The Antimicrobial and Bile Acid Mediated Control of *Clostridium difficile* Infection

by

Andrew William Dempster BSc (Hons)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy



University of Nottingham

UK | CHINA | MALAYSIA

September 2017

Declaraction

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

Andrew W Dempster

September 2017

Abstract

Clostridium difficile is an anaerobic, Gram-positive, endospore forming bacillus and is the leading cause of nosocomial infection. Symptoms range from mild diarrhoea to the potentially fatal intestinal complications pseudomembranous colitis and toxic megacolon. The prerequisite for *C. difficile* infection (CDI) is the perturbation of the healthy microbiota of the gut by broad spectrum antibiotics. It is therefore important to develop therapies which take this in to account, either by minimal disruption of the resident gut microbiota, or by reinstating the protective nature of the gut microbiota. The novel antimicrobial fidaxomicin (FDX) is the first in a new class of macrocyclic antibiotics, with a narrow spectrum of activity for *C. difficile*. FDX exerts its bactericidal activity by binding to RNA polymerase (RNAP) and inhibiting transcription. The minimum inhibitory concentrations were determined for six clinically relevant isolates of *C. difficile* and the effect of the drug on spore germination and outgrowth was assessed. Inhibition of *C. difficile* occurs at low concentrations (0.03 – 0.05 μg/mL) and it was found that FDX does not inhibit the initiation of spore germination, but effectively halts outgrowth at an early stage. The effect of mutations in the β subunit of RNAP were also investigated in terms of susceptibility to FDX and any potential fitness costs incurred to the bacterium. Three separate single nucleotide polymorphisms (SNPs), T3428A, T3428G, G3427T, in the *rpoB* gene were found to confer reduced susceptibilities to FDX. However, the clinical relevance of these mutations is unclear, as mutants appeared to be attenuated in terms of growth, toxin production and virulence in the hamster model of infection.

Clostridium scindens is a member of the healthy gut microbiota and is thought to be a key organism in providing colonisation resistance against *C. difficile*. *C. scindens* is the most active transformer of primary bile acids to secondary bile acids, known to inhibit the growth of *C. difficile*. This occurs due to the gene products of the bile acid inducible

operon and a presently unknown reductive arm of the pathway. An RNA extraction protocol for high quality total RNA from *C. scindens* was developed to aid in the study of the transcriptome of *C. scindens* cultured with the primary bile acid, cholic acid (CA). This has enabled the identification of potential gene candidates for the reductive arm of the bile transformative pathway of *C. scindens*. To further study these potential genes, the ability to transfer DNA in to *C. scindens* is desirable in order to create gene knock outs. The genome of *C. scindens* ATCC 35704 was assembled and annotated and used to identify potential barriers to DNA transfer. The methylation pattern of this strain identified two type I, twelve type II and one type IV restriction methylation (RM) systems. RM systems were further characterised in an effort to circumvent the RM system barrier to DNA transfer.

Acknowledgements

For my Dad, W.B. Dempster (FIBMS), who inspired me to pursue a career in microbiology with science stories around the dinner table, and impromptu on-call visits to his lab.

I would like to thank my Mum for reinforcing my passion for science and her help and support throughout my PhD and the writing process(es), with late night science chats over coffee and the comforts of home. My sisters, Claire and Jenny, for distracting me with photos and videos of my niece and nephews and sending a much-appreciated food to get me through the long days and nights.

I would also like to thank Dr Sarah Kuehne for her unwavering and attentive support over the last 4 years and beyond, particularly in these last months.

Professor Nigel Minton for invaluable advice, supervision and providing the impetus for the final push, with the "Just get it done…" mantra.

And, in no particular order:

Dr Nimitray Joshi for the RNA counselling and pizza hut sessions.

Dr Mark Collery for the toxic advice. We'll always have Barcelona.

Dr Daniela Heeg for teaching me about spores.

Michelle Kelly and Rory Cave for hamster wrangling.

The Leeds Microbiologists: Rob, Duncan, Kate, Siobhan, Lewis.

All my friends and colleagues in the SBRC past and present, particularly those in C02, B60 and B39; Pete, Patrick, Micky, Rob, James, Craig, Bart, Florence, Ronja, Emily.

List of Contents

Declaraction....2
Abstract....3
Acknowledgements....5
List of Figures....11
List of Tables....13
List of Contents....6
Abreviations....14
Chapter 1 Introduction....17
1.1 The Human Microbiome....18
1.2 *Clostridium difficile*....20
1.2.1 History and Phylogeny....20
1.2.2 Epidemiology of *Clostridium difficile* Infection....24
1.2.3 *Clostridium difficile* Toxins....26
1.2.4 *Clostridium difficile* Sporulation....29
1.2.5 *Clostridium difficile* Spores....32
1.2.6 *Clostridium difficile* Spore Germination....35
1.2.5 *Clostridium difficile* Treatment....38
1.3 *Clostridium scindens*....40
1.3.1 History and Discovery....40
1.3.2 *Clostridium scindens* Bile Acid Metabolism....41
1.3.3 Consequences of Bile Acid Transformations....43
1.4 Aims of the Project....46
Chapter 2 Materials and Methods....48
2.1 Bacterial Strains and Plasmids....49
2.1.1 Bacterial Strains....49
2.1.2 Plasmids....51
2.2 Culture Media and Growth Conditions....53
2.2.1 Culture Media....53
2.2.2 Antibiotic Supplements....54
2.2.3 Growth Conditions....54
2.3 Microbiological Methods....55
2.3.1 Bacterial Growth Profiles....55
2.3.2 Minimum Inhibitory Concentration Assay....56
2.3.4 Preparation of Pure Spores of C. difficile....56
2.3.5 Microscopy....56

2.3.6 Preparation of Electrocompetent E. coli 56
2.3.7 Preparation of Electrocompetent C. scindens 57
2.3.8 Conjugal Transfer of Plasmids 57
2.3.9 Spore Germination Initiation Assay 58
2.3.10 Spore Outgrowth Assay 58
2.3.11 Competetive Growth Advantage Assay 59
2.3.12 Rate and Capacity of Sporulation Assay 60
2.3.13 Loss of Heat Resistance Germination Assay 60
2.3.14 Total Toxin Production Assay 60
2.3.15 Cytotoxicity Assay 61
2.3.16 In-Vitro 7α-Dehydroxylation 61
2.3.17 Bile Acid Analysis 62
2.4 *In Vivo* Methods 62
2.4.1 The Hamster Model of Virulence 62
2.4.2 Enumeration of Hamster Cecal Colonisation 63
2.5 Molecular Methods 63
2.5.1 Extraction of Chromosomal DNA 63
2.5.2 Extraction of Total RNA and Purification from *C. scindens* 64
2.5.3 Polymerase Chain Reaction 67
2.5.4 Plasmid Extraction 68
2.5.5 Restriction Digest 68
2.5.6 Dephosphorylation of Digested Vectors 69
2.5.7 Ligation 69
2.5.8 Agarose Gel Electrophoresis 69
2.5.9 Extraction of DNA from Agarose Gels 70
2.5.11 Transformation of E. coli by Electroportation 71
2.5.12 Transformation of *C. scindens* by Electroportation 71
2.5.13 GpC In-Vitro Methylation 72
2.5.14 Quickchange Site Directed Mutagenesis 72
2.5.15 HiFi Assembly 73
2.5.16 Lambda Red Recombineering 74
2.5.17 DNA and RNA Quantification 75
2.5.18 RNA Quality Assessment 75
2.5.19 Extraction and Purification of Total RNA from *C. difficile* 75
2.6 Oligonucleotides 77
2.7 Sequencing Methods 81
2.7.1 Sanger Sequencing 81

2.7.2 Pacific Biosciences Single Molecule Real-Time Sequencing 81
2.7.3 Illumina MiSeq Sequencing 81
2.7.4 RNA Sequencing 82
2.8 Bioinformatics and Data Analysis 82
2.8.1 Characterisation Statistical Analysis 82
2.8.2 *C. scindens* ATCC 45704 Genome Assembly and Annotation 83
2.8.3 RNA Sequencing Analysis 83
Chapter 3 The Effect of The Novel Antibimicrobial Fidaxomicin on Spores of *Clostridium difficile* and The Characterisation of Mutants With Reduced Susceptibility 85
3.1 Introduction 86
3.1.1 Fidaxomicin Discovery and Development 86
3.1.2 Fidaxomicin Pharmacokinetics 86
3.1.3 DNA-dependent RNA Polymerase 87
3.1.4 Fidaxomicin Spectrum of Activity 89
3.1.5 Fidaxomicin Mechanism of Action 90
3.1.6 Fidaxomicin Resistance 92
3.1.7 *C. difficile* Ribotypes of Study 94
3.1.2 Aims and Hypotheses 97
3.2 Results 99
3.2.1 Fidaxomicin Minimum Inhibitory Concentrations 99
3.2.2 The Effect of Fidaxomicin on the Initiation of Spore Germination 101
3.2.3 The Effect of Fidaxomicin on the Outgrowth of Spores 105
3.2.4 Construction of rpoB Mutatants 110
3.2.5 Minimum Inhibitory Concentration 113
3.2.6 Growth Rate 114
3.2.7 Competetive Growth Advantage 116
3.2.8 Rate and Capacity of Sporulation 118
3.2.9 Loss of Heat Resistance Germination Assay 120
3.2.10 Toxin A and B Production 122
3.2.11 Vero Cell Cytotoxicity 123
3.2.12 In vivo Virulence and Colonisation in the Golden Syrian Hamster Model . 125
3.3 Discussion 127
3.3.1 The Effect of Fidaxomicin on Spore Germination and Outgrowth 127
3.3.2 The Effect of rpoB Mutations on Susceptbility to Fidaxomicin 129
3.3.3 The Impact of rpoB Substitutions on the Fitness of *C. difficile* 131
Chapter 4 The Bile Acid Metabolism of *Clostridium scindens* 136

4.1 Introduction 137
4.1.1 Primary Bile Acid Synthesis and The Enterohepatic Circulation 137
4.1.2 Bile Acid Deconjugation 140
4.1.3 Bile Acid 7-Dehydroxylation 140
4.1.4 Aims and Hypotheses 145
4.2 Results 146
4.2.1 The Effect of *C. scindens* Derived Secondary Bile Acids on the Germination of *C. difficile* Spores 146
4.2.2 The Effect of *C. scindens* Derived Secondary Bile Acids on the Outgrowth of C. difficile Spores 150
4.2.1 *C. scindens* Cholic Acid Growth 153
4.2.2 *C. scindens* Growth Curve and In-Vitro Bile Acid Transformations 154
4.2.3 *C. scindens* Growth Curves for RNA Extraction 156
5.2.2 Assembly and Annotation of the *C. scindens* ATCC 35704 Genome 158
4.2.4 Development and Optimisation of RNA Extraction from *C. scindens* 160
4.2.5 Preliminary Identification of Uncharacterised Bile Acid Metabolism Genes 163
4.2.6 Differential Gene Expression of the Bile Acid Inducible Operon 165
4.2.7 Whole Genome Differential Gene Expression 172
4.2.8 Hierarchical Clustering 176
4.3 Discussion 180
4.3.1 The Reponse of *C. difficile* Spores to *C. scindens* Derived Secondary Bile Acids 180
4.3.2 In-Vitro 7α-Dehydroxylation and Bile Acid Analysis 182
4.3.3 Expression of the Bile Acid Inducible Operon 183
4.3.5 Identification of a Novel Bile Acid Dehydratase 184
4.3.4 Identification of Potential Reductive Arm Candidates 186
Chapter 5 *Clostridium scindens* Molecular Characterisation and DNA Transfer 190
5.1 Introduction 191
5.1.1 Bacterial Transformation 191
5.1.2 Mechanisms of Horizontal Gene Transfer 191
5.1.3 Horizontal Gene Transfer by Natural Transformation 192
5.1.4 Horizontal Gene Transfer by Transduction 193
5.1.5 Horizontal Gene Transfer by Conjugation 195
5.1.6 Barriers To Gene Transfer 197
5.1.7 Aims and Hypotheses 203
5.2 Results 204
5.2.1 Initial Attempts to Establish DNA Transfer by Conjugation 204
5.2.2 Initial Attempts to Establish DNA Transfer by Transformation 207

5.2.3 Features of the *C. scindens* ATCC 35704 Genome 208
5.2.4 The CRISPR Cas System of *C. scindens* ATCC 35704 210
5.2.5 Methylome Analysis of *C. scindens* ATCC 35704 212
5.2.6 Restriction Analysis of *C. scindens* ATCC 35704 Total DNA 217
5.2.7 *C. scindens* Lysate Restriction Analysis 219
5.2.8 Transcriptomic Analysis of the Restriction Modification Systems of *C. scindens* ATCC 35704 220
5.2.9 Cloning of Methyltransferase Genes 222
5.2.10 Restriction Analysis of pRBV1_2A and pRBV1_2D 224
5.2.11 Recombineering of the 2A Methyltransferase 227
5.2.12 Final Conjugation Attempts 228
5.3 Discussion 229
5.3.1 Initial DNA Transfer Attempts 229
5.3.2 Genome and Methylome Analysis 231
Chapter 6 Final Conclusions and Perspectives 239
6.1 Conclusions and Outcomes 240
6.2 Future Perspectives 244
7. Bibliography 248
8. Appendices 273
8.1 Bioanalyser Electropherograms 273
8.1.1 Electropherograms for RNA Extrations at 4 Hours 273
8.1.2 Electropherograms for RNA Extrations at 6 Hours 274
8.1.3 Electropherograms for RNA Extrations at 8 Hours 275
8.1.4 Electropherograms for RNA Extrations at 10 Hours 276
8.2 Cholic Acid Induced Differenial Gene Expression 277
8.2.1 Differential Gene Expression: 4 Hours 277
8.2.2 Differential Gene Expression: 6 Hours 287
8.2.3 Differential Gene Expression: 8 Hours 320
8.2.4 Differential Gene Expression: 10 Hours 338

List of Figures

Figure 1.1 A Phylogenetic Tree of Various *Clostridium* Species 23
Figure 1.2 The Pathogenicity Locus of *C. difficile* 27
Figure 1.1: The Process of *C. difficile* Sporulation 30
Figure 1.2 The Structure of the *C. difficile* Spore 34
Figure 1.3 *C. difficile* Spore Germination and Outgrowth 37
Figure 1.4 The Organisation of the *bai* Operon in *C. scindens* ATCC 35704 42
Figure 3.1 Spacefill and Cartoon Crystal Structures of *Thermus thermophilus* 89
Figure 3.2 Orthogonal views of RNAP with highlighted binding sites of RNAP inhibitors 92
Figure 3.17 Prevalence of *C. difficile* Strains in Europe 96
Figure 3.3: FDX Minimum Inhibitory Concentration of *C. difficile* Isolates 100
Figure 3.4.1 Initiation of Germination 103
Figure 3.4.2 Initiation of Germination 104
Figure 3.5 Spore Outgrowth 108
Figure 3.6.1 Phase Contrast Microscopy of Spores and Vegetative Cells 109
Figure 3.6.2 Phase Contrast Microscopy of Germination and Outgrowth of Spores 109
Figure 3.7 Alignments of Wild Type and Mutant rpoB Sequences 112
Figure 3.8 pMTL-YN4-T3428A 112
Figure 3.9 Minimum Inhibitory Concentrations of Mutants 113
Figure 3.10 Mutant and Parental Strain Growth Curves 115
Figure 3.11 Competetive Growth Assays 117
Figure 3.12 Rate and Capacity of Sporulation 119
Figure 3.13 Loss of Heat Resistance of Parental and Mutant Strains 121
Figure 3.14 Parental and Mutant Toxin Production. 124
Figure 3.15 Virulence in the Golden Syrian Hamster Model 126
Figure 3.16 Golden Syrian Hamster Caecal Colonisation 126
Figure 3.18 Cartoon Crystal Structure of Thermus thermophilus RNAP holoenzyme. 133
Figure 4.1 Primary Bile Acid Synthesis and Conjugation 139
Figure 4.2 The Bile Acid Transformation of *C. scindens* strain ATCC 35704 144
Figure 4.3.1 Initiation of Germination 148
Figure 4.3.2 Initiation of Germination 149
Figure 4.5 Cholic Acid Growth Curves for *C. scindens* strain ATCC 35704 154
Figure 4.6 *C. scindens* Growth Rate and In-vitro Bile Acid Transformations 155
Figure 4.7 Growth Rates of *C. scindens* Cultures for RNA Extractions. 157
Figure 4.8 The Genome of *C. scindens* ATCC 35704 159

Figure 4.9 Gene expression of The Bile Acid Inducible Operon at 4, 6, 8 and 10 Hours 166
Figure 4.10 Expression Levels of the Bile Acid Inducible Operon in CA Induced Conditions..... 167
Figure 4.11 A; B Differential gene expression between CA induced and non-induced conditions at 4 and 6 hours..... 174
Figure 4.12 A; B Differential gene expression between CA induced and non-induced conditions at 8 and 10 hours..... 175
Figure 4.13 Whole Transcriptome Heatmap and Hierarchical Clustering for 4, 6, 8 and 10 hour Time Points..... 179
Figure 4.14 The 12α HSDH mediated transformation of DCA to 12-keto-DCA..... 182
Figure 4.15 The BaiE 7α-Dehydratase of *C. scindens* ATCC 3704..... 185
Figure 4.16 The Reaction Catalysed by 3-Ketosteroid Dehydrogenase..... 188
Figure 4.17 The Hypothesised Activity of the FAD binding reductase in the reductive steps of 7α-Dehydroxylation..... 188
Figure 5.1 Mechanisms of Horizontal Gene Transfer..... 196
Figure 5.2 Barriers to Gene Transfer..... 201
Figure 5.3 The Four Types of Bacterial Restriction Modification Systems..... 202
Figure 5.4 The Genome of *C. scindens* ATCC 35704..... 209
Figure 5.5 The CRISPR Loci and Cas operon of *C. scindens* ATCC 35704..... 211
Figure 5.6 The Restriction Modification Systems of *C. scindens* ATCC 35704..... 215
Figure 5.7 The Restriction Modification Systems of *C. scindens* ATCC 35704..... 216
Figure 5.8 Restriction Analysis of Genomic DNA from *C. scindens* ATCC 35704..... 218
Figure 5.9 Lysate Restriction Analysis..... 219
Figure 5.10 Expression Profiles of the RM Systems During Exponential Growth..... 221
Figure 5.11 The Constructed pRBV1-1A Vector..... 223
Figure 5.12 The Constructed pRBV1-2A Vector..... 223
Figure 5.15 AluI Restriction Analysis..... 226
Figure 5.16 AlwI Restriction Analysis..... 226
Figure 5.18 Recombineering of the 2A Methyltransferase..... 228
Figure 5.19 Screening of Recombinant R702-2A Plasmids..... 228
Figure 5.20 The Modular Vector pMTL8X121..... 235
Figure 5.21 The Linear Prophage Sequence ϕClsc35704 of *C. scindens* ATCC 35704.. 236

List of Tables

Table 1.1 The Functions of the Genes of the *bai* Operon 42
Table 2.1 Bacterial Strains 49
Table 2.2 Plasmids 51
Table 2.3 Culture Media 53
Table 2.4 Antibiotic Working Concentrations 54
Table 2.5 PCR Conditions 67
Table 2.6 Ligation Reactions 69
Table 2.7 *C. scindens* Electroporation Settings 72
Table 2.10 HiFi Reaction Mixtures 74
Table 2.11 Primers 78
Table 3.1. RNAP inhibitors with binding sites and proposed mechanisms of action 91
Table 3.2: SNPs and Amino Acid Substitutions Conferring Reduced Susecptibility to FDX in spontaneous mutants of *C. difficile* 93
Table 3.3 Strain Designations and PCR-Ribotypes 94
Table 3.4: Minimum inhibitory concentrations of the selected *C. difficile* strains 100
Table 3.5 MICs of Parental and Mutant Strains 113
Table 4.1 The Steps of 7α-Dehydroxylation of Cholic Acid in *C. scindens* 143
Table 4.2 RNA Sample Details for 4 and 6 Hour Timepoints 161
Table 4.3 RNA Sample Details for 4 and 6 Hour Timepoints 162
Table 4.4 Identified Gene Candidates for the Reductive Arm of 7α-Dehydroxylation . 164
Table 4.5 Differential Expression of the *bai* genes in CA Induced and Non Induced Conditions at 4 Hours 168
Table 4.6 Differential Expression of the *bai* genes in CA Induced and Non Induced Conditions at 6 Hours 169
Table 4.7 Differential Expression of the *bai* genes in CA Induced and Non Induced Conditions at 8 Hours 170
Table 4.8 Differential Expression of the *bai* genes in CA Induced and Non Induced Conditions at 10 Hours 171
Table 4.9 Hierarchical Clustering Dendrogram 3 at 4 Hours 177
Table 4.10 Hierarchical Clustering Dendrogram 3 at 6 Hours 177
Table 4.11 Hierarchical Clustering Dendrogram 2 at 8 Hours 178
Table 4.12 Hierarchical Clustering Dendrogram 3 at 10 Hours 178
Table 5.1 *C. scindens* Conjugation Assay Conditions 206
Table 5.3 Methylation Patterns Detected by SMRT Sequencing 212
Table 5.4 Summary of Putative Characterised Methyltransferase Genes 214
Table 5.5 Final *C. scindens* Conjugation Assays 229

Abreviations

12-keto-DCA	12-keto-Deoxycholic Acid
3-KSD	3-keto Steroid Dehydrogenase
5mC	C5-methylcytosine
6mA	N6-methyladenine
ACE	Allele Coupled Exchange
ATCC	American Type Culture Collection
bai	Bile Acid Inducible
BLAST	Basic Local Alignment Search Tool
BSH	Bile Salt Hydrolases
CA	Cholic Acid
Ca-DPA	Calcium Dipicolonic Acid
Cas	CRISPR associated
CDI	*Clostridium difficile Infection*
CDCA	Chenodeoxycholic Acid
CDT	*Clostridium difficile* Binary Toxin
CFS	Cell Free Supernatant
CFU	Colony Forming Unit(s)
COR	Corallopyronin
CRISPR	Clustered Regularly Interspaced Palindromic Repeat
crRNA	CRISPR RNA
dam	Deoxyadenosine methyltransferase
DCA	Deoxycholic Acid
dcm	Deoxycytosine methyltransferase
ELISA	Enzyme Linked Immunosorbent Assay

EUCLID	European, multicentre, prospective, biannual, point-prevalence study of *Clostridium difficile* infection in hospitalised patients with diarrhoea
FAD	Flavin Adenine Nucleotide
FDA	Food and Drug Agency
FDX	Fidaxomicin
FMT	Faecal Microbiota Transplant
FPKM	Fragments Per Kilobase of exon per Million
GC	Glycocholic Acid
GCDCA	Glycochenodeoxycholic acid
GTD	Glucosyl Transferase Domain
Hfr	High Frequency of Recombination
hsd	Host Specificity Determinant
JCM	Japan Collection of Microorganisms
KLD	Kinase Ligase DpnI
LCA	Lithocholic Acid
MAPQ	Map Quality
MET	Metronidazole
MFS	Major Facilitator Superfamily
MIC	Minimum Inhibitory Concentration
MOPS	Morpholinepropanesulfonic Acid
MRSA	Methicillin Resistant *Staphylococcus aureus*
MTase	Methyltransferase
MYX	Myxopyronin
NTF2	Nuclear Transport Family
OD	Optical Density
PacBio	Pacific Biosciences

PaLoc	Pathogenicity Locus
PAM	Protospacer Adjacent Motif
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PMC	Pseudomembranous Colitis
pMTL	Plasmid Microbial Technology Laboratory
RAST	Rapid Annotation using Subsystems Technology
REA	Restriction Endonuclease Analysis
REase	Restriction Endonuclease
RIF	Rifampicin
RIP	Ripostatin
RM	Restriction Methylation
RNAP	RNA polymerase
RP_0	RNAP-promoter open complex
RP_C	RNAP-promoter DNA complex
rRNA	Ribosomal RNA
SASP	Small Acid Soluble Protein
SCLE	Spore Core Lytic Enzymes
SMRT	Single Molecule Real Time
SNP	Single Nucleotide Polymorphism
TC	Taurocholic Acid
TCDCA	Taurochenodeoxycholic acid
TM	Toxic Megacolon
TRAP	Tripartite ATP-independent Periplasmic
VAN	Vancomycin
VRE	Vancomycin Resistant Enterococci

Chapter 1 Introduction

1.1 The Human Microbiome

The human body plays host to trillions of microscopic organisms. Consisting of eukaryotes, archaea and viruses; bacterial species make up a diverse and dominant constituent of the human microbiome. Over the last 10 - 15 years there has been concerted efforts to characterize and improve our understanding of the complexity of the microbes inhabiting the human body. Advances in next generation sequencing technologies and metagenomic analyses have enabled the sequencing and assembly of whole genomes and subsequent 16 S based taxonomic characterization of the bacteria of the human microbiome. Isolated from human cohorts exhibiting an array of physical characteristics and health statuses, remarkable diversities between even healthy individuals have been observed, conceivably owing to differences in host environmental, dietary and genetic factors as well as primary microbial exposure. Despite these disparities, conservation of the metabolic activities of the complex microbial consortia between individuals is apparent (Huttenhower *et al*., 2012).

In 1972 it was first reported in the literature that the microbial population of the human body outnumbered our own somatic and germ cells by an order of magnitude (Luckey 1972). Sir Samurai Thomas D. Luckey derived the total number of bacterial cells in the human body from the capacity of the alimentary canal, approximately 1 litre, and a known density of 10^{11} bacteria per gram of wet content. Multiplying these figures resulted in the estimation of 10^{14}, or 100 trillion, bacterial cells. Comparison to the total number of human cells in the body, estimated to be at 10^{13}, 10 trillion cells, gave rise to the often-quoted ratio of 10:1 bacterial to human cells. Although ultimately erroneous, Thomas Luckey and his “back of the envelope” calculation can be credited with emphasizing the importance of the microbiota and its significance in health and disease, bringing this area of study to the forefront of the scientific community.

A recent revision of this seminal work by Sender *et al*., based upon current knowledge combining qualitative biology and metabolism, places the ratio closer to 1:1. The number of bacteria residing in all other compartments of the body was found to be negligible compared to that of the colon, in contrast to Luckey's original estimate which took in to account the entire volume of the alimentary canal. Citing MRI data and a number of independent studies, the colonic volume was estimated to be that of 0.4 litres and the number of bacteria residing there, 0.9×10^{11} per gram of wet content. These figures resulted in the revised estimate of approximately 3.8×10^{13} bacterial cells. By reviewing several studies, the number of human cells in the body was estimated to be 3×10^{13}, the vast majority of which are red blood cells. Using the figures of 38 trillion and 30 trillion, a ratio of about 1:1 is arrived at, with each act of defecation potentially resulting in the preponderance of human cells. Although not as remarkable as Luckey's ratio of 10:1, it should be emphasized that this does not serve to diminish the biological importance of the human microbiome (Sender *et al*., 2016). Indeed, the effect of these resident microorganisms on the human body is so profound that the bacteria of the human gut can be thought of as a "new organ" (Bocci, 1992). The multitude of functions performed by this virtual organ includes metabolising indigestible compounds, essential nutrient and vitamin provision, shaping the architecture of the intestine, immune system development and protection against pathogenic bacteria (Hooper & Gordon, 2001).

The combined metabolic effect and the complex homeostatic interactions between bacteria and host are integral to maintaining the optimal health and well-being of an individual. It is hypothesized that shifts in dietary habits over the last 100 years; an increase in the consumption of processed foods and chemicals, as well changing patterns of pharmaceutical treatment regimens (Freedberg *et al*., 2013; Buendgens *et al*., 2014), has resulted in pathological imbalances in the gut microbiota; predisposing an individual to certain ulcerative, neoplastic, inflammatory and infectious diseases (Bengmark, 1998). Pharmaceuticals such as broad-spectrum antibiotics exert a devastating effect

upon the microbial population of the human gastrointestinal tract. The resulting dysbiosis of the commensal bacteria of the gut lumen in the aftermath of these "antimicrobial A-bombs" is a key risk factor for colonisation of the enteric pathogen, *Clostridium difficile* (*C. difficile*).

1.2 *Clostridium difficile*

1.2.1 History and Phylogeny

The genus *Clostridium* represents an immensely diverse number of species. First proposed in 1880 by Adam Prazmowski, the genus has since been used to group a large number of Gram-positive, anaerobic, spore forming rods encompassing proteolytic, solventogenic, cellulolytic, acetogenic and pathogenic organisms.

