassimilated by CoA transferase (CtfAB) conversion to butyryl-CoA, which leads to butanol production (Jones and Woods, 1986; Tashiro and Sonomoto, 2010).

Initially acetoacetyl-CoA is converted to acetoacetate through the CtfAB enzyme, subsequently the latter yields acetone via the enzyme acetoacetate decarboxylase. Likewise to produce butanol, butyrl-CoA is converted to butyraldehyde by alcohol/aldehyde dehydrogenase (AdhE1/AdhE2), subsequently butanol is generated from butyraldehyde via the action of bifunctional aldehyde/alcohol dehydrogenase and butanol dehydrogenase enzyme (AdhE2/BdhABC). Butyrl-CoA is initially produced from Acetyl-CoA by the sequencial catalyses of thiolase (Thl), 3-hydroxylbutyryl-CoA dehydrogenase (Hbd), crotonase (Crt) and butyryl-CoA dehydrogenase-electron transfer flavoprotein (Bcd-EtfAB) forming the intermediates acetoacetyl-CoA, 3-hydroxbutryrl-CoA and crotonyl-CoA respectively as illustrated in Figure 1.3.

The ABE metabolic pathway is regulated by carbon and electron flows such that during acidogenesis, the electrons are accepted by protons while in solventogenesis, the hydrogen produced is oxidised thus, NADH or NADPH (reducing equivalents) are oxidised to generate ethanol and butanol (Tashiro and Sonomoto, 2010).

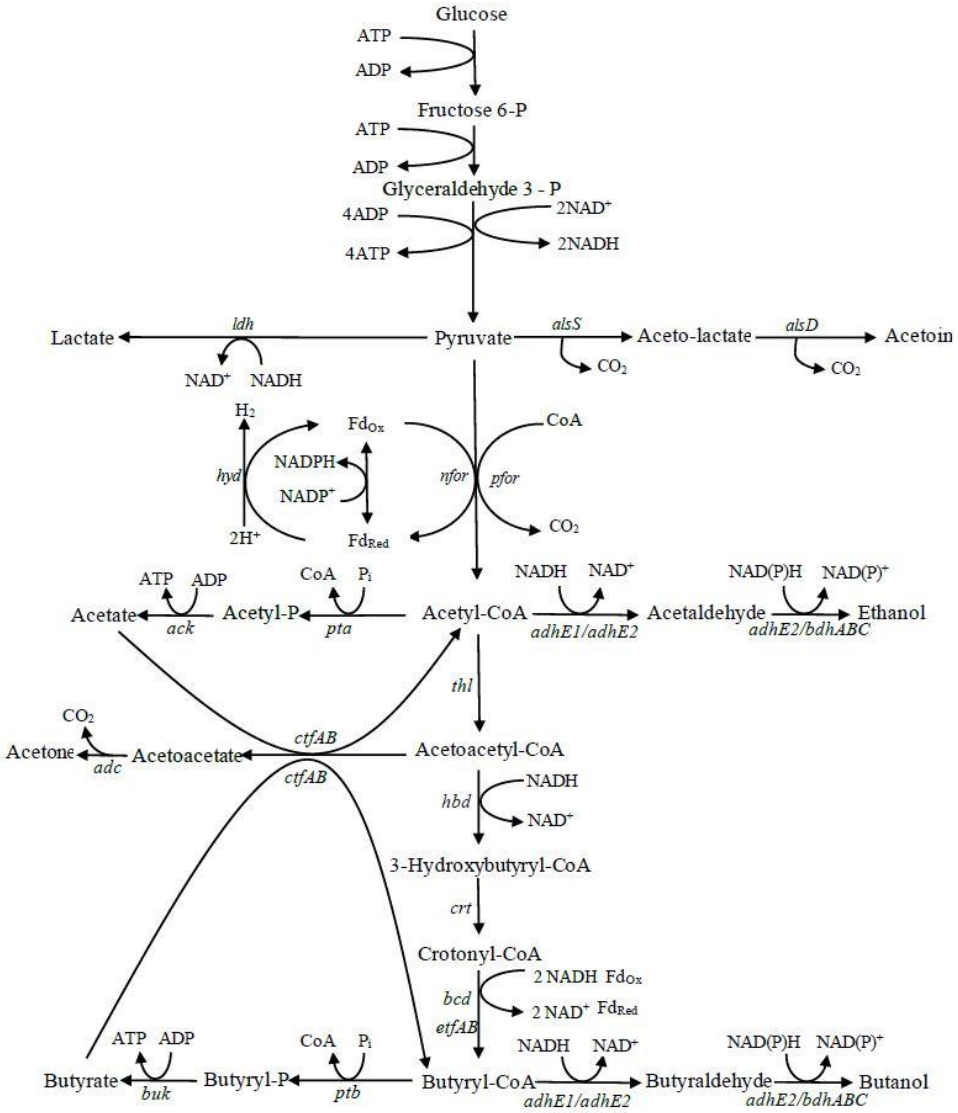

Figure 1.3: Metabolic pathways of glucose metabolism in *C. acetobutylicum* showing genes encoding each metabolic enzyme. Modified from Yoo *et al*., (2019) and Jones and Woods (1986). Fd_{Red} represents reduced ferredoxin whereas Fd_{Ox} represents oxidized ferredoxin. This diagram gives the gene names of the enzymes in italics. Enzymes are indicated by letters as the following: *ldh*, lactate dehydrogenase; *alsS*, acetolactate synthase; *alsD*, alpha-acetolactate decarboxylase; *hyd*, hydrogenase; *nfor*, ferredoxin:$NAD(P)^+$ oxidoreductase; *pfor*, pyruvate:ferredoxin oxidoreductase; *ack*, acetate kinase; *pta*, phosphotransacetylase; *adhE1*, aldehyde dehydrogenase; *adhE2*, bifunctional aldehyde/alcohol dehydrogenase; *bdh*, butanol dehydrogenase; *thl*, thiolase; *adc*, acetoacetate decarboxylase; *ctfAB*, CoA-transferase; *hbd*, 3-hydroxybutyryl-CoA dehydrogenase; *crt*, crotonase; *bcd*, butyryl-CoA dehydrogenase; *etf*, electron transfer flavoprotein; *buk*, butyrate kinase; *ptb*, phosphotransbutyrylase.

1.5 Molecular Biological Tools Developed for Manipulating Genes in *Clostridia*

Gene manipulation can be achieved by gene knock-out, expression regulation and integration. Gene knock-out and gene integration (knock-in) are two potent tools

applied for generating microbial cell factories (Chen and Wang, 2016; Sheng *et al*., 2017). The former can be applied in the inactivation of genes for competitive pathways, thus, favouring or directing the carbon flux to the desired pathway. Gene integration or knock-in is applied in introducing foreign genes that will lead to the synthesis of a new product or exhibition of new traits (Chen and Wang, 2016). Functional genomic studies can only be carried out through mutation (Heap *et al*., 2009b). This is the reason why there is a lot of attention given to developing and testing genetic engineering tools for effective strain manipulation.

Regulation of gene expression is another important transformative tool which can used to control metabolic flux and synthesis of target products. This is achieved by constructing a library of promoters and regulating the transcriptional process through transcription initiation/termination and mRNA degradation (Chen and Wang, 2016).

Over the last 20 years, there has been a variety of molecular tools developed for manipulating genes in *Clostridium* species. Most important of them are discussed below.

1.5.1 ACE Technology

A set of tools used for propelling homologous recombination that could lead to direct inactivation of genes has been of principal use in functional genomic studies in both bacterial and eukaryote species (Heap *et al*., 2009a). One of such molecular biology tools is Allele-Coupled Exchange (ACE) technology. This technique requires the use of a vector to drive DNA integration into bacteria and also contains homology arms in the plasmid that matches the sequence in the organism to direct the location of recombination events that take place. A deletion or damage to the target gene is the result of the second recombination event.

Replicative plasmids are naturally unstable when used for transformation, for this reason, they have limited suitablity for ACE (Heap *et al*., 2012). Exogenous DNA need to be incorporate into the chromosome for it to be stable. Replication deficient or defective plasmids specifically designed with positive selection and counter-selection genes are often used. Their use leads to more stable and irreversible introduction of DNA into the bacterial chromosome.

During ACE, two recombination events occur which are the single crossover and double crossover events. The single crossover could occur at one of the homology arms which leads to incorporation of the entire plasmid, then the double crossover could lead to the creation of a mutant or wild type. According to Faulds-Pain and Wren (2013), the frequency of the second recombination can be increased in two ways. Firstly, the homologous DNA can be designed to be long. Secondly, serially passaging the single crossovers integrants can increase the occurrence of the second recombination. The principle behind this is that single crossovers are unstable, thus members of these integrants have the potential to undergo a second recombination to yield a wild type or mutant genotype and this potential is not lost until a second recombination takes place (Faulds-Pain and Wren, 2013).

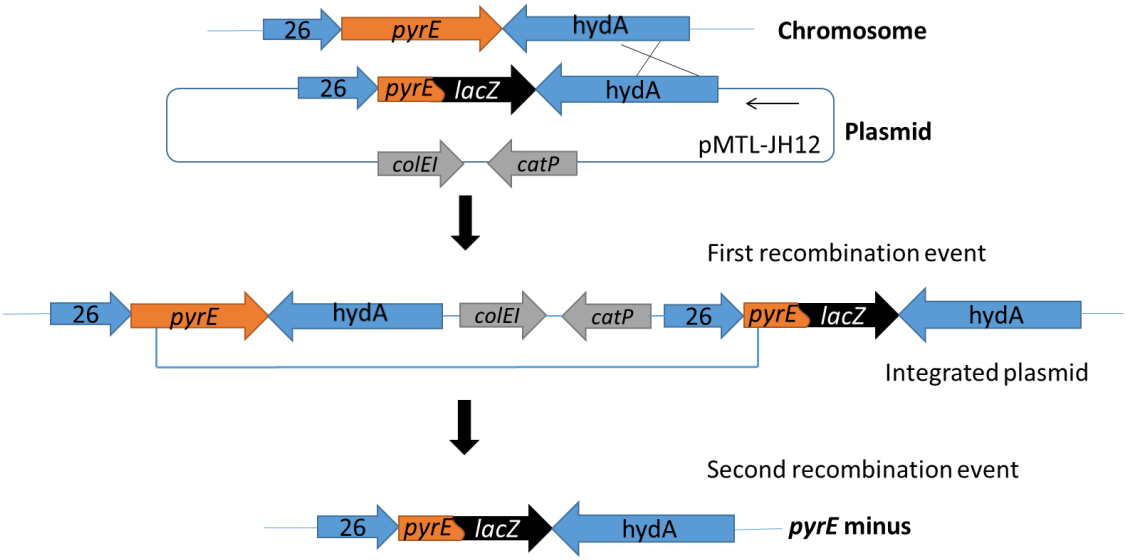

Figure 1.4: *pyrE* mutant generation using Allele-Coupled Exchange. Modified from Heap *et al*., 2012. pMTL-JH12 is designed with left homology arm at *CA_C0026*, a *pyrE* truncated gene fused with *lacZ* multiple cloning site, the right homology arm at *hydA* gene, *catP* antibiotic marker and the Gram-negative origin of replication, ColEI. The first recombination of the chromosome and the plasmid pMTL-JH12 takes place via the right homology arm at *hydA* and leads to plasmid integration. The single crossover integrants are thiamphenicol resistant. The internal portion of the *pyrE* gene directs the second recombination event while the double crossover creates *pyrE* minus which are isolated on FOA.

The *pyrE* gene, which encodes orotate phosphoribosyltransferase, has been inactivated in *C. acetobutylicum* ATCC 824 applying ACE technology (Heap *et al*., 2012). The plasmids were designed with an antibiotic selection marker, thiamphenicol or

chloramphenicol; *codA*, and two homology arms, a 1200 bp right homology arm (long) & a 300 bp left homology arm (short) as illustrated in Figure 1.4. The long homology arm directs the first recombination. After the integration of the replication deficient plasmid into the chromosome, single crossovers were screened and sub-cultured in Clostridium Growth Medium (CGM) to allow for second recombination event. The short homology arm directs the second recombination and the transformed cells were grown on 5-FOA which gives a toxic product in organisms that has the pathway. Colonies on 5-FOA were screened for loss of plasmid by replica plating. The resulting double crossover integrants should be a *pyrE* mutant.

1.5.2 ClosTron

ClosTron is a sophisticated genetic tool designed by the CRG at the University of Nottingham. This is based on group II intron encoded by the *ltrB* gene of *Lactococcus lactis.* ClosTron achieves insertional inactivation by the group II introns propagating between the specific target sites using the disentangling mechanism in RNA (Figure 1.5). The LtrA protein forms a ribonuclear protein complex with the intron which recognises the target sequence and forms base pairs between the specific target site and the excised intron RNA. Subsequently, the intron excises and inserts into the target gene thereby disrupting the gene function (Heap *et al*., 2007) as illustrated in Figure 1.5. The intron can be re-targeted by changing the intron RNA-encoding sequence on the plasmid. A computer algorithm is used to predict the required changes for the targeted intron (Kuehne and Minton, 2012).

To eliminate low frequencies of inactivation, a Retrotransposition-Activated Marker (RAM) was introduced such that acquisition of antibiotic resistance trait confirms integration of the group II intron (Heap *et al.*, 2007).

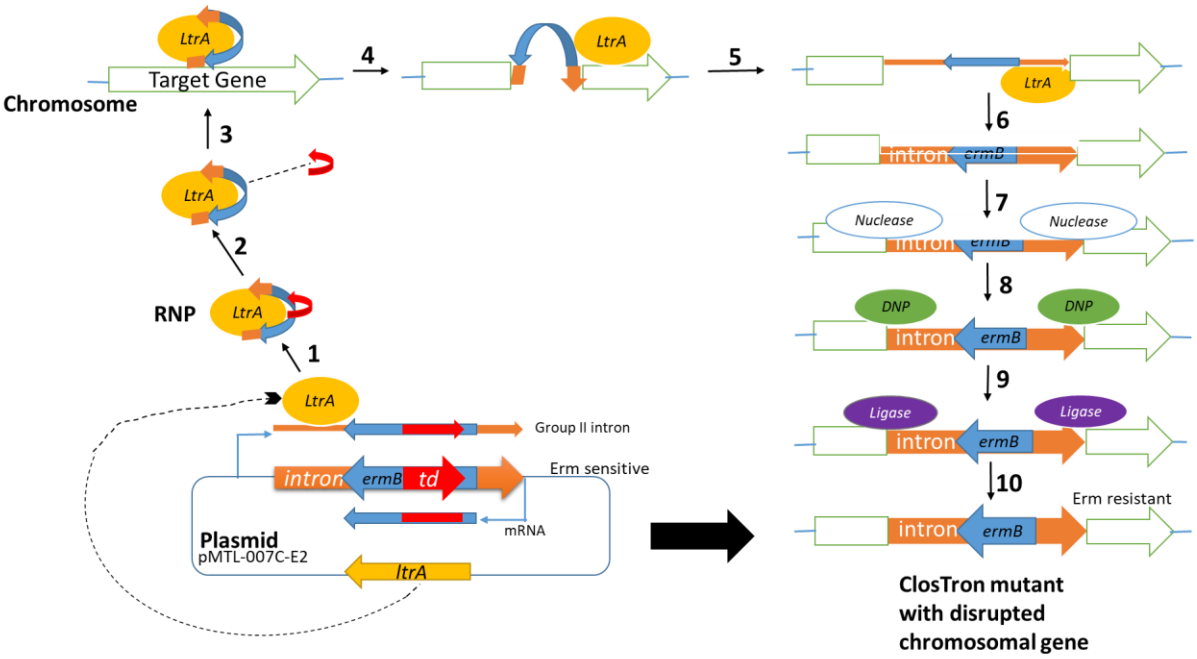

Figure 1.5: Schematic illustration of ClosTron mutant generation. Figure was adapted from Kuehne and Minton (2012). pMTL-007C-E2, the ClosTron plasmid, is composed of *ltrA* gene (yellow) and the group II intron encoding region. The latter encompasses a re-targeted sequence (orange) and *ermB* gene (blue) inactivated by the insertion of *td* (red), a group I intron from phage T4. The *td* group I intron has self-catalytic splicing properties. Therefore having *td* inserted in the opposite direction will allow splicing out to occur only after the mRNA transcription of the opposite DNA strand. Hence, the pMTL-007C-E2 transformants are erythromycin sensitive. (1) LtrA protein (yellow) encoded by *ltrA* gene binds to the mRNA of the group II intron transcript from the opposite DNA strand to form a ribonuclear protein (RNP) complex. (2) The *td* (red) is spliced out. (3) Through the retargeted sequence, RNP recognises and binds to the target gene (white) in the chromosome. (4) LtrA protein nicks the DNA. (5) LtrA protein inserts the RNA and (6) synthesises a complementary DNA strand via reverse transcriptase activity. (7) Host nuclease (white) breakdown the inserted RNA. (8) Host DNA polymerase (DNP) synthesize complementary DNA strand. (9) Host ligase (grey) fill in the gaps of synthesised DNA. (10) The end result is a ClosTron mutant which is erythromycin resistant with group II intron disrupting the function of the target gene.

Heap *et al*. (2010) developed an improved ClosTron system that is easier and faster than the earlier one of Heap *et al*. (2007). First, they provided an intron design tool and made it free to the public at http//:clostron.com. Secondly, they constructed the ClosTron mutagenesis system into the pMTL80000 plasmid. Plasmid cloning tools were also made available at the same website through which members of the public could easily access. Thirdly, transfer of plasmid into the *Clostridium* species was

improved through the appropriate transformation technique (conjugation, electroporation etc). Heap and his co-workers (2010) eliminated the need for an IPTG inducible promoter that was proposed in the earlier work, resulting in introns being expressed immediately after transformation into the host of clostridia. Based on the RAM system, transformed cells were selected using media supplemented with erythromycin. Confirmation of insertions in suspected colonies is done by screening using PCR. Lastly, replica plating is used to confirm the loss of plasmid (Heap *et al*., 2010).

The development of ClosTron as a genetic tool has revolutionised clostridia functional genomic studies. It has been successfully used to create more mutants from strains that are especially difficult to be genetically manipulated. Having the qualities of being efficient, easy to use, reproducible and taking a few days to create a mutant make it an attractive tool. However, it is disadvantaged by having a high probability of polar effects downstream of the target gene and non-polar deletions (Heap *et al*., 2009b; Ehsaan *et al.*, 2016). This poses a major challenge to its use.

1.5.3 CRISPR/Cas Systems

Another molecular biology editing tool worth recognition involves the use of Clustered Regularly Interspaced Short Palindromic Repeats and its associated system known as the CRISPR and Cas genome editing tool. CRISPR Cas systems are naturally found in prokaryotes, with estimated distribution in 90% of archea and 30 to 40% bacteria. These systems employ an adaptive immunity approach to protect against foreign invaders: mobile genetic elements and phages (Westra and Levin, 2020).

A CRISPR locus composes of distinct array of repetitive DNA sequences (up to 50 base pairs) separated by variable spacers sequences (20-50 base pairs). The spacers are transcribed into CRISPR RNA (pre-crRNA) molecules by RNA polymerase. Next, the pre-crRNA is digested by endonucleases into short crRNA known as guide RNA. The guide RNA recognises and base pairs with foreign genetic elements (phages and plasmids) thereby marking them for destruction/cleavage by CRISPR associated system (Cas) proteins (Gophna and Brodt, 2012). Mutations in the protospacer adjacent motifs will lead to a mismatch of the spacer and foreign genetic element, thus there is no cleavage and the host is susceptible to infection.

There are three different subtypes of CRISPR/Cas systems: type I, type II and type III. This classification was based on the peculiar gene they habour. In addition to spacer key proteins Cas1 and Cas2, type I encodes Cas3, type II encodes Cas9 and Cas4 (subtype IIB) or Csn2 (subtype IIA) while type III encodes Cas6 and Cas10 proteins (Shabbir *et al*., 2019).

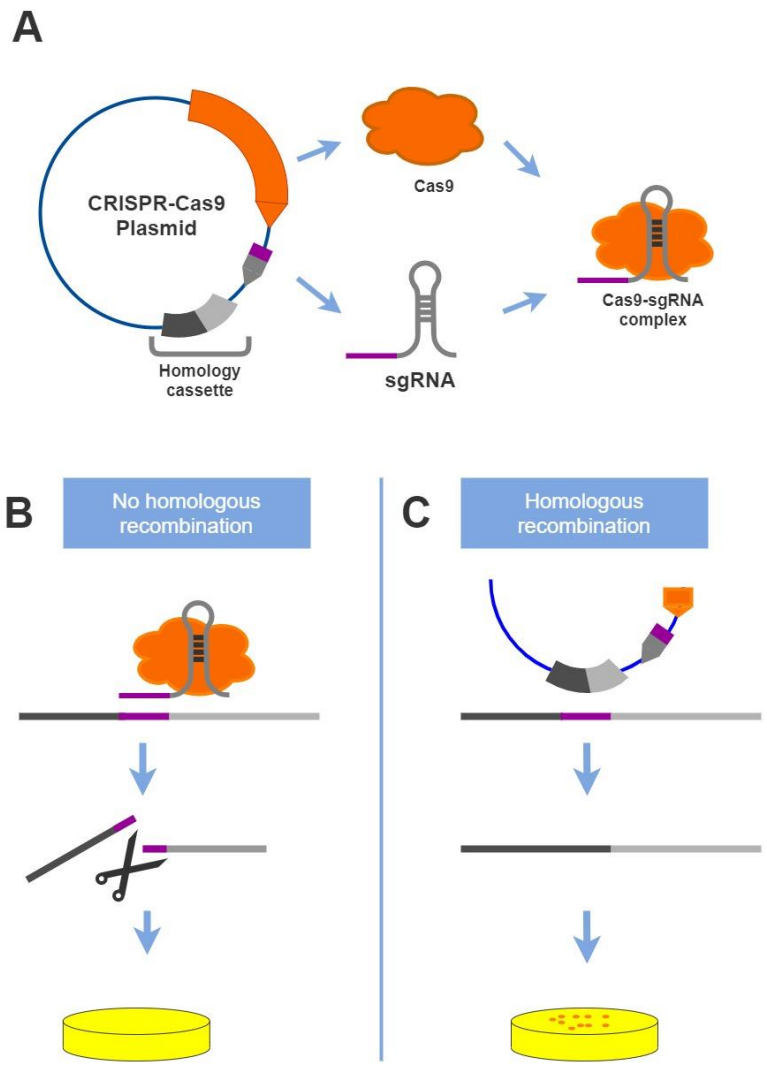

Figure 1.6: CRISPR/Cas9 gene editing. This was adapted from Omorotionmwan, (2019). CRISPR/Cas9 plasmid consists of single guide RNA, Cas9 endonuclease and homology cassette (Figure 1.6A). Cas9 protein guided by the sgRNA cuts the DNA causing double-stranded breaks. This is lethal to the cell (Figure 1.6B). Homologous recombination occurs and this leads to the survival of the cell (Figure 1.6C). If the CRISPR/Cas9 cuts the DNA efficiently, successfully edited cells will bear the antibiotic resistance marker from the plasmid.

Cas9 protein is widely used in molecular engineering. This technology requires the use of a CRISPR/Cas vector encompassing guide RNA, gene encoding Cas9 protein and

the genome editing template (Li *et al*., 2015). An example of how this has been achieved is shown in Figure 1.6. The gRNA array has a short DNA spacer sequence which directs Cas9 proteins to cleave a specific target sequence. The Cas9 endonuclease responsible for cutting the target sequence is often from *Streptococcus pyogenes*. Gene-specific homology arms (upstream and downstream of targeted gene) often serve as the genome editing template and are cloned next to the gRNA in the plasmid. The use of CRISPR/Cas9 gene editing has the advantage of improving accuracy in gene editing without introducing unwanted mutations such as those observed in non-homologous end joining (NHEJ)-dependent double-stranded base repair (Xu *et al*., 2015). When compared with other recombineering techniques, CRISPR/Cas9 has the advantage of being less time consuming, marker-free and high effective targeted gene editing (Jakočiūnas *et al.*, 2016).

1.5.4 Transposons

Apart from the problems of low transformation efficiency, limited availability of shuttle vectors and a strong restriction modification system (Cho and Lee, 2017), another factor affecting the development of biobutanol industrial production is the need for more knowledge of the metabolism which could be elucidated through gene functionality and gene essentiality studies. Though there have been many findings on the physiology of clostridia through the availability of genome sequence, limited genetic tools for the inactivation of the genes identified has prevented the maximum exploitation of the data generated. For functional genetic studies, random mutational tools play a vital role in the study of many bacteria species (Heap *et al*., 2009b). Transposon mutagenesis is the process whereby phages or plasmids are used as vehicles from which a transposon material is extracted and integrated in a bacterial chromosome. Enzymes called transposases encoded on the plasmid are expressed upon transformation and insert transposons into a target gene, interrupting the function of a gene thereby causing mutation. The cut-and-paste transposition mechanism is explained in Figure 1.7.

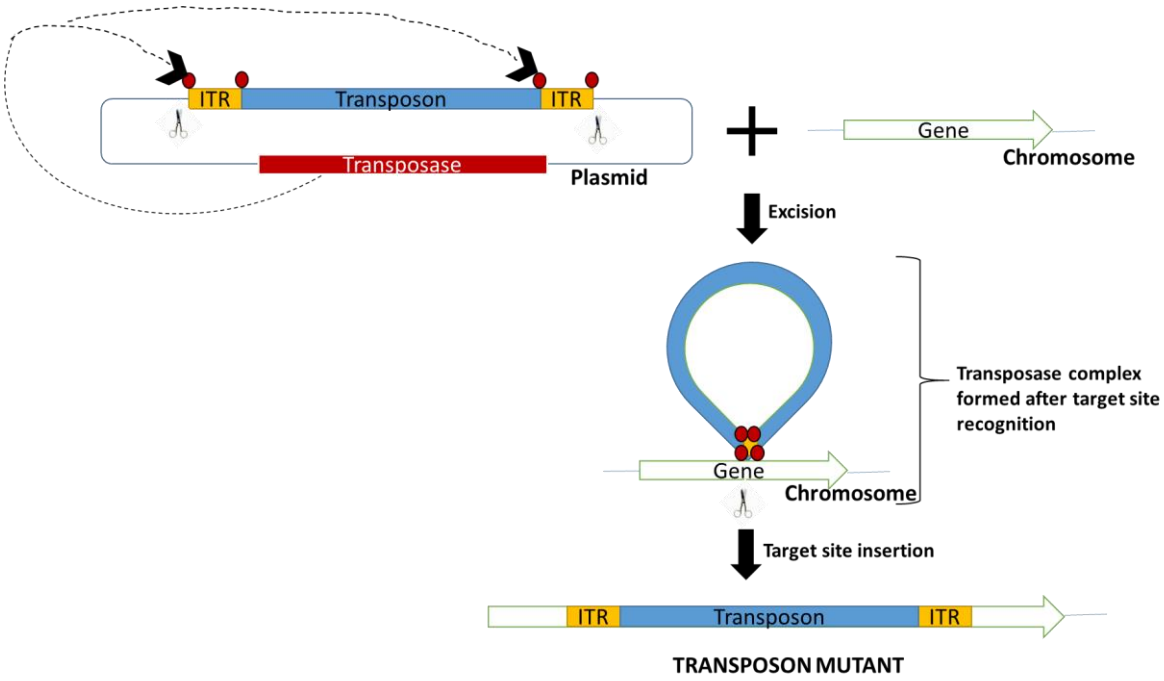

Figure 1.7: Illustration of the cut-and-paste transposition process. Transposase protein (red) is synthesised and excises the transposon at the inverted terminal repeat (ITR) sites (yellow) in the transposon plasmid. Transposase recognises the target gene on the chromosome (white) and forms a transposase complex. The target site is cut and the transposon inserted at the site to create a transposon mutant.

In general, the principle behind transposon mutagenesis is to identify the candidate gene responsible for a phenotype without making assumptions. It involves delivering transposons into bacterial chromosomes to generate libraries of mutants with random insertions. Next, the libraries are screened for mutants that express a particular phenotypic defect. Finally, the genes that have been inactivated due to the mutational inserts are implicated in having a role in expressing the phenotype investigated (Cartman and Minton, 2010). The advantage of this technique lies in its ability to identify multiple genes, novel genes and factors responsible for a particular phenotype as well as essential genes under niche-specific conditions.

Combining transposon mutagenesis with high-throughput sequencing has led to the development of following technologies for genetic studies: Transposon Directed Insertion-site Sequencing (TraDIS), Insertion Sequencing (INSeq), High-throughput Insertion Tracking sequencing (HITs) and Transposon sequencing (Tn-seq) (Goodman *et al.*, 2011; van Opijnen and Camilli, 2013). TraDIS is most commonly used in measurement of cell survival in a population of single transposon mutants, by

fragmentation of the population's gDNA then PCR amplification using primers to selectively enrich the sites in genome adjacent to the transposon inserts (Barquist *et al*., 2016).

1.5.6 Phage-mediated Gene Transfer

Phages are viruses that infect bacteria. They are so abundant that it is predicted that they outnumber bacteria on earth by a factor ten to one (Brüssow *et al*., 2002). Phages are highly specific to their host and this quality serves as an advantage in gene delivery studies. Phage-mediated gene transfer is the process of using phage as gene delivery vector in introducing foreign DNA into a target host. Re-emerging interest in using phage for gene delivery in bacterial host is as a result of the decline in novel antibiotics and increases in antibiotic resistant pathogens, thus increasing the need for phage therapeutic agents. A study by Goh *et al*. (2013) demonstrated the application of phage ϕC2 in the delivery of a transposon element (Tn-*6215*) carrying an erythromycin gene between *C. difficile* strains.

Another aspect of phage genetics employed in genetic engineering studies is the site-specific recombination system. This aspect involves the transformation of a host cell with a vector containing phage integrase/recombinase and foreign gene. This is such that integration of the foreign gene takes place at the recognised phage attachment site of a host genome. There are limited studies on the application of phage-mediated gene transfer and site-specific recombination technologies/tools in genomic studies of *C. acetobutylium*. This might be due to limited number of phages of *C. acetobutylicum* as well as lack of detailed studies on molecular interactions of temperate phages in *C. acetobutylicum*.

1.6 Background of Project

C. acetobutylicum is the most historically important *Clostridium* species used for commercial biobutanol production. There is a great need to overcome the challenges faced by the biobutanol industry if it is to compete favourably with the chemical synthesis. Metabolic engineering of this saccharolytic bacterium can provide vital information that will lead to improved understanding of its ABE pathway, thus solving some of the challenges currently faced by the biobutanol industry. Moreover, advanced genetic tools such as ACE technology, TraDIS and CRISPR, which were not available

during the rise and fall of the biobutanol industry, could be the game changer in fostering beneficial research towards creating an ideal commercial strain to meet the ever-growing energy needs of the world.

Although there have been publications reporting the successful application of transposon mutagenesis in *Clostridium* species (Dembek *et al*., 2015, Dr. Raquel Rodrigues SBRC Nottingham unpublished work), there has not been any work on the use of transposon mutagenesis for the identification of essential genes in *C. acetobutylicum*. In fact, determination of the non-essential and essential genes for viability of a saccharolytic bacterium has not been done using TraDIS. Thus, it is of research and commercial interest to explore and optimise TraDIS as a forward genetic tool for investigating genes in *C. acetobutylicum* as well as novel genes that when mutated can contribute to the butanol tolerance in sub-inhibitory levels. Sub-inhibitory concentrations are concentrations below one capable of inhibiting microbial replication and detectable growth. Several reports showed that exposure of bacteria to antimicrobial concentrations at sub-inhibitory levels induces regulatory responses and alter the expression of different genes (Narimisa *et al*., 2020, Kazi *et al*., 2020). Sub-inhibitory concentrations of antimicrobial agents have a strong effect on the mutation rates which is directly responsible for the emergence of resistant derterminants (Laureti *et al*., 2013). On the other hand, even though there have been records of phage attacks in the ABE industry, there has not been a detailed investigation into the relationship of a lysogenic phage in a saccharolytic bacteria. Through the study of phage genetics, we will foster ideas for building alternative phage vector genetic tools.

1.7 Aim of Project

The overall aim of this project was to use transposon mutagenesis as a high-throughput tool to gain greater insight into the physiological and genetic factors that affect solvent production in *C. acetobutylicum.* In particular, the ultimate goal was to use Transposon-Directed Insertion Sequencing (TraDIS) to simultaneously identify the site of insertion of every transposon in a mutant pool through high-throughput nucleotide sequencing. The project had three aims.

(i) As the implementation of TraDIS requires the production of large mutant libraries, the initial aim of the project was to optimise the number of mutants that could be obtained by increasing transformation frequencies. To achieve this, steps were

taken to remove any barrier presented by the presence of restriction modification (RM) systems in the clostridial recipient through knock-out of the encoding genes. As the *mariner* transposon system used relies on the presence of a functional *tcdR* gene in the clostridial host genome, inactivation of the RM system needed to be undertaken in a *C. acetobutylicum* host carrying *pyrE* mutation. This is so that *tcdR* gene can be inserted at the *pyrE* locus while simultaneously complementing *pyrE*.

(ii) Having maximised the frequency of transformation frequencies, the central aim of the thesis was to maximise the number of mutants that could be generated and establish the optimal conditions for the implementation of TraDIS. The strategy was to generate mutant pools to identify essential and non-essential genes in libraries cultivated in standard clostridial media. Thereafter, mutant libraries could be grown under different niche specific conditions (eg, sub-inhibitory levels of butanol) allowing the identification of genes that were only essential under these conditions.

(iii) During the course of RM mutant generation, evidence emerged that a bacteriophage present as a prophage in the *C. acetobutylicum* genome was functional. A third aim of the thesis was, therefore, to confirm that functional phage particles could be formed and to ascertain whether the vector had potential use as a gene tool, e.g., in transduction.

Chapter three aimed to knock-out the two existing restriction systems in *C. acetobutylicum* using ACE suicide plasmids, compare the transformation efficiencies of single- and double restiction-less mutants as well as determine the electroporation efficiencies of methylation donor pAN2 plasmids.

Chapter four aimed to build transposon libraries of *C. acetobutylicum* in order to identify the genes that are essential and non-essential for viability of saccharolytic *C. acetobutylicum* under standard laboratory conditions.

Chapter five was directed at isolating and characterising the lysogenic phage in *C. acetobutylicum* and determine the relationship of phage in this bacterium.

Chapter 2

Materials and Methods

Chapter 2

2.1 Microbiology

2.1.1 Bacterial Strains

The table below shows all the bacterial strains and their sources used for this study.

Table 2.1: List of bacterial strains, their descriptions and sources

Strain name	Description	Reference/Source
C. acetobutylicum ATCC 824	Type strain, COSMIC variant	SysMo consortium
C. acetobutylicum ATCC 824 Δ*pyrE*	*C. acetobutylicum* Δ*pyrE* mutant; in-frame deletion	Ying Zhang, SBRC culture collection, Synthetic Biology Research Centre Nottingham, University of Nottingham, UK
C. acetobutylicum ATCC 824 Δ*pyrE*Δ*CAC1502*	*C. acetobutylicum* Δ*pyrE* and *CA_C1502* double mutant; in-frame deletion single restriction-less mutant	This study
C. acetobutylicum ATCC 824 Δ*pyrE*Δ*CAC1502* Δ*CAC3535*	*C. acetobutylicum pyrE, CA_C1502* and *CA_C3535* triple mutant; in-frame deletion double restriction-less mutant	This study
C. acetobutylicum ATCC 824 Δ*pyrE*Δ*CAC3535*	*C. acetobutylicum* Δ*pyrE* and *CA_C3535* double mutant; in-frame deletion single restriction-less mutant	This study
*tcdR*Δ*CAC1502* phageless	*C. acetobutylicum* with *tcdR* at *pyrE* locus, *CA_C1502* single restriction-less mutant and the temperate phage absent (*CA_C1867* to *CA_C1951*)	This study
*tcdR*Δ*CAC1502* Δ*CAC3535* phageless	*C. acetobutylicum* with *tcdR* at *pyrE* locus, *CA_C1502 and CA_C3535* double restriction-less mutant and the temperate phage absent (*CA_C1867* to *CA_C1951*)	This study
Δ*pyrE::CRISPR* phageless	*C. acetobutylicum* with complete *pyrE* gene deleted, *CA_C1502 and CA_C3535* double restriction-less mutant and the temperate phage absent (*CA_C1867* to *CA_C1951*)	This study
Δ*pyrE::ACE* phageless	*C. acetobutylicum* with *pyrE* gene lacking 300 bp at 3' end, *CA_C1502 and CA_C3535* double restriction-less mutant and the temperate phage absent (*CA_C1867* to *CA_C1951*)	This study
C. acetobutylicum pMTL85141-51-50	*C. acetobutylicum* with *tcdR* at *pyrE* locus, *CA_C1502* and *CA_C3535* double restriction-less mutant and integrase	This study

	plasmid (containing *CA_C1950* and *CA_C1951*) inserted at the temperate phage loci *attB*	
C. acetobutylicum pMTL85141-67-50	*C. acetobutylicum* with *tcdR* at *pyrE* locus, *CA_C1502 and CA_C3535* double restriction-less mutant and integrase plasmid (containing *CA_C1867*, *CA_C1950 and CA_C1951*)inserted at the temperate phage loci *attB*	This study
C. acetobutylicum pMTL85141-68-50	*C. acetobutylicum* with *tcdR* at *pyrE* locus, *CA_C1502 and CA_C3535* double restriction-less mutant and integrase plasmid (containing *CA_C1867*, *CA_C1868*, *CA_C1950 and CA_C1951*)inserted at the temperate phage loci *attB*	This study
Escherichia coli TOP10	Used for cloning and plasmid storage *hsdR, mcrA, lacZΔM15, endA1, recA1*	Invitrogen, Ltd
Escherichia coli pAN2	*E. coli* TOP10 strain carrying pAN2 plasmid used for methylation Tet^r, Φ3T I gene, p15A origin	Heap *et al*., 2007
Escherichia coli reInterStellar (rIS) pAN2	*E. coli* InterStellar containing pKD46-cm *ara,Δ(lac-proAB) [Φ80d lacZΔM15], rpsL(str), thi, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, dam, dcm* $R702\text{-}Tc^R$, Sm^R, Su^R, Hg^R Tra+, Mob	Dr. Robert Mansfield, Synthetic Biology Research Centre Nottingham, University of Nottingham, UK
Escherichia coli Δ*CAC1502* pAN2	pAN2 with suicide plasmid containing LHA and RHA for in-frame deletion of *CA_C1502*	This study
Escherichia coli Δ*CAC3535* pAN2	pAN2 with suicide plasmid containing LHA and RHA for in-frame deletion of *CA_C3535*	This study
Escherichia coli pMTL85141-51-50	TOP10 with plasmid containing *CA_C1950* and *CA_C1951*, *ΔrepL, catP, colEI RNA II, cd0164* terminator, *lacZ, cpa fdx* terminator	This study
Escherichia coli pMTL85141-67-50	TOP10 with plasmid containing *CA_C1867*, *CA_C1950* and *CA_C1951*, *ΔrepL, catP, colEI RNA II, cd0164* terminator, *lacZ, cpa fdx* terminator	This study
Escherichia coli pMTL85141-68-50	TOP10 with plasmid containing *CA_C1868*, *CA_C1867*, *CA_C1950* and *CA_C1951*, *ΔrepL, catP, colEI RNA II, cd0164* terminator, *lacZ, cpa fdx* terminator	This study

These strains were stored as freezer stocks at -80°C.

2.1.2 Media, Buffers and Supplements

All media used for bacterial growth were supplied by Scientific Laboratory Supplies, Oxoid, UK and chemicals by Sigma-Aldrich Company Ltd, UK unless otherwise stated.

Deionised water was used in the preparation of the media. All media were autoclaved at 121°C at 15 psi for 20 minutes unless otherwise stated.

2.1.2.1 Luria Bertani (LB) Medium

LB broth was prepared as described by Sambrook *et al*. (1989). It contained 10 g tryptone, 5 g yeast extract (Difco Laboratories) and 5 g sodium chloride per litre. Before autoclaving the medium, the pH was adjusted to 7.5. The broth was supplemented with chloramphenicol and tetracycline (see Table 2.2 for concentrations used) for the growth of *E. coli* TOP10 and *E. coli* pAN2 respectively, at 37°C for 16 hours under 200 rpm aerobic condition as according to Heap *et al.* (2007). To prepare LB agar, 10 g/l of bacteriological No. 1 agar was added to the liquid medium before autoclaving. *E. coli* TOP10 and pAN2 were grown aerobically at 37°C for 48 hours on LB agar.

2.1.2.2 2x Yeast Tryptone Glucose (2xYTG) Broth

This broth was prepared as described by Mermelstein *et al.* (1992). It contained 10 g yeast extract, 16 g tryptone (Difco Laboratories) and 20 g glucose per litre of deionised water. The pH was adjusted to 5.2 before it was autoclaved. The pH adjustment is required for the cultivation and preparation of competent cells of *C. acetobutylicum.*

2.1.2.3 *Clostridium* Growth Medium (CGM)

CGM is used for the cultivation of *Clostridium* and other anaerobes. CGM agar medium was prepared as described by Hartmanis and Gatenbeck (1984). It contained 50 g glucose, 1 g yeast extract, 2 g tryptone , 2 g $(NH_4)_2SO_4$, 1 g K_2HPO_4, 0.5 g of KH_2PO_4, 10 mg $MnSO_4 \cdot H_2O$, 0.1 g $MgSO_4 \cdot 7H_2O$, 10 mg $CaCl_2$, 15 mg $FeSO_4 . 7H_2O$, 2 mg $CoCl_2$, 2 mg $ZnSO_4$ and 10 g agar in one litre of deionised water. The mixture was adjusted to pH 7 and then autoclaved.

2.1.2.4 *Clostridium* Basal Medium (CBM)

CBM was prepared as specified by O'Brien and Morris (1971). The medium was prepared by adding 200 mg $MgSO_4 . 7H_2O$, 7.58 mg $MnSO_4 . H2O$, 10 mg $FeSO_4 . 7H_2O$ and 4 g enzymatic casein hydrolysate in one liter of deionized water which was autoclaved. Immediately after sterilising, 0.5 g KH_2PO_4, 0.5 g K_2HPO_4, 1 mg p-aminobenzoic acid, 2 µg biotin and 1 mg thiamine HCL (which were filter-sterilised) and 50 g of sterile glucose per litre were added to the medium before cooling. To

prepare CBM agar, 10 g of bacteriological agar No. 3 was added to the preparation. CBM is used for the propagation of *Clostridium* species.

The media for cultivating *C. acetobutylicum* were put into anaerobic cabinet (supplied by Don Whitley) overnight to ensure that they degassed before use. The anaerobic conditions for the growth of *C. acetobutylicum* were a temperature of 37°C and a gas combination ratio of N_2:H_2:CO_2 (80:10:10). Supplements and antibiotics added to media for this study are listed in Table 2.2.

2.1.2.5 P2 Minimal Medium

P2 medium was used to cultivate *C. acetobutylicum*. To prepare P2, four solutions were initially prepared. Solution 1 comprised of 20 g of glucose in 890 ml of deionised water; Solution 2 comprised of 0.5 g KH_2PO_4 , 0.5 g of K_2HPO_4 and 2.2 g ammonium acetate in deionised 100 ml water with the mixture autoclaved; Solution 3 contained 2 g of $MgSO_4 \cdot 7H_2O$, 0.1 g of $FeSO_4 \cdot 7H_2O$, 0.1 g of NaCl and 0.1 g of $MnSO_4 \cdot H_2O$ in 100 ml deionised water and; Solution 4 comprised of 0.1 g of thiamine, 0.1 g p-aminobenzoic acid and 0.01 g biotin in 100 ml deionised water. Finally, 890 ml of Solution 1, 100 ml of Solution 2, 10 ml of Solution 3 and 1 ml of Solution 4 were all mixed and made up to 1 litre of P2 medium. This formula was described by Baer *et al*. (1987).

2.1.2.6 Electroporation Buffer (EPB)

EPB used for preparing competent cells of *C. acetobutylicum* were made by dissolving 92.4 g of sucrose, 0.6 g sodium phosphate monobasic and 0.71 g sodium phosphate dibasic in 1 litre of deionised water and then it's pH was adjusted to 7.4 and filter sterilised.

2.1.2.7 TAE buffer

TAEx1 buffer used for agarose electrophoresis consists of contained 40 mM Tris, 1 mM EDTA as well as 0.1% (v/v) glacial acetic acid. All images of PCR screening using gel electrophoresis in this study were done with 2-log DNA ladder (0.1 – 10 kbp) from New England Biolab. The letter "L" in the gel images/figures represents this NEB ladder.

Table 2.2: List of supplements and antibiotics added to media

Supplements	Solvent	Final Concentration	Organism	Stock Concentration	Stock storage
Anhydrotetracycline	Water	100 ng/ml	*C. acetobutylicum*	2 mg/ml	-20°*C*
Apramycin	Water	100 µg/ml	*E.coli*	50 mg/ml	-20°*C*
L-Arabinose	Water	100 mM	*C. acetobutylicum* *E. coli*	50 mg/ml	4°*C*
Chloramphenicol	100% Ethanol	12.5 µg/ml (broth) 25 µg/ml (agar)	*E. coli* *E. coli*	25 mg/ml	-20°*C*
Erythromycin	100% Ethanol	500 µg/ml 20 µg/ml	*E. coli* *C. acetobutylicum*	50 mg/ml	4°*C*
5 Flourocytosine	Water	750 µg/ml	*C. acetobutylicum*	20 mg/ml	No stock preservation
IPTG	Water	0-4 mM	*C. acetobutylicum*	1 M	-20°*C*
Kanamycin	Water	50 µg/ml	*E. coli*	50 mg/ml	4°*C*
Tetracycline	70% Ethanol	10 µg/ml	*E. coli* pAN2	20 mg/ml	-20°*C*
Thiamphenicol	100% Ethanol	15 µg/ml	*C. acetobutylicum*	15 mg/ml	4°*C*
Theophylline	DMSO	5 mM	*C. acetobutylicum*	500 mM	No stock preservation
Uracil	Water	20 µg/ml	*C. acetobutylicum*	1 mg/ml	4°*C*

2.1.3 Conditions for Culture Storage

C. acetobutylicum ATCC 824 stock cultures were stored at -80°C in 2xYTG broth with 20% (v/v) glycerol. *E. coli* strains were stored at -80°C using the Microbank™ Long-Term Bacterial and Fungal Storage System.

2.1.4 Bacterial Growth Quantification

Quantification of bacteria growth was carried out by measuring the optical density (OD) at 600 nm wavelength culture using Spectrophotometer BioMate 3 (Thermo Fisher Scientific).

2.1.5 Electro-competent cells

2.1.5.1 Preparation of Electro-competent *E. coli* cells

Five hundred microlitres (500 µl) of overnight of *E. coli* TOP10 (with or without pAN2 plasmid) was added into 50 ml of LB broth. For TOP10, no antibiotic selection was added, however, for pAN2, tetracycline was required (10 µg/ml). The culture was incubated at 37°C with shaking at 200 rpm until an OD_{600} of 0.5 – 0.8 was attained. The cells were kept on ice for 15 minutes for harvesting and then centrifuged at 5000 x g at 4°C for 15 minutes. After discarding the supernatant, the pellet was re-suspended in 100 ml of ice-cold sterile distilled water and centrifuged as above. The pellet was re-suspended in 2 ml of 10% (v/v) glycerol in water and centrifuged as above. The pellet was finally re-suspended in 0.3 ml of 10% (v/v) glycerol. Electro-competent cells were transferred in 40 µl aliquots and stored at -80°C.

2.1.5.2 Preparation of Electro-competent *C. acetobutylicum* ATCC 824 Cells

This technique was designed by Dr. Ying Zhang, University of Nottingham, UK. A heavy loop of *C. acetobutylicum* ATCC 824 Δ*pyrE* was inoculated into 10 ml of 2xYTG plus 200 µl uracil (1 mg/ml stock solution). The inoculum was serially diluted up to 10^{-5} and incubated overnight in an anaerobic cabinet at 37°C. The highest dilution showing growth was selected and inoculated into a sterile conical flask containing 310 ml of 2xYTG supplemented with uracil. The inoculum was further incubated until an OD_{600} of 0.2 – 0.25 was attained. Forty millilitres (40 ml) of the culture was poured into (8) sterile falcon tubes, then the tubes were closed and sealed with paraffin and centrifuged at 4000 rpm at 4°C for 10 minutes with Becker Coulter Allegra K-30R Centrifuge. They were immediately placed on ice and returned to the anaerobic cabinet. The supernatants were discarded while the pellets were re-suspended in cold electroporation buffer (EPB) pH 7.4. Pellets were re-suspended and added together in

a falcon tube. The competent cells were either used immediately or stored at -80°C in 10% (v/v) DMSO.

2.1.6 Electroporation

2.1.6.1 Electroporation of Electro-competent *E. coli* cells

To transform competent *E. coli* cells with plasmid, 4 µl of plasmid was added to 40 µl of electro-competent *E. coli* cells in a 2 mm BioRad electroporation cuvette (previously stored at -20°C). The mixture was exposed to a pulse of 2.5 kV, 25 µF, 200 Ω using BioRad Gene Pulser. Immediately after electroporation, 500 µl of fresh LB was added and the contents of the cuvette transferred to a 1.5 ml eppendorf tube. The cells were incubated at 37°C shaking at 200 rpm for one hour. The cells were serially diluted and used to inoculate LB agar plates supplemented with the required antibiotics.

Methylation of plasmids was carried out by electroporating *E. coli* TOP10 cells containing the pAN2 plasmid with the target plasmids and inoculating the transformed cells into media supplemented with the required antibiotics (Heap *et al*., 2007, Mermelstein and Papoutsakis, 1993). pAN2 vector methylates DNA at specific sites thereby preventing Cac8421 restriction endonuclease from degrading the methylated vector when transferred into *C. acetobutylicum.*

2.1.6.2 Electroporation of Electro-competent *C. acetobutylicum* ATCC 824 with Plasmid

Twenty microgram of pAN2 methylated or unmethylated plasmid DNA was transferred into a 4 mm electroporation cuvette (initially chilled at – 20°C before use) and 590 µl competent cells was added. The cuvette was incubated on ice for 2 minutes and electroporated using a BIORad Gene Pulser with the following settings 2 kV, 25 F and α Ω. One millitre of 2xYTG supplemented with uracil was immediately added to the cuvette and its content transferred to a recovery tube containing 9 ml 2xYTG with appropriate uracil concentration if required. The recovery tube was incubated anaerobically at 37°C for 4 hours. Cells were harvested by centrifugation on Becker Coulter Allegra K-30R centrifuge at 4000 rpm for 10 minutes at room temperature and pellets were suspended in 0.5 ml fresh 2xYTG. Cultures were plated on CGM agar plates with appropriate antibiotic selection and supplement.

2.1.7 TraDIS Procedure

To generate the transposon library in *C. acetobutylicum* ATCC 824, overnight culture of *E. coli* TOP10 carrying the transposon plasmid was subjected to the miniprep procedure to extract the transposon plasmid: pMTL-CW20 and pMTL-CW21. Competent cells of *C. acetobutylicum* ATCC 824 were prepared as described in section 2.1.5.2 and transformed with the pMTL-CW20 or pMTL-CW21 and plated on CGM agar with erythromycin. When using transposon plasmid pMTL-CW20, IPTG at 4 mM (unless stated otherwise) was used as the inducing agent while for pMTL-CW21, aTC at 10 ng/ml was used. After 32 to 36 hours, the transformant colonies were scraped into 20 ml of PBS and serially diluted to 10^{-10}. Hundred microliter (100 µl) of each diluent was inoculated on CGM agar plates containing the following:

- Erythomycin for the selection of the colonies of cells containing the transposon plasmid backbone pMTL-CW20 or pMTL-CW21
- Erythromycin and inducer (IPTG or aTC) for selection of colonies with permissible plasmid loss
- Thiamphenicol and inducer (IPTG or aTC) for selection of putative colonies containing the transposon inserts.

The rest of the undiluted cell suspension was inoculated on forty square (100 x 100) plates containing thiamphenicol and inducer.

After incubating for 48 hours, 100 random colonies from the Tm with IPTG or aTC plates were patch plated on media containing Tm and Erm. This was done to determine the plasmid loss efficiency. Plasmid loss efficiency is the percentage of plasmid harbouring cells that do not grow in the presence of the inducer and antibiotic plasmid backbone. The plasmid loss efficiency, was also derived by calculating:

Calculate:

$$Plasmid\ loss\ efficiency\ = \frac{CFUs\ Erm - \ CFUs\ Erm\ \&IPTG\ or\ aTC}{CFUs\ Erm} \times 100$$

$$Transposon\ mutant\ ratio\ or\ Transposition\ frequency = \frac{CFUs\ Tm\&\ IPTG\ or\ aTC - \ CFUs\ Erm\&IPTG\ or\ aTC}{CFUs\ Erm}$$

$$Resistance\ to\ plasmid\ loss\ = \frac{CFUs\ Erm}{CFUs\ Erm + IPTG\ or\ aTC}$$

Colonies on Tm and IPTG or aTC were collected into a 50 ml falcon tube containing 30 ml CGM and Tm and IPTG or aTC. The transposon library suspension OD_{600} was taken and 1 ml of library was subcultured into 100 ml CGM broth containing Tm and IPTG or aTC (as required) in triplicates. To investigate genes resposonible for butanol and sodium chloride tolerance, the library suspension was also subcultured in 100 ml CGM broth with Tm, IPTG or aTC and 8 g/l of butanol and 8 g/l of sodium chloride in triplicates respectively. The initial OD was at 0.05 - 0.08, the OD was monitored till it reached its stationary phase with OD_{600} of 3.00. After reaching desired OD, then 15 ml of control, butanol and sodium chloride library culture were spun down and gDNA extracted using Sigma Bacterial Genomic DNA extraction kit. Sample of the library before subculture was also spun down and gDNA extracted.

2.1.8 Improved Phage Induction Protocol

Phage induction was done from WT lysogens using mitomycin C and UV light induction. Overnight cultures of *C. acetobutylicum* ATCC 824 *tcdR*Δ*CAC1502*Δ*CAC3535* lysogens were prepared in 20 ml 2xYTG broth pH 7 in an anaerobic cabinet. One in hundred dilutions of the overnight cultures were prepared by transferring 200 µl into 20 ml sterile 2xYTG broth at pH 7 for lysogen culture.

For prophage induction by antibiotics, the OD and pH for the lysogen was monitored every hour until it reached an OD of 0.2. Then mitomycin C was added to the lysogenic cell suspension to a final concentration of 3 µg/ml. It was further incubated for 4 hours while the OD and pH was monitored every one hour. Once the pH fell to 5.5, the lysogen was centrifuged at 4000 rpm for 10 minutes at room temperature. Supernatant was filter-sterilised with a 0.22 µ Sartorius Syringe filter and phage buffer was added to the phage lysate filterate while the pH was adjusted to 7.

Alternatively for prophage induction by UV light, the OD and pH for the lysogen was monitored every hour until it attained an OD of 0.2. Then the cell suspension was exposed to UV light while rotating the falcon tube containing it for one minute. As described in the case of mitomycin C prophage induction, after it was incubated for 4

hours, the phage was harvested by centrifugation and filter-sterilised and phage buffer was added while the pH was adjusted to 7. The supernatants containing free phages particles were kept in the cold room at 4°C.

2.1.9 Plaque Assay

Overnight cultures of *C. acetobutylicum tcdR*Δ*CAC1502*Δ*CAC3535* indicator strain were prepared in 20 ml 2xYTG broth pH 7 in the cabinet. One in hundred dilutions of overnight were prepared by transferring 200 µl into 20 ml sterile 2xYTG broth at pH 7 for indicator culture. The OD_{600} was monitored until it reached 0.2. The culture was centrifuged at 4000 rpm for 10 minutes at room temperature. The pellet was resuspended in broth at pH 7 to OD_{600} of 0.200 while CaCl (10 mM) and Mg_2SO_4 . $7H_2O$ (5 mM) were added.

The phage lysate was used in a plaque assay to test whether phage particles are present. In the anaerobic cabinet, 200 ul of phage supernatant (at various dilutions) was mixed with 1 ml of bacterial indicator culture in 15 ml falcon tubes. Adsorption was allowed for 20 minutes, then 3 ml of 0.5% molten top CGM agar was added to the mixture. The mixture was poured onto the top of a CGM agar plate containing Ca^{2+} (10 mM) and Mg^{2+} (5 mM) and incubated overnight. As a precaution to avoid an uneven top agar, care was taken to avoid moving the plates once they had been added the top agar. The next day, the plates were observed for plaques.

To increase the phage titre, the plaques in the top agar with confluential plaques were scraped off and the material obtained soaked in 10 ml 2xYTG broth for 1 hour before it was spun down and filter-sterlised as described in phage induction protocol section 2.1.8. Then the lysate and its varying dilutions were used to carryout a plaque assay to enumerate and increase the phage titre. This was repeated till the phage titre was above 10^9 pfu/ml.

2.1.10 Characterisation of Bacteriophages by Transmission Electron Microscope

Prior to using transmission electron microscopy (TEM), phage lysates, above $>10^9$ pfu/ml titre, were precipitated as described by Fortier and Moineau (2007). The ammonium acetate precipate technique involved harvesting 1.5 ml phage lysate by centrifugation at 21,000 x *g* for 75 minutes. After centrifugation, 1.4 ml of the supernatant was pipetted and discarded, while the bottom of the tube was not disturbed.

Immediately following this, 1 ml of 0.1 M ammonium acetate solution (from Sigma-Aldrich) was added to the remaining 100 µl of phage lysate. This wash/precipation step with ammonium acetate solution was repeated twice and followed by storage of the precipitated phage particles at 4°C.

To attach the phage particles to the formvar carbon coated copper 200 mesh grid (Agar Scientific), 5 µl of the precipitated phage was transferred on the mesh gridand allowed to stand for 30 seconds. Using Whatman paper, excess liquid was blotted off. Immediately following this, the phages were negatively stained by transferring 10 µl of 2% uranyl acetate to the grid and allowed to stand for 30 seconds. Again the excess liquid was removed with Whatman paper and dried for 5 minutes. Next the phage particles were visualised using Technai T12 BioTWIN Transmission Electron Microscope (100KV) using a Megaview II Soft Imaging System operated by Denise McLean, Senior Scientist Technician, School of Life Sciences, University of Nottingham UK.

2.2 Molecular Biological Techniques

2.2.1 Primers

Primers used for PCR amplification and sequencing in this study were supplied by New England Biolabs (NEB) and listed on Table 2.3.

Table 2.3: List of primers

Primers	**Sequence 5' to 3'**	**Purpose**
Chapter 3		
Cac1501-sf2	GGATTTTTCTCATTTACCCCAGA	*CA_C1502*, restriction modification KO screening
Cac1504-sR1	GCTGTAGTGTAAACTTGTTCTTTG	*CA_C1502*, restriction modification KO screening
Sc7-F	GACGGATTTCACATTTGCCGTTTTGTAAACGAATTGCAGG	Single crossover suicide restriction modification screening
Sc7-R	AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG	Single crossover suicide restriction modification KO screening
Cac3534/F	TTGATGCTCTCGTGTATGCA	*CA_C3535*, restriction modification KO screening

Cac3537/R	CCTCAGTGGGAAGGAATCTTC	*CA_C3535*, restriction modification KO screening
LHA ext F	GAATGTCGTAACCGCTGC	R702 pAN2 recombineering plasmid screening
LHA ext R	ATTCAGAATGCCGAACACC	R702 pAN2 recombineering plasmid screening
Chapter 4		
JH14/F	TAGCACAATTGTATTTGGACTTCTTTAAATAAAAACATGG	*pyrE* deletion, repair or integration at *pyrE* locus screening
JH14/R	TTGATGATGTTTGTCTTGATGACTCAACATGC	*pyrE* deletion, repair or integration at *pyrE* locus screening
ACE/F	CAATGGCAAAGTAATCATAGATGATAAAATGAAAGATAG	*pyrE* deletion, repair or integration at *pyrE* locus Sanger Sequence screening
ACE/R	CTTGCTCACAAATTACTACACGTAAAATACAC	*pyrE* deletion, repair or integration at *pyrE* locus Sanger Sequence screening
Thiolase F	GAACTGTCTTTAGTTTTAATTCTACATTC	PCR screening for transposon plasmid integration at thiolase locus
Thiolase R	GGGGTTAGCATATGCATAAG	PCR screening for transposon plasmid integration at thiolase locus
catP_INV_F1	TATTGTATAGCTTGGTATCATCTCATCATATATCCCCAATTCACC	Inverse PCR transposon screening
catP_INV_R1	TATTTGTGTGATATCCACTTTAACGGTCATGCTGTAGGTACAAGG	Inverse PCR transposon screening
catP_INV_R2	GGCAAGTGTTCAAGAAGTTATTAAGTCGGGAGTGCAGTCGAAGTGG	Inverse PCR transposon Sanger sequence screening
5'Tn1	AATGATACGGCGACCACCGAGATCTACACttcgactgcactcccgacttaataacttc	TraDIS library prep, PCR of library primer and adapter
SplAP5.1	C*AAGCAGAAGACGGCATACGAGATAACGTGATGAGATCGGTCTCGG CATTC*C	TraDIS library prep, PCR of library primer and adapter
SplAP5.2	C*AAGCAGAAGACGGCATACGAGATAAACATCGGAGATCGGTCTCGG CATTC*C	TraDIS library prep, PCR of library primer and adapter
SplAP5.3	C*AAGCAGAAGACGGCATACGAGATATGCCTAAGAGATCGGTCTCGG CATTC*C	TraDIS library prep, PCR of library primer and adapter

SplAP5.4	C*AAGCAGAAGACGGCATACGAGATAGTGGTCAGAGATCGGTCTCGGCATTC*C	TraDIS library prep, PCR of library primer and adapter
SplAP5.5	C*AAGCAGAAGACGGCATACGAGATACCACTGTGAGATCGGTCTCGGCATTC*C	TraDIS library prep, PCR of library primer and adapter
SplAP5.6	C*AAGCAGAAGACGGCATACGAGATACATTGGCGAGATCGGTCTCGGCATTC*C	TraDIS library prep, PCR of library primer and adapter
SplAP5.7	C*AAGCAGAAGACGGCATACGAGATCAGATCTGGAGATCGGTCTCGGCATTC*C	TraDIS library prep, PCR of library primer and adapter
SplAP5.8	C*AAGCAGAAGACGGCATACGAGATCATCAAGTGAGATCGGTCTCGGCATTC*C	TraDIS library prep, PCR of library primer and adapter
SplAP5.9	C*AAGCAGAAGACGGCATACGAGATCGCTGATCGAGATCGGTCTCGGCATTC*C	TraDIS library prep, PCR of library primer and adapter
SplAP5.10	C*AAGCAGAAGACGGCATACGAGATACAAGCTAGAGATCGGTCTCGGCATTC*C	TraDIS library prep, PCR of library primer and adapter
iPCR Taq Seq	AAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTC	TraDIS library illumina sequencing
5' Seq Long HPLC	CTTGAACACTTGCCGAAAAACAATTGTGTCAGACCGGGGACTTATC	TraDIS library illumina sequencing
illumina RPI	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	TraDIS library illumina sequencing
qPCR2.1	AATGATACGGCGACCACCGAG	qPCR for TraDIS library
qPCR2.2	CAAGCAGAAGACGGCATACGA	qPCR for TraDIS library
Chapter 5		
Cac1866F	AGTATGTCAGCAGTATGCAC	Phage genome/phage integrase screening
Cac1871R2	CCTAATGCAATTCTTAAGCC	Phage genome screening
Cac1951R	CTTTCATGTCATTAGTTCTTGGG	Phage genome/phage integrase screening
Cac1949F	TGCACGTTCATAATTTGCCG	Phage genome screening
Cac1951AscI/F2	taattaGGCGCGCCgattattactataattttaag	Phage integrase plasmid cloning
Cac1951FseI/R1	taattaGGCCGGCCctttctccaattgtttttc	Phage integrase plasmid cloning
Cac1867FseI/R2	taattaGGCCGGCCgccgttgcgataatactaata	Phage integrase plasmid cloning
Cac1950AscI/F4	taattcGGCGCGCCggcttccacttatttgtgtta	Phage integrase plasmid cloning
Cac1868FseI/R5	taattaGGCCGGCCgttacaataaaggggaaatac	Phage integrase plasmid cloning

Cac1950AscI/F3	taattcGGCGCGCCgtaaagaatcttagaggtgaa	Phage integrase plasmid Sanger Sequence screening

2.2.2 Chromosomal and Plasmid DNA Extraction and Purification

Genomic DNA and plasmid DNA were extracted by using Sigma-Aldrich GenElute Bacterial genomic DNA kit and Monarch Plasmid Miniprep kit respectively according to manufacturer's instruction.