As a member of the *Clostridium* genus, *C. difficile* is a Gram-positive, anaerobic, endospore forming bacillus. Originally named *Bacillus difficilis*, for the apparent difficulties encountered when trying to culture, the bacterium was first detected in the faeces of healthy neonates in 1935 by the pre-eminent American anaerobists Ivan C. Hall and Elizabeth O'Toole. Hall and O'Toole observed a strict anaerobe capable of forming subterminal elongated plectridial spores and capable of causing disease symptoms in guinea pigs and rabbits. Further studies on *B. difficilis* in 1937 confirmed the production of a thermolabile enterotoxin, lethal in the cat, dog, rat and pigeon. Observations noted that cultured clostridial isolates from infants displayed morphological, cultural and serological characteristics identical to that of *B. difficilis* (Snyder, 1937). Subsequently, *B. difficilis* was reclassified as belonging to the genus *Clostridium*, and the binomial classification of *Clostridium difficile* was established.

In 1962 *C. difficile* was isolated from patients exhibiting gas gangrene, abscesses and from the blood, pleural and peritoneal fluids of infants. Although there was no evidence that *C. difficile* was the causative agent in these cases, all isolates were found to be lethal in the guinea pig model (Smith & King, 1962).

It was 40 years after the initial isolation of the bacterium that was to become *C. difficile* that an etiological association was established between the organism and disease symptoms. In 1978, Bartlett and colleagues isolated cytopathic toxin producing *C. difficile* from patients presenting with antibiotic associated pseudomembranous colitis (PMC) (Bartlett et al., 1978).

More recently, subsequent reclassifications for *C. difficile* have been proposed based upon the observed diversity of the clostridia. It was not until the advent of molecular technologies in the 1970s and 80s (Johnson & Francis, 1975; Tanner *et al*., 1981; Woese, 1987; Ludwig et al., 1988) and the developments in 16 S ribosomal RNA (rRNA) sequencing in the 1990s (Cato & Stackbrandt, 1989; Lawson *et al*., 1993; Rainey & Stackbrandt, 1993; Stackebrandt *et al*., 1999;) that the degree of phylogenetic diversity within the genus was recognised.

In 1994, Collins and colleagues confirmed the heterogenous nature of the genus by further 16 S sequencing and comparisons within the clostridia and other species of low GC content Gram-positive genera. Taxonomic restructuring was proposed using 16 S rRNA sequences as a hierarchical structure for re-classification of species within the genus (Collins *et al*. 1994). The group containing the type species and first described of the clostridia, *Clostridium butyricum*, was designated as rRNA Cluster I, known as *Clostridium sensu stricto*, from the latin sēnsū strictō, meaning 'in the narrow sense'. Of the 238 known species of clostridia, less than 80 lies in Cluster I and can be classified as *Clostridium sensu stricto*. *C. difficile* is located in Cluster XI, shown to be phylogenetically distant from Cluster I (Figure 1.1) (Collins *et al*., 1994). Based upon these vast

differences, it was proposed in 2013 by Yutin & Galperin that the name *Peptoclostridium* be used as a genus to encompass all the organisms in Cluster XI (Yutin & Galperin, 2013), thus proposing another name change to *Peptoclostridium difficile* (*P. difficile*). However, it has been shown that Cluster XI exhibits too great a phylogenetic depth to be incorporated in to a single genus (Yarza *et al*., 2014) and the proposed name changes also do not fall in line with the Bacteriological Code. This would represent significant difficulties in altering the ubiquitous name of the world renowned 'C. diff'. To mitigate these potential problems, Lawson and Rainey have proposed the new genus name of *Clostridioides*, with *Clostridioides difficile* designated the type species (Lawson & Rainey, 2016). For the remainder of this document, it is left up to the reader to decide upon the etymology of the "C" in *C. difficile*, and what it represents.

Since the discovery of *C. difficile* in 1935, extensive study has expanded the knowledge and understanding of the diagnosis, epidemiology and pathology of *C. difficile*, which has become the pre-eminent cause of antibiotic associated diarrhoea in the developed world.

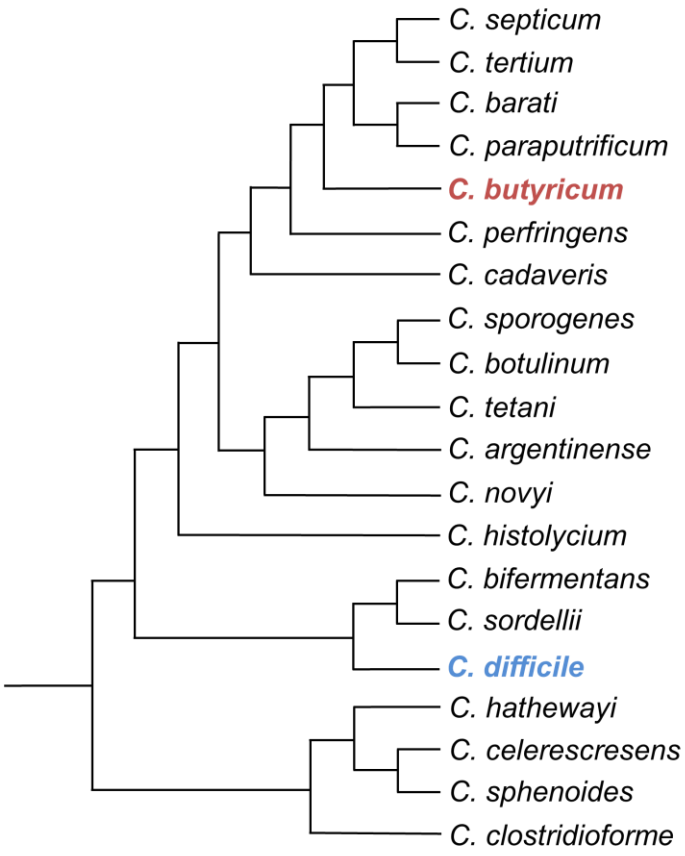


Figure 1.1 A Phylogenetic Tree of Various *Clostridium* Species: The 16 S based rRNA relationship of *C. difficile* to various other Clostridium species, showing the phylogenetic distance between *C. diffiile* and the type strain *C. butyricum* (adapted from Elsayed & Zhang, 2004)

1.2.2 Epidemiology of Clostridium difficile *Infection*

For epidemiological purposes, various typing methods for *C. difficile* isolates have been developed, the most commonly used being polymerase chain reaction (PCR) ribotyping. This technique relies on the banding patterns observed from PCRs performed with primer pairs complementary to the 16 S and 23 S subunits of rRNA. Identifying banding patterns are based upon inter-strain diversity in copy number of the rRNA operon and the polymorphisms in the intergenic spacer region between the 16 S and 23 S subunits of rRNA, with higher copy numbers resulting in the potential for more diversity in these spacer regions (Gürtler, 1993; Bidet *et al*, 1999).

Over the past three decades, the epidemiological landscape of *C. difficile* has evolved, with more severe and frequent cases in healthcare facilities and clear changes in clinical presentation, responsiveness to treatment and disease outcome. As of 2010, *C. difficile* has overtaken Methicillin-resistant *Staphylococcus aureus* (MRSA) as the primary cause of healthcare acquired infections in the United States (Miller *et al.,* 2011), and cases of community acquired CDI being increasingly reported amongst groups outside of the healthcare setting, lacking the traditional risk factors associated with the disease (Khanna & Pardi, 2010).

The increase in cases and disease severity appears to be correlated with the emergence of so called "hyper virulent" strains, classified as PCR ribotype 027, also characterised as restriction endonuclease analysis BI (REA BI) and North American pulsed-field type 1 (NAP-1), (Cartman *et al.,* 2010).

As well as carrying the genes for toxin A (TcdA) and toxin B (TcdB) on the pathogenicity locus (PaLoc), these epidemic strains also produce a third, binary toxin (CDT), which is thought to contribute to the disease severity, possibly as a colonisation factor by promoting adherence to epithelial cells (Gerding *et al*., 2014). Ribotype 027 strains are

also reported to be highly resistant to fluoroquinolones due to mutations in the *gyrA* gene (Drudy, 2007), and produce more toxin due to a frameshift mutation in the *tcdC* toxin repressor gene (Warny *et al*., 2005; Dupuy *et al.,* 2008). It has also been reported that variations in the *tcdB* sequence in ribotype 027 strains are responsible for encoding a more potent, hypertoxic variant of TcdB (Lanis *et al*., 2013).

The first recorded incidence of a PCR ribotype 027 *C. difficile* strain was in 1985 in Paris (Popoff *et al*., 1988). Popoff and colleagues isolated a strain of *C. difficile* from a hospitalised patient suffering with PMC. The causative strain was found to produce a binary toxin with actin ADP-ribosyltransferase activity, as well as the entero and cytotoxins, TcdA and TcdB. This strain was designated CD196 and retrospectively represents a historical non-epidemic isolate of the ribotype 027 lineage. The first cases of CDI associated with a ribotype 027 strain in the UK were the Stoke Mandeville Hospital outbreaks between 2003 and 2006. This strain, designated R20291, was associated with 498 cases of CDI and 128 deaths (Health Care Commission, 2007). The epidemic R20291 strain and the historical, non-epidemic CD196 strain have afforded the opportunity to study the evolution of the ribotype 027 lineage. Genomic comparisons between these two strains have revealed five genetic regions in the R20291 genome not present in CD196. It is hypothesised that these newly acquired genes are the reason for the observed disparity in virulence between these two strains (Stabler *et al*., 2009). It was also reported by Stabler and colleagues that there is significant genetic diversity between the studied ribotype 027 strains and *C. difficile* 630, a ribotype 012 strain, thought to be a factor in the phenotypic differences and disease severities between the two ribotypes.

The rise in incidences of CDI can also be attributed to the fluidity of the *C. difficile* genome. Further studies on the genomes of CD196, R20291 and 630 reveal the accumulation of inversions and point mutations, thought to have been acquired in response to changes in environment and human activities, showing the continued evolution of *C. difficile* strains (Stabler *et al*., 2010).

The recent EUCLID (European Multicentre Prospective Biannual Point Prevalence Study of *Clostridium difficile* Infection in Hospitalized Patients with Diarrhoea) study found that the appearance of ribotype 027 strains is inversely correlated with the level of testing. This suggests that a higher level of testing is indicative of increased awareness and infection control, resulting in a reduction of such outbreaks and an overall decrease in ribotype 027 incidences in Europe. Interestingly, the study has also revealed a more diverse pool of strains across Europe, increasing from 65 to 138 different ribotypes in 2014, with no single strain dominating in 14 out of the 20 participating countries (Davies *et al*., 2014).

Work conducted in the Netherlands indicates the emergence of a new "hypervirulent" strain ribotype 078. The toxins, TcdA and TcdB are present, along withthe CDT binary toxin (Goorhuis *et al*., 2008). Ribotype 078 strains have also been shown to harbour a novel transposon, further highlighting the fluid nature of the *C. difficile* genome (Corver *et al*., 2012). These 078 strains are routinely isolated from pigs, presenting a possible reservoir for community acquired CDI and a potential food borne disease hypothesis (Hensgen *et al.*, 2012).

1.2.3 Clostridium difficile *Toxins*

The symptoms of *C. difficile* infection (CDI) range from asymptomatic colonisation and mild self-limiting diarrhoea to more severe symptoms, leading to the potentially fatal intestinal complications, pseudomembranous colitis (PMC) and toxic megacolon (TM). This pathology is principally due to the two main virulence factors, toxin A (TcdA) and toxin B (TcdB). These toxins are part of a group of structurally and functionally related toxins, the large clostridial cytotoxins.

TcdA and TcdB are encoded on the 19.6 Kb chromosomal Pathogenicity Locus (PaLoc) (Figure 1.2) found in toxigenic strains of *C. difficile*, occupying large open reading frames of 8,133 and 7,098 base pairs (bp) respectively (Voth & Ballard, 2005). Three other open reading frames are situated on the pathogenicity locus, *tcdR*, *tcdE* and *tcdC* (Just & Gerhard, 2004). The *tcdR* gene is located upstream of *tcdB* and encodes an alternative RNA polymerase sigma factor, responsible for directly activating expression of toxin genes in response to certain environmental stimuli such as temperature and nutrient conditions (Mani & Dupuy, 2001). A putative phage-like holin protein is encoded by *tcdE*, which is thought to have a role in mediating the extracellular release of TcdA and TcdB via permeabilisation of the bacterial cytoplasmic membrane (Govind & Dupuy, 2012; Govind *et al*., 2015). Although there is also evidence to dispute this hypothesis (Olling *et al*., 2011).

The farthest downstream gene, *tcdC* is transcribed divergently to the other genes in the locus and is hypothesised to encode a repressor of toxin expression through interference with the *tcdR*-RNA polymerase complex (Carter *et al*., 2011). The TcdC repressor is highly active during phases of exponential cell growth, where it represses toxin production, shunting metabolic focus to cellular processes involved in growth (Dupuy *et al.*, 2008). However further work also suggests a disassociation between *tcdC* genotype and toxin production (Cartman *et al.,* 2012). The function of TcdC therefore remains unclear.

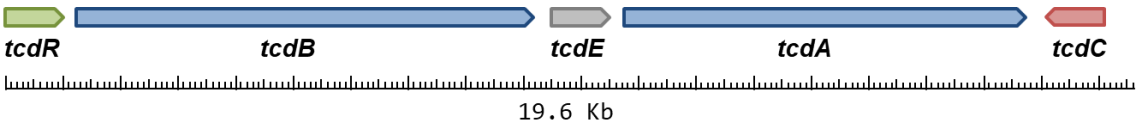


Figure 1.2 The Pathogenicity Locus of *C. difficile*: The 19.6 PaLoc of *C. difficile* encompassing the *tcdR* activator, *tcdE* holin, *tcdC* repressor and toxins A and B, *tcdA* and *tcdB*.

TcdA and TcdB can be structurally divided in to two sections, the catalytic subunit responsible for enzymatic activity and the translocation subunit, which is involved in delivering the catalytic subunit in to target cells. The C termini of both TcdA and TcdB contain multiple combined repetitive oligopeptides enabling the toxins to bind to sugar moieties on the surface of target cells. After binding, the toxins are internalised via clathrin mediated endocytosis (Just & Gerhard, 2004). Endosomal acidification in the cytoplasm of the target cell induces conformational changes in the delivery/pore forming domain, revealing hydrophobic regions which insert in to the membrane of the endosome, forming pores. This allows the glucosyltransferase domain (GTD) to protrude in to the cytosol, where, mediated by host cell cytosolic factors, the autoproteolytic domain located adjacent to the GTD cleaves the catalytic domain, releasing it in to the cytosol (Rupnik *et al.*, 2009). Once free of the endosome, the GTD inactivates Rho GTPases by monoglucosylation. Glucosylation of the membrane associated Rho GTPases disrupts a multitude of downstream signalling pathways, resulting in cytoskeletal rearrangement, cell rounding and subsequent apoptosis (Just & Gerhard, 2004). In CDI, it is the necrosis of epithelial cells in the colon which causes fluid secretion and inflammation, resulting in the characteristics symptoms of the disease. Although previously reported that only toxin B is essential for virulence (Lyras *et al*., 2009), it is now accepted that both toxins A and B are important in *C. difficile* associated disease (Kuehne *et al*., 2010; Kuehne *et al.,* 2013).

The CDT binary toxin is present in various strains, including ribotype 027, is unrelated to TcdA and TcdB, and belongs to the family of clostridial binary toxins. CDT is encoded on the chromosomal 6.2 Kb CDT locus which is comprised of three genes, *cdtA*, *cdtB* and *cdtR*. These genes encode for CdtA the catalytic domain, CdtB the translocation and binding domain and CdtR the response regulator, respectively. During infection, CdtB binds to the target cell wall via the lipolysis-stimulated lipoprotein receptor

(Papatheodorou *et al.*, 2011). This allows the translocation of the catalytic component, CdtA, in to the cytosol, where ADP-ribosylation of actin, resulting in cytoskeletal disruption and eventual cell death (Gerding *et al*., 2014).

1.2.4 Clostridium difficile *Sporulation*

A key element in the successful infection cycle of *C. difficile* is the ability of the bacterium to sporulate. This process of cell differentiation is an ancient adaptation employed by members of the genera *Clostridium* and *Bacillus* as well as other members of the phylum of low GC content, Gram-positive organisms; the Firmicutes. The most well studied and characterised of the spore forming bacteria is *Bacillus subtilis* (*B. subtilis*), from which a paradigm has been generated to aid in the elucidation of the sporulation pathway in other spore forming bacteria, such as *C. difficile*.

The factors which initiate sporulation remain largely unknown, but it is probable that nutrient starvation and the consequential inability to maintain vegetative cell growth is a significant trigger (Setlow, 2003). Other factors may involve quorum sensing or external stress factors. In response to these stimuli, the vegetative cell switches from binary fission to asymmetric growth. This polar cell division results in the formation of two compartments, the mother cell and the smaller forespore. Once DNA segregation is completed, the larger mother cell engulfs the forespore, forming the spore cortex and the inner and outer coats. The process of sporulation concludes, when the mother cell lyses, releasing the mature spore, (Figure 5).

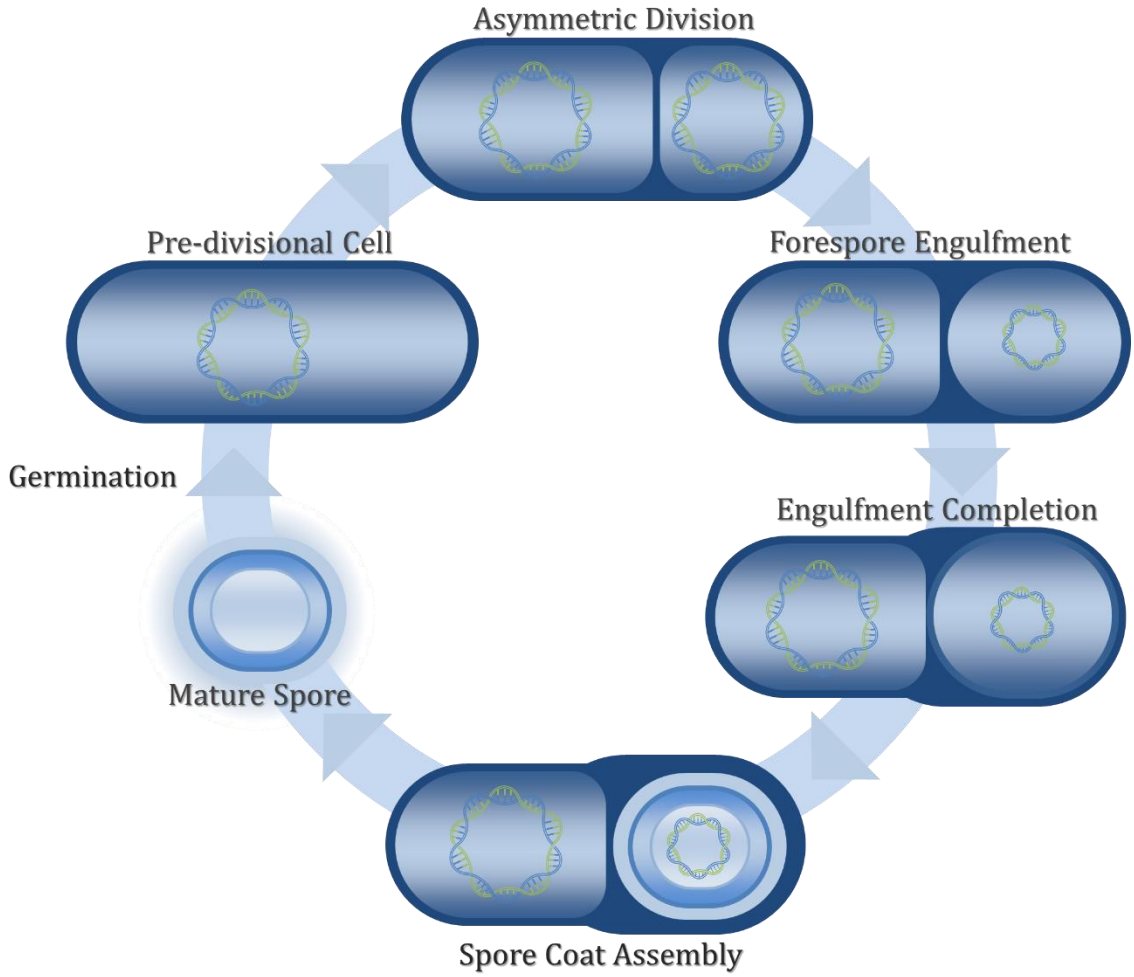


Figure 1.1: The Process of *C. difficile* Sporulation: The vegetative cell initiates sporulation in periods of stress. DNA is replicated and segregated in asymmetric cell division in to the mother cell and forespore. The Forespore compartment is engulfed by the mother cell, leading to the specialised differentiation and construction of the spore coat layers, preceding mother cell death and spore release.

B. subtilis initiates sporulation via a complex variation of two component signal transduction, which is responsible for the autophosphorylation and activation of the master transcriptional regulator of sporulation, Spo0A. This signal transduction system, termed the sporulation phosphorelay, is composed of five sensory histidine kinases: KinA, B, C, D and E (Higgins & Dworkin, 2013). The genes for this phosphorelay are not conserved in *C. difficile*, however bioinformatic approaches have identified five orphan histidine kinases which are likely to be involved in the cellular decision to enter sporulation. These orphan kinases are encoded at the gene loci CD1352, CD1492,

CD1579, CD1949 and CD2492 in the genome of *C. difficile* strain 630 (Underwood *et al.*, 2009).

To date, three of these orphan kinases have been characterised in strain 630; CD1492, CD2492 and CD1579. By measuring the incorporation of radioactive phosphate, Underwood and colleagues revealed that the cytosolic CD1579 directly phosphorylates the master regulator Spo0A *in vitro*. Due to the membrane spanning nature of CD1492 and CD2492, these histidine kinases are less amenable to cloning, overexpression and purification, but the conservation of certain key amino acid residues in these proteins, suggests that CD1492 and CD2492 are also able to phosphorylate Spo0A with specificity comparable to CD1579. Insertional inactivation of CD2492 resulted in a 3.5-fold decrease in sporulation *in vitro* and in mutants with inactivated Spo0A, sporulation is completely abolished (Underwood *et al*., 2009). These findings suggest a number of histidine kinases are responsible for the activation of Spo0A, which occurs by direct phosphorylation. Sporulation in *C. difficile* therefore deviates from the proposed *B. subtilis* model, likely due to the evolutionary divergence of the two genera.

Although there is no conservation of genes upstream of Spo0A, both *C. difficile* and *B. subtilis* employ the same regulatory network of four specific sigma factors. The morphological changes which occur during sporulation are accompanied by the spatial and sequential activation of specific transcription regulators, the RNA polymerase sigma factors, σF, σE, σG and σK. This intracellular signalling cascade is responsible for the activation of the transcription of the genes required for spore morphogenesis (Fimlaid *et al.*, 2013).

Morphological characterisation of mutants in the *sig* genes encoding for these transcription regulators and the use of fluorescent markers has highlighted important differences from the *B. subtilis* model of sporulation (Pereira *et al.*, 2013). The temporal activities of these specific sigma factors during the developmental events are conserved

between *C. difficile* and *B. subtilis*, with σF and σE mediating early gene expression and σG and σK controlling activities at later stages. However, with *C. difficile* sporulation, there is less of an association between the sigma factors than with the *B. subtilis* model, proposing less intercompartmental communication between sigma factors in the mother cell and forespore.

1.2.5 Clostridium difficile *Spores*

Sporulation concludes with the formation of the bacterial endospore. The intrinsic resistance of these spores, enables the persistence and ability to withstand extreme environmental conditions in which the vulnerable vegetative cell would perish, making it an ideal agent of infection.

Multiple structural layers and factors contribute to the resistance properties of the spore. The low water content, high levels of calcium chelated dipicolonic acid (Ca-DPA) and small acid soluble proteins (SASPs) which saturate and protect the spore DNA (Raju *et al.*, 2006). The inner most compartment is the core, which houses DNA, RNA and enzymes. The core is surrounded by the inner membrane and the mother cell derived germ cell wall. The inner membrane is impermeable to small molecules, including water, preventing the infiltration of genotoxic molecules (Figure 1.2)

The germ cell wall is coated by the spore cortex. The cortex is composed of a specialised peptidoglycan, unique from that of peptidoglycan found in the germ cell wall. The cortex layer is surrounded by the mother cell derived outer membrane and the spore coat. The proteinaceous coat layer is composed of a multitude of proteins unique to the spore and is critical for spore resistance. Spores of *C. difficile* and *B. subtilis* share structural similarities but only 18 orthologues of the 70 *B. subtilis* spore proteins have been found in *C. difficile*, this suggests differences in coat composition between the two genera

(Henriques & Moran, 2007). Recently, six genes encoding for spore coat proteins were identified; *cotA*, *cotB*, *cotCB*, *cotD*, *cotE* and *sodA* (Permpoonpattana *et al.*, 2011). Mutation of the *cotA* gene resulted in the production of malformed spores, indicating the importance of the CotA protein in coat assembly and stabilisation. The rest of these coat proteins are predicted to have enzymatic activity, three of which have been confirmed: CotD, a manganese catalase, SodA, a superoxide dismutase and CotE with bifunctional enzymatic peroxiredoxin and chitinase activities. (Permpoonpattana *et al.*, 2013)

The outermost coat of the *C. difficile* spore is the exosporium. The composition of this is hair-like layer is unclear, but genes paralogous to the *Bacillus anthracis* collagen-like glycoproteins have been identified in the *C. difficile* strain 630 genome, although the function of these proteins in the exosporium is uncertain (Pizarro-Guajardo *et al.*, 2014). It has been hypothesised that the exosporium could be important in pathogenicity, mediating spore surface interactions with host.

The resilience conferred by the structure of the spore allows the organism to lay dormant until conditions favourable for the spore to germinate are encountered.

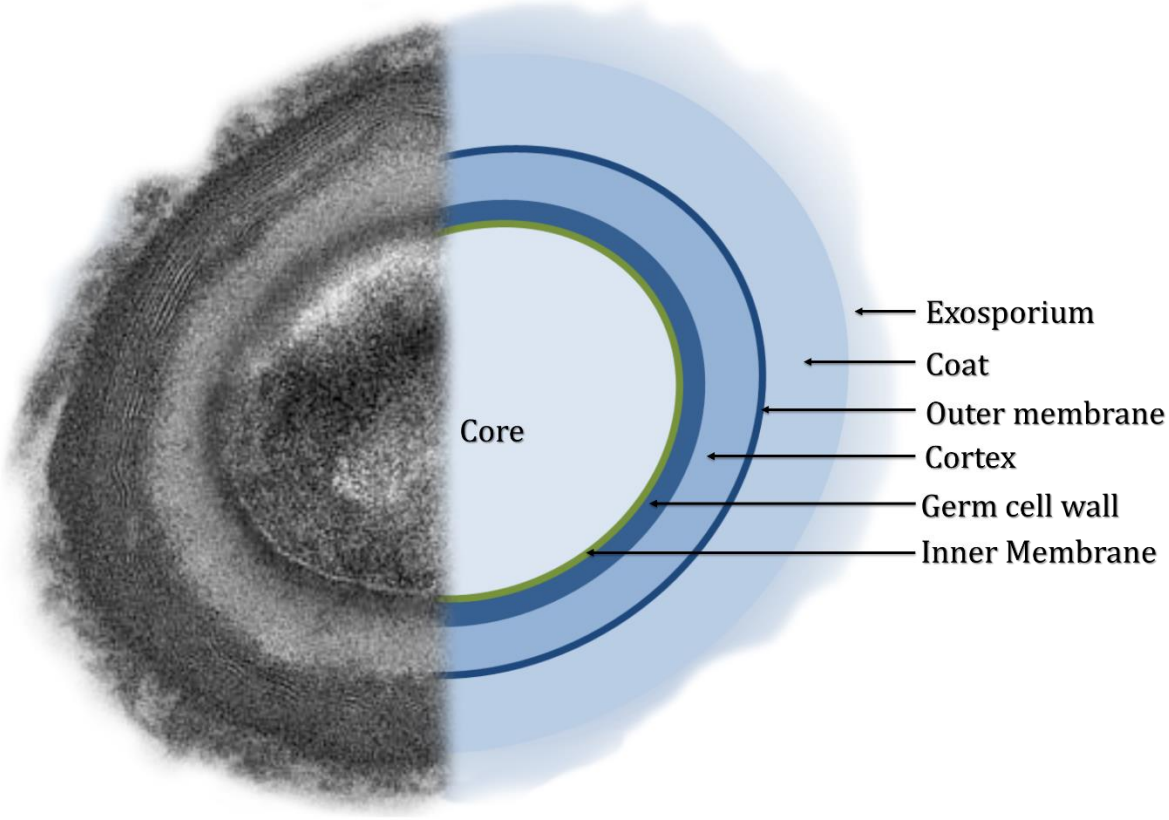


Figure 1.2 The Structure of the *C. difficile* Spore: A transmission electron micrograph of a *C. difficile* spore showing the core, inner membrane, germ cell wall, cortex, outer membrane, coat and exosporium layers. TEM adapted from Permpoonpattana *et al.*, 2013.