2.2.3 Polymerase Chain Reaction (PCR)

The DNA polymerases used for PCR DNA amplification were DreamTaq Green PCR Master Mix and Phusion High Fidelity Master Mix unless otherwise stated. The PCR machine used was Applied Biosystems Gene Amp PCR system 2400 thermal cycler. PCR reactions were prepared and carried out according to the manufacturer's instructions.

2.2.4 Inverse PCR

First, the gDNA of the transposon mutants generated as described under section 2.2.2 were digested with HindIII at 37°C for 1 hour and then heat inactivated at 80°C for 25 minutes. Next, ligation was set up with the digest product for 1 hour at 22°C and then heat inactivated at 65°C for 25 minutes. This allowed self-ligation of the digested sticky ends. Next, the heat-inactivated ligation product was used to set up PCR using CatP_INV_F1 and CatP_INV_R1 primers with Q5 High-Fidelity DNA Polymerase Master Mix x2 according to the manufacturer's instruction. After this, the mixture were amplified applying a touchdown protocol as in Table 2.4.

Table 2.4: Inverse PCR touchdown details for screening transposon mutants

Steps	Temperature °C	Time	Stage of cycles
Initial Denaturation	95	60s	
Denaturation	95	10s	Stage 1 Temperature decreases by 1°C per cycle until 60°C is reached. 10 cycles
Annealing	70-60	30s	
Extension	72	360s	
Denaturation	98	10s	Stage 2 25 cycles
Annealing	60	30s	
Extension	72	360s	
Final extension	72	300s	
	4	Pause	

The touchdown products were analysed by gel electrophoresis, cut out and extracted as described as sections 2.2.5 and 2.2.6.

2.2.5 Agarose Gel Electrophoresis

To prepare agarose gels for electrophoresis, 0.8 g of agarose was dissolved in 100 ml of X1 TAE buffer from Fisher Scientific. Ten microlitres (10 µl) of SYBR Safe gel stain from Invitrogen was added to the mixture before pouring into the tanks and allowed to cool. PCR products, plasmid DNA and DNA restriction-digested fragments were mixed with 6X purple loading dye (NEB), when required, before being separated along with 2-Log DNA ladder (NEB) by gel electrophoresis at 100 volts for 50 minutes. The DNA bands were viewed under ultraviolet light and then photographed using a BioRad Gel Doc™ XR Imaging system.

2.2.6 DNA Extraction and Purification from Gel

Fragments of DNA bands were excised from the agarose electrophoresis gel with a sterile scapel and extracted using Monarch DNA Gel Extraction kit.

2.2.7 DNA Sanger Sequencing

PCR fragments and Plasmid DNA were sent for sequencing along with sequencing primers to Source Bioscience (Nottingham, UK).

2.2.8 DNA Quantification

Quantification of DNA samples were carried out using a NanoDrop Lite Spectrophotometer (Thermo Scientific™) at 260 nm. The DNA samples were measured against a blank of water while the A260/280 ratio was used to determine the purity. However, DNA for Illumina sequencing was measured using Invitrogen™ Qubit 4 Fluorometer (Thermo Fisher Scientific). To detect RNA contamination, DNA samples were analysed on agarose gel as described in section 2.2.5.

2.2.9 Cloning Techniques

2.2.9.1 Restriction Enzymes

The buffers and restriction enzymes used for molecular biology were supplied by NEB. All the restriction digestion reactions were carried out according to the manufacturer's instruction.

2.2.9.2 Ligation Enzyme

T4 DNA ligase (NEB) was used to ligate digested and purified DNA fragments according to instructions by the manufacturer. Ligation reactions were incubated on thawing ice overnight so as to allow a temperature gradient.

2.2.9.3 Membrane Dialysis

After ligation, the products were dialysed for 30 minutes over water using a Merck MF-Millipore™ (membrane pore size of 0.025 μm) which were then used for electroporation.

2.2.9.4 Illumina Sequencing

Whole Genome sequencing was done by MicrobesNG (Birmingham, UK) (http://www.microbesng.uk) using Illumina-based.

2.2.10 Plasmid Constructs and Genomic Mutagenesis of *C. acetobutylicum*

The following below are the list of plasmids used in this study (Table 2.5). All the plasmids used in this study were according to the pMTL80000 modular series.

Table 2.5: List of plasmids

Plasmids	Relevant Features	Source
pMTL-CAC1502.KO	Suicide Restriction Modification gene *CA_C1502* KO vector	Dr. Muhammad Ehsaan, University of Nottingham, UK Ehsaan *et al*., 2016
pMTL-CAC3535.KO	Suicide Restriction Modification *CA_C3535* gene KO vector	Dr. Muhammad Ehsaan, University of Nottingham, UK
pMTL85141	pIM13, ColE1, catP (CmR/TmR) containing modular plasmid	Dr. John Heap, University of Nottingham, UK Heap *et al*., 2009a
R702 pAN2 recombineering	R702 recombineering plasmid for pAN2 methylation	Dr. Robert Mansfield, University of Nottingham, UK
pMTL-CW20	Transposon delivery plasmid	Dr. Craig Woods, University of Nottingham, UK Woods *et al*., 2022
pMTL-CW21	Transposon delivery plasmid	Dr. Craig Woods, University of Nottingham, UK
pMTL-GL15	Suicide transposon delivery plasmid	Dr. Gareth Little, University of Nottingham, UK
pMTL-ME6C-tcdR	*pyrE* complementation and backbone for *tcdR* integration	Dr. Muhammad Ehsaan, University of Nottingham, UK
pMTL8315_RibocasE_pyrE	CRISPR Cas9 full *pyrE* KO	Dr. Bunmi omoriotionmwan,

		University of Nottingham, UK
pMTL-JH12	ACE partial *pyrE* gene KO	Dr. John Heap, University of Nottingham, UK
pMTL85141ΔrepL	Suicide plasmid	This study
pMTL85141-51-50	Suicide Phage integrase plasmid	This study
pMTL85141-67-50	Suicide Phage integrase plasmid	This study
pMTL85141-68-50	Suicide Phage integrase plasmid	This study

2.2.10.1 Suicide Knock-Out Plasmid

The knock-out plasmids for restriction modification systems *CA_C1502* and *CA_C3535* were suicide vectors; pMTL-CAC1502.KO and pMTL-CAC3535.KO. These plasmids made based on pMTL-SC7515 with the Gram-positive origin of replication deleted and replaced with left and right homology arms of 750 bp each. The homology arms are sequences adjacent genes *CA_C1502* and *CA_C3535*. CodA functions as counter selection marker. The plasmid maps for pMTL-CAC1502.KO (as seen in Figure 2.1) which is similar to pMTL-CAC3535.KO except for the differences in homology arm sequences are as shown Figure 2.1. They were obtained from the Synthetic Biology Research Group (SBRC Nottingham, University of Nottingham, UK) culture collection having been generated by Dr. Muhammad Ehsaan (University of Nottingham, UK).

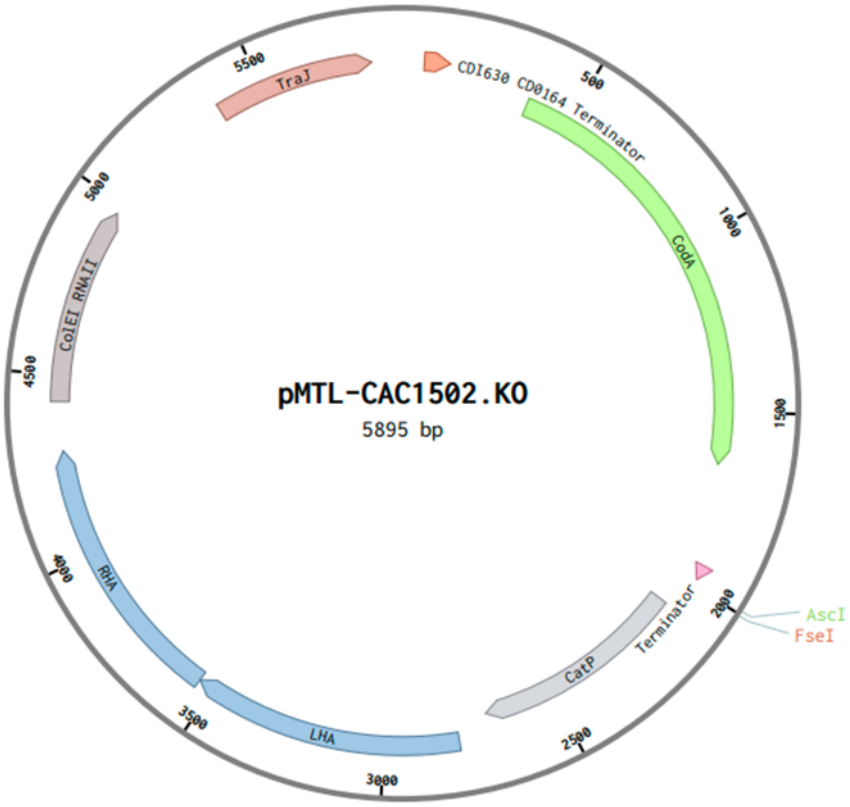

Figure 2.1: The allelic- exchange suicide restriction modification system knock-out plasmid. pMTL-CAC1502.KO has the following components: (i) LHA, left homology arm (ii) RHA, right homology arm (iii) ColE RNAII, the replication region of the *E.coli* plasmid ColE1 (iv) *catP*, the *catP* gene of *C. perfringens* conferring thiamphenicol resistance (v) *traJ*, transfer function of plasmid RP4 *oriT* region (vi) *codA*, the cytosine deaminase gene of *E. coli*, (vii) the promoter in front of *codA* is the P_{fac} promoter of pMTL007 (viii) cpA fdx, a transcriptional terminator of the ferredoxin gene of *C. pasteurianum* and (ix) CDI630 CD0164, a transcriptional terminator isolated from downstream of the *C. difficile* strain 630 CD0164 gene.

2.2.10.2 R702 pAN2 Plasmid for Methylation

The R702 pAN2 plasmid contains the R-factor responsible for transfer function and which expresses the Φ3TI methyltransferase gene. It serves as a methylation donor plasmid for the protection of vectors from cleavage of indigenous restriction endonucleases in *C. acetobutylicum*. It was designed with P_{BAD}-araC located upstream of methyltransferase gene so that methylation is controlled with the presence of L-arabinose. There is an apramycin resistant antibiotic marker downstream the methyltransferase gene as seen in Figure 2.2. This plasmid was obtained by plasmid mini prepping *E. coli* reInterStellar R702 pAN2 (*dam*$^-$, *dcm*$^-$) from SBRC Nottingham culture collection.

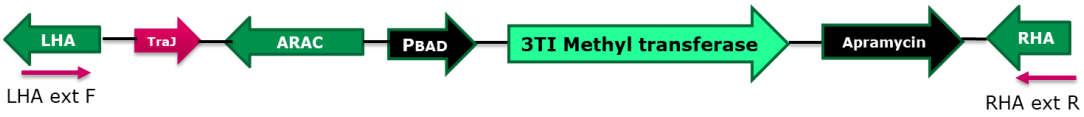

Figure 2.2: Representation of the methylation cassette in R702 pAN2 plasmid. The positions of the screening primers LHA ext F and RHA ext R are indicated in the figure. The methylation cassette in R702 pAN2 contains the following components: (i) LHA, left homology arm (ii) *traJ*, transfer function of plasmid RP4 *oriT* region (iii) *araC*, which encodes AraC protein that regulates the P_{BAD} in *E. coli* (iv) P_{BAD}, promoter part of the arabinose operon from *E. coli* (v) φ3TI methyltransferase, gene from *Bacillus subtilis* phage φ3TI (vi) apramycin, antibiotic marker (vii) RHA, right homology arm.

2.2.10.3 Allele-Coupled Exchange *pyrE* Repair Plasmid

The *pyrE* repair plasmid with *tcdR* integration at the *pyrE* locus used for this study was pMTL-ME6C-tcdR, an ACE *pyrE* complementation pseudosuicide plasmid (Ehsaan *et al*., 2016). It was designed with left and right homology arms adjacent the *pyrE* site of the *C. acetobutylicum* genome, which directs recombination at the *pyrE* site. This plasmid was obtained from SBRC Nottingham culture collection.

2.2.10.4 CRISPR/Cas9 Gene Deletion Plasmids

pMTL8315_RibocasE_J23119_MCS is a theophylline riboswitch CRISPR/Cas9 vector created by Ines Cañadas (Cañadas *et al*., 2019). The vector has a synthetic promoter, J23119, responsible for driving the expression of sgRNA. It also encodes a Cas9 gene from *Streptococcus pyogenes* under the control of the promoter of the *Clostridium sporogenes* ferrodoxin gene, fused to a theophylline-dependant riboregulator, P_{fdx}-rbE. pMTL8315_RibocasE_J23119_*pyrE* cloned from this plasmid was generated by inserting sgRNA and left and right homology arms targeting *pyrE*, between the SpeI and AscI restriction sites. The vector was used to knock-out the entire *pyrE* gene in phage-less *C. acetobutylicum*. After *C. acetobutylicum* was transformed with pMTL8315_RibocasE_J23119_*pyrE*, it was plated on CGM+Tm containing 5 mM theophylline (induced) and CGM+Tm (non-induced) media. It was expected that there would be more growth on the non-induced plates when compared to the induced plate. After incubating for 48 hours, the colonies on the non-induced plates were streaked on CGM+Tm with theophylline. Colonies on the non-induced were PCR screened for *pyrE* mutants using primer pair JH14/F and JH14/R.

2.2.10.5 Allele-Coupled Exchange Procedure for *pyrE* KO

A special replicative plasmid, pMTL-JH12, was available for the Allele-Coupled Exchange. The vector has an origin of replication cloned from *Bacillus* plasmid pIM13 and a selective marker, *catP*, which confers thiamphenicol or chloramphenicol resistance to its host (Heap *et al*., 2012). The long homology arm (RHA) which contains 1200 base pairs directs the initial recombination, hence, single crossover integrants generated are products of recombination at this locus. The long homology arm (fragment of *hydA*) is located downstream of the *pyrE* gene in *C. acetobutylicum*. The shorter homology arm which consist of a 300 base pair internal fragment of the *pyrE* gene drives the second recombination.

After electroporation of the *tcdR@pyrE*Δ*CAC1502*Δ*CAC3535* phage-less strain with pMTL-JH12, the transformants were spread on CGM media containing uracil and thiamphenicol in the given concentration in Table 2.2 to give single crossover colonies. Selected single crossover colonies were passaged twice to allow for second recombination event to occur. The second recombination event cleaves the plasmid, leading to inactivation of *pyrE* gene. The *pyrE* mutants were counter-selected on media containing 5-FOA and uracil. The presence of the *pyrE* gene renders *C. acetobutylicum* ATCC 824 uracil prototroph and converts 5-FOA to 5-fluorouridine monophosphate (5-FUMP) which is toxic to the cell (Heap *et al*., 2012). Replica plating with and without thiamphenicol and PCR with JH14/F and JH14/R primers were used to screen for *pyrE* mutants.

2.2.10.6 Phage Integrase Recombination Plasmid

The Phage Integrase Recombination vectors pMTL85141-51-50, pMTL85141-67-50 and pMTL85141-68-50 were created by deleting the Gram-positive origin of replication (441 bp) from the pMTL85141 plasmid backbone using FseI and AscI and then cloning the phage integrase and *attP* insert from the *C. acetobutylicum* genome into same site using the same restriction enzymes. The pMTL85141 plasmid is from the pMTL80000 modular series (www.clostron.com). The *tcdR@pyrE*Δ*CAC1502*Δ*CAC3535* phage-less was transformed with these vectors. The plasmid map and structure of the pMTL85141 is given in Figure 2.3.

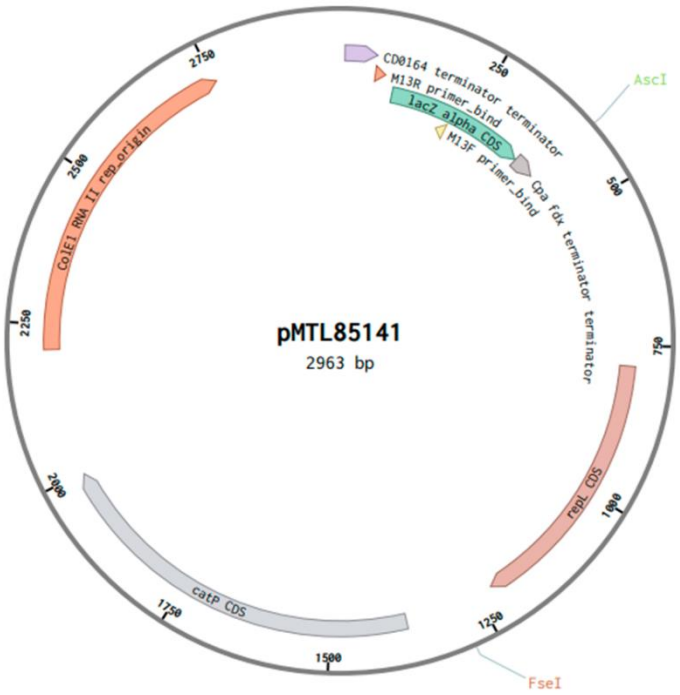

Figure 2.3: The pMTL85141 plasmid backbone for phage integrase vector cloning. This plasmid contains: *ColEI RNA II*, Gram-negative origin of replication; *catP*, chloramphenicol/thiamphenicol marker; *repL*, Gram-positive origin of replication; cpA fdx, a transcriptional terminator of the ferredoxin gene of *C. pasteurianum*; lacZ, region where the multiple cloning site is; and CD0164, a transcriptional terminator isolated from downstream of the *C. difficile* strain 630 CD0164 gene.

2.3 Solvent Analysis

Batch fermentation broth was set up in conical flasks containing 100 ml CBM broth. Overnight cultures were inoculated into reduced medium to an initial OD_{600} of 0.05. Samples were collected at intervals, then OD_{600} and pH were measured at room temperature. For solvent analysis using HPLC and GC techniques, 1 ml of sample collected was centrifuged at 13,000 x g for 5 minutes at 4°C. Next, cell-free supernatants were preserved before analysis by storage in 2 ml screw top tubes at –80°C.

2.3.1 High Performance Liquid Chromatography (HPLC)

Supernatant from Section 2.3 was thawed on ice and an equal volume (300 µl) of 80 mM valeric acid (Sigma-Aldrich, UK) in 0.005 M H_2SO_4 (internal standard) was added and vortexed for 10 seconds. The mixture were filtered using 0.2 µm syringe filters into HPLC vials containing 300 µl inserts with split caps. Glucose was analysed using a Dionex UltiMate 3000 HPLC system (Thermo Fisher Scientific, UK) equiped with a Aminex HPX-87H Ion Exclusion column (BioRad, UK), a refractive index (RI)

detector and photo diode array detector at a flow rate of 0.5 ml/minute at 30°C with 0.005 M H_2SO_4 as mobile phase. Standard concentrations ranging from 0 – 300 mM were used for the analysis. The signal analysis was done using the Chromeleon 7.2 Chromatography Data System (Thermo Fisher Scientific, UK).

2.3.2 Gas Chromatography (GC)

Supernatant from Section 2.3 was thawed on ice and 500 µl was transferred to a sterile eppendorf tube. Next, 5 µl of 10 M H_2SO_4 was mixed into the sample followed by 500 µl of propyl propionate containing 50 mM valeric acid as internal standard (549 µl of valerate was added to 100 ml propyl propionate). The mixture was vortexed and centrifuged, then 300 µl from top layer was transferred into GC vials bearing an insert and covered with screw cap. Acetone, ethanol, butanol, acetate and butyrate were analysed using Thermo Focus GC fitted with TR-FFAP (30 m) column bearing 0.25 mm internal diameter and a Flame Ionization Detector.

2.4 Bioinformatics Tools

2.4.1 Primer Design

Primers design, secondary structure, primer dimer, melting point and percentage GC content were determinied while ordering the primers from the Sigma-Aldrich online https://www.sigmaaldrich.com/GB/en/search/custom-dna-oligos?focus=products&gclid=Cj0KCQjw8_qRBhCXARIsAE2AtRYUXzLbta-2l1PhgdRHeCLVbP4jtAOh7AO0GOnqwYFce8A2yhxmrQ4aAnOQEALw_wcB&page=1&perpage=30&sort=relevance&term=custom%20dna%20oligos&type=product_name.

2.4.2 Plasmid Map Construction

Plasmid map construction and sequencing data handling were performed using benchling.com online.

2.4.3 Illumina Sequencing Data Analysis

Whole genome sequencing data for *C. acetobutylicum* from Illumina sequencing was analysed using the CLC Genomics Workbench 10 (Qiagen Digital Insights). The Basic Variant tool was used to analyse, compare, visualise and identify SNPs, and deletions and phage DNA induction sequencing data.

2.5.4 Database Search

BLASTn search under Nucleotide Collections (nr/nt) database and BLASTx search under Non-redundant Protein database of the National Centre for Biotechnology Information (NCBI) were performed for the identification of gene and protein sequences. The searches were done on https://blast.ncbi.nlm.nih.gov/Blast.cgi. For the phage annotation, the search was also done on Phage Search Tool Enhanced Release (PHASTER) https://phaster.ca/ (Arndt *et al*., 2016) and rapid annotation using subsystems (RAST) pipeline https://rast.nmpdr.org/ (Aziz *et al*., 2008).

2.5.5 Data Analysis and Visualisation

GraphPad Prism for Windows (GraphPad Software, USA) was used for data analysis, graphical presentation and visualisation of data.

Multiple Sequence Alignment Comparison by Log-extraction (MUSCLE) software tool (Output format Clustal W) was used for comparing the large terminase amino acid sequence of ACK1 phage with that of other phages to shed light on the type of the ACK1 phage genome termini (Edgar, 2004). This genetic tool generates neighbour-joining cladogram of the large terminase amino acid sequence of ACK1 phage and others of known termini (https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/).

2.5.6 Library Preparation Process for TraDIS Sequencing

The library prep protocol was carried out by Dr. Craig Woods, Cynthia Akaluka and Dr. Bashir Rumah, University of Nottingham UK as described in Figure 2.4. The MiSeq Sequencing of library mutants was done by University of Nottingham DeepSeq facility UK while the bioinformatics analysis of sequencing data for successful TraDIS studies was done by Dr. Craig Woods, University of Nottingham UK. Fifty microlitres (50 µl) containing 2 µg of a genomic pool DNA was sonicated to about 400 bp size. In the *C. acetobutylicum* ATCC 824 4132880 bp chromosome plus pSOL1, fragmenting the genome to 400 bp means that only one fragment out of above 10,000 fragments from a cell will contain the transposon-tag linked to the unique genome insertion-site sequence. A good library prep protocol should enrich these junctions prior to sequencing.

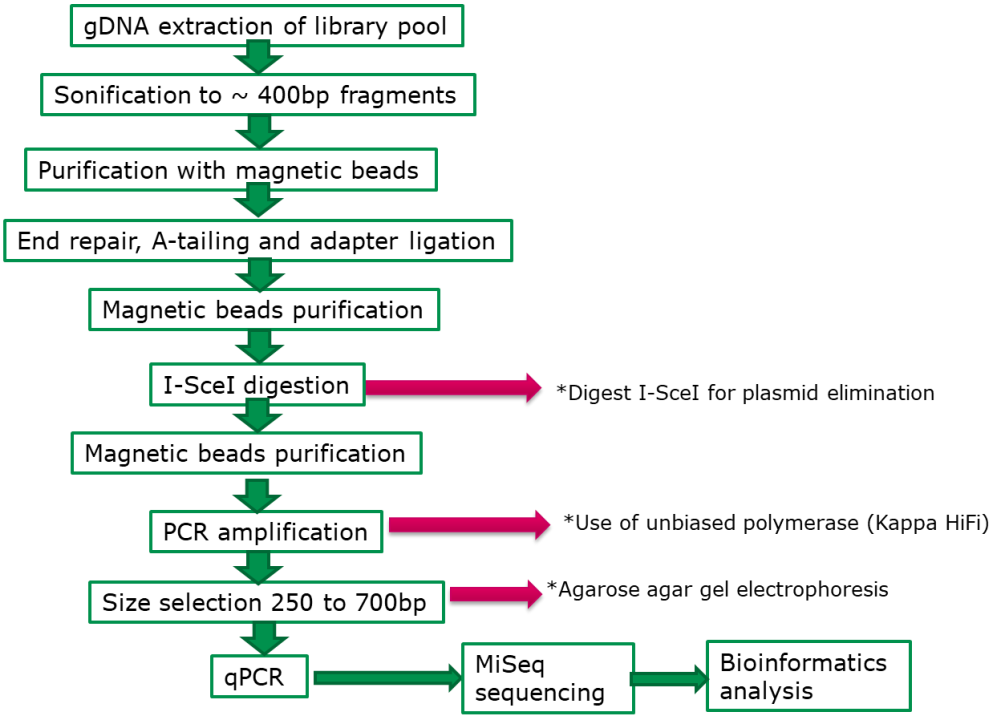

Figure 2.4: Schematic illustrating workflow for transposon mutant gDNA library preparation. The pink arrow shows areas of the procedure that were modified to improve the TraDIS yield.

The next step was purification of fragments using NEB sample purification magnet x1 beads, followed by end repair, A-tailing and then adapter ligation step. The 1x magnet beads purification step was done after each step listed above. Splinkerette is designed into the adapters to increase the capture of the transposon-chromosome junction and reduce hybridisation of the 3'-5' primers during amplification (Barquist *et al*., 2016). Subsequently, after adapter-ligation, 1x magnetic bead purification step was done before I-SceI digestion. The I-SceI digestion step was done to eliminate plasmid reads which is followed by another 1x magnetic beads purification step and subsequent PCR amplification step which enriched the transposon-chromosome junctions.

Kapa HiFi polymerase was used because it has been shown to allow the amplifications of transposon-chromosome junctions without bias to GC (Quail *et al.*, 2012). The strategy involves amplifications of the transposon-chromosome junctions with P5 and P7 primers sequences which are recognised and bind to the flow cell during MiSeq sequencing platform. P5 and P7 were incorporated during the PCR amplification stage of adapter ligated/I-SceI treated samples such that P5 (containing 5Tn1 sequences) annealed to the transposon sequence while P7 (indexed SpIAP5.x primer) contained a

unique region for multiplexing on samples. To optimise the amplicon stage, cycle numbers were increased from 20 to 25 for *C. acetobutylicum* samples.

Following transposon insertion-site amplification step, the quality and quantity of the amplicons was assessed using Bioanalyser (Agilent 2100 high sensitivity) and Qubit prior to size selection of 250-700 bp using agarose gel electrophoresis. The DNA fragments were cut, extracted and measured again with Bioanalyser and Qubit before sending them to the University of Nottingham DeepSeq for qPCR analysis and Miseq Sequencing. Bioinformatics analysis to process, map and analyse transposon insertion sequencing data was carried out by Dr. Craig Woods, University of Nottingham using Bio-TraDIS analysis pipeline software (Barquist *et al*., 2016). It involved searching the reads for transposon tag (an 11 bp sequence on the transposon terminal), then the adjoining sequence was then mapped to the reference genome of *C. acetobutylicum* ATCC 824 (accession NC_003030) and pSOL1 (accession NC_001988).

Chapter 3

Improving the transformation of *Clostridium acetobutylicum* by countering restriction modification systems

Chapter 3

3.1 Introduction

Genetic studies of any industrial microorganism rely on the introduction of DNA into host cells by transformation to bring about the desired genetic changes. In the case of saccharolytic *C. acetobutylicum*, the goal is to improve ABE fermentation. Gram-positive anaerobic clostridia are generally more difficult to transform than Gram-negative bacteria because they have a thick complex network of peptidoglycan which is often also protected by an S layer (Kolek *et al*., 2016). To overcome these barriers and allow passage of exogenous DNA through the cytoplasm in Gram-positive bacteria, various genetic transformation techniques such as protoplast transformation (Reid *et al*., 1983), conjugation (Purdy *et al*., 2002) and electroporation (Ehsaan *et al*., 2016; Pyne *et al*., 2013; Minton *et al*., 2016) have been developed and applied. Currently, the most common technique used in transforming *C. acetobutylicum* is electroporation of the cells. Despite the application of this technique, the transformation efficiency in *Clostridium* is still poor compared to Gram-negative bacteria. This is due to the second critical barrier that must be overcome when delivering DNA into the Gram-positive bacteria, the endogenous restriction modification system (Pyne *et al*., 2013).

Electroporation is an effective method for the transformation of cells only if the DNA introduced is protected from the activities of restriction modification system (Jennert *et al*., 2000). Bacterial restriction modification (RM) systems prevent the transformation of cells with foreign DNA. RM systems recognise specific DNA sequences in foreign DNA, then cleave the unprotected exogenous DNA thereby preventing it from replicating. On the other hand, endogenous DNA is methylated thereby protecting it from cleavage (Suzuki *et al*., 2012). An example of research on improving transformation efficiency by deleting the restriction modification system in solventogenic *Zymomonas mobilis* was demonstrated by Kerr *et al*. (2011). Knocking out a type IV restriction endonuclease ZYO0028 and type I restriction endonuclease gene, ZYO01933, led to 60-fold and 30-fold increase respectively, in transfer efficiency when *Z. mobilis* ZM4 is transformed with unmethylated plasmid (Kerr *et al*., 2011). The knock-out of the restriction modification genes in *C. acetobutylicum* has previously been published (Dong *et al*., 2010, Ehsaan *et al*., 2016; Croux *et al*., 2016). By increasing the genetic accessibility of *C. acetobutylicum* ATCC 824, the burden of work associated with methylating plasmid *in vivo* using *E. coli* before transformation is

reduced, while making transposon mutagenesis studies using suicide vector more attractive.

3.1.1 Restriction Modification Systems in *Clostridium acetobutylicum*

Various restriction modification systems have been characterised in bacteria and archaea and they are divided into four types: type I, II, III and IV. Type I restriction systems have multisubunit proteins which function as a single protein complex. It often consists of two restriction subunits, two methylation subunits and one DNA site sequence recognition subunit. Type II restriction methylation systems have separate individual genes encoding restriction endonucleases and methylases. Type III restriction modification systems have two genes encoding subunit proteins which play the role of either restriction or in DNA recognition and modification. The two protein subunits are required for restriction activity as well as ATP hydrolysis. Finally, type IV restriction systems comprise of one or two genes that encode proteins that cleave only methylated/modified DNA (REBASE Genome, 2020).

C. acetobutylicum has a type II restriction modification system. Possession of a type II RM system gives the organism the ability to cleave specific recognised sequences of foreign DNA using its restriction endonuclease, while its methyltransferase methylates endogenous double-stranded DNA protecting it from cleavage. Methyltransferase enzymes in general produce N^6-methyladenine (^{6m}A), 5-methylcytosine (^{5m}C) and N^4-methylcytosine (^{4m}C) (Suzuki *et al*., 2012). Figure 3.1 shows all the restriction modification genes in *C. acetobutylicum* comprising 5 methylases and 2 restriction endonucleases. It has two sets of type II RM system as shown in Figure 3.2A and Figure 3.2B. The first one is Cac824I, a strong restriction system including restriction endonuclease which recognises and digests 5' GCNGC 3'sequences (Mermelstein *et al*., 1992). It is encoded in the 714 bp gene *CA_C 1502* and upstream of the restriction endonuclease is the methylase encoded by *CA_C1501* (Figure 3.2A). Cac824II is the second type II RM system which has dual function of restriction endonuclease and methyltransferase as seen in Figure 3.2B. It is encoded by 2982 bp gene, *CA_C 3535,* and recognises and cleaves the sequence 5´CTGAAG 3´ (Ehsaan *et al*., 2016; Croux *et al*., 2016). To successfully overcome this barrier, shuttle vectors were methylated *in vitro* before use for transformation (Mermelstein and Papoutsakis, 1993).

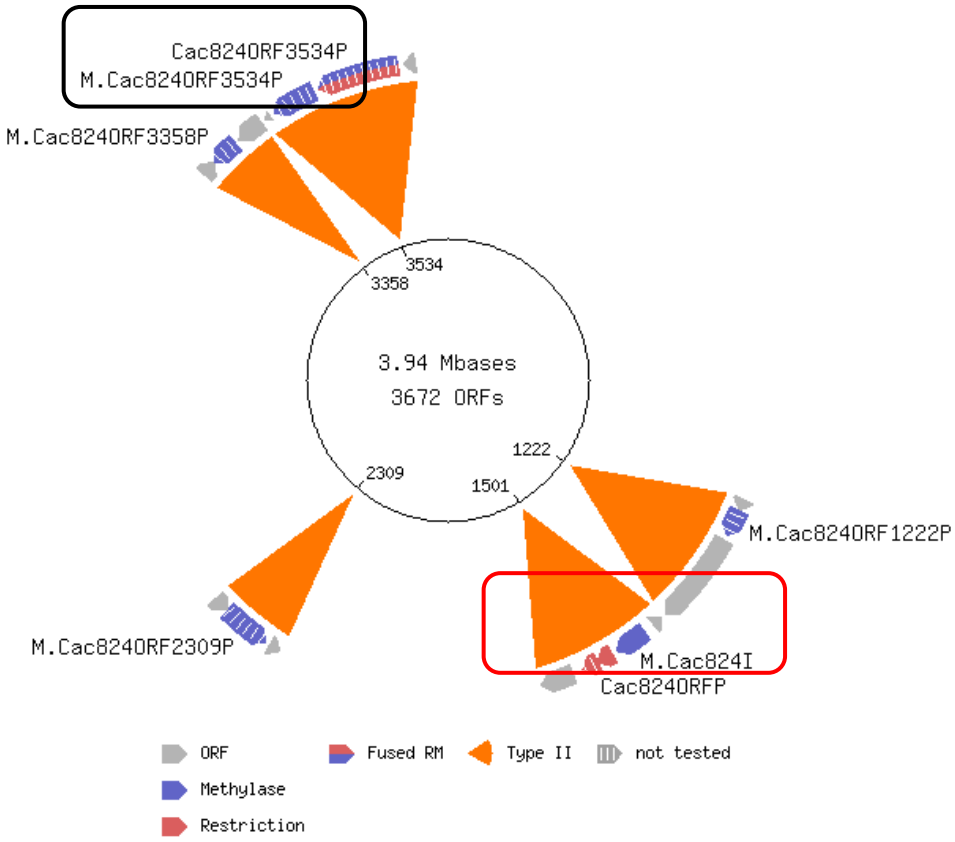

Figure 3.1: *Clostridium acetobutylicum* ATCC 824 restriction endonucleases and methylase genes. The Cac824I type II restriction modification system highlighted in red has restriction endonuclease and methylase genes separated while Cac824II type II restriction modification system highlighted in black has a restriction endonuclease gene fused with the methylase gene. Source of figure: REBASE © Genomes[101]

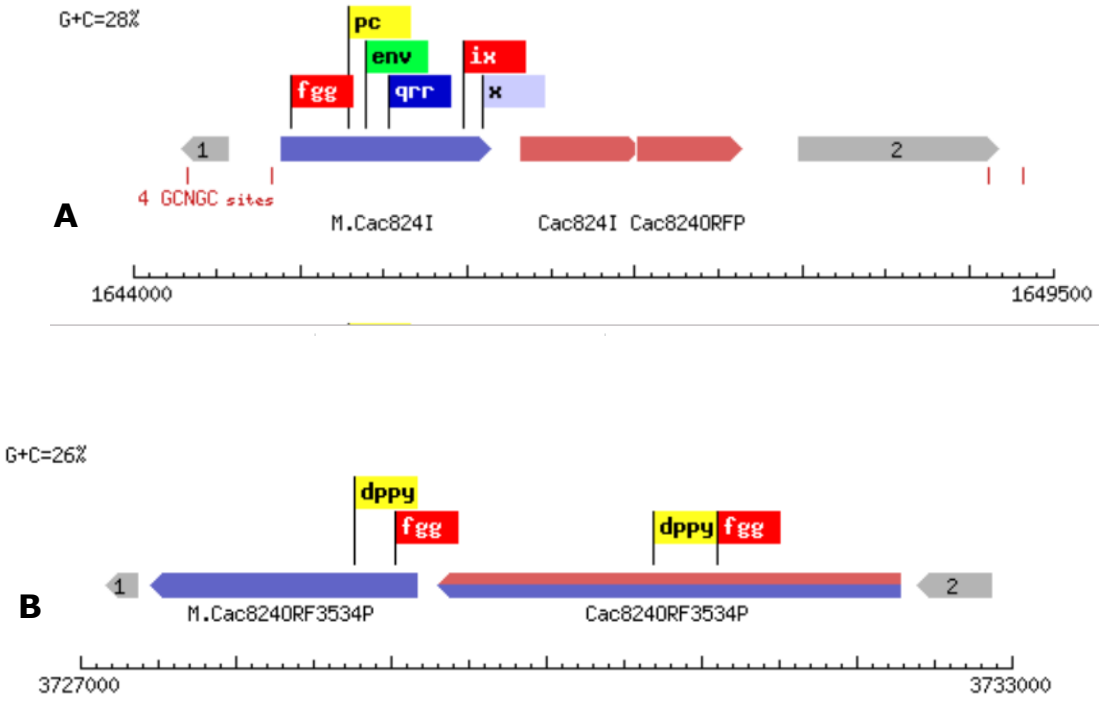

Figure 3.2: Schematics of type II restriction modification systems in *C. acetobutylicum* ATCC 824 from Restriction Enzyme Database (REBASE). Figure 3.2A shows type II Cac824I restriction endonuclease (in red) and methylase (in blue) genes at its upstream loci. Figure 3.2B shows type II fused Cac824II restriction endonuclease (red and blue) and methylase (blue) genes at its downstream loci. Unrelated genes flanking the restriction modification genes are shown in grey. Conserved signature motifs for binding and catalytic activity are illustrated by dppy, fgg, qrr, pc, ix and x.

Cac824I is not commercially available, but Fnu4HI (isolated from *Fusobacterium nucleatum*) also recognises 5' GCNGC 3', cleaving just after the sequence (Leung *et al*., 1979). This isoschizomer may therefore be used to test that the site has been appropriately methylated and protected from cleavage. Two isoschizomer of Cac824II are commercially available, AcuI and Eco571, both of which cleaves downstream of the restriction recognition sequence 5' CTGAAG 3' (Samuelson *et al*., 2006).

3.1.2 Methylation of Exogenous DNA for Transformation of *C. acetobutylicum*

Shuttle vectors prepared in standard *E. coli* donors are likely to transform *C. acetobutylicum* poorly, if at all, because 5′GCNGC3' is a common sequence motif (Mermelstein *et al*., 1992) and will be cleaved by Cac824I when transformed into *C. acetobutylicum*. Since *E. coli* shuttle vectors are the most developed and easiest

recombinant technology vectors tools, methylation of foreign DNA in *E. coli* to mimic the target host strategy was instigated to overcome any problems.

In the host-mimicking strategy developed by Mermelstein and Papoutsakis (1993), Φ3TI methyltransferase of *Bacillus subtilis* bacteriophage Φ3TI was employed in protecting foreign DNA from the Cac824I RM system. The enzyme has two recognition domains, the first one recognises the 5' GGCC 3' sequence and second one recognises the 5' GCNGC 3' sequence. 5-methylcytosine is the result of Φ3TI methyltransferase methylation of the internal cytosine at each recognition sequences. Mermelstein and Papoutsakis (1993) showed efficient transformation performance in *C. acetobutylicum* after *E. coli-C. acetobutylicum* shuttle vectors were maintained in *E. coli* in the presence of pAN1 methylase gene (Mermelstein and Papoutsakis 1993). Since this DNA methylation with methyltransferase has been proven to prevent digestion of vectors by Cac824I restriction endonuclease, numerous genetic studies have been conducted using this method (Pyne *et al*., 2013; Moon *et al*., 2016).

3.1.3 Knock-out Restriction Modification Systems

Besides mimicking host systems by methylating plasmids to be delivered into *Clostridium* spp, deletion of the RM genes is another way of increasing its transformation performance. In fact recently, a lot of emphasis has been laid on the identification of RM systems and applying genetic manipulation tools to knock-out these restriction systems, thereby eliminating the need for methylating vectors (Minton *et al*., 2016).

3.1.4 Previous Studies on the KO of Restriction Modification System in *C. acetobutylicum*

Dong and his co-workers (2010) carried out an experiment to knock-out Cac824I from *C. acetobutylicum* using the ClosTron procedure, and observed that the mutant can be transformed with unmethylated plasmid unlike the wild type. Ehsaan *et al*. (2016) reported an in-frame deletion of Cac824I in *C. acetobutylicum* ATCC 824 using a *codA*-based suicide plasmid and showed that the transformation of this mutant using unmethylated plasmid gave a transformation efficiency of 1.5×10^4 transformants per µg DNA which was similar to that of methylated plasmid DNA.

More recently, the knock-out of *CA_C3535* in addition to *CA_C1502* was achieved by applying *upp*/5-FU counter-selection technique as described by Croux *et al*. (2016) and an improved transformation efficiency by a further 8-fold increase was reported. Though Croux *et al.* (2016) compared the transformation efficiency of Δ*CA_C1502* and Δ*CA_C1502*Δ*upp*Δ*CA_C3535*, there was no report comparing the transformation efficiency of Δ*CA_C1502* with Δ*CA_C3535* strain.

Comparing the transformation performance of mutants having each type of RM systems deletions would give a clearer picture of the impact of the individual RM systems. It is worthy to prove the prediction by REBASE (Roberts *et al*., 2007) that *CA_C1502* is the main restriction modification system in *C. acetobutylicum* ATCC 824 and not *CA_C3535*.

3.2 Hypothesis and Aim of Chapter

Restriction modification systems in *Clostridium* spp are highly specific and diverse, therefore, different shuttle vectors have to be designed for manipulating different species. In addition, the process for methylating DNA *in vivo* in *E. coli* for different experiments is a laborious one. These challenges may hinder the development of effective and efficient transformation procedures in *Clostridium* spp. Increasing the frequency of transformation in *C. acetobutylicum* will highly benefit the planned TraDIS study which was the ultimate goal of this chapter.

However, it was pertinent for TraDIS research reported in this work (chapter 4) that the in-frame deletion of restriction modification genes was undertaken in a Δ*pyrE* background. There were plans to integrate the *tcdR* cargo gene at the *pyrE* locus for TraDIS studies as described in the next chapter. The transposon system previously developed for use in *C. acetobutylicum* was reliant on an orthogonal system based on the *C. difficile* TcdR alternate sigma factor and its cognate promoter from the *tcdB* toxin gene, P_{tcdB} (Zhang *et al*., 2015). The latter promoter was positioned in front of the *mariner* transposase gene on the transposon delivery vehicle, while the *tcdR* gene was integrated into the genome of the *C. acetobutylicum* host at the *pyrE* locus. As the *E.coli* donor lacked *tcdR*, the absence of production of the transposase in the donor meant that transposition did not occur until the transposon plasmid was introduced into the modified *C. acetobutylicum* recipient carrying the *tcdR* gene in its genome.

Transposition of the envisaged TraDIS system was reliant on a transposon promoter and sigma factor.

The first goal was to create an in-frame deletion of *CA_C1502*, the main restriction system, in a *pyrE* mutant of *C. acetobutylicum* ATCC 824. The effect of disrupting *CA_C1502* in the *pyrE* mutant strain was going to be be determined. This study aimed to address questions such as: Is the transformation performance of *ΔCA_C1502* mutant improved by the KO? Is the transformation efficiency of *ΔCA_C1502* mutant using methylated and unmethylated plasmid similar?

Following the work of Ehsaan *et al*. (2016), this project aimed to create a second mutant by knocking out the second restriction endonuclease system which is *CA_C3535*, both in the *pyrE* mutant as well as the Δ*pyrE*Δ*CAC1502* strain. Using an in-frame deletion approach, the project will give a better understanding on the extent of improvement in the transformation of the *C. acetobutylicum* Δ*pyrE*Δ*CAC1502* by knocking out the second type II restriction modification gene, *CA_C3535*. Would the Δ*pyrE*Δ*CAC3535* strain accept unmethylated plasmid? Furthermore, the transformation efficiencies of individual knock-outs of Δ*pyrE* mutant, single restriction-less mutants; Δ*pyrE*Δ*CAC3535* and double restriction-less mutants; Δ*pyrE*Δ*CAC1502*Δ*CAC3535* was compared. It is worth noting that building a TraDIS library for gene essentiality and butanol tolerance investigations was the main aim of the overall PhD project, which was going to benefit from the improvement in the transformation efficiency of the organism achieved as described in the next chapter.

3.3 Results and Discussion

3.3.1 Creation of Single Crossovers for *CA_C1502* KO in *pyrE* mutant

There is no commercial kit available for the *in vitro* methylation of foreign DNA to be transformed. However, deletion of the type II RM system (Cac824I and Cac824II) encoded by *CA_C1502* and *CA_C3535* did result in improved transformation efficiency of *C. acetobutylicum* (Croux *et al*., 2016). To carry out single and double RM systems knock-outs by in-frame deletion, two replicative defective plasmids made available by Dr. Muhammad Ehsaan (SBRC Nottingham, University of Nottingham); pMTL-CAC1502.KO for *CA_C1502* and pMTL-CAC3535.KO for *CA_C3535* (Ehsaan *et al*., 2016). The features of the plasmids were as described in Materials and Method section

2.2.10.1 and Figure 2.1. The Δ*pyrE* strain of ATCC 824 was made applying the ACE technique using the plasmid pMTL-JH14 (Heap *et al*., 2012).

The medium for the cultivation of *C. acetobutylicum pyrE* mutants was always supplemented with uracil since it is uracil auxotroph. Following electroporation of *C. acetobutylicum* Δ*pyrE* with the pMTL-CAC1502.KO plasmid, the recovered cells were plated on medium supplemented with uracil and thiamphenicol to select for single crossovers. Cells that were resistant to thiamphenicol were those cells with plasmid integrated into the chromosome. Screening for single crossovers was done using two pairs of primers. The first pair of primers were Cac1501-sF2, a forward primer complementary to a region of chromosomal DNA of the *C. acetobutylicum* Δ*pyrE* and Sc7-R, a reverse primer which bound to a region on the plasmid as shown in Figure 3.5. The second pair of primers are Cac1504-sR1, a reverse primer complementary to sequences on the chromosomal DNA of the *pyrE* mutant and Sc7-F, a forward primer binding to the KO plasmid. In Figure 3.5, the positions of Sc7-F and Sc7-R on the suicide plasmid were indicated. The results for screening 4 transformants for *CA_C1502* single crossover are as shown in Figure 3.3A while the results for screening transformants for another 6 single crossover integrants are shown in Figure 3.3B. They were all right single crossover integrants. The expected PCR product sizes for single crossover integrants depended on primer pairs used for screening (Table 3.1).

Table 3.1: Expected DNA fragment sizes for right and left single crossover integrants

Primer pairs	Right Single Crossover Band Size	Left Single Crossover Band Size
Cac-1501-sF2 + Sc7-R	2370 bp	1662 bp
Sc7-F + Cac-1504-sR1	1688 bp	2396 bp

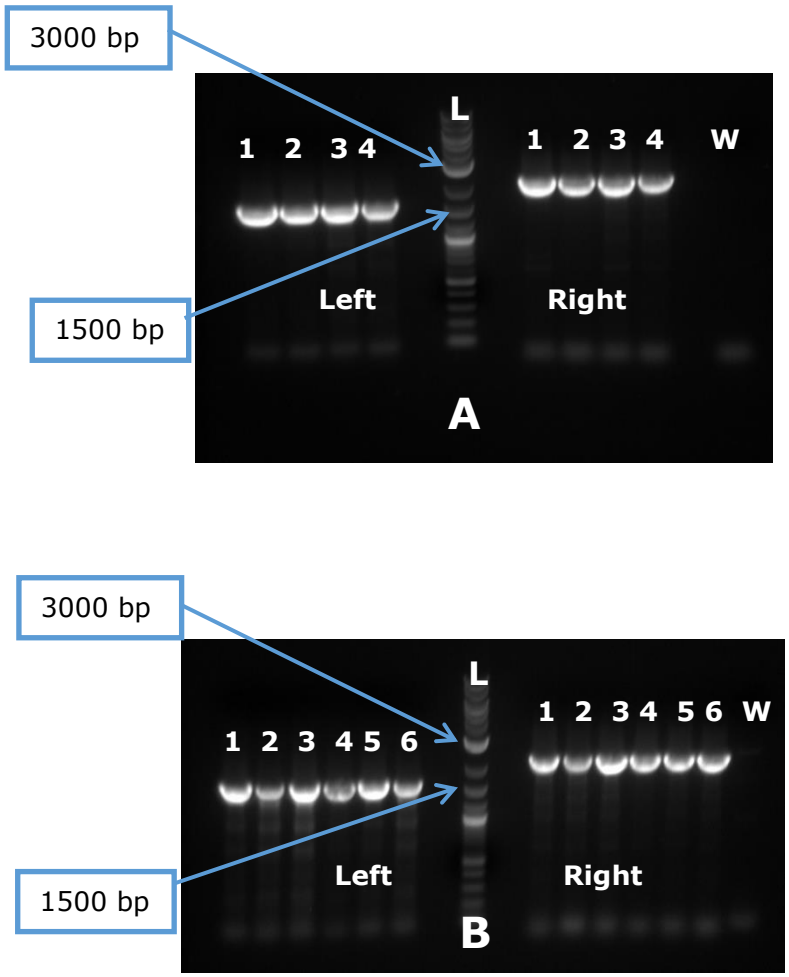

Figure 3.3: PCR screening of 10 single crossover colonies for integration of plasmid pMTL-CAC1502.KO. L represents NEB ladder 0.1-10 kbp. In Figure 3.3A, lanes 1-4 on the left show PCR products generated with primers Sc7-F and Cac-1504-sR1. The resulting DNA fragment is 1688 bp. Lanes 1-4 on the right hand side show PCR products generated with Cac-1501-sF2 and Sc7-R primers, which resulted in a DNA fragment of 2370 bp. In Figure 3.3B, lanes 1-6 on the left show PCR products generated with primers Sc7-F and Cac-1504-sR1, resulting in DNA fragments of 1688 bp. Lanes 1-6 on the right show PCR products generated with Cac-1501-sF2 and Sc7-R primers, which resulted in a DNA fragment of 2370 bp. W represents negative control, which used wild type gDNA as template and did not show any product. This suggests that the ten colonies were right hand single crossover integrants.

3.3.2 Creation of Double Crossover Mutants for *CA_C1502* KO

Following the confirmation of the single crossovers mutants for *CA_C1502* (See section 3.3.1), four colonies were passaged on CGM containing the supplement uracil to allow for second recombination. Then the cells were grown overnight on 2xYTG supplemented with uracil, serially diluted to 10^{-4} and plated on CBM containing uracil and 5-fluorocytosine (5-FC). Following growth on the counter selection marker (5-FC),

the 5-FC resistant colonies were patch plated onto medium containing uracil with and without thiamphenicol. Colonies which were sensitive to thiamphenicol were screened by PCR using flanking primers Cac1501-sF2 and Cac1504-sR1, which are the forward and reverse primers for *CA_C1502*, respectively. This pair of primers should amplify a DNA fragment 2319 bp if the *CA_C1502* gene is intact or a 1611 bp fragment if the *CA_C1502* gene has been deleted as illustrated in Figure 3.6.

Although the method for the creation of single crossover mutants was successful, there were some hurdles faced in knocking out *CA_C1502*. First, a very low frequency of the second recombination event was observed. After patch plating 800-1000 colonies from the 5-FC media transferring into CGM plus uracil medium with and without thiamphenicol, only 2 to 3 individual colonies were observed to have lost their thiamphenicol resistance. When the 2 or 3 colonies were screened by PCR with Cac1501-sF2 and Cac1501-sR1 primers followed by gel electrophoresis, it was observed that they reverted back to wild type (Figure 3.4A, C &D). In addition, double DNA fragment sizes from the PCR product also indicated mixed colonies of wild type and mutant. A further 8 colonies were screened in attempt to isolate a *CA_C1502* mutant but only 4 single DNA fragment sizes showing wild type were observed while the other 4 double DNA fragment sizes indicated mixed cultures of mutant and wild type (Figure 3.4B).

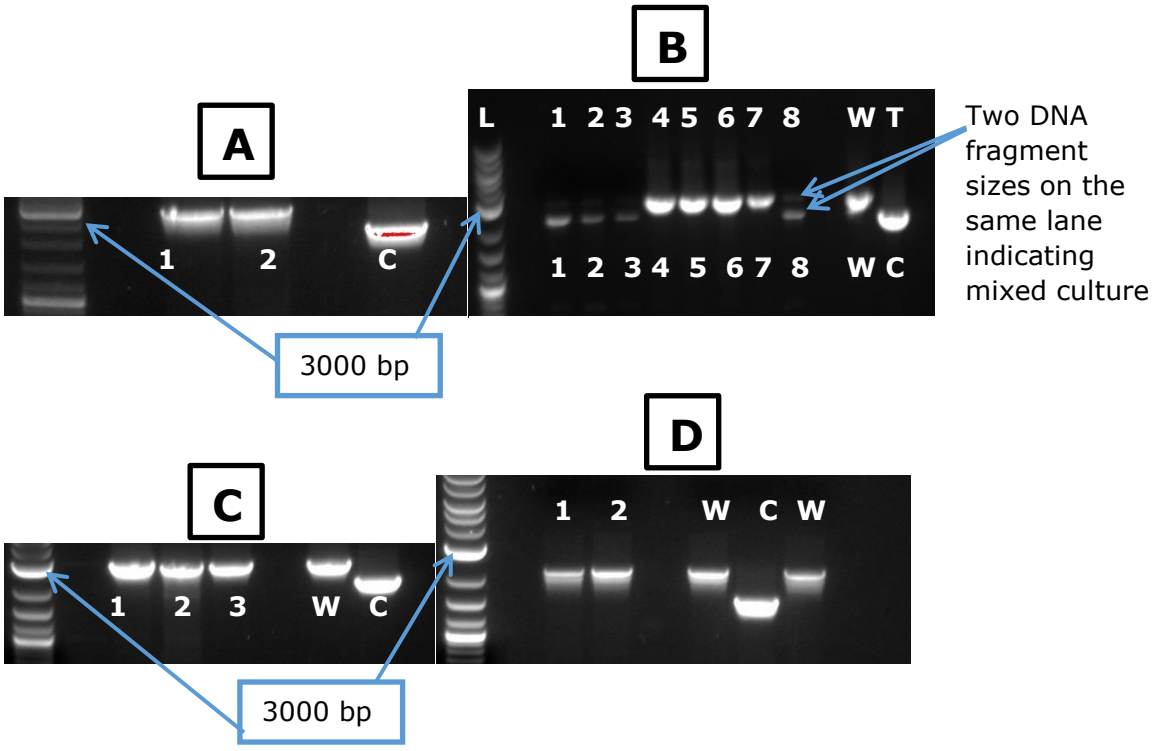

Figure 3.4: PCR analysis revealing reversion of single crossover pMTL-CAC1502.KO integrants back to wild type. L represents NEB ladder 0.1-10 kbp. Images A, B, C and D show PCR screening of integrants for double crossover mutants. In A, C and D, lanes 1, 2 and/or 3 are amplified DNA fragments of colonies that reverted to wild type after the assumed double crossover event, with W and C indicate wild type and positive control for mutant respectively. In gel picture B, lanes 1, 2, 3 & 8 have double DNA fragment sizes which indicate mixed cultures of wild type and Δ*CA_C1502* mutants while 4, 5, 6 and 7 indicates bands of colonies that reverted back to wild type after the selection for double crossovers.

To tackle the problem of low success rates of obtaining double crossover mutants as described in Figure 3.4, the following changes were made;

(1) The single crossover integrants were passaged for longer periods (48 hours) twice to allow the second recombination event to take place.

(2) The concentration of 5-FC was increased from 500 µg/ml to 1000 µg/ml to further increase the toxic effect on single crossover integrants.

(3) The incubation time for single crossover integrants on media containing 5-FC was reduced from 48 hours to 24 hours. This was to increase the chances of picking up colonies which does not experience any toxic effect from exposure to 5-FC, i.e. non-specific effects.

The application of these changes to the method contributed to the isolation of 5 colonies (out of 350 patch-plated colonies) which had lost their thiamphenicol resistance. This was an improvement from previous efforts. Three (3) of the 5 colonies exhibited the expected DNA fragment size of 1611 bp for *CA_C 1502* mutant while two colonies had double DNA fragment sizes which appeared to consist of mixed cultures of wild type and *CA_C 1502* mutant (Figure 3.6). Note that only one of these DNA fragment sizes are visible in lanes 1 and 2.

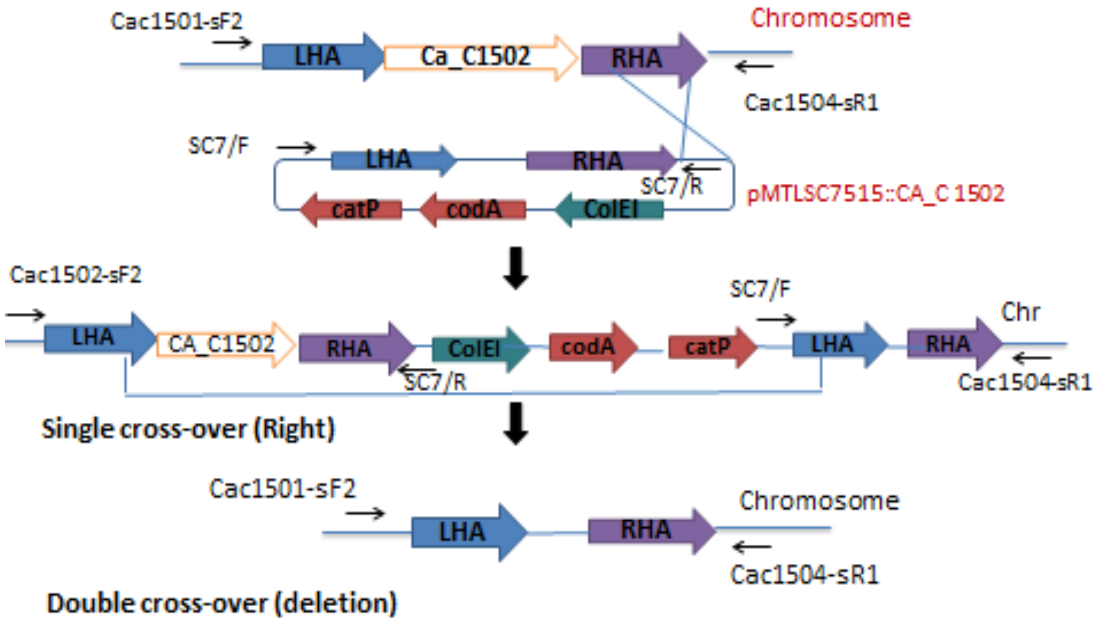

Figure 3.5: Schematics of right hand single crossover integration and double crossover with deletion of *CA_C1502* gene. Relative positions of forward and reverse primers for single and double crossover screening are shown.

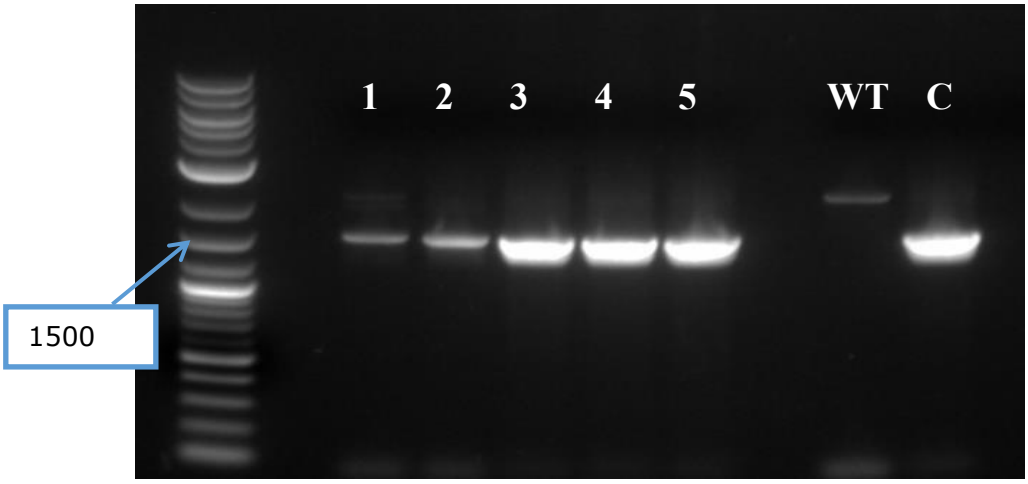

Figure 3.6: Isolation of mutants with in-frame deletion of Cac824I. PCR screening of five (5) putative double crossover mutant colonies in lanes 1-5, WT is the wild type and C is the positive control using Cac-1501-sF2 and Cac-1504-sR1 primers. The expected DNA fragment sizes are 1611 bp for *CA_C1502* mutants and 2319 bp for wild type. Two bands occurred in lanes 1 and 2 but only one band is clearly visible here.

Following double crossover, the three mutant's DNA fragments shown in lanes 3, 4 and 5 in Figure 3.6, were purified and confirmed for deletion of *CA_C1502* as expected using Sanger sequencing. The strains were labelled Δ*pyrE*Δ*CAC1502* strain 3, Δ*pyrE*Δ*CAC1502* strain 4 and Δ*pyrE*Δ*CAC1502* strain 5 respectively and stored in 20% glycerol solution at -80°C.

3.3.3 Creation of a Double Crossovers *CA_C3535* KO in the *CA_C1502* Mutant Strain

To knock out *CA_C3535*, competent cells of *C. acetobutylicum* Δ*pyrE*Δ*CAC1502* were made and transformed with pMTL-CAC3535.KO suicide plasmid. Following transformation by electroporation, the recovered cells were plated on medium supplemented with uracil and thiamphenicol to select for single crossovers. Single crossover integrants were cells which had been transformed with the designed plasmid that conferred thiamphenicol resistance to them due to the presence of *catP* gene. Four single crossover colonies were passaged on CGM containing the supplement uracil to allow for second recombination. Then the cells were grown overnight on 2xYTG plus uracil, serially diluted to 10^{-4} and plated on CBM containing uracil and 5-FC. Following growth on the counter selection marker (5-FC), colonies that grew on the media were screened by patch-plating to isolate colonies that had undergone the second recombination. Out of 300 colonies, 9 colonies that had lost their thiamphenicol resistant were further screened using PCR with flanking primers Cac3534/F and

Cac3537/R, which are forward and reverse for *CA_C3535* respectively. This pair of primers should amplify a 4,352 bp if the *CA_C3535* gene is intact or a 1370 bp fragment if the *CA_C3535* gene has been knocked out as illustrated in Figure 3.7. The DNA fragments in lanes 1, 3 and 4 as shown in Figure 3.7 were cut and extracted from the gel, and confirmed for deletion of *CA_C3535* as expected using Sanger sequencing.

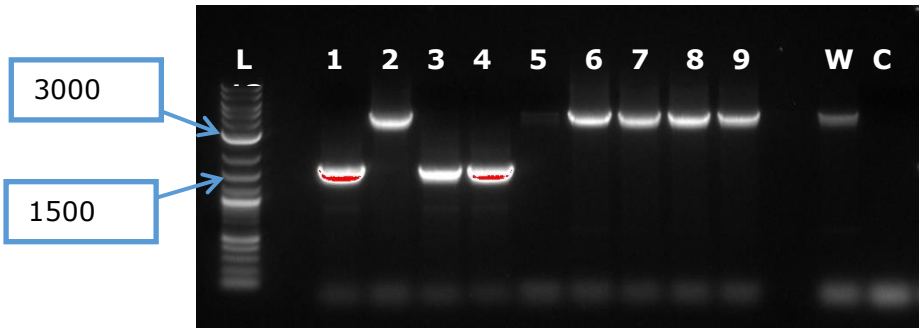

Figure 3.7: PCR screening for occurrence of *CA_C3535* mutations in 9 thiamphenicol sensitive colonies. L represents NEB ladder 0.1-10 kbp. In lanes 1, 3 and 4, a DNA fragment size of approximately 1,370 bp is observed, corresponding to the deletion of *CA_C3535*, while lanes 2, 5, 6, 7, 8 and 9 show a DNA fragment of 4,352 bp, corresponding to the wild type genotype. WT indicates a control consisting of wild type DNA yielding a fragment of 4,352 bp while C is the negative control without the inclusion of DNA.

3.3.4 Creation of Double Crossovers of *CA_C3535* KO in the Δ*pyrE* Strain

Preparation of *C. acetobutylicum* Δ*pyrE* competent cells and transformation of the cells with pMTL-CAC3535.KO suicide plasmid was done as described in Materials and Methods 2.1.5.2 and 2.1.6.2. After transformation by electroporation, the recovered cells were plated on medium supplemented with uracil and thiamphenicol to select for single crossovers. Four putative single crossover colonies were passaged on CGM supplemented with uracil, to allow for occurrence of the second recombination. Subsequently, the cells were grown overnight on 2xYTG plus uracil, serially diluted to 10^{-4} and plated on CBM containing 5-FC and uracil.

Colonies that grew on the media containing the counter selection marker (5-FC) were screened by patch-plating to isolate colonies that were thiamphenicol sensitive. Out of 200 colonies screened, 6 had lost their thiamphenicol resistance and were further screened using PCR and gel electrophoresis in attempt to isolate an in-frame deletion

of *CA_C3535* (Figure 3.8). The DNA fragments of the two putative obtained Δ*pyrE*Δ*CAC3535* mutants were extracted and confirmed by Sanger sequencing.

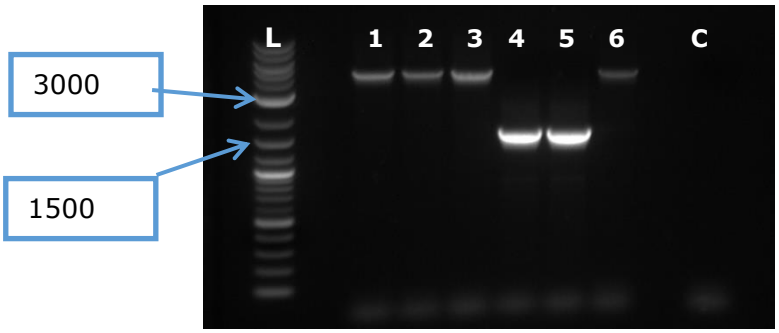

Figure 3.8: PCR screening of putative *CA_C3535* deletions generated in the Δ*pyrE* strain. L represents NEB ladder 0.1-10 kbp. Lanes 4 and 5 show expected DNA fragment sizes of 1,370 bp, corresponding to a *CA_C3535* deletion while lanes 1, 2, 3 and 6 shows a DNA fragment of 4,352 bp, corresponding to the wild type DNA fragment size. C is the negative control without inclusion of DNA.

3.3.5 Determination of the Transformation Efficiency of all Mutants Created

In order to test the hypothesis that knocking out the host's restriction modification systems would improve *C. acetobutylicum* Δ*pyrE* strain transformation efficiency including the acceptance of unmethylated plasmid DNA, all the RM mutants constructed in this study (Figure 3.9) were transformed with the replicative plasmids, pMTL85141 (created by Dr. John Heap, University of Nottingham UK) and pMTL85141 which was methylated *in vivo* by co-expression of pAN2-encoded methylase in *E. coli*.

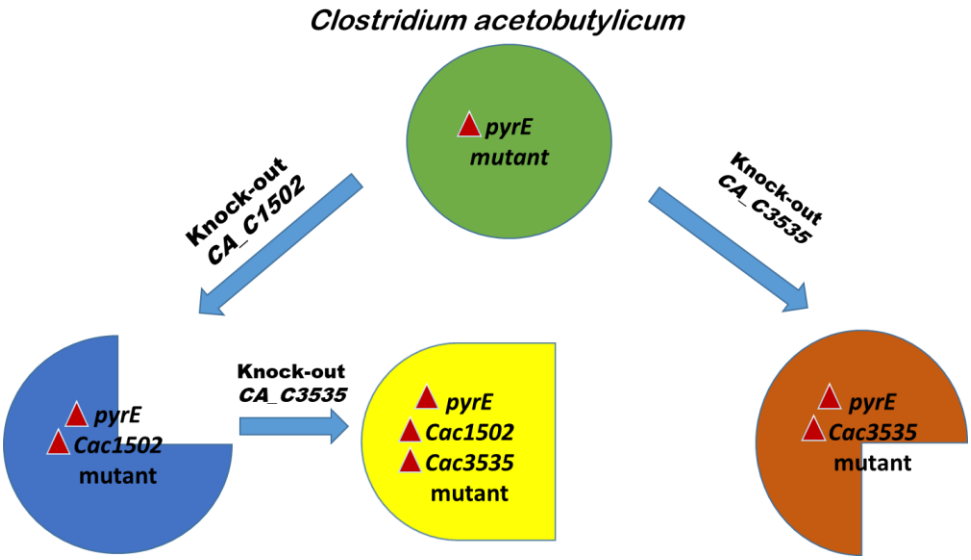

Figure 3.9: All single and double restriction-less mutants of the *Clostridium acetobutylicum* Δ*pyrE* strain created by in-frame deletion using a suicide plasmid.

Competent cells of the Δ*pyrE* strain, Δ*pyrE*Δ*CAC1502*, Δ*pyrE*Δ*CAC1502*Δ*CAC3535* and Δ*pyrE*Δ*CAC3535* mutants were prepared (see Materials and Methods 2.1.5.2) and electroporated in triplicates with 7500 ng of pMTL85141 propagated in *E. coli* (unmethylated) and pMTL85141 methylated in *E. coli* TOP10 pAN2.

When Δ*pyrE* was electroporated with unmethylated pMTL85141, there were no transformants generated. This is because the *E. coli* vector has many *G+C* rich or Cac824I restriction sites which are easily attacked. However, electroporation of the Δ*pyrE*Δ*CAC1502* mutant produced transformants in media containing thiamphenicol which showed that knocking out *CA_C1502* gene enabled *C. acetobutylicum* to accept unmethylated plasmid DNA. This agrees with the research by Dong *et al*. (2010) who reported that by knocking out the Cac824I restriction system using ClosTron insertions in *C. acetobutylicum*, unmethylated plasmid could be taken up by the bacteria. Similar findings were published by Ehsaan *et al*. (2016) and Croux *et al*. (2016). This showed that Cac824I was an active RM system in *C. acetobutylicum* that digested unmethylated DNA preventing its integration into the *C. acetobutylicum* chromosome.

On the other hand, the Δ*pyrE*Δ*CAC3535* mutant was also successfully transformed with unmethylated pMTL85141 plasmid (Table 3.2 and Figure 3.10) which also indicated that the type II restriction system Cac824II had a role to play in digesting unmethylated DNA. Though the *CA_C1502* gene encoding the Cac824I restriction system is labelled the main restriction system in *C. acetobutylicum* (Mermelstein and Papoutsakis, 1993), there is no published work reporting that deletion of *CA_C3535* mutant allows transformation of *C. acetobutylicum* of unmethylated plasmid.

Therefore, there may be no need to methylate any vector *in vivo* using *E. coli* co-harbouring pAN1 or pAN2 methylation donors before being introduced into Δ*pyrE*Δ*CAC1502* and Δ*pyrE*Δ*CAC3535* mutants. Eliminating the methylation step in the transformation process saves time and overcomes the problem of *Clostridium* genes which may be toxic if expressed in *E. coli* (Ehsaan *et al.*, 2016). The low transformation efficiency of Δ*pyrE*Δ*CAC3535* when transformed with unmethylated replicative plasmid as compared with that of Δ*pyrE*Δ*CAC1502* and Δ*pyrE*Δ*CAC1502*Δ*CAC3535* may suggest that it may not be ideal for gene studies using suicide plasmids (Table 3.2). These findings support the view of Cac824I being the main RM in *C. acetobutylicum*.

Table 3.2: Transformation efficiencies for *C. acetobutylicum* RM mutants in cfu/ µg plasmid DNA (10^3).

Δ*pyrE* Meth	Δ*pyrE* Unmeth	Δ*pyrE* Δ*CAC3535* Meth	Δ*pyrE* Δ*CAC3535* Unmeth	Δ*pyrE* Δ*CAC1502* Meth	Δ*pyrE* Δ*CAC1502* Unmeth	Δ*pyrE* Δ*CAC1502* Δ*CAC3535* Meth	Δ*pyrE* Δ*CAC1502* Δ*CAC3535* Unmeth
6.49	0	6.98	0.25	11.72	15.69	14.09	36.89
±2.94	±0	±2.63	±0.28	±3.89	±7.25	±4.51	±7.81

n=3, plasmid concentration =7.5 µg

Meth means methylated pMTL85141 while Unmeth means unmethylated pMTL85141.

In the first row, the mean cfu/µg plasmid DNA (10^3) of the transformation efficiencies of *C. acetobutylicum* RM mutants in triplicates is shown. The second row indicates the standard deviation of the RM mutants' transformation efficiencies in triplicates in cfu/µg plasmid DNA (10^3).

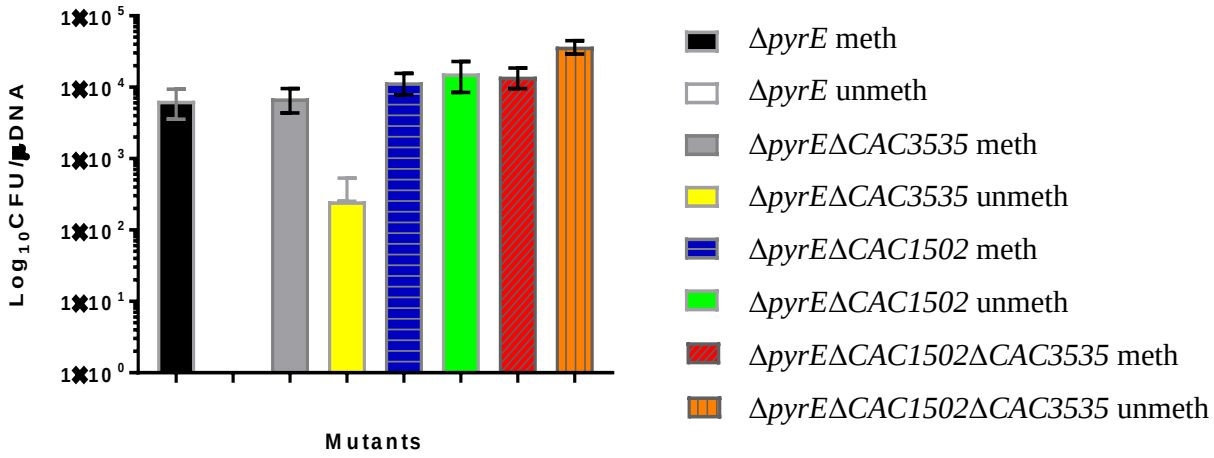

Figure 3.10: Transformation efficiencies of *C. acetobutylicum* RM mutants as number of transformants/µg plasmid. Meth means methylated pMTL85141 plasmid and unmeth means unmethylated pMTL85141 plasmid. All the single and double restriction-less mutants created in this study were transformed with methylated and unmethylated plasmid pMTL85141.

The data represented in Table 3.2 and Figure 3.10 shows that the transformation efficiencies of Δ*pyrE*, Δ*pyrE*Δ*CAC3535* and Δ*pyrE*Δ*CAC1502* with methylated plasmid were similar and the unpaired t-test done using Graph Prism 8 indicated that there was no significant difference ($p>0.05$) between them. The transformation efficiencies of Δ*pyrE*, Δ*pyrE*Δ*CAC3535* and Δ*pyrE*Δ*CAC1502* with methylated plasmid were 6.49 (±2.94) x 10^3 cfu/µg, 6.98 (±2.63) x 10^3 cfu/µg and 11.72 (±3.89) x 10^3 cfu/µg (Table 3.2). Similar result was found for wild type with methylated and

Δ*CAC1502* using unmethylated plasmid by Ehsaan *et al*. (2016) and Croux *et al*. (2016). Though the p value of the transformation efficiency of Δ*pyrE*Δ*CAC1502* with methylated and unmethylated plasmid was high showing that there was no significant difference, there was a slight increase in the transformation efficiency when the Δ*pyrE*Δ*CAC1502* was transformed with unmethylated compared to methylated plasmid (Table 3.2). Previous works by Dong *et al.* (2010) and Croux *et al*. (2016) also reported a slight increase in transformation efficiency for the single mutant when transformed with unmethylated plasmid. This could be due to the fact that methylation of a plasmid adds extra mass/weight to the DNA strand and since the nanodrop instrument used determines the concentration of DNA based on absorbance of UV light, this means that the actual amount of methylated DNA will be less than unmethylated DNA even when they are the same concentration measured on a nano drop. Dong *et al.* (2010) suggested that the increase in transformation efficiency for *C. acetobutylicum* SMB009 Δ*CAC1502* mutant with unmethylated may be because it is difficult to ascertain the absolute amount of methylated pIMP1 as it is a mixture of pIMP1 and pAN1. Thus the actual amount of methylated pIMP1 is less than the measured plasmid DNA from *E. coli*.

Transforming Δ*pyrE*Δ*CAC1502*Δ*CAC3535* with unmethylated plasmid gave a 2.4-fold increase when compared to Δ*pyrE*Δ*CAC1502*. The creation of a completely restriction-less mutant *(*Δ*pyrE*Δ*CAC1502*Δ*CAC3535*) gave an almost 6-fold increase in transformation efficiency for unmethylated pMTL85141 when compared to the *C. acetobutylicum* Δ*pyrE* strain. Croux *et al*. (2016) reported an 8-fold increase in transformation efficiency using unmethylated plasmid with the knock-out of *CA_C3535* in Δ*CA_C1502* strain which is in agreement with the data presented here.

The results of this study agree with the prediction by REBASE (Roberts *et al*., 2007) that *CA_C1502* is the main restriction modification system in *C. acetobutylicum* ATCC 824 and not *CA_C3535*. Though Croux *et al*. (2016) used a replicative plasmid for the deletions of the two RM systems, this report shows that a suicide plasmid can achieve the same deletions. By increasing the genetic accessibility of *C. acetobutylicum* ATCC 824, the inconvenience of methylating plasmid *in vivo* using *E. coli* before transformation is removed, while also making large scale transposon mutagenesis studies with suicide vectors more attractive.

3.3.6 Comparing the Methylation System in R702 with TOP10 pAN2 Methylation System

Aside from improving the transformation efficiency by knocking out the restriction modification systems, a search for an improved methylation donor strain was also undertaken. Two methylation donors consisting of poly(A) nuclease 2 (pAN2) from the φ3TI methyltransferase gene of *B. subtilis* phage Φ3TI, were compared. The first methylation donor was *E. coli* reInterStellar carrying the R-factor R702 into which had been inserted the φ3TI methyltransferase gene carried by plasmid pAN2, replacing the R702 spectinomycin resistant gene, *aadA1*. This host was constructed by Dr. Rob Mansfield, University of Nottingham UK. R702::*Φ3T* has the antibiotic markers conferring resistance to apramycin (100 µg/ml) and kanamycin (50 µg/ml). It also has an arabinose inducible system with *araC-*p_{BAD} upstream of φ3TI methyltransferase gene and apramycin marker downstream of the methylase. However, since *E. coli* reInterStellar is very prone to mutations (due to being a *dcm dam* double mutant), the goal was to introduce the R702::*Φ3T* into *E. coli* TOP10 and use this methylation donor to protect *E. coli-C. acetobutylicum* shuttle vectors. The second methylation donor strain used was *E. coli* TOP10 containing pAN2 (Heap *et al*., 2007) from the SBRC Nottingham culture collection. It is based on plasmid pMTL007 and harbours the φ3TI metyltransferase gene of *B. subtilis* phage φ3TI which, however, is tetracycline resistant and chloramphenicol sensitive (Heap *et al*., 2007). This served as the control methylation vector since it has been used to protect pMTL85141 as described in the previous section.

The reason for carrying out this research was to first compare the R702::*Φ3T* methylation donor with control pAN2 methylation donor from the culture collection. This would give an insight into which donor plasmid performs better and which was important for the following TraDIS research to be done on wild type *C. acetobutylicum* ATCC 824. Secondly, if a tetracycline inducible transposon delivery vector is to be used for TraDIS studies in *C. acetobutylicum*, propagating *E. coli* containing the control pAN2 tetracycline marker and tet-system transposon delivery plasmid in media containing tetracycline may lead to early transposition in *E. coli* or an unstable plasmid. This may negatively affect its transformation into wild type *C. acetobutylicum*.

3.3.7 Electrotransformation Efficiency of Methylation Donor Strains Carrying R702::*Φ3T* or pAN2

E. coli reInterStellar habouring the R702::*Φ3T* vector with kanamycin and apramycin markers was cultured at 37°C to lose temperature sensitive pKD46-cm. Vector pKD46-cm was previously used for recombineering the *Φ3TI* methyltranferase into pR702 and since recombineering was no longer needed, it was desirable to eliminate it from the *E. coli* reInterStellar strain.

E. coli reInterStellar was miniprepped and the extracted plasmid (R702::*Φ3T*) was electroporated into the *E. coli* TOP10 strain. The colonies formed on LB containing kanamycin and apramycin were screened using primers LHAextF (338) and LHAextR (244) (Figure 3.11).

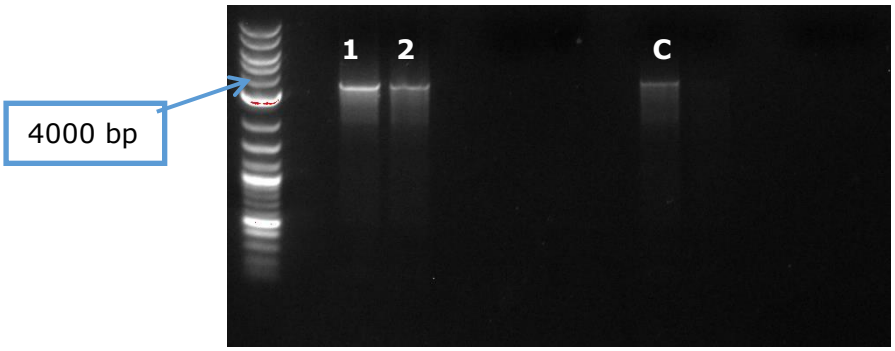

Figure 3.11: PCR screening of TOP10 colonies for R702 methylase using primers 338 and 244. Lanes 1 and 2 are both PCR products of screening two transformant colonies containing R702 while plasmid C is positive control R702 from *E. coli* reInterStellar. The DNA fragments in lanes 1 and 2 matched the expected size 3,804 bp.

Competent cells of the confirmed TOP10 R702 were made in the presence of 10 mM L-arabinose to activate the methylase cassette and they were subsequently transformed with pMTL85141. After extracting the putative protected pMTL85141, they were tested for protection from Cac824I restriction endonuclease by digesting 1 µg of plasmid DNA with SatI (Fnu4HI) restriction recognition enzyme as shown in Figure 3.12 below.

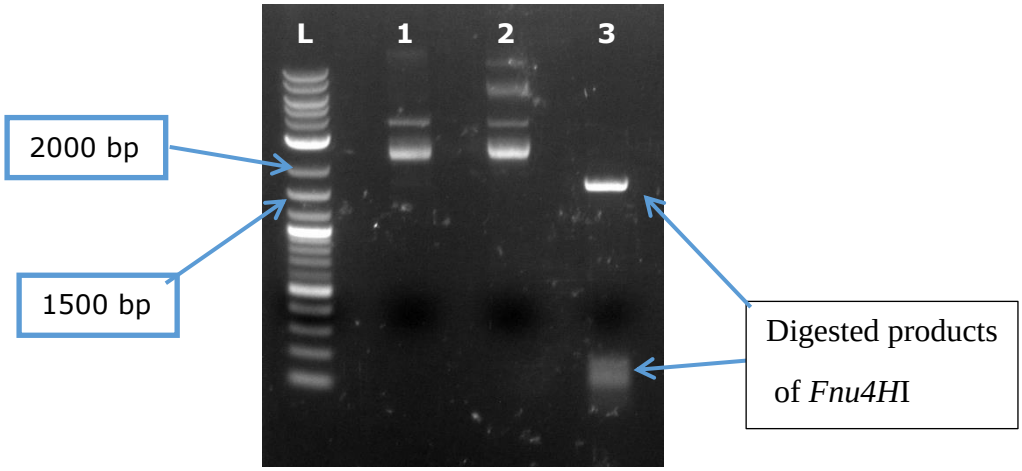

Figure 3.12: Gel image showing SatI (Fnu4HI) treatment of R702::*Φ3T* protected pMTL85141, pAN2 protected pMTL85141 and unmethylated pMTL85141. L represents NEB ladder containing 0.1-10 kbp. Lane 1 represents pMTL85141 methylated *in vivo* by R702::*Φ3T*, lane 2 is pMTL85141 methylated *in vivo* by control pAN2 while lane 3 represents the digestion products of unmethylated pMTL85141. No plasmid digestion is observed in lanes 1 and 2 with the expected DNA fragment size of 2,963 bp, lane 3 has the expected digestion products of unmethylated pMTL85141: 1736 bp (large DNA fragment) and a number of smaller DNA fragments sizes ranging from 3 to 206 bp.

The result displayed in Figure 3.12 indicated that the methylated pMTL85141 was immune to digestion by Fnu4HI restriction enzyme from *Fusobacterium nucleatum* which recognises the same site and is equally inhibited by methylation; it is an isoschizomer of Cac824I. This result means that *in vivo* methylation of pMTL85141 with R702::*Φ3T* was successful. There were 15 restriction recognition sites for 5' GCNGC 3'. The digested products expected were of 1736, 206, 179, 155, 143, 118, 80, 74, 73, 65, 56, 51, 18, 3, 3 and 3 base pairs in length. The blurred DNA fragments at the bottom of the gel image in Figure 3.12 corresponded to the 15 small digested fragments of the plasmid.

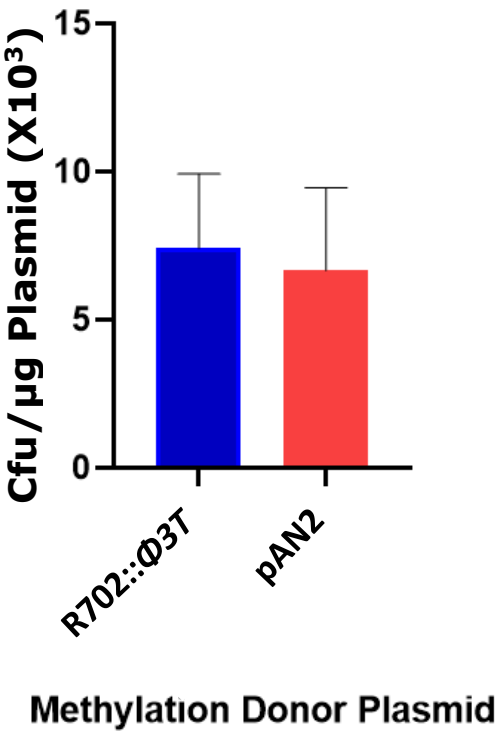

Figure 3.13: Electrotransformation efficiency of plasmids prepared in *E. coli* TOP10 methylation donors carrying R702::*Φ3T* or pAN2. Both *E. coli* strains containing pMTL85141 were grown in LB broth in the presence of apramycin plus 10 mM L-arabinose for R702::*Φ3T* and chloramphenicol plus tetracycline for control-pAN2. Data shown for *E. coli* carrying either R702::*Φ3T* or pAN2 represents the average values for four independent transformations of *C. acetobutylicum* wild type. Standard error of the mean of each sample is represented by error bars in the graph.

Result shown in Figure 3.13 displays a comparison of the transformation efficiencies in cfu/µg of pMTL85141 prepared in the methylation donors carrying either R702::*Φ3T* or pAN2. The unpaired t-test performed using Graph Prism 8 indicated that there was no significant difference ($p>0.05$) between the DNA prepared in either donor.

3.4 Conclusion

Cac824I is said to be the main barrier to transformation of *C. acetobutylicum* with *E. coli-C. acetobutylicum* shuttle plasmids (Croux *et al*., 2016, Mermelstein *et al*., 1992), the results of this current study support this statement. While there are some publications reporting the in-frame deletions of *CA_C1502* gene (Cac824I), the transformation efficiency of double restriction-less mutant of *C. acetobutylicum* was evaluated by Croux *et al.* (2016). They recorded an improvement of 8-fold increase with unmethylated plasmid in transformation efficiency when *CA_C3535* was knocked out in Δ*CA_C1502*. However, they did not report whether a single Δ*CA_C3535* mutant

could be transformed by unmethylated plasmid. This study shows that the *CA_C3535* mutant can be transformed with unmethylated pMTL85141 although the transformation efficiency decreased 28-fold compared to transformation with methylated plasmid.

Transformation with unmethylated plasmid also gave a slightly higher transformation efficiency for Δ*pyrE*Δ*CAC1502* and Δ*pyrE*Δ*CAC1502*Δ*CAC3535* mutants than with methylated plasmid. The *E. coli* TOP10 methylation donor strain co-harbours pAN2 in addition to the plasmid to be methylated, in this case pMTL85141. This means the product of the miniprep was a mixture of pAN2 and methylated pMTL85141 and may be responsible for the observed lower transformation efficiency in methylated plasmids. Ultimately, knocking out *CA_C3535* gene from *CA_C1502* mutant gave almost 2.4-fold increase in transformation efficiency which is comparably less than the 8-fold increase observed by Croux *et al.* (2016) for knocking out *CA_C3535* from the *CA_C1502* mutant. In addition, knocking out both *CA_C1502* and *CA_C3535* gave almost 6-fold increase when compared to the transformation efficiency of the parent strain.

Apart from the importance of improving the transformation efficiency of the strain, there were plans to repair *pyrE* while inserting *tcdR* at the *pyrE* loci. Insertion of the *tcdR* gene into the genome of *C. acetobutylicum* was required in order to use transposon delivery plasmids in which the transposase was under the control of the P_{tcdB} promoter. Expression of the sigma factor in a cell transformed with the transposon vector will drive the transposition activity. Furthermore, comparing the electroporation efficiency of the two methylation donor plasmids was undertaken in search of a better methylation donor for the transposon delivery vectors in chapter 4. Although there was no significant difference in transformation efficiency between the *E. coli* methylation donors carrying either pR702::*Φ3T* or pAN2 (Tet), the former has the advantage of an alternative antibiotic marker to tetracycline for methylating anhydrotetracycline-inducible transposon delivery plasmids for TraDIS studies.

Key points about this chapter

- Knocking out both *CA_C1502* and *CA_C3535* gave almost 6-fold increase compared to the transformation efficiency of the parent strain
- There was no significant difference in the transformation efficiencies of plasmid prepared in *E. coli* methylation donors carrying either pR702::*Φ3T* or pAN2

Chapter 4

Investigating essential and non-essential genes using TraDIS

Chapter 4

4.1 Introduction

Forward genetic studies are used for random gene inactivation to elucidate the exact role of specific genes implicated to a particular phenotype (Cartman and Minton, 2010). Transposon Directed Insertion-site Sequencing (TraDIS), a type of forward genetic approach, has been applied to identify genes essential for sustaining the cell viability in *Samonella typhi* (Langridge *et al*., 2009), *Clostridium difficile* (Dembek *et al*., 2015) and *Burkholderia pseudomallei* K96243 (Moule *et al*., 2014). There are limited works on forward genetic studies in *C. acetobutylicum* (Zhang *et al*., 2016). In addition to this, there has not been any published TraDIS research used to identify essential genes in saccharolytic clostridia. There is a need for research in this area because this approach can be applied in unveiling novel genes and genetic information that gives invaluable insights to understanding the metabolism for biotechnological and commercial purposes. Besides this, TraDIS can be applied in identifying genes that are dispensable and can thus be knocked out from the genome.

4.1.1 Transposon Mutagenesis Versus Chemical Mutagenesis

Transposons, also known as jumping genes, are sequences of DNA that can move from one site in a genome to another, which can result in loss-of-function mutations. Transposase acts like DNA scissors; it is an enzyme that cuts through double-stranded DNA to remove the transposon. Most transposons are non-replicating and thus most use a cut–and-paste rather than copy–and-paste mechanism. Typically the composite transposon consists of a transposase gene flanked by inverted terminal repeats (ITR) which are recognised, cleaved and transported to a target site on a genome by the enzyme. Since they are naturally occurring and are found in all organisms from bacteria to eukaryotes, they have contributed to the evolution of many genomes. Transposons are also employed as genetic tools in introducing foreign DNA into a target genome (Muñoz-López and García-Pérez., 2010).

In the late 1940s, Barbara McClintock discovered the concept of transposon when she observed chromosome breaks in maize that can change their position on the chromosome. Depending on the area of insertion, she observed that their insertion can control the expression of maize genes. Today transposon studies have been incorporated into the study of various genetic networks in vast areas of molecular

biological studies (Biémont, 2010). Transposon mutagenesis tools are applied in the identification of genes and/or pathways for diseases, studies on development of embryos, regulatory genome studies as well as gene therapy (Muñoz-López and García-Pérez., 2010). Though research using chemical mutagens was conducted earlier than transposon studies, the later has some advantages in its application as a genetic tool over the former. The advantages of transposon mutagenesis over chemical mutagenesis include: (1) changes caused by chemical mutagens might be more difficult to track/locate while sequencing of mutants can be used to identify the loci of transposon in a genome; (2) single mutations per bacterial cell can be created using transposon mutagenesis but it is often not the case for chemical mutagenesis; (3) inserting antibiotic marker in the transposable element/composite could allow selection of mutants from wild type while it might be difficult to separate in chemical mutagens; (4) also a useful gene can be inserted and expressed in transposon mutants (Siegrist and Rubin, 2009).

Taking into consideration the advantages and disadvantages of transposons in the study of genomes, an ideal transposon should have the following qualities; does not excise and translocate locally (highly stable), has high efficiency of insertion, can exit without leaving remnants and also has a high cargo capacity (Muñoz-López and García-Pérez., 2010). Other advantages include its application in identifying multiple genes associated with niche adaptations such as virulence (Wong *et al*., 2016) and identifying novel factors and genes involved in genetic network interaction/regulation (Freed, 2017) such as sporulation.

On the other hand, the potential disadvantages of transposon studies are low stability or low mutation in the host and it may lead to polar effects. Creating large libraries may be labour intensive and multiple insertions may occur which may make characterisation difficult (McCurrie, 2011). It is very important that for the tools to be carefully selected for transposon mutagenesis studies to overcome these challenges.

4.1.2 Transposon Mutagenesis Studies

Early forms of transposon mutagenesis studies used for some species of *Clostridium* were conjugative transposons. Blouzard *et al*. (2010) developed a functional Tn*1545* transposon mutagenesis system for *C. cellulolyticum.* This improved transposon system consists of conjugation with two vectors; one vector carried the integrase-encoding gene Tn*916,* whereas the other is a suicide vector which carried the attachment site of

Tn*1545*. Transposon mutants were generated and shown to have random insertion. Blouzard *et al*. (2010) proposed that this technique can be used to identify essential genes involved in plant cell wall degradation for the development of new approaches for strain engineering. Conjugative transposon Tn*5397* was also designed for gene delivery in *C. difficile* (Mullany *et al*., 1990) since the only means of transfer of recombinant DNA for this organism is through conjugation (Cartman and Minton, 2010). An earlier work by Babb *et al*. (1993) used conjugative transposons Tn*916* and Tn*925::916* and created a transposon mutagenesis system for the industrial strain *C. acetobutylicum* P262. Babb and his co-workers (1993) isolated and characterised two mutants which were sporulation deficient and resistant to metronidazole as well as a mutant lacking the ability to produce butanol and acetone. They also observed that the Tn*925::916* transposon gave multiple inserts. Moreover, Wang *et al.* (2006) reported that Tn*5397* had preferred insertion sites while Hussain *et al*. (2005) reported that Tn*916* gave multiple inserts in a single transposon mutant and these pose a big challenge for transposon studies which depends on the generation of pools of individual random mutants.

The disadvantages of inserting multiple copies in the same chromosome (Liu *et al*., 2013) and the large size result in low efficiency of the transformation process (Heap *et al*., 2009b) making conjugative transposon plasmids less attractive. Later on, Mu phage was introduced as a shuttle vector into *C. perfringens* to make single transposon insertions of which 43% were in rRNA regions. The low number of transformants generated makes it unattractive for transposon studies. A third and more efficient random mutagenesis system which is Ez Tn*5*-based was developed which has lower occurrence of insertion in rRNA regions although it is still as high as 18% compared with the Mu phage (Liu *et al*., 2013). Generally, these systems require a lot of time to create many mutants and there is need for the exact gene annotations because of insertions at non-coding regions like rDNA (Barquist *et al*., 2013).

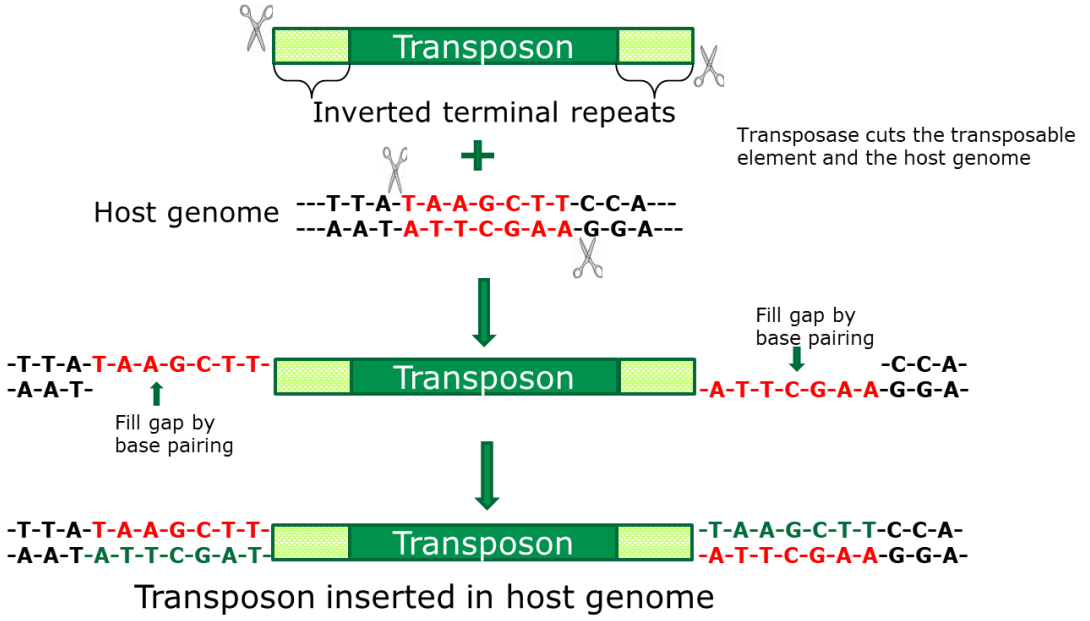

Figure 4.1: Illustration of the mechanism for cut-and-paste transposition. This was dapted from Muñoz-López and García-Pérez., 2010.

Ideally, a good transposon tool should be capable of high transposition rate and high non-specific insertion efficiency. *Himar1* transposon is from the mariner family and was originally isolated from *Haematobia irritans*, the horn fly (Lampe *et al*., 1999). It has two domains; a carboxyl domain for catalytic activity and a helix-turn-helix motif for recognising and binding to the inverted terminal repeats (ITR). The C9 variant derived from point mutations of *Himar* transposase became widely used in research because of its high transposition activity and cut-and-paste mechanism (Figure 4.1) (Muñoz-López and García-Pérez., 2010). Based on these advantages, Cartman and Minton (2010) designed a novel *mariner*-based transposon system to study *C. difficile* R20291 epidemic strain. *Himar1* C9, the mariner-transposase used for the study, has insertion preference for TA sequences which are common in clostridia. Thus, the insertion site requirement of *Himar1* C9 is an advantage for studies in TA-rich *C. difficile*. The transposon element functions by inverted terminal repeats on both ends which inserted at TA target sites (Cartman and Minton, 2010). Cartman and Minton (2010) reported that 98.3% of mutants carried single random inserts, which was an improvement from previous conjugative transposon systems (Hussain *et al*., 2005) which had multiple insertions. Although no new genes were identified, the experiment was applied in isolating sporulation defective mutants with insertions at *cspBA* encoding germination protease genes as well as an auxotrophic mutant with insertion

at *pyrB*, which encodes for pyrimidine biosynthesis protein (aspartate carbomoyltransferase chain).

A plasmid-based *mariner* transposon system was also designed and employed to investigate gliding motions in *C. perfringens* (Liu *et al*., 2013). Liu *et al.* (2013) used a transposase controlled by an inducible promoter, a replicating plasmid which had a negative selection marker to indicate elimination of transposon plasmid as soon as transposition event had taken place. In another research work, a *mariner* transposon system was designed for *C. acetobutylicum* ATCC 824 and other *Clostridium* spp. (Zhang *et al*., 2015). It is known as a universal *mariner* transposon system in which the transposition events are controlled by an IPTG inducible promoter. Thus, the inclusion of IPTG in the growth media controlled transposition events such that random inserts are produced. *Mariner* transposon mutagenesis tackles some of the problems experienced with conjugation transposon but there is still a need for high-throughput screening which can save time and give more comprehensive data on gene essentiality and fitness.

4.1.3 Previous Studies on Gene Essentiality and Conditional Studies Using TraDIS

Horizontal gene transfer between bacteria and phages has contributed to the expansion of genes in the genomes of bacteria. Most free living organisms have larger genomes than those organisms living in restricted ecological niches. This is because evolution and the molecular interactions of such organisms in extreme/restricted niches do not select for the maintenance of unused genes. However, there are a series of genes which are hardly lost nor gained because they encode proteins that are responsible for the survival of the cell under specific conditions, known as essential genes. Essential genes are of interest in research because they are required for the basal bioprocesses for growth in media (Fang *et al*., 2005).

Transposon mutagenesis is one of the approaches used in determining the essential genes in the bacterial genome. For example, it has been applied in identifying 620 essential genes in *E. coli* (Gerdes *et al*., 2003), 670 essential genes in *Haemaphilus influenzae* (Akerley *et al*., 2002), 265–350 genes from *Mycobacterium genitalium* (Hutchison *et al*. 1999), 400 essential genes from *Pseudomonas aeruginosa* (Jacobs *et al*., 2003), 383 essential genes from *Burkholderia cenocepacia* strain J2315 (Wong *et al*., 2016) and 404 essential genes in *C. difficile*

(Dembek *et al*., 2015). Essentiality of a gene is often relative to variable experimental conditions and output. The essential genes for survival of cells in laboratory conditions with abundant nutrients may be different from its survival in competition with other organisms for limited nutrients or stress conditions. Stress is another factor that could affect the fitness of mutating genes essential for survival in a particular experimental condition. Furthermore, a gene might not be essential for growth under standard laboratory conditions but deletion of that gene could negatively affect the fitness of the mutant (Fang *et al*., 2005).

Langridge *et al*. (2009) developed a high-throughput sequencing system which will exploit the advantage of genomic data explosion for functional studies. It is a high-throughput technology named, "TraDIS" which means Transposon Directed Insertion-site Sequencing. Using Tn*5* derivative, a large pool of mutants of about 1.1 million individuals is generated into libraries such that the non-essential genes are saturated with transposon elements. Then the genome of the mutant pool is extracted for Illumina sequencing. Illumina sequencing generates 100 bp sequence reads from the large DNA population. This technology will allow high-throughput sequencing of all the mutants in one experiment. The principle behind this is the shearing of genomic DNA at random and ligation of oligonucleotide linkers, then the fragments are amplified using universal primers, finally the complementary oligonucleotide sequencing primers are employed in sequencing-by-synthesis of the amplified fragments. The sequencing primer anneals to the transposon, hence sequencing the region around it allows for the determination of the genes with insertion. Genes without transposon insertions are likely to be essential for survival of the organism as illustrated in Figure 4.2. It was applied in mapping out 370,000 individual transposon inserted-sites in *Salmonella enterica* serovar Typhi strain CVD 908-*htrA* strain (Langridge *et al*., 2009). Transposon Directed Insertion-site Sequencing has since been applied to various organisms and transposon studies in a consortium of *in vivo* and *in vitro* conditions (Barquist *et al*., 2016).

In addition to gene essentiality studies, TraDIS is applied in identifying genes that contribute to survival under a particular stress condition (see Figure 4.2). Conditional essential studies for resisting high concentrations of bile salt were done for *Salmonella enterica* serovar Typhi strain CVD 908-*htrA, typi* (Langrigde *et al*., 2009). This was

done by comparing the TraDIS sequencing data of simultaneously cultured highly dense transposon pool with and without ox bile salt. Langrigde *et al.* (2009) identified 169 genes that contribute to bile salt tolerance including 30 hypothetical genes and regulatory genes which had not been previously linked with bile tolerance in literature. Research such as this contributed to insights on the gene function for bacterial lifestyle.

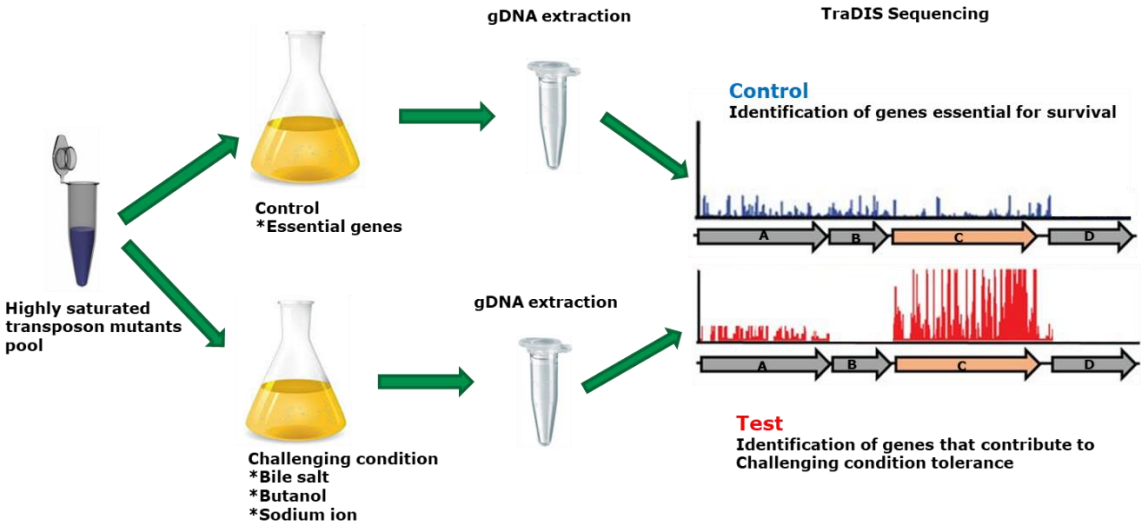

Figure 4.2: Illustration of TraDIS strategy for investigating gene essentiality and conditional gene studies. A large pool size of library mutant ensures high genomic saturation to enable high statistical prediction of essential genes. TraDIS control result shows that insertion in gene D was fatal, thus it is an essential functional gene while TraDIS test condition show that insertion in gene B was fatal under a particular condition, thus suggesting it is essential for tolerance in bile salts, butanol or sodium ion condition.

Finally, the advantages of TraDIS studies on an organism cannot be over-emphasised. In the development of commercial industrial strains, it will be interesting to know the minimal genes required for basal bioprocesses such that superfluous cargo genes can be dispensed with. There are a lot of hypothetical genes with unknown function in *C. acetobutylicum* and TraDIS could help in gene functionality assignment. Besides, limited understanding of the physiology, genetics and biochemical activities of saccharolytic bacteria are part of the factors impeding the proliferation of the ABE industry. Forward genetics involving high-throughput analysis has been shown to give faster and quicker information about large scale interrogation of genes.

4.1.4 Butanol Tolerance Investigations with *C. acetobutylicum*

Accumulation of toxic substances that inhibit butanol production is one of the hindrances to high titre in the ABE pathway (Jones and Woods, 1986). The typical ABE pathway for *C. acetobutylicum* produces acetone, butanol and ethanol at the ratio of 3:6:1, which shows that butanol is produced at a higher ratio than other products (Bankar *et al*., 2013). Butanol has been shown to be one of the main inhibitory substances in the ABE pathway (Liyanage *et al*., 2000; Bankar *et al*., 2013). High butanol concentrations affect the fatty acids and phospholipids components of the *C. acetobutylicum* cell membrane (Bankar *et al*., 2013). Toxic levels of butanol inhibit glucose uptake, nutrient transport and ATPase activity in the cell membrane which leads to reduction in cell energy (Tomas *et al*., 2003). Papoutsakis *et al*. (2008) proposed that the cost of production in ABE fermentation will be cut to half if the butanol produced is raised to 19 g/l. This showed the importance of using butanol tolerant strains in ABE fermentation process.

Liyanage *et al*. (2000) published their investigations for butanol tolerance in *C. beijerinckii* BR54, a strain derived from the wild type strain. They conjugated transposon Tn*1545* to create transposon mutants and cultured them in media containing 1.2% w/v of butanol. They reported that a single insert of transposon element in the intergenic region between *accC*, which encodes biotin carboxylase and *gldA*, which encodes glycerol dehydrogenase increased the strain's tolerance to butanol and its titre. A study on isopropanol tolerance in *C. acetobutylicum* using Adaptive Laboratory Evolution (ALE) approach revealed three individual colonies that grew in isopropanol concentrations at 40 g/l. Sequenced data of the three strains revealed the presence of 26 mutations common to all. Two of these mutations were in *CA_C3368*, putative permease and *CA_C0437*, sensory transduction histidine kinase responsible for the phosphorylation and activation of Spo0A respectively (Foulquier *et al*., 2019).

In another study done by Tomas *et al.* (2004), the effects of 0.25% v/v and 0.75% v/v butanol concentrations on the transcriptional profile of *C. acetobutylicum* ATCC 824 with overexpressed *groESL* in pGROE1 and ATCC 824 with pSOS95del (empty plasmid not containing *groESL*) strains were analysed to investigate butanol tolerance/stress genes. Plasmid pGROE1 contains an overexpressed gene, *groESL*, encoding a class I heat shock protein from *C. acetobutylicum* while pSOS95del does not contain the *groESL* genes. When both DNA array transcriptional profiles were

compared, genes identified with opposite expression dynamics were considered to be associated with butanol tolerance. *artP*, *rlpA* and hermin permease encoding genes were observed to fall under the category of opposing expressions. Furthermore, all the main stress proteins as well as solvent forming genes had similar expression with pGROE1 which indicated that they were butanol stress response genes. Tomas *et al.*'s (2004) findings suggested that even when stress is caused by a solvent product, it could trigger the transcription of solventogenic genes.

Although improved butanol tolerant strains of *C. acetobutylicum* have been investigated by exposure to high doses of butanol and isolating resistant strains (Liyanage *et al*., 2000; Foulquier *et al*., 2019) and DNA array transcriptional studies for identifying butanol tolerance genes in *C. acetobutylicum* (Tomas *et al*., 2004), there has not been any work on the use of transposon mutagenesis for the identification of essential genes that contribute to butanol tolerance in *C. acetobutylicum*. Moreover, determination of essential genes of a saccharolytic bacteria has not been done using TraDIS. Thus, it would be interesting and vital to explore these essential genes in *C. acetobutylicum* as well as novel genes that when mutated can contribute to the butanol tolerance in sub-inhibitory levels. TraDIS presents a great tool for high-throughput screening of butanol tolerance, as mutants are not screened one at a time which can be time consuming and labour intensive. Producing and analysing millions of single random mutants in a single experiment will contribute to an effective and comprehensive gene functionality study. By observing the distribution of insertion sites of libraries when normalised for gene length, predictions of which genes are essential under certain conditions (butanol tolerance and sporulation deficient) can be made (Barquist *et al.*, 2016). Using a suicide plasmid for TraDIS will ensure total plasmid loss after transposition event, consequently, tackling the problem of plasmid reads being included in transposon sequence reads.

4.1.5 Sodium Ion Tolerance Investigations in *C. acetobutylicum*

Another major reason for the decline in ABE Industry was that feedstock/substrates (corn and molasses) were expensive (Jones and Woods, 1986). The cost of the feedstock contributes to 60-70% of the overall cost of ABE production (Qureshi and Blaschek, 2001). Alternative cost-effective substrates such as lignin, hemicellulose and cellulose (Lignocellulose) biomass in the form of agricultural wastes are in abundance. Biomass

from agricultural wastes are considered good sources of feedstock for ABE fermentation because of their high sugar content, high availability and low cost (Abo *et al*., 2019). However, they first need to be pretreated and then hydrolysed enzymatically to release the fermentable sugars because they are highly resistant to biological and chemical degradation. There are several researched pretreatment methods though they lead to the accumulation of a wide range of toxic compounds which may be expensive to remove and may cause a reduction in the fermentable sugars (Birgen *et al*., 2019).

Amongst the various pretreatment methods, alkali pretreatment using sodium hydroxide (NaOH) is most widely used because of its suitability and effectiveness for the fragmentation of various agricultural wastes and materials. Sodium hydroxide separates the bonds between the carbohydrates and the lignin, disrupts the lignin structure, decreases the degree of polymerisation and crystallinity and increases the internal surface area (Bensah and Mensah, 2013; Gao, 2016). Silverstein *et al*. (2007) reported that sodium hydroxide pretreatment yielded the highest level of lignin conversion (65.63% for NaOH 2% (w/v), 121°C /15 psi, 90 min) as well as cellulose conversion (60.8%). In another report, a combination of 2% (*w*/*v*) NaOH pretreatment and enzymatic hydrolysis of rice straw gave 0.6 g/g reducing sugar with 60.75% saccharification (Rahnama *et al*., 2014). Additionally, dilute NaOH treatment of substrate appears to be more effective in breaking the ester bonds and removal of lignocellulose (Gaspar *et al*., 2007).

The main drawback with using dilute NaOH pretreatment is that its accumulation is difficult to remove downstream, thus can hinder/inhibit the downstream biomass fermentation (Wang, 2012). Therefore, a neutralisation step is required to maintain the optimum pH for the enzymatic hydrolysis of lignocellulosic materials to be used subsequently for ABE fermentation. Alkalis or acids are used to neutralise and adjust the pretreated substrate to pH 5.0 and pH 6.5 for enzymatic hydrolysis and ABE fermentation respectively. However, the products of neutralisation, sodium salts, affect the efficiency of both processes (Yang *et al*., 2019). Many studies have demonstrated the inhibitory effects of sodium salts on ABE fermentation (Yang *et al*., 2019; Zhao *et al*., 2016).

Yang *et al*. (2019) showed that sodium inhibited the ABE fermentation by decreasing the solvent concentration from 12.8 g/l to 0–10.7 g/l at various concentrations. Zhao *et*

al. (2016) proclaimed that high sodium ion conditions postpone cell growth, affect the final biomass and slow down the growth rate of *C. acetobutylicum* ATCC 824. These negative effects on cells are as a result of inducing high osmotic pressure on cell membrane which may lead to the cell resources being channelled towards controlling membrane transporters. Zhao *et al*. (2016) observed a lag in cell growth and decrease in solvent production when *C. acetobutylicum* was grown in the presence of sodium chloride. Butanol production decreased from 16.90 ± 0.12 g/l to 11.26 ± 0.17 g/l representing a 34% reduction.

Using a sodium salt tolerant strain for ABE fermentation would be beneficial when feedstocks are low cost substrates like agricultural wastes. Identifying genes that contribute towards sodium ion tolerance will be a step in the right direction for creating a sodium ion tolerant strain.

4.1.6 Suicide Versus Conditional Transposon Plasmid

The preferred plasmid for transposon studies is a suicide plasmid. A suicide plasmid is a delivery vector that can be taken up by the host but is incapable of replicating in *C. acetobutylicum* because it lacks a Gram-positive origin of replication. This means that when a cell is transformed with a suicide transposon delivery plasmid, the transposase is expressed and transposition of the transposable element into the genome takes place, however, the next generation of cells loses the plasmid. Plasmid loss is a great advantage in TraDIS study to avoid contamination of transposon reads and complicating sequence results. Another advantage of using a suicide plasmid is that each mutant colony contains independent transposon insertions. A disadvantage with the use of suicide plasmids for TraDIS studies is that the transformation frequency is low in most bacteria.

When pseudo-suicide transposon plasmids are used, the replication of the plasmids in cytoplasm is slower compared to that of the recipient chromosome, therefore there will be more representation of transposon mutants than cells containing plasmids (Cartman and Minton, 2010). Conditional replicative transposon plasmids are used as an alternative to suicide and pseudo-suicide transposon vectors (Zhang *et al*., 2015). The advantage of using conditional replicative plasmid for transposon delivery is that it does not depend on low electroporation efficiency as only one transformant colony carrying

the delivery plasmid under non-permissive conditions can be allowed to replicate. Subsequently, plasmid loss and the transposition event can be permitted in the daughter cells in the presence of the inducer (Guilhot *et al*, 2009).

4.1.7 Inducible Promoter System in *C. acetobutylicum*

When using a replicative plasmid, it is very important that the expression of the transposase is controlled such that the transposition event does not occur during the vector propagation in *E. coli* or immediately after transformation into the host. Controlling the expression of the transposase as well as the Gram-positive origin of replication will ensure that effective transposition and plasmid loss. Using a conditional plasmid for random mutagenesis studies will help eliminate the problem of early transposition occurring shortly after the host takes up the plasmid, which can lead to lack of randomness in the library mutant pool (Zhang *et al*., 2015).

Cartman and Minton (2010) used a pseudo-suicide plasmid pMTL-SCI containing a P_{tcdB} promoter to control the expression of the mariner transposase *Himar1* C9 and this was transferred by conjugation. The specialised sigma factor TcdR in *C. difficile* controls the transcription of toxin B. The promoter P_{tcdB} allows transposition to take place only in *C. difficile* and not in the *E. coli* during propagation of the vector. The major problem with using this plasmid was that as a pseudo-suicide plasmid with segregational instability, it required passaging for a few generations before the plasmid was lost. Another report by Zhang *et al*. (2016) developed a xylose inducible plasmid for random mutagenesis for *C. acetobutylicum*. They put the expression of the mariner transposase *Himar1* C9 under the control of promoter P_{araE} of the *C. acetobutylicum CA_C1339* gene (putative sugar-proton symporter, *araE*). Their data demonstrated that 91.6% of the colonies on the mutant selection plates had the transposon insert.

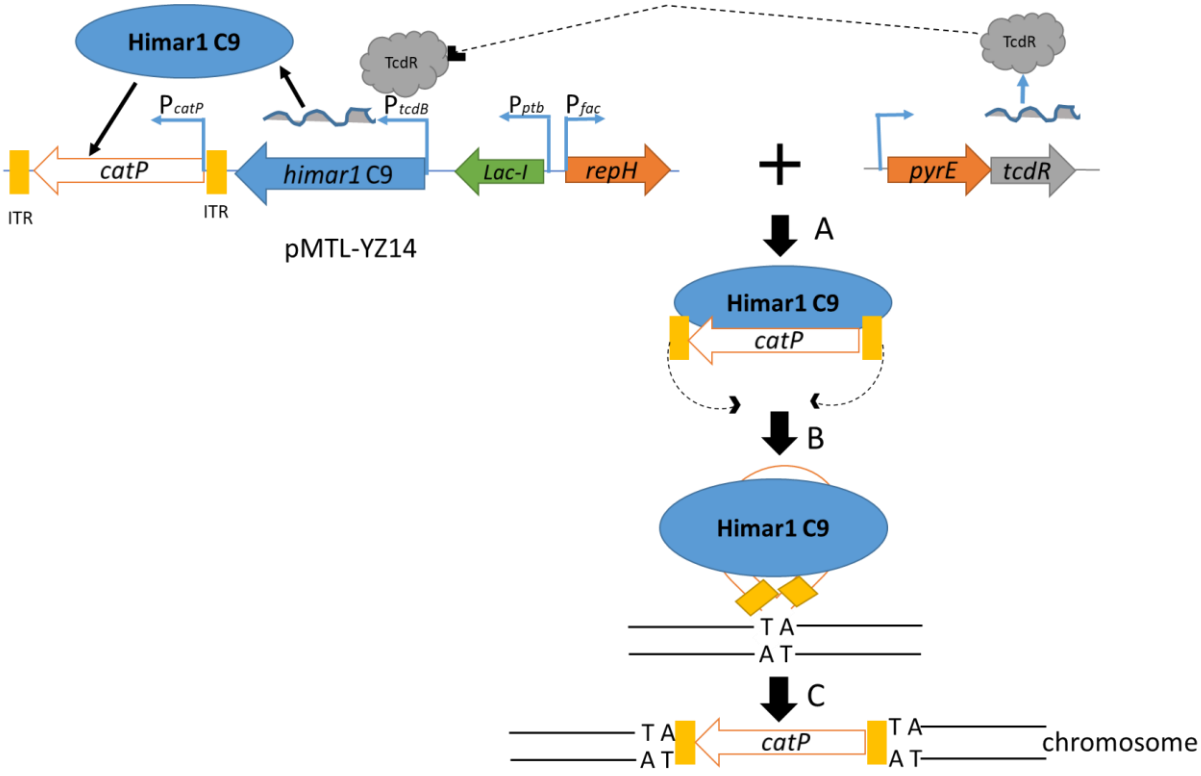

Figure 4.3: Mode of transmission of inducible *Himar1* C9 transposase in pMTL-YZ14. The transposon plasmid, pMTL-YZ14, contains a conditional plasmid stability system of *lacI* arrangement such that the LacI repressor protein blocks or allows the transcription from the promoter P_{fac} to Gram-positive origin of replication (*repH*). LacI binds to P_{fac} thereby blocking transcription from the strong promoter. The presence of IPTG lead to binding of the LacI repressor protein making it unavailable to block transcription through P_{fac}. pMTL-YZ14 also has a minitransposon of *catP* gene flanked by two inverted transposon repeats (ITR) and the *Himar1* C9 transposase put under control of the P_{tcdB}, such that it is only activated in the presence of *tcdR*. (A) *Himar1* C9 transposase is expressed and recognises the ITRs and binds to the minitransposon forming a transposase-transposon complex. (B) The excised complex recognises a -TA- sequence at a target site of the host chromosome. (C) The transposase causes a double strand break on the target site and inserts the minitransposon. Adapted from Zhang *et al*. (2015).

In a previous study, a β-D-1-thiogalactopyranoside (IPTG)-controllable conditional replicative transposon vector, pMTL-YZ14, was developed and shown to achieve random insertion in clostridia recipients (Zhang *et al*., 2015). The plasmid also has P_{tcdB} promoter controlling the expression of the mariner transposase. The Gram-positive origin of replication (*repH*) was put under transcriptional control of a strong promoter P_{fac} which caused plasmid instability due to increased transcription of the origin. Furthermore a conditional transcription for the origin of replication was created by

introducing the *lacI* gene from *E. coli* with a promoter P_{ptb} upstream of P_{fac}-repH origin of replication. The *lacI* gene encodes the LacI repressor protein that binds to the lacO region on the P_{fac} promoter, thereby blocking the transcription from P_{fac}. Therefore, the absence of IPTG should repress transcription of the Gram-positive origin of replication and ensure plasmid maintenance. The presence of IPTG drove plasmid loss in the host by preventing the LacI repressor protein from binding to P_{fac} thus, there is transcription of the origin from the strong promoter P_{fac} (as illustrated in Figure 4.3). However, sequencing of mutant library created with pMTL-YZ14 indicated the presence of plasmid contamination (Dr. Gareth Little, University of Nottingham UK; unpublished). Dr. Gareth Little created a suicide plasmid from pMTL-YZ14 of which he deleted the Gram-positive origin of replication, *repH*. The suicide plasmid named pMTL-GL15 has all the features of pMTL-YZ14 except *repH*. This suicide plasmid was choosen for this transposon study because it has the advantage of preventing contamination of transposon reads with plasmid reads.

Another inducible system that has not been explored in the literature in creating transposon mutants of *C. acetobutylicum* is the anhydrotetracycline (aTC) inducible system. However, the aTC inducible system has been explored in *C. acetobutylicum* for ß-glucuronidase gene expression (Dong *et al*., 2012), controlling the expression of CRISPR Cas9 for gene editing (Wasels *et al*., 2017) and inducible expression of autolysin genes (Yang *et al*., 2013).

4.2 Hypothesis and Aim of Chapter

The aim of this chapter is to demonstrate the use of TraDIS in *C. acetobutylicum* ATCC 824 in determining genes that are essential and non-essential for viability under standard laboratory conditions. This chapter also aims at demonstrating the use of TraDIS to elucidate genes responsible for butanol and sodium ion tolerance under standard laboratory conditions.

4.3 Results and Discussion

4.3.1 Screening for *tcdR* Insert *pyrE* Repair in Single and Double Restriction-less Mutants

To be able to use the transposon suicide delivery plasmid, pMTL-GL15, in which the *Himar1* C9 transposase gene is controlled by P_{tcdB}, it was important to insert *tcdR* gene into *C. acetobutylicum*. Single and double restriction-less Δ*pyrE* mutants (as previously

obtained from chapter 3) were transformed with pMTL-ME6C-*tcdR* by electroporation followed by single crossover integration of the plasmid using ACE. The pseudo-suicide plasmid contains the *tcdR* gene with its ribosome binding site (RBS) downstream of the remaining part of the *pyrE* gene. Transformants were grown on CGM agar supplemented with thiamphenicol and uracil. Then a second recombination was allowed by growing on CGM broth without uracil and thiamphenicol. Subsequently, the putative double crossover transformants were inoculated on P2 minimum agar without any supplements added. Colonies which had lost their antibiotic resistance and had their prototrophy restored, were screened for *tcdR* insertion and *pyrE* repair using flanking primers JH14/F and JH14/R (as seen in Figure 4.4). Transcription of the *tcdR* gene was controlled by the *CA_C0025* promoter, as in the *tcdR* strain 3011 created in Zhang *et al.* (2015).

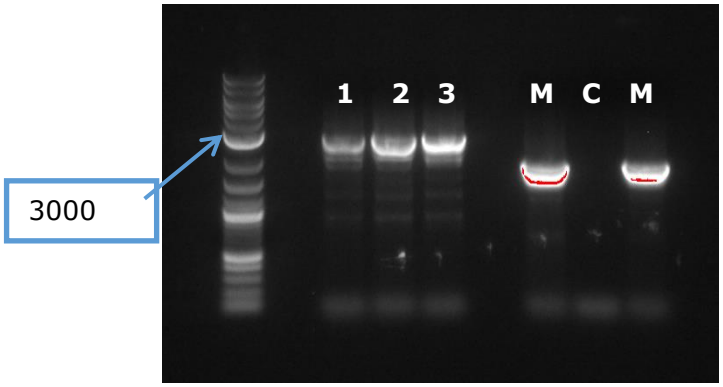

Figure 4.4: PCR screening of *pyrE* repair and *tcdR* insertion in *C. acetobutylicum* Δ*CA_C1502* using primers JH141/F and JH141/R. The expected DNA fragment size of 2880 bp is observed in lanes 1, 2 and 3. M is the restriction-less Δ*pyrE*Δ*CAC1502* strain with expected DNA fragment size at 1689 bp and C is the negative control which is just water with no DNA. DNA fragments were cut, purified, Sanger sequenced and shown to be of expected nucleotide sequence.