1.2.6 Clostridium difficile *Spore Germination*

Germination is the process resulting in the return of the spore to a vegetative cell. As the detrimental properties associated with CDI are only exhibited during vegetative cell growth, germination is a key event in the *C. difficile* infection cycle. Once ingested, the highly resistant spores are able to survive passage through the hostile, acidic environment of the stomach and pass in to the gut, where germination occurs. These dormant spores sense and respond to their environment by interacting with specific small molecule ligands, called germinants. The binding of germinants to germination receptors located on the inner spore membrane, results in the initiation of germination, (Hudson, *et al.,* 2001). The process of germination is largely conserved amongst spore forming bacteria; however, the means by which germination is initiated differs. As with sporulation, knowledge of *C. difficile* germination is based upon the study of *B. subtilis*, for which the germination receptors GerA, GerB and GerK have been characterised (Paidhungat & Setlow 2000). These receptors are encoded by the genes *gerA*, *gerB* and *gerK*, which reside on three tricistronic operons, collectively known as the *gerA* family of operons. Bioinformatic analysis has determined that there are no homologues of the *gerA* family present in the genome of *C. difficile* 630 (Paredes-Sabja *et al.,* 2011, Sebaihia *et al*., 2006). Therefore, spores of *C. difficile* must utilise a unique set of receptors and germinants in the recognition of germination suitable conditions.

The nutrient germinants recognised by the receptors of *B. subtilis* include L-alanine or a combination of L-asparagine, glucose, fructose and potassium ions. Germination is not only initiated by the binding of endogenous agonists, it may also be triggered by mechanical treatments, such as high pressures and abrasion (Black *et al.*, 2005). Sorg and Sonenshein found that the conjugated bile acid, taurocholic acid (TC), is a germinant for *C. difficile* spores, with the amino acid glycine as a co-germinant *in vitro* (Sorg & Sonenshein, 2008). Although it should be noted that the specific response to these

germinants is strain specific (Heeg & Burns, 2012). More recently, in 2013, Francis and colleagues used chemical mutagenesis and parallel genome sequencing to identify SNPs which result in germination phenotypes in *C. difficile*. It was found that mutations in the *cspC* gene abolished the initiation of spore germination in response to bile acids. It can therefore be hypothesised that CspC is a bile acid germinant receptor in *C. difficile* spores (Francis *et al.*, 2013).

Once a germination receptor binds a germinant, the spore ceases dormancy and irreversibly proceeds through germination. The events of these initial stages are comprised of the release of monovalent cations; H^+, Na^+ and K^+ and Ca-DPA (Setlow, 2003). Following this is the depolymerisation of the spore cortex peptidoglycan by spore core lytic enzymes, (SCLEs). SCLEs have been identified in *B. subtilis*, CwlJ and SleB, and in *Clostridium perfringens*, SleC and SleM (Galperin *et al.*, 2012). Homologues of the genes encoding for CwlJ, SleB and SleC, have been found in the *C. difficile* genome. The SCLE, SleC, has been found essential for nutrient mediated germination of *C. difficile* spores (Burns *et al.*, 2010) and subsequent return to vegetative growth.

The unique structure of the peptidoglycan which composes the spore cortex facilitates recognition by the SCLEs, leaving the germ cell wall intact. During the degradative step of core hydrolysis, the spore is rehydrated to the water content levels of a vegetative cell. Following this rehydration, cellular metabolism resumes, and the germinated spore grows out in to the vegetative cell.

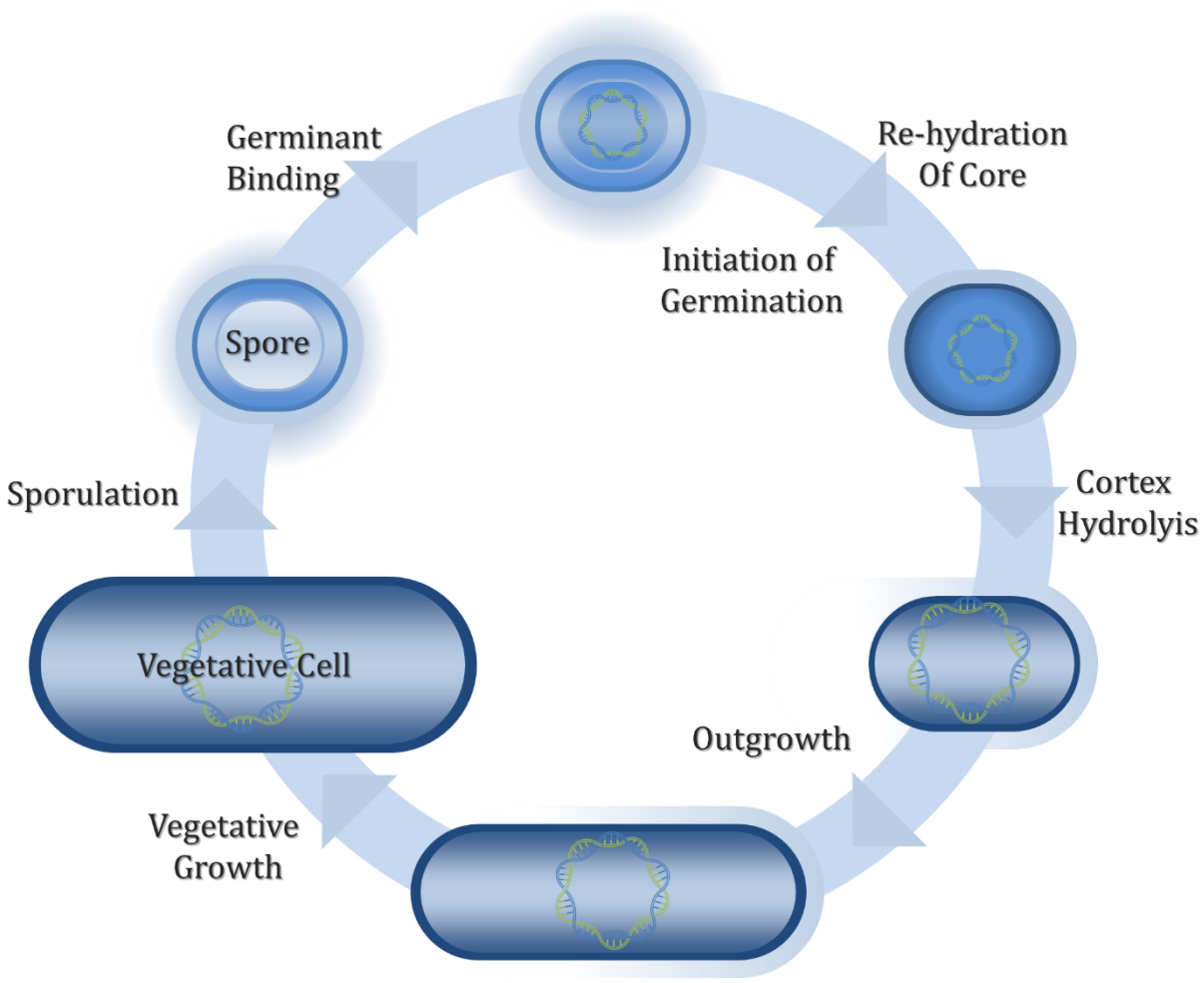


Figure 1.3 ***C. difficile*** **Spore Germination and Outgrowth:** The spore senses and responds to favourable conditions by the binding of a germinant to a specific germinant receptor. The core becomes rehydrated concomitant with the release of Ca-DPA. Spore cortex lytic enzymes degrade the cortex, allowing the spore to enter outgrowth and return to vegetative cell growth.

1.2.5 Clostridium difficile *Treatment*

Treatment for CDI has essentially remained the same for the last 30 years, largely due to the fact that no significant resistance has developed to the two main drugs of choice, metronidazole and vancomycin. However, there is a high potential for resistance to develop, with a large portion of the *C. difficile* genome consisting of mobile genetic elements, mainly conjugative transposons, which are often implicated in the transfer of genes for antimicrobial resistance mechanisms (Valiente *at al.*, 2014). Indeed, growing resistance to macrolides, fluoroquinolones and other common antibiotics has been observed in the epidemic, ribotype 027 strains (Tenover *et al*., 2011)

For mild cases of CDI, metronidazole is the preferred choice for treatment. This is mainly due to the cost of the drug, 57p per treatment regimen in contrast to £60 for vancomycin therapy as of 2014. The other perceived disadvantage to treatment with vancomycin is the selection of intestinal vancomycin resistant enterococci (VRE). Although data suggests that both drugs select for VREs and that the risk factors associated with CDI also promote intestinal colonisation by VREs and other multidrug resistant organisms such as MRSA (Fujitani *et al.*, 2011).

For more severe disease, vancomycin is the recommended starting therapy. Vancomycin has ideal pharmacokinetics for treatment of infection localised to the gut lumen. Levels of vancomycin reach about 1000 μg/mL in the stool as a result of the poor absorption of the drug from the colon. Vancomycin has been found to be significantly more effective than metronidazole at removing *C. difficile* from the stool as indicated by *C. difficile* density culture (Al-Nassir *et al.*, 2008).

Relapse has been reported after treatment with both metronidazole and vancomycin possibly due to deleterious effects upon the normal microbiota by these drugs (Lewis *et al.,* 2015). Relapse rates appear to be rising, with 28.9 % of patients over 65 suffering

relapses in 1991, increasing to 58.4 % in 2004 in Quebec, Canada (Pepin *et al*., 2005). Relapse of infection has been shown the be more likely in patients initially infected with a PCR ribotype 027 (Marsh *et al*., 2012).

A relatively new, promising antimicrobial for the treatment of CDI is fidaxomicin (FDX). FDX was approved for use by the US Food and Drug Agency (FDA) in 2011 and data from a number of studies shows that the drug is more effective in reducing second recurrence of CDI, four weeks post treatment (Drekonja *et al.*, 2011, Cornely *et al.*, 2012). The main disadvantage of FDX therapy is the cost, which is estimated at > $2,800 per treatment; however, savings are recouped in the reduction of recurrence, avoidance of government-imposed fines and freeing up beds for other treatments, (Goldenberg *et al.*, 2016).

Due to the growing poor response of CDI to conventional antimicrobial chemotherapy and the emergence of epidemic strains, other non-antibiotic treatment options are being investigated, such as intravenous immunoglobulins, probiotics and faecal microbiota transplant (FMT) (Cocanour, 2011). Whilst efficacy data on probiotics and immunoglobulin therapies is limited (Surawicz *et al.*, 2013), FMT therapy has shown promise in the treatment of obstinate CDI. FMT which was first described in 1958, (Eiseman *et al*., 1958), aims to restore the microbial balance of the gut with faecal suspensions from healthy donors (Youngster *et al.,* 2014; McCune *et al.,* 2014; Weingarden *et al.,* 2015). A review of 27 cases of obstinate and recurring CDI found that out of 317 patients, 92 % of these saw a resolution of the disease through FMT (Gough *et al.,* 2011).

A more attractive and elegant form of this treatment would involve the characterisation of members of the gut microbiota which confer colonisation resistance against *C. difficile*. Beneficial components of the microbiota could then be administered to patients to restore a healthy homeostasis for therapeutic benefit. Trevor Lawley and colleagues developed a mouse model of dysbiosis allowing the identification of six phylogenetically

diverse bacteria of the murine gut. When these bacteria were administered to mice with CDI, the pathogenic dysbiosis was disrupted enabling the regrowth of commensal bacteria associated with a healthy intestinal homeostasis and subsequent clearing of *C. difficile* (Lawley *et al*., 2012).

1.3 *Clostridium scindens*

1.3.1 History and Discovery

As a member of the clostridia, *Clostridium scindens* (*C. scindens*) is a Gram-positive, anaerobic, endospore forming bacillus. Before being formally identified, the organism's characteristic steroid desmolytic metabolism was observed and reported in early studies. In 1957, Nabarro and colleagues noticed increased levels of excreted 17-keto steroids in patients receiving cortisol rectal infusion treatment for ulcerative colitis (Nabarro *et al.*, 1957). Nabarro hypothesised that this was due to bacterial modification of hydrocortisone in the colon, converting the C21 hydrocortisone to C19 androgen steroids. Later, in 1959, Wade *et al*. further implicated the bacterial residents of the gut in the transformation of cortisol, when a reduction in the levels of excreted 17-keto steroids was observed after oral administration of neomycin (Wade *et al*., 1959).

In 1971 Gustafsson and Eriksson detected discrepancies in excreted steroids amongst germ free and conventional rats, corroborating the hypothesis of the participation of gut microorganisms in these transformations (Eriksson & Gustafsson, 1971). In 1981, Bokkenheuser and colleagues isolated an organism from the human faeces capable of cleaving the side chain of corticoids, (Cerone-McLernon *at al.*, 1981). Through conventional microbiological examination, this desmolase synthesising bacterium was

characterised as belonging to the genus *Clostridium*, designated Strain 19. Finally, in 1985, the name *Clostridium scindens* was proposed, from the Latin scindo meaning to split or dissect, a reference to the characteristic desmolytic activity which lead to the bacterium's discovery (Morris *et al.*, 1985).

1.3.2 Clostridium scindens *Bile Acid Metabolism*

In addition to the aforementioned desmolytic activity, *C. scindens* also possesses the ability to transform primary bile acids in to secondary bile acids, through the process of 7α/β-dehydroxylation. This bio-transformation occurs via a complex, multistep enzymatic pathway involving the gene products of the bile acid inducible (*bai*) operon. Eight open reading frames are localised on the polycistronic 12.2 Kb *bai* operon, as well as additional monocistronic 3α hydroxysteroid dehydrogenase *baiA* genes (Ridlon *et al*., 2016) (Figure 1.4).

The oxidative steps of the 7α/β-dehydroxylation pathway and the functions of the participating gene products were elucidated by the cloning and expression of the genes from *Eubacterium* sp. strain VPI 12708 (Mallonee & Hylemon 1996). However, 16 S rDNA sequence analysis revealed that this bacterium belongs to genus *Clostridium*, phylogentically located in cluster XIVa. Strain VPI 12708 was formally identified as belonging to the *C. scindens* species on the basis of DNA-DNA homology values to *C. scindens* strain JCM 6567 of over 70 %, indicating a common lineage (Kitahara *et al.*, 2000). Whilst the oxidative arm of the bile transformative pathway has been elucidated, the reductive steps remain unknown (Table 1.1).

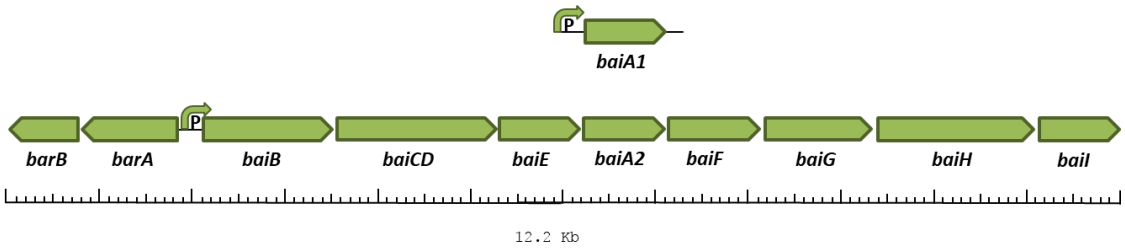


Figure 1.4 The Organisation of the *bai* Operon in *C. scindens* ATCC 35704: The 12.2 Kb polycistronic bai operon and monocistronic *baiA1*, encoding the genes of the oxidative arm of the 7α/β-dehydroxylation pathway (adapted from Ridlon *et al.,* 2016).

Table 1.1 The Functions of the Genes of the *bai* Operon

Gene	Function
barA/barB	Bile Acid Response Regulator
baiB	Bile acid Co-enzyme A Ligase
baiCD	7α-Hydroxysteroid Dehydrogenase
baiE	7α-Dehydratase
baiA1/A2	3α-Hydroxysteroid Dehydrogenase
baiF	Co-enzyme A Transferase
baiG	H^+ Dependent Bile Acid Transporter
baiH	7β-Hydroxysteroid Dehydrogenase
baiI	7β -Dehydratase

The 7α/β-dehydroxylation pathway results in the transformation of the primary bile acids cholic acid (CA) and chenodeoxycholic (CDCA) acid to the secondary bile acids deoxycholic acid (DCA) and lithocholic acid (LCA), respectively. A more detailed explanation of the pathway can be found in 4.1.3.

1.3.3 Consequences of Bile Acid Transformations

It has been observed that the secondary bile acid DCA, induces the germination of *C. difficile* spores *in vitro* (Sorg & Sonenshein, 2008). However, in the presence of this secondary bile acid, outgrowth of the germinated spores is inhibited. These vulnerable pre-vegetative cells halted by the bacteriostatic effects of deoxycholic acid eventually perish, unable to colonise and establish infection. Inhibition of vegetative cell growth may be due to the hostile environment generated by toxic bile compounds. Many bacterial species are sensitive to bile acids and bile salts (Begley *et al.*, 2005). It has been found that the hydrophobic bile salts, taurodeoxycholic acid and deoxycholic acid are significantly more bacteriostatic than that of the hydrophilic taurocholic acid and chenodeoxycholic acid. This signifies the importance of the hydrophobicity of these molecules with respect to antimicrobial activity (Sung *et al.*, 1993).

It is thought that *C. difficile* has evolved mechanisms to contend with the toxic effects of certain bile acids, such as taurocholic acid, which is exploited as a germination factor (Sorg & Sonenshein, 2008). It has been suggested that *Lactobacillus* sp. employ bile salt hydrolases (BSH) to detoxify bile salts (Alander *et al.*, 1999). Disputing this, a dissociation between the ability to withstand the toxicity of bile salts and the expression of such BSHs in Lactobacilli has also been proposed, which is suggestive of species and strain specific mechanisms of resistance to bile salts. It is therefore an important survival factor for microorganisms in the mammalian gut to resist the toxic effects of bile acids and bile salts (Moser & Savage, 2001).

Antibiotic associated perturbations in murine microbiota, leading to alterations of the bile acid pool, has been shown to affect susceptibility to CDI (Buffie *et al*., 2014, Theriot *et al.,* 2015). Buffie and colleagues administered the antibiotics clindamycin, ampicillin and enrofloxacin, resulting in shifts in the diversity of the murine gut population and

consequently, differential colonisation resistance in mice challenged with *C. difficile* spores. *C. scindens* was identified as a candidate for colonisation resistance by correlating this diversity with susceptibility to *C. difficile*. Enrofloxacin had the least detrimental effect on *C. difficile* colonisation resistance, with ampicillin and clindamycin having the most pronounced effects. This indicates the sensitivity of *C. scindens* to antibiotics of the penicillin and lincosamide classes, which have been implicated as risk factors for CDI (Dene` ve *et al*., 2008).

This association of *C. scindens* and *C. difficile* colonisation resistance was further reinforced when Buffie and colleagues observed enhanced resistance to CDI in the mouse model with the probiotic administration of *C. scindens* (Buffie *et al*., 2014)

It has also been observed that the addition of *C. scindens* to a defined microbial consortia incapable of 7α/β-dehydroxylation resulted in a significant reduction in *C. difficile* pathogenesis, associated with microbial alterations to the murine bile acid pool (Studer *et al.,* 2016).

Microbial derived secondary bile acids have been shown to have a significant effect upon the germination, outgrowth, vegetative growth and toxin production in a panel of clinically relevant *C. difficile* strains, further implicating the importance of the bile acid pool in CDI (Thanissery *et al*., 2017).

To further highlight the importance of *C. scindens* in the healthy gut, high resolution profiling of the gut microbiome and correlation analysis has identified consistent negative associations between *C. difficile* colonisation and the presence of *C. scindens* in multiple patient populations in the community (Daquigan *et al.,* 2017).

The bio-transformative activity of *C. scindens* may therefore have the potential to form the basis of a model of colonisation resistance against *C. difficile*. It can be hypothesised that the presence of *C. scindens* in the gut of healthy individuals provides colonisation resistance against *C. difficile* through the capacity of *C. scindens* to generate secondary

bile acids by the uptake and transformation of free, unconjugated primary bile acids. *C. scindens* actively transports primary bile acids and through 7α/β-dehydroxylation, yields secondary bile acids, which are then exported in to the extracellular environment of the large bowel.

These bile acid transformative mechanisms can be strongly implicated in obstinate recurrent CDI, whereby the protective gut flora is unable to recover after antimicrobial treatment. *C. difficile* spores persist in the gut, germinate and grow out unimpeded due to the low levels of 7α/β-dehydroxylating bacteria, such as *C. scindens*.

1.4 Aims of the Project

The previously detailed work emphasizing the importance of *C. scindens* in colonisation resistance against *C. difficile* has significant implications for future CDI therapies; both in terms of the rational design or indication of narrow spectrum antimicrobials and the potential to utilise *C. scindens* as a live biotherapeutic. Based upon this, the aims of the project are to investigate these two approaches to controlling *C. difficile* colonisation of the human gut.

The effects of the novel antimicrobial, fidaxomicin, on spore germination and outgrowth will be investigated in six clinically relevant strains of *C. difficile* to establish at which point during the processes of germination and outgrowth inhibition occurs. Mutants with reduced susceptibilities to the drug will be created in the epidemic R20291 strain using allelic exchange mutagenesis. These mutations will be characterised in terms of growth, sporulation, germination, toxin production and lethality in the hamster model of infection, in comparison to the wild type strain, to assess the clinical significance of strains with reduced susceptibility to FDX.

The effect of *C. scindens* derived secondary bile acids on *C. difficile* germination and outgrowth will be investigated in 6 clinically relevant strains, to reinforce the link between *C. scindens* 7α -dehydroxylation and colonisation resistance against *C. difficile* in the gut. To further study the pathway and associated genes, the transcriptomics of *C. scindens* during bile acid induction and metabolism will be investigated. This will allow the identification of possible gene candidates for the reductive arm of the 7α-dehydroxylation pathway.

To further study the *C. scindens* mediated colonisation resistance against *C. difficile*, it is important to establish tools for genetic manipulation in this organism. To do this, efforts to establish DNA transfer in *C. scindens* was undertaken. The genome sequence and methylome status of *C. scindens* strain ATCC 35704 will be determined to aid in the

identification of significant barriers to gene transfer, enabling the development of a strategy to circumvent these barriers.

Chapter 2 Materials and Methods

2.1 Bacterial Strains and Plasmids

2.1.1 Bacterial Strains

A list of bacterial strains used in this work are summarised in table 2.1, below.

Table 2.1 Bacterial Strains

Strain	Description	Source
C. difficile R20291 *ΔpryE*	Ribotype 027 *pyrE* deficient derivative	Ng *et al.*, 2013
C. difficile R20291::*ermB*	Ribotype 027 Erythromycin resistant derivative	Kelly *et al.*, 2016
C. difficile DH1916	Ribotype 027	Ed Kuijper, Leiden University
C. difficile R12801	Ribotype 106	Ed Kuijper, Leiden University
C. difficile Serogroup F	Ribotype 017	Ed Kuijper, Leiden University
C. difficile EK 28	Ribotype 078	Ed Kuijper, Leiden University
C. difficile EK 15	Ribotype 014	Ed Kuijper, Leiden University
C. difficile L 26	Ribotype 001	Ed Kuijper, Leiden University
C. scindens ATCC 35704	Type strain 20 alpha-hydroxysteroid dehydrogenase, 7 betadehydrogenase, 7 alpha-dehydroxylase, steroid desmolase	Jason Ridlon, Virginia Commonwealth University
C. scindens VPI 12708	20 alpha-hydroxysteroid dehydrogenase, 7 betadehydrogenase, 7 alpha-dehydroxylase.	Jason Ridlon, Virginia Commonwealth University
C. scindens SA14	Clinical isolate	Jason Ridlon, Virginia

		Commonwealth University
C. scindens SO76	Clinical isolate	Jason Ridlon, Virginia Commonwealth University
C. scindens SO77	Clinical isolate	Jason Ridlon, Virginia Commonwealth University
C. scindens KS11	Clinical isolate	Jason Ridlon, Virginia Commonwealth University
C. scindens G11	Clinical isolate	Jason Ridlon, Virginia Commonwealth University
C. scindens C132	Clinical isolate	Jason Ridlon, Virginia Commonwealth University
E. coli TOP10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lacX74 nupG recA1 araD139* Δ(*ara-leu*)*7697 galE15 galK16 rpsL*(Str^r) *endA1* λ-	Invitrogen
NEB 5-alpha	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
E. coli CA434	*E. coli HB101 [F- mcrB mrr hsdS20(rB-mB-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(Smr) glnV44λ-] R702*	Williams *et al*., 1990
E. coli S17-1	TpR SmR recA, thi, pro, hsdR-M+RP4: 2-Tc:Mu: Km Tn7	Simon *et al*., 1983
E. coli NEB Express	*fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10*	New England Biolabs
E. coli SExpress	*fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-*	Woods, 2016 (unpublished)

	210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10, R702	
E. coli Stellar	*F–, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ(lacZYA-argF) U169, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, λ–*	Clontech
E. coli InterStellar	*F–, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ(lacZYA-argF) U169, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, λ–*, R702	Millard, 2016 (unpublished)
E. coli RE InterStellar	*F–, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ(lacZYA-argF) U169, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, λ–*, R702, KD46cm	Millard, 2017 (unpiblished)

2.1.2 Plasmids

Table 2.2 Plasmids

Plasmid	Description	Source
pBMH	Vector containing insert to introduce the T3428A SNP in *rpoB*. Amp^R	Biomatik
pMTLYN4-TA	Allelic exchange vector to introduce the T3428A SNP in *rpoB*. Cm^R	This study
pMTLYN4-TG	Allelic exchange vector to introduce the T3428G SNP in *rpoB*. Cm^R	This study
pMTLYN4-GT	Allelic exchange vector to introduce the G3427T SNP in *rpoB*. Cm^R	This study
pMTL82151	Modular shuttle vector pBP1 Gram + replicon Cm^R	Heap *et al*., 2009
pMTL83151	Modular shuttle vector pCB102 Gram + replicon Cm^R	Heap *et al*., 2009
pMTL84151	Modular shuttle vector pCD6 Gram + replicon	Heap *et al*., 2009

	Cm^R	
pMTL85151	Modular shuttle vector pIM13 Gram + replicon Cm^R	Heap *et al*., 2009
RBV1	Recombineering vector aad9 homology arms araC inducible promoter Erm^R, Amp^R	Mansfield, 2017 (unpublished)
RBV1_1A	Recombineering vector Containing type 1 methylase and specificity Erm^R, Amp^R	This study
RBV1_1B	Recombineering vector Containing type 1 methyltransferase and specificity Erm^R, Amp^R	This study
RBV1_2A	Recombineering vector Type 2 A methyltransferase Erm^R, Amp^R	This study
RBV1_2B	Recombineering vector Type 2 B methyltransferase Erm^R, Amp^R	This study
RBV1_2C	Recombineering vector Type 2 C methyltransferase Erm^R, Amp^R	This study
RBV1_2D	Recombineering vector Type 2 D methyltransferase Erm^R, Amp^R	This study
pKD46	Lambda Red recombinase expression plasmid Cm^R	Datsenko & Wanner, 2000

2.2 Culture Media and Growth Conditions

2.2.1 Culture Media

Bacterial culture media is listed in table 2.3, below. Dehydrated media were sourced from Oxoid, unless otherwise stated. All culture media were sterilised by autoclaving at 121 ^{o}C at 100 kPa above atmospheric pressure for 15 minutes.