4.3.2 Design of TraDIS Experiment

The process of determining the essential genes for survival of *C. acetobutylicum* and investigating genes responsible for butanol and sodium tolerance requires the optimisation of different elements and stages of the experiment. The different elements needed for this study include:

A. Selection of an ideal TraDIS strain which does not have SNPs or deletions when compared to wild type
B. Selection of a transposon plasmid that functions effectively and efficiently in the bacterium,
C. Determining the concentrations of the butanol and sodium ion to subject the library with
D. Work flow of the transposon library generation procedure for efficient plasmid loss and unbiased random insertion
E. Work flow of library preparation process for TraDIS sequencing

4.3.3 Selecting an Ideal TraDIS Strain which does not have SNPs or Deletions

The first stage in the preparation of a transposon library was selecting a good strain for TraDIS studies. The fermentation profiles of wild type *C. acetobutylicum* and single and double RM KO strains with *tcdR* integrated were evaluated by completing 72 h fermentations with glucose as the substrate in CBM media. Samples were taken at 0 h, 4 h, 8 h, 24 h, 48 h and 72 h to analyse using GC and HPLC to determine their acetone, butanol, ethanol, acetic and butyric acid as well as glucose consumption profile (Figure 4.5). To validate that the selected strains had the same genotype as the wild type, Illumina whole genome sequence was carried out. Two strains of the tested strains that were observed to have no significant SNP or indels compared to the wild type and were selected for further studies and designated 52A (*tcdR*Δ*CAC1502*) and 51A (*tcdR*Δ*CAC1502*Δ*CAC3535*).

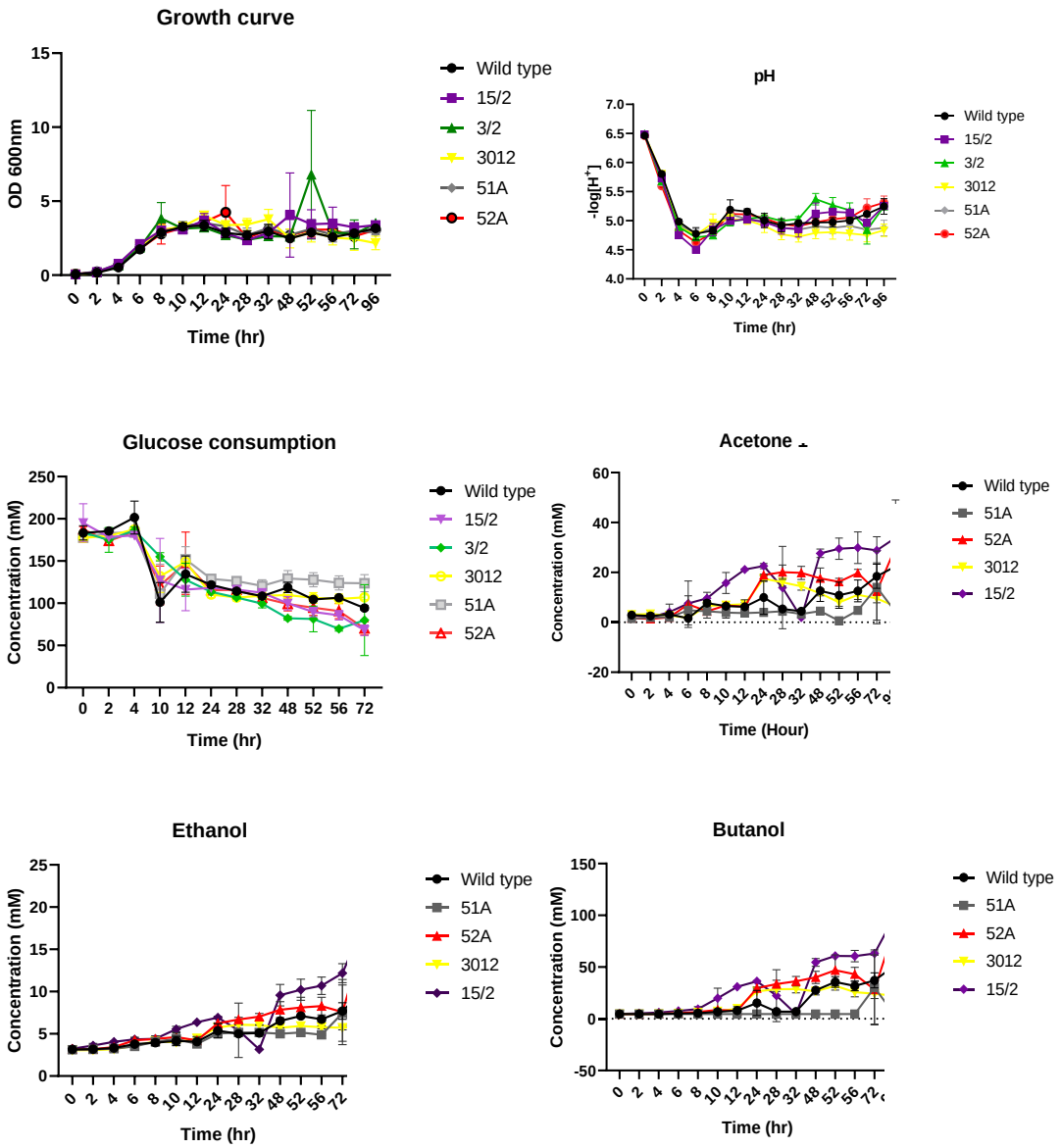

Figure 4.5: Fermentation profile analysis for wild type and restriction-less *tcdR* strains. This data shows that growth curve, pH, glucose consumption, acetone, ethanol and butanol yields over time for *C. acetobutylicum* ATCC 824 wild type and *tcdR* strains: 51A, 52A, 3012, 3/2 and 15/2. The values for glucose consumption and solvents are expressed in mM per litre. The 51A and 52A strains have *tcdR* gene inserted at *pyrE* locus and are *CA_C1502* mutants. The 3012 strain, created by Zhang *et al*. (2016), was from wild type strain which has *tcdR* gene inserted at the *pyrE* locus. The 15/2 and 3/2 strains were both CA_C1502 single restriction-less mutants with *tcdR* gene inserted at the *pyrE* locus and phagefree.

4.3.4 Selection of the Transposon Plasmid that Functions Effectively and Efficiently in *C. acetobutylicum*

4.3.4.1 Transposon Suicide Plasmid

Having deleted the type II RM systems in the previous chapter and inserted *tcdR* with *pyrE* repaired successfully, it was necessary to determine whether the 5.6-fold increase in transformation efficiency (achieved in chapter 3) was sufficient to construct a library pool of one million transposon mutants using a suicide plasmid. Using a suicide delivery vector for transposon library construction is actually preferable to replicative plasmid because it involves a one-step procedure in selecting transformants. The genome of the suicide transformants should contain either the composite transposon part or the entire plasmid. Transposon suicide plasmid pMTL-GL15 created by Dr. Gareth Little in SBRC Nottingham was used for this study. pMTL-GL15 has a mariner transposase which is advantageous feature to use with TA-rich *Clostridium* spp. because it propels transposition into TA sequences in the genome. It has the P_{tcdB} promoter to control the expression of the *Himar1* C9 transposase which means that its expression will only take place if the host already carries the *tcdR* gene. pMTL-GL15 is a derivative of pMTL-YZ14 (Zhang *et al.*, 2015) of which the Gram-positive conditional origin of replication has been deleted. Its transposable element consists of *catP*, an antibiotic marker, flanked by two inverted terminal repeats (ITR).

First, pMTL-GL15 was extracted from *E. coli* TOP10 and digested using NdeI and EcoR1 before sequencing to confirm that the transposon was intact and had not moved to other parts of the vector. TraDIS strain 52A was transformed with pMTL-GL15 in triplicates and with a control replicative plasmid, pMTL85141 via electroporation. The average transformation efficiency using the control replicative plasmid (pMTL85141) was 3 (±1.5) x 10^4 cfu/µg while that of pMTL-GL15 was an average of 50 (±13) cfu/µg. This result is comparable with studies by Lamberg *et al.* (2002) which showed that the transposition efficiency for replication defective transposon delivery plasmid was near three orders of magnitude less than the plasmid electroporation frequency. Because of the low transformation efficiency of pMTL-GL15, a conditional replicative transposon plasmid seemed like a better alternative in building a large transposon library.

4.3.4.2 Transposon Conditional Replicative Plasmid, pMTL-CW20

Circumventing the poor transformation efficiency of the pMTL-GL15 suicide plasmid, a replicative transposon would need some important features to facilitate plasmid loss, enable selection of transposon mutants from transformants, prevent transposition in *E. coli* and ensure plasmid stability. Based on these needs, a conditional replication plasmid created by Dr. Craig Woods (SBRC Nottingham, University of Nottingham UK), pMTL-CW20, was selected for TraDIS studies in *C. acetobutylicum*. pMTL-CW20 (Figure 4.6) exhibits a conditional Gram-positive origin of replication, P_{fac}-pCB102 and a conditional *Himar1* C9 transposase expressed via the P_{tcdB} promoter. Being a conditional replicon ensures plasmid loss while the presence of P_{tcdB} allows transposition to take place only in host organism carrying the *tcdR* gene and not in *E. coli*. It has an erythromycin antibiotic marker plasmid backbone which selects for transformants successfully replicating the plasmid. However, its composite transposon/transposable element consists of *catP*, an antibiotic marker, flanked by two inverted terminal repeats (ITR) which will be inserted at random in host genome during transposition. Two I-SceI sites flank the composite transposon insert while there is *codA* gene downstream of the transposon. Although the I-SceI recognition sites were built into the pMTL-CW20, they were not transferred to the genome during transposition and this helped in eliminating vectors during TraDIS library prep.

In addition, P_{thl}-Tn-*codA* arrangement ensures the plasmid stability in *E. coli* thereby preventing the transposon to excise and translocate locally which is a problem often experienced in other transposon vectors. It is worthy to note that many attempts to clone *codA* downstream of a strong promoter such as P_{thl} have been unsuccessful in *E. coli* at the SBRC research group. If the composite transposon translocates to another location, transcription of *codA* through P_{thl} even in the absence of 5-FC could lead to the death of the *E. coli* cell. To further ensure the stability of the plasmid in *E. coli*, a ferrodoxin gene (*fdx*) terminator is present downstream of the composite transposon (between *catP* and ITR) to prevent the transcription due to read through and expression of *codA*, which is under the control of the thiolase gene promoter, P_{thl}.

Another important trait is the *lacI* repressor gene that allows for IPTG-mediated control. The P_{fac} expression of the Gram-positive origin of replication (*repH*) was repressed by

the incorporation of the *lacI* repressor gene which was put under the transcriptional control of the P_{ptb}.

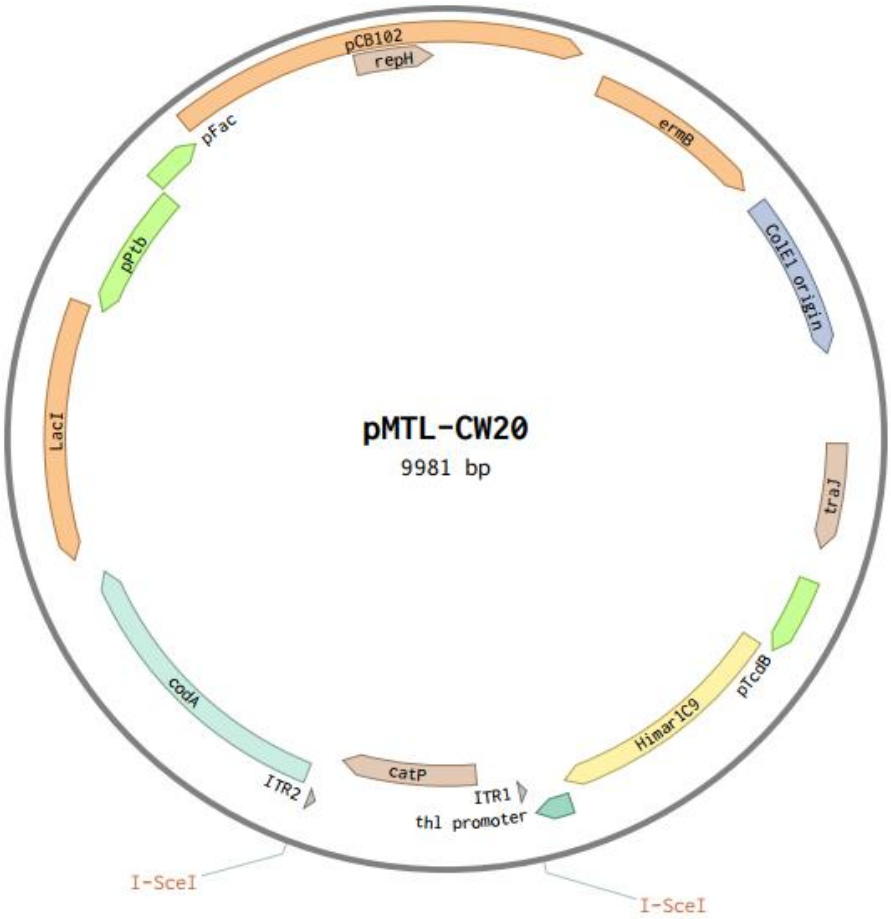

Figure 4.6: Map of conditional replicative vector pMTL-CW20. This plasmid has the following components: (i) ColE1 RNAII, the replication region of the *E.coli* plasmid ColE1; (ii) *ermB*, erythromycin antibiotic marker; (iii) pCB102, Gram-positive origin of replication, *repH*; (iv) P_{fac}, promoter of pMTL007 immediately upstream of the Gram-positive origin of replication; (v) P_{ptb} promoter, which is the promoter of the phosphotransbutyrylase (*ptb*) gene of *C. acetobutylicum*, immediately upstream of the *lacI* gene; (vi) *lacI* gene, encodes the lactose repressor protein; (vii) *codA*, the cytosine deaminase of *E. coli*; (viii) two I-SceI sites flanking the ITR components; (ix) two ITRs, inverted terminal repeats flanking *catP*; (x) the *catP* gene from *C. perfringens* conferring thiamphenicol resistance; (xi) P_{thl}, thiolase promoter downstream of the second ITR flanking *catP*; (xii) *Himar1* C9, mariner transposase; (xiii) P_{tcdB}, promoter from *C. difficile*; and finally (xiv) *traJ*, transfer function of the plasmid RP4 *oriT*.

4.3.4.3 Transposon Conditional Replicative Plasmid, pMTL-CW21

Another replicative plasmid was tested for its efficiency in creating a transposon pool in *C. acetobutylicum*. The conditional plasmid was pMTL-CW21, also created by Dr. Craig Woods. pMTL-CW21 has similar features to pMTL-CW20 with respect to the type of transposase, antibiotic backbone, composite transposon, P_{thl}-Tn-*codA* and Gram-negative origin of replication as shown in Figure 4.6. However unlike pMTL-CW20, it uses an anhydrotetracycline (aTC) inducible system to control the expression of the *Himar1* C9 transposase as well as the plasmid loss. As seen in Figure 4.7, it has a $P_{xylA/tetO}$ promoter controlling the transcription of the *Himar1* C9 transposase while $P_{tetR/tetO}$ promoter incorporated immediately upstream the tetracycline repressor gene (*tetR*) and next to the pCB102 derived origin of replication, drives the expression of both. Thus in the absence of the analog inducer aTC, the repressor protein is produced, transposon activity as well as Gram-positive replication are repressed, which is the permissive condition. On the other hand, when present aTC binds to the inactivated TetR repressor protein, thus allowing for the expression of the transposase and causing plasmid instability, this is the non-permissive condition. Plasmid loss is caused by the transcription read-through from the promoter $P_{tetR/tetO}$ through pCB102 (*repH*).

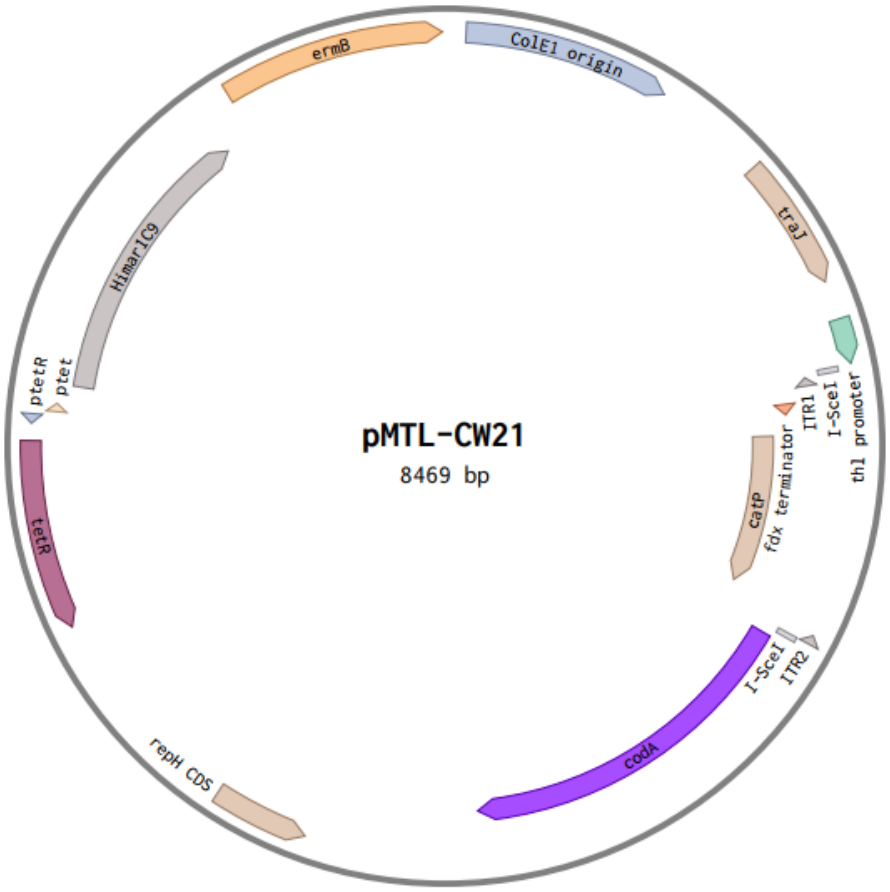

Figure 4.7: Map of conditional replicative vector pMTL-CW21. This plasmid has the following components: (i) ColE1 RNAII, the replication region of the *E.coli* plasmid ColE1; (ii) *ermB*, erythromycin antibiotic marker; (iii) *Himar1* C9, mariner transposase; (iv) P_{tet}, immediately upstream the himar transposase; (v) P_{tetR}, promoter upstream the tetracycline encoding repressor gene *tetR*; (vi) *tetR*, encoding the tetracyclyine repressor protein; (vii) pCB102, Gram-positive origin of replication, *repH*; (viii) *codA*, the cytosine deaminase of *E. coli*; (ix) two I-SceI sites flanking the ITR components; (x) two ITRs, inverted terminal repeats flanking the *catP*; (xi) the *catP* gene of *C. perfringens* conferring thiamphenicol resistance; (x) *fdx*, terminator after *catP* marker; (xi) *thl* promoter, thiolase promoter downstream of the second ITR; and lastly (xii) *traJ*, transfer function of plasmid RP4 *oriT*.

One advantage of using this plasmid over pMTL-CW20 is that there is no need to insert *tcdR* gene into the host target strain prior to transformation and thus the wild type can be used. This is an advantage as it is uncertain whether knocking out the restriction modification genes and inserting *tcdR* have created bias in selective fitness of mutant genes or insertion of the chosen transposon due to polar effects. In this case, *C. acetobutylicum* ATCC 824 wild type (WT) was the target host with this plasmid.

pMTL-CW21 has 26 Cac824I (5' GCNGC 3') recognition sites. Therefore to overcome the restriction modification system in the wild type, methylation of the plasmid was required. Despite using pAN2, which contains the ϕ3TI methyltransferase gene of *B. subtilis* phage ϕ3TI, for successful methylation in chapter 3, it had one major disadvantage in current study. The antibiotic marker gene on the backbone of the pAN2 plasmid confers tetracycline resistance thus, requiring tetracycline inclusion in the media when propagating any plasmid to be methylated in *E coli* TOP10 (pAN2). When pMTL-CW21 was purified from *E. coli* TOP10 pAN2 and transformed into *C. acetobutylicum* WT, there were no transformants on erythromycin selection plates. It was not clear why this happened. It might be because the presence of tetracycline used to maintain the pAN2 drove the expression of the *Himar1* C9 transposase which might have led to transposition occurring earlier in *E. coli* or a relocation of the composite transposon to another part of the plasmid.

To circumvent this problem, the *E. coli* TOP10 R702::*ϕ3TI* strain which had an apramycin marker (created in chapter 3) was used for the methylation of pMTL-CW21. After purifying pMTL-CW21, methylation was confirmed by an attempt to digest it with Fnu4HI enzyme (an isoschizomer of Cac824I) which revealed that it was not digested by the enzyme.

4.3.5 Determine the Sub-Inhibitory Concentration of Butanol and Sodium Ion for TraDIS

Once a strain and plasmid had been selected for TraDIS studies, there was a need to determine the sub-inhibitory concentration of butanol and sodium chloride that were to be used for the TraDIS studies. In order to investigate the genes that are responsible for innate butanol tolerance, the TraDIS library was to be subjected to varying concentrations of butanol conditions. Evaluation of the concentration to be used was done ahead so that once the library was generated, it would be subcultured immediately in the chosen concentrations of butanol and sodium ion without previously storing them in at -80°C. It has been observed that storing cells at -80°C could lead to losing some of the cells as a result of freeze injury (Sourek, 1974). Losing some of the mutants in the library through freezing before subjecting them to sub-inhibitory concentrations of butanol and sodium ion may not give a true result of representation of genes responsible for *C. acetobutylicum*'s survival under such challenging conditions.

The concentration of butanol to be used should be chosen to negatively affect growth of the bacteria whilst not killing the bacteria and allow the expression of the genes required for tolerance. In essence, the overrepresentation of butanol tolerance mutants would in turn lead to identification of butanol tolerant genes in the final sequencing data. To first identify the sub-inhibitory concentration of butanol to subject the library to, the wild type was grown in CGM media containing different concentrations of butanol 0 g/l (0% v/v), 2 g/l (0.25% v/v), 4 g/l (0.49% v/v), 6 g/l (0.74% v/v), 8 g/l (0.98% v/v), 10 g/l (1.22% v/v), 12 g/l (1.46% v/v), 14 g/l (1.7% v/v), 16 g/l (1.93% v/v) and 18 g/l (2.18% v/v). The OD_{600} was taken at different time intervals for 96 hours as seen in Figure 4.8.

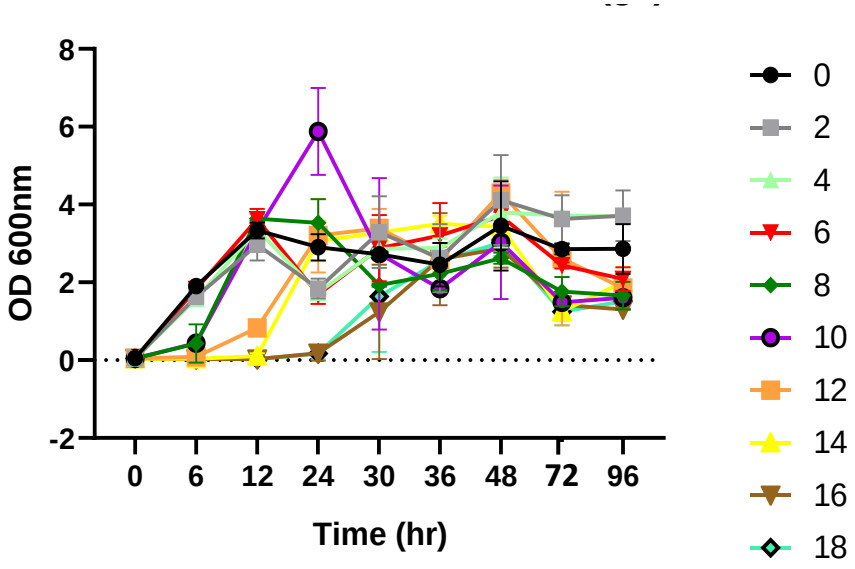

Figure 4.8: Growth of *C. acetobutylicum* ATCC 824 in the presence of various concentrations of butanol. *C. acetobutylicum* ATCC 824 was grown in butanol concentrations ranging from 0 to 18 g/l in triplicates and the OD_{600} was taken at 0 to 96 hours. The error bars indicates the standard deviation of OD_{600} at each sample time.

A lag in growth was observed in the first 6 hours when a butanol concentration of 8 g/l (0.98% v/v) and greater was used (Figure 4.8). There was a further delay in growth after 12 hours with a concentration between 12 g/l (1.46% v/v) to 18 g/l (2.18% v/v). A similar trend was reported by Tomas *et al.* (2004) when culturing *C. acetobutylicum* ATCC 824 carrying the *groESL* overexpressed plasmid and ATCC 824 carrying a *groESL*-free control plasmid in 0% (v/v), 0.25% (v/v) and 0.75% (v/v) butanol challenge conditions. Growth inhibition increased with increasing concentrations of butanol. This also agrees with the report that butanol completely inhibits the growth of

C. acetobutylicum at 14 g/l (Ounine *et al*., 1985). However, the growth observed here after the 12th hour for concentrations of 14 g/l and more might be as a result of evaporation of the butanol at 37°C through the cotton plugged opening of conical flasks. Like alcohol, butanol has been predicted to affect membrane-linked fluidity. It has been shown to cause a change in the lipid composition of the cell membrane by reducing the unsaturated to saturated fatty acid ratio (U/S) which in turn affects membrane function (Lepage *et al*., 1987). Butanol has been shown to have various inhibitory effect on the growth of *C. acetobutytlicum* by affecting its ability to maintain internal pH, inhibiting membrane ATPase activity, then reduces intracellular levels of ATP which leads to a decrease in glucose uptake (Bowles and Ellefson, 1985). However, there are limited genetic studies including extensive investigation towards identifying *C. acetobutylicum* genes that contribute to butanol tolerance and stress response.

In a previous study published by Tomas *et al.* (2004), microarray-based transcriptional profiles of *C. acetobutylicum* ATCC 824 carrying overexpressed *groESL* (pGROE1) and ATCC 824 with *groESL*–free control plasmid (pSOS95del) in butanol challenged cultures relative to unchallenged condition identified 104 and 154 genes differentially expressed in 0.75% (v/v) respectively. Given that butanol concentration of 8 g/l (equivalent to 0.98% v/v) seem to be the lowest concentration affecting bacterial growth in Figure 4.8 at the 6 hour mark, it was selected as the butanol concentration to subject the library to. It was expected that transposon mutants attenuated when culturing the library pool culture in butanol concentration of 8 g/l are outcompeted while others with increased growth are overrepresented in the population. Quantification of the TraDIS results would then reveal the prevalence of each mutant fitness in the selected condition (Chao *et al*., 2016).

In a separate experiment, wild type was grown in CGM broth media containing different concentrations of sodium chloride; 0, 2, 4, 6, 8, 10, 15, 20 and 25 g/l. The OD_{600} was taken at different time intervals for 96 hours as seen in Figure 4.9.

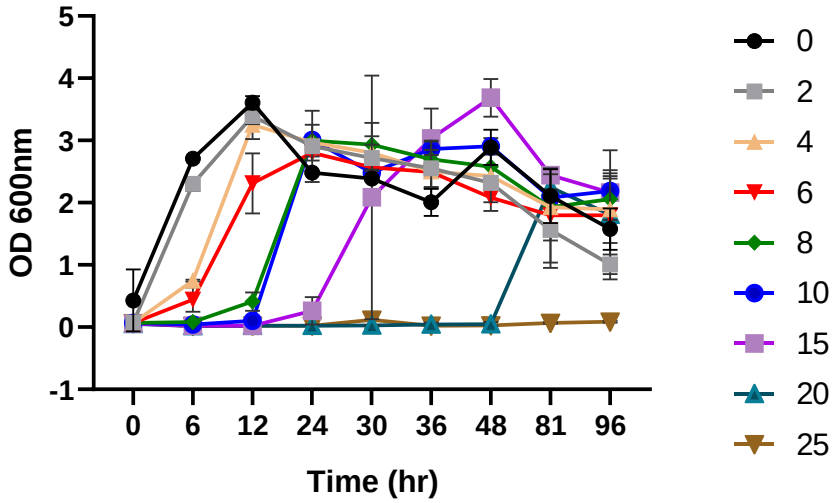

Figure 4.9: Growth of *C. acetobutylicum* ATCC 824 at various concentrations of sodium chloride. *C. acetobutylicum* ATCC 824 was grown in sodium chloride concentrations ranging from 0 to 25 g/l in triplicates and the OD_{600} was taken at 0 to 96 hours. The error bars indicates the standard deviation of OD_{600} at each sample time.

As observed in Figure 4.9, there was a growth lag of the *C. acetobutylicum* culture with increasing concentrations of NaCl. *C. acetobutylicum* did not grow at 25 g/l and therefore that it is the only concentration tested that completely inhibited the bacterial growth. This result is in agreement with Zhao *et al.* (2016) that the growth of *C. acetobutylicum* is totally inhibited in glucose culture containing 25 g/l sodium chloride. Maddox *et al.* (1995) has shown that the toxic effect of NaCl on *C. acetobutylicum* is mainly caused by the sodium ion and not the chloride ion as 10 g/l caused 40% growth inhibition but there was no significant difference in growth in ammonium chloride of 10 g/l. There was a significant delay for the first 6 hours between sodium chloride concentrations of 4 g/l to 25 g/l, while growth occurred for 4 g/l and 6 g/l at the 6th hour mark, growth for 8 g/l to 25 g/l did not occur at that time point. High sodium conditions adversely affect cell homeostasis by inducing high osmotic pressure. Cells may adapt to high sodium ion challenges by channelling more energy towards controlling membrane transporter and channels (Zhao *et al.*, 2016). Eight g/l was selected for both sodium ion and butanol concentrations as conditions to subject the transposon library to. This decision was made with the hypothesis that such conditions may be fatal for mutants with transposon insertions in stress response genes that are responsible for survival in butanol or sodium ion conditions. Therefore, genes without insertions in the library pool following butanol or sodium exposures compared to that

of the control (without butanol or sodium ion supplementation) would be suggested as essential genes for the survival of *C. acetobutylicum* in that challenging condition.

4.3.6 Work Flow of Transposon Library Generation Procedure

Figure 4.10 below shows a flow chart of the general transposon mutagenesis procedure. Arrows to the right indicate some elements altered to improve on the biological procedure. Supercompetent cells of *C. acetobutylicum* ATCC 824 were prepared (as described in Materials and Methods section 2.1.5.2) and electroporated with pMTL-CW20 or pMTL-CW21 transposon delivery vectors (as described in Materials and Methods section 2.1.6.2 and 2.1.7). Transformed cells were recovered by plating onto freshly prepared CGM agar supplemented with erythromycin. Incubation was done for 36 hours instead of 48 hours and then the transformants were pooled together into CGM liquid and serially diluted. The dilutions were plated on plates containing Tm and induction agent (IPTG/aTC) and incubated at 37°C. Serial dilutions of transformants were also plated on media containing Tm, Erm and Erm+inducer to allow calculation of transposition frequency. After incubation, colonies on mutant selection plates were screened for attributes of transposition via replica plating to determine plasmid loss efficiency.

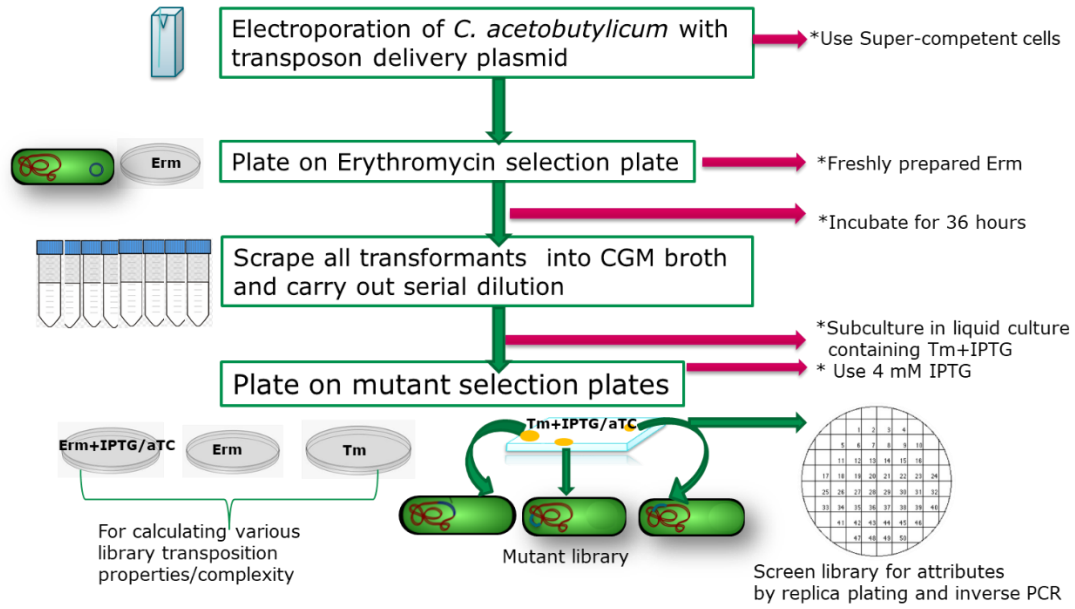

Figure 4.10: Schematic illustration of TraDIS library construction, indicating areas modified to improve library creation. The procedure is illustrated starting from the transforming the *C. acetobutylicum* to the screening of colonies on mutant selection plates. The pink arrow shows areas of the procedure that were modified to improve the frequency of transposition events and plasmid loss.

Properties of an effective transposon delivery vector can be measured by plasmid loss efficiency, transposition frequency and resistance to plasmid loss. Since the presence of the inducer (aTC or IPTG) causes a transcriptional read through from the strong promoter to the Gram-positive origin of replication, the resulting instability of plasmid means that the number of colony forming units per ml on Erm plates will be higher than those on Tm+IPTG/aTC plates. It is also expected that the number of cfu/ml on Tm+IPTG/aTC is more than those on Erm+IPTG/aTC as plasmid retention in the presence of the inducer should be low.

Plasmid loss efficiency is the percentage of plasmid harbouring cells that do not grow in the presence of the inducer and antibiotic plasmid backbone. It is also calculated through replica plating as the percentage of population on mutant selection plates that have lost the plasmid antibiotic marker backbone. Apart from determining plasmid loss efficiency through patch plating, it was determined by calculating the number of colony forming unit per ml on Erm +IPTG/aTC plates substracted from the cfu/ml on Erm plates, and then divided by the cfu/ml on Erm plates multiplied by 100 (see Materials and Method section 2.1.7).

Transposition frequency is expressed as comparability between the estimated population on mutant selection plates considering colony forming units on non-permissive plasmid selection and the estimated population of plasmid-harbouring colonies. The transposition frequency or transposon mutant ratio of library generation was calculated as shown in Materials and Method section 2.1.7.

4.3.7 Library Preparation Process for TraDIS Sequencing

The gDNA library preparation protocol are as described in Material and Methods section 2.5.6 and it is similar to a refined technique developed by Barquist *et al.* (2016) except that an I-SceI digestion step was included before the PCR amplification step. The library prep protocol was carried out on *C. acetobutylicum* ATCC 824 transposon library mutants by Dr. Craig Woods, Cynthia Akaluka and Dr. Bashir Rumah, University of Nottingham UK while the bioinformatics of sequencing result using Bio-TraDIS Analysis toolkit was done by Dr. Craig Woods.

4.3.8 First Attempt at TraDIS

4.3.8.1 Transposon Library Creation - 1st Attempt

The first attempt to create a library was with pMTL-CW20 using 1 mM of IPTG as an inducer for plasmid loss. Three transformations of 52A strain (*tcdRΔCac1502*) with pMTL-CW20 were carried out as described in Materials and Methods section 2.1.7. Several attempts to demonstrate high transposition and plasmid loss efficiency were unsuccessful as there was no significant order difference between the cfu/ml on Erm+IPTG and Tm+IPTG plates. The calculated colonies on both Erm+IPTG and Tm+IPTG plates ranged from 2.5 $x10^3$ to 1.5 $x10^4$ cfu/ml. Therefore, to facilitate transposition and plasmid loss when inducing with 1 mM IPTG, the modified version of the library creation method in Figure 4.10 was applied. After scraping the transformants from Erm selective plates, they were subcultured in 200 ml of liquid CGM+Tm+IPTG (1 mM) for 10 hours prior to plating on mutant selection plates to promote plasmid loss and transposition. Colonies on the mutant selection plates for the liquid subcultured colonies were observed to have different sizes; both big and small sizes. The plasmid loss of the library created was 99.99% by calculation and 95% by replica plating while the transposition frequency was 3.3 $\times 10^{-3}$ per plasmid carrying cell.

4.3.8.2 Sequencing data for pMTL-CW20 - 1st attempt

The library pool gDNA was extracted then library prepped and sent for sequencing at the University of Nottingham DeepSeq facility. The bioinformatics analysis of sequencing data are as shown in Table 4.1.

Table 4.1: TraDIS pMTL-CW20 library sequencing result for 1st attempt

		Chromosome		Megaplasmid		Total	
No of Mutants	**% Mapped**	**Unique Insertion Sites**	**Seq Len/UIS**	**Unique Insertion Sites**	**Seq Len/UIS**	**Total Unique Insertion Sites**	**Total Seq Len/Total UIS**
32,308	95.58	4836	814.90	398	482.41	5234	789.62

Seq Len/UIS means sequencing length per unique insertion site.
Total Seq Len/Total UIS means total sequencing length per total unique insertion site.

The number of cfu mutants isolated, percentage of reads that mapped to genome, number of unique insertion sites and sequencing length per unique insertion sites for both chromosome, pSOL1 and total of both for TraDIS pMTL-CW20 library are all

shown on the Table 4.1. About 95% (a total of 604,336 reads) of sequence reads were mapped to the *C. acetobutylicum* ATCC 824 reference chromosome (accession NC_003030) and pSOL1 (accession NC_001988). Figure 4.11A and Figure 4.11B show reads mapping across the reference chromosome and megaplasmid. A random distribution of the tagged reads throughout the genome (see Figure 4.11) was observed, which supports the conclusion by Cartman and Minton (2009) that mariner transposase is ideal for G+C low organism like *C. acetobutylicum* as it ensures random insertion around the genome. Bioinformatic analysis of the library also revealed that there were no insertions in 1975 genes (data not shown) and there was an average of one insertion in every 789 base pairs (as shown in Table 4.1). Although the library created is not densely saturated enough to identify the genes essential for survival under the chosen standard *in vitro* conditions, it showed that the technique is promising for creating a saturated transposon library.

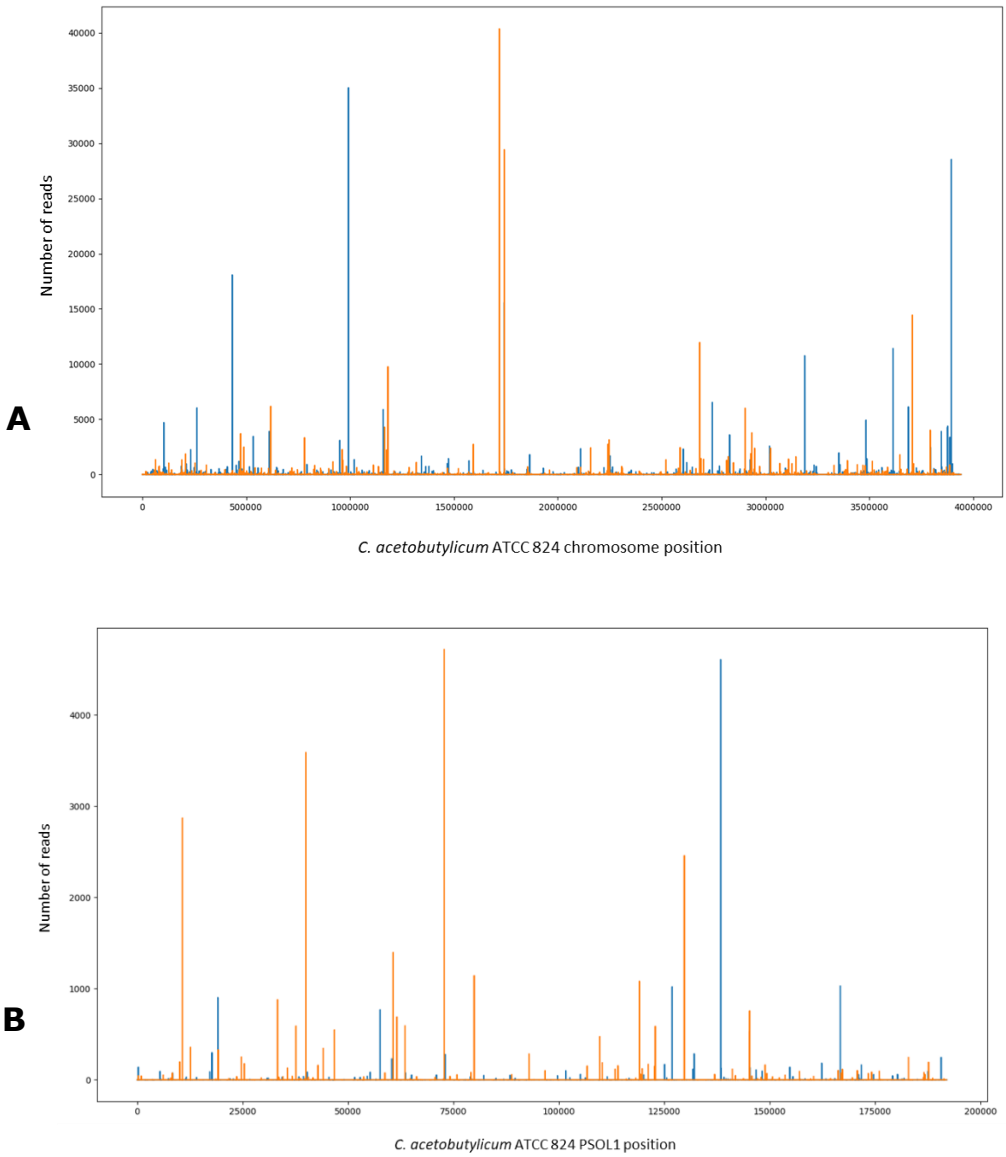

Figure 4.11: Transposon tagged maps and unique insertion gene frequency for first attempt pMTL-CW20 TraDIS library sequencing. Figure 4.11A shows the transposon-tagged reads mapped across *C. acetobutylicum* ATCC 824 reference chromosome. X-axis is the chromosome position of the reference *C. acetobutylicum* ATCC 824 while y-axis shows the number of reads mapped to a particular site. Vertical orange lines are reverse insertion while blue is for forward insertions. **Figure 4.11B** shows the transposon-tagged reads mapped across *C. acetobutylicum* ATCC 824 reference pSOL1. As with Figure 4.11A, x-axis is the pSOL1 position of the reference *C. acetobutylicum* ATCC 824 while y-axis shows the number of reads mapped to a particular site.

Furthermore, three genes and one position were overrepresented, they were 1718725 positions (between *CA_C1573* and *CA_C1574*), *CA_C0862* (transmembrane protein), *CA_C1598* (hypothetical protein) and *CA_C3688* (membrane associated methyl-

accepting chemotaxis protein) with 40359, 35560, 30291 and 28557 transposon tag reads mapped to the loci, respectively. The sum of these four reads gave 22.3% of the total reads mapped to chromosome. Despite having the highest number of reads, these three genes did not have the highest number of unique insertions.

Possible reasons for this happening may be, first, that there was a problem with library creation procedure which may lead to exponential growth of selected mutants such as the mutants of the three genes having 22.3% of the total chromosome mapped reads. Early transposition could lead to dominance of the early mutant over late mutants especially if transposition occurred in transformants on plasmid selection plates (Erm) before induction. Controlling this would be difficult as the *tcdR* switch was in *C. acetobutylicum* and once it had been transformed with the plasmid, early transposition was possible. This suggested that a shorter incubation period might be required. Secondly, the different colony sizes (big and small) observed on the mutant selection plates, may indicate overrepresentation of some mutants and create prejudice of the fitness of some genes. Not surprisingly, this could be caused by the treatment of the transformants in liquid culture containing Tm + IPTG prior to plating on mutant selection plates. The liquid culture pretreatment might have fostered plasmid loss but it seemed to positively select the fast growing mutants under tested condition. Elimination of this step was needed because competition for nutrients in liquid culture may have greater effect on the weaker mutants than when the mutants grow on solid media where their nutrient consumption and waste excretion do not directly affect one another's growth. Colonies on a solid media experience diffusion limitation of substrate and inhibitors unlike bacteria growing in liquid media and this reduces the competition amongst colonies (Jeanson *et al*., 2015).

4.3.8.3 Troubleshooting the Plasmid Loss Challenge (1)

To eliminate the liquid culture pretreatment with Tm and IPTG, increasing the concentration of IPTG administered was considered. In order to determine the concentration of IPTG required to achieve high efficient plasmid loss on solid media, the transformants were cultured in CGM agar containing Tm and increasing concentrations of IPTG: 0 mM, 1 mM, 2 mM and 4 mM were tested. After 48 hours of growing the putative transposon mutants, the cfu/ml of each plate of varying IPTG concentrations was counted and the plasmid loss efficiency was determined through

percentage of colonies that lost the plasmid as evaluated by replica-plating (Figure 4.12).

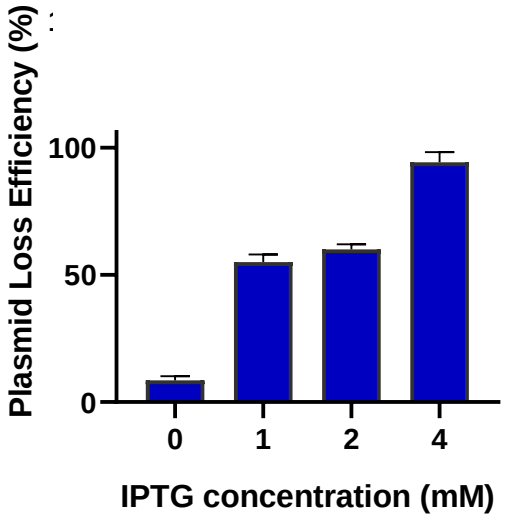

Figure 4.12: Plasmid loss efficiency of pMTL-CW20 in *C. acetobutylicum* using varying concentrations of IPTG.

The results presented in Figure 4.12 indicated that the IPTG inducible system was a bit leaky as some colonies on Tm plates were Erm sensitive even in the absence of IPTG, the inducing agent. The absence of IPTG is the condition upon which the P_{fac}-pCB102 origin of replication is meant to be stable. Chloramphenicol acetyltransferase (CAT) assay carried out by Zhang *et al.* (2015) reported a significant *lac*-based expression in *C. acetobutylicum* in the absence of IPTG. In a previous study by Heap *et al.* (2007) also using *catP* as reporter gene, the CAT assay showed that there was notable expression without added IPTG. The problem with this in *C. acetobutylicum* was that there was a combination of the P_{fac} becoming unstable and transposition being triggered immediately as the plasmid entered the bacteria. To tackle this problem, the transformants on initial selective media were harvested after 32 to 36 hours of incubation instead of 48 hours as shown in Figure 4.10.

Nevertheless, as the concentration of IPTG is increased, an increase in plasmid loss efficiency was observed as the condition renders the origin of replication more unstable. As observed in Figure 4.12, increasing the concentration from 2 mM to 4 mM led to an increase in plasmid loss efficiency from 60% to 94%. Zhang *et al.* (2015) achieved 80% plasmid loss of pMTL-YZ14 in *C. acetobutylicum* using 1mM IPTG although *C. sporogenes* had 100% plasmid loss with the same plasmid. It was not clear why plasmid

loss was more efficient in *C. sporogenes* than in *C. acetobutylicum*. Therefore, 4 mM IPTG was selected as the concentration for inducing plasmid loss in transformants of pMTL-CW20.

Despite eliminating the liquid pretreatment step, reducing the incubation time of transformants on selective media and increasing the concentration to IPTG to 4 mM, several attempts made to create high density transposon library with high plasmid loss efficiency were unsuccessful. Because it was not clear what might have caused this, a decision was made to test library creation with pMTL-CW21.

4.3.9 Second Attempt at TraDIS

4.3.9.1 Transposon Library Creation -2nd Attempt

The procedure in Materials and Methods section 2.1.7 was followed in creating a transposon library pool using wild type *C. acetobutylicum* ATCC 824 with methylated pMTL-CW21. Plasmid loss and transposition was achieved by growing the transformants on CGM with Tm and aTC (100 ng/ml). The estimated population of the generated pool was 1.2 million mutants. An average of ninety-nine percent (99%) plasmid loss efficiency both by calculation and replica plating was achieved with 100 ng/ml aTC. The transposition frequency achieved was 1.85×10^{-3}. This plasmid loss seemed more promising than that of pMTL-CW20, hence, further assays to determine what concentration of aTC was more effective in driving transposase and plasmid loss was not carried out. Besides, there has been successful reports in literature of 100 ng/ml aTC used for transposon induction in *C. acetobutylicum* (Dong *et al*., 2012). After harvesting the colonies (slurry) from the mutant selection plates, they were cultured in 100 ml CGM broth containing Tm and aTC to maintain the transposon mutant status to an initial OD_{600} of 0.1.

4.3.9.2 Sequencing Data for pMTL-CW21 - 2nd Attempt

The gDNA of the library pool was extracted, prepped and sent for sequencing at the DeepSeq facility (the University of Nottingham).

Table 4.2: TraDIS pMTL-CW21 library sequencing result for 2nd attempt

			Chromosome		Megaplasmid		Total	
Cell	**No of Mutants**	**% Mapped**	**Unique Insertion Sites**	**Seq Len/UIS**	**Unique Insertion Sites**	**Seq Len/UIS**	**Total Unique Insertion Sites**	**Total Seq Len/Total UIS**
SL	1.2m	0.88	1386	2843	130	1477	1516	2726
SU	1.2m	3.27	2132	1848	170	1129	2302	1795

SL means slurry which is DNA from cells washed from Tm+aTC plates
SU means subcultured slurry in liquid media containing Tm+aTC
1.2m is 1.2 million mutant
Seq Len/UIS means sequencing length per unique insertion site
Total Seq Len/Total UIS means total sequencing length per total unique insertion site

The number of cfu mutants isolated, percentage of reads that mapped to genome, number of unique insertion sites and sequencing length per unique insertion sites for both chromosome, pSOL1 and total of both for TraDIS second attempt library sequencing are all shown on the Table 4.2. Despite demonstrating effective plasmid loss efficiency through replica plating and calculation, the sequencing results indicated that 95.21% of the slurry and 92.37% of the subcultured library pool mapped to the plasmid, pMTL-CW21, which means that there was a large residue of plasmid DNA in the library. The increase in the number of unique insertion sites in the subcultured pool from 1516 to 2302 as described in Table 4.2 might be a result of the continuing transposition activity carried out by residue plasmids in liquid media rather than overrepresentation of mutations that are essential for cell growth in laboratory conditions. A transposon library pool created using pMTL-YZ14 in *C. beijerinckii* gave a poor complexity TraDIS sequencing result and a large percentage of plasmid reads (Dr. Gareth Little of SBRC Nottingham, unpublished) just as this library.

4.3.9.3 Troubleshooting the Plasmid Reads Challenge (2)

Figure 4.13 shows the mapping of the trimmed TraDIS data to the plasmid pMTL-CW21. The Figure 4.13 reveals that even though there were mappings to the entire plasmid as indicated by the horizontal blue line from *ColE1*origin to *ermB*, there were regions that had more reads as exhibited by the blue and red peaks.

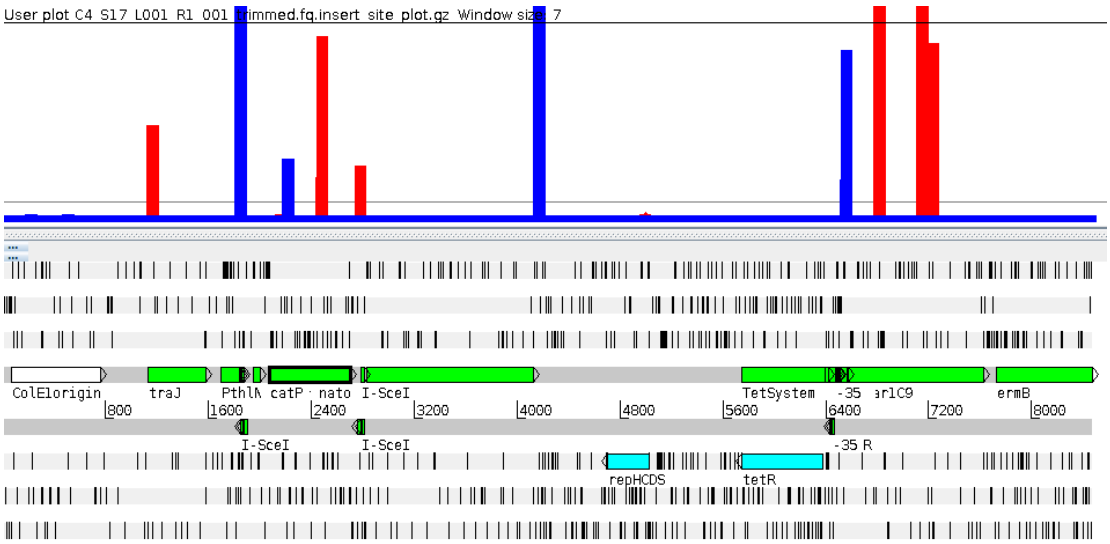

Figure 4.13: Trimmed sequencing plasmid reads mapping to the pMTL-CW21 transposon delivery plasmid. This data was analysed using Artemis software tool. This shows the relative height of peaks of aligned trimmed reads to pMTL-CW21. The dominant horizontal stretched dark blue part of the diagram represents trimmed reads from the beginning to the end of the transposon delivery plasmid. The dark blue and red horizontal peaks represents increased number trimmed reads mappings in a particular locus on the plasmid. Dark blue is for forward reads while red is for reverse reads. The vertical bars indicate genes of the plasmid were the trimmed reads mapped to: the white bar is ColE1 origin og replication, the lemon green are forward genes while the light blue are reverse genes in the transposon delivery plasmid.

4.3.9.4 Probable Reasons Why There were Plasmid Reads in TraDIS Data

To investigate this problem, a number of hypotheses were proposed and then different aspects of the experiment were assessed in attempt to identify the cause of low plasmid loss efficiency and improve the mapping of the sequencing reads to the genome.

4.3.9.4.1 Loss of Activity of Erythromycin

The first hypothesis for more Tm^R Erm^S colonies on mutant selection plates but low transposition events was that the erythromycin used for screening might have lost its activity. This would mean that there might be a false result about plasmid loss on replica plating due to instability of erythromycin. To test this theory, erythromycin was replaced with a more stable analogue, clarithromycin. The library creation, gDNA library extraction, library prep, sequencing and bioinformatics analysis were all repeated by Cynthia Akaluka and Dr. Craig Woods. However, this did not solve the problem as above 90% of library pool still mapped to the plasmid (data not shown).

4.3.9.4.2 Poor Digestion of I-SceI

I-SceI treatment step in the library prep may not have been effective at plasmid degradation, which could have caused the large number of plasmid sequences in the TraDIS pool. The main goal of introducing I-SceI recognition sites in the delivery vector was to eliminate plasmid reads before the library sequencing stage. Poor digestion of I-SceI sites could defeat this goal. A repeat of the library preparation with a control of transposon library pool from *C. difficile* was created by Marcel Te Vrugt (PhD student, SBRC Nottingham, University of Nottingham UK) and showed that the I-SceI digestion step was effective for other cases but not the one generated here. The generated *C. difficile* library was observed to contain only 373 reads in the plasmid which indicated a decent level of plasmid elimination.

4.3.9.4.3 Low Transposition Frequency

This is a case where the number of successful insertions in the genome was so low that the few surviving plasmid-based reads with I-SceI sites in them would have been amplified. Based on the sequencing data, this seemed more plausible. It could also explain why the PCR cycles during library preparation had to be increased from 20 to 25 to amplify anything effectively. The only problem with this explanation is how to relate it with cfu/ml results from the validation plates. Looking at the information in section 4.3.9.1, the calculated results for plasmid loss efficiency and rate of transposition frequency should be sufficient to invalidate this line of thought. Transposition frequency of 1.85×10^{-3} is comparable with existing literature (Zhang *et al*., 2016; Cartman and Minton, 2010). The ratio of cfu/ml on aTC+Tm to aTC+Erm plates was 340, which means that for every cell that permitted the plasmid in the presence of aTC, there are 340 cells that lost the plasmid in the presence of aTC. Thus, in 1.2 million library pool, there should be only 3,529 plasmid harbouring cells resisting plasmid loss. This ratio does not depict low transposition frequency due to plasmid retention, therefore this does not explain the problem.

Another way to investigate the diversity of unique insertions in the genome was to do an inverse PCR of selected transposon mutants. Twenty-four random colonies, on the mutant selection plates, were selected from those with confirmed plasmid loss (Erm^s) and inverse PCR was carried out as described in Materials and Method section 2.2.4.

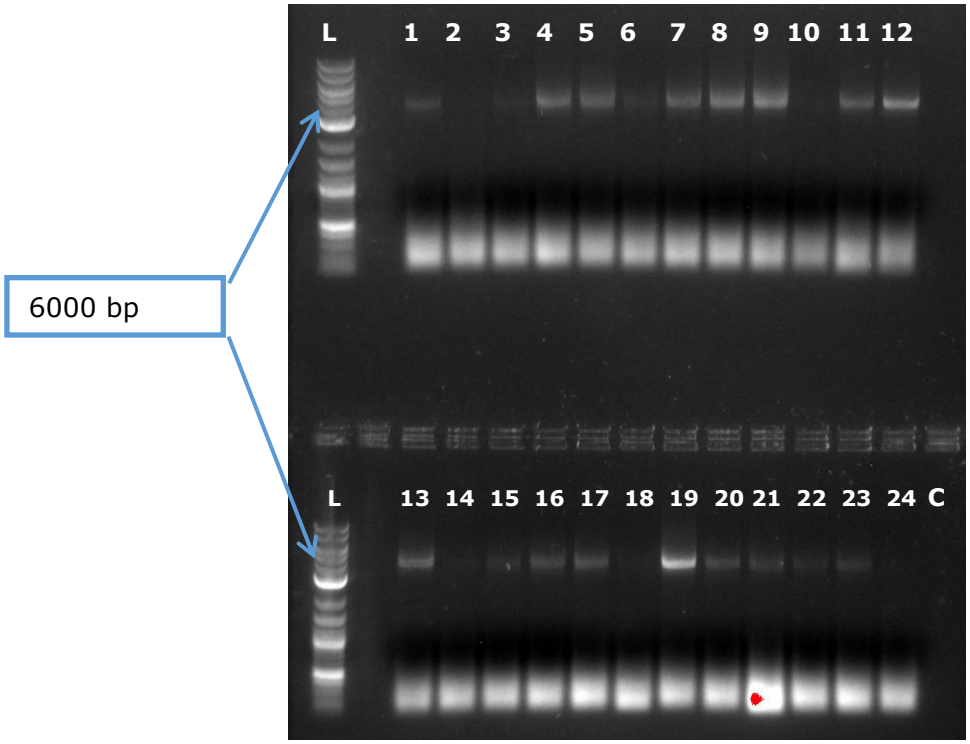

Figure 4.14: Inverse PCR fragments for 24 randomly selected transposon mutant colonies. Lanes 1 -24 represent the results for 24 selected transposon mutant colonies while C is the control water. PCR were run with primers catP-INV-F1 and catP-INV-R1. The quality of a good library should be determined by different bands sizes for each transposon mutant. However, this result depicts similar DNA fragments in all the lanes (note although a few lanes show faint or no PCR bands here, they appear to be similar band sizes on the computer screen). The band size is estimated to be above 6000 bp.

The gel image in Figure 4.14 indicated that the frequency of transposition events occurring in the pool were very low. It was expected that there should be different DNA fragment sizes in each lane, however, the picture shows the same band size in all 24 mutant colonies. When comparing the inverse PCR products with that from transposon mutagenesis vector pMTL-CW21 and Erm + aTC colonies, it was observed that they were all similar band sizes (data not shown). This may indicate that plasmid integration took place in the *C. acetobutylicum* genome instead of transposition or the plasmid replicates in the cell. Six bands from the clones were cut and PCR products extracted and sequenced with Sanger sequencing. Bioinformatics analysis results showed that the catP-INV-R2 primer did not align to any of the sequences which may indicate that the PCR amplified the plasmid backbone.

4.3.9.4.4 Plasmid Integration

To investigate the possibility that there could be plasmid integration event occurring after electroporation or during the mutant selection stage. Homologous recombination is a problem experienced when using delivery vectors to carry out genetic engineering in a bacterial cell. Studying the TraDIS trimmed sequencing reads mapped to plasmid pMTL-CW21 in Figure 4.13, there is an observed peak in the thiolase promoter region. This raises a critical question: Is there homologous recombination of the pMTL-CW21 at the thiolase gene in *C. acetobutylicum*? Homologous recombination occurring in the bacteria could explain the reason for identical clones in inverse PCR but it does not explain why most colonies on mutant selection plates are thiamphenicol resistant but erythromycin sensitive. Knowing that the promoter originates from the acetyl-CoA transferase gene in *C. acetobutylicum*, screening for homologous recombination into the genome at the acetyl-CoA transferase locus was done. Since determining the genome integration site of the plasmid was difficult using the TraDIS plasmid trimmed reads, direct screening of putative mutants' gDNA were done. Two flanking primers to acetyl-CoA transferase gene, Thiolase F and Thiolase R, were designed. Ten (10) independent colonies from aTC+Tm with confirmed Erm^s were amplified with the flanking primers. The gel picture in Figure 4.15 displays the results of thiolase screening.

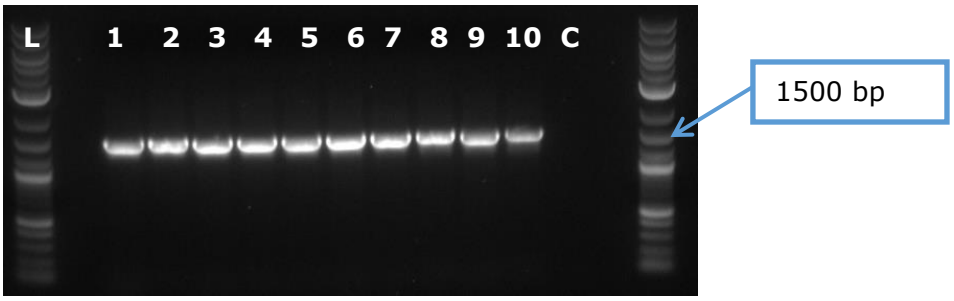

Figure 4.15: PCR screening reveals no insertion at acetyl-CoA transferase gene locus. Expected band size 1,514 bp for no integration. The DNA fragment sizes in all the lanes are similar. C represents the negative control which is running the PCR in the absence of DNA. There was no evidence of integration in the thiolase promoter region.

The suspicion of thiolase promoter homologous recombination into the bacterial genome does not seem to have been the case. Amongst the ten colonies selected from mutant selection plates, all of them matched the expected DNA fragment size of 1514 bp indicating no evidence of integration into the thiolase region.

4.3.9.4.5 Problems with Plasmid or Bacteria Strain Used

Another idea that needed to be tested was whether the *ermB* gene was lost from its original position in the plasmid. If this occurred in a large part of the pMTL-CW21 extracted from *E. coli*, this could explain the large percentage of the Erm sensitive colonies on mutant selection plates and a few mutants with unique insertion. Also contamination of the TraDIS selected strain with a plasmid bearing *catP* gene could cause this problem. Therefore it was decided that a third attempt at biological library creation was to be carried out.

4.3.10 Third Attempt at TraDIS

4.3.10.1 Transposon Library Creation - 3rd Attempt

Keeping in mind all the precautionary techniques applied in Section 4.3.9.4 to overcome the problem of plasmid reads in TraDIS, a third attempt was made at building a transposon library pool using fresh *C. acetobutylicum* ATCC 824 wild type and transposon delivery plasmid pMTL-CW21. Nine transformations of wild type strain with pMTL-CW21 were incubated on Erm selection plate for 36 hours. The transformants were flooded with CGM broth then inoculated on large square petri-plates containing mutant selection media CGM, Tm and aTC (100 ng/ml). To determine the quality of the library, serial dilutions of the transformants were plated on different media containing Tm, Tm + aTC, Erm and Erm + aTC to determine the relevant calculations as shown in Table 4.3. Following incubation for 48 hours, approximately 500,000 colonies were observed on the mutant selection plates.

Table 4.3: Transposon pMTL-CW21 library creation result for 3rd attempt.

Plasmid	Resistance to plasmid loss	Transposition Frequency	Plasmid Loss efficiency (%)	Plasmid loss by Replica plating (%)
CW21	2.5×10^{2}	1.67×10^{-4}	99.999998%	95%

4.3.10.2 Analysis of Inverse PCR Products for 3rd Transposon Library

Having observed that replica plating was not a sufficient quality control test to determine transposition effectivity, an inverse PCR test was carried out as described before. It was done to test the randomness and effectivity of the transposition activity. The results as shown in Figure 4.16 depict different DNA fragment sizes of the inverse PCR products from 25 individual colonies of mutant selection plates Tm^R Erm^S. There was a double band in lane 17 which could indicate the colony of two transposon mutants or a double transposition in the same cell genome.

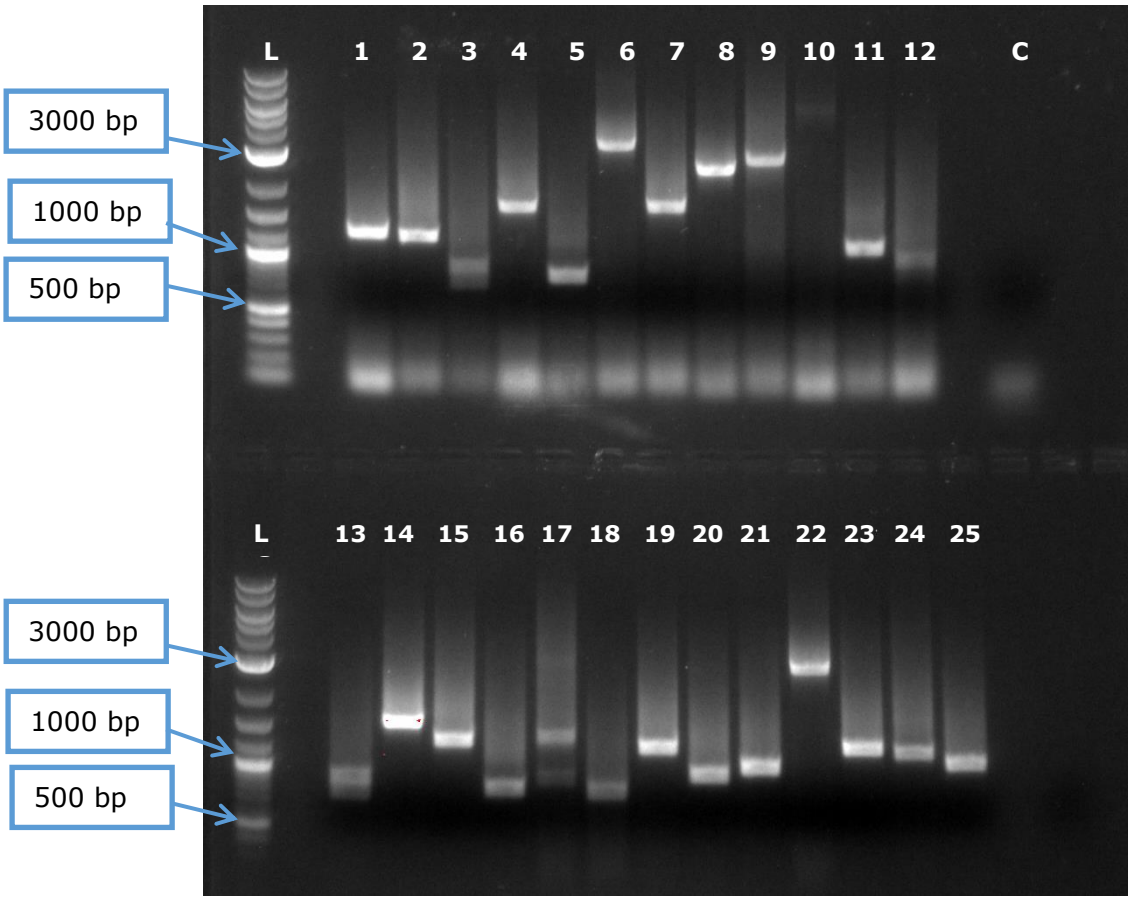

Figure 4.16: Inverse PCR screening obtained for 25 transposon mutant colonies with random insertions. L represents NEB ladder 0.1-10 kbp. Lanes 1 to 25 show different sizes of DNA fragments. C is the negative control with no DNA added.

To further confirm the randomness of the transposition, each DNA fragment was cut out from the agarose gel, extracted and Sanger sequenced. Locations of the transposon insertions are shown in Table 4.4. Out of the 25 colonies tested, all had unique insertion

locations except colonies/lanes 12 and 17A which had identical insertions in *CA_C2171* gene while the fragments in lane 10 mapped to pMTL-CW21. Like the case in the second attempt, cat-INV-R2 could not bind to the inverse PCR product of lane 10 so it was difficult to determine whether it was plasmid integration or mere plasmid remnant. The data suggested that the majority of the cells in the library were indeed transposon insertion mutants and there was no abundance of plasmid vector.

Table 4.4: Insertion positions of transposons in the *C. acetobutylicum* wild type chromosome

Number	Position of Transposon Insertion and Their Encoded Protein Product
1	*greA* gene (Transcription elongation factor)
2	*CA_C2150* (Flagella biosynthesis protein FliP)
3	*CA_C2958* (PTS System enzyme IIBC component galactitol/fructose specific)
4	*CA_C1852* (Magnesium and cobalt transport protein)
5	*CA_C1474* (Proline glycine betaine ABC transport system periplasmic component)
6	*CA_C3086* (Protein containing cell adhesion domain)
7	*CA_C2724* (Hypothetical protein)
8	*CA_C3411*(Homology with plant auxin-response GH3-like protein)
9	*CA_C0603* (superfamily I DNA and DNA helicase)
10	Plasmid integration
11	*CA_C2682* (Hypothetical protein)
12	*CA_C2171* (Predicted glycosyltransferase)
13	*CA_C0699* (Spore photoproduct lyase SpIB)
14	*CA_C2984* (Hypothetical protein)
15	*CA_C1577* (N-characterised conserved protein)
16	*CA_C0063* (Phage-related protein)
17A	*CA_C2171* (Predicted glycosyltransferase)
17B	*CA_C2445* (AICAR transformylase domain of PurH-like protein)
18	Between *CA_C 1065* (Uncharacterised protein, related stage III sporulation protein AH of *Bacillus* sp) and *CA_C 1066* (Hypothetical protein)
19	*CA_C3456* (Highly conserved protein containing a domain related to cellulose catalytic domain and thioredoxin domain)
20	*CA_C1794* (Predicted permease)
21	Between *CA_C1058* (CDS-predicted membrane protein) and *CA_C1059* (No annotation)
22	*CA_C1227* (Hypothetical protein)
23	*CA_C0460* (CBS-domain containing protein, YHDP *B. subtilis* ortholog)
24	*CA_C2445* (AICAR transformylase domain of PurH-like protein)
25	*cheB* gene (Chemotaxis protein CheB)

4.3.10.3 Growing Transposon Pool in Challenging Conditions

Since both the replica plate test and calculated plasmid loss efficiency data suggested a high rate of transposition events, the library was subcultured in media with and without butanol and high sodium ion conditions. This was done as soon as the library was created to prevent freezing before the library was subjected to the respective conditions, in order to avoid any possible negative effect of the low temperatures on the viability of the mutants. To achieve this, the OD_{600} of the library pool (slurry) was measured and used to inoculate 100 ml CGM broth with added Tm and aTC such that the starting OD_{600} was 0.1. Triplicates of the following conditions were prepared:

a. Control without any added agent for investigating essential genes as well as non-essential genes for survival in under standard laboratory condition
b. 8 g/l butanol for investigating genes for butanol tolerance
c. 8 g/l sodium chloride for investigating genes that promote sodium ion tolerance.

All the cultures grown in these conditions were monitored by taking the optical density. There was an observed growth retardation in cultures containing butanol and sodium ions with time when compared with the control. The respective TraDIS stock library samples were prepped and sent to the University of Nottingham DeepSeq facility for qPCR and Illumina sequencing.

4.3.11 Analysis of TraDIS Sequenced Data

A summary of the mapping statistics of the transposon library stock are shown in Table 4.5. Combining the stock triplicate runs for standard conditions only gave a total of 2,992,009 mapped reads with 90% mapping to the genome and 3% mapping to the transposon delivery plasmid. The low number of reads mapping to the plasmid indicates that the I-SceI digestion step was effective in removing the residual transposon plasmids from the library. A random distribution of the tagged reads throughout the genome was observed, which supports the conclusion by Cartman and Minton (2010) that mariner transposase is ideal for G+C low organism like *C. acetobutylicum* as it ensures random insertion around the genome. There were 28,641 and 2,003 unique insertions sites identified on the chromosome and pSOL1 respectively. This means that for the *C. acetobutylicum* ATCC 824 genome of 4,132,880 bp, 30644 unique insertion sites gave an average of one insertion in every 135 bp (Table 4.5).

Table 4.5: Summary of the mapping statistics of TraDIS library MiSeq sequencing data

Sample	Reads Mapped	% Mapped	Unique Insertion Sites	Average distance between insertions (bp)
Stock Library	2992009	90	30644	135

Bioinformatics analysis of the trimmed reads allowed for the identification of the number of unique insertion sites in each polypeptide-encoding gene, therefore by dividing this number by gene length, the insertion index of an ORF could be determined. The density frequency distribution of the insertion indexes yielded a bimodal plot (Figure 4.17). Usually, the frequency plot of highly dense library pools gives a clear bimodal curve enabling differentiation of genes required for cell survival in the first peak close to zero (essential genes) from those genes that are not required for cell survival in the second peak far right (Moule *et al*., 2014; Barquist *et al*., 2013, Langridge *et al*., 2009). Polypeptide-coding ORFs with insertion index above 0.0001 were assigned as probable candidates of essential genes for sustaining the bacterial cell life under the tested conditions. Of which 537 of the total annotated genome were assigned essential genes and did not have any transposon insertion. Insertion index value between 0.0001 and 0.0003 were assigned ambiguous region of which no genes fell in that category.

Using this criteria of insertion index above 0.0001, 537 genes (see appendix Table 7.1) out of the 3989 fulfilled the putative essential gene category which corresponds to 13.5% of the annotated genes. This is comparable in magnitude as 7.8% of the genes are essential in *Salmonella enterica* serovar Typhimurium (Barquist *et al*., 2013), 8.5% of the genes are essential in *Burkholderia pseudomallei* (Moule *et al*., 2014), 11% genes are essential in *C. difficile* (Dembek *et al*., 2015) and 13% genes are essential in *C. sporogenes* (Dr Raquel Rodrigues unpublished work). Amongst the 72 common aminoacyl-tRNA biosynthesis genes, 68 of them were considered essential. The four aminoacyl-tRNA biosynthesis genes with insertions were *CA_Ct027* (tRNA-lysine), *CA_Ct051* (tRNA-tyrosine), *CA_Ct055* (tRNA-alanine) and *CA_Ct061* (tRNA-asparagine) and all have another tRNA synthetase gene that could compensate for them in the putative essential list.

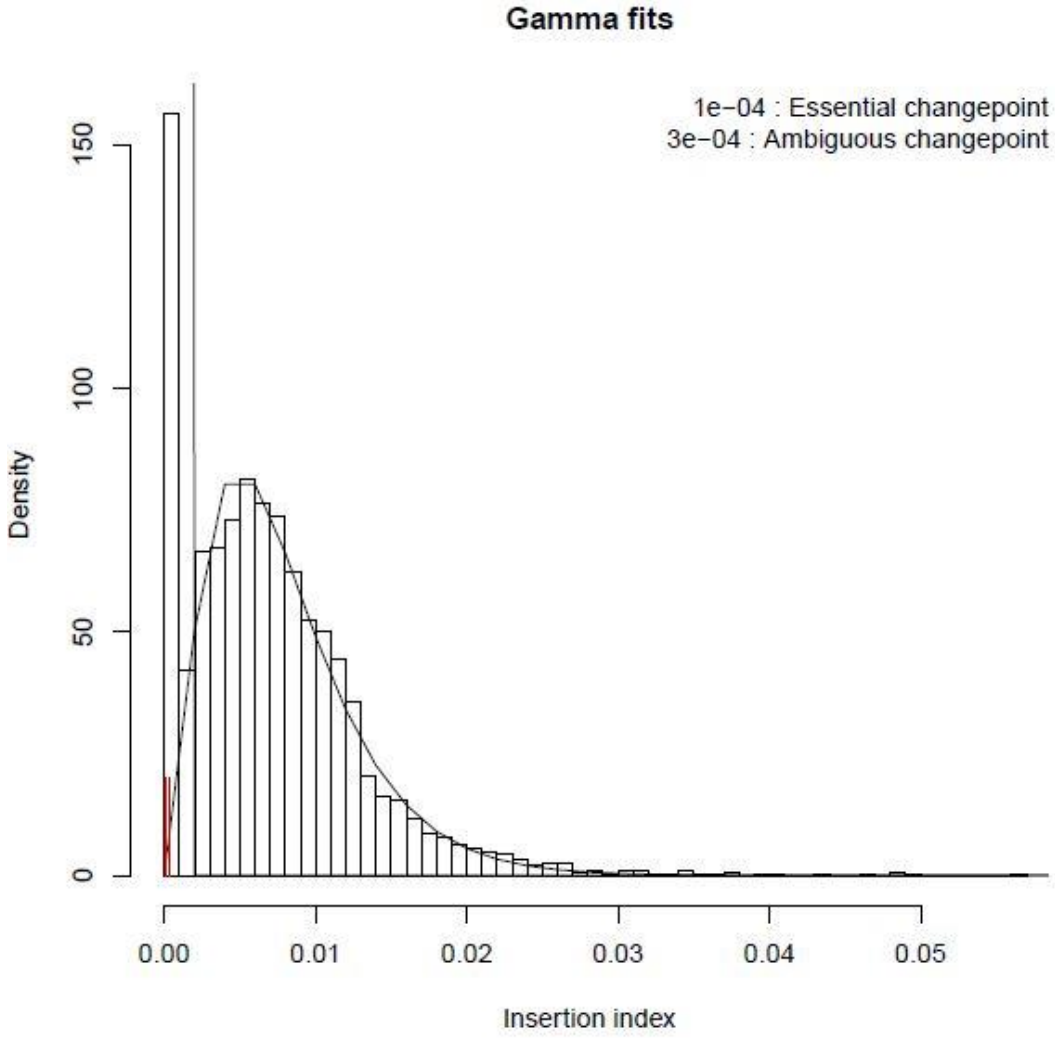

Figure 4.17: Frequency distribution plot of gene insertion index for the summary sets of TraDIS pool. The vertical black lines indicate the fitted gamma distributions of the insertion index from the TraDIS stock library. The gene insertion index is derived by dividing the number of unique insertions in a gene by the gene length. The broad peak on the right represents genes with several insertions while the long peak on the left represents potentially essential genes as they have no insertions. The red lines indicate the cut-off of insertion index for essential genes. This was generated using Bio-TraDIS toolkit approach as described by Barquist *et al*., 2016.

The functions of the candidate essential genes required for survival were identified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Accordingly, genes that code for components of DNA replication, transcription and translation machinery, carbon metabolism, amino acid biosynthesis and metabolism, aminoacyl-tRNA biosynthesis, purine metabolism, pyrimidine metabolism, lysine biosynthesis and protein export were identified as essential for life (see Table 4.6). Comparing the

predictive KEGG essential pathways of *C. difficile* (Dembek *et al*., 2015) with *C. acetobutylicum* is quite informative. As expected, both bacteria have essential genes that fall under categories required for key cellular functions such as cell division, replication, transcription and translation. Interestingly, 185 out of the 528 predicted essential chromosome genes encode proteins with unknown function based on detectable homologs including 19 conserved proteins, 156 hypothetical proteins and 10 genes with ATPase, peptidase and DNA-binding domain. This is not surprising as 32% of the annotated genome of *C. acetobutylicum* ATCC 824 belong to the conserved gene strings (Nölling *et al*., 2001).

Table 4.6: KEGG functional categories for candidate essential genes of *C. acetobutylicum* ATCC 824. Comparing the KEGG pathway categories for essential genes of *C. difficile* from Dembek *et al.* (2015) with that of *C. acetobutylicum* ATCC 824. na = no essential genes added in this category

KEGG Database	*C. difficile*	*C. acetobutylicum*
Ribosomal	46	56
Aminoacyl-tRNA biosynthesis	23	82
Metabolic pathways	104	22
Fatty acid metabolism	12	9
Peptidoglycan biosynthesis	12	6
DNA replication	9	7
Purine metabolism	18	5
Terpenoid backbone biosynthesis	9	5
RNA polymerase	4	3
Protein export	7	4
Lysine biosynthesis	9	2
Pyrimidine metabolism	14	2
Homologous recombination	8	6
Secondary metabolite biosynthesis	37	na
Oxidative phosphorylation	8	8
Mismatch repair	7	1
Folate biosynthesis	5	4
Carbon metabolism	18	8
D -Glutamine and D -glutamate metabolism	3	1
Glycerophospholipid metabolism	6	4
Microbial metabolism in diverse environments	25	na
Biotin metabolism	4	1
Propanoate metabolism	6	na
Pantothenate and coenzyme A biosynthesis	5	5
Amino acid biosynthesis	na	7
Quorum Sensing	na	2
RNA degradation	na	3
Sulphur Relay System	na	2
Amino sugar and nucleotide sugar metabolism	na	3
Amino acid Metabolism	na	10
Transcription factor	na	7
Translation Factor	na	7
Transcriptional regulators	na	10
Messenger RNA biogenesis	na	2
Chromosome and associated Proteins	na	9
Transfer RNA biogenesis	na	4

Chaperone and folding catalysts	na	4
Cell growth	na	7
Peptidases and inhibitors	na	1
Protein kinase	na	1
Ribosome biogenesis	na	6
Oxidoreductase activity	na	2
Transporter	na	6
Signalling and Cellular Processes	na	2
Energy Metabolism	na	3
Two component system	na	2
Bacterial chemotaxis	na	1

An obvious set of genes which is expected to be saturated with transposon tags are in pSOL1 (*CA_P001* to *CA_P178*). When degeneration of *C. acetobutylicum* occurs, there is loss of the megaplasmid (Ehsaan *et al*., 2016; Sabathe *et al*., 2002). Previous research had isolated pSOL1 mutants of *C. acetobutylicum* (Foulquier *et al*., 2019). However, 9 out of the 537 polypeptide-encoding ORF designated essential genes (which had no insertions) were from the megaplasmid. There are a total of 178 genes in pSOL1, thus indicating 5% false positives. This validates the essential gene list at 95% accuracy. Therefore, there might be a few ORFs in the 529 designated essential gene list which might not be essential. Nevertheless, it appears that the 9 pSOL1 genes without transposon insertions were small genes within the range 84 bp to 576 bp when the mean length in pSOL1 is 725 bp. It might be that compared to larger genes, the smaller genes had lower chances of having a transposon insertion or were not effectively captured. This might be due to the presence of fewer TA (preferred insertion sites of mariner transposase) or underrepresentation of these transposon mutants which makes it difficult to be captured during library prep. The coverage was not high enough to saturate the non-essential genes and depict with high confidence that the list of genes in Table 4.6 are all essential.

Another indication that coverage was not sufficient to tag all the non-essential genes is that all the genes without transposon inserts were all assigned essential by the Bio-TraDIS Analysis Pipeline script, however, in previous studies genes with few insertions can also be assigned as essential (Dembek *et al*., 2015, Dr. Raquel Rodrigues unpublished work). In addition, an average insertion density of 1 in 135 bp is quite poor especially when compared to previous studies achieving an average of one insertion in

every 54 bp (Dembek *et al*., 2015) and one in every 35 bp (Dr. Raquel Rodrigues unpublished work).

Despite the low saturation transposon-mutagenesis density of the library, many genes that have been previously knocked out from *C. acetobutylicum* by our SBRC Nottingham, University of Nottingham, UK team were not on the putative essential gene list. These genes include: *CA_C1502* and *CA_C3535* (restriction modification genes); *agrA, agrB, agrC* and *CA_C0082* (putative quorum sensing regulatory); *pyrE* (orotate phosphoribosyltransferase): *pheA* (prephenate dehydratase); *argH* (argininosuccinate lyase); *hbd* (3-hydroxybutyryl-CoA dehydrogenase); and *purD* (phosphoribosylamine-glycine ligase). In a recent study in *C. acetobutylicum*, a mariner transposase plasmid was used to create a transposon library for phenotypic screening which led to the discovery of the sRNA gene, (*CA_C2384*), as a crucial gene for cell growth (Yang *et al*., 2020). However, *CA_C2384*, which has 198 bp, was amongst the putative essential genes list. This false positive result could be as a result of reduced fitness or poor saturation of small genes. In spite of this, the evidence provided in this chapter showed that essential genes for survival of *C. acetobutylicum* can be investigated using the TraDIS and transposon vector elucidated herein. A complete list of putative essential genes are attached in the Appendix A (Table 7.1).

4.4 Conclusion

To acquire better understanding of cellular processes in *C. acetobutylicum* ATCC 824, the study of cells with the minimum number of genes required for survival could decode the complex biochemical reactions processes underlying its metabolism. Knowing which genes are dispensable and indispensable and their functions is vital in creating an ideal industrial strain for commercial solventogenic processes. Recent advances in transposon mutagenesis and high-throughput sequencing techniques have contributed to the robust assessment of gene essentiality in various bacteria. Transposon Directed Insertion-site Sequencing (TraDIS) is a forward genetics approach which has been applied to identify genes required for sustaining the cells' life in *Samonella typhi, Clostridium difficile, Burkholderia pseudomallei* K96243 (Langridge *et al*., 2009, Dembek *et al*., 2015, Moule *et al*., 2014) and a host of other pathogenic microorganisms. TraDIS has also been used in investigating conditional essential genes such as those responsible for bile tolerance in *S. typi* (Langridge *et al*., 2009) and

sporulation for *C. difficile* (Dembek *et al.*, 2015). However, there has not been any published research in which TraDIS was used to identifying essential genes in solventogenic bacteria.

The key element for transposon mutagenesis delivery was the selection of the vector. Though knocking out the two restriction systems, *CA_C1502* and *CA_C3535* genes, have been successful (chapter 3), the increase in transformation efficiency was not sufficient to build a large transposon mutants pool using a suicide plasmid. Replicative plasmids with conditional origin of replications and inducible condition/system for transposition, pMTL-CW20 and pMTL-CW21, were demonstrated to be superior for generating a large pool of transposon mutants.

For pMTL-CW20, 1 mM IPTG inducing agent did not seem to generate effective plasmid loss during the selection of mutants, thus an additional liquid culture step of transformants in the presence of IPTG and Tm was introduced before plating them on mutant selection plates. This technique seemed to have further boosted plasmid loss efficiency, but the TraDIS sequencing results revealed that there might have been dominance of some mutants over others in the pool during the liquid culture stage. The idea that some mutants were created in the liquid culture stage and might have had a head-start in multiplication posed a challenge because a good biological library creation step should eliminate biases of overrepresentation of a set of mutants over others which could yield false fitness results. Despite experiencing improved calculated plasmid loss efficiency from 68% to 99% when the IPTG concentration was increased from 2 mM to 4 mM, it proved difficult to generate a large scale transposon library with 4 mM IPTG.

To eliminate the described potential biases in transposon library creation and solve the observed plasmid loss challenge, an attempt to build a library pool using *C. acetobutylicum* ATCC 824 wild type and aTC inducible transposon delivery vector pMTL-CW21 was done. The calculated transposition frequency for pMTL-CW21 was significantly higher than that of pMTL-CW20. This means that the aTC inducible system seemed to have had a tighter transcriptional control than the IPTG system. In spite of the considerable higher plasmid loss demonstrated both by calculation and replica plating, sequencing results showed that plasmid loss challenges persisted. The reason for this persistence was unclear.

Carrying out inverse PCR step on random colonies on mutant selection plate prior to library preparation and sequencing steps helped reveal and authenticate the quality and diversity of the transposon pool. A third attempt at building a transposon mutagenesis library pool using wild type and pMTL-CW21 showed promising results as the inverse PCR product Sanger sequencing revealed random transposition events occurring in the genome. Though this step is expensive and time consuming, the inverse PCR step was valuable step for confirming the effectiveness of the inducing agent in promoting high transposition events and plasmid loss.

Given that the challenge of demonstrating effective plasmid loss and transposition event was solved, the library pool was sequenced and analysed. Despite the distribution insertion indexes giving a clear bimodal plot, the non-essential genes in the TraDIS library was not saturated enough to identify essential genes with high statistical confindence. The list of the 537 potentially essential genes contained 9 genes from the megaplasmid which are known to be not essential. The COVID-19 lockdown affected the progress of this research in creating a highly dense TraDIS library.

This experiment stands as a proof of concept for the prospect of applying Transposon Directed Insertion-site Sequencing of a highly saturated library pool in *C. acetobutylicum* for deducing essential genes. The COVID-19 lock down interrupted the course of this work but there are indications that a highly saturated library could be produced with the pMTL-CW21 transposon delivery vector. Apart from gene essentiality investigations, TraDIS analysis of a saturated mutant pool under certain stress/specific conditions could possibly enable the annotation of uncharacterised proteins. *C. acetobutylicum* has a large number of genes with unknown functions and elucidating their role in the cell metabolism and fitness are all important aspects of developing a commercial strain for solvent production.

Key points about this chapter

- Despite the deletion of both restriction modification systems, the transformation efficiency was not sufficient to build a highly saturated transposon library using a suicide delivery vector
- pMTL-CW20 and pMTL-CW21 were demonstrated to have potential in generating highly saturated pools of transposon mutants for TraDIS studies

- Introducing a liquid culture treatment step of pMTL-CW20 transformants lead to the dominance of some mutants over others in the pool
- aTC inducible system present in pMTL-CW21 seem to be more tightly regulated inducible system than IPTG inducible system present in pMTL-CW20
- Using TraDIS, a list of 537 genes with an insertion index of less than 0.0001 insertion index were considered to be potential candidates for essential genes.

Chapter 5

Isolation and characterisation of lysogenic phage

Chapter 5

5.1 Introduction

Phages out-number bacterial populations by almost a factor of 10 (Brüssow *et al*., 2002). There are an estimated 10^{31} bacteriophages worldwide, therefore for every bacterium there are 10 phages (Brüssow *et al*., 2002). The ubiquitous nature of phage indicates that their impact cannot be overestimated (Steczkiewicz *et al*., 2021). It is well-known that phages shape the bacterial communities, their virulence and evolution. Phage biologists understand the importance of unravelling the complexity of phage-host relationship (Fortier and Sekulovic, 2013) in a bid to solving global problems. In other words, while carrying out studies to understand a bacterium, it is vital to study its temperate and virulent phage(s).

In the last decade, there has been a boom in the publications on phage, however, the drive has mostly been directed towards phage therapy as an alternative to antibiotics in the clinical world (Górski *et al*., 2018; Furfaro *et al*., 2018; Necel *et al*., 2020). Although the influence of phage on the virulence of pathogenic bacteria is well studied, there are surprisingly little genetic studies of phage of saccharolytic bacteria, which are of great importance in production of alternative fuels. One of the major challenges faced by the Acetone Butanol Ethanol (ABE) industry was recurrent phage attacks (Jones and Woods, 1986). Figure 5.1 shows the recorded phage attacks in early ABE industry such that the facility at Terre Haute, Indiana USA experienced 12 attacks a year which contributed to relocating the facility to Pretoria. However, there is surprisingly little information on the lifestyle of phage within the bacteria. Despite the precautionary techniques of constantly developing phage-resistant strains and strict sanitary techniques, phage attacks were often responsible for fermentation failures (Jones *et al*., 2000). When those phage reports were made, scientists did not have the current advanced microbiology tools now available to us. Progress towards phage genetics could guide schemes towards the development of commercial phage-resistant strains and development of vectors for metabolic engineering. The study of their evolution of genomes, their replication and transcription could give insights to the development of biotechnology products (Harada *et al*., 2018).

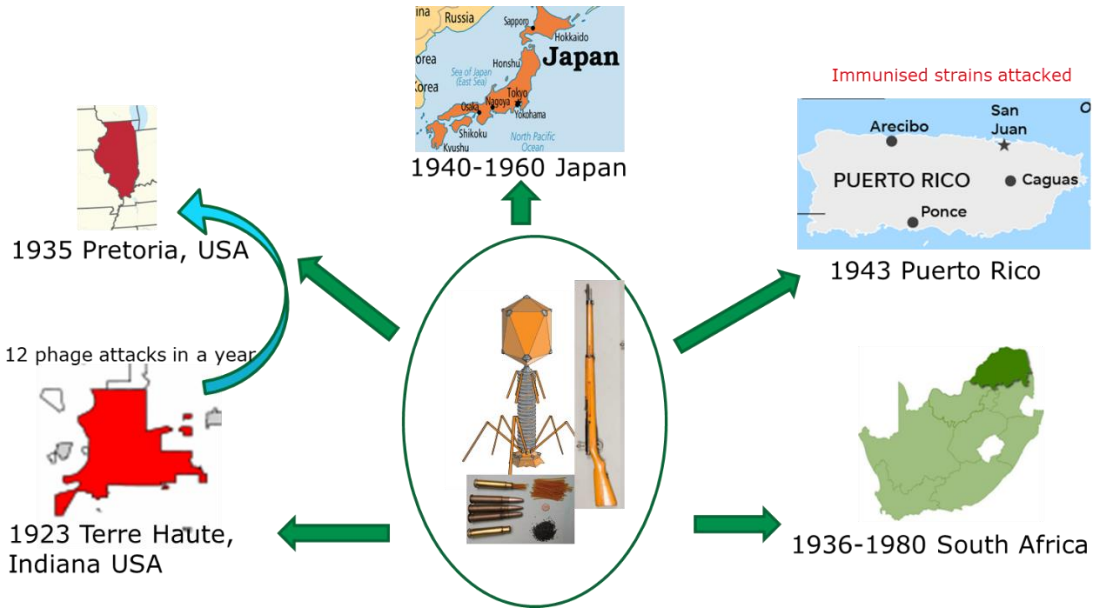

Figure 5.1: Phage attacks in the ABE industries in various countries. Source of information Jones *et al.* (2000). The ABE facility in Terre Haute, Indiana USA experienced about 12 attacks in a year. The recurrent phage attacks was one of the major factors for moving the facility to Pretoria.

5.1.1 Life Cycle of a Phage

In order to explore phage for molecular research purposes, it is vital to know the intricacies of phage operations (Yang *et al*., 2014). The phage–host interactions are so complex that scientists have only been able to scratch the surface of uncovering their characteristics. The most common phage life cycles are the lytic cycle (also known as the virulent cycle) and the lysogenic cycle, (also known as the temperate cycle). During phage infection, the phage attaches to the bacterial cell surface using specialised phage tail tip complex or baseplate. These specialised tail structures recognise specific receptors binding protein on the host cell wall and is the first point of contact with the host. As soon as its tail fibers bind to the receptors, the phage digests the cell wall mechanically as well as enzymatically using lysozymes (Kizziah *et al*., 2020). This leads to the penetration of the plasma membrane and injection of phage genome into the host cell while the "empty-headed" capsid remains outside the cell (Kizziah *et al*., 2020). At this stage the phage can proceed into the lytic cycle or lysogenic cycle.

During the lytic cycle of a virulent phage like T4 of *E. coli*, the phage uses the host machinery to replicate its DNA and produce phage encoded proteins, including capsids. Phage synthesis begins with the expression of endonucleases to degrade the host

chromosomal DNA. The phage DNA forms a circular permutation to avoid degradation by the host restriction modification systems. In some cases during replication, phages assemble as a nucleus-like structure to segregate the DNA and DNA assembly proteins from other cell components. Phage-encoded spindle-like tubulin, PhuZ in *Pseudomonas chlororaphis* and *FtuZ* in *E. coli*, positions the phage nucleus in the midcell during replication (Chaikeeratisak *et al*., 2017). Mutations in this structure could lead to a compromise on infectivity (Chaikeeratisak *et al.,* 2017). Replicated DNA are then packaged into capsids to form progeny which accumulates while phage-encoded lysin lyses the cell wall and free virions burst out of the cell. Virulent phages always lead to the death of the bacteria, on the other hand temperate phages can activate the lytic mode or lysogenic cycle where attacks do not lead to the death of the bacteria (Fortier and Sekulovic, 2013).

Alternatively in the lysogenic cycle, after the injection of the phage DNA into the cell in a linear form, it becomes circular and replicates in plasmid form like observed for P1 of *E. coli* or integrates into the chromosome by site-specific recombination like lambda phage of *E. coli* (Lobocka *et al.,* 2004). There are also unusual forms of temperate phages like N15 of *E. coli* that exists/replicates in linear form with covalently closed ends (Ravin, 2015). The integrated phage genome replicates simultaneously with the host chromosome by existing in a dormant state. A phage in a dormant state is known as a prophage. (Fortier and Sekulovic, 2013). Prophages do not degrade the cell and no new progeny is formed. Environmental factors, such as stress, antibiotics and UV irradiation, could trigger the host to release an SOS response as a result of DNA damage, this may lead to the prophage's excision and the production of free progenys. Lysogens are resistant to superinfection by a homologous virus but not by heterologous one. There is little understanding of why a temperate phage enters the lytic cycle or lysogenic cycle.

A third cycle in the phage-host relationship called pseudolysogeny was proposed by Ripp and Miller (1998), in which infecting viral genomes exist in an unstable relationship with starved bacterial cells. The condition was dependent of the availability of nutrients. In their experiment, when nutrients are replenished, phage form true lysogens in *Pseudomonas aeruginosa* strains or free virons. However, under nutrient depleted conditions, the phage genome did not commit to integration into host genome

(lysogeny cycle) nor creating free virions (lytic cycle) (Ripp and Miller, 1998). In another study, a pseudolysogenic relationship was demonstrated between *Salmonella enterica* serovar Typhimurium LT2 and phage P22 (Cenens *et al.*, 2013) where the phage genome consistently remained unintegrated in the *S. enterica* serovar Typhimurium LT2 chromosome and was asymmetrically segregated into its siblings. They proposed that pseudolysogenic condition could give rise to siblings that are lysogens as well as those which have lost the phage genome (Cenens *et al*., 2013).

5.1.2 Classification of Bacteriophages According to Their Morphology and Genetic Materials

Based on the type of genetic information stored in their capsids, phages are divided into four major forms; single-stranded RNA phages, double stranded-RNA phages, single-stranded DNA phages, double- stranded DNA phages (Harada *et al*., 2018) (see Figure 5.2). Most of the phages published on International Committee of Taxonomy of Viruses (ICTV) database belong to the order *Caudovirales* as they are tailed, double-stranded phages (Adams *et al*. 2015). *Myoviridae, Podoviridae* and *Siphoviridae* are the families under the order *Caudovirales,* whose classification is based on tail morphology. *Myoviridae* have a long contractile tail, *Podoviridae* have short non-contractile tails while *Siphoviridae* have a long non-contractile tail. They generally comprise an icosahedral head containing the genome connected to a flexible tail. Their genome sizes range from 18 kb up to 150 kb (Lavigne *et al.*, 2012). Other families of phages without tails that have double-stranded DNA are *Corticoviridae, Fuselloviridae*, *Tectiviridae, Plasmaviridae*, *Rudiviridae* and *Lipothrixviridae* (Harada *et al*., 2018) as shown in Figure 5.2.

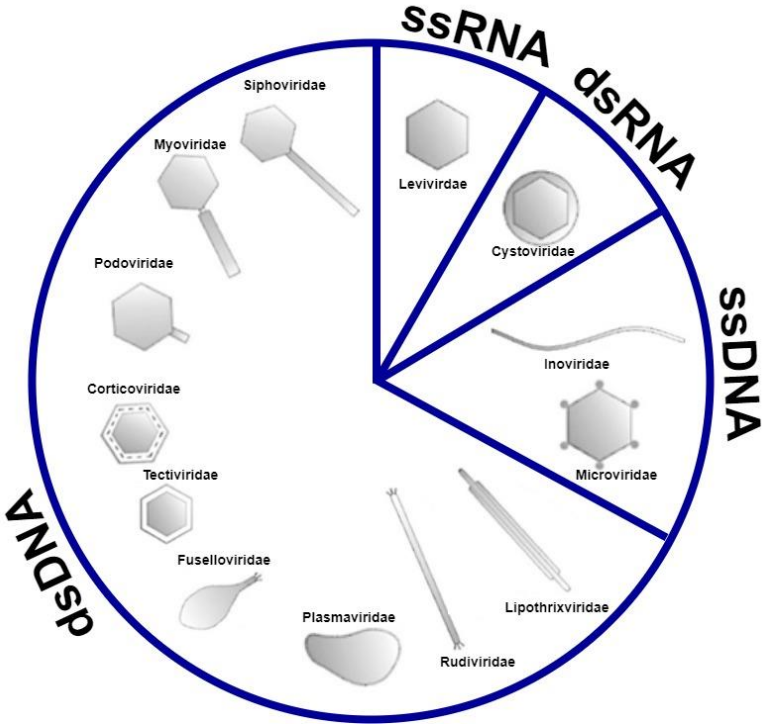

Figure 5.2: Classification of phages according to their genetic material and morphology. Modified from Harada *et al*., 2018**.**

5.1.3 Conditions That Affect Phage Infecting Host Cells

The most common/typical phage morphology comprises an icosahedral head with the genetic material embedded in it, a tail sheath and tail fibre protruding at the tail base. Infection of bacteria by phage occurs upon adsorption of the bacteriophage tail fibre to the bacterial host (Harada *et al*., 2018). Adsorption occurs by the recognition of nonspecific receptors in the bacterial cell wall and binding to it via electrostatic forces reversibly, followed by binding of phage capsid protein to specific bacterial surface receptor irreversibly (Harada *et al*., 2018). Keys to successful infection start with conditions that affect the phage and bacteria physiology.

Ionic strength of an environment can promote or deter phage adsorption to the bacterial cell surface. Many phages have been known to have increased infectivity in the presence of divalent ions like calcium and magnesium (Bandara *et al*., 2012; Chhibber *et al*., 2014; Tominaga and Enomoto, 1986). Foddai et al. (2009) observed a 2-fold increase in D29 mycobacteriophage infectivity on *Mycobacterium avium* ssp. *paratuberculosis* when $CaCl_2$ was increased from 1 mM to 2 mM.

Various explanation have been proposed for the effect of metals on host-phage interactions during infection. Foddai *et al.* (2009) suggests that calcium ion stimulates the ATP driven activity connected with tail contraction, thus promoting the injection of the phage genome into the host cell. Cvirkaitė-Krupovič (2010) and his colleagues proposed that PM2 phage-induced opening of the host outer membrane in *Pseudoalteromonas* host cells is dependent on the concentration of calcium ions being at least 2.5 mM. Calcium ions are suggested to polarise the host cell membrane thus promoting phage DNA transfer. The maintenance of the phage protein structures is greatly influenced by metal ions in the environment. Some phages encode metal-dependent hydolyases responsible for lyses of the bacterial cell membrane to allow the progeny escape after completion of the lytic cycle (Mikoulinskaia *et al*., 2009). In another report, it was shown that during infection, P2 phage penetrated the host cell wall using tail iron spike proteins (Browning *et al*., 2012). Magnesium and manganese ions have also been reported to replace the role of calcium ions in P2 and other 936-group phage infectivity suggesting phage adaptability (Mahaony *et al*., 2015). Calcium ions and magnesium ions have been shown to be vital for efficient inversion reactions by phage DNA integrases from the invertase family (Johnson, 2015).

The pH of the environment is another factor that influences phage infection. Kerby *et al.* (1946) have shown that the extreme pH affected the stability of the T7 phage particle while infecting its host, *E. coli*. They observed that the optimum pH for stability in T7 phage infection on *E. coli* was pH 6 to 8 and for the stability of the physical phage particle was pH 5-10. They showed that the phage particle degraded into two components at extreme alkalinity and acidity (Kerby *et al*., 1946). Phage PM2 loses complete activity when placed at pH 5 for one hour at 37°C (Chatain-Ly, 2014). Another factor influencing phage attachment is the temperature of the environment. In as much as temperature affects the growth of the bacteria, when the bacterial fitness is negatively affected, it in turn hinders phage attachment to the bacterial cell surface receptors. Temperature also affects lysis enzyme activity in degrading cell membrane and wall to allow new progeny to escape the cell. Temperatures as low as 4°C did not allow T4 host cell lysis, while temperatures above 45°C rendered T4 phage inactive (Taj *et al*., 2014). The optimum temperature for the phage was found to be 37°C.

Bacteria developed mechanisms to prevent phage attacks through their restriction modification systems (RMS) and Clustered Regularly Interspaced Short Palindromic

Repeats (CRISPR) systems. Restriction endonucleases of RMS recognise specific short sequences of foreign DNA and cleave it while methylases of the RMS methylate the host gene thereby protecting it from self-restriction (Pyne *et al*., 2016). If a phage is methylated by host methylase genes, it will be protected from restriction and its progeny will inherit that protection. In such a case, the progeny will carry on infection without being destroyed by the host endonucleases (Enikeeva *et al*., 2010). The determinant of whether a phage infection is successful or not is dependent on the competition between host endonucleases and methylases, depending on which is able to reach the phage genome first.

Some phages have evolved to evade host restriction systems by having little or no restriction recognition sites for host endonuclease cleavage, carrying phage-encoding methylase genes that methylate their phage DNA or encoding anti-restriction proteins that deactivate the host endonuclease (Enikeeva *et al*., 2010). Some temperate and virulent phages hold one or more methylase genes to protect them from host endonucleases (Murphy *et al*., 2013). For example, *Bacillus* temperate phage SPβ encode multi-specific methylases that methylate 2 recognition sites such that when induced, the methylase genes get activated and protect the excised phage genome (Murphy *et al*., 2013). Similarly, phage 933W carrying Shiga toxin (Stx2) genes require its Dam methylase genes for maintaining lysogeny in *E. coli* 0157:H7 (Murphy *et al*., 2008). Murphy *et al.* (2013) hypothesised that the origin of the methylase systems in phages is a result of horizontal gene transfer (HGT) between lytic or lysogenic phages and their host DNA encoding restriction modification systems. As phages evolve, they keep the acquired genes which will be of benefit to them, such as DNA methylases.

Similarly, CRISPR systems involve Cas (CRISPR associated) protein endonucleases to attack invading DNA. CRISPR systems confer adaptive immunity by acquiring sequence tags or protospacers flanked by short palindromic repeats, often 24-47 bp, from previous phage attacks or plasmids. CRISPR sequences acquired serve as a guide for Cas9 endonucleases to 'attack' future phage genome or plasmid DNA (Pyne *et al*., 2016). The bacterial genome is protected because it lacks an additional highly conserved sequence in the CRISPR array called protospacer adjacent motifs (PAM) (Tao *et al*., 2018). As new spacer sequences are acquired from phage attacks, old spacer sequences are excised from prokaryotic genome. This shows the highly ever-adapting nature of CRISPR systems (Pyne *et al*., 2016). The CRISPR Cas9 type II system from

Streptococcus pyogenes is the simplest CRISPR system studied (Tao *et al.*, 2018). The main CRISPR Type1-B locus in saccharolytic *Clostridium pasteuranium* has been identified (Pyne *et al*., 2016). Some phages have evolved systems to evade the CRISPR attacks by mutating the protospacers or protospacers adjacent motifs, e.g., *Streptococcus thermophilus* phage (Deveau *et al*., 2008) or encoding anti-CRISPR genes, e.g., in *Pseudomonas* phages (Bondy-Denomy *et al*., 2013). There are no endogenous CRISPR system present in *C. acetobutylicum* ATCC 824 (Wasels *et al*., 2017).

5.1.4 Using Phage Integrases for Site-specific Recombination Technology

Temperate bacteriophages are gifted by nature with enzymes that catalyse the cleavage, translocation and ligation of the phage genome into the host genome. These enzymes are called phage integrases or site-specific recombinases. There are a number of advantages to the deployment of site-specific recombination (SSR) tools for genetic engineering over homologous recombination technology (HT) like ACE. Firstly, SSR does not require a matching sequence of a certain minimum length as recombination is controlled by recombinases which form protein-DNA and protein-protein complexes leading to DNA exchange. HT requires longer identical sequences for DNA-DNA interactions that lead to DNA loci exchange. Secondly, site-specific recombinase enzymes (except transposition reactions and some 'unconserved' types) form recombinants without costing the cell energy or need for co-factors while HT may require DNA synthesis and use of host ligase to join the homologous DNA at the respective loci (Stark, 2014).

To put it in perspective, an example of recombinases which do not require identical recognition sequences at the phage and bacterial chromosome attachment sites in order to carry out their recombination activity is *E. coli* lambda phage recombinase. However, CRE and FLP catalyse recombination at identical phage DNA attachment sites (*attP*) and bacteria DNA attachment sites (*attB*) sites (Stoll *et al*., 2002). In addition, some recombinases act autonomously and do not require cofactors, for example, the CRE and FLP-FRT tyrosine recombinases from phage P1 and *Saccharomyces cerevisiae*, respectively. An exception is the lambda phage recombinase that requires some co-factors which has made it undesirable as a genetic tool (Stoll *et al.,* 2002). Furthermore, the production of irreversible reactions by unidirectional recombinases is another

property that makes it an attractive genetic engineering tool. This means that after recombination, the resulting *attL* and *attR* sites cannot be substrates for homologous recombination (Stoll *et al*., 2002). One major disadvantage of using this system as a genetic tool is that the bacteria DNA attachment site must be present or first introduced at the genomic location for recombination.

In certain cases, recombinase enzymes require some accessory genes and recognition sites in addition to *attB* and *attP* to carryout recombination. In some other cases, the presence of the recombinase and its recognition *attB* and *attP* sites is enough for the reaction to take place like Tp901-1 integrase phage of *Lactococcus lactis* (Christiansen *et al*., 1994). TP901-1 integrase is different because the integrase activity is carried out by a resolvase-like protein that uses a serine residue to form covalent bond instead of tyrosine. The resolvase-like protein introduces double stranded cuts to create 2 bp overhangs at the *attP* site as well as the *attB* site. The serine residue of the integrase dimer binds to the cleaved sites of *attB*, this forms a "synaptic complex" by binding with another recombinase dimer bond to cleaved complementary *attP* site (Stark, 2014). The recombinase tetramer complex catalyses the exchange of the cleaved sites by dimer rotation, pairing bound strands and ligating the staggered ends to generate a recombinant with *attR* and *attL*. The hybrid attachment sites contain half of the sequences in *attB* and half of *attP* (Fogg *et al*., 2014). TP901-1 integrase has a common sequence core of 5' TCAAT '3 which is shared with all the recombination sites (Breüner *et al*., 2001).

According to Stark (2014), all known serine recombinases cut at the recognition site to create a 2 bp staggered overlapping sequence in the opposite direction, unlike tyrosine recombines which cut to create 6 to 8 bp single-stranded sticky-end overhangs. Apart from TP901-1 resolvase-like integrase, another type of serine recombinase system is DNA invertases like Gin from bacteriophage Mu in *E. coli*. Gin with the help of FIS (the host factor for inversion stimulation) and the recombination enhancer catalyses site-specific inversion of DNA between inverted repeat sequences (Klippel *et al.,* 1988). Gin recombinase together with FIS/recombinational enhancers catalyse inversion of 3015 bp known as the G segment of the phage genome (Johnson, 2015). The G segments contains part of the tail fibre genes, therefore inversion alters the tail fibre which can lead to an increase in the phage host range (Johnson, 2015). Gin

recombinase creates single strand nicks at the top strand and bottom strand of the imperfect palindromic inverted repeats sequence (IR) 5' TTCCTGTAAACC/ GA/CCTTTTGGATAA 3' in a staggered manner to yield 2 bp overhang like resolvase-like recombinases (Klippel *et al.,* 1988). Serine residues of the Gin dimer forms a phosphodiester bond with the 5' nicked region of IR (Klippel *et al*., 1988). FIS activates the conformational change in Gin for the process of inversion to take place. Formation of synaptic tetramer complexes precede a 180°C rotation of the dimer for the exchange and inversion, then with the aid of the host factor recombination enhancer binding to FIS, ligation of inverted staggered ends generates a recombinant with *attR* and *attL* (Ritacco *et al*., 2013). Very low inversion rates occur without the recombination enhancer in DNA invertase recombination (Ritacco *et al*., 2013).

The major difference between resolvase-like recombinases and DNA invertase recombinases is that while resolvases delete/integrate double-stranded DNA, DNA invertases invert double stranded DNA. In addition, some resolvase-like recombinases do not require host factors and recombination enhancing genes to carry out their activity, although some DNA invertases require host factors, FIS and recombination enhancers to catalyse the recombination reactions (Johnson, 2015).

Unlike the serine recombinases, tyrosine recombinases, like lambda phage integrases use tyrosine residues to form covalent bonds with DNA. Tyrosine recombination follows a two-step reaction. Two of its four monomers create single-stranded breaks of about 6 to 8 bp distance at the *attP* and *attB* identical sites, then new partners from the phage and bacteria bind to the tyrosine residue of the recombinase enzyme. The strands are subsequently exchanged and ligated to form Holliday junctions close to the middle of the overhangs (Breüner *et al*., 2001). In the second step, the Holliday is resolved by the other two integrase monomers cleaving the Holliday junction intermediate which is proceeded by exchanging and ligating the strands to catalyse recombination of the phage into the bacterial genome (Breűner *et al*., 2001; Ritacco *et al*., 2013).

5.1.5 Phage-host Interactions and Evolutionary Conclusions

Horizontal gene transfer is recognised as a rapid means of evolution (Brüssow *et al*., 2004). Transduction, transformation and conjugation are the three typical mechanisms of transfer of genetic material between bacteria, of which phages employ transduction as an evolutionary weapon for the transfer of DNA/genetic information. In the past,

there were arguments that all phage-encoding genes were switched off in the lysogenic state, and thus the prophages played a silent role in the life of bacteria. Nevertheless, the concept of lysogenic conversion has been proven and disputes this old hypothesis. Lysogenic conversion shows that some parts of a phage genome can be expressed while in lysogeny. Some prophages carry genes that change the phenotype of their host thereby affecting its fitness. In fact for bacteria like Shigella toxin-producing *E. coli, Vibrio cholerae, C. botulinum* and *Corynebacterium diphtheriae*, their virulence is highly dependent on the prophage-encoded toxin (Brüssow *et al*., 2004).

On this basis, lysogenic conversion is not limited to virulence genes, other non-pathogenic bacteria benefit from prophage relationship too (Canchaya *et al*., 2003). Research analysing and comparing the presence of specific prophages sequences in a large group of closely related bacteria revealed that prophages contribute to differences in pathogenic and non-pathogenic strains (Canchaya *et al*., 2003). Examples of fitness factors encoded in prophages that may be expressed during lysogeny include lipopolysaccharide-modifying enzymes, ADP-ribosyl transferase toxins, type III effector proteins, serum resisting-proteins, detoxifying enzymes and hydrolytic enzymes (Brüssow *et al*., 2004). It is very common to find these fitness genes or phage-encoded genes with no known function in Gram-positive bacteria with low GC (Brüssow *et al*., 2004).

Apart from lysogenic conversion, other means by which prophages can affect the phenotype or fitness of their hosts include disruption of host genome during integration. Bombay and his co-workers (2013) identified 15 genes that were disrupted as a result of phage integration in the genomes of *Salmonella enterica* and *E. coli*. In another case, *Pseudomonas* temperate phages DMS3, D3112 and MP22 disrupted virulence of its host (Drulis-Kawa *et al*., 2015). A third means of prophage influence on the host is through phage induction of a competing strain as well as rearranging host genome (Brüssow *et al*., 2004). Finally, the fitness of the host can be affected by the protection from phage attacks. It may offer immunity against attack from related phages which use similar *attB* sites but cannot prevent infections by other phages which use other *attB* sites in the host genome. Evolution and the reoccurring acquisition of many phages may lead to an inactivated prophage losing its function due to SNPs or other mutations or incomplete prophage genome. Such phages cannot be induced and their residues are seen in host genomes (Brüssow *et al*., 2004). The long term evolutionary relationship

between phages and bacteria is evident in the numerous isolated phage-like integrase genes found in bacterial chromosomes as leftover from previous integration events (Brüssow *et al*., 2004).

5.1.6 Exploring Phage Studies for the Development of Industrial Strains

Some of these qualities or fitness methods can be explored in designing an industrial strain for butanol production. For obvious reasons, the presence of a prophage could cost the cell some energy due to replication of phage genome and DNA translocation to daughter cells. For example, lysogenic *Clostridium madisonii* 4J9 strain, when used for solvent production, was found to grow slower than the non-lysogenic strain 214 and its solvent production was reduced as well (Jones *et al*., 2000).

Despite this shortcoming, phage loci can be manipulated to carry beneficial genes that will improve solvent production, while eliminating unnecessary genes that are extra baggage. Compared to *E. coli*, *C. acetobutylicum* does not transform as efficiently. Despite the existence of several plasmids created for effective genetic manipulation of genes in *C. acetobutylicum*, there is a limit to the size of insert they can carry into the bacterial cell. The larger the plasmid, the lower the transformation efficiency (Karimi *et al*., 2016). Phages are highly specific for their target and their self-regulating systems, such as a limitation of propagation to specific cells, make them versatile and modifiable carriers in targeted delivery systems (Karimi *et al*., 2016; Harada *et al*., 2018). Phages as nano-vehicles can be explored for delivering entire novel metabolic pathway beneficial for solvent production because it can allow the transfer of larger DNA sequences than transformation with plasmids.

In another prospect, identification of the phage integrase and its attachment sites in both bacteria and phage can be explored as a gene editing tool. Site-specific recombination technology is well-studied and has been explored in many bacteria and higher organisms. On the other hand, mutating the *attB* can be explored to prevent infection by temperate phages in industrial important strains.

5.1.7 Previous Studies of Phage in *Clostridium acetobutylicum* and Other Saccharolytic Bacteria

The recent boom in publications on phage studies are focussed on studying phages for phage therapy applications. These studies concentrate on creating lytic phages and

using its phage-encoded proteins like holin and endolysins to fight pathogenic bacteria in the human body (Drulis-Kawa *et al*., 2012). However, there are limited publications on saccharolytic bacteriophage studies and its exploitation for advances in biofuel technology industry despite the fact that the ABE industry is one of the oldest commercial fermentation industries. Phage studies could be the key to developing the ideal commercial strain for alternative fuel production that could compete with the fossil fuel derived products. There is need for more research driven at understanding phage mechanisms and unlocking the treasures of the encoded proteins for biotechnology purposes.

In the early years of the ABE industry, the industry was tormented by re-occurring phage attacks. Scientists kept working on strains that were resistant to phage attacks and a few were patented (Jones *et al*., 2000). The method of creating these phage resistant saccharolytic strains was to expose them to phages for a period of time and then test the survivors for susceptibility to phage infection using plaque assay. If plaques were not obtained, the strain was designated, as "phage resistant" strains. The problem with this method is that it could lead to the creation of lysogenic bacteria which can enter the lytic cycle when subjected to stress. Besides this, when used for ABE fermentation, the phage resistant strains were further subjected to phage attacks. The phage attacks were characterised as causing slow growth of bacteria and poor yield of solvents. During this era of ABE industrial boom, phage CA1 of the *Podoviridae* family was isolated from *C. acetobutylicum* and then in 1983, Reid and his co-workers worked on this phage (Reid *et al*., 1983). *C. acetobutylicum* phage CA1 isolated from a fermenter bacteriophage outbreak on strain P262 (now reclassified as *Clostridium saccharobutylicum*) was directed at transforming the phage DNA into *C. acetobutylicum* protoplast (Reid *et al*., 1983). They reported the release of matured viral particles by the transformed protoplasts. Phage CA1 had an icosahedral head, a very short tail connected with 12 tail fibres at the end. It was shown to have a pseudolysogenic relationship with its host (Jones *et al*., 2000).

Another publication described the isolation of virus-like particle from *C. acetobutylicum* NCIMB 6444 strain (Kim and Blaschek, 1991), now named *C. beijerinckii* NCIMB 6444. They stated that the filamentous virus-like particle was single-stranded DNA about 6.6 kb and constantly expelled from solventogenic bacterial cell wall into the culture (Kim and Blaschek, 1991). In another publication in 1993,

Kim and Blaschek ligated the 6.6 kb filamentous single-stranded phage DNA of the *C. beijerinckii* NCIMB 6444 strain with *E. coli* plasmid pAK102 (5 kb) and showed that the phagemid (hybrid) was functional in *E. coli*, *C. beijerinckii* and *C. perfringens* as they all were transformed by the phagemid.

For the first 46 years of early ABE production, there were several reports of phage attacks on the original *C. acetobutylicum* strain isolated by Chaim Weizmann in the National Chemical Products (NCP) factory in Germiston, South Africa plant. Phages from the family *Myoviridae* and *Siphoviridae* were isolated from sluggish growing cultures. Attempts to grow some of them in the lab were not successful. At a point in time, the facility in NCP closed down for ten days to clean and disinfect the plants because of the phage attacks (Jones *et al*., 2000). Other bacteriophage infections include virulent HM phage variants from *Clostridium saccharoperbuylacetonicum* in Japan and phages A, B, C and D infecting *Clostridium madisonii* (now known as *C. beijerinckii)* at the plant in Puerto Rico. Some of these phages mentioned were not characterised (Jones *et al*., 2000).

5.2 Hypothesis and Aim of Chapter

To isolate, characterise, analyse the genome, integration system and transduction capabilities of a lysogenic phage in *C. acetobutylicum* ATCC 824 and investigate whether insertion of *tcdR* gene in this *Clostridium spp*. is responsible for inducing phage.

5.3 Results and Discussions

5.3.1 How Phage Activity Was Discovered

Previous work described in chapter four had raised the question: is the insertion of *tcdR* in *C. acetobutylicum* responsible for phage induction? In order to carry out transposon mutagenesis described in chapter four, *tcdR* was inserted downstream of the *pyrE* gene while simultaneously repairing the *pyrE* gene in *C. acetobutylicum* Δ*pyrE*Δ*CA_C1502* strain using ACE. To select an ideal TraDIS strain with no significant SNPs, whole genome sequencing was carried out on the parent strain, *pyrE::tcdR* repaired strain as well as restriction-less strains. Bioinformatic analysis of the sequencing data revealed the absence of data between positions 2,016,384 to 2,069,759 in the *tcdR* inserted strain as seen in Figure 5.3A below. However, other "*non-tcdR*" strains had the data in that

region still intact (as seen in Figure 5.3B and Figure 5.3C), in fact the region had more coverage than other parts of the genome for most of the sequenced strains (Figure 5.3B).

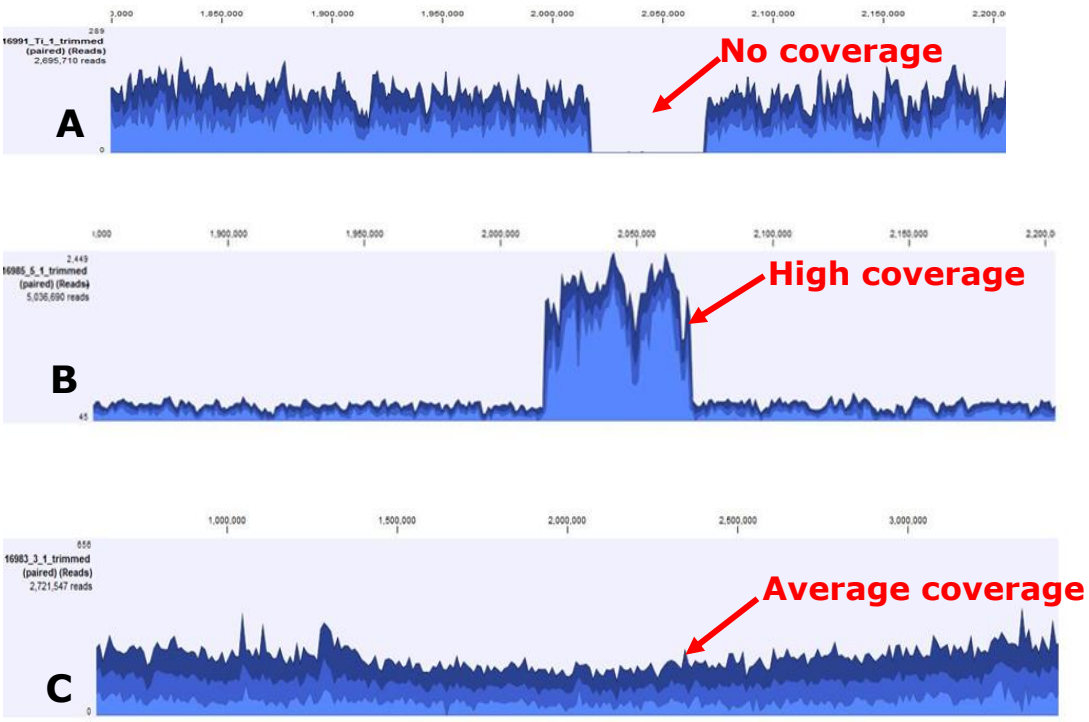

Figure 5.3: Sequencing coverage of phage-related region in different *C. acetobutylicum* mutants. This was done with CLC Genomics Workbench bioinformatics software tool by Qaigen. Figure 5.3A shows that *tcdR*Δ*Cac1502* did not have sequences in the region between positions 2,016,384 to 2,069,759. Figure 5.3B shows the same region in the *pyrE* mutant strain shows considerably higher than average sequence coverage between positions 2,016,378 to 2,067,764. Figure 5.3C shows coverage for the Δ*pyrE*Δ*CAC1502* strain 3 which shows similar coverage for region 2,016,378 to 2,067,764 compared to surrounding region. The aggregated mapped reads on the Read tracks are indicated by different shades of blue. From top to bottom, the dark blue represents maximum coverage value, the royal blue (middle shade) represents average coverage value while the light blue represents the minimum coverage value.

While going through the genes missing in the *tcdR* insert mutant, it became apparent that the region had phage-related annotations. It became obvious that the loss of 53,380 bp was as a result of phage induction. This phage was named Phage ACK1. This prompted the following questions. Was the insertion of *tcdR* in *C. acetobutylicum* responsible for loss of the lysogenic phage in the respective mutant? Why is there a high sequencing coverage of phage genes in some mutants but low coverage in other strains? What type of phage is present in this host?

5.3.2 Is the Insertion of the *tcdR* Gene Responsible for Inducing Phage ACK1?

Tao *et al*. (2010) reported that a host source sigma factor, *sigH*, regulates phage integration and excision in *Staphylococcus aureus*. The sigma factor recognises the promoter upstream of the phage integrase, thus directs transcription which leads to lysogeny stabilisation. Discovering that genes in a host could regulate the phage-encoding genes, raised questions as to whether the introduction of the foreign sigma factor, TcdR, from *C. difficile* could have facilitated the excision of the prophage in *C. acetobutylicum*. To investigate this, 25 colonies of *C. acetobutylicum* with *tcdR* inserted at the *pyrE* locus, from 5 independent transformations (achieved by transforming *pyrE* five times), were grown overnight in 2xYTG and gDNA extracted and screened. Genomic DNA from the *tcdR* mutants were screened for lysogenic phage using primers pair Cac1866F and Cac1871R2, in addition, another primer pair was used to confirm the presence of the phage: Cac1949F and Cac1951R2. Figure 5.4 shows the location of the primers for prophage screening.