Table 2.3 Culture Media

Brain Heart Infusion Supplemented (BHIS)
-37 g brain heart infusion, 20 g yeast extract and 1 g L-cysteine in 1 L dH_2O. For solid plate media brain heart infusion supplemented with agar at 10 g l^{-1} is used.
Tryptose Yeast Extract medium (TY)
-30g tryptose, 20 g yeast extract and 1 g thioglycolate in 1 L dH_2O.
Fructose Broth Media (FBM)
-40 g proteose peptone, 5 g sodium phosphate dibasic, 1 g potassium phosphate monobasic, 2 g sodium chloride, 0.1 g magnesium sulphate, 6 g fructose, 1 g sodium taurocholate, 0.5 g L-cysteine, 0.03 g neutral red. For solid media, Oxoid No.1 was added to a concentration of 2 %, omitting neutral red.
Lysogeny Broth (LB)
-10g Bacto tryptose (BD Biosciences), 5 g yeast extract, 10 g sodium chloride in 1L dH2O. For solid plate media Oxoid agar No. 1 added at a concentration of 1.5 %.
Super Optimal with Catabolite repression (SOC) medium (Invitrogen)
-2 % tryptone, 0.5 % yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulphate, and 20 mM glucose.
Fastidious Anaerobe Agar (FAA)
Peptone 23 g, Sodium Chloride 5 g, Soluble Starch 1 g, Agar No. 2 12 g, Sodium Bicarbonate 0.4 g, Glucose 1 g, Sodium Pyruvate 1 g, L-Cysteine 0.5 g, Haemin 0.01 g, Vitamin K 0.001 g, L-Arginine 1.0 g, Sodium Pyrophosphate 0.25 g, Sodium Succinate 0.5 g in 1 L dH_2O with 50 mL Defibrinated Horse Blood added after autoclaving.

2.2.2 Antibiotic Supplements

Antibiotic stocks were all filter sterilised using 0.2 µm Minisart NML syringe filters (Sartorius) and stored in the recommended conditions. Antibiotic working concentrations are listed in table 2.4, below.

Table 2.4 Antibiotic Working Concentrations

Antibiotic	***E. coli* Working Concentration**	***Clostridia* Working Concentration**
Ampicillin	50 µg/mL	N/A
Apramycin	50 µg/mL	N/A
Cefoxitine	N/A	8 µg/mL
Chloramphenicol	25 µg/mL	N/A
D-Cycloserine	250 µg/mL	N/A
Erythromycin	50 µg/mL	N/A
Fidaxomicin	N/A	1.0 µg/mL
Kanamycin	50 µg/mL	N/A
Spectinomycin	250 µg/mL	250 µg/mL
Thiamphenicol	N/A	15 µg/mL

2.2.3 Growth Conditions

E. coli strains were cultured aerobically at 37 ^{o}C and liquid cultures were incubated with agitation at 200 G.

Clostridia were cultured anaerobically in an anaerobic work station (Don Whitley Scientific MG1000) with an internal atmosphere of CO_2:H_2:N_2 (10:10:80 v/v/v) at 37^{o}C.

All media were reduced for a minimum of 4 h for solid media and 8 h for liquid media before culturing.

2.3 Microbiological Methods

All methods are carried out in triplicate with biological replicates from independent experiments, unless otherwise stated.

2.3.1 Bacterial Growth Profiles

Overnight cultures were set up by inoculating pre-reduced BHIS broth media from strains pre-streaked on to solid BHIS media. Cultures were incubated overnight for approximately 18 hours and were standardised to the lowest OD_{600}. These cultures were then used to inoculate pre-reduced broth by adding 20 µl in to 180 µl of BHIS in a 96 well microplate (black with clear F-bottom/chimney well, Greiner Bio-One). The microplate was sealed using a plate sealer (R&D Systems) and transferred to a pre-warmed Promega GloMax-Multi+ Detection System plate reader. The plate was incubated at 37 ^{o}C and optical density readings were taken at a wavelength of 600 nm after a 6 second period of orbital shaking. Readings were taken every hour over a 24 h period.

Growth curves for RNA extraction were performed by inoculating pre-reduced BHIS broth with standardised overnight cultures in a Don Whitley Anaerobic workstation. Cultures were incubated as detailed in 2.2.3. Samples were taken at 1 h intervals every hour for 12 h and finally at 24 h, with OD_{600} determined using a Jenway 7310 spectrophotometer.

2.3.2 Minimum Inhibitory Concentration Assay

The broth dilution method was used to ascertain minimum inhibitory concentrations (Stalons & Thornsberry, 1975). A 96 well microplate (black with clear F-bottom/chimney well, Greiner Bio-One) is set up with antibiotics diluted in BHIS media, in a range of 0 to 32 μg/ml. Independent overnight cultures are standardised to the lowest OD_{600} and used to inoculate the antibiotic supplemented broth in the 96 well microplate. Cultures are then incubated anaerobically as described in 2.2.3, for 24 h. The microplate is then transferred to a Promega GloMax-Multi+ Detection System plate reader, and OD_{600} measurements taken.

2.3.4 Preparation of Pure Spores of C. difficile

Overnight cultures are spread on to BHIS plates in 100 μL volumes and incubated anaerobically as described in 2.2.3, for five days. Plates are then stored at 4 ^{o}C for 2 – 3 h before growth material is scraped off and resuspended in ice cold double distilled water (ddH_2O). Crude spore preparations are stored overnight at 4 ^{o}C to promote release of spores from mother cells. Spores are then washed by centrifugation at 4 ^{o}C in ice cold ddH_2O until the desired purity is reached. Spore preparation purity is assessed by observation under phase-contrast microscopy (2.3.5).

2.3.5 Microscopy

Microscopic observation was carried out using a Nikon Eclipse Ci Microscope and photographs taken using Nikon DS3 software.

2.3.6 Preparation of Electrocompetent E. coli

Overnight *E. coli* cultures were added at a ratio of 1:100 to 200 mL of pre-warmed LB broth. Cultures were incubated at 37 °C with agitation until OD_{600} 0.3 was reached, when cells were harvested by centrifugation at 4,000 G, 4 °C for 15 minutes. Cells are washed in ice cold ddH_2O three times, in successively decreasing volumes of 40 mL, 20 mL combining pellets, and finally 10 mL. The final pellet is resuspended in 1 mL 4-Morpholinepropanesulfonic acid (MOPS) with 10 % glycerol, and 50 µL aliquots made for subsequent usage.

2.3.7 Preparation of Electrocompetent C. scindens

Overnight *C. scindens* cultures were added at a ratio of 1:100 to 4 x 50 mL volumes of pre-warmed, pre-reduced BHIS broth. Cultures were incubated anaerobically at 37 °C with occasional agitation until an optical density at 600 nm (OD_{600}) of 0.5 was reached, when cells were harvested by centrifugation anaerobically at 4,000 G, 4 °C for 15 minutes. Cells were washed in ice cold electroporation buffer (270 mM sucrose, 0.6 mM Na2HPO4, 4.4 mM NaH2PO4, 10 mM MgCl2 adjusted to pH 6) three times anaerobically, in successively decreasing volumes of 40 mL, 20 mL, combining pellets, and finally 10 mL. The final pellet was resuspended in 1 mL MOPS with 10 % glycerol, and 50 µL aliquots made for subsequent usage.

Table 5.2 *C. scindens* Electroporation Settings

Voltage	Resistance	Capacitance	Decay
1500	∞	25	Exponential
1800	∞	25	Exponential
2000	∞	25	Exponential
2200	∞	25	Exponential

2.3.8 Conjugal Transfer of Plasmids

A 1 mL sample of overnight stationary phase culture of *E. coli* donor cells were centrifuged at 4,000 G and washed in PBS twice to remove residual antibiotic. Washed donor pellets were taken in to the anaerobic cabinet where they were resuspended in 200 µL of the recipient organism. The resulting conjugal mixture was plated on to BHIS agar plates in discrete 20 µL spots and incubated for a variety of incubation times; 4, 8, 12 or 24 h. The mating mixtures were then harvested and suspended in PBS, before being spread on to BHIS agar plates supplemented with the appropriate antibiotics to select for the recipient strain. Colonies observed were re-streaked to purity on the same selective media. The method was originally described in Purdy *et al*., 2002.

2.3.9 Spore Germination Initiation Assay

Spores were prepared for the strains to be tested, as detailed in 2.3.4. Spores were heat treated at 65 ^{o}C for 30 minutes, centrifuged for 1 minute at 12,000 G and supernatant discarded. Spore pellets were then resuspended in 500 µL of BHIS and 75 µL added to the wells of a 96 well microplate (black with clear F-bottom/chimney well, Greiner Bio-One). Media supplemented with 2 % taurocholic acid (Sigma-Aldrich) and 2 x concentration of substance to test was prepared and 75 µL added to each well containing spores, diluting the test solution 1:2, resulting in a final volume of 150 µL. The microplate is then transferred to a Promega GloMax-Multi+ Detection System plate reader, and OD_{600} measurements taken after a 6 second period of orbital shaking, every minute for 10 minutes and then every 10 minutes until 80 minutes was reached.

2.3.10 Spore Outgrowth Assay

Spores were prepared for the strains to be tested, as detailed in 2.3.4. Spores were heat treated at 65 °C for 30 minutes. Spore preparations were then standardised to an OD_{600} of 0.5 in a total volume of 1 mL. A 96 well microplate (black with clear F-bottom/chimney well, Greiner Bio-One) was prepared in an anaerobic workstation (Don Whitley Scientific) with 180 µL of pre-reduced BHIS per well, supplemented with 1 % taurocholic acid. Spore preparations were used to inoculate each well, diluting 1:10. The microplate was sealed using a plate sealer (R&D Systems) and transferred to a pre-warmed Promega GloMax-Multi+ Detection System plate reader. The plate was incubated at 37 °C and optical density readings were taken at a wavelength of 600 nm after a 6 second period of orbital shaking. Readings were taken every hour over a 15 h period.

2.3.11 Competetive Growth Advantage Assay

Independent overnight cultures of mutant and parental strains were grown overnight and the OD_{600} standardised to the lowest reading. Mutant and parental strains were inoculated together at 1:100 dilution into BHIS broth. Samples were taken at 0, 2, 4, 6, 8, 10 and 24 h, diluted and plated on to BHIS agar plates supplemented with cefoxitin at 8 µg/mL and D-cycloserine at 250 µg/mL (as detailed in table 2.4) for the total viable count, and on to BHIS agar plates supplemented with cefoxitin at 8 µg/mL and D-cycloserine at 250 µg/mL and fidaxomicin at twice the appropriate MIC (Table 3.4) for the total mutant count. After 24 h of incubation, colonies were counted and colony forming units per mL (CFU/mL) were calculated.

2.3.12 Rate and Capacity of Sporulation Assay

Overnight cultures of *C. difficile* strains were sub-cultured to ensure no spore carry over. This starter culture was then used to inoculate BHIS broth media at a dilution of 1:100. Samples were taken at 0, 24, 48, 72, 96 and 120 h, heat treated and plated on to BHIS agar plates supplemented with 0.1 % taurocholic acid. Colonies were counted, and heat resistant CFU/mL calculated, (Burns *et al*., 2010).

2.3.13 Loss of Heat Resistance Germination Assay

Spores were prepared for the strains to be tested, as detailed in 2.3.4. Spore preparations were standardised to OD_{600} 0.5 and used to inoculate pre-reduced BHIS broth supplemented with 1 % taurocholic acid. Samples were taken at 0 h and then every 30 minutes over a 4 h period. Samples were heat treated at 65 ^{o}C for 30 minutes and plated on to BHIS agar supplemented with taurocholic acid. Colonies were enumerated, and CFU/mL calculated. The percentage decrease in recovered colonies was calculated as an indication of spore germination due to a loss of heat resistant of germinated spores in the culture.

2.3.14 Total Toxin Production Assay

C. difficile strains were grown in TY broth, overnight cultures were subcultured at a dilution 1:100 into fresh media and supernatants of strains were collected at 24, 48 and 72 h. The production of toxin A/B was quantified by a standard curve using a toxin A/B ELISA assay according to the manufacturer's instructions (Techlab TOX A/B ELISA™). The

standard curve was produced using *C. difficile* toxin A and B standards (The Native Antigen Company), serially diluted 1:2 from 125 ng/mL to 1.95 ng/mL.

2.3.15 Cytotoxicity Assay

C. difficile strains were cultured in TY broth. Overnight cultures were sub-cultured 1:100 into fresh media and sampled at 0, 24, 48 and 72 h. At each time point, OD_{600} was measured, samples centrifuged and the supernatant filter sterilised and stored at –20 °C. The supernatants were serially diluted, added to confluent monolayers of Vero (African green monkey kidney) cells and incubated for 24 h before scoring under a microscopic observation. The toxin endpoint titre was the dilution at which the Vero cells resembled a negative control, exhibiting no cell rounding.

2.3.16 In-Vitro 7α-Dehydroxylation

BHIS broth was supplemented with cholic acid (Sigma-Aldrich) at a concentration of 200 μM. Independent overnight cultures of *C. scindens* strain ATCC 35704 were standardised to the lowest OD_{600} and used to inoculate pre-reduced BHIS broth, with and without CA supplementation, at a dilution of 1:100. Samples were taken at 4, 8 and 24 h for bile acid analysis and *C. difficile* germination initiation and outgrowth assays (2.3.9, 2.3.10). Samples for RNA extractions were taken at 4, 6, 8 and 10 h.

2.3.17 Bile Acid Analysis

Bile acid analysis was performed by collaborators at the École Polytechnique Fédérale de Lausanne (EPFL). Analyses were conducted on an Agilent 6530 Accurate-Mass Q-TOF LC/MS mass spectrometer coupled with an Agilent 1290 UHPLC system (Agilent Technologies, USA). Bile acids were quantified using standard calibrations curves. Standard stock solutions were prepared at a concentration of 10 mM in methanol. Bile acid solutions were mixed together and diluted with 50:50 ammonium acetate (5 mM):MeOH to construct standard curves between 5 and 5000 nM.

2.4 *In Vivo* Methods

All animal procedures were carried out in accordance with the United Kingdom Home Office Inspectorate under the Animals (Scientific Procedures) Act 1986. All experiments were approved by the University of Nottingham Ethics committee and carried out under licence number 40/3950. (Kuehne *et al*., 2013; Kelly *et al*., 2016)

2.4.1 The Hamster Model of Virulence

Virulence was assessed in infection assays using female Golden Syrian hamsters, as described previously, (Kelly *et al*., 2016). Groups of eight hamsters were pre-treated with clindamycin for 7 days prior to being challenged with 10^5 *C. difficile* spores. The primary output was survival or prolonged survival. Hamsters were monitored for disease symptoms and general condition, and euthanised when a predetermined threshold of

symptom severity and/or worsening of condition was met, according to a previously published scoring system, (Kuehne *et al*., 2010).

2.4.2 Enumeration of Hamster Cecal Colonisation

Hamster caeca were homogenised in a Precellys 24 homogeniser (Bertin Instruments) and centrifuged. The resulting supernatant was heat-treated at 65 °C prior to serial dilution and plating on D-cycloserine–cefoxitin fructose agar to enumerate. The number of colonies was counted after 2–3 days of incubation, as detailed in 2.2.3.

2.5 Molecular Methods

2.5.1 Extraction of Chromosomal DNA

Chromosomal DNA for cloning applications was extracted using GenElute Bacterial Genomic DNA kits (Sigma-Aldrich) according to the manufacturer's instructions for Gram-positive bacteria, with optional RNA removal step after step 2 and deviations in elution volume when required, detailed below.

1. *Pellet 1.5 mL of an overnight bacterial broth culture by centrifuging for 2 minutes at 12,000–16,000 × g. Remove the culture medium completely and discard.*
2. *Resuspend the pellet thoroughly in 200 µL of Lysozyme Solution (prepared in step 1b) and incubate for 30 minutes at 37 °C.*
3. *Add 500 µL of the Column Preparation Solution to each pre-assembled GenElute Miniprep Binding Column seated in a 2 mL collection tube. Centrifuge at 12,000 × g for 1 minute. Discard the eluate.*
4. *Add 200 µL of ethanol (95–100 %) to the lysate and mix thoroughly by vortexing for 5–10 seconds.*

5. *Transfer the entire contents of the tube into the binding column and Centrifuge at ≥ 6500 × g for 1 minute. Discard the eluate.*
6. *Add 500 µL of Wash Solution 1 (W0263) to the column and centrifuge for 1 minute at ≥ 6500 × g. Discard the eluate.*
7. *Add 500 µL of Wash Solution to the column and centrifuge for 3 minutes at maximum speed (12,000–16,000 × g) to dry the column.*
8. *Pipette 200 µL of the Elution Solution (B6803) directly on to the center of the column; centrifuge for 1 minute at ≥ 6500 × g to elute the DNA.*

For next generation sequencing applications, chromosomal DNA was extracted using phenol-chloroform extraction, detailed below.

1. *Pellet 1 – 5 mL of bacterial culture at 12,000 G for 2 minutes, discard supernatant.*
2. *Incubate pellets with 180 µL 10 mg/mL lysozyme solution at 37 °C for 30 minutes.*
3. *Add 25 µL 10 mg/mL proteinase K solution, 85 µL d.H2O and 110 µl 10 % (w/v) SDS solution incubated at 65°C for 30 min*
4. *Add equal volumes of Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v/v) and transfer to 5PRIME phase lock heavy tubes (Quanta Bio) and centrifuge for 3 minutes at 20,000 G.*
5. *Transfer the top phase to a new phase lock tube and repeat step 4 a further two times.*
6. *Transfer the top phase in to a fresh tube and add 40 µL sodium acetate and 800 µL of ice cold ethanol.*
7. *Precipitate DNA at – 80 °C for 1 – 2 h.*
8. *Centrifuge samples for 15 minutes at 12,000 g an 4 °C.*
9. *Discard supernatent and wash in 70 % Ethanol, repeat step 8.*
10. *Centrifuge to remove residual ethanol and air dry pellet.*
11. *Resuspend DNA pellet in nuclease free water and quantify.*

2.5.2 Extraction of Total RNA and Purification from C. scindens

Total RNA was extracted using an adapted and optimised protocol utilising RNAprotect reagent (Qiagen), FastRNA PRO BLUE kit (MP Biomedicals), RNeasy Mini kit (Qiagen) and the Turbo DNA-free reagents (Ambion). The protocol is detailed below.

Harvesting Cells

1. Transfer culture tubes on ice to a cooled centrifuge. Spin down cultures - 5000 G for 2 - 5 minutes at 4 °C – critical for centrifuge to be cooled appropriately. Centrifugation parameters depend upon culture volume used and OD of time point.

2. Discard supernatant quickly and carefully so as not to disturb the pellet. Immediately snap freeze pellet in liquid nitrogen, place on dry ice and transfer to -80 °C freezer when all pellets have been prepared. If processing multiple tubes, rotate order of snap freezing for consistency between samples.
Work as quickly as possible to minimise the effects of cold shock.

RNA Potect and FastRNA Pro Blue

Perform the following lysis steps (1 – 3) as quickly and cold as possible.
1. Remove samples from -80 °C freezer and place on ice. Quickly, before thawing, re-suspend and combine pellets in 1 ml RNA protect.
Maximum of 3 samples at a time, to process quickly and avoid RNA degradation. Tubes must not be too cold (i.e. if placed on dry ice) to avoid RNA protect freezing, leaving pellets exposed to thawing.
Pellets may be split at this point and frozen at –80 °C if required.
Incubate at room temperature for 5 minutes and centrifuge for 1 minute at 13,000 G at 4 °C.

2. Discard supernatant and re-suspend pellets in 1 ml RNA pro solution on ice. Transfer the whole mixture in to RNA pro matrix tube on ice and homogenise at 6400 G for 45 seconds.

3. Centrifuge the homogenised sample for 15 minutes at 13,000 G at 4 °C.

Place chloroform on ice and prepare 1.5 ml centrifuge tubes containing 500 µL of absolute ethanol for each sample.

4. Transfer the solution, avoiding cell debris and matrix to a fresh 1.5 ml micro centrifuge tube and leave at room temperature for 5 minutes.

5. In a fume hood, add 300 µL of ice cold chloroform to each sample and mix by vortex for ~6 seconds; incubate at room temperature for 5 minutes to allow dissociation of nucleoprotein complexes.

6. Centrifuge samples for 15 minutes at 13,000 G at 4 °C.

7. Carefully* Transfer the upper phase to the micro centrifuge tubes.

Optional DNA shearing step: Pass the solution through a 12G needle at lesat 10 times to shear genomic DNA, easing the gDNA removal step.

Mix sample by inversion with 500 µL of absolute ethanol.

* using a P200 pipette, P1000 aspiration is too violent and may disturb the bottom phase. P200 allows removal of more of the upper phase with more control.

8. Incubate at -80 °C for 1-2 h to precipitate the RNA.

9. Centrifuge samples for 15 minutes at 13,000 G at 4 °C.

Prepare 70 % ethanol in DEPC RNase free water.

11. Discard supernatant and wash pellet in 500 µL of 70 % ethanol, centrifuge for 15 minutes at 13,000 G at 4 °C.

12. Discard supernatant, removing excess with a p200 pipette and air dry pellet at room temperature for 30 minutes. Avoid air drying for longer periods to avoid difficulties in solubilising RNA pellet.

13. Re-suspend pellet in 50 µL of DEPC water and store at -80 °C.

Genomic DNA Removal

1. Add 5 µL of Ambion Turbo DNase buffer and 2 µL (increase if more for rigorous treatment is required) of Ambion Turbo DNase to each sample, spin down briefly and mix.

2. Incubate samples at 37 °C for 30 minutes with occasional agitation.

3. Repeat step 1.

Clean Up
RNeasy Mini Kit

Prepare RLT buffer with β-mercaptoethanol (3 µL of 2-mercaptoethanol to 400 µL RLT buffer, can increase amount of mercaptoethanol if excessive RNase contamination is suspected).

1. Add 350 µL of the RLT 2-mercaptoethanol mixture and 200 µL of absolute ethanol to the DNase treated sample and mix by inversion.

2. Apply the mixture to an RNeasy column and centrifuge for 15 seconds at 10,000 G.

3. Discard flow through, add 500 µL of RPE wash buffer to the column and centrifuge for 15 seconds at 10,000 G, discard flow-through.

4. Repeat wash step 3.

5. Centrifuge the column at 10,000 G for 2 minutes to dry the column.

6. Using the RNase free tubes supplied with the kit, elute RNA from the column in 17 µL of DEPC water, stand for 3 minutes to allow liquid to soak in to the membrane before centrifuging at 10,000 G for 1 minute.

7. Repeat elution step a further two times to attain a final volume of ~50 µL, store at -80 °C

2.5.3 Polymerase Chain Reaction

Polymerase chain reactions (PCR) for non-proof reading applications were performed in a total volume of 20 µL, comprising of 1 µL of each primer (Table 2.10), 0.5 µL of GoTaq polymerase (Promega), 10 µL of FailSafe buffer 2 x PreMix E (Cambio), 6.5 µL of ddH_2O and 1 µL of template DNA. Reaction conditions are detailed in Table 2.5.

PCRs for proof reading applications were performed in a total volume of 20 µL, comprising of 1 µL of each primer (Table 2.10), 0.5 µL of Phusion high-fidelity polymerase (NEB), 10 µL of FailSafe buffer 2 x PreMix E (Cambio), 6.5 µL of ddH_2O and 1 µL of template DNA. Reaction conditions are detailed in Table 2.5.

PCRs for site directed mutagenesis were carried out using Q5 Hot Start High-Fidelity polymerase (NEB) in a total volume of 20 µL. The reactions were comprised of 1 µL of each primer (Table 2.10), 0.5 µL Q5 polymerase, 5 µL Q5 reaction buffer, 11.5 µL of ddH_2O and 1 µL of template DNA. Reaction conditions are detailed in Table 2.5.

Colony PCR was performed as with the Phusion polymerase protocol, except for an extra step of 98 ^{o}C for 15 minutes at the start of the protocol.

Table 2.5 PCR Conditions

		GoTaq		**Phusion**	
Reaction step		Temperature	Time	Temperature	Time
Initial Denaturation		95 °C	2 mins	98 °C	30 secs
Denaturation	25 cycles	95 °C	1 min	98 °C	10 secs
Annealing		specific	1 min	specific	30 secs
Extension		72 °C	1 min/ Kb	72 °C	15 secs / Kb
Final Extension		72 °C	5 mins	72 °C	5 min
Stop		15 °C	∞	15 °C	∞
		Q5			
Reaction step		Temperature	Time		
Initial Denaturation		98 °C	30 secs		
Denaturation	25 cycles	98 °C	10 secs		
Annealing		specific	30 secs		
Extension		72 °C	30 secs/Kb		
Final Extension		72 °C	2 mins		
Stop		10 °C	∞		

2.5.4 Plasmid Extraction

Plasmids were extracted from 1 – 5 mL of overnight bacterial culture using NEB Monarch Plasmid Miniprep kits (NEB #T1010). Extractions were performed according to the manufacturer's instructions, detailed below.

1. *Pellet 1–5 mL bacterial culture by centrifugation for 30 seconds. Discard supernatant.*
2. *Resuspend pellet in 200 µl Plasmid Resuspension Buffer (B1). Vortex or pipet to ensure cells are completely resuspended. There should be no visible clumps.*
3. *Add 200 µl Plasmid Lysis Buffer (B2) , gently invert tube 5–6 times, and incubate at room temperature for 1 minute. Color should change to dark pink, and solution will become transparent and viscous. Do not vortex.*
4. *Add 400 µl of Plasmid Neutralization Buffer (B3) , gently invert tube until neutralized, and incubate at room temperature for 2 minutes. Sample is neutralized when color is uniformly yellow and precipitate forms. Do not vortex.*
5. *Centrifuge lysate for 2–5 minutes. For culture volumes > 1 mL, we recommend a 5 minute spin to ensure efficient RNA removal by RNase A. Pellet should be compact; spin longer if needed.*
6. *Carefully transfer supernatant to the spin column and centrifuge for 1 minute. Discard flow-through.*
7. *Re-insert column in the collection tube and add 200 µl of Plasmid Wash Buffer 1. Centrifuge for 1 minute. Discarding the flow-through is optional.*
8. *Add 400 µl of Plasmid Wash Buffer 2 and centrifuge for 1 minute.*
9. *Transfer column to a clean 1.5 mL microfuge tube. Use care to ensure that the tip of the column does not come into contact with the flow-through. If there is any doubt, re-spin the column for 1 minute.*
10. *Add ≥ 30 µl DNA Elution Buffer to the center of the matrix. Wait for 1 minute, then spin for 1 minute to elute DNA. Nuclease-free water (pH 7–8.5) can also be used to elute the DNA. Yield may slightly increase if a larger volume of DNA Elution Buffer is used, but the DNA will be less concentrated. For larger size DNA, (≥ 10 kb), heating the elution buffer to 50°C prior to use can improve yield.*

2.5.5 Restriction Digest

All restriction enzymes were sourced from NEB and stored at -20 °C before and after use. Reactions were incubated at 37 °C for 1 – 2 h, with 1 µL of restriction endonuclease and

2 µL of appropriate buffer, to a final volume of 20 µL. Digests were subsequently run on agarose gels to visualise, (2.5.8), and remove the restriction endonucleases.

2.5.6 Dephosphorylation of Digested Vectors

To prevent re-ligation of digested vectors, backbones were dephosphorylated with Antarctic phosphatase (NEB). To digested vector reactions of 20 µl, 1 µl of phosphatase and 1.2 µl of Antarctic phosphatase buffer was added and incubated at 37 ^{o}C for 30 minutes. Dephosphorylated digests were subsequently run on agarose gels (2.5.8), removing the restriction endonucleases and Antarctic phosphatase, halting these reactions.

2.5.7 Ligation

Ligation reactions were carried out using T4 ligase (NEB) and T4 ligase buffer, in 2:0 (negative control) and 2:6 reaction ratios as detailed in table 2.6. Reactions were incubated overnight in ice water. Ligation reactions were dialysed on 0.025 µm Millipore filters (Millipore) for 25 minutes prior to transformation.

Table 2.6 Ligation Reactions

	Ratio 2:0	Ratio 2:6
Vector	4 ul	4 ul
Insert	0 ul	12 ul
Ligase	1 ul	1 ul
Ligation buffer	2 ul	2 ul
dH_2O	13 ul	1 ul

2.5.8 Agarose Gel Electrophoresis

DNA was visualised on 1 % agarose gels (Sigma-Aldrich) in 1 x TAE buffer (50 x TAE: 242 g Tris free base, 18.61 g Disodium EDTA, 57.1 ml Glacial Acetic Acid, ddH2O to one L),

with SyberSafe (Invitrogen) at 1 x concentration to allow visualisation under blue light. Prior to loading the samples, 6 x purple loading dye (NEB) was added. Fragment size was estimated using the appropriate DNA marker (NEB), with gels being run at 100 V for the appropriate time for fragment migration. DNA was visualised, and images captured using a Gel Doc XR system (BioRad).