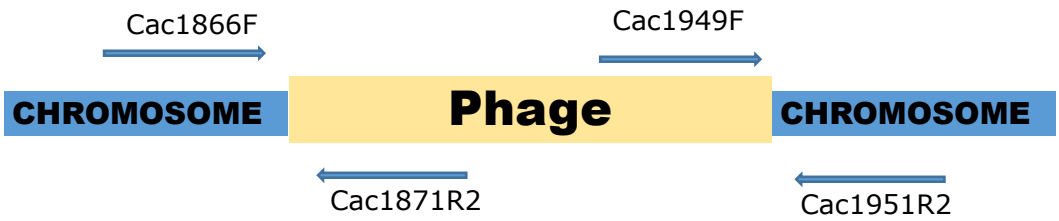

Figure 5.4: Position and location of primers used for phage screening. Forward primers Cac1866F and Cac1951R2 were designed to anneal to the host chromosome while Cac1871R2 and Cac1949F anneal on the phage. The expected band size for Cac1866F/Ca1871R2 is 2827 bp and Cac1949F/Cac1951R2 is 2922 bp.

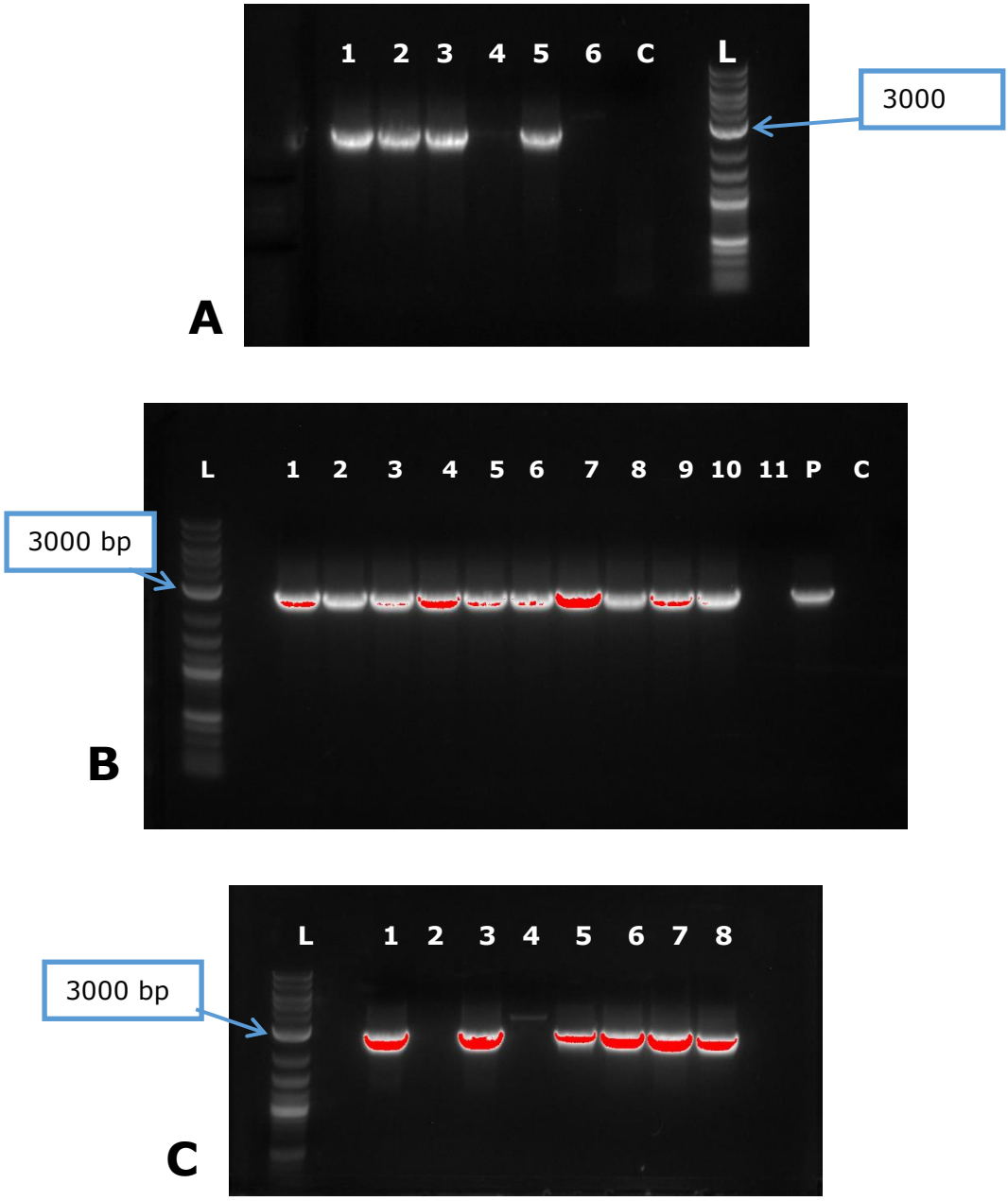

Figure 5.5: Twenty-five *tcdRΔCAC1502* mutants from 5 Independent transformations screened for the presence of lysogenic phage. L stands for 2-log DNA ladder (0.1 – 10 kbp) by New England Biolab. C is the phage-less sequenced *tcdRΔCac1502* strain as positive control while P is Δ*pyrE* strain which is the negative control. Expected DNA fragment is 2,827 bp for the presence of phage but no fragment if there is no phage. In Figure 5.5A, lanes 1, 2, 3 and 5 have the expected DNA fragment 2,827 bp while 4 and 6 have no fragment indicating an absence of lysogenic phage. In Figure 5.5B, lanes 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 have the expected DNA fragments 2,827 bp while lane 11 has no fragment indicating an absence of lysogenic phage. In Figure 5.5C, lanes 1, 3, 5, 6, 7 and 8 have the expected DNA fragment of 2,827 bp while lanes 2 and 4 do not have any DNA fragment.

Out of twenty-five (25) *tcdR* mutant colonies screened in Figures 5.5A, B and C, five (5) of them had no lysogenic phage. The remaining 20 *tcdR*Δ*Cac1502* mutants were all prophage positive. On the other hand, restriction-less mutants of *C. acetobutylicum* from different transformation batches were screened for the presence of phage genome, they all carried the prophage (data not shown). Investigations on the presence of phage was also carried out on a Δ*phe*Δ*arg*Δ*pyrE* strain (triple auxotrophic mutants) with same mutations but in different independently obtained variants created by Dr. Bunmi Omorotionmwan from SBRC Nottingham, University of Nottingham UK. They all tested positive for the presence of phage in the host genome (Figure 5.6).

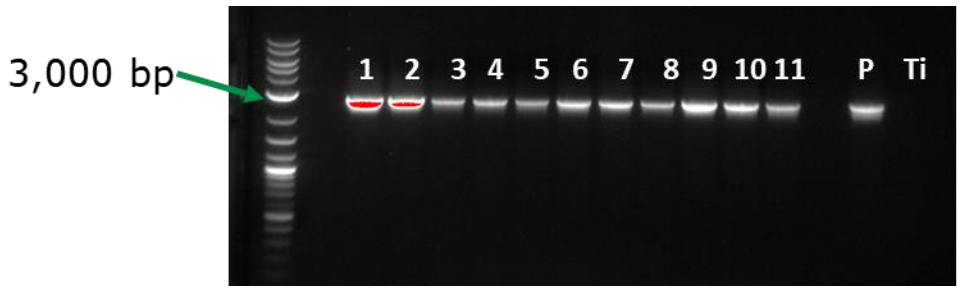

Figure 5.6: Screening of *C. acetobutylicum* Δ*phe*Δ*arg*Δ*pyrE* strains revealed the presence of the lysogenic phage ACK1 in all independent isolates. L represents 2-log DNA ladder (0.1 – 10 kbp) by New England Biolab. The expected DNA fragment sizes are 2827 bp. Lanes 1 to 11 are the triple mutants (Δ*phe*Δ*arg*Δ*pyrE* strain) showing expected DNA fragment sizes for prophage positive, P is the positive control, *C. acetobutylicum pyrE* mutant while Ti is the negative control, confirmed phage-less mutant.

The answer to the question "was the insertion of *tcdR* into the restriction-less mutant responsible for the excision of phage DNA in the *tcdR*Δ*Cac1502* strain?" is likely no, although the data remain inconclusive. Obviously some of the screened *tcdR*Δ*Cac1502* carried prophages in their genome but the fact that excision occurred in several of the *tcdR* mutants while none of the strains without the *tcdR* gene examined had lost the phage is certainly intriguing. The existence of pseudolysogenic relationship where temperate phages do not commit to lytic or lysogenic lifestyle and segregation can lead to daughter cells that are lysogenics and non-lysogenics (Cenens *et al*., 2013) raises the possibility that such an event could have occurred during ACE *tcdR* insertion. If this is the case, then the probability of isolating phage sensitive colonies during *tcdR* insertion and plasmid loss screening increases.

5.3.3 Gene Map of Phage ACK1 of *Clostridium acetobutylicum* ATCC 824

Phage ACK1 in the genome of *C. acetobutylicum* ATCC 824 comprises 53,380 base pairs which has 100% gene identity with 3 other *C. acetobutylicum* strains; EA2018, DSM1731 and LJ4 as identified by BLAST. Considering that for a long time, *C. acetobutylicum* ATCC 824 has been one of the most studied solventogenic *Clostridia* (Croux and García, 1991), it was therefore an ideal strain for this phage study. In 2001, Nölling and his co-workers who sequenced the whole genome of *C. acetobutylicum* ATCC 824 reported that it seemed to have two cryptic prophages. The first prophage they described was located between *CA_C 1113* and *CA_C1197* with approximately 90 kbp and 85 genes while the second one was postulated to be located between *CA_C1878* and *CA_C1957* with approximately 60 kb and approximately 79 genes. From this project, it is clear that they described Phage ACK1 as the second prophage.

The ACK1 phage genome occupies the region between 3' end of *CA_C1866* gene encoding putative 4-cysteine ferrodoxin and 3' end of *CA_C1951* gene encoding the site-specific recombinase, DNA invertase pin homolog based on my analysis but not *CA_C1878* to *CA_C1957* as proposed by Nölling *et al.* (2001). The *CA_C 1951* gene encodes the phage encoding integrase. The *CA_C1866* encodes 4-Cysteine ferrodoxin protein in the host, bearing iron binding domain and electron binding activity, however, its biological process or function in *C. acetobutylicum* is unknown. Since *CA_C1951* seems to be the functional integrase of the phage, the excision of phage resulting in a sequence exchange at the 3' end of *CA_C1866* and *CA_C1951* might have annulled any negative effect on phage and bacterial genes.

The phage genome consists of 84 open reading frames with a GC content of 31.64%. Only 22 of the 84 predicted proteins have similarity with proteins with known functions on nBLAST. Therefore to determine the roles of other proteins, nucleotide and amino acid sequence analysis were employed. With the use of bioinformatics analyses, 36 ORFs were assigned putative functions (see Figure 5.7). All the genes in Phage ACK1 are listed in the Appendix B (Table 7.2).

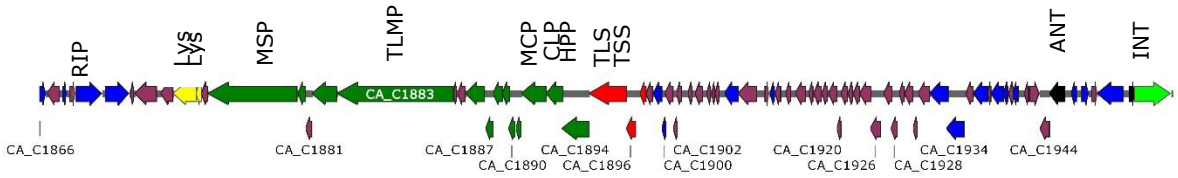

Figure 5.7: Gene map of *Clostridium* phage ACK1 drawn to scale. The putative functions of selected open reading frames were given above the respective gene. Colour coded phage module as identified by bioinformatics analysis: blue are DNA replication and transcription genes; yellow are lysis genes; green are structural genes (head and tail proteins); red are packaging genes; black are lysogeny control genes; lemon green is integrase; purple are hypothetical genes. Some selected genes were annotated: RIP, (replication initiation protein); Lys (lysin); MSP (minor structural protein); TLMP (tape length measure protein); MCP (major capsid protein); CLP (clpP serine protease); HPP (head portal protein); TLA (terminal large subunit); TSS (terminal small subunit); ANT (anti-repressor); INT (integrase).

Bioinformatics analysis enabled the determination of the phage's functional gene arrangement for integrase, DNA recombination, modification and replication, DNA packaging proteins, head and tail proteins, lysis proteins and lysogeny associated functions. It was difficult to establish synteny with other bacteriophage genomes because it showed low similarity with others. There seemed to be similarity with the organisation of clusters of genes with gene architectures of the family *Siphoviridae*. PHASTER genome assessment showed the most common hit with 8 genes is *Paenibacillus larvae* bacteriophage Harrison from the family *Siphoviridae*, in which 3 genes were part of the structural genes region, 3 genes were the packaging genes and 2 genes were in the region encoding hypothetical protein region. Based on this mosaic arrangement of genes, the map was divided into 5 parts; DNA replication and transcription, phage lysis genes, structural genes, packaging genes and integrase gene.

5.3.3.1 Phage DNA Replication and Transcription Regulators

ACK1 has its own replication system. In fact it seems that there are more genes linked to DNA replication and transcription than there are for the other four structures. The mosaic arrangement suggests that genes *CA_C1867* to *CA_C1875* as well as *CA_C1898* to *CA_C1950* encode DNA replication, transcription and lysogeny proteins. Although there are some hypothetical proteins in these regions that do not have any similarity with PHASTER and NCBI nucleotides sequences and proteins, there are also genes having similarity with transcription regulators characterised by XRE, however, their

precise function is unknown. For example, *CA_C 1867* and *CA_C 1869* are annotated as phage-related XRE family transcription regulators, which are similar to a family of DNA-binding transcription regulators of XRE helix-turn-helix domain in multiple *Clostridium* species including *C. sporogenes* and *C. botulinum*. However, *CA_C 1868* has a DUF4532 conserved domain with unknown function. *CA_C 1871* encodes DNA translocase FtsK. According to Crozat and Griange (2010), FtsK converts the chemical energy from hydrolysed ATP to transport DNA. FtsK are SOS-regulated translocases which serve as a DNA pump during phage genome packaging into the prohead, DNA recombination and repair and phage induction (Wolfe *et al*., 2014). The ACK1 phage encodes its own replication initiation protein (*CA_C 1872)* then uses the host replication machinery for viral replication.

The region *CA_C1898* to *CA_C1950* encodes some DNA replication and transcription regulator proteins, lysogeny regulators as well as many hypothetical proteins. *CA_C1950* has a conserved domain similar to Cro/C1type helix-turn-helix domain HTH_26 from pfam13443 family. Cro and C1 type proteins are transcription repressors that regulate the switch to lysogenic and lytic cycle by binding to operators with differential affinity in lambda phage (Takeda *et al*., 1977). *CA_C1945* encodes another putative lysogeny regulator, an anti-repressor protein. Other putative DNA replication transcription and metabolism genes in this large region include genes encoding thymidine kinase, DNA polymerase, single-stranded binding protein, helicase loader, transcription regulator AbrB, replisome organiser, translation initiaton factor and mazG-like protein.

To predict the presence of tRNA in the phage genome, tRNAscan-SE was used to analyse the genome and one pseudo-tRNA was identified between genes *CA_C1904* and *CA_C1905* with the tRNA and Anticodon undetermined (Lowe and Chan, 2016). This suggests that phage ACK1 uses the host tRNA for amino acid synthesis.

5.3.3.2 Phage Lysis Genes

The ACK1 phage *lyc* gene encodes lysozyme, *N*-acetyl-β-d-muramidase, the enzyme is responsible for degrading the cell wall specifically the pepitoglycan and allows for the injection of the viral DNA. It has a peptidoglycan-binding glycosyl domain and also catalyses the degradation of the cell wall in the lytic cycle for the new progeny to escape their host (Harada *et al*., 2018). One of the striking features of this phage is the gene

CA_C1877, located upstream of *lyc*. BLASTn hits of *CA_C1877* indicate that it shows some similarities with a region of an autolytic lysozyme gene 1,4-beta-N-acetylmuramidase, *lyc*, located at position 641317 to 642291 in the host chromosome. The region of similarity covers 26.5% of the entire *lyc* gene with 98% query coverage and 87.4% gene identity. This gene of this lysozyme in the host has been sequenced and characterised by Croux and García, (1991). This is the only gene in the ACK1 phage's genome that shares similar sequence with the host. This may be evidence of recombination and gene transfer which occurred between the phage and host, thus it may have played a role in bacteria genome evolution.

5.3.3.3 Phage Structural Genes

CA_C1879 is the second largest gene in the phage genome. It occupies almost 8% of the genome. The conserved domain on BLASTx identifies regions with similarity to N-terminal of phage minor structural protein, phage endopeptidase tail and pentapeptide repeats. The largest ORF in the genome is *CA_C1883* which is 5460 bp in length and remarkably occupies more than 10 percent of the entire phage genome. It has been assigned a putative function of phage tail tape measure protein and has similarities with *Clostridium perfringens* phage phiSM101 tail tape measure protein by PHASTER. This agrees with Hatfull *et al.* (2008) that the largest gene in a phage genome is often the tail tape measure gene. However, there have been a few reports of the minor tail protein being larger than the tape tail measure protein for some phages (Swanson *et al*., 2012).

Phage putative structural proteins genes appear to span from *CA_C1879* to *CA_C1892* with putative functions including minor structural protein, tail tape measure protein, major tail protein, prohead protease, head and tail joining protein, head and tail adaptor protein, DNA packing protein and major capsid protein. There are few hypothetical proteins in this region which are encoded by *CA_C1880, CA_C1881, CA_C1882, CA_C1884* and *CA_C1885.*

5.3.3.4 Phage Packaging Genes

The *CA_C1893* gene encodes a putative ClpP family serine protease which is recognised in *E. coli* phages to play the essential role of maturational cleavage of phage capsid precursors. This is important for the stability of the capsid for phage DNA packaging initiation (Liu and Mushegian, 2004). This gene is followed by the phage head portal protein gene, *CA_C1894* which is known to play essential roles in different

stages of phage DNA packaging. The terminal large and small subunits are encoded *CA_C1895* and *CA_C1896* respectively. The small terminases are responsible for initiating DNA packaging, they recognise and bind to specific *pac* or *cos* sites where packaging starts (Rao and Fiess, 2008). In T4 phage, the large terminase provides nuclease activity for cleaving the phage genome to activate phage packaging into pro-capsids and cleaving phage DNA to terminate packaging. It also has ATP hydrolysis activity for DNA translocation to empty prohead capsids during packaging (Sun *et al*., 2012).

5.3.3.5 Phage Integrase

The identification and study of the integrases and their attachment sites in bacteria and phages are important in the use of site-specific recombination tools for genetic engineering and system biology. Integrases are vital enzymes for the recognition, excision and integration of the phage into its host (Fogg *et al*., 2014). Running BLASTx analysis of ACK1 phage integrase (*CA_C1951)* showed that although it has above 90% coverage and 60% identity with several uncharacterised bacterial recombinases in *C. botulinum*, *C. sporogenes*, *Clostridium acidisoli*, *Clostridium magnum*, *Clostridium kluyveri*, *Clostridium cellulovorans* and *Clostridium tyrobutylinum*, it has some similarities with *Lactococcus* phage TP901-1 integrase from *Lactococcus lactis* subsp *cremonis* with 39% identity and an E value 3e-17. PHASTER identifies *CA_C1951* as similar to *Enterococcus* phage EFC-1 site-specific recombinase from *Enterococcus faecalis* KBL101 with an E value 1.22e-36 and *Lactococcus* phage 28201 recombinase from *Lactococcus lactis* with an E value 6e-38. These hits by BLASTx and PHASTER were all serine recombinases (Yoon and Chang, 2015; Stoll *et al*., 2002). The conserved domain of their integrase is identified as a resolvase of the serine recombinase superfamily (Figure 5.8). The division of integrases into serine and tyrosine recombinase is based on the type of conserved amino acid residue which reacts with the DNA phosphodiester and forms a covalent bond with the DNA strand as recombination takes place (Stark, 2014). Serine recombinases which are autonomous, not requiring cofactors, and have unidirectional activity as a result mediating recombination at non-identical *attB* and *attP* sites are ideal for designing integration tools (Stoll *et al*., 2002).

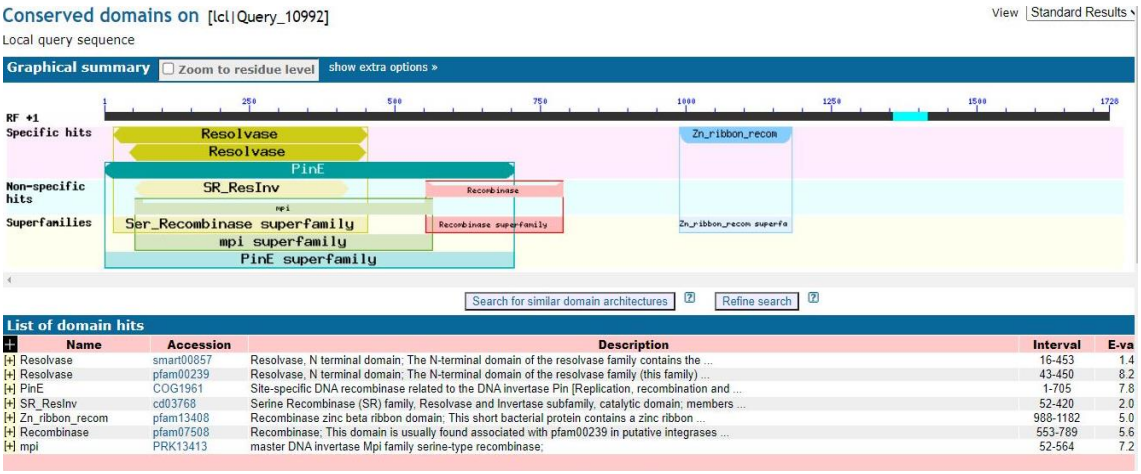

Figure 5.8: Conserved protein domains identified in *CA_C1951* recombinase by NCBI BLASTx search. This BLASTx results of reference proteins identifies *CA_C 1951* recombinase as a resolvase of the superfamily serine recombinase. Pin E domain suggests that it may be closely related to the DNA invertase pin responsible for DNA replication, recombination and repair. Image taken from NCBI BLASTx results page (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?RID=4SPP180B016&mode=all) following query with *CA_C1951* using standard settings.

5.3.4 Site-specific Recombination Region in *C. acetobutylicum* ATCC 824 Phage ACK1

Like *Lactococcus* phage TP901-1 integrase (1st gene), *Enterococcus* phage EFC-1 site-specific recombinase (last gene) and *Lactococcus* phage 28201 serine recombinase (1st gene), the *Clostridium* phage ACK1 integrase is located at the extreme end of the phage genome. It is well known that the integration functional sequences like *attP* and integrases are tightly clustered (Christiansen *et al*., 1994). ACK1 Integrase is allocated the last gene of the phage genome. Amongst all the phage serine recombinases, *Lactococcus* phage TP901-1 integrase has been well studied (Stoll *et al*., 2002; Christiansen *et al*., 1994; Breüner *et al*., 2001). *Lactococcus* phage Tp901-1 serine recombinase has been identified as the only phage protein required for recombination to take place and classified as a resolvase type recombinase (Breüner *et al*., 2001). Instead of 6 - 8 bp overhang cuts like tyrosine recombinases in the double-stranded *attP* and *attB* sites, Tp901-1 recombinase cuts a staggered 2 bp single strand overhangs (Breüner *et al*., 2001). Tp901-1 has been used widely as a Recombination Mediate Cassette Exchange technique to facilitate genetic engineering in *Drosophila melanogaster* (Voutev and Mann, 2019) and mouse artificial chromosome (Tomimastu *et al*., 2017) and human cells (Stoll, 2002). It will be very interesting to see if the phage

recombinase in *C. acetobutylicum* phage ACK1 can be used for recombinant engineering in its host. To achieve this, *attB*, *attP*, *attL* and *attR* sites should be identified.

The CLC workbench sequencing assembly revealed that the phage ACK1 genome integrated in *C. acetobutylicum* ATCC 824 contains 53,386 base pairs of double-stranded DNA with 6 bp terminal repeats represented by the sequence 5' AGATGC 3' (Figure 5.10). Furthermore, when the gap in the phage-less *tcdR* strain is joined together, it forms the sequence 5' AGATGCAGATGC 3', this may be the recognition specific site for insertion of the phage (Figure 5.9). Lambda phage has 48,490 pairs of linear double-stranded DNA with 12-base single sequences forming cohesive ends that bind with each other to circularize the genome, resulting in a genome of 48,502 bp in its circularised form (Casjens and Hendrix, 2015). It is likely that the recombinase cuts the phage and bacteria in a staggered manner to create 2 bp overhangs like in TP901-1 recombinase of similar homology and other known serine recombinases (Stoll, 2014).

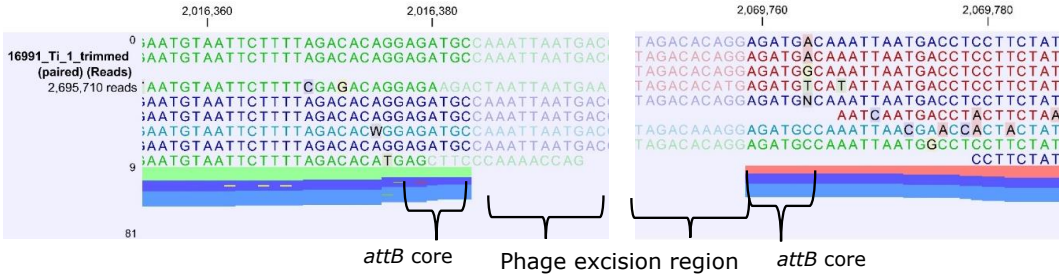

Figure 5.9: Alignment of trimmed mutant reads covering the phage integration region in a *TcdRΔCAC1502* phage-free strain. This image was obtained from CLC Genomic Workbench bioinformatics analysis software. The faded region indicates no phage genome alignment. This figure shows no read data coverage in the phage integration region between positions 2,016,384 to 2,069,759. Observe that position 2,016,384 has the 6 bp sequence 5' AGATGC 3' in front of it while position 2,069,759 has the the 6 bp sequence 5' AGATGC 3' following it. This indicates that the bacterial attachment (*attB*) has the common 6 bp core sequence. The horizontal layered colours represent aggregated coverage value on the Read tracks.

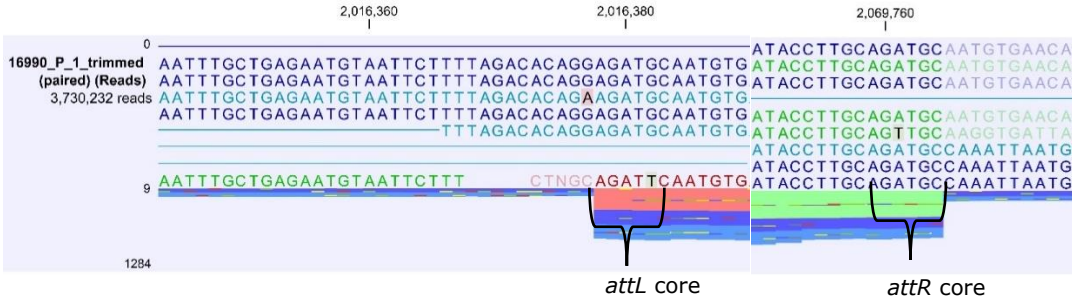

Figure 5.10: Alignment of trimmed *pyrE* mutant reads covering the phage integration region in a lysogen strain. This image was obtained from CLC Genomic Workbench bioinformatics analysis software. The horizontal layered colours represent aggregated coverage value on the Read tracks. It shows higher read coverage in the phage integration region between positions 2,016,378 to 2,069,764 than other surrounding regions. Observe that the position 2,016,378 has the 6 bp sequence 5' AGATGC 3' following it and this is the *attL* core sequence. Observe that the position 2,069,764 has the the 6 bp sequence 5' AGATGC 3' in front which is the *attR* core sequence.

To confirm that the exact core sequence in the *attB* and *attP* site is 5' AGATGC 3', the gDNA of *tcdR*Δ*CAC1502* phage-less mutant was screened by PCR using primers Cac1866F and Cac1951R2 (Figure 5.11).

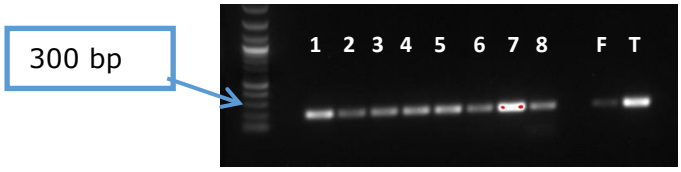

Figure 5.11: Screening the attachment sites in phage-free and phage-containing *C. acetobutylicum* strains with Cac1866F and Cac1951R2. Lanes 1 to 8 are *tcdR* phage-less strains, F is control Δ*pyrE*Δ*CAC1502* mutant 4 which showed high sequence coverage of the phage-related region while T is the control *tcdR*Δ*CAC1502* phage-less strain. Expected DNA fragment size is 252 bp. Presence of a faint DNA fragment size for Δ*pyrE*Δ*CAC1502* mutant 4 indicates that there are cells in the colony that do not have the phage integrated in its host genome.

The DNA fragments were excised and extracted from the gel, and then sent for Sanger sequencing. Sequencing results of the DNA fragments in Figure 5.11 showed that the common core at *attB*, *attL* and *attR* is 5' AGATGC 3'. This suggests that it may have a similar system to that of TP901-1 which has a common core 5-bp region, 5' TCAAT 3' where integration takes place followed by a 1 bp mismatch and identical region 5' (T/C)AAGGTAA 3' (Christainsen *et al.*, 1994). The Phage ACK1 recognition *attP* is

located within the integrase gene, as in *Myxococcus xanthus* phage Mx8 recombinase (Műller *et al*., 2006).

Figure 5.12 explains the putative site-specific recombination model for *Clostridium* phage ACK1 recombinase. The *attB* is 5' ACACAGG-AGATGC-CCAAATT 3' while the *attP* is 5' ...ACCTGC-AGATGC-AATGTGA... 3'. Recombinase is likely to create a 2 bp overhang at both *attP* and *attB* common core, the serine recombinase dimers binds nicked *attP* and another recombinase dimer binds the *attB* crossover site and synaptic complex is formed by the recombinase tetramer. The recombinase tetramer complex catalyses the exchange of cleaved site and ligates the staggered ends to generate a recombinant with *attL* 5' ...ACACAGG-AGATGC-AATGTGA... 3' and *attL* 5' ...ACCTGC-AGATGC-CCAATT... 3'. The hybrid attachment sites contain half of the sequences in *attB* and half of *attP* as reported by Fogg *et al*., (2014). ACK1 Integrase is unlikely to be a DNA invertase type because it lacked the imperfect palindromic symmetry feature commonly observed in DNA invertases (Johnson, 2015).

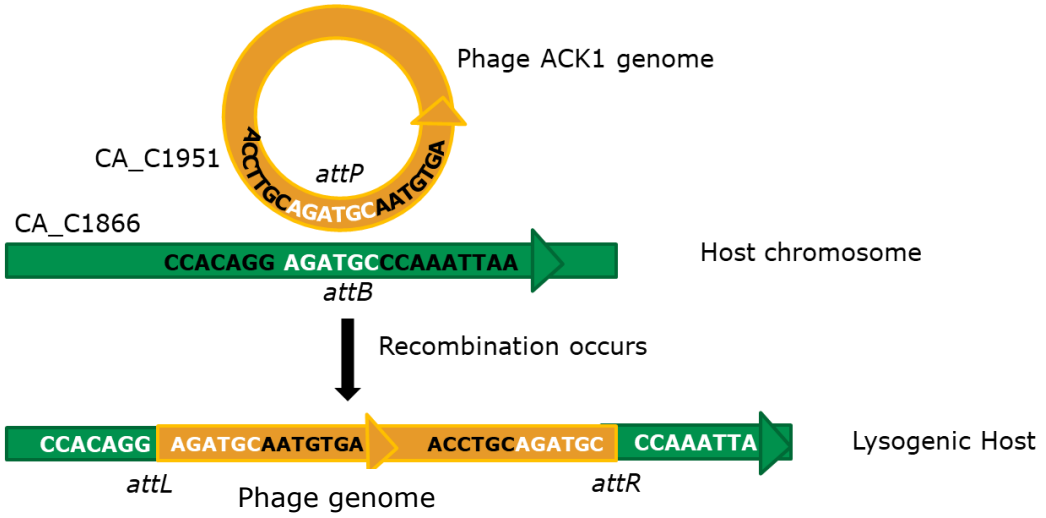

Figure 5.12: Putative site-specific recombination model for *Clostridium* phage ACK1 recombinase. The *attB* site is 5'…CCACAGG-AGATGC-CCAAATT…3'while the *attP* site is 5'…ACCTGC-AGATGC-AATGTGA...3', *attL* site is 5'…CCACAGG-AGATGC-AATGTGA…3' and *attR* site is 5'…ACCTTGC-AGATGC-CCAAATTA…3'. The complete sequences of the *attB, attP, attR and attL* sites was not determined. The dot dot dot indicates omitted sequences which are unknown. The minimum sequences making up each recombination site were not determined.

5.3.5 Phage-host Relationship of *Clostridium* Phage ACK1 in *C. acetobutylicum* ATCC 824

Three model phages of *E. coli* that display different lifestyles (phage-host interactions) in the bacteria are T4, lambda and P1. The study of these bacteriophages has contributed towards advances in molecular biology tools. While T4 is a strict virulent phage, lambda is a temperate phage that integrates into the host genome and P1 is a temperate phage that forms low copy plasmid in the host cell (Lobocka *et al*., 2004). The gel images in Figure 5.13A and Figure 5.13B show that the phage ACK1 circularises in the host cytoplasm like the P1 phage and integrates into the host genome like lambda, respectively. Comparing the read coverage depth of the phage genome and bacterial genome observed in the *C. acetobutylicum* Δ*pyrE* mutant whole genome sequencing result indicates it exists as a self-replicating plasmid, creating many copies about 18-fold in quantity of the host genome in the cytoplasm. The Δ*pyrE*Δ*CAC1502* strain 3 lysogenic mutant which displayed an average sequence coverage in the phage region compared to surrounding region (Figure 5.3C), exhibited a very faint band when screened for circularisation with primers Cac1871R2 and Cac1949F (Figure 5.13A).

To determine when the phage transits from integrated host to circularised self-replicating plasmid form, the Δ*pyrE*Δ*CAC1502* strain 3 lysogenic mutant was subcultured in 2xYTG broth and screened with Cac1871R2 and Cac1949F every 2 hours for 12 hours. For every 2 hours of PCR screening, there were clear bands indicative of the circular forms of the phage were persistent and replication of phage plasmid occurred. On the other hand, screening of the lysogenic bacteria with Cac1866F and Cac1951R2 gave faint bands which could mean that in a colony of *C. acetobuylicum*, there are cells that do not have phage genome integrated in them but may still exist in plasmid form. Just to confirm that the circular form of phage seen in the Illumina sequencing of gDNA extract is not a result of free virons released from the cell during culture, plaque assay was carried out on the filtered lysogenic *C. acetobutylicum* culture with *tcdR* phage-less indicator strain using a previously developed and improved protocol as described in Materials and Methods section 2.1.9. There were no plaques formed on a 0.5% soft agar overlay.

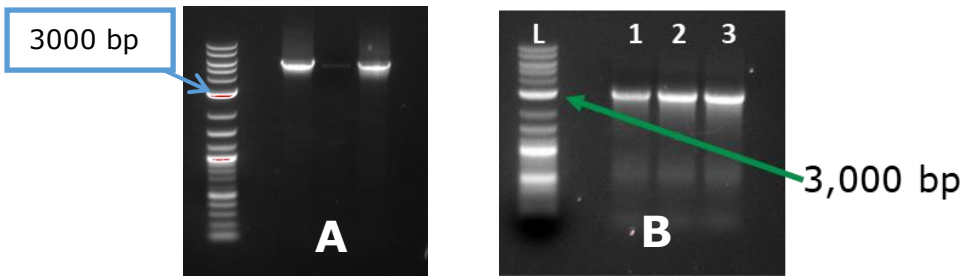

Figure 5.13: PCR screening for circularised phage ACK1 genomes in different *C. acetobutylicum* host. In both images, L represents 2-log DNA ladder (0.1 – 10 kbp) by New England Biolab, lane 1 is the *C. acetobutylicum pyrE* mutant, lane 2 is Δ*pyrE*Δ*CAC1502* strain 3 and lane 3 is *ΔpyrEΔCAC1502* strain 4 (both restriction-less mutants created in chapter 3). Figure 5.13A, test for circularised phage ACK1, the expected band size is 5,534 bp. The faint band in lane 2 shows that despite the CLC biosequencing data analysis showing lysogeny with comparable average read coverage depth with host genome (refer to Figure 5.3C), a few of the phage genomes exist as circular plasmid. Figure 5.13B, Test for integrated phage into the host genome, the expected band size is 2922 bp.

Based on this, there are four proposed life styles which phage ACK1 genome may adopt in its host; A) Integration of the phage into the host genome (Figure 5.14A); B) Integration as well as plasmid circular forms of phage (Figure 5.14B), C) Plasmid form of phage in the host cytoplasm without integration (Figure 5.14C), D) Packaging free virions as a result of lytic cycle activation (see section 5.16). However, it is not known which path the phage will follow when it enters *C. acetobutylicum* ATCC 824.

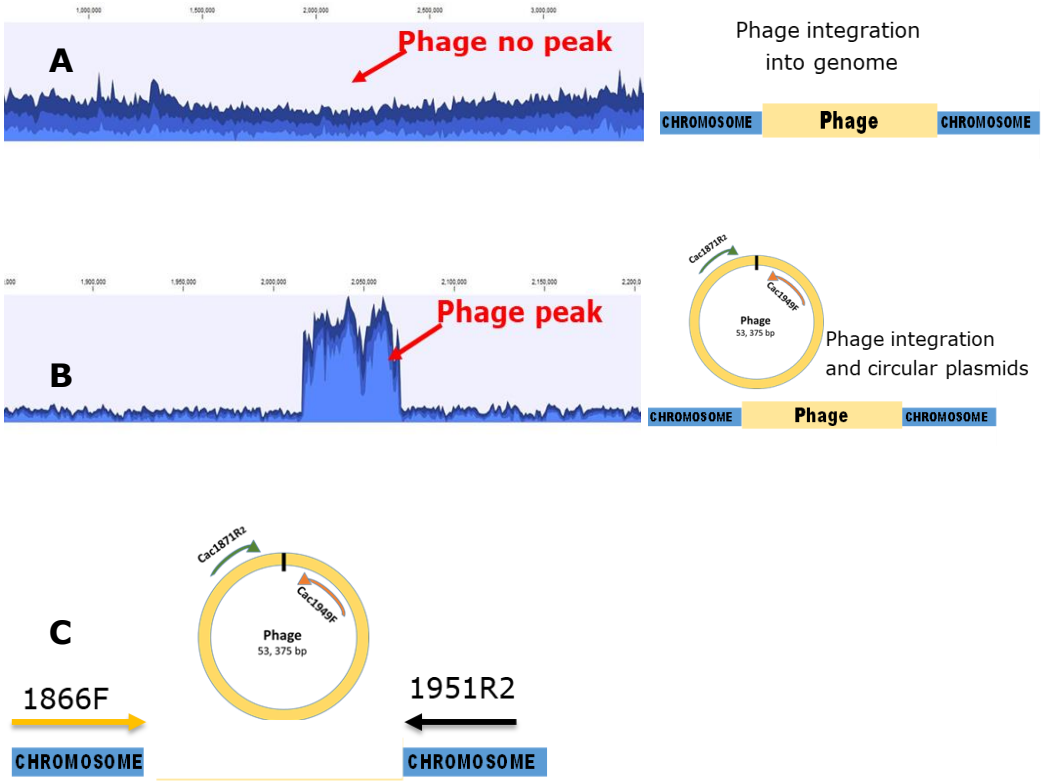

Figure 5.14: The different lifestyles of phage ACK1 in *C. acetobutylicum* ATCC 824. Figure 5.14A, phage genome integration into host genome. Figure 5.14B, circularised plasmid form with integration. Figure 5.14C, circularised plasmid form without integration.

The evidence of a possible pseudotemperate relationship puts the isolation of *tcdR* mutants that had lost the phage into perspective. In 2013, Cenens and his co-workers had proposed a pseudolysogenic condition which could lead to segregation of phage P22 host daughter cells into lysogens and non-lysogens. In other words, phage sensitive cells can be isolated from a population of pseudolysogens (Cenens *et al*., 2013). Though the ACK1 phage was observed to multiply up to 18-fold copy numbers compared to the host chromosome in some lysogens, it was also shown to integrate in the host genome with little plasmid formed and some cells in the population lacking the integrated phage. Therefore, segregation of the one with little plasmid formed or phage-free lysogen as a result of selecting one mutant during ACE technique of *tcdR* insertion is a likely possibility. Cenens and his co-workers (2013) suggested that pseudolysogenic state could be a mechanism to prevent an abortive integration or replication event due to starvation.

5.3.6 Improving on the Protocol for Phage Induction and Plaque Assay

Since there are only a few publications of phage assays on *C. acetobutylicum*, it was challenging to replicate a plaque assay for phage propagation in this organism. The publication of the isolation of a virus-like particle from *C. acetobutylicum* NCIB 6444 strain by Kim and Blaschek (1991) stated that they were not successful in demonstrating phage CAK1 induction and infectivity. Another publication of *Clostridium* phage CA1 of *C. saccharobutylicum* P262 (formerly named *C. acetobutylicum* P262) isolated from a fermenter bacteriophage outbreak on strain P262 was directed at transforming the phage DNA into *C. saccharobutylicum* P262 protoplast (Reid *et al*., 1983). Although Reid and his co-workers reported the release of matured viral particles by the transformed protoplasts, it lacked details on the plaque assay technique used in the isolation.

Using the phage induction and overlay plaque assay protocol developed for *C. difficile* (SBRC Nottingham, University of Nottingham, UK), induction of phage with 3 mg/ml mitomycin C and UV light was done. To detect if it was successful, a plaque assay using 0.5% CGM overlay with 1 ml bacterial culture and 200 µl putative phage solution also diluted to 10^{-8} was carried out. The phage solution was cell-free supernatant after mitomycin C induction. There was no plaque observed after 24 hours incubation. The problem could have been that;

1. The concentration of mitomycin C was not sufficient or too high to activate an SOS response in the cell in response to DNA damage.
2. The *C. acetobutylicum* strain was not sensitive to mitomycin C, thus its DNA was not damaged by it.
3. It could be that products of solventogenesis (especially acetone) were interfering with the activity of mitomycin C.
4. The strain is resistant to infection by phage.

The chemical and physical conditions of growth media can affect the plaque formation (Lillehaug, 1997). The plaque assay was improved by carrying out the following adjustments. In the first experiments the possible benefits of divalent metals was explored through the addition of 10 mM $CaCl_2$ and 5 mM of $Mg_2SO_4 . 7H_2O$ into both the soft overlay agar and the bottom agar. Due to the solvent-producing nature of *C. acetobutylicum*, it was observed that the low pH as the result of acids produced may

have affected phage attachment. To tackle this, the growth medium (both top and bottom) was buffered with 10 mM of Tris-HCl (pH 7.6). Secondly, the phage induction was not carried out overnight, rather it was performed by sub-culturing the lysogen in mitomycin for 4 hours while monitoring the pH to ensure that it did not drop below 5.5. Thirdly, an additional step was introduced that involved the centrifugation of the sub-cultured indicator cells, before resuspending them in broth supplemented with 10 mM $CaCl_2$ and 5 mM of $Mg_2SO_4 . 7H_2O$ at pH 7 and then mixing them with the phage lysate. This was undertaken to rid the mixture of solvents produced by the host and adjust the pH close to neutral. Foddai *et al*. (2009) showed that plaque formation was improved by washing *Mycobacterium avium* subsp. *paratuberculosis* prior to infection with D29 mycobacteriophage because it gets rid of the lipid substance built up and released during bacterial growth that might mask receptors for phage attachment.

When these adjustments were made, plaques were observed as seen in Figure 5.15B. See chapter 2 section 2.1.9 for the improved plaque assay methodology. This shows that phage ACK1 is not a cryptic phage as Nölling and his co-workers (2001) suggested because the phage lyses its hosts and forms active phage particles as evidence by the plaques on soft agar. A plaque was approximately 1 mm in diameter. To increase the phage titre, plaque overlay plate dilutions with confluential plaques were scraped and put in 10 ml of CGM pH 6-7. The phage suspension was incubated for one hour in an anaerobic cabinet. The mixture was then centrifuged at 5000 x g for 10 minutes at room temperature, the supernatant was then filtered with a P 0.22 µm filter and the filtrate containing phage lysate used to repeat plaque assay. These steps were repeated until the phage titre was 10^9.

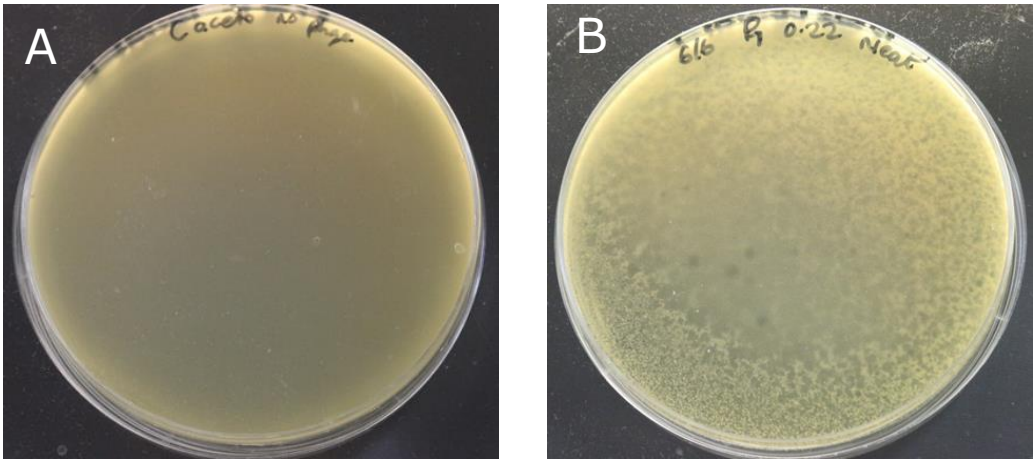

Figure 5.15: Phage ACK1 plaques. Figure 5.15A, no plaques observed when *C. acetobutylicum* (*tcdR*Δ*CAC1502*Δ*CAC3535* strain) grown on a plate with no phage lysate. Figure 5.15B, plaques are visible on plates when the same phage indicator strain was grown with phage lysate.

5.3.7 Characterisation of Phage ACK1

To characterise the phage, the method published by Fortier and Moineau (2007) was used to precipitate the phage lysate (containing > 10^{-9} pfu/ml) with 0.1 mM ammonium acetate and stored at 4°C. The precipitated phage was negatively stained with 2% uranyl acetate and viewed under a transmission electron microscope (TEM). TEM images revealed a phage with an elongated icosahedral head with a length of 123±13 nm, a width of 58±8 nm and a tail length of 312 ±21 nm (Figure 5.16A). A long-tailed phage was expected because the tail tape measure gene is 10% of the entire phage genome and the largest phage gene. This agrees with Katsura and Hendrix report in 1984 that the tail tape measure gene length corresponds to the actual length of the phage tail. There were no visible tail fibres in the images, indicating they may be too small to be captured by TEM.

Phage CAK1 isolated from *C. beijerinckii* NCIMB 6444 (formerly named *C. acetobutylicum* NCIMB 6444) as filamentous structure with single-stranded DNA (Kim and Blaschek, 1991). Another phage, CA1, isolated from *C. saccharobutylicum* P262 (formerly named *C. acetobutylicum*) has an icosahedral head with a diameter spanning 40 to 42 nm and a short tail with visible tail fibres (Reid *et al*., 1983). Phage CA1 is likely from the *Myoviridae* family. The phage from this study, Phage ACK1 as seen in Figure 5.16A, is proposed to be of the family *Siphoviridae* of the order *Caudovirales*. Phages of this family have long flexible non-contractile tails. The

modular arrangement is in line with that of *Siphoviridae*. For lysogenic *Siphoviridae* phages of low GC content Gram-positive bacteria, the functional module is divided into packaging, head structural, tail structural, lysis, integrases/recombination, lytic/lysogenic switch control, excision, and DNA replication genes (Stevens *et al*., 2011).

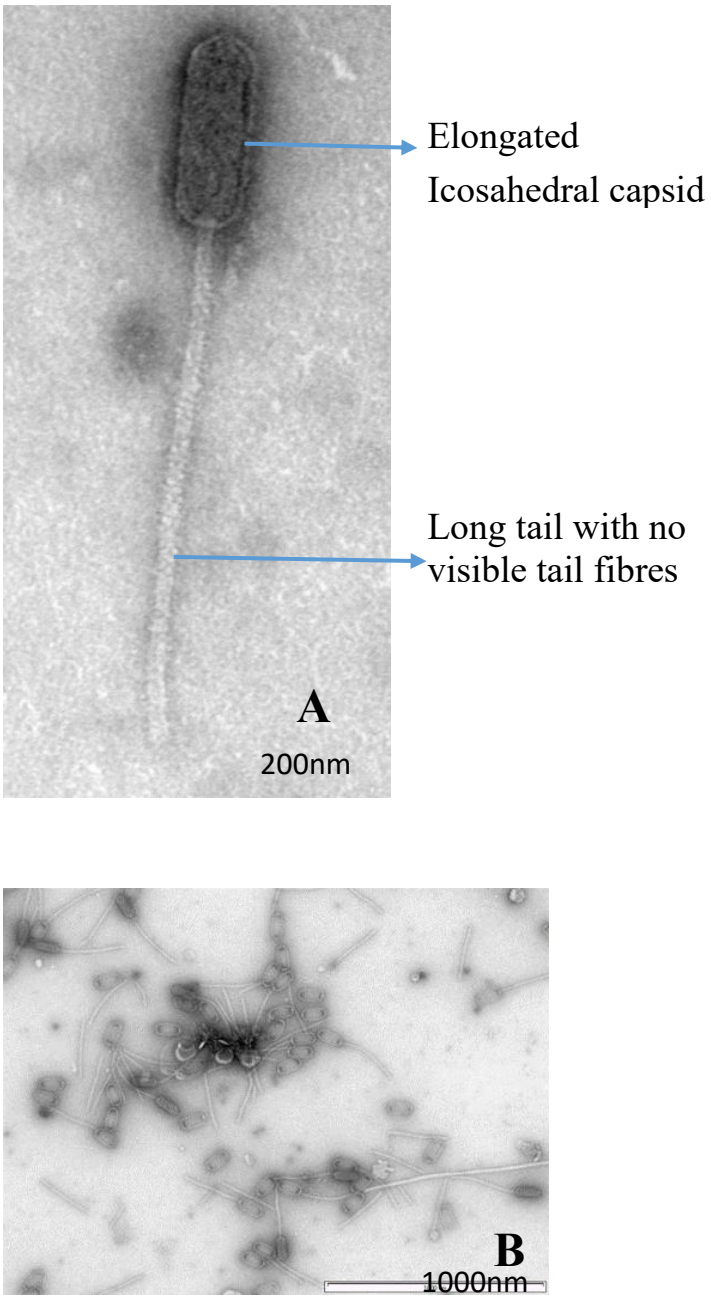

Figure 5.16: Transmission electron microscope images of phage ACK1. Figure 5.16A, *C. acetobutylicum* ATCC 824 ACK1 phage is from the family *Siphoviridae*; it has an elongated capsid, long tail and linear double stranded DNA. Figure 5.16B, image of clusters of ACK1 phages.

5.3.8 Phage ACK1: Transducing or Non-transducing Phage?

Given that the *CA_C1877* phage gene shares 87% identity with 26.5% of *lyc*, the *C. acetobutylicum* host's autolytic enzyme gene, it can be hypothesised that transduction is responsible for this similarity. Comparing the region of similarity in *lyc* with *CA_C1877*, there are no gaps but 28 mismatches occurred out of 252 base pairs. Is it possible that the phage transduced the *lyc* gene then evolution and environment caused the mutations/mismatches observed over time? If it was transduced, was it likely through generalised or specialised transduction? Considering that specialised transduction is phage packaging host DNA close to its locus as a result of an error in packaging, the *lyc* host gene is not located close to the phage genome integration point. Therefore, generalised transduction which involves packaging and transmitting DNA from any host locus (Wagner and Waldor, 2002) could be an explanation for the similarity. Generalised transduction has been described for host and phage virulence genes (Wagner and Waldor, 2002) but not for saccharolytic bacteria.

To test the hypothesis of generalised transducing capability, the phage and host were assessed for factors that lead to failed transduction. The first reason is that the presence of a CRISPR system can prevent transduction (Weinberger and Gilmore, 2012), nevertheless there is no native CRISPR/Cas system in *C. acetobutylicum* (Wasels *et al*., 2017). Secondly, host restriction modification systems can interrupt transduction. Although there are no identified phage-encoding methylase genes and there are 68 restriction recognition sites for Cac824I and 4 restriction recognition sites for Cac824II in phage ACK1, fortunately the main restriction modification gene, *CA_C1502*, has been knocked out from the *tcdR* phage-less mutant (See chapter 3). Besides phage progeny to be used for transduction is a product of successful plaque formation/phage infection, therefore it should be carrying methylation protection from the previous host (Enikeeva *et al*., 2010). Thirdly, transduction is prevented when abortive infection occurs. In support of this hypothesis, using the improved plaque assay described in Section 2.1.9, infection has been shown to be successful in *C. acetobutylicum* ATCC 824 thus providing a good positive control. Since these three factors do not seem like a problem to interfere with transduction capabilities of the phage, transduction test was designed as illustrated in Figure 5.17.

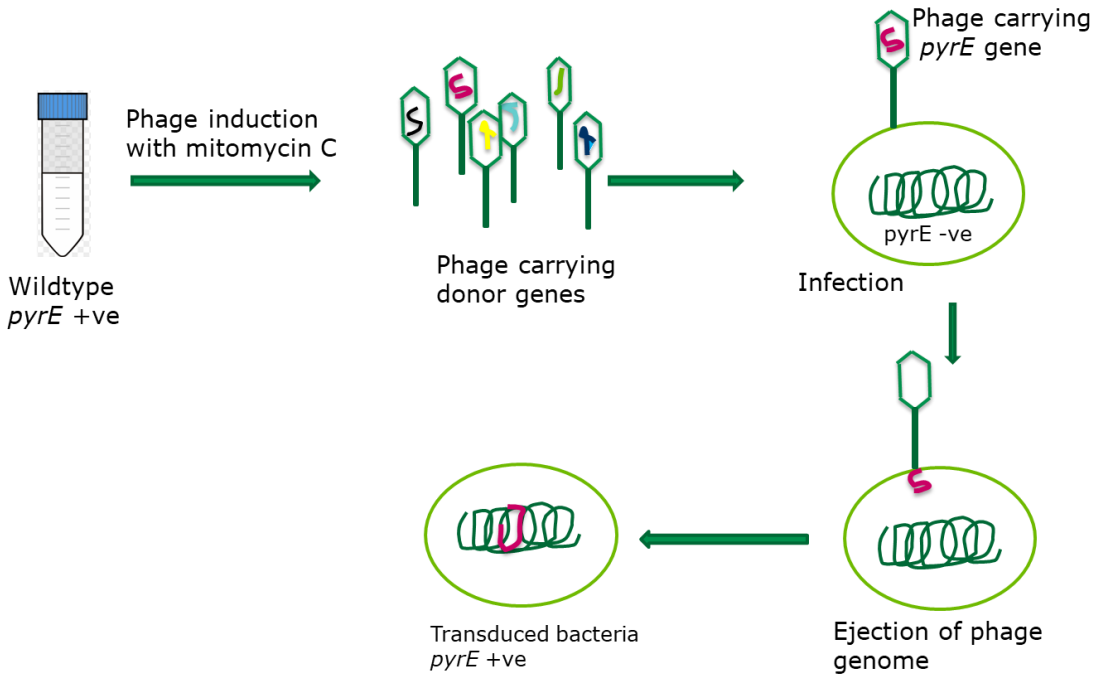

Figure 5.17: Demonstration of transduction test technique. Phage induction on wild type *C. acetobutylicum* ATCC 824 will yield a phage lysate containing a putative cocktail of host DNA including *pyrE* genes. When a *pyrE* mutant is infected with this phage lysate cocktail, phages carrying the *pyrE* gene will transduce the gene into the host thereby creating a *pyrE* positive strain that does not require uracil inclusion in the medium for growth.

The process of testing for generalised transduction in phage ACK1 strain was undertaken by first creating two *pyrE* auxotrophic mutants in a *tcdR*Δ*CAC1502*Δ*CAC3535* phage-free background. The first *pyrE* mutant, Δ*pyrE*::CRISPR phageless, was created by knocking out the full *pyrE* gene using CRISPR RiboCas system in *tcdR*Δ*CAC1502*Δ*CAC3535* while the second *pyrE* mutant, Δ*pyrE*::ACE phageless, was created by knocking out part of the *pyrE* gene using ACE in *tcdR*Δ*CAC1502*Δ*CAC3535*. The CRISPR RiboCas *pyrE* full gene KO plasmid was created by Dr. Bunmi Omorotionmwan while the ACE *pyrE* plasmid for partial KO was created by Dr. Muhammad Ehsaan.

The procedure for creating the Δ*pyrE*::CRISPR phageless and Δ*pyrE*::ACE phageless from *tcdR*Δ*CAC1502*Δ*CAC3535* mutant was explained in Materials and Methods section 2.2.10.4 and 2.2.10.5, respectively. The *pyrE* gene was chosen for this experiment because its acquisition can be selected for as conversion of the test strain to prototrophy, i.e. ability to grow on minimal media lacking uracil. Next phage transduction was tested by infecting *pyrE* mutants (partial gene KO and entire *pyrE*

gene KO) with phage induced from *C. acetobutylicum* wild type as described by Olson (2016). Two controls of *pyrE* mutants without phage infection and *pyrE* mutants infected with phage but plated on media supplemented with uracil were used. Adsorption was allowed for 20 minutes before pouring the phage-indicator bacteria mixture on the P2 minimal media. This technique was tried twice but there were no colonies observed on P2 minimal media without uracil added.

In another approach to detecting transduction, a double soft overlay of the bacteria-phage infection was prepared and incubated overnight before spreading the plaques on minimal media. No colonies were observed on the P2 minimal media without uracil. Based on the methods applied herein, no evidence was obtained to support the notion that ACK1 could be used for phage transduction.

5.3.9 Prediction of ACK1 Packaging Using Multiple Sequence Alignment Comparison by Log-extraction (MUSCLE)

Phage large terminases have two major components; N-terminal adenosine triphosphatase (ATPase) and C-terminal nuclease activity subunits. The primary function of phage terminases is to recognise phage DNA sequence for packaging and cleave it to create a type of DNA terminus structure (Casjens and Gilcrease, 2009). Six major categories of phage end/terminus structures are (i) 3' protuding single-stranded COS ends without circular permutation or terminal repeats (e.g., HK97 phage), (ii) 5'protruding single-stranded COS ends without circular permutation or terminal repeats (e.g., λ phage), (iii) short or long exact direct terminal repeats without circular permutation (e.g., T7 phage), (iv) circularly permuted direct terminal repeats (e.g., P22 phages), (v) terminal host DNA (e.g., Phage Mu) and (vi) covalently bound terminal protein (Φ29-like phage) (Casjens and Gilcrease, 2009). The type of terminus is indicative of the DNA replication strategy adopted as well as the mechanism of DNA packaging carried out by the phage terminase. Casjen and Gilcrease (2009) showed that the type of DNA terminus of a phage can be determined by bioinformatically classifying the large terminase amino acid sequences of various phages to recognisable separate groups. Another previous study by Cheepudom *et al.* (2015) has also demonstrated that this technique can be used to accurately determine the type of phage DNA terminus.

Since the ACK1 phage has not been proven to be transducing by the test shown in section 5.3.8, it would be interesting to know the type of its DNA terminus as it will give more information on whether ACK1 is capable of transduction or not. Figure 5.18 displays the result of neighbour-joining cladogram tree comparing the large terminase of the *C. acetobutylicum* ACK1 phage with that of 54 phages with known terminase ends. These phage large terminases in this tree grouped were as described by Cheepudom *et al.* (2015) except those of *C. acetobutylicum* ACK1 and *C. perfringens* ATCC 3626 phage. Based on the neighbour-joining cladogram tree in Figure 5.19, ACK1 phage large terminase cleaves the phage genome to give 3' protruding single-stranded end DNA at the beginning and end for packaging. This confirms that ACK1 phage is not capable of carrying out generalised and specialised transduction. This predicts that the DNA replication style is similar to that of HK97 phage which is likely to form a rolling circle and then a concatemeric DNA that gets packaged into the capsid (Casjens and Gilcrease, 2009).

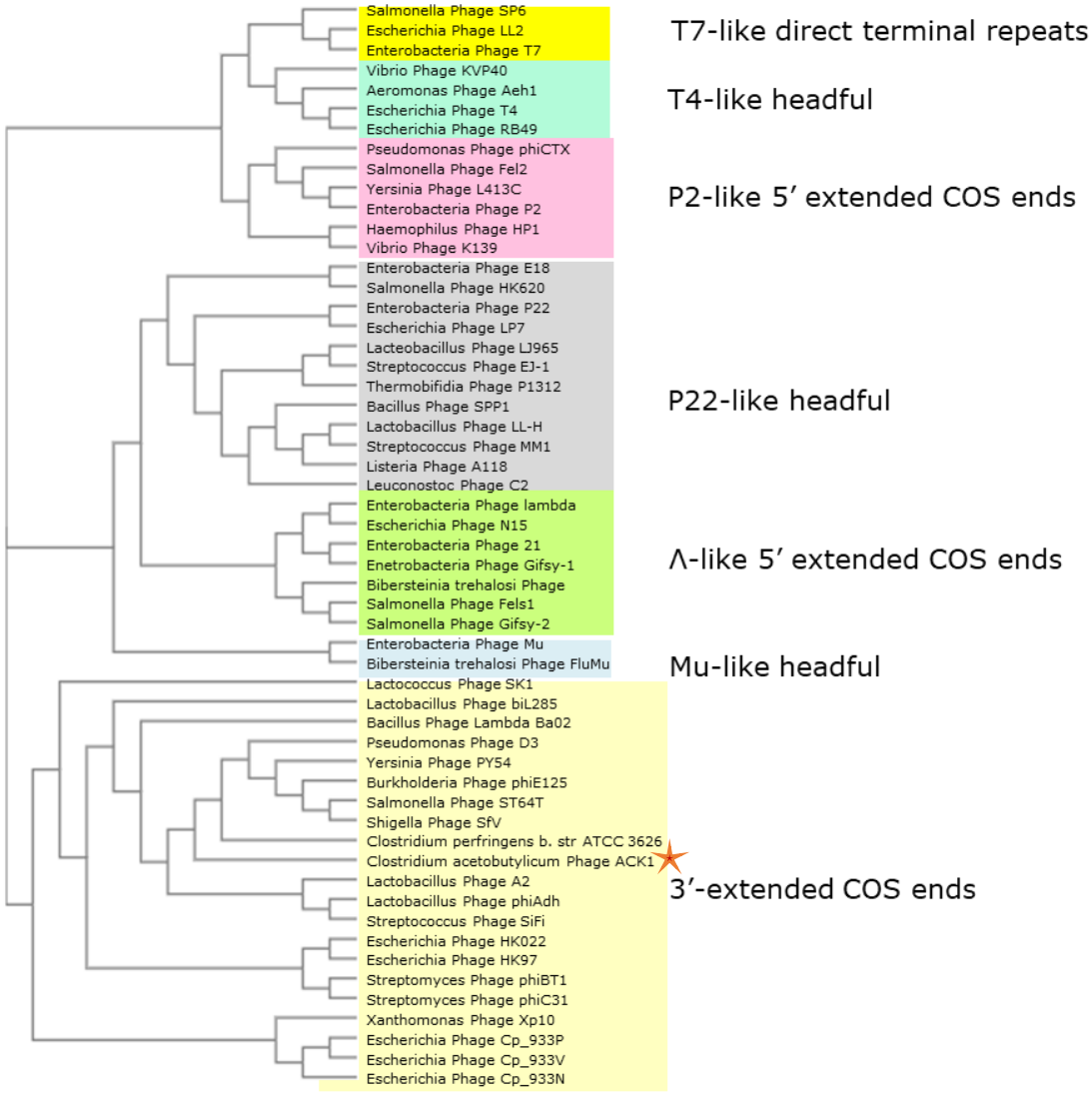

Figure 5.18: Neighbour-joining cladogram tree of the large terminase subunit of ACK1 Phage of *C. acetobutylicum* compared with 54 phages with known DNA termini. *Clostridium acetobutylicum* ACK1 phage is marked with a red asterisk. Except *Clostridium acetobutylicum* ACK1 phage and *Clostridium perfringens* ATCC 3626 phage, the large terminase amino acid sequences were classified as described by Cheepudom *et al.* (2015). Multiple Sequence Alignment Comparison by Log-extraction (MUSCLE) software tool (output format CLUSTAL W) was used for DNA termini study (Edgar, 2004). (https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/).

5.3.10 Development of Plasmid-based Integration System Utilising the Site-specific Recombinase and *attP* Locus in Phage ACK1.

Previous work by Breüner *et al.* (2001) has shown that Tp901-1 phage integrase is the only gene needed to carryout recombination. Tp901-1 integrase shares some similarities with that of phage ACK1. Having identified the integrase and its *attB*/*attP* sites, three integration plasmids were constructed containing the integrase and *attP*.

The plasmid was derived from the pMTL80000 modular plasmid series (www.clostron.com), pMTL85141 is as described in Materials and Methods, Figure 2.3. The Gram-positive origin of replication was deleted by digesting it with FseI and AscI and self-ligating the larger DNA fragment to create pMTL85141*ΔrepL*. Then the integrase with embedded *attP* site in *C. acetobutylicum* was cloned with three sets of PCR primers: Cac1950AscI/F2 and Cac1951FseI/R1; Cac1950AscI/F2 and Cac1867FseI/R2; Cac1950AscIF4 and Cac1868FseIR5 with their respective restriction sites included using *C. acetobutylicum* wild type genome containing circularised phage plasmid as template.

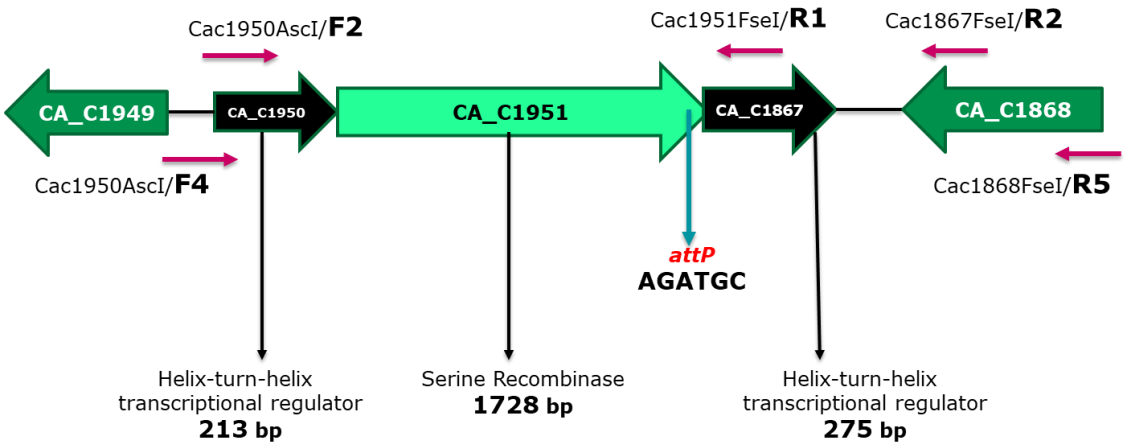

Figure 5.19: Primer locations for phage integrase cloning. Primers Cac1951FseI/R1 and Cac1950AscI/F2 were used to generate integrase plasmid pMTL85141-51-50. Primers Cac1867FseI/R2 and Cac1950AscI/F2 were used to generate integrase plasmid pMTL85141-67-50 while Cac1868FseI/R5 and Cac1950AscI/F4 were used to generate integrase plasmid pMTL85141-68-50.

Primer set Cac1951FseI/R1 and Cac1950AscI/F2 amplified the *CA_C1951* integrase gene with the putative *attP* site and the *CA_C1950* transcriptional regulator gene upstream which is likely to have the *CA_C1951* promoter embedded in it or located upstream of the gene as seen in Fig 5.19. Primer set Cac1867FseI/R2 and Cac1950AscI/F2 amplified the *CA_C1951* integrase gene with the putative *attP* site, the *CA_C1950* transcriptional regulator gene upstream and the *CA_C1867* transcription regulator XRE gene downstream of the *CA_C1951*. Furthermore, the third primer set Cac1868FseI/R5 and Cac1950AscI/F4 (as shown in Fig 5.19) amplified the *CA_C1951* integrase gene with the putative *attP* site, the *CA_C1950* transcriptional regulator gene upstream, the *CA_C1867* transcriptional regulator XRE gene and the *CA_C1868* (unknown function) both downstream of *CA_C1951*. The PCR products of the three

reactions were run on a gel and the bands were cut and purified. Following PCR product purification, they were digested at their respective restriction sites *Fse*I and *Asc*I. Cleaned digests of 1951FseIR1-1950AscIF1 PCR fragments were ligated with digested pMTL85141Δ*repL* plasmid to yield pMTL85141-51-50 (Figure 5.20). In a similar way, cleaned digests of PCR fragments 1867FseIR2-1950AscIF2 and 1868FseIR5-1950AscIF4 were ligated with digested pMTL85141Δ*repL* plasmid to yield pMTL85141-67-50 and pMTL85141-68-50, respectively (Figure 5.21 and Figure 5.22).

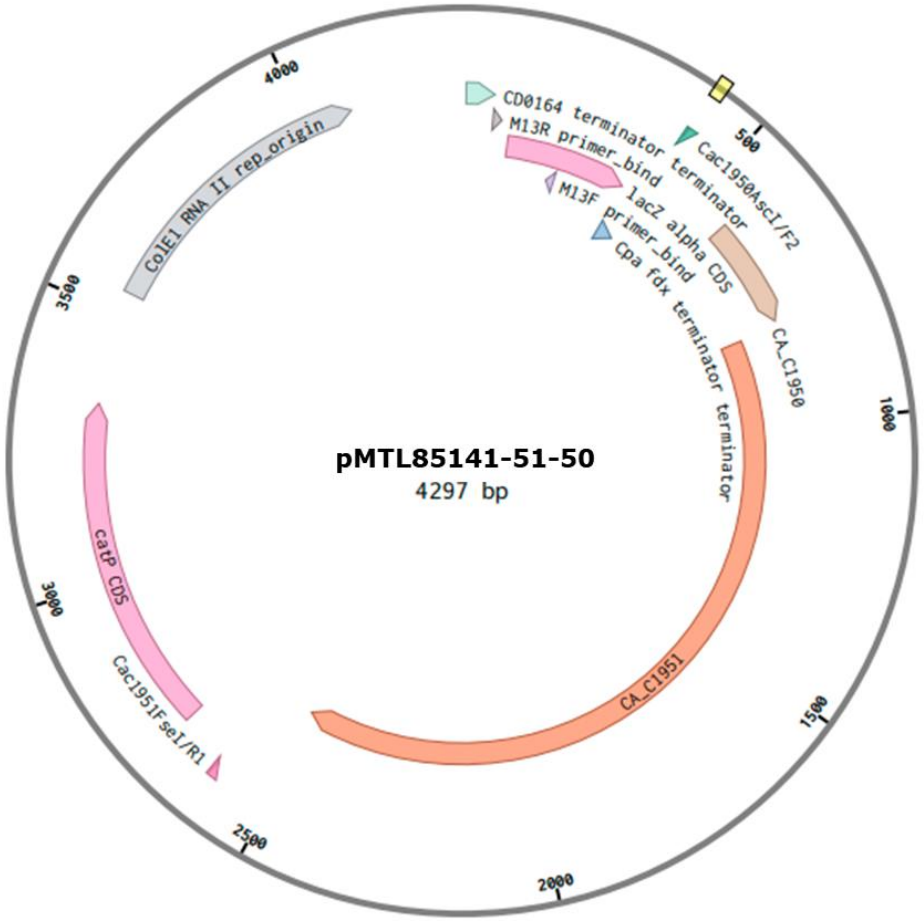

Figure 5.20: Plasmid map of integrase suicide plasmid pMTL85141-51-50.

This was derived from the pMTL80000 modular plasmid series. The plasmid contains: *catP*, chloramphenicol/thiamphenicol marker; ColE1 RNA II, Gram-negative origin of replication; *fdx* promoter from *C. sporogenes* followed by *lacZ* region where the multiple cloning site is. *repL* was deleted by digesting with FseI and AscI and replaced with the integrase *CA_C1951* gene with the *attP* site embedded in it and XRE-like protein-encoding *CA_C1950*.

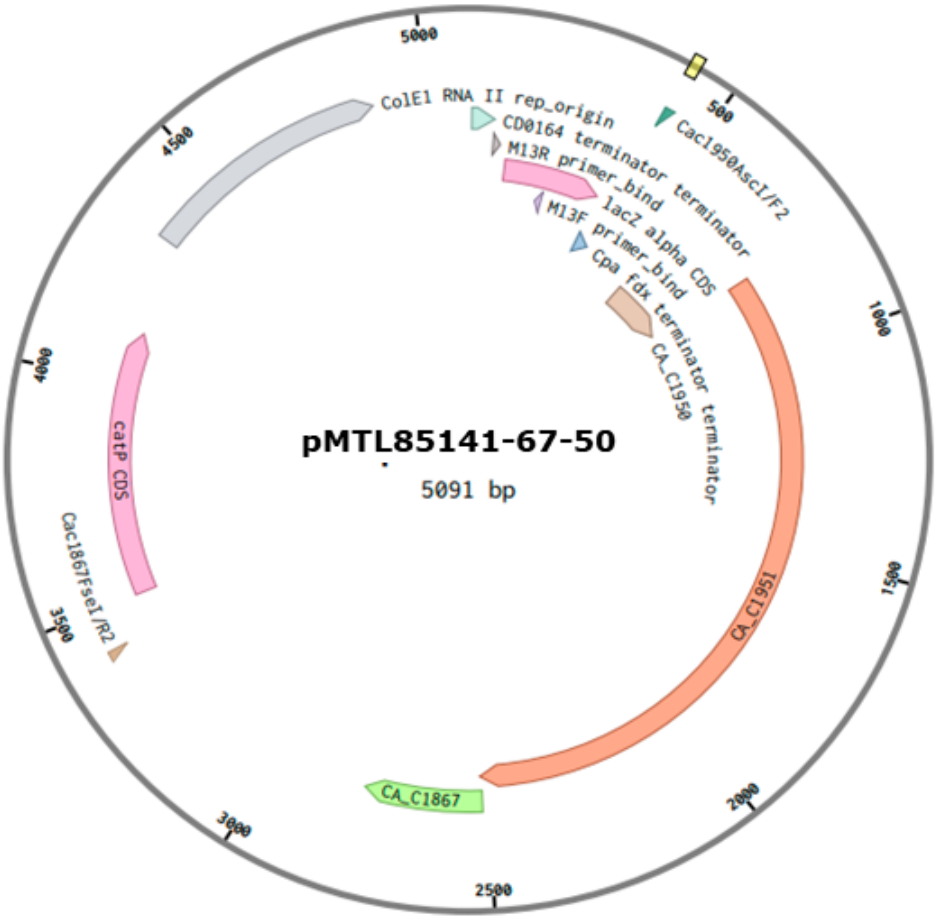

Figure 5.21: Plasmid map for integrase suicide plasmid pMTL85141-68-50

This was derived from the pMTL80000 modular plasmid series. The plasmid contains: *catP*, chloramphenicol/thiamphenicol marker; ColE1 RNA II, Gram-negative origin of replication; *fdx* promoter from *C. sporogenes* followed by *lacZ* region where the multiple cloning site is. *repL* was deleted by digesting with FseI and AscI and replaced with the *CA_C1950*, the integrase *CA_C1951* with the *attP* site embedded in it and the transcriptional regulator gene *CA_C1867*.

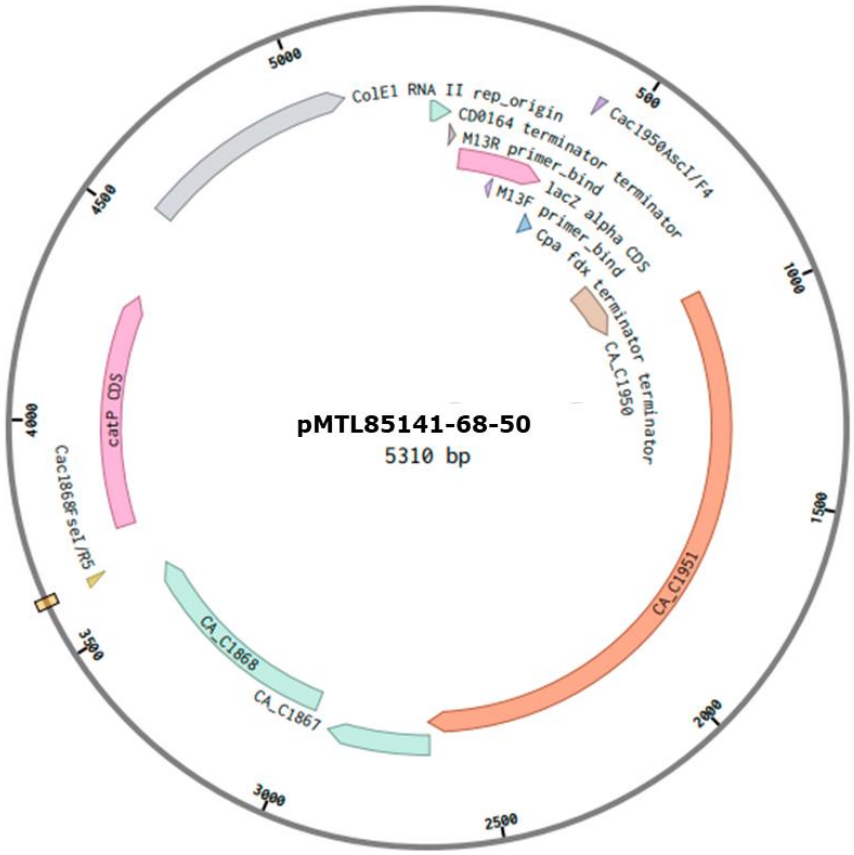

Figure 5.22: Plasmid map for integrase suicide plasmid pMTL85141-68-50

This was derived from the pMTL80000 modular plasmid series. The plasmid contains: *catP*, chloramphenicol/thiamphenicol marker; ColE1 RNA II, Gram-negative origin of replication; *fdx* promoter from *C. sporogenes* followed by *lacZ* region where the multiple cloning site is. *repL* was deleted by digesting with FseI and AscI and replaced with the *CA_C1950* gene, the integrase *CA_C1951* with the *attP* site embedded in it, the transcriptional regulator gene *CA_C1867* and the *CA_C1868 gene* of unknown function.

After the cloning of the three suicide integrase plasmids in *E. coli* TOP10, they were extracted using a Monarch miniprep kit. To determine autonomous recombinase properties of ACK1 phage integrase, the *tcdR*Δ*CAC1502*Δ*CAC3535* phage indicator mutant was successfully transformed with pMTL85141-67-50 and pMTL85141-68-50 but not pMTL85141-51-50 via electroporation. Three transformants of pMTL85141-67-50 and five transformants of pMTL85141-68-50 were screened for the site-specific recombinase delivery vector using genome primers flanking *attB*, Cac1866F and Cac1951R2 as shown in Figure 5.23.

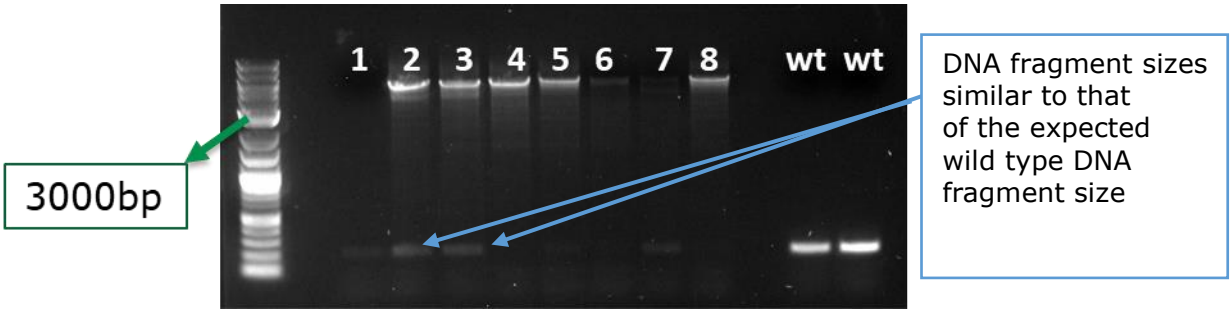

Figure 5.23: PCR screening of *C. acetobutylicum* transformants for recombination of pMTL85141-67-50 and pMTL85141-68-50. Screening was done with primers Cac1866F and Cac1951R2 flanking the *attB* site. Lane 1 represents one of the pMTL85141-67-50 transformants but it has no band. Lanes 2 and 3 represnt pMTL85141-67-50 colonies with expected DNA fragment sizes of 5,343 bp. Lanes 4,5,6,7 and 8 represent pMTL85141-68-50 colonies with expected DNA fragment sizes of 5,562 bp. WT is wild type strain with expected band size is 252 bp and/or 53,632 bp (which cannot be achieved via the employed PCR method).

The results obtained confirmed the proposed role of ACK1 phage integrase and demonstrated functionality of integrative recombination between the *attB* site of the bacterial genome and the *attP* site in a vector. By carrying out an integration test with the three vectors, it indicated that ACK1 phage integrase activity requires *attB*, phage integrase and *CA_C1950* (213 bp) located upstream and/or its promoter and *CA_C1867* (275 bp) located downstream. It showed that the *CA_C1868* gene is not required for integrative recombination to occur. These phage sequences required for integrative recombination are somewhat comparative to that of TP901-1 phage integrase from *Lactococcus lactis* which has the *attP* site located downstream of the phage integrase. For integration to take place, it required the TP901-1 phage integrase gene (1467 bp) and 425 bp sequence upstream of the integrase (Christainen *et al*., 1996). However, the integrase can function with the replacement of 425 bp with a promoter which indicated that the 425 bp upstream region was not necessary (Stoll *et al*., 2002). In contrast, tyrosine recombinases like lambda integrase requires the *attP* site, the integrase as well as accessory DNA-binding proteins: integration host factor proteins (IHF), excision recombination proteins (Xis) and other proteins in which some are required for integration, some for excision and others are required for both (Landy, 2015).

ACK1 site-specific recombination has great potential as a genetic tool for the integration of heterologous DNA into the *C. acetobutylicum* genome. It will ensure single-copy integration/recombination and minimise inactivation of host genes and

insertional mutations with a potential for yielding unlimited cargo capacity. Since the recombination process for ACK1 does not require many phage-encoded genes, it makes development of a simple gene cargo plasmid feasible. Further work is needed to determine the stability of the integration as there were DNA fragment sizes similar to those of the wild type DNA fragments in the PCR integrants screening (Figure 5.23). This suggested that the inserted sequences had been removed in some of the cells present in the population. In addition, the minimial sequences required at *attB* and *attP* for recombination to occur need to be investigated. Nevertheless exploring this engineering technique may be an interesting way to introduce genes to improve the ABE metabolic pathway.

5.4 Conclusion

There is limited research on the genetics of phages of *C. acetobutylicum*, the model organism for industrial-scale butanol production. Exploitation of phage tools could potentially add to the synthetic biologist's arsenal in understanding this important industrial bacterium. Here we report on the characterisation of a temperate phage isolated from *C. acetobutylicum* ATCC 824, designated phage ACK1. Phage ACK1 is 53,380 bp spanning from *CA_C1866* to *CA_C1951* and not as originally annotated *CA_C1878* to *CA_C1953* (Nölling *et al.,* 2001). Experiments to determine whether insertion of a gene (*tcdR*) encoding the *C. difficile* sigma factor TcdR were responsible for phage excision were inconclusive. Although there is a possibility that a pseudolysogenic relationship exists between the phage and host, which could potentially lead to isolation of a phage-less strain, the trigger for phage excision remains unclear.

Four lifestyles of phage ACK1 were observed in *C. acetobutylicum* ATCC 824. The first is integration of the phage into the host genome; the second is integration as well as plasmid circular forms of the phage; the third is a plasmid form of the phage in the host cytoplasm without integration; and last is lytic cycle activation forming virons. It is unknown which path ACK1 will follow after infecting *C. acetobutylicum* ATCC 824. The ACK1 phage genome was found to circularise and form multiple copy plasmids in host cell. This shows a good potential for genetic engineering by introducing genes at the phage locus.

Through bioinformatics analysis with BLAST and PHASTER, putative gene functions were assigned to the phage genes. It has 84 ORFs with 48 ORFs of unknown function. The mosaic architecture of ACK1 comprises gene sets arranged into phage DNA replication, transcription and metabolism, phage lysis, phage structure, phage packaging and phage integrase.

ACK1 appears to have its own replication initiation protein and other DNA replication and transcription genes like DNA polymerase, transcription regulation factor XRE as well as putative lysogeny regulators like anti-repressor and Cro/C1-like transcription regulators. Its tail tape measure gene is the longest gene in the genome which is not surprising since it has a long tail as observed under TEM. Some putative phage packaging and lysis genes were assigned functions.

ACK1 integrase has a conserved resolvase domain which indicates the attachment sites are most likely to be cut in a staggered manner creating 2 bp overhangs like most resolvase-like integrases. It shares some similarity with that of TP901-1 integrase from *Lactococcus lactis*, which is known to be the only gene required for recombination to occur. The phage attachment and bacterial attachment sites were determined through a combination of Illumina whole genome sequencing and Sanger sequencing of cloned sites and confirmed through the integrase plasmid test. It was observed that the core specific sequence in *C. acetobutylicum* shared by *attB*, *attP*, *attL* and *attR* is 5' AGATGC 3'. This core sequence is integrated at the 3' end of the integrase.

Studying the attachment sequences revealed that it was not perfectly or imperfectly palindromic, thus this eliminates the potential of the integrase to be from the DNA invertase family. It is a resolvase-like integrase which has an advantage over tyrosine recombinase. Serine recombinases often do not need divalent ions like magnesium to drive recombination, rather they depend on the amino acid side chain. Exploring the potential for applying recombination technology in this strain, the integrases were cloned into a plasmid and tested for recombination into phage-less host. The result showed that the phage-encoded ORF required for integrative recombination activity were phage integrase and most likely the flanking proteins: *CA_C1867* and *CA_C1950* and/or its promoter. The two flanking ORFs are predicted to encode helix-turn-helix DNA transcriptional proteins that posess DNA-binding motifs and may pay an essential role during recombination activity.

The 26.5% of the *lyc* gene in the host showed 87% identity with the phage gene, *CA_C 1877*, which may suggest there is transduction potential in ACK1. However, after using two transduction methods, evidence of transducing activity was not found. To shed light on the type of ends of the ACK1 phage genome, a neighbour-joining cladogram tree of the large terminase amino acid sequence of ACK1 phage compared with 54 phages of known genome terminal structures was generated and analysed with MUSCLE software tool using the output format Clustal W. ACK1 phage was categorised as having 3' single-stranded cohesive extension at its ends which suggests that it is a non-transducing phage and the mechanism of DNA replication is by a rolling circle strategy which forms contacemeric substrate that becomes packaged.

The plaque assay protocol for the ACK1 phage was successfully improved on by the inclusion of the following steps: (i) adding buffer and metal ions in the top and bottom agar; (ii) washing indicator strains to remove the accumulated solvent; (iii) adjusting the pH of phage and bacteria culture, and; (iv) limiting the subculture time during phage induction with mitomycin C and uv-irradiation. It was observed that the plaques were about 1 mm in size and attachment of ACK1 to host receptors may have been pH sensitive. TEM images of the phages indicate that the phage belongs to the *Siphoviridae*. It has an elongated icosahedral head with a long tail without visible tail fibres. It has a linear double-stranded DNA which circularises in the host.

Key points about this chapter

- The characterised phage ACK1 isolated from *C. acetobutylicum* ATCC 824 is also present in 3 other *C. acetobutylicum* strains; EA2018, DSM1731 and LJA
- The phage genome comprises 53,380 bp of double-stranded DNA with GC content of 31.64%
- It contains 84 ORFs of which 36 were ascribed a putative function
- The genome integrase contains the phage attachment site at the 3' terminus and the core attachment site sequence for *attB, attP, attL* and *attR* is 5' AGATGC 3'
- The integrase is a resolvase type of the serine recombinase family.
- The phage genome can integrate into the host chromosome and/or exists in a circular autonomous form.

- Phage attachment to host is dependent on presence of divalent ions and pH 6-7. It forms relatively small uniform plaques of 1 mm.
- ACK1 phage is of the family *Siphoviridae*; elongated capsid with long non-contractile tail without visible tail fibres.
- ACK1 phage DNA has 3’ single-strand extension cohesive ends, thus, it is a non-transducing phage.

Chapter 6

Conclusions and Further Work

Chapter 6

6.0 Conclusions and Further Work

C. acetobutylicum is the model organism for industrial-scale butanol production. Creating a road map for genetic engineering of *C. acetobutylicum* is dependent on the full annotation of the genome being accessible and genetic tools for introducing DNA into the genome being available (Minton *et al*., 2016). Although there is easy access to the full annotation of *C. acetobutylicum* genome on NCBI, many genes in the genome have unknown function. Furthermore, there is need to improve the efficiency of DNA transfer into the cells and create more effective genetic engineering tools. To develop a commercial strain for industrial biobutanol production, it is important to gain better understanding of the essential and non-essential genes as well as their functions in metabolic pathway.

The aims of this thesis was to improve the transformation efficiency of *C. acetobutylicum*, to use TraDIS studies of the genome to determine essential and non-essential genes and characterise the temperate phage for phage genetic study and tool development.

6.1 Improving the Transformation Efficiency of *C. acetobutylicum*

The first aim of this project was to improve the DNA transfer efficiency of the strain by deleting the restriction modification systems. This organism has type II restriction modification (RM) systems that prevent the transformation of cells with foreign DNA, thus DNA has to be methylated to circumvent the system. The data showed that Cac824I is the main restriction modification system and that mutants of only Cac824II (the second RMS) could accept unmethylated plasmid DNA. Almost 6-fold improvement in transformation efficiency was achieved when both RM systems genes were deleted. In a recent study by Foulquier *et al*. (2019), treating cultured non-sporulating *C. acetobutylicum* ATCC 824 Δ*cac1502*Δ*cac3535*Δ*upp*Δ*pSOL* with 150 µg/ml of lysozyme, before the centrifugation and wash steps in making competent cells, significantly improved the transformation efficiency when electroprated with unmethylated pCons2.1. It improved from 142 transformants per µg of unmethylated plasmid DNA without lysozyme treatment to 6.5×10^3 transformants per µg of unmethylated plasmid DNA with lysozyme treatment which is comparable with the

transformation efficiency of the double restriction-less sporulating Δ*upp* strain (Foulquier *et al*., 2019). It would be interesting to see if the transformation efficiency of a restriction-less sporulating strain with an unmethylated plasmid can be improved further by the lysozyme treatment during the competent cells' technique. Perhaps the cell wall of *C. acetobutylicum* is another important barrier to plasmid DNA uptake that should be paid attention to.

6.2 Investigating the Essential Genes for Survival of *C. acetobutylicum*

The aim of this thesis was to elucidate genes that are essential for survival in standard laboratory conditions using TraDIS study. A suicide transposon plasmid was tested on the double restriction-less strains and the transformation efficiency was not sufficient to create a large pool of transposon mutants. Two conditional replicative transposon plasmids, pMTL-CW20 and pMTL-CW21 which use the IPTG inducible system and aTC inducible system, respectively, were tested. Various aspects of the protocol for library creation and preparation were altered to improve and investigate plasmid loss (which is one of the main challenges of using the replicative plasmid).

The TraDIS results aided the identification of 537 potentially essential genes in *C. acetobutylicum* for survival in standard laboratory conditions. The inclusion of genes from the megaplasmid in the essential gene list indicates that the non-essential genes were not saturated enough with the transposon. Further work needs to be done by creating and sequencing a high density library of transposon mutants for high statistical analysis of essential genes of *C. acetobutylicum* ATCC 824. This thesis showed that the plasmids used and method applied have great potential in gene functionality studies.

6.3 Isolation and Characterisation of Lysogenic Phage

In order to develop robust genetic engineering tools for *C. acetobutylicum* ATCC 824, exploitation of phage tools could potentially add to synthetic biologist's arsenal in understanding the bacteria. A temperate phage in *C. acetobutylicum* was isolated, characterised and putative annotation/function was assigned to some of the genes. The phage is of the family *Siphoviridae*, with a long tail and elongated capsid. It was further identified as non-transducing and has 3'-extended cohesive DNA ends. To facilitate the development of a phage tool, the *attP* site, *attB* site and integrase were identified. To test the potential for a site-specific recombination tool, the *attP* site and integrase gene were cloned into a suicide vector, then a phage-less strain was transformed with the

cloned vector. The ability to integrate DNA at the *attB* can be explored in introducing genes that will benefit the pathway for biobutanol production. To improve the efficiency of insertion of large genes, knocking out the RM as well as lysozyme treated competent cells protocols could be applied. The advantage of exploring a site-specific recombination gene-editting tool in creating a tailor-made industrial strain is that it eliminates the need for plasmid loss and maintaining a plasmid by introducing antibiotics into the large fermentation feedstock. Antibiotics are expensive and would add to the cost of fermentation. It also reduces the probability of a phage attack at that *attB* site since the sites is already occupied with desirable genes. The segregational stability of genes inserted at that locus needs to be investigated if this technique is to be used for cloning.

References

Abo, B. O., Gao, M., Wang, Y., Wu, C., Wang, Q. & Ma, H. (2019). Production of butanol from biomass: recent advances and future prospects. *Environmental Science and Pollution Research International*, **26:**20164–20182.

Adams, M. J., Lefkowitz, E. J., King, A. M., Bamford, D. H., Breitbart, M., Davison, A. J., Ghabrial, S. A., Gorbalenya, A. E., Knowles, N. J., Krell, P., Lavigne, R., Prangishvili, D., Sanfaçon, H., Siddell, S. G., Simmonds, P. & Carstens, E. B. (2015). Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses. *Archive of Virology*, **160**(7):1837-1850.

Akerley, B. J., Rubin, E. J., Novick, V. L., Amaya, K., Judson, N. and Mekalanos, J. J. (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenza, Proceedings of the National Academy of Sciences*, **99**(2):966-971.

Atmadjaja, A., Harding, A. & Jenkinson, E. (2016). *Sporulation in solventogenic Clostridium sp.: Investigating the link between endospore formation and solvent production in the industrial ABE fermentation,* 14th International Conference on the Genetics, Physiology and Synthetic Biology of Solvent- and Acid-forming *Clostridia,* Darmouth College, Hanover, NH. August 28-31, 2016.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research,* Jul 8;**44**(W1):W16-21.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K, Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, **9**:75.

Babb, B. L., Collett, H. J., Reid, S. J. & Woods, D. R. (1993). Transposon mutagenesis of *Clostridium acetobutylicum* P262: isolation and characterization of solvent deficient and metronidazole resistant mutants. *FEMS Microbiology Letters,* **114** (3):343-348.

Baer, S. H., Blaschek, H. P. & Smith, T. L. (1987). Effect of Butanol Challenge and Temperature on Lipid Composition and Membrane Fluidity of Butanol-Tolerant *Clostridium acetobutylicum*. *Applied and Environmental Microbiology,* **53:** 2854-2861.

Baer, S. H., Bryant, D. L. & Blaschek, H. P. (1989). Electron Spin Resonance Analysis of the Effect of Butanol on the Membrane Fluidity of Intact Cells of *Clostridium acetobutylicum*, *Applied and Environmental Microbiology*, **55**(10):2729-2731.

Balan, V. (2014). Current challenges in commercially producing biofuels from lignocellulosic biomass. *International Scholarly Research Notices: Biotechnology*, **2014**:1-31.

Bandara, N., Jo, J., Ryu, S. & Kim, K. (2012). Bacteriophages BCP1-1 and BCP8-2 require divalent cations for efficient control of *Bacillus cereus* in fermented foods. *Food Microbiology*, **31**:9–16.

Bankar, S. B., Survase, S. A., Ojamo, H. & Granström, T. (2013). Biobutanol: the outlook of an academic and industrialist. *RVC Advances*, **3**:24734-24757.

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

Barquist, L., Mayho, M., Cummins, C., Cain, A. K., Boinett, C. J., Page, A. J., Langridge, G. C., Quail, M. A., Keane, J. A. & Parkhill, J. (2016). The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics*, **32**(7):1109–1111.

Bensah, E. C. & Mensah, M. (2013). Chemical pretreatment methods for the production of cellulosic ethanol: technologies and innovations. *International Journal of Chemical Engineering*, **2013**:719607.

Biémont, C. (2010). A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics*, **186**(4):1085–1093.

Birgen, C., Dürre, P., Preisig, H. A. & Wentzel, A. (2019). Butanol production from lignocellulosic biomass: revisiting fermentation performance indicators with exploratory data analysis. *Biotechnology for Biofuels*, **12**:167.

Blackburn, M., Mazzacano, C. A. S., Fallon, C. & Black, S. H. (2014). *Oil in our ocean: A review of the impact of oil spills ion marine invertebrates.* The Xerces Society for Invertebrates Conservation, Portland, Oregon, pp4.

Blouzard, J. C., Valette, O., Tardif, C. & de Philip, P. (2010). Random mutagenesis of *Clostridium cellulolyticum* by using a Tn1545 derivative. *Applied and Environmental Microbiology*, **76**:4546–4549.

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, **493**(7432):429-432.

Bowles, L. K. & Ellefson, W. L. (1985). Effects of butanol on *Clostridium acetobutylicum*. *Applied and Environmental Microbiology*, **50**(5): 1165–1170.

Breüner, A., Brøndsted, L. & Hammer, K. (2001). Resolvase-like recombination performed by the TP901-1 integrase. *Microbiology*, **147**(8):2051–2063.

Browning, C., Shneider, M. M., Bowman, V. D., Schwarzer, D. & Leiman, P. G. (2012). Phage pierces the host cell membrane with the iron-loaded spike. *Structure*, **20**:326–339.

Brüssow, H. & Hendrix, R. W. (2002). Phage genomics: small is beautiful. *Cell*, **108**:13–16.

Brüssow, H., Canchaya, C. & Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews: MMBR*, **68**(3):560–602.

Cañadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R. & Minton, N. P. (2019). RiboCas: A universal CRISPR-based editing tool for *Clostridium*. *ACS Synthetic Biology*, **8**(6):1379-1390.

Carbon Disclosure Project (CDP) Climate Change Report (2016). Nordic edition. pp 1-35. Available at https://b8f65cb373b1b7b15febc70d8ead6ced550b4d987d7c03fcdd1d.ssl.cf3.rackcdn.com/cms/reports/documents/000/001/234/original/CDP-Nordic2016-v13-MR-Links.pdf?1478707498

Cartman, S. T. & Minton, N. P. (2010). A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. *Applied and Environmental Microbiology*, **6**:1103-1109.

Casjens, S. R. & Gilcrease, E. B. (2009). Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods in Molecular Biology*, **502**:91–111.

Casjens, S. R. & Hendrix, R. W. (2015). Bacteriophage lambda: early pioneer and still relevant. *Virology*, **479-480**:310–330.

Cenens, W., Makumi, A., Mebrhatu, M. T., Lavigne, R. & Aertsen, A. (2013). Phage–host interactions during pseudolysogeny Lessons from the Pid/dgo interaction. *Bacteriophage*, **3**:1, e25029.

Cenens, W., Mebrhatu, M. T., Makumi, A., Ceyssens, P. J., Lavigne, R., Van Houdt, R., Taddei, F. & Aertsen, A. (2013). Expression of a novel P22 ORFan gene reveals the phage carrier state in *Salmonella typhimurium*. *PLoS Genetics*, **9**(2):e1003269.

Chao, M. C., Abel, S., Davis, B. M. & Waldor, M. K. (2016). The design and analysis of transposon insertion sequencing experiments. *Nature Reviews. Microbiology*, **14**(2): 119–128.

Cheepudom, J., Lee, C. C., Cai, B. & Meng, M. (2015). Isolation, characterization, and complete genome analysis of P1312, a thermostable bacteriophage that infects *Thermobifida fusca*. *Frontiers in Microbiology*, **6**:959.

Chen, C. K. & Blaschek, H. P. (1999). Effects of acetone on the molecular physiological of *Clostridium beijerinckii* NCIMB 8052 solvent production and strain degeneration. *Applied and Environmental Microbiology*, **65**:499-505.

Chen, H. & Wang, H. (2016). *Technologies for Biochemical Conversion of Biomass*. London: Academic Press, pp1-132.

Chaikeeratisak, V., Nguyen, K., Egan, M. E., Erb, M. L., Vavilina, A. & Pogliano, J. (2017). The phage nucleus and tubulin spindle are conserved among large *Pseudomonas* phages. *Cell Reports*, **20**:1563–1571.

Chhibber, S., Kaur, T. & Kaur, S. (2014). Essential role of calcium in the infection process of broad-spectrum methicillin-resistant *Staphylococcus aureus* bacteriophage. *Journal of Basic Microbiology*, **54**:775–780.

Chiao, J. S. & Sun, Z. H. (2007). History of acetone-butanol-ethanol fermentation industry in China: Development of continuous production technology. *Journal of Molecular Microbiology and Biotechnology*, **13**(1-3):12-14.

Cho, C. & Lee, S. Y. (2017). Efficient gene knockdown in *Clostridium acetobutylicum* by synthetic small regulatory RNAs. *Biotechnology and Bioengineering,* **114**(2):374-383.

Christiansen, B., Johnsen, M. G., Stenby, E., Vogensen, F. K., & Hammer, K. (1994). Characterization of the lactococcal temperate phage TP901-1 and its site-specific integration. *Journal of Bacteriology,* **176**(4):1069–1076.

Coyle, E. D. & Simmons, R. A. (2014). *Understanding the Global Energy Crisis*. Indiana: Purdue University Press, pp1-304.

Croux, C. & García, J. L. (1991). Sequence of the *lyc* gene encoding the autolytic lysozyme of *Clostridium acetobutylicum* ATCC 824: comparison with other lytic enzymes. *Gene*, **104**(1):25-31.

Croux, C., Nguyen, N., Lee, J., Raynaud, C., Saint-Prix, F., Gonzalez-Pajuelo, M., Meynial-Salles, I. & Soucaille, P. (2016). Construction of a restriction-less, marker-less mutant useful for functional genomic and metabolic engineering of the biofuel producer *Clostridium acetobutylicum*. *Biotechnology of Biofuels*, **9**:23.

Crozat, E. & Grainge, I. (2010). FtsK DNA translocase: the fast motor that knows where it's going. *Chembiochem*, **11**(16):2232–2243.

Cvirkaite-Krupovic, V., Krupovic, M., Daugelavicius, R. & Bamford, D. H. (2010). Calcium ion-dependent entry of the membrane-containing bacteriophage PM2 into its Pseudoalteromonas host. *Virology*, **405**(1):120-128.

Dembek, M., Barquist, L., Boinett, C. J., Cain, A. K., Mayho, M., Lawley, T. D., Fairweather, N. F. & Fagan, R. P. (2015). High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *mBio,* **6**(2):e02383.

Desvaux, M., Guedon, E., & Petitdemange, H. (2000). Cellulose catabolism by *Clostridium cellulolyticum* growing in batch culture on defined medium. *Applied and Environmental Microbiology*, **66**(6):2461–2470.

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P. & Moineau S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology*, **190**(4):1390-400.

Dong, H., Zhang, Y., Dai, Z. & Li, Y. (2010). Engineering *Clostridium* strain to accept unmethylated DNA. *PLoS ONE,* **5**(2):e9038.

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**:59-67.

Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., Delattre, A. S. & Lavigne, R. (2012). Learning from bacteriophages - advantages and limitations of phage and phage-encoded protein applications. *Current Protein and Peptide Science*, **13**(8):699–722.

Dürre, P. (2007). Biobutanol: an attractive biofuel. *Journal of Biotechnology,* **2**:1525-1534.

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics, 5 (**1**):113.

Ehsaan, M., Kuit, W., Zhang, Y., Cartman, S. T., Heap, J. T., Winzer, K. & Minton, N. P. (2016). Mutant generation by allelic exchange and genome resequencing of the biobutanol organism *Clostridium acetobutylicum* ATCC 824. *Biotechnology for Biofuels*, **9**:4.

Elshahed, M. S. (2010). Microbiological aspects of biofuel production: Current status and future Directions. *Journal of Advanced Research,* **1**:103-111.

Enikeeva, F. N., Severinov, K. V. & Gelfand, M. S. (2010). Restriction-modification systems and bacteriophage invasion: who wins? *Journal of Theoretical Biology*, **266**(4):550–559.

Fang, G., Rocha, E. & Danchin, A. (2005). How essential are non-essential genes? *Molecular Biology and Evolution,* **22**(11):2147–2156.

Faulds-Pain, A. & Wren, B. W. (2013). Improved bacterial mutagenesis by high-frequency allele exchange, demonstrated in *Clostridium difficile* and *Streptococcus suis*. *Applied and Environmental Microbiology*, **79**(15):4768-4771.

Foddai, A., Elliott, C. T. & Grant, I. R. (2009). Optimization of a phage amplification assay to permit accurate enumeration of viable *Mycobacterium avium* subsp. paratuberculosis cells. *Applied and Environmental Microbiology,* **75**(12):3896–3902.

Fogg, P. C., Colloms, S., Rosser, S., Stark, M. & Smith, M. C. (2014). New applications for phage integrases. *Journal of Molecular Biology,* **426**(15):2703–2716.

Fortier, L. C. & Moineau, S. (2007). Morphological and genetic diversity of temperate phages in *Clostridium difficile*. *Applied and Environmental Microbiology,* **73**(22):7358-7366.

Fortier, L. C. & Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, **4**(5):354–365.

Foulquier, C., Huang, C. N., Nguyen, N. P., Thiel, A., Wilding-Steel, T., Soula, J., Yoo, M., Ehrenreich, A., Meynial-Salles, I., Liebl, W. & Soucaille, P. (2019). An efficient method for markerless mutant generation by allelic exchange in *Clostridium acetobutylicum* and *Clostridium saccharobutylicum* using suicide vectors. *Biotechnology and Biofuels*, **12**:31.

Freed N. E. (2017). Creation of a dense transposon Insertion library using bacterial conjugation in enterobacterial strains such as *Escherichia coli* or *Shigella flexneri*. *Journal of Visualized Experiments,* **127**: 56216.

Furfaro, L. L., Payne, M. S. & Chang, B. J. (2018). Bacteriophage therapy: Clinical trials and tegulatory hurdles. *Frontiers in Cellular and Infection Microbiology*, **8**:376.

Gao, K. (2016). *ABE fermentation from low cost substrates.* Graduate Program in Chemical and Biochemical Engineering. PhD thesis from The University of Western Ontario. Pp 1-146: https://ir.lib.uwo.ca/etd

Gaspar, M., Kalman, G. & Reczey, K. (2007). Corn fiber as a raw material for hemicellulose and ethanol production. *Process Biochemistry*, **42**(7):1135-1139.

Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balázsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabási, A. L., Oltvai, Z. N. & Osterman, A. L. (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *Journal of Bacteriology,* **185**(19):5673-5684.

Kizziah, J. L., Manning, K. A., Dearborn, A. D. & Dokland, T. (2020). Structure of the host cell recognition and penetration machinery of a *Staphylococcus aureus* bacteriophage. *PLOS Pathogens*,**16**(2):e1008314.

Goh, S., Hussain, H., Chang, B. J., Emmett, W., Riley, T. V. & Mullany, P. (2013). Phage ϕC2 mediates transduction of Tn6215, encoding erythromycin resistance, between *Clostridium difficile* strains. *mBio,* **4**(6):e00840-13.

Goodman, A. L., Wu, A. & Gordon, J. I. (2011). Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nature Protocols*, **6**(9):1969-1980.

Gophna, U. & Brodt, A. (2012). CRISPR/Cas systems in archaea: what array spacers can teach us about parasitism and gene exchange in the 3rd domain of life, *Mobile Genetic Elements*, **2**(1):63-64.

Górski, A., Międzybrodzki, R., Łobocka, M., Głowacka-Rutkowska, A., Bednarek, A., Borysowski, J., Jończyk-Matysiak, E., Łusiak-Szelachowska, M., Weber-Dąbrowska, B., Bagińska, N., Letkiewicz, S., Dąbrowska, K. & Scheres, J. (2018). Phage therapy: What have we learned?. *Viruses*, **10**(6):288.

Guilhot, C., Jackson, M. & Gicquel, B. (2009). Mobile genetic elements and plasmids: Tools for genetic studies (ed) Ratledge, C. & Dale, J. W. In *Mycobacteria: Molecular Biology and Virulence*, John Wiley and Sons pp32.

Hallegatte, S., Bangalore, M., Bonzanigo, L., Fay, M., Kane, T., Narloch, U., Rozenberg, J., Treguer, D. & Vogt-Schilb, A. (2016). *Shock waves: Managing the impacts of climate change on poverty.* Washington DC: World Bank, pp1- 207.

Harada, L. K., Silva, E. C., Campos, W. F., Del Fiol, F. S., Vila, M., Dąbrowska, K., Krylov, V. N. & Balcão, V. M. (2018). Biotechnological applications of bacteriophages: State of the art. *Microbiological Research*, **212-213**:38-58.

Hartmanis, M. G. & Gatenbeck, S. (1984). Intermediary metabolism in *Clostridium acetobutylicum*: levels of enzymes involved in the formation of acetate and butyrate. *Applied and Environmental Microbiology*, **47**:1277-1283

Hatfull, G. F., Cresawn, S. G. & Hendrix, R. W. (2008). Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. *Research in Microbiology*, **159**:332–339.

Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P. & Minton, N. P. (2007). The ClosTron: a universal gene knock-out system for the genus *Clostridium. Journal of Microbiological Methods*, **70:**452–464.

Heap, J. T., Pennington, O. J., Cartman, S. T. & Minton, N. P. (2009a). A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods*, **78**:79–85.

Heap, J. T., Cartman, S. T., Pennington, O. J., Cooksley, C. M., Scott, J. C., Blount, B., Burns, D. A. & Minton, N. P. (2009b). *Development of genetic knock-out systems for Clostridia*. (Eds.) Brüggemann, H. & Gottschalk, G. In: Molecular Biology in the Post-genomic Era *Clostridia*, Norfolk: Caister Academic Press, pp1-230.

Heap J. T., Kuehne, S. A., Ehsaan M., Cartman, S.T., Cooksley C. M., Scott, J. C. & Minton N. P. (2010). The ClosTron: mutagenesis in *Clostridium* refined and streamlined. *Journal of Microbiological Methods*, **80:**49–55.

Heap, J. T., Ehsaan, M., Cooksley, C. M., Ng, Y. K., Cartman, S. T., Winzer, K. & Minton, N. P. (2012). Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research*, **40**(8): e59.

Hina, F. A., Priya, K., Sudakar, B. T., Devabalaji, K. R., Rekha, M., Rajalakshmi, K. & Shilaja, C. (2014). Problems in conventional energy sources and subsequent shift to green energy. *International Journal of Innovative Research in Science, Engineering and Technology*, **3**(1):1618-1624.

Hutchison, C. A., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z. Q., Richter, R. A., Strychalski, E. A., Sun, L. Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., Glass, J. I., Merryman, C., Gibson, D. G. & Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. *Science*, **351**(6280):aad6253.

Hutchison, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O. & Venter, J. C. (1999). Global transposon mutagenesis and a minimal Mycoplasma genome. *Science*, **286**(5447):2165-2169.

Ho, C. R., Shylesh, S. & Bell, A. T. (**2016)**. Mechanism and kinetics of ethanol coupling to butanol over hydroxyapatite. *American Chemical Society Catalysis*, **6**(2):939-948.

Hossain, M. F., Hossain, S. & Uddin, M. J. (2017). Renewable energy: Prospects and trends in Bangladesh, *Renewable and Sustainable Energy Reviews,* **70**:44–49.

Hussain, H. A., Roberts, A. P. & Mullany, P. (2005). Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Delta*erm*) and demonstration that the conjugative transposon Tn*916*DeltaE enters the genome of this strain at multiple sites. *Journal of Medical Microbiology*, **54**:137–141.

IEA (International Energy Agency) (July 2020). *Global Energy Review 2020: The impacts of the Covid-19 crisis on global energy demand and CO2 emissions*, France. https://webstore.iea.org/download/direct/2995

Macrotrends (2020). Interactive charts of West Texas Intermediate (WTI or NYMEX) crude oil prices per barrel back to 1946. https://www.macrotrends.net/1369/crude-oil-price-history-chart
(Retrieved 8th August, 2020).

International Energy Agency (IEA) (2016). *World Energy Outlook 2016 Executive Summary*. 75739 Paris Cedex 15, France. pp 1-13.

International Energy Outlook (IEO) (2021). *International Energy Outlook Narrative with Projections to 2050,* Independent Stastistics & Analysis U.S. Energy Information Administration. https://www.eia.gov/outlooks/ieo/consumption/sub-topic-01.php
(retrieved 9th March, 2022)

Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., Chun-Rong, L., Guenthner, D., Bovee, D., Olson, M. V. & Manoil, C. (2003). Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America,* **100**(24):14339-14344.

Jakočiūnas, T., Jensen, M. K. & Keasling, J. D. (2016). CRISPR/Cas9 advances engineering of microbial cell factories. *Science Direct*, **34**:44-59.

Jennert, K. C. B., Tardif, C., Young, D. I. & Young, M. (2000). Gene transfer to *Clostridium cellulolyticum* ATCC 35319, *Microbiology,* **146**:3071-3080.

Jin, C., Yao, M., Liu, H., Lee, C. F. F. & Ji, J. (2011). Progress in the production and application of n-butanol as a biofuel. *Renewable and Sustainable Energy Reviews*, **15**(8):4080-4106.

Jeanson, S., Floury, J., Gagnaire, V., Lortal, S., & Thierry, A. (2015). Bacterial colonies in solid media and foods: A review on their growth and interactions with the micro-environment. *Frontiers in Microbiology*, **6**:1284.

Johnson R. C. (2015). Site-specific DNA Inversion by Serine Recombinases. *Microbiology Spectrum*, **3**(3):1–36.

Jones, D. T. & Woods, D. R. (1986). Acetone-butanol fermentation revisited. *Microbiology Reviews*, **50**:484-524.

Jones, D., Shirley, M., Wu, X. Y. & Keis, S. (2000). Bacteriophage infections in the industrial acetone butanol (AB) fermentation process. *Journal of Molecular Microbiology and Biotechnology*, **2**:21-26.

Joshi, S. M., Waghmare, J. S., Sonawane, K. D. & Waghmare, S. R. (2015). Bio-ethanol and bio-butanol production from orange peel waste. *Biofuel*, **6**(1-2):55-61.

Karimi, M., Mirshekari, H., Moosavi, B. S. M., Bahrami, S., Moghoofei, M. & Hamblin, M. R. (2016). Bacteriophages and phage-inspired nanocarriers for targeted delivery of therapeutic cargos. *Advance Drug Delivery Review*, **106**:45–62.

Katsura, I. & Hendrix, R. W. (1984). Length determination in bacteriophage lambda tails. *Cell*, **39**(3 Pt 2):691-698.

Kazi, M. I., Schargel, R. D. & Boll, J. M. (2020). Generating transposon insertion libraries in Gram-negative bacteria for high-throughput sequencing. *Journal of Visualized Experiments*, **161**:10.3791/61612.

Kerby, G. P., Gowdy, R. A., Dillon, E. S., Dillon, M. L., Csâky, T. Z., Sharp, D. G. and Beard, J. W. (1949). Purification, pH stability and sedimentation properties of the T7 bacteriophage of *Escherichia coli*. *Journal of Immunology*, **63**(1):93-107.

Kerr, A. L., Jeon, Y. J., Svenson, C. J., Rogers, P. L. & Neilan, B. A. (2011). DNA restriction-modification systems in the ethanologen, *Zymomonas mobilis* ZM4. *Applied Microbiology Biotechnology*, **89**(3):761-769.

Kim, A. Y. & Blaschek, H. P. (1991). Isolation and characterization of a filamentous viruslike particle from *Clostridium acetobutylicum* NCIB 6444. *Journal of Bacteriology,* **173**(2):530-535.

Klippel, A., Mertens, G., Patschinsky, T., & Kahmann, R. (1988). The DNA invertase Gin of phage Mu: formation of a covalent complex with DNA via a phosphoserine at amino acid position 9. *The EMBO journal*, **7**(4):1229–1237.

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*, **9**:14.

Koonin, E. V. & Galperin, M. Y. (2003). *Sequence - evolution - function: Computational approaches in comparative genomics.* Boston: Kluwer Academic, pp1-230.

Kuehne, S. A., Heap, J. H., Cooksley, C. M., Cartman, S. T. & Minton, N. P. (2011). ClosTron-mediated engineering of *Clostridium*. (Ed.) Williams, J. A. In: *Strain Engineering: Protocols*. New York: Springer Protocol Humana Press, pp 389-407.

Kuehne, S. A. & Minton, N. P. (2012). ClosTron-mediated engineering of *Clostridium*. *Bioengineered*, **3**(4):247-254.

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

Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J. & Robertson, H. M. (1999). Hyperactive transposase mutants of the *Himar1* mariner transposon. *Proceedings of the National Academy of Sciences of the United States of America*, **96**(20):11428–11433.

Landy, A. (2015). The λ Integrase site-specific recombination pathway. *Microbiology Spectrum*, **3**(2): MDNA3–2014.

Langridge, G. C., Phan, M., Turner, D. T., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., Wain, J., Parkhill, J. & Turner, A. K. (2009). Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants. *Genome Research*, **19**:2308–2316.

Laureti, L., Matic, I., & Gutierrez, A. (2013). Bacterial responses and genome instability induced by subinhibitory concentrations of antibiotics. *Antibiotics*, **2**(1):100–114.

Lavigne, R., Molineux, I. J. & Kropinski, A.M. (2012). Caudovirales. (Ed). King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J. In: *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, Elsevier, pp29-45.

Lee, S. Y., Park, J. H., Jang, S. H., Nielsen, L. K., Kim, L. K., J. & Jung, K. S. (2008). Fermentative butanol production by *Clostridia. Biotechnology and Bioengineering*, **101**(2):209–228.

Lepage, C., Fayolle, F., Hermann, M. & Vandecasteele, J. P. (1987). Changes in membrane lipid composition of *Clostridium acetobutylicum* during acetone-butanol fermentation: effects of solvents, growth temperature and pH free. *Journal of General Microbiology*, **133**(1):103-110.

Leung, D. W., Lui, A. C., Merilees, H., McBride, B. C. & Smith, M. (1979). A restriction enzyme from *Fusobacterium nucleatum* 4H which recognizes GCNGC. *Nucleic Acids Research*, **6**(1):17–25.

Li, Y. F., Lin, Z. Q., Huang, C., Zhang, Y., Wang, Z., Tang, Y. J., Chen, T. & Zhao, M. X. (2015). Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. *Metabolic Engineering*, **31**:13-21.

Lillehaug, D. (1997). An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. *Journal of Applied Microbiology*, **83**:85–90.

Liu, J. & Mushegian, A. (2004). Displacements of prohead protease genes in the late operons of double-stranded-DNA bacteriophages. *Journal of Bacteriology*, **186**(13):4369-4375.

Liu, H., Bouillaut, L., Sonenshein, A. L. & Merville, S. B. (2013). Use of a mariner-based transposon mutagenesis system to isolate *Clostridium perfringens* mutants deficient in gliding motility. *Journal of Bacteriology*, **195**:629–636.

Liyanage, H., Young, M. & Kashket, E. R. (2000). Butanol tolerance of *Clostridium beijerinckii* NCIMB 8052 associated with down-regulation of *gldA* by antisense RNA, *Journal of Molecular Microbiology & Biotechnology*, **2**(1):87-93.

Łobocka, M. B., Rose, D. J., Plunkett, G., Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M. B. & Blattner, F. R. (2004). The genome of bacteriophage P1. *Journal of Bacteriology*, **186**(21):7032-7068.

Lowe, T. M. & Chan, P. P. (2016). tRNAscan-SE On-line: search and contextual analysis of transfer RNA genes. *Nucleic Acids Research*, **44**:W54-57.

Chatain-Ly, M. H. (2014). The factors affecting effectiveness of treatment in phages therapy. *Frontiers in Microbiology*, **5**:51.

Maddox I. S., Qureshi N. & Roberts-Thomson, K. (1995). Production of acetone-butanol-ethanol from concentrated substrates using *Clostridium acetobutylicum* in an integrated fermentation-product removal process. *Process Biochemistry*, **30**:209–215.

Mahony, J., Tremblay, D. M., Labrie, S. J., Moineau, S. & van Sinderen, D. (2015). Investigating the requirement for calcium during lactococcal phage infection. *International Journal of Food Microbiology*, **201**:47-51.

McCurrie, K. A. L. (2011). Development of transposon based tools for the investigation of virulence factors in *Clostridium difficile*, PhD thesis from University College London Page 72.

Mermelstein, L. D., Welker, N. E., Bernnett, N. G. & Papoutsakis, E. T. (1992). Expression of clones homologous fermentative genes in *Clostridium acetobutylicum* ATCC. *Biotechnology*, **10**:190-195.

Mermelstein, L. D. & Papoutsakis, E. T. (1993). *In vivo* methylation of *Escherichia coli* by the *Bacillus subtilis* phage *phi-3t-I* methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology*, **59**:1077-1081

Mikoulinskaia, G. V., Odinokova, I. V., Zimin, A. A., Lysanskaya, V. Y., Feofanov, S. A. & Stepnaya, O. A. (2009). Identification and characterization of the metal ion-dependent L-alanoyl-D-glutamate peptidase encoded by bacteriophage T5. *FEBS Journal*, **276**(24):7329-7342.

Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M., Liew, F., Kelly, M. L., Sheng, L., Schwarz, K. & Zhang, Y. (2016). A roadmap for gene system development in *Clostridium*. *Anaerobe*, **41**:104–112.

Moon, H. G., Jang, Y., Cho, C., Lee, J., Binkley, R. & Lee, S. Y. (2016). One hundred years of clostridial butanol fermentation minireview – Biotechnology & Synthetic Biology, *FEMS Microbiology Letters*, **363**:fnw001.

Moule, M. G., Hemsley, C. M., Seet, Q., Guerra-Assunção, J. A., Lim, J., Sarkar-Tyson, M., Clark, T. G., Tan, P. B., Titball, R. W., Cuccui, J., & Wren, B. W. (2014). Genome-wide saturation mutagenesis of *Burkholderia pseudomallei* K96243 predicts essential genes and novel targets for antimicrobial development. *mBio*, **5**(1):e00926-13.

Mullany, P., Wilks, M., Lamb, L., Clayton, C., Wren, B. & Tabaqchali, S. (1990). Genetic analysis of a tetracycline resistance element from *Clostridium difficile* and its conjugal transfer to and from *Bacillus subtilis*. *Journal of General Microbiology*, **136**:1343–1349.

Müller, S., Shen, H., Hofmann, D., Schairer, H. U. & Kirby, J. R. (2006). Integration into the phage attachment site, *attB*, impairs multicellular differentiation in *Stigmatella aurantiaca*. *Journal of Bacteriology*, **188**(5):1701–1709.

Muñoz-López, M., & García-Pérez, J. L. (2010). DNA transposons: nature and applications in genomics. *Current Genomics*, **11**(2):115–128.

Murphy, J., Mahony, J., Ainsworth, S., Nauta, A., & van Sinderen, D. (2013). Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. *Applied and Environmental Microbiology*, **79**(24):7547–7555.

Murphy, K. C., Ritchie, J. M., Waldor, M. K., Løbner-Olesen, A. & Marinus, M. G. (2008). Dam methyltransferase is required for stable lysogeny of the Shiga toxin (Stx2)-encoding bacteriophage 933W of enterohemorrhagic *Escherichia coli* O157:H7. *Journal of Bacteriology*, **190**(1):438–441.

Myllyvirta, L. & Thieriot, H. (2020). 11.000 air pollution-related deaths avoided in Europe as coal, oil consumption plummet. Available in: https://energyandcleanair.org/wp/wp-content/uploads/2020/04/CREA-Europe-COVID-impacts.pdf. (Accessed May 2020).

Narimisa, N., Amraei, F., Kalani, B. S., Mohammadzadeh, R. & Jazi, F. M. (2020). Effects of sub-inhibitory concentrations of antibiotics and oxidative stress on the expression of type II toxin-antitoxin system genes in *Klebsiella pneumoniae*, *Journal of Global Antimicrobial Resistance*, **21**:51-56.

NCBI BLASTx 2020 reference protein results page following query with *CA_C1951* using standard settings https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?RID=4SPP180B016&mode=all

Necel, A., Bloch, S., Nejman-Faleńczyk, B., Grabski, M., Topka, G., Dydecka, A., Kosznik-Kwaśnicka, K., Grabowski, L., Jurczak-Kurek, A., Wołkowicz, T., Węgrzyn, G. & Węgrzyn, A. (2020). Characterization of a bacteriophage, vB_Eco4M-7, that effectively infects many *Escherichia coli* O157 strains. *Scientific Reports,* **10:**3743.

Nölling, J., Breton, G., Omelchenko, M. V., Makarova, K. S., Zeng, Q., Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y. I., Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V. & Smith, D. R. (2001). Genome sequence and comparative analysis of the solvent producing bacterium *Clostridium acetobutylicum*. *Journal of Bacteriology*, **183**(16):4823-4838.

O'Brien, R.W. & Morris, J. G. (1971). Oxygen and the growth and metabolism of *Clostridium acetobutylicum. Journal of General Microbiology*, **68**:307-318.

Ohimain, E. I. & Izah, S. C. (2017). A review of biogas production from palm oil mill effluents using different conFigureurations of bioreactors. *Renewable and Sustainable Energy Reviews* **70***:*242–253.

Omorotionmwan, B. B. (2019). *Metabolic engineering of Clostrdidum acetobutylicum and product extension*, PhD thesis submitted to the University of Nottingham, October 2019, pp26.

Onderdonk, A. B. & Garrett, W. S. (2020). Diseases caused by *Clostridium*. (ed) Bennett, J. E., Dolin, R. & Blaser, M. J. In: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, Ninth Edition, Elsevier, pp 246. 2960-2968.e2.

Ounine, K., Petitdemange, H., Raval, G. & Gay, R. (1985). Regulation and butanol inhibition of D-xylose and D-glucose uptake in *Clostridium acetobutylicum*. *Applied and Environmental Microbiology*, **49**(4):874-878.

Papoutsakis, E.T. (2008). Engineering solventogenic *Clostridia*. *Current Opinion in Biotechnology*, **19**:420-429

Pasteur, L. (1862).Quelques rèsultats nouveaux relatifs aux fermentations acètique et butyrique. *Bulletin De La Societe De Chimique De Paris*, **52**:3.

Patakova, P., Maxa, D., Rychtera, M., Linhova, M., Fribert, P., Muzikova, Z., Lipovsky, J., Paulova, L., Pospisil, M., Sebor, G. & Melzoch, K. (2011). Perspectives of biobutanol production and use. (Ed.) Aurelio, M. In: *Biofuel's Engineering Process Technology*, Czeh Republic: InTech, pp 243-266.