2.5.9 Extraction of DNA from Agarose Gels

DNA fragments were excised from gels and purified using Zymoclean Gel DNA Recovery kits (Zymo Research), in accordance with the manufacturer's instructions, detailed below, deviating in elution volumes when required.

1. *Excise the DNA fragment from the agarose gel using a razor blade, scalpel or other device and transfer it into a 1.5 mL microcentrifuge tube.*
2. *Add 3 volumes of ADB to each volume of agarose excised from the gel (e.g. for 100 µL (mg) of agarose gel slice add 300 µL of ADB).*
3. *Incubate at 37-55 °C for 5-10 minutes until the gel slice is completely dissolved.*
4. *Transfer the melted agarose solution to a Zymo-Spin™ Column in a Collection Tube.*
5. *Centrifuge for 30-60 seconds. Discard the flow-through.*
6. *Add 200 µL of DNA Wash Buffer to the column and centrifuge for 30 seconds. Discard the flow-through. Repeat the wash step.*
7. *Add ≥ 6 µL DNA Elution Buffer4 or water5 directly to the column matrix. Place column into a 1.5 mL tube and centrifuge for 30-60 seconds to elute DNA.*

2.5.10 PCR Product Purification

PCR products were purified using the Monarch PCR & DNA Cleanup Kit (NEB) according to the manufacturer's instructions detailed below, deviating in elution volumes when required.

1. *Dilute sample with DNA Cleanup Binding Buffer. Mix well by pipetting up and down or flicking the tube. Do not vortex. A starting sample volume of 20–100 µl is recommended. For smaller samples, TE can be used to adjust the volume. For diluted samples larger than 800 µl, load a portion of the sample, proceed with Step 2, and then repeat as necessary.*
2. *Insert column into collection tube and load sample on to column and close the cap. Spin for 1 minute, then discard flow-through.*
3. *Re-insert column into collection tube. Add 200 µl DNA Wash Buffer and spin for 1 minute. Discarding flow-through is optional.*
4. *Repeat wash (Step 3).*
5. *Transfer column to a clean 1.5 mL microfuge tube.*
6. *Add ≥ 6 µl of DNA Elution Buffer to the center of the matrix. Wait for 1 minute, then spin for 1 minute to elute DNA.*

2.5.11 Transformation of E. coli by Electroportation

Electro-competent *E. coli* cells were thawed on ice before the addition of 1 – 10 ng of plasmid DNA. Electroporation mixtures were added to pre-cooled electroporation cuvettes with a 2 mm aperture (BioRad). Bacteria were electroporated using a BioRad MicroPulser at 2.5 kV voltage, 25 µF capacitance and 200 Ω resistance; program EC 2.

Cells were recovered in SOC broth and incubated at 37 ^{o}C with agitation for at least 1 h, prior to plating on LB agar, supplemented with the appropriate antibiotic.

2.5.12 Transformation of C. scindens *by Electroportation*

Electro-competent *C. scindens* cells were thawed on ice in an anaerobic chamber, before the addition of 1 – 10 ng of plasmid DNA. Electroporation mixtures were added to pre-cooled electroporation cuvettes with a 4 mm aperture (BioRad). Bacteria were

electroporated using a BioRad MicroPulser and a BioRad Gene Pulser Xcell ShockPod Cuvette Chamber at a range of settings detailed in table 2.7.

Cells were recovered in pre-reduced and warmed BHIS broth, anaerobically at 37 °C with occasional agitation for at 2 – 4 h, prior to plating on to BHIS agar supplemented with the appropriate antibiotic.

Table 2.7 *C. scindens* Electroporation Settings

Voltage	Resistance	Capacitance	Decay
1500	∞	25	Exponential
1800	∞	25	Exponential
2000	∞	25	Exponential
2200	∞	25	Exponential

2.5.13 GpC In-Vitro Methylation

The *In vitro* methylation of plasmids to be transformed in to *C. scindens* was carried out in a reaction volume of 20 µL, comprised of nuclease free water 14 µL, 10 x NEBuffer 2 two µL, SAM two µL, plasmid DNA one µL, GpC methyltransferase 1 µL. Reactions were incubated at 37 °C for 1 h, with a 20 minute inactivation step at 65 °C.

2.5.14 Quickchange Site Directed Mutagenesis

Primers for site directed mutagenesis were designed using NEB basechanger tool (https://nebasechanger.neb.com/), with 5' back-to-back annealing ends (Table 2.8). Reaction mixtures were set up in 25 µL volumes, with 12.5 µL of Q5 Hot Start High-Fidelity polymerase, 1.25 µL of each primer, 1 µL of template DNA and 9 µL of nuclease

free water. The PCR cycling conditions according to the manufacturer's instructions are detailed in table 2.8, annealing temperatures provided by NEB basechanger tool.

Table 2.8 Site Directed Mutagenesis PCR protocol

Reaction Step		Temperature	Time
Initial Denaturation		98 °C	30 secs
Denaturation	25	98 °C	10 secs
Annealing	cycles	Reaction specific	30 secs
Extension		72 °C	30 secs/Kb
Final Extension		72 °C	2 mins
Hold		15 °C	∞

Kinase, ligase and *Dpn*I (KLD) reactions were set up according to the manufacturer's instructions (Table 2.9).

Table 2.9 KLD Reaction Conditions

Reaction Step	Volume	Final Concentration
PCR product	1 µL	
2 x KLD reaction buffer	5 µL	1 x
10 x KLD reaction mix	1 µL	1 x
Nuclease-free water	3 µL	

A total of 5 µL of the KLD reaction was added to 50 µL of NEB 5-alpha chemically competent cells and incubated on ice for 30 minutes, before heat shock at 42 °C for 30 seconds. Cells were then incubated on ice for 5 minutes before recovery in SOC media at 37 °C with gentle agitation for 1 h. Recovered cells were plated on to LB agar plates supplemented with chloramphenicol at 25 µg/mL.

2.5.15 HiFi Assembly

Primers for HiFi assembly were designed with 5' overlapping regions using the NEBuilder online tool (https://nebuilder.neb.com/). Vector backbones were linearised with appropriate restriction enzymes and HiFi mixtures were prepared as per the manufacturer's instructions, described below in table 2.10.

Table 2.10 HiFi Reaction Mixtures

	Recommended Amount of Fragments Used for Assembly		
Recommended DNA Molar Ratio	2-3 Fragment vector:insert 1:2	4-6 Fragments vector:insert 1:1	Positive Control
Total Amount of Fragments	0.03 – 0.2 pmols	0.02 – 0.5 pmols	10 µL
HiFi mastermix	10 µL	10 µL	10 µL
ddH_2O	10-x µL	10-x µL	0
Total Volume	20 µL	20 µL	20 µL

Samples were incubated at 50 °C for 15 minutes for 2-3 fragment assembly, or 60 minutes for 4-6 fragments, prior to transformation in to chemically competent NEB 5-alpha cells or electrocompetent *E. coli* TOP 10.

2.5.16 Lambda Red Recombineering

PCR fragments were generated using Phusion polymerase, with 500 bp homology arms flanking the desired insert and an apramycin resistance cassette. These PCR products were then electroporated (2.5.11) in to the appropriate *E. coli* recipient cells harbouring the arabinose induced pKD46 Lambda Red recombinase plasmid. Cells were recovered in SOC media and incubated overnight at 30 °C to maintain the pKD46 plasmid, after

which recovery cultures were plated on to LB agar plates supplemented with apramycin. Plates were incubated at 37 °C to lose the pKD46 plasmid, until colonies were observed. Colonies were screened by colony PCR (2.5.3), using an internal and external primer pair, to confirm successful recombination (Table 2.10).

Recombinants were screened for loss of pKD46 plasmid by negative selection on LB agar plates supplemented with chloramphenicol.

2.5.17 DNA and RNA Quantification

DNA for next generation sequencing was quantified fluorometrically using a Qubit 3.0 Fluorometer using the Qubit dsDNA HS assay kit (Life Technologies Q32854) according to the manufacturer's instructions.

For general cloning applications, DNA was quantified using microspectrophotometry on a NanoDrop Lite spectrophotomer (Thermo Fisher Scientific) according to the manufacturer's instructions.

RNA for RNA sequencing was quantified fluorometrically using a Qubit 3.0 Fluorometer using the RNA BR assay kit (Life Technologie Q10210) according to the manufacturer's instructions.

2.5.18 RNA Quality Assessment

Total RNA quality was assessed using the Agilent RNA 6000 Nano Kit (Agilent Technologies, 5067-1511) on the Agilent 2100 Bioanalyzer, according to the manufacturer's instructions.

2.5.19 Extraction and Purification of Total RNA from C. difficile

1. Extraction of RNA using the FastRNA pro blue extraction protocol

a) Add 2 volumes of RNA protect (Qiagen) to culture, vortex and incubate for 5 minutes
b) Centrifuge the samples at 5000 x g for 10 minutes at 4 °C to pellet cells (pellets can be frozen immediately at -80 °C)
c) Resuspend the pellets in 1-3 mL of RNA pro solution. Completely resuspend cells by pipetting or vortexing.
d) Transfer 1 mL of sample to matrix tube
e) Process in the homogeniser at 6400 G for 40 seconds
f) Remove and centrifuge at 13000 G for 15 minutes at 4 °C
g) Transfer solution to fresh Eppendorf avoiding cell debris and matrix
h) Incubate sample at room temperature for 5 minutes to increase RNA yield
i) Add 300 µL chloroform (cold). Vortex 10 seconds
j) Incubate 5 minutes at room temperature to permit nucleoprotein dissociation
k) Centrifuge at 13000 G for 15 minutes at 4°C
l) Transfer upper phase to fresh tube containing 500 µL 100% EtOH, mix by inverting 5 times, incubate at least 30 minutes (up to 2 hours) at -20 °C to precipitate RNA
m) Centrifuge at 13000 G for 15 minutes at 4 °C
n) Remove and discard supernatant
o) Add 500 µL 75% EtOH (made with DEPC water)
p) Centrifuge at 13000 G for 15 minutes at 4 °C
q) Remove supernatant and air dry pellet (< 10 minutes)
r) Resuspend pellet in 50 µL DEPC water
s) RNA can be stored at -80 °C or continue with DNase treatment and clean up

2. DNase treatment and purification

a) Use Turbo DNase from Ambion to remove DNA following kit protocol

b) clean up as described below

3. RNeasy purification (QIAGEN)

a) To DNase treated sample add
   350 µL RLT buffer (add β-merceptothanol)

   200 µL absolute EtOH

b) Mix, then apply to column
c) Centrifuge for 15 seconds at 10,000 G
d) Transfer to new collection tube
e) Add 500 µL RPE wash buffer
f) Centrifuge for 15 seconds at 10,000 G
g) Repeat wash 500 µL RPE wash buffer
h) Spin at 10000 G for 2 minutes to dry column
i) Elute from column with 17 µL DEPC water, allow to stand for 3 minutes
j) Centrifuge for 1 minute at 10000 G

Repeat twice to give a total volume of 50 µl

2.6 Oligonucleotides

Oligonucleotide primers were designed using Benchling (https://benchling.com) with annealing temperatures calculated using the NEB Tm calculator (http://tmcalculator.neb.com/#!/). All primers were ordered and synthesised by Sigma-Aldrich and Eurofins MWG. The Primers used in this study are detailed in table 2.11.

Table 2.11 Primers

Oligonucleotide	Sequence 5' – 3'	Function
rpoB_F	GTAGCAACATCAGTATTCGATGGTGC (26)	Forward *rpoB* screening primer
rpoB_R	CATATCAAAGTCATCACTTTCAAAATCTATTTCTTCTTC (39)	Reverse *rpoB* screening primer
TG_QC_F	ATCATTTAAGGTCTTATAAAGGAACTTC (28)	Forward primer for quickchange reactions; TG construct
TG_QC_R	TCTGGGATTCCTGGTCA (18)	Reverse primer for quickchange reactions; TG construct
GT_QC_F	ATCATTTAAGTTTCTTATAAAGGAACTTCAAAGTTTATGC (40)	Forward primer for quickchange reactions; GT construct
GT_QC_R	TTCTGGGATTCCTGGTCA (19)	Reverse primer for quickchange reactions; GT construct

baiCD_F	GGWTTCAGCCCRCAGATGTTCTTTG (25)	Forward *baiCD* screening primer
baiCD_R	GAATTCCGGGTTCATGAACATTCTKCKAAG (30)	Reverse *baiCD* screening primer
1A_RBVHF_F	GTTTCTCCATACCCGTTTTTTTGGGGAGTGCAAACGACTGAATATAAGTGAAAACCTAAGAAA (63)	Forward HiFi primer for MTase 1A
1A_RBVHF_R	CCGGAATCATATGACTCGAGTTTACGATAATCTGACCTTGCCGGTCAGG (49)	Reverse HiFi primer for MTase 1A
1B_RBVHF_F	GTTTCTCCATACCCGTTTTTTTGGGAGGAGAATTATTATGAGTATATCGGCTATAATTAAATCAGTGC (68)	Forward HiFi primer for MTase 1B
1B_RBVHF_R	CCGGAATCATATGACTCGAGTTCAGCATTCCTCGCCAGCTATCTCC (46)	Reverse HiFi primer for MTase 1B
2A_RBVHF_F	GTTTCTCCATACCCGTTTTTTTGGGAAGTGGCGAGGTTATAAAAATAGTACTAGC (55)	Forward HiFi primer for MTase 2A
2A_RBVHF_R	CCGGAATCATATGACTCGAGTCTATAGATCAAACAGCCGCTTCGCG (46)	Reverse HiFi primer for MTase 2A
2B_RBVHF_F	GTTTCTCCATACCCGTTTTTTTGGGTCATTGAGGCGAATAGAATGAATCAATATAATGAAATC (63)	Forward HiFi primer for MTase 2B
2B_RBVHF_R	CCGGAATCATATGACTCGAGTTTACTTACACCTCCATAACCCATCCTGG (49)	Reverse HiFi primer for MTase 2B
2C_RBVHF_F	GTTTCTCCATACCCGTTTTTTTGGGTAGAGATTGAAGGGGAAAAGGCGTGATGG (54)	Forward HiFi primer

		for MTase 2C
2C_RBVHF_R	CCGGAATCATATGACTCGAGTTTATGCTATAACATTCTCTACGTTATCAGCAACTTGTCG (60)	Reverse HiFi primer for MTase 2C
2D_RBVHF_F	GTTTCTCCATACCCGTTTTTTTGGGGAAGCAGCAGCAAGAACTATTTATTTGAATAAAACATGTTTTAATGG (71)	Forward HiFi primer for MTase 2D
2D_RBVHF_R	CCGGAATCATATGACTCGAGTTTAGCATGTTTTTGATTTCCAAAAGTAATATGTATAATCTATTGGC (68)	Reverse HiFi primer for MTase 2D
RE_LHA_F	TTCTACGGCACGTTTGAAGGCGCGC (25)	Forward primer to amplify recombineering cassette
RE_HA_R	GCAGCGACTTCCACGATGGG (20)	Reverse primer to amplify recombineering cassette
RE_F	GCCATTCAGAGAAGAAACCAATTG (24)	Recombineering forward screening primer
RE_R	ATTCAGAATGCCGAACACC (19)	Recombineering reverse screening primer

2.7 Sequencing Methods

2.7.1 Sanger Sequencing

Plasmid DNA and PCR products were sequenced using Sanger sequencing performed by Source Bioscience, Nottingham, UK.

2.7.2 Pacific Biosciences Single Molecule Real-Time Sequencing

Pacific Biosciences (PacBio) Single Molecule Real-Time (SMRT) sequencing was performed by Génome Québec on a PacBio RSII sequencer using a sheared large insert library and a single SMRT cell.

2.7.3 Illumina MiSeq Sequencing

Illumina MiSeq sequencing was performed by the University of Nottingham DeepSeq facility.

Libraries were prepared using the Illumina TruSeq PCR Free Library Preparation Kit (Illumina). 2 μg of genomic DNA was fragmented using a Carvars S2 sonicator and subsequently purified using Agencourt Ampure XP beads (Beckman Coulter). DNA fragments were end repaired for 30 minutes at 30 ^{o}C and desired fragment size selected using Ampure XP beads. Y adapters for Illumina sequencing were ligated to fragments and then incubated for 5 minutes at 70 ^{o}C. Surplus un-ligated adapters were removed by successive rounds of Ampure XP bead purification. Purified adapter ligated DNA was quantified using qPCR (Kapa Biosystems, KK4824).

Prepared DNA libraries were loaded and run on the MiSeq sequencing platform. Sequencing was completed using Illumina MiSeq 500 bp V2 SBS chemistry.

2.7.4 RNA Sequencing

RNA sequencing was carried out by the University of Nottingham Deepseq facility.

1µg of Total RNA was used for rRNA depletion using Ribo-Zero rRNA Removal Kit (Illumina, MRZMB126). Illumina stranded whole transcriptome sequencing libraries were prepared using NEBNext Ultra Directional RNA library prep kit for Illumina (NEB, E7420S). Dual Index Primers Set 1 was used to generate barcoded multiplex libraries (NEB, E7600S). Library quality control was performed using the bioanalyser HS kit (Agilent biotechnologies, 5067-4626). Libraries were quantified using qPCR (Kapa Biosystems, KK4824). Equimolar library pool was loaded for sequencing according to manufacturer's instructions. Sequencing was performed on the Illumina NextSeq500 sequencing platform to generate 2 x 75 bp paired end reads.

2.8 Bioinformatics and Data Analysis

2.8.1 Characterisation Statistical Analysis

All statistical analyses were carried out using GraphPad Prism 7.0, with One-way or two-way ANOVA, followed by Dunnett's multiple comparison test or Sidak's multiple comparison test. All P values were multiplicity adjusted.

2.8.2 C. scindens *ATCC 45704 Genome Assembly and Annotation*

The *C. scindens* ATCC 35704 genome was assembled using a combination of PacBio SMRT and Illumina MiSeq technologies, utilising CLC Genomic Workbench 9.5.3 in house (Qiagen).

SMRT sequencing was utilised to generate a single contig to which Illumina MiSeq paired end reads were mapped and a consensus sequence extracted in CLC Genomic Workbench.

The assembled sequence was annotated automatically using the Rapid Annotation Subsystem Technology (RAST) server and by manual curation using the NCBI Basic Local Alignment Tool (BLAST) nucleotide and protein tools.

2.8.3 RNA Sequencing Analysis

RNAseq analysis was carried out by the Deep Seq facility at the University of Nottingham.

A Filtering pipeline was used to exclude reads with low sequencing scores as well as reads aligned to adapter sequences. Raw reads were trimmed by the "skewer" tool (Jiang *et al*., 2014) for processed next generation sequencing paired-end sequence reads.

Reads which passed these filters were mapped on to the *C. scindens* ATCC 35704 reference genome in the context of known gene exon coordinates using the hisat2 mapping tool (Kim *et al*., 2015). Mapped reads were recorded in BAM formatted alignment files and companion BAM index files.

Read alignments were further filtered according to their map quality (MAPQ), a probabilistic metric for the accuracy of read mapping. Calculations were based upon; the

repeat structure of the reference, the base quality of the read and the sensitivity of the alignment algorithm.

For gene expression, read counts per gene were calculated using the htseq-count program, determining the number of uniquely aligned reads per gene (http://www-huber.embl.de/users/anders/HTSeq/doc/count.html).

Counts were normalised by calculating the fragments per kilobase of exon per million reads mapped (FPKM), with rRNA excluded and reads not uniquely aligned and below a MAPQ score of 20.

Principal component analysis (PCA) was calculated to assess the relatedness of biological replicates. This statistical procedure uses an orthogonal transformation to convert a set of observations of possibly correlated variables in to a set of values of linearly uncorrelated variables called principal components. According to the PCA results, pairs of differential expression comparisons were decided upon.

Differential gene expression was calculated using the DESeq package (Anders & Huber, 2010), implementing a single analytical approach. DESeq analyses the variance between biological replicates in order to model the expression values of individual genes within the group of replicates.

Differentially expressed genes were detected using a p-value of < 0.5 and fold-change of > 2 as DESeq thresholds. Genes were filtered out with an average FPKM of < 1.

The heatmaps and hierarchical clustering dendrograms were generated using the 'hclust' R package with a Euclidian distance metric. The *bai* operon differential expression heatmaps were produced using GraphPad Prism 7.0.

Chapter 3 The Effect of The Novel Antibimicrobial Fidaxomicin on Spores of *Clostridium difficile* and The Characterisation of Mutants With Reduced Susceptibility

3.1 Introduction

3.1.1 Fidaxomicin Discovery and Development

Fidaxomicin (FDX) is a macrolide antimicrobial and the first in a new class of macrocyclic antibiotics. It is the first new drug to be approved for the treatment of *C. difficile* associated colitis in over three decades, having endured a particularly convoluted route from discovery to market.

After initial discovery in 1972 at Gruppo Lepetit in Italy, as a fermentation product of a novel species of actinomycete isolated from soil, (Parenti *et al.,* 1975), it wasn't until over a decade later that development continued at Abbott Laboratories in the USA, where the drug was discovered independently from a novel subspecies of *Dactylosporangium autantiacum*, named tiacumicin B (Theriault *et al.,* 1986). Tiacumicin B was out licensed to Optimer Pharmaceuticals in 2000, where it was designated OPT-80. Optimer took OPT-80 through clinical trials and to market after FDA approval in 2011, indicated in the treatment of CDI, as the trade name Dificid under the generic name of fidaxomicin. FDX is marketed in Europe as Dificlir by Astellas Pharmaceuticals.

3.1.2 Fidaxomicin Pharmacokinetics

In the initial phase IA and IB studies conducted by Shue *et al.* on behalf of Optimer, the bioavailability of FDX and its active metabolite OPT-118 was investigated. Healthy volunteers were administered either a single dose of FDX followed up by a second larger dose, or a fixed dosage regimen over a ten day period. In both cases, no FDX was detected in the urine and plasma levels of the drug were close to the lower limit of quantification. Conversely, the levels of FDX recovered from faeces was high, showing a dose dependent

relationship (Shue *et al.*, 2008). These findings were substantiated in the subsequent phase IIA trials, with the mean faecal concentrations at the conclusion of therapy being > 1 mg/kg, well in excess of reported MICs for *C. difficile* (Louie *et al.*, 2011).

The data from these early clinical studies reveals the poor systemic bioavailability of FDX, which without further manipulation, would not be a suitable for systemic treatments. The drug persists at the intracolonic site of infection at high concentrations, exhibiting significant post antibiotic effects. Bacterial inhibition continues for 5.5-10 hours following the last dosage of FDX, in contrast with the 0-1.5 hours post antibiotic effect observed with vancomycin (Babakhani *et al.*, 2011).

3.1.3 DNA-dependent RNA Polymerase

In bacteria, all classes of RNA are synthesised by a single DNA-dependent RNA polymerase (RNAP). Essential for cell viability, RNAP is a proven target for antimicrobial therapy, with FDX belonging to a group of antibiotics, termed RNAP inhibitors. The RNAP catalytic core is comprised of the α_2 subunit, β and β' subunits and the ω subunit with the association of transcription initiating σ factors making up the RNAP holoenzyme (Figure 3.1). These σ factors are required for promoter recognition and unwinding, and *de-novo* RNA synthesis.

The α_2 subunit is a dimerization of two α molecules (α^{I} and α^{II}), each coded for by the *rpoA* gene. These two molecules fulfil different roles within the core enzyme, with one located adjacent to the β subunit and the other, the β' subunit.

The β and β' subunits form the catalytic core of RNAP, both containing determinants for the non-sequence specific binding of template DNA and the DNA/RNA hybrid, forming

the transcription bubble of approximately 10-12 bp (Kent *et al*., 2009). The larger β' subunit is encoded by *rpoC* and the smaller β subunit by *rpoB*.

The ω subunit, is the smallest constituent of the RNAP core enzyme and coded for by the *rpoZ* gene. It has been shown, through cross linking studies, that ω binds to β' helping to recruit and maintain the subunit's conformational structure in the core enzyme assembly (Gentry & Burges, 1993).

There are a variable number of sigma factors with different promoter recognition properties responsible for regulating the transcription of unique sets of genes. Sigma factors transiently associate with the core enzyme allowing an organism to express genes in response to environmental conditions.

During the initiation of transcription, the RNAP core enzyme combines with a promoter specific σ factor to form the holoenzyme. This enables the locating and binding of specific promoter DNA, forming the closed RNAP-promoter DNA complex (RP_C). The next step is the melting of the promoter DNA, separating the coding and non-coding strands, transitioning to the RNAP open promoter complex (RP_O) and forming the transcription bubble. Incoming nucleoside triphosphates pair with the unwound template DNA in the active site of the enzyme sequentially, as the transcription bubble progresses along the DNA in the process called elongation, until the terminator is reached.

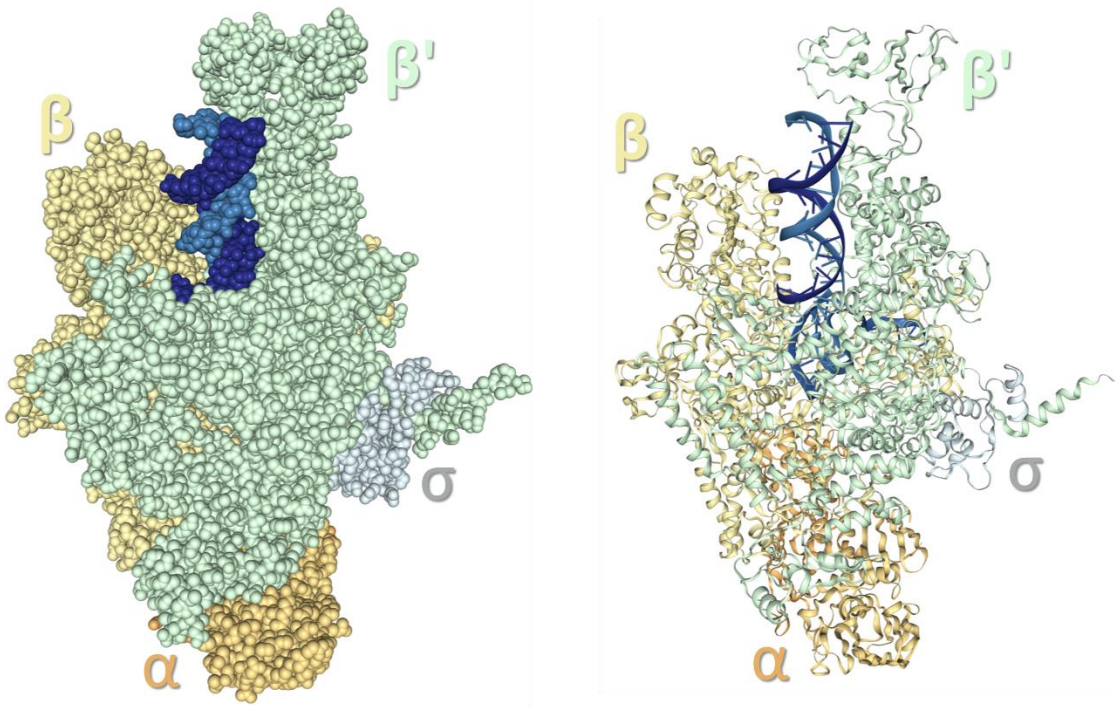


Figure 3.1 Spacefill and Cartoon Crystal Structures of *Thermus thermophilus* RNAP with DNA in the Active Cleft: α subunit in **orange**, β subunit in **yellow** β' subunit in **green**, σ subunit in **grey** and DNA in **blue**. PDB; 4WQS.

3.1.4 Fidaxomicin Spectrum of Activity

The highly conserved bacterial RNAP subunits allows efficacy over a broad spectrum of Gram-positive and Gram-negative bacterial infections, whilst the lack of similarity to eukaryotic RNAPs I, II and III enables selective toxicity. Despite the level of conservation between the components of the RNAP core enzyme amongst bacteria, there is variation due to the imposed constraints by the complex regulatory networks modulating gene expression, allowing for the possibility of both broad and narrow spectrum activity of RNAP inhibitors (Lane & Darst, 2010). FDX is reported to have narrow spectrum activity, predominantly active against Gram-positive anaerobes and relatively poor activity against the anaerobic Gram-negative organisms inhabiting the gut, (Credito *et al.,* 2004,

Finegold *et al.,* 2004). FDX has been shown to have less perturbing effect on the intestinal microbiota than vancomycin (Louie *et al*., 2009).

Bacterial RNAP is a relatively underexploited drug target. Until FDX was approved, the rifamycins were the only other class of RNAP inhibitors in clinical use. Rifampicin (RIF), a potent RNAP inhibitor, is indicated as a first line therapy in the treatment of mycobacterial infections such as tuberculosis and leprosy. RIF inhibits protein synthesis by preventing the extension of the RNA chain exiting the RNAP holoenzyme, limiting the transcript to 2 - 3 nt in length. This effect is exerted through the binding of the drug within the cleft of the β subunit, greater than 12 Å away from and adjacent to the active site (figure 3.2) (Campbell *et al*., 2001).

3.1.5 Fidaxomicin Mechanism of Action

In contrast to the rifamycins, FDX belongs to a group of RNAP inhibitors that are referred to as "switch-region inhibitors" (Srivastava *et al*., 2011). Switch inhibitors bind to and inactivate structural components known as "switches" on the β and β' subunits of the RNAP core enzyme. The switch region is segmented in to five components, switches 1 – 5, which undergo conformational changes required for the loading and retention of melted promoter DNA in the active cleft of the enzyme. This occurs by a hinge and clamp like mechanism (Tupin *et al*, 2010). The switch region represents a promising target for antimicrobial therapy with the clinical efficacy of the rifamycin class under the threat by the emergence of resistant bacteria (Chopra, 2007).

The natural antibiotics myxopyronin (MYX), corallopyronin (COR) and ripostatin (RIP) are further examples of RNAP switch inhibitors. The binding sites within RNAP for MYX, COR and RIP were identified by isolation and subsequent sequencing of resistant *E. coli* clones following mutagenesis of the RNAP genes. Reduced susceptibility was conferred

by substitutions in the *rpoB* and *rpoC* genes. There is comprehensive overlap between the binding sites for RIP, COR and MYX due to near identical mechanisms of action. This mechanism is due to interference with RNAP clamp conformational states and DNA contact by interaction with switches 1 and 2 in the core enzyme (figure 3.1, table 3.2) (Mukhopadhyay *et al.,* 2008).

The FDX binding site has been elucidated by the sequencing of *B. subtilis* (Gualtieri *et al.*, 2006), *Mycobacterium tuberculosis* (Kurabachew *et al*., 2008) and *Enterococcus faecalis* (Gualtieri *et al*., 2009) mutants with reduced susceptibilities to the FDX. As with the aforementioned switch inhibitors, substitutions in the *rpoB* and *rpoC* genes were implicated, but the substituted residues indicated a binding site distinct to RIP, COR and MYX, with minimal overlap. Mutations clustered in the N-terminus of the β' subunit and the C-terminus of the β subunit (Figure 3.2). The mechanism of action is partially similar, but it is also proposed that FDX may interfere with the growing RNA chain, through interacting with an area encompassing the RNA exit channel and parts of switches 2 and 3 in the β subunit (Srivastava *et al*., 2011) (Table 3.1). Though the exact mechanism of action is still unclear, it is known that FDX traps RNAP at a closed, RP_C, complex intermediate during isomerisation to the transcriptionally functional open promoter complex, RP_O. Although all of these antibiotics target RNAP, due to their unique binding sites and consequent mechanisms of action, no cross resistance has been observed (Srivastava *et al*., 2011, Babakhani, 2014).

Table 3.1. RNAP inhibitors with binding sites and proposed mechanisms of action.

RNAP Inhibitor	Binding Site	Proposed Mechanism of Action
Rifampicin	Adjacent to the active site	Prevents RNA chain extension
Ripostatin Corallopyronin Myxopyronin	Switch 1/2	Interfering with clamp opening and promoter binding
Fidaxomicin	Switch 2/3, RNA exit channel	Interfering with clamp opening, promoter binding, RNA exit

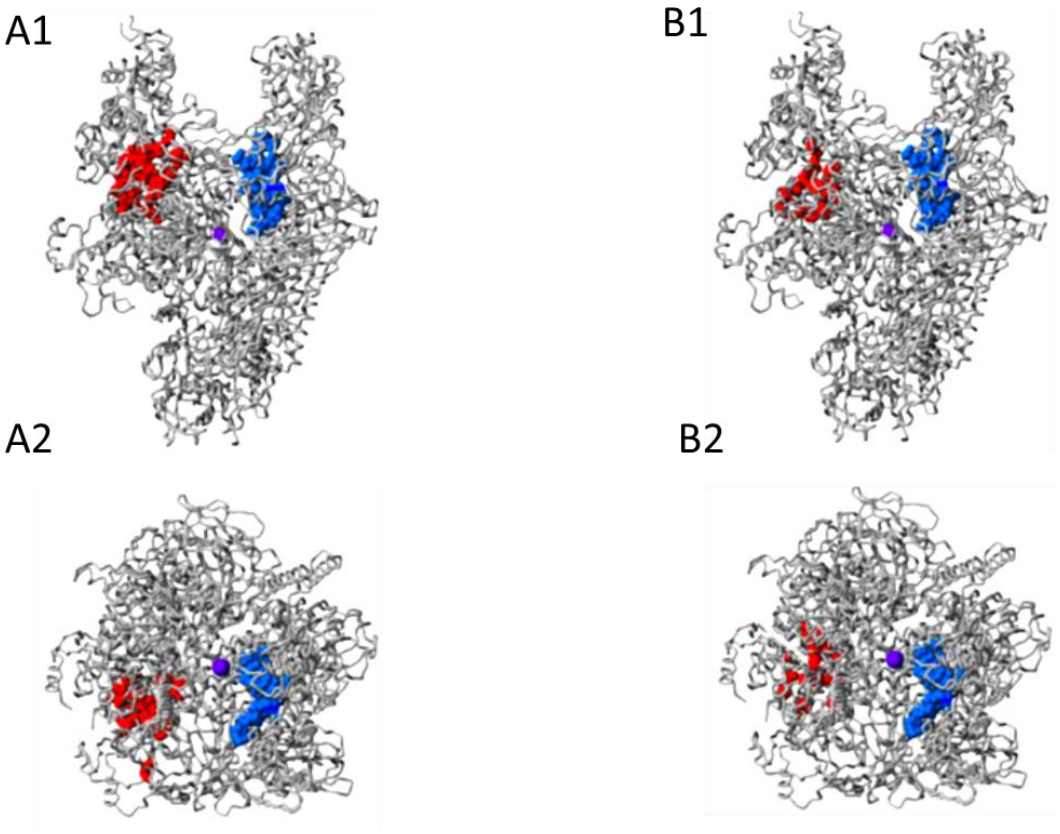


Figure 3.2 Orthogonal views of RNAP with highlighted binding sites of RNAP inhibitors. **A1** and **A2**: The binding site for **FDX** at the N-terminus of the switch-2 and C-terminus of switch 3 regions shown in red. **B1** and **B2**: The binding site for **MYX**, **COR** and **RIP** at the switch 1 and C-terminus of switch 2 shown in red. **RIF** binding site adjacent to the active site shown in blue. The purple sphere represents Mg^{2+} bound to the active centre. (adapted from Ho *et al*., 2009)

3.1.6 Fidaxomicin Resistance

Reduced susceptibility to FDX has been shown to develop under selective pressure of fidaxomicin via serial passage (Leeds *et al.,* 2013), but to date there has only been one reported case of resistance emerging in the clinical setting. During phase III clinical trials, *C. difficile* isolates were recovered at baseline and end of therapy. During the trial, no resistance was observed. Although once the therapy concluded, a recurrent case from the trial yielded one isolate exhibiting an elevated FDX MIC of 16 µg/mL (Goldstein *et al*.,

2011). Sequencing of this isolate revealed a possible association between the observed reduced susceptibility and a point mutation in the *rpoB* gene; substituting the valine residue at 1143 with glycine.

Seddon and colleagues isolated laboratory derived, spontaneous mutants with reduced susceptibilities to FDX by exposing dense cultures of *C. difficile* to FDX at 4 and 8 times the MIC. Sequencing revealed mutations in the *rpoB* and *rpoC* genes, one of which resulted in the previously identified glycine substitution of valine at the 1143 residue in the β subunit, (Table 3.2). (Seddon *et al.,* 2011)

Table 3.2: SNPs and Amino Acid Substitutions Conferring Reduced Susecptibility to FDX in spontaneous mutants of *C. difficile*. Adapted from Seddon *et al.*, 2011

Gene	Nucleotide Change	Amino Acid Change	MIC
rpoB	T3428A	Val1143Asp	>64
rpoB	T3428G	Val1143Gly	16
rpoB	G3427T	Val1143Phe	2
rpoC	G709T	Asp237Tyr	2

3.1.7 C. difficile *Ribotypes of Study*

At the time this study was initiated, the *C. difficile* strains selected were six of the most prevalent and clinically relevant PCR-ribotypes in Europe (Table 3.3). When studying the germination of *C. difficile* spores, it is vital to sample a diverse selection of strains, due to the diverse responses in rate and extent between different strains. (Heeg & Burns *et al*., 2012).

Table 3.3 Strain Designations and PCR-Ribotypes

Strain Designation	Ribotype
R12801	106
Serogroup F	017
DH1916	027
L26	001
EK15	014
EK28	078

Coincident with the *C. difficile* PCR-ribotype 027 Stoke Mandeville hospital outbreaks between 2003 and 2005, a surveillance programme was initiated to assess the true extent of the prevalence of ribotype 027 strains in English hospitals. This surveillance study identified PCR-ribotypes 106, 027 and 001 as the most prevalent strains, accounting for 26 %, 25 % and 24 % respectively, from 881 symptomatic patients in English hospitals (Brazier *et al*., 2007). In a subsequent hospital surveillance of *C. difficile* cases conducted by Brazier and colleagues covering 2007 to 2008, PCR-ribotype 027 strains accounted for 41.3 % of all isolates, followed by ribotype 106 and 001 strains representative of 20.2 % and 7.8 %, respectively (Brazier *et al.,* 2008).

A Europe-wide hospital-based surveillance study conducted in 2008 revealed PCR-ribotypes strains 014, 001 and 078 were the 3 most prevalent ribotypes in Europe, accounting for 15.7 %, 9.3 % and 7.8 %, respectively (Bauer *et al.,* 2011). In the same study, PCR-ribotypes 106, 027, and 017, accounted for 5.1 %, 4.8 % and 3.5 % of isolates in Europe.

The recent 2016 EUCLID surveillance study (European, multicentre, prospective, biannual, point-prevalence study of *Clostridium difficile* infection in hospitalised patients with diarrhoea) reported a change in the landscape of *C. difficile* ribotypes across Europe. PCR-ribotype 027 strains are now reported to be most represented amongst clinical isolates in Europe with a 19 % prevalence. PCR-ribotypes 001 and 014 occupy the second and third most prevalent in Europe, representing 11 % and 10 % of all isolates, respectively. PCR-ribotypes 078, 017 and 106 have shown a sharp decline in incidence across Europe. Since being the third most commonly isolated ribotype in 2008, ribotype 078 strains currently only account for 3 % of isolates, being the seventh most prevalent. PCR-ribotype 106 strains account for 0.6 % and ribotype 017 strains less than 5 % in Southern Europe alone (Davies *et al*., 2016).

Of the strains investigated in this study, PCR-ribotypes 027 and 001 have increased in prevalence, whilst ribotypes 078, 106, 017 and 014 have declined. PCR-ribotypes 027, 001, 014 and 078 remain significantly prevalent in the current European *C. difficile* epidemiological landscape, (Figure 3.17).

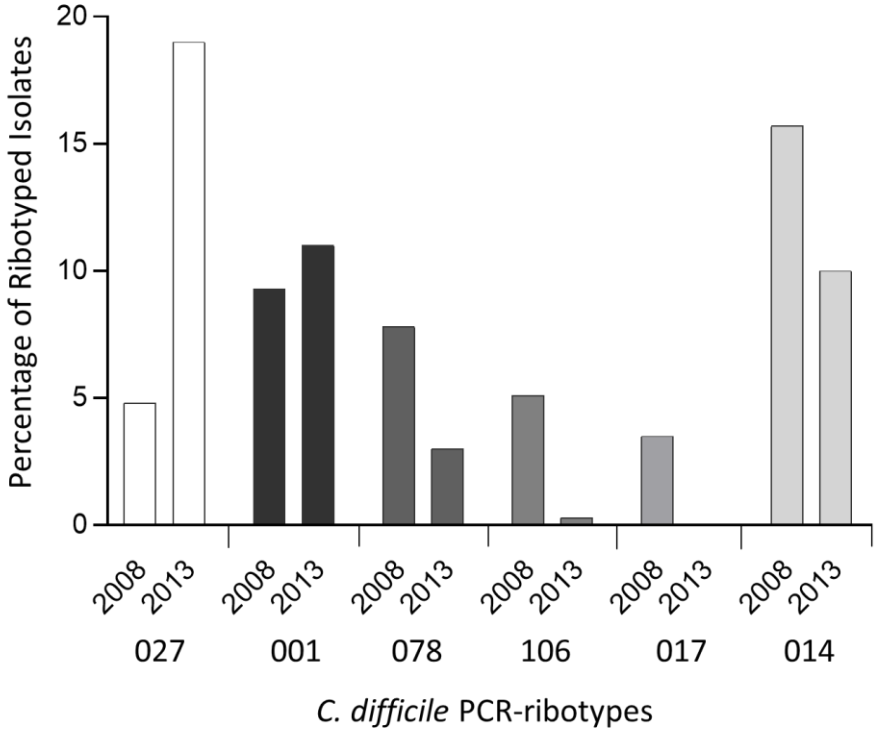


Figure 3.17 Prevalence of *C. difficile* Strains in Europe: The European prevalence of the PCR-ribotypes used in this study. Data adapted from Bauer *et al*., 2008 and Davies *et al*., 2016.

3.1.2 Aims and Hypotheses

The first aim of this work is to determine the effect of FDX on the germination of a selection of *C. difficile* clinical isolates. To accomplish this, the initiation of spore germination and the outgrowth of spores will be assayed separately. This will allow a better understanding of whether FDX inhibits the germination of *C. difficile* spores and at which point the inhibition occurs. The hypothesis is that the RNAP inhibiting drug, FDX, would not halt the gene expression independent process of germination initiation, but would inhibit the vegetative outgrowth of germinated spores.

The second aim is to provide definitive proof that previously identified nucleotide substitutions in the *rpoB* gene of *C. difficile* confers a reduction in susceptibility to FDX. This will be achieved through the creation of isogenic mutants harbouring only the intended SNPs in the *rpoB* gene through allelic exchange. As well was demonstrating the relationship between these target modifications and reduced susceptibility to FDX, this will also aid in validating the hypothesis that the binding site of FDX is located in the switch region of the RNAP core enzyme and the mechanism of action of the drug.

It is also hypothesised that the modifications to the *rpoB* gene conferring resistance to FDX would incur a fitness cost. To test this hypothesis, the fitness of these mutants compared to the parental strain will also be assessed in terms of growth, competition, sporulation, germination toxin production and virulence *in vivo* to predict the importance of these mutations in disease and the clinical setting.

It should be noted that in the following results section, the *in vivo* hamster study was carried out by Michelle L. Kelly and Rory Cave.

The author would also like to acknowledge Dr Mark M. Collery, who was instrumental in assisting with the characterisation assays in this chapter.

Portions of this work have been published in The Journal of Antimicrobial Chemotherapy (Kuehne *et al*., 2017)

3.2 Results

3.2.1 Fidaxomicin Minimum Inhibitory Concentrations

The FDX MIC values for a range of *C. difficile* clinical isolates were established for use in subsequent germination and outgrowth assays. These values were ascertained using the broth dilution method as described in 2.3.2. Cultures were grown for 24 hours in a range of concentrations of FDX (0 - 5.0 μg/mL), with the MIC being taken as the lowest concentration of FDX at which the OD_{600} of the culture did not exceed the starting OD_{600} (Figure 3.3).

Preliminary data indicated that the strains DH1916, R12801 and Serogroup F appeared to exhibit higher MICs than the strains L26, EK15 and EK28, and so separate concentration ranges were used in these assays.

DH1916, R12801 and the Serogroup F strain showed a slightly elevated MIC of 0.05 μg/mL, when compared to strains L26, EK15 and EK28 which were inhibited at a concentration of 0.03 μg/mL. The strains can therefore be broadly divided in to two groups in terms of the determined FDX MICs (Table 3.4).

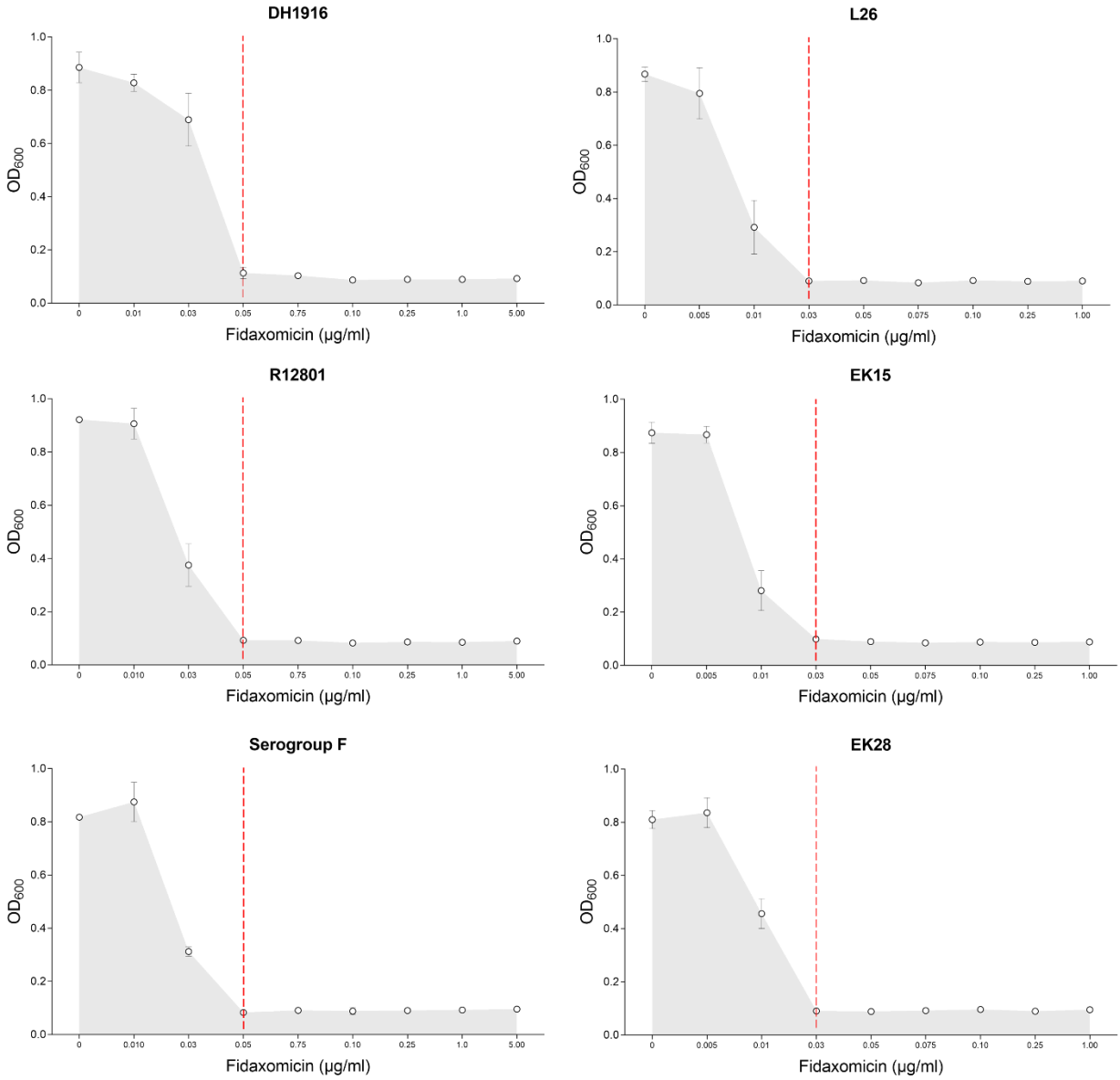


Figure 3.3: FDX Minimum Inhibitory Concentration of *C. difficile* Isolates: Final optical density measurements of *C. difficile* cultures incubated for 24 h at a range of concentrations of fidaxomicin. The red dashed line delineates the MIC for each strain. Plotted values are mean ± standard deviation (SD).

Table 3.4: Minimum inhibitory concentrations of the selected *C. difficile* strains

Strain Designation	Ribotype	MIC (µg/mL)
R12801	106	0.05
Serogroup F	017	0.05
DH1916	027	0.05
L26	001	0.03
EK15	014	0.03
EK28	078	0.03

3.2.2 The Effect of Fidaxomicin on the Initiation of Spore Germination

To investigate the effect of FDX on the initiation of *C. difficile* spore germination, spore preparations of at least 95% purity were prepared for each strain (2.3.4). Spores were heat treated at 65 ^{o}C for 30 minutes and incubated in BHIS broth, supplemented with 1 % TC, 1 % TC and FDX at MIC, 2 × MIC and 4 × MIC. BHIS without germinant supplementation was used as a negative germination control. OD_{600} measurements were taken every minute for 10 minutes and then every 10 minutes, over an 80 minute period (2.3.9). Germination initiation was detected as a decrease in OD_{600} as the spores rehydrate, becoming less optically dense and losing spore specific characteristics.

A decrease in OD_{600} was observed in all cases, except in the negative control, indicating that in the presence of FDX at concentrations in excess of MIC, spores are able to progress through the initial stages of TC mediated germination (Figures 3.3.1 and 3.3.2).

DH1916 spores showed an average drop in OD_{600} of 19.4 %, with an average final OD_{600} measurement of 0.81. The most significant average decrease of 11 % (OD_{600} drop of 0.12) was observed between 9 and 10 minutes (Figure 3.4.1).

R12801 spores showed average drop of 20.8 % in OD_{600}, with an average final OD_{600} measurement of 0.79. The most significant average decrease of 17 % (OD_{600} drop of 0.18) was observed between 0 and 3 minutes (Figure 3.4.1).

Serogroup F spores showed an average drop in OD_{600} of 18.6 %, with an average final OD_{600} measurement of 0.81. The most significant average decrease of 14 % (OD_{600} drop of 0.14) was observed between 0 and 3 minutes (Figure 3.4.1).

L26 spores showed an average drop in OD_{600} of 19.3 %, with an average final OD_{600} measurement of 0.81. The most significant average decrease of 14 % (OD_{600} drop of 0.14) was observed between 0 and 3 minutes (Figure 3.4.2).

EK15 spores showed an average drop in OD_{600} of 20.3 %, with an average final OD_{600} measurement of 0.8. A gradual decrease of 19.9 % (OD_{600} drop of 0.2) was observed between 0 and 7 minutes (Figure 3.4.2).

EK15 spores showed an average drop in OD_{600} of 20.3 %, with an average final OD_{600} measurement of 0.84. A gradual decrease of 15.2 % (OD_{600} drop of 0.15) was observed between 0 and 10 minutes (Figure 3.4.2).

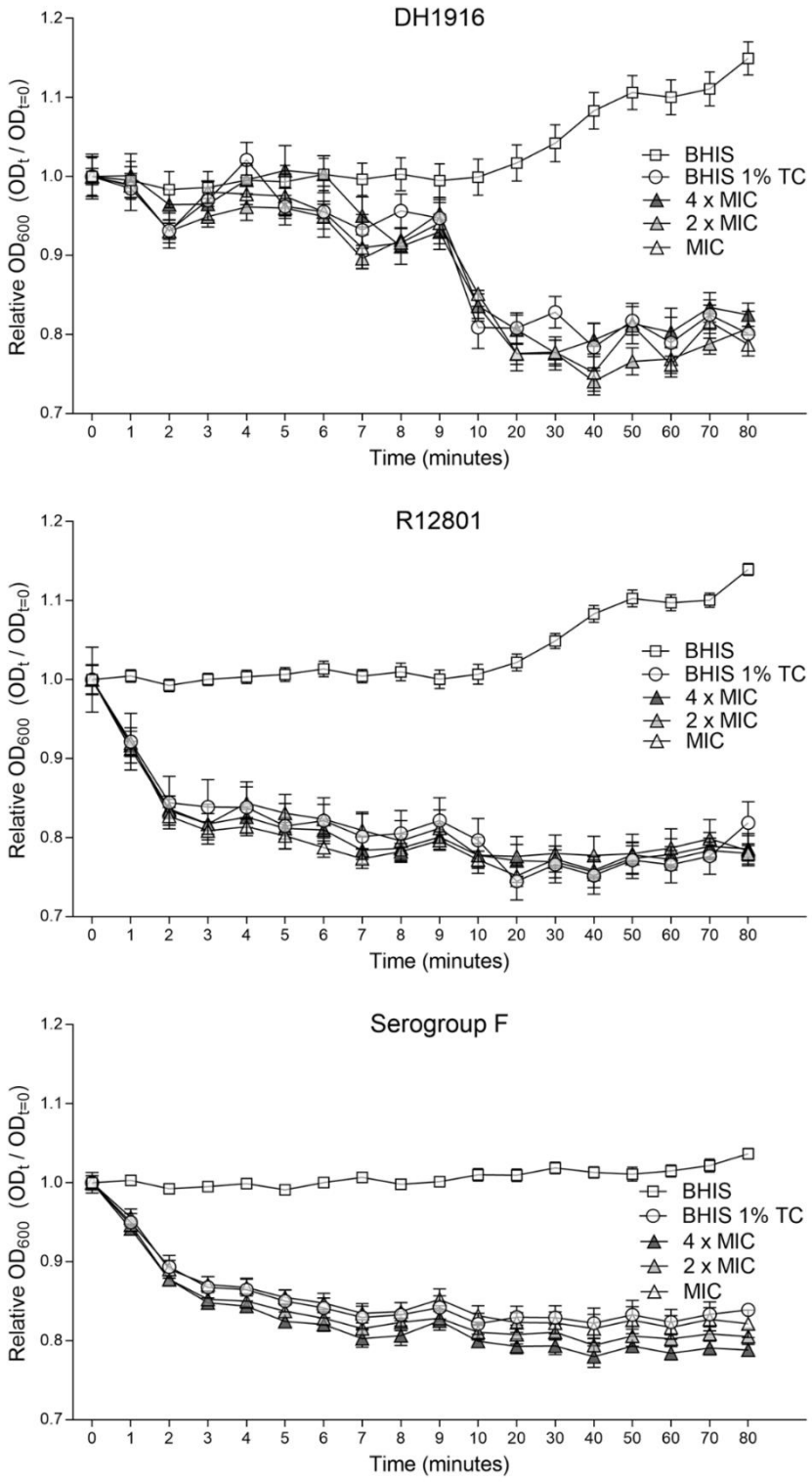


Figure 3.4.1 Initiation of Germination: Spores of *C. difficile* isolates were challenged with FDX at MIC, 2 and 4 × MIC in the presence of the germinant TC. The initation of germination for all strains proceeded at all concentrations of FDX with observed decreases in OD_{600}, with no drop in OD_{600} observed for the negative no germinant controls. Plotted values are mean ± SD, normalised to the starting OD_{600} value at $T_{(0)}$.