Purdy, D., O'Keefe, T. A. T., Elmoore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowki, M. & Minton, N. P. (2002). Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular Microbiology*, **46**:439–452.

Pyne, M. E., Moo-Young, M., Chung, D. A., & Chou, C. P. (2013). Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. *Biotechnology for Biofuels*, **6**:50-70.

Pyne, M. E., Liu, X., Moo-Young, M., Chung, D. A. & Chou, C. P. (2016). Genome-directed analysis of prophage excision, host defence systems, and central fermentative metabolism in *Clostridium pasteurianum*. *Scientific Reports*, **6**:26228.

Quail, M. A., Otto, T. D., Gu, Y., Harris, S. R., Skelly, T. F., McQuillan, J. A., Swerdlow, H. P. & Oyola, S. O. (2011). Optimal enzymes for amplifying sequencing libraries. *Nature Methods*, **9**(1):10-11

Qureshi, N. & Blaschek, H. P. (2001). Recent advances in ABE fermentation: hyper-butanol producing *Clostridium beijerinckii* BA101. *Journal of Industrial Microbiology and Biotechnology*, **27**:287-291.

Rahnama, N., Foo, H. L., Rahman, N. A. A., Ariff, A. & Shaff, U. K. M. (2014). Saccharification of rice straw by cellulase from a local *Trichoderma harzianum* SNRS3 for biobutanol production. *BMC Biotechnology*, **14:**103.

Rao, V. B. & Feiss, M. (2008). The bacteriophage DNA packaging motor. *Annual Review of Genetics*, **42**:647–681.

Ravin, N. V. (2015). Replication and maintenance of linear phage-plasmid N15. *Microbiology Spectrum*, **3**(1):PLAS-0032-2014.

REBASE genome (2020) http://rebase.neb.com/rebase/rebase.html

Reid, S. J., Allcock, E. R., Jones, D. T. & Woods, D. R. (1983). Transformation of *Clostridium acetobutylicum* protoplasts with bacteriophage DNA. *Applied and Environmental Microbiology*, **45**(1):305-307.

Ripp, S. & Miller, R. V. (1998). Dynamics of the pseudolysogenic response in slowly growing cells of *Pseudomonas aeruginosa* free. *Microbiology*, **144**(144):2225-2232.

Ritacco, C. J., Kamtekar, S., Wang, J. & Steitz, T. A. (2013). Crystal structure of an intermediate of rotating dimers within the synaptic tetramer of the G-segment invertase, *Nucleic Acids Research*, **41**(4):2673–2682.

Roberts, R., Vincze, T., Posfai, J. & Macelis, D. (2007). REBASE–enzymes and genes for DNA restriction and modification. *Nucleic Acids Research*, **35**: D269.

Sabathé, F., Croux, C., Cornillot, E. & Soucaille P. (2002). *amyP*, a reporter gene to study strain degeneration in *Clostridium acetobutylicum* ATCC 824. *FEMS Microbiology Letters*, **210**(1):93-98.

Salameh, Z. (2014). *Renewable Energy System Design*. Amsterdam: Academic Press, Elsevier Inc. pp1-369.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Samuelson, J., Xu, S. & O'Loane, D. (2006). New England Biolabs I. Method for cloning and expression of *Acu*I restriction endonuclease and *Acu*I methylase in *E. coli*. US patent No.7,011,966.

Sauer, M. (2016). Industrial production of acetone and butanol by fermentation—100 years later. *FEMS Microbiology Letters*, **363**(13):fnw134.

Schleifer, K. (2009). Phylum XIII. Firmicutes Gibbons and Murray 1978, 5 (Firmacutes [sic] Gibbons and Murray 1978, 5). (Ed). De Vos, P., Garrity, G. M., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A., Schleifer, K. & Whitman, W. B. In: *Bergey's Manual of Systematic Baceriology*, Springer New York, pp 736-827.

Sheng, L., Kovács, K., Winzer, K., Zhang, Y., & Minton, N. P. (2017). Development and implementation of rapid metabolic engineering tools for chemical and fuel production in *Geobacillus thermoglucosidasius* NCIMB 11955. *Biotechnology for Biofuels*, **10**:5.

Shoham, Y., Lamed, R. & Bayer, E. A. (1999). The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends in Microbiol*ogy, **7:**275–281.

Shabbir, M. A. B., Shabbir, M. Z., Wu, Q., Mahmood, S., Sajid, A., Maan, M. K., Ahmed, S., Naveed, U., Hao, H. & Yuan, Z. (2019). CRISPR-cas system: biological

function in microbes and its use to treat antimicrobial resistant pathogens. *Annals of Clinical Microbiology and Antimicrobials*, **18**(21).

Siegrist, M. S. & Rubin, E. J. (2009). Phage transposon Mutagenesis. (Eds).Parish T. & Brown, A. In: *Mycobacteria Protocols. Methods in Molecular Biology*, Human Press, Totawa, NJ. Vol 465.

Silverstein, R. A., Chen, Y., Sharma-Shivappa, R. R., Boyette, M. D. & Osborne, J. (2007). A comparison of chemical pretreatment methods for improving saccharification of cotton stalks. *Bioresources Technology*, **98**(16):3000-3011.

Smith, K. R., Bruce, N. G., Balakrishnan, K., Adair-Rohani, H., Balmes, J., Chafe, Z., Dherani, M., Hosgood, H. D., Mehta, S., Pope, D. & Rehfuess, E. (2014). HAP CRA risk expert group. Millions dead: how do we know and what does it mean? Methods used in the comparative risk assessment of household air pollution. *Annual Review of Public Health*, **35**:185–206.

Sourek, J. (1974). Long-term preservation by freeze-drying of pathogenic bacteria of the Czechoslovak national collection of type cultures. *International Journal of Systematic Bacteriology*, **24**(3): 358-365.

Sreekumar, S., Baer, Z. C., Pazhamalai, A., Gunbas, G., Grippo, A., Blanch, H. W., Clark, D. S. & Toste, D. F. (2015). Production of an acetone-butanol-ethanol mixture from *Clostridium acetobutylicum* and its conversion to high-value biofuels, *Nature Protocols*, **10**(3):528-537.

Stark, W. (2014). The serine recombinases. *Microbiology Spectrum: American Society for Microbiology Press*, **2**(6):1-16.

Steczkiewicz, K., Prestel, E., Bidnenko, E. & Szczepankowska, A. K. (2021). Expanding diversity of firmicutes single-strand annealing proteins: A putative role of bacteriophage-host arms race. *Frontiers in microbiology*, **12**:644622.

Stevens, R. H., Ektefaie, M. R. & Fouts, D. E. (2011). The annotated complete DNA sequence of *Enterococcus faecalis* bacteriophage φEf11and its comparison with all available phage and predicted prophage genomes. *FEMS Microbiology Letters*, **317**:9–26.

Stoll, S. M., Ginsburg, D. S. & Calos, M. P. (2002). Phage TP901-1 site-specific integrase functions in human cells. *Journal of Bacteriology*, **184**(13):3657-3663.

Sun, S., Gao, S., Kondabagil, K., Xiang, Y., Rossmann, M. G. & Rao, V. B. (2012). Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proceedings of the National Academy of Sciences*, **109**(3):817-822.

Suzuki, H. (2012). Host-mimicking strategy of DNA methylation for improved transformation. (Ed) Dricu, A. In: *Methylation -from DNA, RNA and Histones to Diseases and Treatment*, Croatia: Intech, pp1-301, pp 219-224.

Swanson, M. M., Reavy, B., Makarova, K. S., Cock, P. J., Hopkins, D. W., Torrance, L., Koonin, E. V. & Taliansky, M. (2012). Novel bacteriophages containing a genome of another bacteriophage within their genomes. *PlosOne*, **7**(7):1-12.

Taj, M., Ling, J., Bing, L., Zhang, Q., Hassani, I., Hassani, T., Samreen, Z. & Yunlin, W. (2014). Effect of dilution, temperature and pH on the lysis activity of T4 phage against *E. coli* BL21. *Journal of Animal and Plant Sciences*, **24**:2014-1252.

Takeda, Y., Folkmanis, A. & Echols, H. (1977). Cro regulatory protein specified by bacteriophage lambda. Structure, DNA-binding and repression of RNA synthesis. *Journal of Biology and Chemistry*, **252**(17):6177–6183.

Tao, L., Wu, X. & Sun, B. (2010). Alternative sigma factor sigmaH modulates prophage integration and excision in *Staphylococcus aureus*. *PLoS Pathogen*, **6**:e1000888.

Tao, P., Wu, X. & Rao, V. (2018). Unexpected evolutionary benefit to phages imparted by bacterial CRISPR-Cas9. *Science Advances*, **4**(2):eaar4134.

Tashiro, Y. & Sonomoto, K. (2010). Advances in butanol production by *Clostridia*. (ed.) Vilas A. M. In: *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology (Microbiology Book Series, Volume 2)***,** Spain: Formatex Research Center, pp 1383-1394.

Tomas, C. A., Welker, N. E. & Papoutsakis, E. T. (2003). Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Applied and Environmental Microbiology*, **69**(8):4951–4965.

Tomas, C. A., Beamish, J. & Papoutsakis, E. T. (2004). Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *Journal of Bacteriology*, **186**(7):2006-2018.

Tomimatsu, K., Kokura, K., Nishida, T., Yoshimura, Y., Kazuki, Y., Narita, M., Oshimura, M. & Ohbayashi, T. (2017). Multiple expression cassette exchange via TP901-1, R4, and Bxb1 integrase systems on a mouse artificial chromosome. *FEBS open bio*, **7**(3):306–317.

Tominaga, A. & Enomoto, M. (1986). Magnesium-dependent plaque formation by bacteriophage P1*cin*C(-) on *Escherichia coli C* and *Shigella sonnei*. *Virology*, **155**(1): 284-288.

UK Standards for Microbiology Investigations (2015). *Identification of Clostridium species, the standards unit, microbiology services, Public Health England bacteriology –Identification*, Issue date 25th June, 2015 ID **8**(4):1-27.

Urrutia-Pereira, M., Mello-da-Silva, C. A., & Solé, D. (2020). COVID-19 and air pollution: A dangerous association? *Allergologia et Immunopathologia*, S0301-0546(20)30109-9. Advance online publication.

Van der Lelie, D., Taghavi, S., McCorkle, S. M., Li, L. L., Malfatti, S. A., Monteleone, D., Donohoe, B. S., Ding, S., Adney, W. S., Himmel, M. E. & Tringe, S. G. (2012). The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PLoS ONE*, **7**(5):e36740.

Van Opijnen, T. & Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology*, 11(7):435–442.

Voutev, R. & Mann, R. S. (2019). TP901-1 Phage recombinase facilitates genome engineering in *Drosophila melanogaster*. *G3 (Bethesda)*, **9**(4):983–986.

Wagner, P. L. & Waldor, M. K. (2002). Bacteriophage control of bacterial virulence. *Infection and Immunity*, **70**(8) 3985–3993.

Wang, H., Smith, M. C. M. & Mullany, P. *(*2006*).* The conjugative transposon Tn*5397* has a strong preference for integration into its *Clostridium difficile* target site. *Journal of Bacteriology*, **188***:*4871–4878.

Wang, Z. (2012). Photonanocatalyst aided alkaline pretreatment and Raman spectroscopic characterization of corn stover biomass. Graduate Theses and Dissertation submitted to Iowa State University, Paper 12711.

Wasels, F., Jean-Marie, J., Collas, F., López-Contreras, A. M. & Ferreira, N. L. (2017). A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *Journal of Microbiology Methods*, **140**:5-11.

Weinberger, A. D. & Gilmore, M. S. (2012). CRISPR-Cas: To take up DNA or not—that is the question. *Cell Host & Microbe*, **12(**2):125-126.

Wells, C. L. & Wilkins, T. D. (1996). *Clostridia*: Sporeforming anaerobic bacilli. (ed.). Baron S. In: *Medical Microbiology*. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston.

Westra, E. R. & Levin B. R. (2020). It is unclear how important CRISPR-Cas systems are for protecting natural populations of bacteria against infections by mobile genetic elements. *Proceedings of National Academy of Sciences of the United States of America (PNAS)*, **117**(45):27777-27785.

Wolfe, A., Phipps, K. & Weitao, T. (2014). Viral and cellular SOS-regulated motor proteins: dsDNA translocation mechanisms with divergent functions. *Cell & Bioscience*, **4**:31.

Wong, Y. C., Ghany, M.,A. E., Naeem, R., Lee, K. W., Tan, Y. C., Pain, A. & Nathan, S. (2016). Candidate essential genes in *Burkholderia cenocepacia* J2315 Identified by genome-wide TraDIS. *Frontiers in Microbiology*, **7**:1288.

Woods, C., Humphreys, C. M., Tomi-Andrino, C., Henstra, A. M., Köpke, M., Simpson, S. D., Winzer, K. & Minton, N. P. (2022). Required gene set for autotrophic

growth of *Clostridium autoethanogenum*. *Applied Environmental Microbiology*, Mar 14:e0247921

Wu, Y. D., Xue, C., Chen, L. J., Wan, H. H. & Bai, F. W. (2015). Transcriptional analysis of micronutrient zinc-associated response for enhanced carbohydrate utilization and earlier solventogenesis in *Clostridium acetobutylicum*. *Scientific Reports*, **5**:16598.

Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., Van Nostrand, J. D., He, Z. & Zhou, J. (2015). Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. *Applied and Environmental Microbiology*, **81**(13):4423-4431.

Yang, H., Ma, Y., Wang, Y., Yang, H., Shen, W. & Chen, X. (2014). Transcription regulation mechanisms of bacteriophages. *Recent Advances and Future Prospects Bioengineered*, **5**(5): 300–304.

Yang, L., Bao, G., Zhu, Y., Dong, H., Zhang, Y. & Li, Y. (2013). Protein cell & discovery of a novel gene involved in autolysis of *Clostridium* cells. *Protein & Cell*, **4**(6):467–474.

Yang, M., Wang, J., Nan, Y., Zhang, J., Li, L., Liu, G., Vepsäläinen, J., Kuittinend, S. & Pappinend, A. (2019). Effect of salts formed by neutralization for the enzymatic hydrolysis of cellulose and acetone–butanol–ethanol fermentation. *Royal Society of Chemistry Advance,* **9:**33755-33760.

Yang, Y, Zhang, H., Lang, N., Zhang, L., Chai, C., He, H., Jiang, W. & Gu Y. (2020). The small RNA sr8384 is a crucial regulator of cell growth in solventogenic *Clostridia*. *Applied and Environemental Microbiology*, **86**:13.

Yoo, M., Nguyen, N. P. & Soucaille, P. (2020). Trends in systems biology for the analysis and engineering of *Clostridium acetobutylicum* metabolism. *Trends in Microbiology,* **28**(2):118-140.

Yoon, B. H. & Chang, H. (2015). Genomic annotation for the temperate phage EFC-1, isolated from *Enterococcus faecalis* KBL101. *Archive of Virology,* **160:**601–604.

Zhang, Y., Grosse-Honebrink, A. & Minton, N. P. (2015). A universal mariner transposon system for forward genetic studies in the genus *Clostridium*. *PLoS One,* **10** (4):e0122411.

Zhang, Y., Xu, S., Chai, C., Yang, S., Jiang, W., Minton, N. P. & Gu, Y. (2016). Development of an inducible transposon system for efficient random mutagenesis in *Clostridium acetobutylicum*. *FEMS Microbiology Letters*, **363**:fnw065.

Zhao, X., Condruz, S., Chen, J. & Jolicoeur, M. A (2016). Quantitative metabolomics study of high sodium response in *Clostridium acetobutylicum* ATCC 824 acetone-butanol-ethanol (ABE) fermentation. *Scientific Reports,* **6**:28307

Appendix A: Supplementary Information for Chapter 4

Table 7.1: Canditates for *C. acetobutylicum* ATCC 824 essential gene list. These genes had no insertions in them (insertion index < 0.0001). Genes designated CA_C are from the chromosome while genes designated CA_P are from pSOL1.

locus_tag	gene_name	function
CA_C0004	*recF*	RecF ABC family ATPase
CA_C0005	*CA_C0005*	Uncharacterized small conserved protein ortholog of YAAB B.subtilis
CA_Ct001	*CA_Ct001*	Aminoacyl-tRNA biosynthesis (tRNA - methionine)
CA_Ct002	*CA_Ct002*	Aminoacyl-tRNA biosynthesis (tRNA - alanine)
CA_Ct003	*CA_Ct003*	Aminoacyl-tRNA biosynthesis (tRNA - serine)
CA_Ct004	*CA_Ct004*	Aminoacyl-tRNA biosynthesis (tRNA - serine)
CA_Ct005	*CA_Ct005*	Aminoacyl-tRNA biosynthesis (tRNA - arginine)
CA_Ct006	*CA_Ct006*	Aminoacyl-tRNA biosynthesis (tRNA - arginine)
CA_C0038	*CA_C0038*	Hypothetical protein
CA_C0041	*CA_C0041*	Uncharacterized small conserved protein homolog of yfjA/yukE B.subtilis
CA_C0054	*CA_C0054*	Hypothetical protein
CA_C0058	*CA_C0058*	Hypothetical protein
CA_C0064	*CA_C0064*	Hypothetical protein
CA_C0079	*CA_C0079*	Hypothetical protein
CA_Ct007	*CA_Ct007*	Aminoacyl-tRNA biosynthesis (tRNA - serine)
CA_Ct008	*CA_Ct008*	Aminoacyl-tRNA biosynthesis (tRNA - serine)
CA_Cs001	*CA_Cs001*	scrRNA
CA_C0126	*CA_C0126*	Uncharacterized conserved protein YbaB family
CA_C0127	*recR*	Recombination protein (recR)
CA_C0128	*CA_C0128*	Hypothetical protein
CA_C0131	*CA_C0131*	Hypothetical protein
CA_Cr009	*CA_Cr009*	5S ribosomal RNA
CA_Ct009	*CA_Ct009*	Aminoacyl-tRNA biosynthesis (tRNA - lysine)
CA_C0158	*glmS*	Glucoseamine-fructose-6-phosphate aminotransferase (gene glmS)
CA_C0172	*CA_C0172*	Predicted membrane protein
CA_Cr015	*CA_Cr015*	5Se Ribosomal RNA
CA_Ct010	*CA_Ct010*	Aminoacyl-tRNA biosynthesis (tRNA - phenylalanine)
CA_C0251	*CA_C0251*	Uncharacterized Fe-S protein (fragment)
CA_Cr018	*CA_Cr018*	
CA_Ct011	*CA_Ct011*	Aminoacyl-tRNA biosynthesis (tRNA - lysine)
CA_C0276	*CA_C0276*	Uncharacterized possible metal-binding protein
CA_Cr024	*CA_Cr024*	5Sh Ribosomal RNA
CA_Cr027	*CA_Cr027*	5Si Ribosomal RNA
CA_Ct012	*CA_Ct012*	Aminoacyl-tRNA biosynthesis (tRNA - phenylalanine)
CA_Ct013	*CA_Ct013*	Aminoacyl-tRNA biosynthesis (tRNA - cysteine)
CA_C0292	*CA_C0292*	Hypothetical protein
CA_C0300	*CA_C0300*	DNA-polymerase III gamma subunit
CA_C0301	*CA_C0301*	Signal peptidase II (PSP1) homolog YAAT B.subtilis ortholog
CA_C0303	*CA_C0303*	Ferredoxin

CA_C0308	*CA_C0308*	Cell wall-associated hydrolase
CA_C0310	*abrB*	Regulators of stationary/sporulation gene expression abrB B.subtilis ortholog
CA_C0334	*CA_C0334*	Hypothetical protein CF-21 family
CA_C0340	*CA_C0340*	Hypothetical protein
CA_C0341	*CA_C0341*	Predicted membrane protein
CA_C0344	*CA_C0344*	Hypothetical protein
CA_C0361	*kduD*	2 deoxy-D-gluconate 3-dehydrogenase
CA_C0373	*CA_C0373*	ABC transporter ATP binding protein
CA_C0387	*CA_C0387*	Hypothetical protein
CA_C0389	*CA_C0389*	Hypothetical protein
CA_C0396	*CA_C0396*	3-hexulose-6-phosphate synthase (D-arabino-3-hexulose-6-phosphate formaldehyde lyase)
CA_C0400	*CA_C0400*	Hypothetical protein CF-19 family
CA_C0401	*CA_C0401*	Hypothetical protein - CF-24 family
CA_C0410	*CA_C0410*	Uncharacterized small conserved protein - homolog of YUKE/YFJA
CA_C0411	*CA_C0411*	Uncharacterized small conserved protein - homolog of YUKE/YFJA
CA_C0434	*CA_C0434*	Putative enzyme of deoxyxylulose pathway (terpenoid biosinthesis) YgbB family similar to yacN B.subtilis
CA_C0438	*CA_C0438*	Predicted metal-binding protein
CA_C0439	*CA_C0439*	Hypothetical protein
CA_C0450	*CA_C0450*	Response regulator (CheY-like and HTH domains)
CA_C0455	*CA_C0455*	Hypothetical protein
CA_C0480	*nrdD*	Oxygen-sensitive ribonucleoside-triphosphate reductase nrdD
CA_C0501	*bacA*	Bacitracin resistance protein (bacA)
CA_C0511	*CA_C0511*	Predicted P-loop containing kinase similar to B.subtilis yvcJ
CA_C0512	*CA_C0512*	Uncharacterized conserved protein YbhK/UPF0052 family
CA_C0521	*CA_C0521*	Hypothetical protein
CA_C0545	*CA_C0545*	Uncharacterized small protein homolog of Treponema (6136517)
CA_C0549	*CA_C0549*	Predicted transcriptional regulator
CA_C0589	*birA*	Biotin-(acetyl-CoA carboxylase)ligase
CA_C0593	*ribH*	Riboflavin synthase beta-chain
CA_C0613	*CA_C0613*	Spore coat protein F
CA_C0614	*CA_C0614*	Spore coat protein F
CA_C0626	*trpS*	Tryptophan-tRNA synthetase trpS
CA_C0646	*leuS*	Leucyl-tRNA synthetase
CA_Ct014	*CA_Ct014*	Aminoacyl-tRNA biosynthesis (tRNA - glutamine)
CA_C0655	*CA_C0655*	Hypothetical protein CF-26 family
CA_C0669	*CA_C0669*	Hypothetical protein
CA_C0677	*CA_C0677*	uncharacterized conserved protein YdjX/UPF0043 family
CA_C0678	*CA_C0678*	CBS domains
CA_C0681	*nrgB*	Nitrogen regulatory protein PII gene nrgB
CA_C0684	*CA_C0684*	CBS domains
CA_Ct015	*CA_Ct015*	Aminoacyl-tRNA biosynthesis (tRNA - leucine)
CA_C0708	*CA_C0708*	Putative transcriptional regulator
CA_C0749	*CA_C0749*	HD hydrolase superfamily protein
CA_C0763	*CA_C0763*	Transcriptional regulator MarR family

CA_C0764	*CA_C0764*	NADPH-dependent glutamate synthase beta chain
CA_C0817	*CA_C0817*	Hypothetical protein
CA_C0846	*CA_C0846*	Hypothetical proteinCF-29 family
CA_C0891	*CA_C0891*	Hypothetical protein
CA_C0892	*CA_C0892*	DAHP synthase related protein
CA_C0893	*CA_C0893*	Prephenate dehydrogenase
CA_C0894	*aroB*	3-dehydroquinate synthetase
CA_C0895	*aroA*	5-enolpyruvylshikimate-3-phosphate synthase
CA_C0898	*aroK*	Shikimate kinase
CA_C0899	*CA_C0899*	3-dehydroquinate dehydratase II
CA_C0900	*CA_C0900*	Possible Zn-finger containing protein
CA_C0901	*CA_C0901*	Probably O-sialoglycoprotein endopeptidase
CA_C0932	*CA_C0932*	Hypothetical protein
CA_C0942	*hisI_1*	Phosphoribosyl-AMP cyclohydrolase
CA_C0945	*ssb*	Single-stranded DNA-binding protein
CA_C0952	*CA_C0952*	Hypothetical protein
CA_C0956	*CA_C0956*	Hypothetical protein
CA_C0961	*CA_C0961*	Cobyric acid synthase CobQ
CA_Cr028	*CA_Cr028*	16Sj ribosomal RNA
CA_Ct016	*CA_Ct016*	Aminoacyl-tRNA biosynthesis (tRNA - asparagine)
CA_C0965	*CA_C0965*	1-acyl-sn-glycerol-3-phosphate acyltransferase
CA_C0971	*citB*	Aconitase A
CA_C0977	*asnC*	Transcriptional regulator Lrp family
CA_C0983	*CA_C0983*	Hypothetical protein
CA_C0989	*CA_C0989*	General secretion family related protein
CA_C0990	*gltX*	Glutamyl-tRNA synthetase
CA_C0994	*CA_C0994*	Predicted membrane protein
CA_C1030	*CA_C1030*	FeoA-like protein involved in iron transport
CA_C1032	*CA_C1032*	Predicted transcriptional regulator
CA_C1034	*CA_C1034*	Hypothetical protein
CA_C1035	*CA_C1035*	Hypothetical protein
CA_C1041	*argS*	Arginyl-tRNA synthetase
CA_C1051	*CA_C1051*	Uncharacterized conserved protein (possible membrane)
CA_C1056	*CA_C1056*	Hypothetical protein
CA_C1071	*CA_C1071*	Transcriptional regulator AcrR family
CA_C1089	*CA_C1089*	Serine kinase/phosphatase HPr
CA_C1093	*CA_C1093*	Hypothetical secreted protein
CA_C1099	*CA_C1099*	P-loop kinase (uridine kinase family)
CA_Ct017	*CA_Ct017*	Aminoacyl-tRNA biosynthesis (tRNA - leucine)
CA_C1104	*CA_C1104*	Hypothetical protein
CA_C1105	*CA_C1105*	Hypothetical protein
CA_C1111	*CA_C1111*	Hypothetical protein CF-10 family
CA_C1112	*CA_C1112*	Hypothetical protein CF-11 family
CA_C1138	*CA_C1138*	Hypothetical protein
CA_C1156	*CA_C1156*	Hypothetical protein
CA_C1157	*CA_C1157*	Hypothetical protein

CA_C1171	*CA_C1171*	Hypothetical protein
CA_C1172	*CA_C1172*	Predicted integrase of XerC/XerD family diverged
CA_C1174	*CA_C1174*	Hypothetical protein
CA_C1175	*CA_C1175*	Hypothetical protein CF-35 family
CA_C1179	*CA_C1179*	Hypothetical protein
CA_C1196	*CA_C1196*	Hypothetical protein
CA_C1211	*CA_C1211*	Hypothetical protein
CA_C1217	*CA_C1217*	Guanylate kinase (P-loop type)
CA_C1236	*CA_C1236*	Hypothetical protein
CA_C1240	*maf*	Maf protein ortholog putative inhibitor of septum formation
CA_C1242	*mreB*	mreB
CA_C1244	*CA_C1244*	Predicted membrane protein
CA_C1250	*CA_C1250*	Hypothetical protein
CA_C1251	*rodA*	Cell division protein rodA/ftsW/spoVE family
CA_C1257	*rpl21*	Ribosomal protein L21
CA_C1258	*CA_C1258*	Hypothetical protein
CA_C1259	*rpl27*	Ribosomal protein L27
CA_C1260	*obg*	SPO0B-associated GTPase obg
CA_C1273	*CA_C1273*	Uncharacterized protein YQEN B.subtilis homolog
CA_C1274	*rpsT*	Ribosomal protein S20
CA_C1281	*grpE*	Molecular chaperone GrpE
CA_C1283	*dnaJ*	Molecular chaperones DnaJ (HSP40 family)
CA_C1288	*rpsU*	Ribosomal protein S21
CA_C1289	*CA_C1289*	Uncharacterized conserved protein YqeY B.subtilis ortholog
CA_C1293	*CA_C1293*	Uncharacterized conserved predicted metal-binding protein
CA_C1299	*dnaG*	DNA primase YQFG ortholog DNAG
CA_C1300	*sigA*	RNA polymerase sigma factor RPOD
CA_C1306	*CA_C1306*	Hypothetical protein
CA_C1324	*CA_C1324*	Uncharacterized predected metal-binding protein
CA_C1327	*CA_C1327*	Hypothetical protein
CA_C1331	*CA_C1331*	Short-chain alcohol dehydrogenase
CA_C1337	*CA_C1337*	Spore coat protein COTJB
CA_C1350	*CA_C1350*	Hypothetical protein
CA_C1398	*CA_C1398*	Uncharacterized protein from YqgV/UPF0045 family
CA_C1409	*CA_C1409*	Predicted membrane protein
CA_C1413	*CA_C1413*	Similar to C-terminal fragment of toxic anion resistance protein terA
CA_C1418	*CA_C1418*	Nitroreductase family protein
CA_C1428	*CA_C1428*	Hypothetical protein
CA_C1439	*CA_C1439*	Uncharacterized conserved protein
CA_C1442	*CA_C1442*	Hypothetical protein
CA_C1446	*CA_C1446*	Hypothetical protein
CA_C1449	*CA_C1449*	Hypothetical protein
CA_C1463	*CA_C1463*	Predicted transcriptional regulator
CA_C1477	*CA_C1477*	Hypothetical protein
CA_C1483	*CA_C1483*	Predicted transcriptional regulator
CA_C1486	*CA_C1486*	Uncharacterized protein ortholog of YKUS B.subtilis

CA_C1498	*CA_C1498*	Predicted flavoprotein
CA_C1500	*CA_C1500*	Predicted membrane protein
CA_C1545	*deoC*	Deoxyribose-phosphate aldolase
CA_C1546	*deoA*	Pyrimidine-nucleoside phosphorylase
CA_C1547	*trxA*	Thioredoxin trx
CA_C1551	*CA_C1551*	Nitroreductase family protein
CA_C1583	*CA_C1583*	Predicted P-loop ATPase
CA_C1629	*CA_C1629*	Putative intracellular protease/amidase ThiJ family
CA_C1636	*CA_C1636*	Uncharacterized protein homolog of B. firmus (2654481)
CA_C1638	*CA_C1638*	Hypothetical protein CF-37 family(almost identical)
CA_C1641	*CA_C1641*	Ankyrin repeats containing protein
CA_C1659	*CA_C1659*	Predicted membrane protein
CA_C1663	*CA_C1663*	Hypothetical protein
CA_C1679	*CA_C1679*	Hypothetical protein
CA_C1680	*CA_C1680*	Predicted endonuclease involved in recombination
CA_C1681	*CA_C1681*	Hypothetical protein
CA_C1684	*CA_C1684*	TYPA/BIPA type GTPase
CA_C1691	*CA_C1691*	Predicted glycosyltransferase
CA_C1693	*ftsZ*	Cell division GTPase FtsZ
CA_C1711	*CA_C1711*	Predicted GTPase
CA_C1712	*gpsA*	Glycerol 3-phosphate dehydrogenase
CA_C1717	*CA_C1717*	Uncharacterized conserved protein YLOA ortholog
CA_C1718	*CA_C1718*	Guanylate kinase YLOD B.subtilis ortholog
CA_C1719	*rpoZ*	RNA polymerase-associated protein RpoZ omega subunit (YLOH B.subtilis ortholog)
CA_C1720	*CA_C1720*	Flavoprotein involved in panthothenate metabolism YLOI B.subtilis ortholog
CA_C1723	*fmt*	Methionyl-tRNA formyltransferase
CA_C1728	*CA_C1728*	Serine/threonine protein kinase Pkn2 family (YLOP B.subtilis ortholog)
CA_C1731	*CA_C1731*	Predicted nucleotide-binding protein YLOS B.subtilis ortholog
CA_C1733	*rpmB*	Ribosomal protein L28
CA_C1738	*kdtB*	Phosphopantetheine adenylyltransferase
CA_C1743	*askA*	Acetate kinase
CA_C1744	*CA_C1744*	Predicted Zn-finger-like protein possible nucleic acid binding
CA_C1745	*rpmF*	Ribosomal protein L32
CA_C1747	*acpA*	Acyl carrier protein ACP
CA_C1750	*CA_C1750*	SpoVS-related protein
CA_C1754	*ffh*	Signal recognition particle GTPase Ffh
CA_C1755	*rpsP*	Ribosomal protein S16
CA_C1758	*trmD*	TRNA-(guanine-N1)-methyltransferase TrmD
CA_C1759	*rplS*	Ribosomal protein L19
CA_C1787	*rpsB*	Ribosomal protein S2
CA_C1790	*frr*	Ribosome recycling factor
CA_C1791	*CA_C1791*	Undecaprenyl pyrophosphate synthase
CA_C1795	*CA_C1795*	1-deoxy-D-xylulose 5-phosphate reductoisomerase
CA_C1798	*CA_C1798*	Uncharacterized protein YhbC family

CA_C1799	*nusA*	Transcription terminator NusA
CA_C1800	*CA_C1800*	Uncharacterized conserved protein YLXR B.subtilis homolog
CA_C1802	*infB*	Translation IF2 GTPase
CA_C1808	*pnpA*	Polyribonucleotide nucleotidyltransferase
CA_C1811	*CA_C1811*	Periplasmic serine protease YMFB B.subtilis ortholog
CA_C1814	*pgsA*	Phosphatidylglycerophosphate synthase
CA_C1815	*recA*	RECA recombinase ATPase
CA_C1816	*CA_C1816*	HD superfamily hydrolase YMDA B.subtilis ortholog
CA_C1817	*CA_C1817*	Stage V sporulation protein spoVS
CA_C1820	*CA_C1820*	Phosphocarrier Protein (Hpr)
CA_C1830	*CA_C1830*	Hypothetical protein
CA_C1831	*CA_C1831*	Protein containing Zn-finger domain
CA_C1834	*CA_C1834*	Host factor I protein Hfq
CA_C1844	*CA_C1844*	Hypothetical protein
CA_C1846	*motA*	Flagellar motor component MotA
CA_C1861	*CA_C1861*	Fragment flavodoxin oxidoreductase
CA_C1864	*CA_C1864*	Hypothetical protein
CA_C1866	*CA_C1866*	Putative 4-Cys ferredoxin
CA_C1967	*CA_C1967*	Hypothetical protein
CA_C1983	*CA_C1983*	Hypothetical protein
CA_C1985	*CA_C1985*	Hypothetical protein
CA_C1998	*CA_C1998*	ABC-type transport system ATPase component
CA_C2029	*CA_C2029*	Hypothetical protein
CA_C2034	*CA_C2034*	Hypothetical protein
CA_C2043	*CA_C2043*	Hypothetical protein
CA_C2055	*CA_C2055*	Predicted transcriptional regulator
CA_C2066	*CA_C2066*	Integrase/recombinase XerD family
CA_C2080	*CA_C2080*	Predicted geranylgeranyl pyrophosphate synthase
CA_C2081	*xseB*	Exonuclease VII small subunit
CA_C2083	*folD*	Tetrahydrofolate dehydrogenase/cyclohydrolase FolD
CA_C2085	*CA_C2085*	Uncharacterized protein from alkaline shock protein family YQHY B.subtilis ortholog
CA_C2094	*efp*	Translation elongation factor P
CA_C2099	*CA_C2099*	Hypothetical protein
CA_C2117	*pfs*	Nucleoside phosphorylase
CA_C2120	*CA_C2120*	Uncharacterized protein YlmF B.subtilis ortholog
CA_C2125	*divIB*	Cell division septal protein divIB/FtsQ
CA_C2126	*spoVE*	Stage V sporulation protein E FtsW/MrdB/SpoVE family
CA_C2127	*mraY*	Phospho-N-acetylmuramoyl-pentapeptide transferase MraY
CA_C2128	*murF*	UDP-N-acetylmuramyl pentapeptide synthase
CA_C2136	*CA_C2136*	Hypothetical protein
CA_C2138	*CA_C2138*	Exopolyphosphatase
CA_C2141	*CA_C2141*	Hypothetical protein
CA_C2144	*CA_C2144*	Uncharacterized protein YPFA B.subtilis ortholog
CA_C2149	*fliQ*	Flagellar biosynthesis protein FliQ

CA_C2153	*flbD*	Flagellar protein flbD
CA_C2206	*fliS*	Flagellar protein FliS
CA_C2213	*CA_C2213*	Hypothetical protein
CA_C2218	*cheY*	Chemotaxis signal receiving protein CheY
CA_C2227	*CA_C2227*	Phosphoserine phosphatase family enzyme
CA_C2231	*murG*	Undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc transferase MurG
CA_C2233	*CA_C2233*	TRNA 5-methylaminomethyl-2-thiouridylate)-methyltransferase PP-loop ATPase
CA_C2236	*CA_C2236*	Uncharacterized conserved protein of YjeB/RRF2 family
CA_C2242	*CA_C2242*	Predicted transcriptional regulator arsE family
CA_C2259	*CA_C2259*	Predicted transcriptional regulator of MarR-family
CA_C2269	*aspS*	Aspartyl-tRNA synthetase
CA_C2272	*CA_C2272*	Predicted Zn-dependent hydrolase of metallo-beta-lactamase superfamily
CA_C2284	*ruvB*	Holliday junction specific DNA helicase subunit ruvB
CA_C2285	*ruvA*	Holliday junction specific DNA helicase subunit ruvA
CA_C2307	*CA_C2307*	Sigma factor F inhibitor spoIIAB (anti-sigma factor)
CA_C2310	*CA_C2310*	Methenyl tetrahydrofolate cyclohydrolase (serine cycle enzyme)
CA_C2311	*CA_C2311*	Nitroreductase family protein
CA_C2312	*CA_C2312*	Rhamnosyl transferase rfbQ
CA_C2319	*CA_C2319*	Hypothetical protein
CA_C2321	*CA_C2321*	Predicted glycosyltransferase
CA_C2328	*CA_C2328*	Polysaccharide ABC transporter ATPase component
CA_C2331	*CA_C2331*	DTDP-4-dehydrorhamnose 3 5-epimerase
CA_C2332	*spsJ*	DTDP-D-glucose 4 6-dehydratase
CA_C2333	*spsI*	DTDP-glucose pyrophosphorylase
CA_C2337	*CA_C2337*	Phosphomannomutase
CA_C2355	*CA_C2355*	Hypothetical protein
CA_C2356	*pheT*	Phenylalanyl-tRNA synthetase (beta subunit)
CA_C2357	*pheS*	Phenylalanyl-tRNA synthetase (alpha subunit)
CA_C2360	*rpmL*	Ribosomal protein L35
CA_C2361	*infC*	Translation initiation factor IF3
CA_C2362	*thrS*	Threonyl-tRNA synthetase
CA_C2365	*sspA*	Small acid-soluble spore protein
CA_Ct018	*CA_Ct018*	Aminoacyl-tRNA biosynthesis (tRNA - cysteine)
CA_Ct019	*CA_Ct019*	Aminoacyl-tRNA biosynthesis (tRNA - glycine)
CA_Ct020	*CA_Ct020*	Aminoacyl-tRNA biosynthesis (tRNA - phenylalanine)
CA_Ct021	*CA_Ct021*	Aminoacyl-tRNA biosynthesis (tRNA - aspartate)
CA_Ct022	*CA_Ct022*	Aminoacyl-tRNA biosynthesis (tRNA - valine)
CA_C2373	*CA_C2373*	Predicted membrane protein
CA_C2381	*dapD*	Tetrahydrodipicolinate N-succinyltransferase
CA_C2384	*CA_C2384*	Hypothetical protein
CA_C2399	*vals*	Valyl-tRNA synthetase
CA_C2402	*CA_C2402*	TPR repeats containing protein
CA_C2417	*CA_C2417*	Hypothetical protein
CA_C2481	*CA_C2481*	Predicted kinase from adenilate kinase family FLAR-like protein

CA_C2502	*CA_C2502*	Predicted permease
CA_C2508	*CA_C2508*	Nitroreductase family protein
CA_C2516	*CA_C2516*	Hypothetical protein
CA_C2545	*CA_C2545*	Hypothetical protein
CA_C2557	*CA_C2557*	Hypothetical protein
CA_C2564	*CA_C2564*	Hypothetical protein
CA_C2565	*CA_C2565*	NifU-related domain containing protein
CA_C2568	*CA_C2568*	Predicted transcriptional regulator
CA_C2569	*CA_C2569*	NimC/NimA family protein
CA_C2586	*CA_C2586*	Predicted membrane protein
CA_Ct023	*CA_Ct023*	Aminoacyl-tRNA biosynthesis (tRNA- leucine)
CA_C2595	*CA_C2595*	Hypothetical protein
CA_C2626	*fabG*	Possible 3-ketoacyl-acyl carrier protein reductase
CA_C2634	*CA_C2634*	Ferric uptake regulator (FUR family) YGAG B.subtilis ortholog
CA_C2636	*CA_C2636*	Predicted GTPase YSXC B.subtilis ortholog
CA_C2646	*sipS*	Signal peptidase I
CA_C2658	*glnA*	Glutamine synthetase type III
CA_Ct024	*CA_Ct024*	Aminoacyl-tRNA biosynthesis (tRNA - arginine)
CA_Ct025	*CA_Ct025*	Aminoacyl-tRNA biosynthesis (tRNA - OTHER)
CA_Ct026	*CA_Ct026*	Aminoacyl-tRNA biosynthesis (tRNA - proline)
CA_Ct028	*CA_Ct028*	Aminoacyl-tRNA biosynthesis (tRNA - arginine)
CA_Ct029	*CA_Ct029*	Aminoacyl-tRNA biosynthesis (tRNA - glycine)
CA_Ct030	*CA_Ct030*	Aminoacyl-tRNA biosynthesis (tRNA - glycine)
CA_Ct031	*CA_Ct031*	Aminoacyl-tRNA biosynthesis (tRNA - leucine)
CA_Ct032	*CA_Ct032*	Aminoacyl-tRNA biosynthesis (tRNA - lysine)
CA_Ct033	*CA_Ct033*	Aminoacyl-tRNA biosynthesis (tRNA - glutamine)
CA_Ct034	*CA_Ct034*	Aminoacyl-tRNA biosynthesis (tRNA - histidine)
CA_Ct035	*CA_Ct035*	Aminoacyl-tRNA biosynthesis (tRNA - lysine)
CA_Ct036	*CA_Ct036*	Aminoacyl-tRNA biosynthesis (tRNA - glycine)
CA_Ct037	*CA_Ct037*	Aminoacyl-tRNA biosynthesis (tRNA - arginine)
CA_Ct038	*CA_Ct038*	Aminoacyl-tRNA biosynthesis (tRNA - glycine)
CA_Ct039	*CA_Ct039*	Aminoacyl-tRNA biosynthesis (tRNA - glycine)
CA_Ct040	*CA_Ct040*	Aminoacyl-tRNA biosynthesis (tRNA - leucine)
CA_Ct041	*CA_Ct041*	Aminoacyl-tRNA biosynthesis (tRNA - lysine)
CA_Ct042	*CA_Ct042*	Aminoacyl-tRNA biosynthesis (tRNA - glutamine)
CA_Ct043	*CA_Ct043*	Aminoacyl-tRNA biosynthesis (tRNA - histidine)
CA_Ct044	*CA_Ct044*	Aminoacyl-tRNA biosynthesis (tRNA - arginine)
CA_Ct045	*CA_Ct045*	Aminoacyl-tRNA biosynthesis (tRNA - lysine)
CA_Ct046	*CA_Ct046*	Aminoacyl-tRNA biosynthesis (tRNA - proline)
CA_C2664	*CA_C2664*	Predicted phosphoesterase YSNB B.subtilis ortholog
CA_C2671	*CA_C2671*	Glu-tRNAGln amidotransferase subunit C
CA_C2672	*CA_C2672*	Predicted membrane protein
CA_C2680	*pgi*	Glucose-6-phosphate isomerase
CA_C2683	*cotF*	Related to spore coat protein F
CA_C2694	*CA_C2694*	Hypothetical protein
CA_C2698	*CA_C2698*	Hypothetical protein

CA_C2699	*CA_C2699*	Hypothetical protein
CA_C2701	*guaB*	IMP dehydrogenase
CA_C2703	*groEL*	Chaperonin GroEL HSP60 family
CA_C2704	*groES*	Co-chaperonin GroES HSP10 family
CA_C2707	*CA_C2707*	8-oxoguanine-DNA-glycosylase
CA_C2708	*hbd*	Beta-hydroxybutyryl-CoA dehydrogenase NAD-dependent
CA_C2709	*etfA*	Electron transfer flavoprotein alpha-subunit
CA_C2710	*etfB*	Electron transfer flavoprotein beta-subunit
CA_C2712	*crt*	Crotonase (3-hydroxybutyryl-COA dehydratase)
CA_C2715	*CA_C2715*	Hypothetical protein
CA_C2740	*hisS*	Histidyl-tRNA synthetase
CA_C2744	*CA_C2744*	Predicted membrane protein
CA_C2777	*CA_C2777*	Glutaredoxin
CA_C2778	*CA_C2778*	Rubredoxin
CA_C2785	*CA_C2785*	Hypothetical protein
CA_C2790	*CA_C2790*	Hypothetical protein
CA_Ct047	*CA_Ct047*	Aminoacyl-tRNA biosynthesis (tRNA - tryptophan)
CA_C2812	*CA_C2812*	Hypothetical protein
CA_C2838	*CA_C2838*	Predicted nucleotide-binding protein YjeE family
CA_C2856	*metK*	S-adenosylmethionine synthetase
CA_Ct048	*CA_Ct048*	Aminoacyl-tRNA biosynthesis (tRNA - tryptophan)
CA_Ct049	*CA_Ct049*	Aminoacyl-tRNA biosynthesis (tRNA - threonine)
CA_Ct050	*CA_Ct050*	Aminoacyl-tRNA biosynthesis (tRNA - valine)
CA_C2865	*atpD*	FoF1-type ATP synthase beta subunit
CA_C2866	*atpG*	FoF1-type ATP synthase gamma subunit
CA_C2867	*atpA*	FoF1-type ATP synthase alpha subunit
CA_C2868	*atpH*	FoF1-type ATP synthase delta subunit
CA_C2869	*atpF*	FoF1-type ATP synthase B subunit
CA_C2870	*atpE*	FoF1-type ATP synthase C subunit
CA_C2871	*atpB*	FoF1-type ATP synthase A subunit
CA_C2872	*CA_C2872*	Predicted membrane protein in FoF1-type ATP synthase operon
CA_C2874	*CA_C2874*	UDP-N-acetylglucosamine 2-epimerase
CA_C2875	*tagO*	Undecaprenyl-phosphate
CA_C2880	*CA_C2880*	Ribose 5-phosphate isomerase RpiB
CA_C2881	*CA_C2881*	Protein-tyrosine-phosphatase YWLE B.subtilis ortholog
CA_C2882	*CA_C2882*	Predicted translation factor (SUA5)
CA_C2884	*prfA*	Protein chain release factor A
CA_C2886	*CA_C2886*	Uncharacterized conserved protein predicted metal-dependent enzyme YQHQ B.subtilis ortholog
CA_C2888	*rpmE*	50S ribosomal protein L31
CA_C2892	*ctrA*	CTP synthase (UTP-ammonia lyase)
CA_C2895	*ddlA*	D-alanine-D-alanine ligase
CA_C2897	*CA_C2897*	Predicted UDP-glucuronosyltransferase YPFP B/subtilis ortholog
CA_C2902	*CA_C2902*	Isopentenyl monophosphate kinase IPK
CA_C2915	*panC*	Pantoate--beta-alanine ligase
CA_C2924	*thiS*	Uncharacterized protein possibly involved in thiamine biosynthesis

CA_C2990	*csp*	Cold shock protein
CA_C3004	*folA*	Dihydrofolate reductase
CA_C3018	*CA_C3018*	Rubrerythrin
CA_C3023	*CA_C3023*	Hypothetical protein
CA_C3060	*CA_C3060*	CPSC/CAPB subfamily ATPase
CA_C3073	*CA_C3073*	Sugar transferase involved in lipopolysaccharide synthesis
CA_C3074	*CA_C3074*	Uncharacterized conserved protein
CA_C3075	*buk*	Butyrate kinase BUK
CA_C3076	*ptb*	Phosphate butyryltransferase
CA_C3094	*CA_C3094*	Uncharacterized conserved protein associated with phosphate permease
CA_C3097	*rpsI*	Ribosomal protein S9
CA_C3098	*rplM*	Ribosomal protein L13
CA_C3102	*CA_C3102*	ABC-type transporter ATPase component (cobalt transporters subfamily)
CA_C3103	*rplQ*	Ribosomal protein L17
CA_C3105	*rpsD*	Ribosomal protein S4
CA_C3106	*rpsK*	Ribosomal protein S11
CA_C3108	*rpmJ*	Ribosomal protein L36
CA_C3109	*infA*	Translation initiation factor IF-1
CA_C3110	*CA_C3110*	Ribosomal protein L14E (archaeal/eukaryotic type)
CA_C3111	*CA_C3111*	Hypothetical protein
CA_C3112	*adk*	Adenylate kinase
CA_C3113	*seqY*	Preprotein translocase subunit SecY
CA_C3115	*rpmD*	Ribosomal protein
CA_C3115	*rpmD*	Ribosomal protein L30
CA_C3116	*rpsE*	Ribosomal protein S5
CA_C3118	*rplF*	Ribosomal protein L6
CA_C3119	*rpsH*	Ribosomal protein S8
CA_C3120	*rpsN*	Ribosomal protein S14
CA_C3121	*rplE*	Ribosomal protein L5
CA_C3122	*rplX*	Ribosomal protein L24
CA_C3123	*rplN*	Ribosomal protein L14
CA_C3125	*rpmC*	Ribosomal protein L29
CA_C3126	*rplP*	Ribosomal protein L16
CA_C3128	*rplV*	Ribosomal protein L22
CA_C3129	*rpsS*	Ribosomal protein S19
CA_C3130	*rplB*	Ribosomal protein L2
CA_C3131	*rplW*	Ribosomal protein L23
CA_C3132	*rplD*	Ribosomal protein L4
CA_C3133	*rplC*	Ribosomal protein L3
CA_C3135	*CA_C3135*	Uncharacterized C4-type Zn-finger containing protein
CA_C3136	*tufA*	Elongation Factor Tu (Ef-Tu)
CA_C3139	*rpsG*	Ribosomal protein S7
CA_C3140	*rpsL*	Ribosomal protein S12
CA_C3143	*rpoB*	DNA-dependent RNA polymerase beta subunit
CA_C3144	*CA_C3144*	Hypothetical protein

CA_C3145	*rplL*	Ribosomal protein L7/L12
CA_C3146	*rplJ*	Ribosomal protein L10
CA_C3147	*rplA*	Ribosomal protein L1
CA_C3148	*rplK*	Ribosomal protein L11
CA_C3150	*secE*	Preprotein translocase subunit SecE
CA_C3151	*rpmG*	L33
CA_Ct052	*CA_Ct052*	Aminoacyl-tRNA biosynthesis (tRNA - threonine)
CA_Ct053	*CA_Ct053*	Aminoacyl-tRNA biosynthesis (tRNA - glycine)
CA_C3152	*CA_C3152*	DNA-dependent RNA polymerase sigma subunit
CA_C3153	*CA_C3153*	Uncharacterized conserved protein YACP B.subtilis ortholog
CA_C3175	*CA_C3175*	Hypothetical protein
CA_C3178	*proS*	Prolyl-tRNA synthetase
CA_C3184	*CA_C3184*	4-diphosphocytidyl-2-methylerithritol synthase (Sugar Nucleotide Phosphorylase family)
CA_Ct054	*CA_Ct054*	Aminoacyl-tRNA biosynthesis (tRNA - asparagine)
CA_Ct056	*CA_Ct056*	Aminoacyl-tRNA biosynthesis (tRNA - methionine)
CA_C3195	*CA_C3195*	Glycyl-tRNA synthetase
CA_C3196	*CA_C3196*	Hypothetical protein
CA_Ct057	*CA_Ct057*	Aminoacyl-tRNA biosynthesis (tRNA - asparagine)
CA_C3197	*lysS*	Lysyl-tRNA synthetase
CA_C3200	*CA_C3200*	Predicted transcriptional regulator homolog of Bvg accessory factor
CA_C3201	*CA_C3201*	Formate--tetrahydrofolate ligase
CA_C3202	*ftsH*	ATP-dependent Zn protease FTSH
CA_C3203	*hprT*	Hypoxanthine-guanine phosphoribosyltransferase
CA_Ct058	*CA_Ct058*	Aminoacyl-tRNA biosynthesis (tRNA - methionine)
CA_Ct059	*CA_Ct059*	Aminoacyl-tRNA biosynthesis (tRNA - methionine)
CA_Ct060	*CA_Ct060*	Aminoacyl-tRNA biosynthesis (tRNA - leucine)
CA_C3206	*CA_C3206*	Ribosomal protein S1 domain family protein
CA_C3207	*CA_C3207*	Hypothetical protein
CA_C3210	*CA_C3210*	Predicted RNA-binding protein YABO B.subtilis ortholog
CA_C3211	*hbs*	DNA binding protein HU
CA_C3222	*gcaD*	UDP-N-acetylglucosamine pyrophosphorylase
CA_C3223	*spoVG*	Stage V sporulation protein G
CA_C3225	*murC*	UDP-N-acetylmuramate-alanine ligase
CA_Ct071	*CA_Ct071*	Aminoacyl-tRNA biosynthesis (tRNA - proline)
CA_Ct072	*CA_Ct072*	Aminoacyl-tRNA biosynthesis (tRNA - proline)
CA_C3252	*proC*	Pyrroline-5-carboxylate reductase
CA_C3254	*proA*	Gamma-glutamyl phosphate reductase
CA_C3255	*CA_C3255*	Hypothetical protein
CA_C3288	*CA_C3288*	Iron-regulated ABC transporter ATPase subunit (SufC) VEG296 B.subtilis ortholog
CA_C3290	*CA_C3290*	Iron-regulated ABC-type transporter membrane component (SufB)
CA_C3292	*CA_C3292*	NifU homolog involved in Fe-S cluster formation
CA_C3318	*CA_C3318*	Hypothetical protein
CA_C3324	*CA_C3324*	Predicted transcriptional regulator
CA_Ct061	*CA_Ct061*	Aminoacyl-tRNA biosynthesis (tRNA - asparagine)
CA_C3438	*CA_C3438*	Transcriptional regulator

CA_C3479	*CA_C3479*	Hypothetical protein
CA_C3480	*CA_C3480*	Hypothetical protein
CA_C3481	*CA_C3481*	Transcriptional regulator AcrR family
CA_C3497	*CA_C3497*	Hypothetical protein
CA_C3527	*CA_C3527*	Ferredoxin
CA_C3537	*CA_C3537*	Fragment of SECA (fragment)
CA_C3539	*murA*	UDP-N-acetylglucosamine enolpyruvyl transferase
CA_C3568	*accA*	Acetyl-CoA carboxylase alpha subunit
CA_C3569	*accD*	Acetyl-CoA carboxylase beta subunit
CA_C3570	*accC*	Biotin carboxylase
CA_C3571	*fabZ*	Hydroxymyristoyl-(acyl carrier protein) dehydratase
CA_C3574	*fabG*	3-ketoacyl-acyl carrier protein reductase
CA_C3575	*fabD*	Malonyl CoA-acyl carrier protein transacylase
CA_C3576	*CA_C3576*	Dioxygenase related to 2-nitropropane dioxygenase
CA_C3580	*CA_C3580*	Dioxygenase related to 2-nitropropane dioxygenase
CA_Ct062	*CA_Ct062*	Aminoacyl-tRNA biosynthesis (tRNA - threonine)
CA_Ct063	*CA_Ct063*	Aminoacyl-tRNA biosynthesis (tRNA - aspartate)
CA_Ct064	*CA_Ct064*	Aminoacyl-tRNA biosynthesis (tRNA - valine)
CA_Ct065	*CA_Ct065*	Aminoacyl-tRNA biosynthesis (tRNA - glutamate)
CA_Ct066	*CA_Ct066*	Aminoacyl-tRNA biosynthesis (tRNA - threonine)
CA_Ct067	*CA_Ct067*	Aminoacyl-tRNA biosynthesis (tRNA - aspartate)
CA_Ct068	*CA_Ct068*	Aminoacyl-tRNA biosynthesis (tRNA - valine)
CA_Ct069	*CA_Ct069*	Aminoacyl-tRNA biosynthesis (tRNA - glutamate)
CA_C3592	*CA_C3592*	Hypothetical protein CF-27 family
CA_C3599	*CA_C3599*	Hypothetical protein
CA_C3621	*fer*	Ferredoxin
CA_C3633	*CA_C3633*	Hypothetical protein
CA_C3639	*CA_C3639*	CRO repressor-like DNA-binding protein
CA_C3654	*CA_C3654*	Heavy-metal binding protein (similar to N-terminal domain of MerA)
CA_C3691	*CA_C3691*	Hypothetical protein
CA_C3703	*CA_C3703*	Hypothetical protein
CA_C3707	*CA_C3707*	Uncharacterized protein homolog of Bacillus firmus (gi:2654481)
CA_C3713	*CA_C3713*	Uncharacterized conserved protein yukE/yfjA B.subtilis family
CA_Ct070	*CA_Ct070*	Aminoacyl-tRNA biosynthesis (tRNA - threonine)
CA_C3721	*CA_C3721*	Hypothetical secreted protein
CA_C3723	*ssb*	Single strand DNA binding protein SSB
CA_C3724	*rpsF*	Ribosomal protein S6
CA_C3725	*CA_C3725*	Uncharacterized protein homolog of Streptococcus pneumoniae (gi:2109447)
CA_C3738	*rnpA*	RnpA ribonuclease P protein component
CA_C3739	*rpmH*	L34
CA_P007	*CA_P007*	Hypothetical protein
CA_P012	*pedb*	Pediocin immunity protein
CA_P031	*hlyU*	Transcriptional activator HLYU
CA_P061	*CA_P061*	Hypothetical protein
CA_P087	*CA_P087*	HTH transcriptional regulator TetR/AcrR family

CA_P100	*CA_P100*	HTH transcriptional regulator TetR/AcrR family
CA_P132	*CA_P132*	Hypothetical protein
CA_P139	*CA_P139*	Possible D-alanyl carrier protein
CA_P169	*CA_P169*	Hypothetical protein

Appendix B: Supplementary Information for Chapter 5

Table 7.2: *C. acetobutylicum* ATCC 824 bacteriophage ACK1 gene list

ORF	start	stop	strand	Putative function	Bacteria genome
1	3	278	+	Transcriptional regulator (phage-related) (Xre family)	CA_C1867
2	935	327	-	Hypothetical protein	CA_C1868
3	1094	1297	+	Predicted transcriptional regulator	CA_C1869
4	1413	1688	+	Hypothetical protein	CA_C1870
5	1741	2958	+	FtsK/SpoIIIE family protein	CA_C1871
6	3118	4206	+	Replication initiation protein REP	repB
7	4449	4273	-	Hypothetical protein	CA_C1873
8	5498	4464	-	TPR repeat	CA_C1874
9	6281	5718	-	Hypothetical protein, CF-36 family	CA_C1875
10	7356	6301	-	Autolytic lysozyme, N, acetyl muramidase with peptidoglycan binding domain	lyc
11	7631	7374	-	Putative Autolytic lysozyme	CA_C1877
12	7909	7643	-	Hypothetical protein	CA_C1878
13	12150	7939	-	Phage minor Structural protein	CA_C1879
14	12556	12182	-	Hypothetical protein, CF-35 family	CA_C1880
15	12809	12543	-	Hypothetical protein	CA_C1881
16	14036	12828	-	Hypothetical protein	CA_C1882
17	19497	14038	-	Phage tail length tape-measure protein T	CA_C1883
18	19686	19531	-	Hypothetical protein	CA_C1884
19	20042	19713	-	Hypothetical protein	CA_C1885
20	21016	20096	-	Phage major tail protein	CA_C1886
21	21384	21061	-	Prohead protease	CA_C1887
22	21812	21381	-	Phage head-tail joining protein	CA_C1888
23	22160	21813	-	Phage head-tail adaptor protein	CA_C1889
24	22440	22123	-	Phage head-tail joining protein	CA_C1890
25	22717	22432	-	Hypothetical protein	CA_C1891
26	23918	22728	-	Phage major capsid protein	CA_C1892
27	24690	23920	-	ClpP protease family serine protease	CA_C1893
28	25912	24653	-	Phage head portal protein	CA_C1894
29	27699	25909	-	Phage terminase, large subunit	CA_C1895
30	28129	27674	-	Phage terminase, small subunit	CA_C1896
31	28616	28305	-	Phage-associated homing endonuclease	CA_C1897
32	28886	28623	-	Hypothetical protein	CA_C1898
33	29358	28981	-	Thymidine kinase recombinase related endonuclease	CA_C1899

34	29530	29351	-	Phage regulatory Com-like protein, containing Zn-finger	CA_C1900
35	29871	29518	-	Hypothetical protein	CA_C1901
36	30103	29906	-	Hypothetical protein	CA_C1902
37	30281	30063	-	Hypothetical protein	CA_C1903
38	30779	30585	-	Hypothetical protein	CA_C1904
39	31274	30924	-	Hypothetical protein	CA_C1905
40	31684	31484	-	Hypothetical protein	CA_C1906
41	31863	31699	-	Hypothetical protein	CA_C1907
42	32066	31869	-	Predicted membrane protein	CA_C1908
43	32937	32320	-	Ribonuclease D-related protein	rND
44	33825	32950	-	Predicted membrane protein	CA_C1910
45	34214	34435	+	Hypothetical protein	CA_C1911
46	34678	34466	-	Uncharacterized phage-related protein	CA_C1912
47	34949	34731	-	Zn-finger containing protein	CA_C1913
48	35452	35180	-	Hypothetical protein	CA_C1914
49	36152	35625	-	Hypothetical protein	CA_C1915
50	36475	36284	-	Hypothetical protein	CA_C1916
51	36864	36523	-	Hypothetical protein	CA_C1917
52	37168	36947	-	Hypothetical protein	CA_C1918
53	37616	37215	-	Single-stranded DNA-binding protein	CA_C1919
54	37803	37621	-	Zn-finger containing protein	CA_C1920
55	38087	37803	-	Hypothetical protein	CA_C1921
56	38372	38175	-	Hypothetical protein	CA_C1922
57	38716	38384	-	Hypothetical protein	CA_C1923
58	39220	38717	-	Hypothetical protein	CA_C1924
59	39677	39444	-	Hypothetical protein	CA_C1926
60	40196	39819	-	Hypothetical protein, CF-31 family	CA_C1927
61	40455	40189	-	Hypothetical protein	CA_C1928
62	40786	40571	-	Hypothetical protein	CA_C1929
63	41194	40823	-	Uncharacterized protein, MazG family	CA_C1930
64	41378	41184	-	Hypothetical protein	CA_C1931
65	41996	41469	-	Hypothetical protein	CA_C1932
66	42707	42006	-	Helicase loader DnaI	DNAC
67	43632	42793	-	Phage replication initiation protein/Replisome Organiser	CA_C1934
68	44044	43733	-	Hypothetical protein	CA_C1935
69	44744	44055	-	Translation initiation Factor	CA_C1936
70	44882	44745	-	Hypothetical protein	CA_C1937
71	45462	44887	-	HD superfamily Hydrolase protein	CA_C1938
72	45659	45465	-	Transcriptional regulator protein	CA_C1939
73	45913	45710	-	Hypothetical protein	CA_C1940
74	46157	45918	-	Transition state regulatory protein AbrB	abrB
75	46637	46449	-	Hypothetical protein	CA_C1942
76	46721	47209	+	Hypothetical protein	CA_C1943
77	47638	47198	-	Hypothetical protein	CA_C1944
78	48386	47652	-	Phage antirepressor protein	CA_C1945

79	48937	48653	-	Phage related transcriptional regulator (Xre family)	CA_C1946
80	49145	49516	+	Transcriptional regulator, Xre family	CA_C1947
81	49665	49910	+	Hypothetical protein	CA_C1948
82	51111	49921	-	Possible TPR-repeat contaning protein	CA_C1949
83	51445	51657	+	Cro/C1 type helix-turn-helix domain from pfam13443 family transcriptional regulator	CA_C1950
84	51668	53395	+	Resolvase from the family Serine Recombinase	CA_C1951