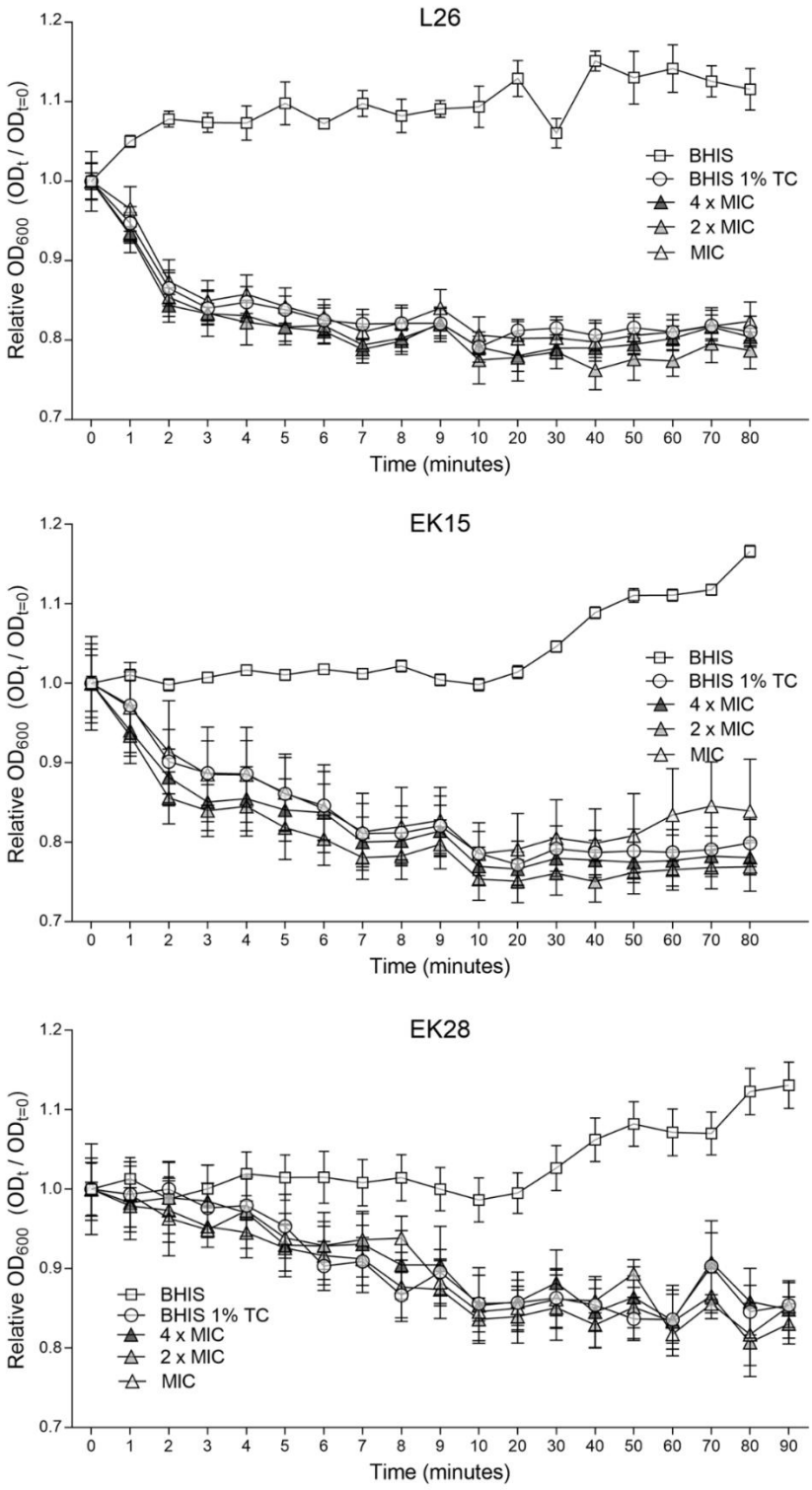


Figure 3.4.2 Initiation of Germination: Spores of *C. difficile* isolates were challenged with FDX at MIC, 2 and 4 × MIC in the presence of the germinant TC. The initation of germination proceeded for all strains at all concentrations of FDX with observed decreases in OD_{600}, with no drop in OD_{600} observed for the negative no germinant controls. Plotted values are mean ± SD normalised to the starting OD_{600} value at $T_{(0)}$.

3.2.3 The Effect of Fidaxomicin on the Outgrowth of Spores

To investigate the effect of FDX on the outgrowth of *C. difficile* spores, spore preparations of at least 95 % purity were prepared for each strain (as outlined in 2.3.4). Spores were heat treated at 65 ^{o}C for 30 minutes and incubated in BHIS broth, supplemented with 1 % TC, 1 % TC and FDX at ¼ MIC, ½ MIC and MIC. OD_{600} measurements were taken every hour over a 15 h period (2.3.10).

Spores for all strains incubated in the presence of the germinant TC germinated and grew out, returning to vegetative growth. An initial drop in OD_{600} was observed for the DH1916, R12801, Sergroup F and EK28 strains between the 0 and 1 hour time points. This could be explained as a loss of optical density of the cultures due to the germination of spores prior to outgrowth.

All strains incubated with TC and FDX at MIC were inhibited, as with ½ MIC except for strain EK15 which exhibited delayed outgrowth at 10 h. All strains showed delayed outgrowth when incubated with 1/4 MIC (Figure 3.5)

DH1916 spores incubated with TC germinated and grew out. Germinated spores progressed from lag phase at 4 h entering exponential and finally stationary phase at 9 h, reaching a peak OD_{600} of 0.94. The outgrowth of spores incubated at both MIC and ½ MIC was inhibited. Spores incubated at ¼ MIC appeared to exhibit delayed outgrowth, with the OD_{600} of these cultures starting to increase at the 14 h time point (Figure 3.5).

R12801 spores incubated with TC germinated and grew out. Germinated spores progressed from lag phase at 7 h entering exponential and finally stationary phase at 11 h, reaching a peak OD_{600} of 0.87. Spores incubated at ¼ MIC showed a slightly delayed outgrowth, entering exponential phase at 10 h and finally stationary phase at 13 h, reaching a peak OD_{600} of 0.85. The difference between growth data for the TC and ¼ MIC

conditions was statistically significant (P <0.0001). The outgrowth of spores incubated at both MIC and ½ MIC was inhibited (Figure 3.5).

Serogroup F spores incubated with TC spores germinated and grew out. Germinated spores progressed from lag phase at 7 h entering exponential and finally stationary phase at 12 h, reaching a peak OD_{600} of 0.94. Spores incubated at ¼ MIC exhibited delayed outgrowth, with the OD_{600} of these cultures starting to increase at the 13 h time point. The outgrowth of spores incubated at both MIC and ½ MIC was inhibited (Figure 3.5).

L26 spores incubated with TC germinated and grew out. Germinated spores progressed from lag phase at 4 h entering exponential and finally stationary phase at 9 h, reaching a peak OD_{600} of 0.85. Spores incubated at ¼ MIC showed a slightly delayed outgrowth, entering exponential phase at 6 hours and finally stationary phase at 11 h, reaching a peak OD_{600} of 0.8. The difference between growth data for the TC and ¼ MIC conditions was statistically significant (P <0.0001). The outgrowth of spores incubated at both MIC and ½ MIC was inhibited (Figure 3.5).

EK15 spores incubated with TV germinated and grew out. Germinated spores progressed from lag phase at 4 h entering exponential and finally stationary phase at 8 h, reaching a peak OD_{600} of 0.9. Spores incubated at ¼ MIC showed a slightly delayed outgrowth, entering exponential phase at 6 h and finally stationary phase at 10 h, reaching a peak OD_{600} of 0.92. The difference between growth data for the TC and ¼ MIC conditions was statistically significant (P <0.0001). Spores incubated at ½ MIC exhibited further retardation of outgrowth, with the OD_{600} of cultures starting to increase at 11 h. The outgrowth of spores incubated at MIC was inhibited (Figure 3.5).

EK28 spores incubated with TC germinated and grew out. Germinated spores exited lag phase at 5 h, entering exponential and finally stationary phase at 9 h, reaching a peak OD_{600} of 0.61. Spores incubated at ¼ MIC exhibited slightly delayed outgrowth, entering exponential at 8 hours, and finally stationary phase at 13 h, reaching a peak OD_{600} of 0.61.

The difference between growth data for the TC and ¼ MIC conditions was statistically significant ($P < 0.0001$). Spores incubated at ½ MIC appeared to initiate outgrowth at the 14 h time point, with an increase in OD_{600} of the culture. The outgrowth of spores incubated at MIC was inhibited (Figure 3.5).

Microscopic analysis (2.3.5) of cultures after incubation with FDX reveals that spores appear to be phase dark, consistent with a loss of spore specific characteristics, whereas those incubated without FDX are able to return to vegetative cell growth (Figure 3.6.1/2)

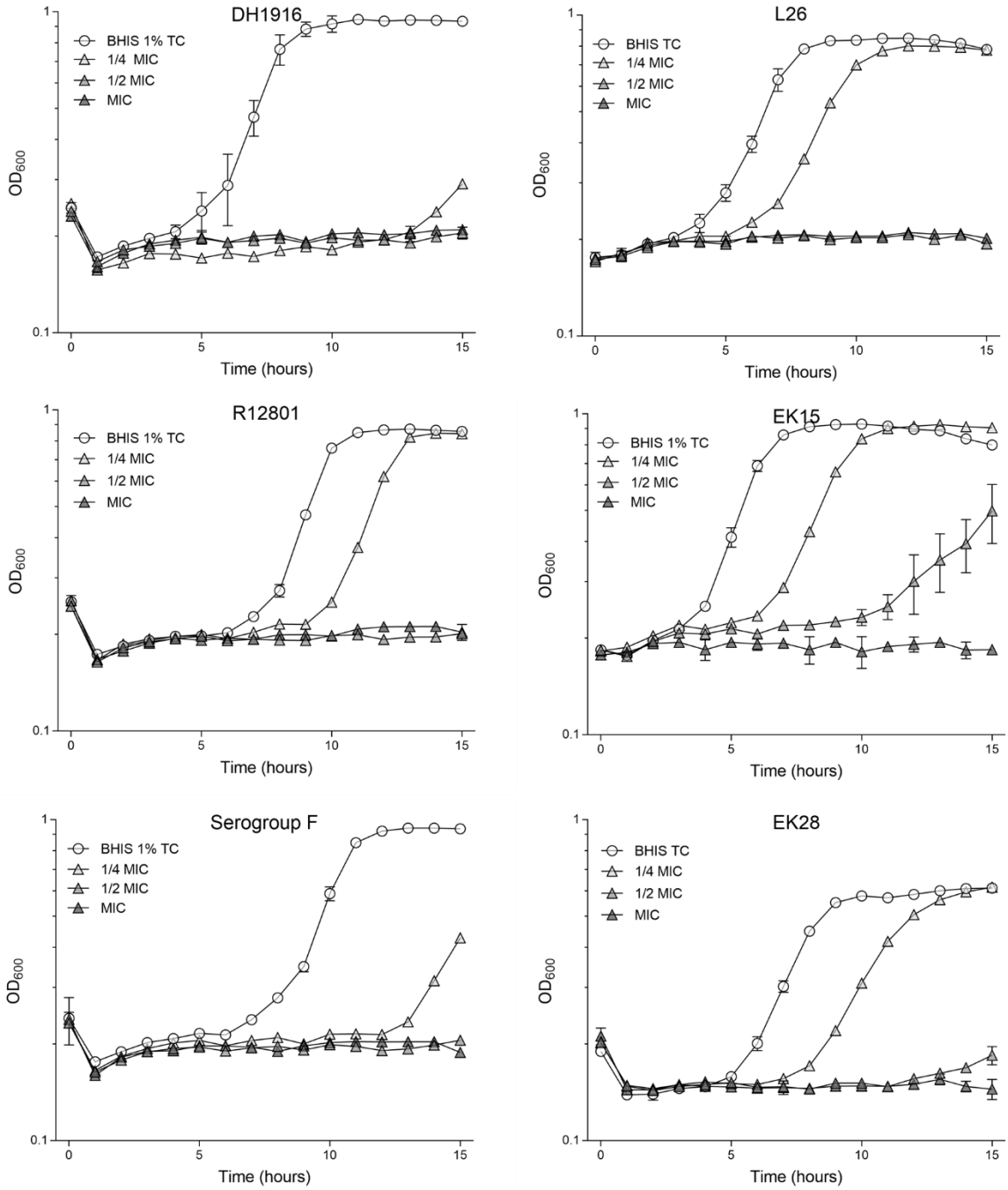


Figure 3.5 Spore Outgrowth: Spores of *C. difficile* isolates were challenged with FDX at MIC, ½ and ¼ MIC in the presence of the germinant TC. Spores for all strains incubated with TC germinated and grew out, with an initial drop in OD_{600} for strains DH1916, R12801, Serogroup F and EK28. All strains incubated with FDX at MIC were inhibited in outgrowth, and at least delayed at ½ and ¼ MIC. Plotted values are mean ± SD.

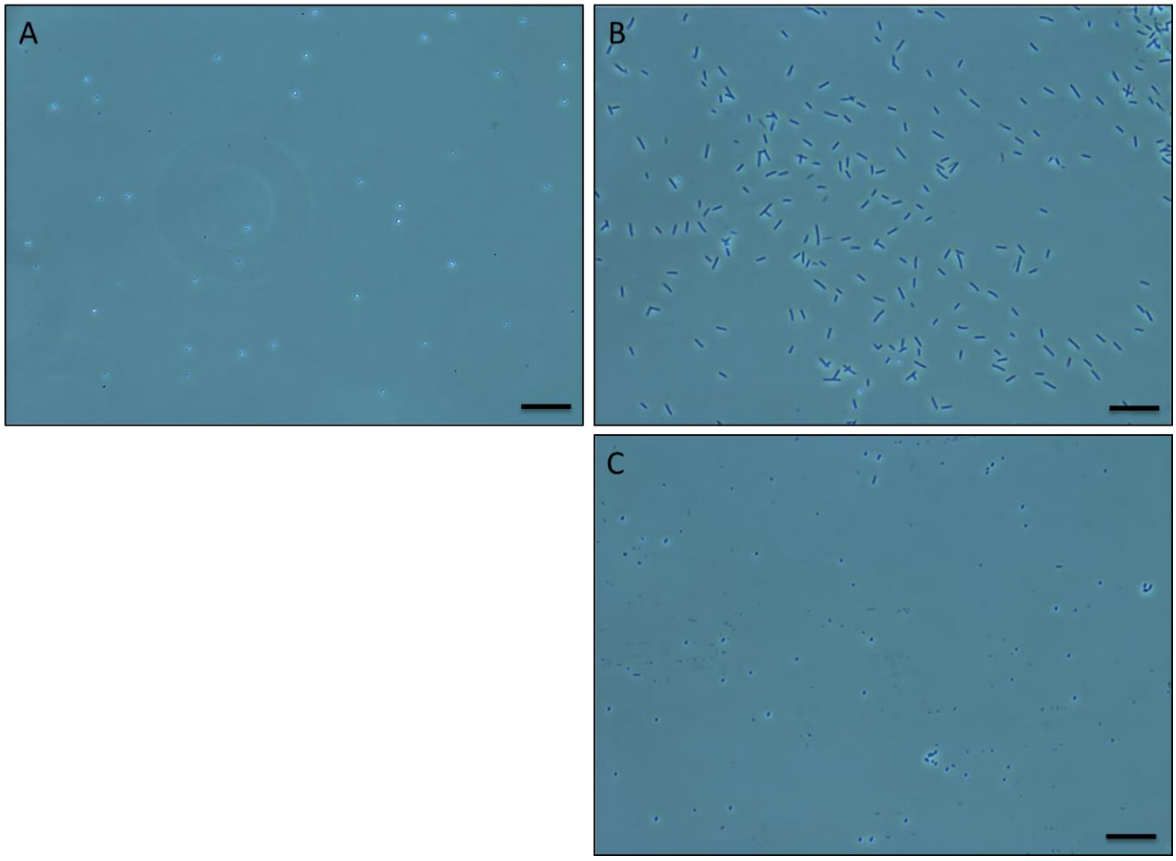


Figure 3.6.1 Phase Contrast Microscopy of Spores and Vegetative Cells: A: Phase bright spores prior to incubation with TC, **B**: Vegetative cells after incubation with TC and no FDX, outgrowth is permitted. **C**: Spores which have been incubated with TC and FDX at MIC. Germination is initiated but spores are unable to grow out in to vegetative cells, remaining phase dark as pre-vegetative structures. At 40 times magnification, black bars representing 10 µm.

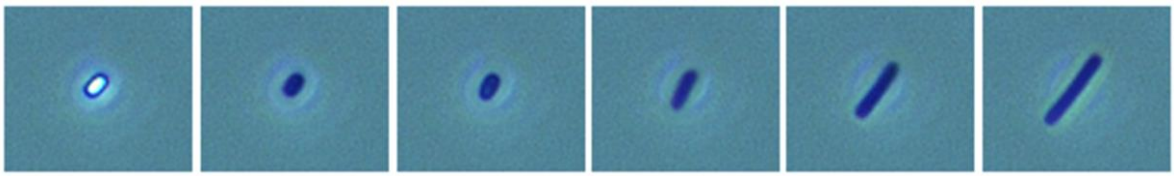


Figure 3.6.2 Phase Contrast Microscopy of Germination and Outgrowth of Spores: Composite image of *C. difficile* at different stages of TC mediated germination and outgrowth, progresing from a phase bright spore to a vegetative cell.

3.2.4 Construction of rpoB Mutatants

The desired single nucleotide changes (Table 3.2) were introduced in to the genome of *C. difficile* strain R20291 by point mutation using allelic exchange methodology (Ng *et al.*, 2013). The vector pMTL-YN4 was employed, which carries a functional copy of the *pyrE* gene, enabling counter selection for double cross over events when used in conjunction with a *pyrE* deficient strain.

An allelic exchange cassette, flanked by the restriction sites MluI and SacI was synthesised and supplied by Biomatik. The sequence comprised the *rpo* promoter and 500 bp regions of homology to *rpoB*, either side of the desired nucleotide change for T3428A. After digestion with MluI and SacI, the DNA fragment was ligated in to the vector pMTL-YN4 following appropriate digestion and dephosphorylation (2.5.6), to yield pMTL-YN4-T3428A (Figure 3.8). The correct construct was verified with restriction digest and confirmed with Sanger sequencing (Figure 3.7).

Site directed mutagenesis (2.5.14) was used to create the remaining two allelic exchange vectors, T3428G and G3427T, which were verified by Sanger sequencing (2.7.1).

Each of the three plasmids harbouring the verified allelic exchange cassettes were transformed in to *E. coli* CA434 conjugal donor cells by electroporation (2.5.11) and selected for by plating on to LB agar plates supplemented with chloramphenicol. The plasmids were transferred in to *C. difficile* R20291 Δ*pyrE* by conjugation (2.3.8) and transconjugants subsequently selected for by plating on to BHIS agar supplemented with D-cycloserine, cefoxitine and thiamphenicol. Based upon the hypothesis that double cross-over clones would be resistant to FDX, purified colonies were then re-streaked on to BHIS supplemented with FDX at 0.2 μg/mL, (4 × MIC of the parental strain). This allowed for rapid selection of double cross over clones. Growing colonies were then re-

streaked to purity, genomic DNA extracted (2.5.1) and PCR performed to amplify the region of *rpoB* harbouring the SNPs. The mutants were verified by Sanger sequencing.

Mutants were confirmed for plasmid loss, genomic DNA extracted by phenol-chloroform extraction (2.5.1) and genome sequences obtained. The mutant genomes were compared to that of the parental strain, confirming that the only changes in the genome were the desired SNPs in *rpoB*.



Figure 3.7 Alignments of Wild Type and Mutant rpoB Sequences: *rpoB* genes for each mutant were sequenced using Sanger sequencing and the sequences aligned to the parental R20291 *rpoB* sequence using the Clustal Omega algorithm. The intended SNPs are highlighted for each mutant sequence.

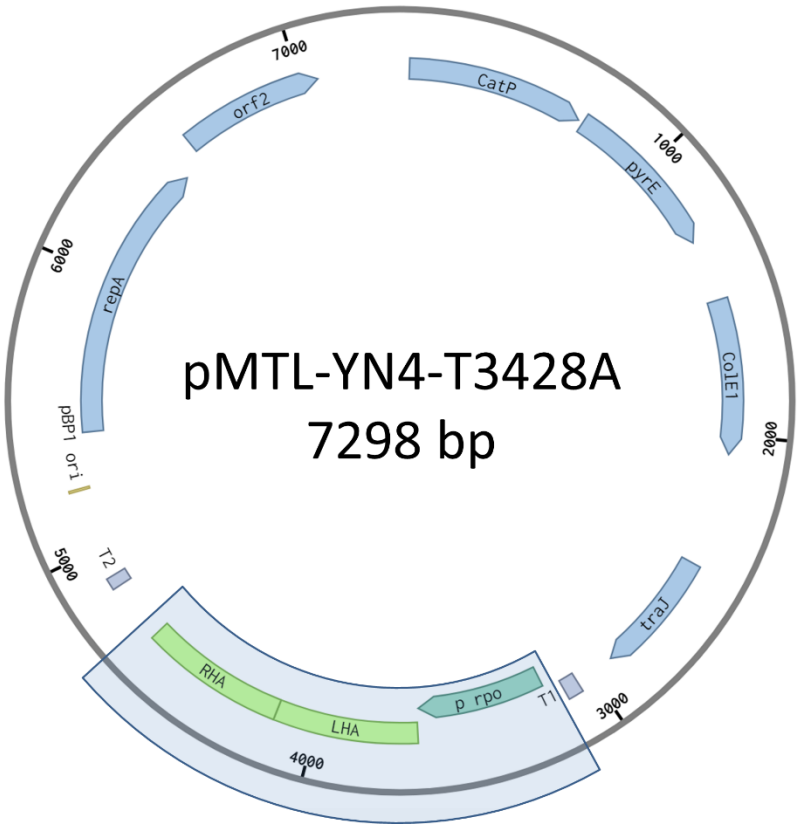


Figure 3.8 pMTL-YN4-T3428A: The constructed allelic exchange vector to introduce the nucleotide change, T3428A, in to the R20291 genome. Allelic exchange cassette comprised of left and right homology arms and the desidered nucleotide change highlighted in the blue segment.

3.2.5 Minimum Inhibitory Concentration

The MIC of FDX was determined for each mutant and the parental strain using the broth dilution method as described in 2.3.2. Cultures were inoculated at a range of concentrations of FDX, with the OD_{600} measured post incubation after 24 hours. The MIC was determined to be the lowest concentration of FDX at which the OD_{600} of the culture did not exceed the starting OD_{600} (Figure 3.9). The parental strain R20291 Δ*pyrE* has an MIC of 0.05 µg/mL with elevated MICs of > 32.0 µg/mL, 8.0 µg/mL and 3.0 µg/mL for the mutants T3428A, T3428G and G3427T respectively (Table 3.5).

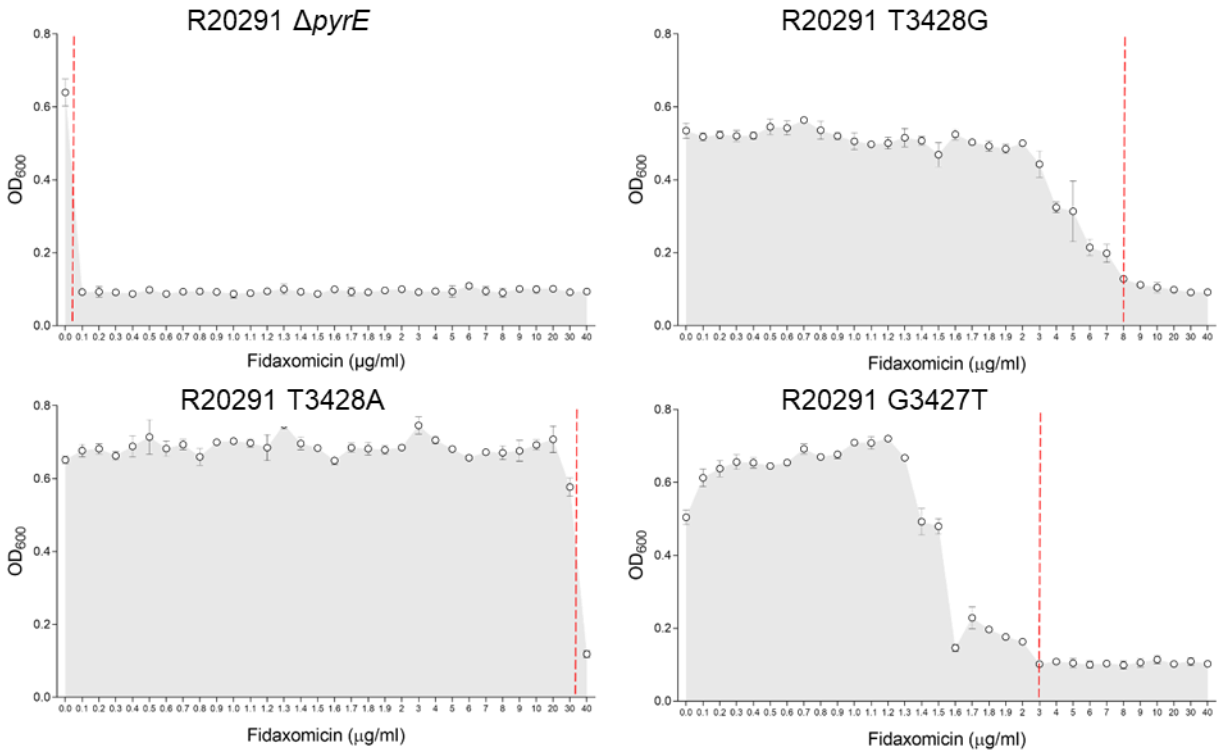


Figure 3.9 Minimum Inhibitory Concentrations of Mutants: Final optical density measurements of *C. difficile* R20291 Δ*pyrE* WT, R20291-T3428A, R20291-T3428G and R20291-G3427T mutants cultures incubated for 24 h at a rising range of concentrations of fidaxomicin. The red dashed line delineates the MIC for each strain. Plotted values are mean ± SD.

Table 3.5 MICs of Parental and Mutant Strains

Strain Designation	Ribotype	MIC (µg/mL)
R20291 Δ*pyrE*	027	0.05
R20291-T3428A	027	> 32.0
R20291-T3428G	027	8.0
R20291-G3427T	027	3.0

3.2.6 Growth Rate

To determine the potential effect of the SNPs in the *rpoB* gene on the fitness of the mutants relative to the parental strain in terms of growth rate, growth curve assays were set up (2.3.1). OD_{600} measurements were taken every hour over a 24 h period.

All three of the mutant strains appeared to be attenuated in terms of growth rate compared to the parental strain, with P values < 0.0001.

The parental strain entered exponential phase at 3 h before reaching stationary phase at 8 h and a peak OD_{600} of 0.89 at 9 h (Figure 3.10).

The T3428A mutant entered exponential phase at 3 h, stationary phase at 6 h, reaching a peak OD_{600} of 0.78 at 13 h. (Figure 3.10).

The T3428G mutant entered exponential phase at 3 h, stationary phase at 6 h, reaching a peak OD_{600} of 0.67 at 13 h (Figure 3.10).

The G3427T mutant entered exponential phase at 3 h, stationary phase at 6 h, reaching a peak OD_{600} of 0.64 at 12 h (Figure 3.10).

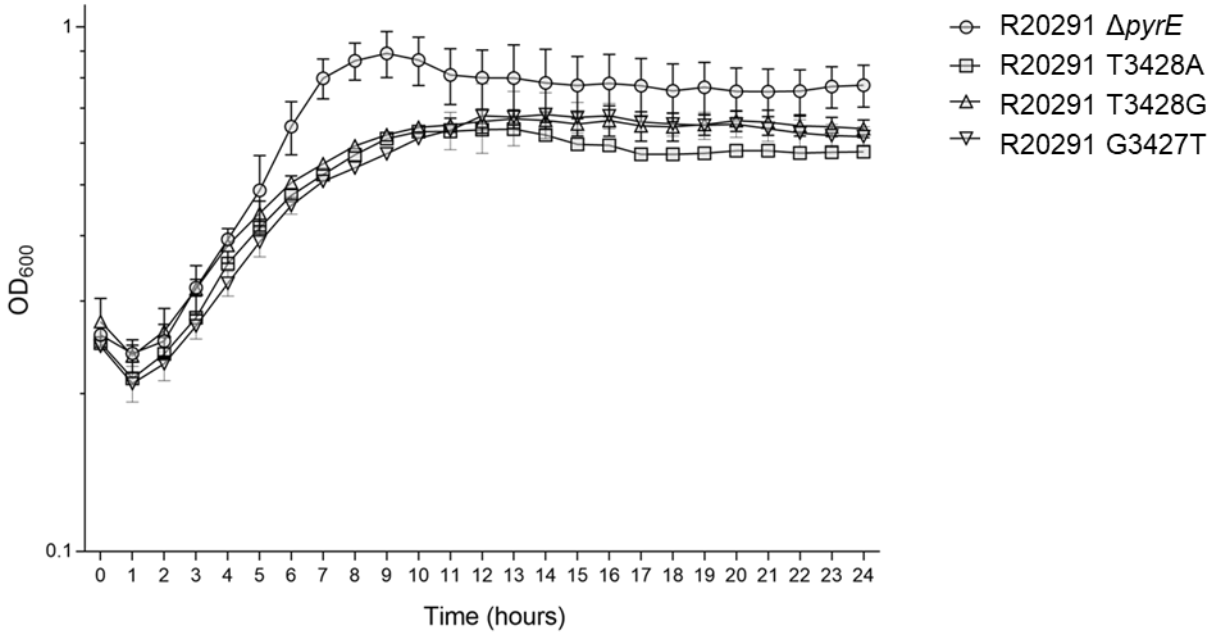


Figure 3.10 Mutant and Parental Strain Growth Curves: The rates of growth of parental strain *C. difficile* R20291 ΔpyrE and mutants R20291-T3428A, R20291-T3428G and R20291-G3427T mutants. Optical density measurements taken over a period of 24 hours. The parental strain reaching a higher OD_{600} than each of the mutant strains, which were significantly attenuated (P values > 0.0001). Plotted values are mean ± SD.

3.2.7 Competetive Growth Advantage

To assess whether the parental or mutant strain had a competitive advantage in terms of growth, each mutant and the parental strain were set up in co-culture (2.3.11). The clinical relevance of resistant mutant strains may be diminished if they are able to be outcompeted by wildtype strains.

Samples were taken, diluted and plated on to non-selective media to ascertain the total viable cell count and media supplemented with FDX at 2 × MIC to ascertain the mutant viable cell count, at 0, 2, 4, 6, 8, 10 and 24 hours.

The mutants appeared to be outcompeted by the parental strain when grown in competition, with statistical significance ($P<0.001$). Comparable generation times were observed up until the 6 h time point, when the viable cell count for the parental strain increased over the mutants in all cases. All strains achieved a peak number of viable cells at the 10 h time points (Figure 3.11).

In the T3428A mutant/parental co-culture, at 10 h the parental strain numbered 2.71×10^8 viable cells compared to the T3428A mutant with a viable cell count of 2.13×10^7. At the 24 h time point, the parental strain numbered 1.86×10^5 compared to the T3428A mutant with a viable cell count of 8.73×10^3 (Figure 3.11).

In the T3428G mutant/parental co-culture, at 10 h the parental strain numbered 3.13×10^8 viable cells compared to the T3428G mutant with a viable cell count of 3.27×10^7. At the 24 h time point, the parental strain numbered 1.34×10^4 compared to the T3428G mutant with a viable cell count of 9.89×10^3 (Figure 3.11).

In the G3427T mutant/parental co-culture, at 10 h the parental strain numbered 2.91×10^8 viable cells compared to the G3427T mutant with a viable cell count of 2.23×10^7. At

the 24 h time point, the parental strain numbered 7.67×10^4 compared to the G3427T mutant with a viable cell count of 3.43×10^3 (Figure 3.11).

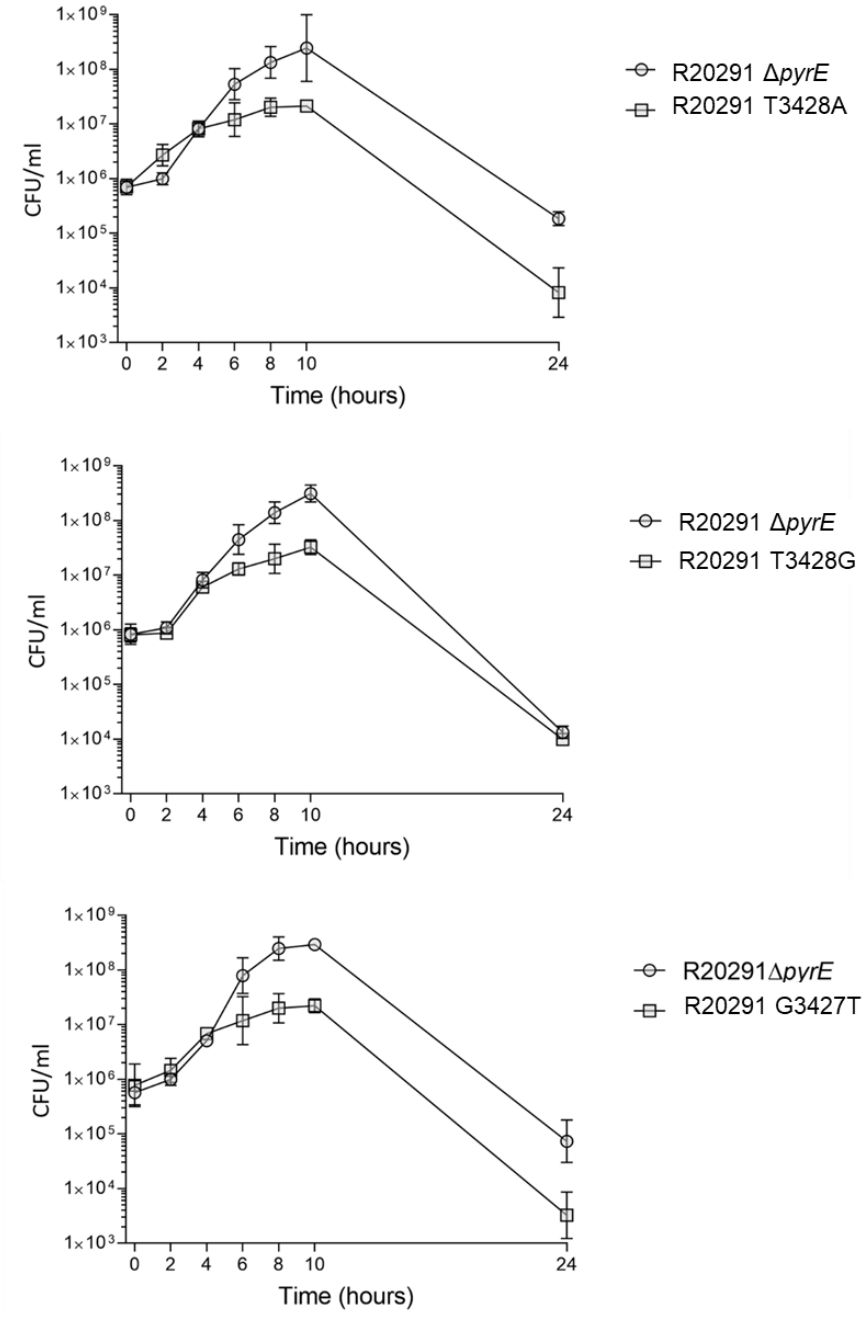


Figure 3.11 Competetive Growth Assays: Parental strain; R20291 Δ*pyrE* and mutants; R20291-T3428A, R20291-T3428G and R20291 G3427T were grown in competetion. Viable cell counts for mutant and parental strains were compared. In each case, the mutant strain reached significantly lower cell counts compared to the parental strain. Plotted values are mean ± SD.

3.2.8 Rate and Capacity of Sporulation

To investigate the potential effect the mutations in *rpoB* on the rate of sporulation and capacity to form spores, sporulation assays were carried out for the parental and mutant strains as described in 2.3.12. Sporulation is a critical process in disease transmission, therefore if the mutant strains are impaired in their ability to produce spores, this would reduce the clinical relevance of such mutant strains.

Heat resistant CFU were not observed for neither the parental nor mutant strains at the 0 and 24 h time points, above the detection limit of the assay.

Heat resistant CFU were observed for the parental and T3428G strains at 48 h, with T3428A delayed until 72 h and G3427T exhibiting the longest delay of 120 h (Figure 3.12).

T3428G showed comparable sporulation rate and capacity to that of the parental strain at the 48, 72 and 96 h time points (P values from 0.1 - 0.4). At 120 h, the number of heat resistant CFU/mL observed for the T3428G mutant compared to the parental strain was significantly reduced (P = 0.0105).

The T3428A mutant began to form spores at the 72 h time point. At the 72 and 96 h time points, heat resistant CFU/mL was significantly reduced compared to that of the parental strain (P = 0.0217, 0.0182). There was no significant difference when comparing the heat resistant CFU of the parental strain and the T3428A mutant at 120 h, (P = 0.1913)

The G3427T mutant appears to be significantly attenuated in terms of sporulation compared to the parental strain and the T3428A and T3428G mutants. The first time point at which heat resistant CFU were observed was at 120 h,and was significantly less than that of the parental strain (P = 0.0008).

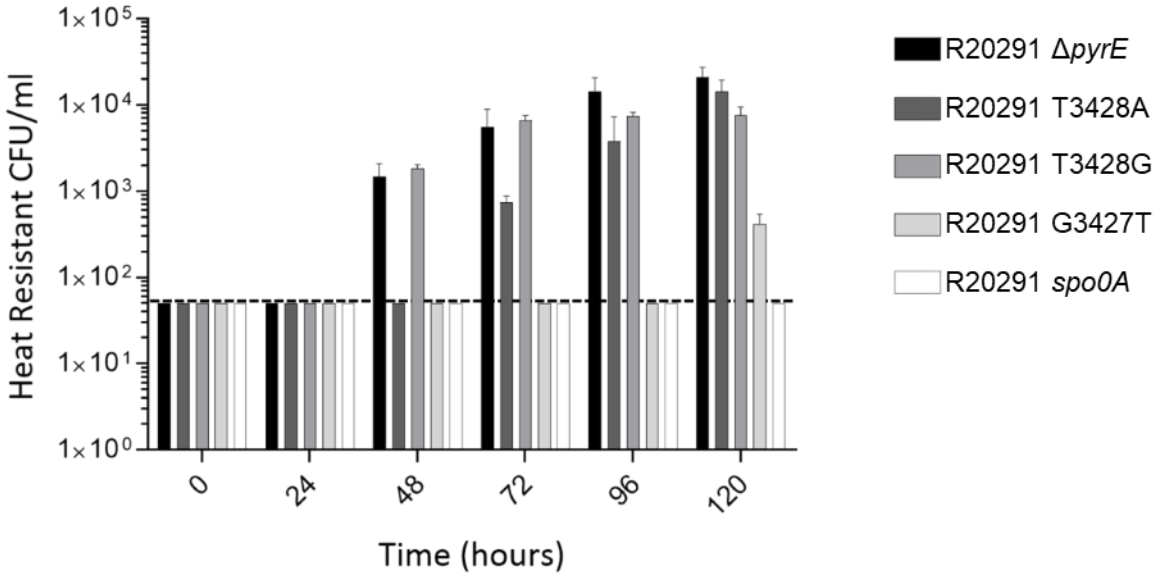


Figure 3.12 Rate and Capacity of Sporulation: The sporulation rate and capacity to form spores for the parental strain; R20291 Δ*pyrE*, and mutants; R20291-TA, R20291-TG, R20291-GT and R20291 spo0A. The TA and GT mutants were delayed in forming spores, with GT the most attenuated. Plotted values are mean ± SD. The dashed line deliniates the detection limit of the assay of 50 CFU/mL

3.2.9 Loss of Heat Resistance Germination Assay

A loss of heat resistance assay was used to measure the rate of germination of the parental and mutant strains. As with sporulation, germination is a key step in the infection cycle of *C. difficile.* Spores with impaired or delayed germination will result in a reduced population of vegetative cells and subsequent reductions in toxin production and spore shedding.

This particular assay was used over the previously utilised germination initiation assay due to the relative difficulty in obtaining sufficiently pure spore preparations for the R20291 strain.

Spores for the parental and mutant strains were incubated with the germinant TC. Samples were taken over a 240 minute period, heat treated at 65 ^{o}C for 30 minutes and plated on to BHIS agar supplemented with TC (2.3.13). Colonies observed are an indication of un-germinated spores in the primary culture which are able to survive heat treatment. Therefore, the percentage decrease in CFU/mL is an indication of the rate of germination and outgrowth of spores.

All strains showed a reduction in heat resistant CFU over the course of the assay, indicating germination and return to vegetative cell growth (Figure 3.13).

The parental strain showed a 91.25 % percentage decrease in heat resistant CFU over 240 minutes (Figure 3.13).

The T3428A mutant was slightly attenuated when compared to the parental strain, with a 80 % decrease in heat resistant CFU over 240 minutes, although not statistically significant (P = 0.9999) (Figure 3.13).

The T3427G mutant also appeared to be attenuated when compared to the parental strain, with the final time point at 240 minutes indicating a 41.46 % loss in heat

resistance, although the previous time point at 180 minutes showed an outlying reduction of 58.54 %. These data are also not statistically significance when compared to the parental strain (P = 0.1674) (Figure 3.13).

The G3427T mutant was the most attenuated in terms of the loss of heat resistance and hence rate of germination. After 240 minutes a 41.58 % drop in heat resistance was observed with a germination profile significantly attenuated when compared to the parental strain (P = 0.0382) (Figure 3.13).

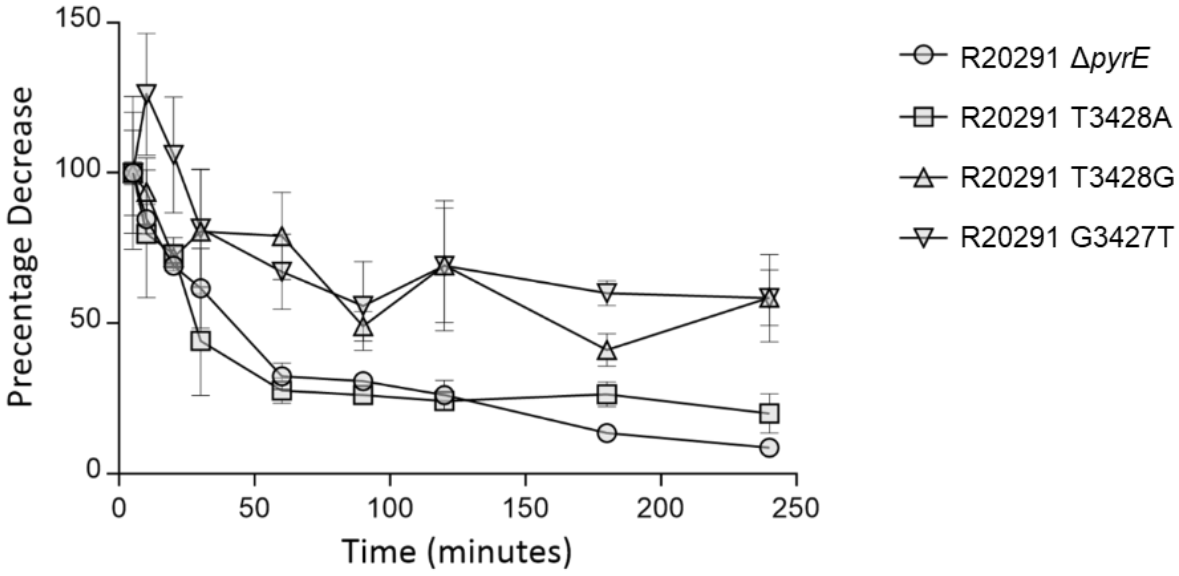


Figure 3.13 Loss of Heat Resistance of Parental and Mutant Strains: The loss of heat resistance of parental and mutant cultures over 240 minutes. The percentage decrease in heat resistant CFU at each time point taken as an indication of the rate of germination. The T3428G and G3427T mutant strains appeared to be delayed in germination. The parental and T3428G strains germinated at a comparable rate. Plotted values are mean ± SD

3.2.10 Toxin A and B Production

Toxins A and B are the primary virulence factors resulting in the characteristic symptoms of the disease caused by *C. difficile*, it is therefore important to establish the toxin producing ability of mutant strains resistant to FDX to assess the clinical relevance of these mutant strains.

To measure toxin A and B production, supernatant samples were obtained at 24, 48 and 72 hours for each mutant grown in TY broth (2.3.14), toxin was measured and quantified using toxin A/B ELISA and standard curve analysis. Samples were adjusted for OD_{600}.

At 24 h, all three mutants showed a lag in toxin A/B production compared to the parental strain.

At 48 h, the T3428A and G3427T mutants were significantly attenuated in terms of toxin A/B production when compared to the parental strain ($P = 0.0001$). The T3428G mutant, appeared to produce a comparable titre of toxin A/B to the parental strain ($P = 0.1065$).

At 72 h, the parental strain had showed 45-fold increase in toxin production over the T3428A mutant ($P = 0.0004$), 26-fold over the G3427T mutant ($P = 0.0004$) and 4-fold over the T3428G mutant ($P = 0.0019$), with statistical significance (Figure 3.14 A).

3.2.11 Vero Cell Cytotoxicity

For cytotoxicity assays, supernatant samples were obtained at 8, 24 48 and 72 h for each mutant, grown in TY broth. Prepared supernatants were added to Vero (African green monkey kidney) cells and incubated for 24 h before reading (2.3.15) (Figure 3.14 B).

At 8 h, all the mutant strains were significantly attenuated in terms of cytotoxicity compared to the parental strain ($P < 0.0001$). G3427T appeared to be the most attenuated with no observed cytotoxicity to Vero cells.

At 24 h, the T3428G mutants, exhibited cytotoxicity comparable to that of the parental strain ($P = 0.0897$). The cytotoxicity remained significantly attenuated for the T3428A and G3427T mutants when compared to the parental strain ($P = 0.0001$).

At 48 h, all mutant strains were significantly attenuated in terms of cytotoxicity compared to the parental strain, ($P = 0.0001$). Mutant TG was significantly more cytotoxic to Vero cells when compared to both T3428A and G3427T mutants ($P = 0.0001$).

At 72 h all mutant strains were significantly attenuated in terms of cytotoxicity compared to the parental strain, (P values from 0.0002 to 0.0006). Mutant G3427T was significantly attenuated when compared to T3428G ($P = 0.0341$), but no significant difference in cytotoxicity when compared to T3428A ($P = 0.4130$). Both of the mutants T3428A and T3428G exhibited a statistically comparable level of cytotoxicity at 72 h ($P = 0.1701$).

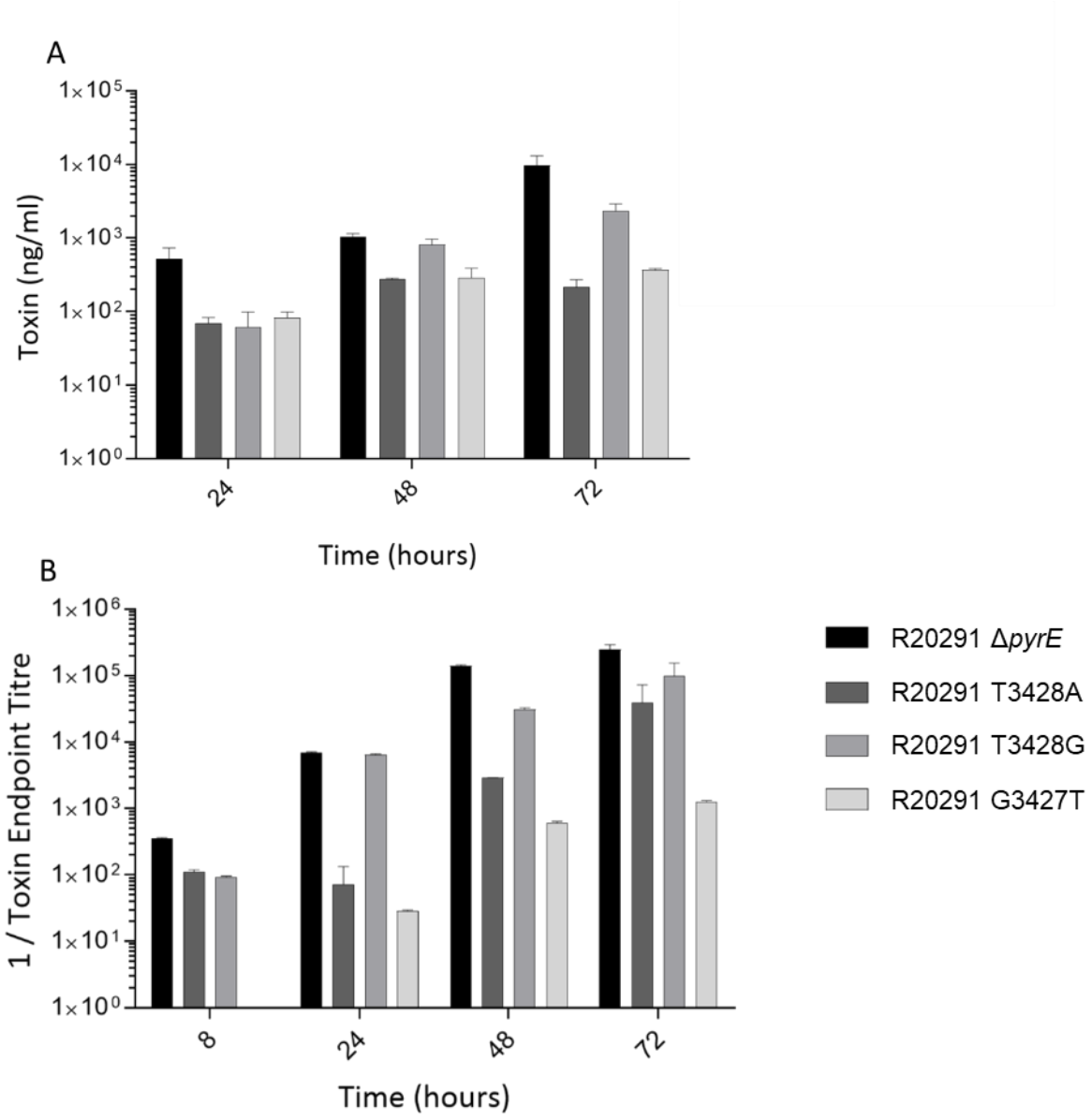


Figure 3.14 Parental and Mutant Toxin Production. A : Toxin A/B production measured by ELISA **B**: Vero cell cytoxocitiy assay. The G3427T mutant was the most attenuated in terms of toxin production and cytotoxicity, followed by the T3428G mutant. The T3428A mutant strain was the least attenuated when compared to the paretal strain. Plotted values are mean ± SD.

3.2.12 In vivo Virulence and Colonisation in the Golden Syrian Hamster Model

The *in vivo* virulence of the mutant strains was assessed to establish a link between the observed fitness costs and disease in the hamster model of infection. The hypothesis being that due to the fitness costs, the mutant strains will be less virulent.

For the *in vivo* work, the three mutants were remade in the R20291::*ermB* strain using the same methodology as before. The genome sequences for all three R20291::*ermB* mutants were obtained and compared to the parental as with the previous set of mutants.

All three of the mutant strains exhibited a statistically significant decrease in virulence, when compared to the parental strain in the hamster model of infection, (P values from 0.0001 – 0.0087) (Figure 3.15). T3428A was the least virulent with a mean time of 53.5 h until clinical endpoint. Followed by G3427T with a mean of 51.25 h and T3428G with a mean of 40 h. The parental strain was the most virulent with a mean time of 31.2 h until clinical endpoint.

There were no significant differences in the mutant capacity to colonise the hamster caecum when compared to the parental strain (Figure 3.16).

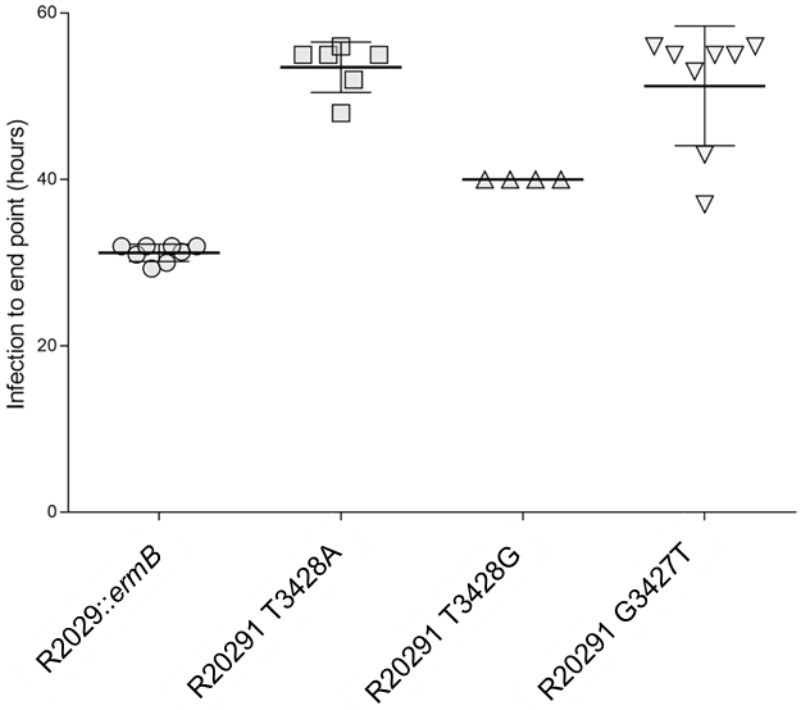


Figure 3.15 Virulence in the Golden Syrian Hamster Model: Time from infection with spores to clinical endpoint for parental R20291::ermB and mutant strains T3428A, T3428G and G3427T in the hamster model of infection. All mutant strains were significantly delayed in initiating the clinical end point. Plotted values are mean ± SD.

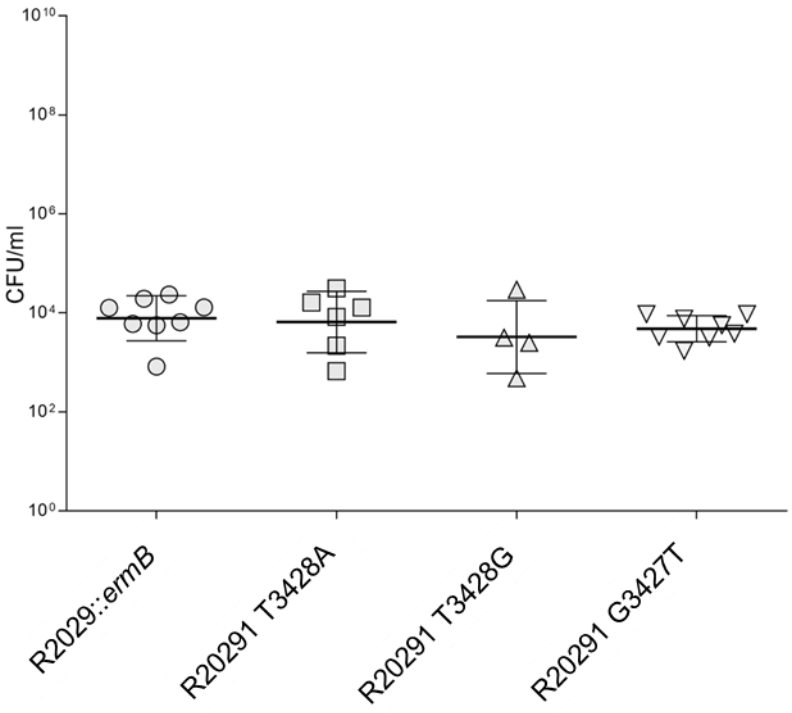


Figure 3.16 Golden Syrian Hamster Caecal Colonisation: *C. difficile* CFU/ml recovered from caecal samples for all strains after sacrifice. Colonisation levels of hamster caeca were statistically comparable. Plotted values are mean ± SD.

3.3 Discussion

3.3.1 The Effect of Fidaxomicin on Spore Germination and Outgrowth

The initiation of germination assays confirms that FDX does not inhibit the initial stages of germinating *C. difficile* spores of the strains of study, *in vitro* (Figures 3.4.1 and 3.4.2). As previously stated, a decrease in the OD_{600} of spore cultures is due to spores becoming less optically dense and losing birefringence, as a direct result of progressing through the stages of germination.

The outgrowth assays show that FDX inhibits the outgrowth of germinated spores of the *C. difficile* ribotypes of study, *in vitro*. The OD_{600} of cultures supplemented with FDX at MIC did not increase, and growth was at least delayed at ¼ and ½ MIC for all strains studied (Figure 3.5). Furthermore, spores incubated with FDX and the germinant TC exhibited phase dark, ellipsoid morphologies when observed under phase contrast microscopy, as opposed to spores incubated only with the germinant, which grew out in to phase dark vegetative cells (Figure 3.6.1).

These data suggest that spores progress through both stage I and stage II of germination; irreversible processes involving the release of Ca^{2+}-DPA, cortex hydrolysis and core expansion (Setlow, 2003). After this stage of germination, the spore enters the period of time known as outgrowth, when metabolism is resumed, and macromolecule synthesis occurs, beginning the outgrowth of the now fully hydrated spore. High energy compounds such as ATP are lacking in the dormant spore, although lower energy compounds such as ADP, AMP, NAD and CoA are present. The catabolism of these compounds meets the initial energy requirements of the spore during outgrowth. RNA synthesis, utilising stored nucleotides in the spore, begins soon after the initiation of germination. The genes involved in the early stages of outgrowth are not yet fully understood (Paidhungat & Setlow, 2002).

The low water content of the dormant spore and the large deposits of dipicolonic acid chelated with calcium ions are vital to spore integrity and resistance (Setlow, 2003). Consequently, it can be hypothesised that due to the increased permeability of these pre-vegetative spores, FDX is able to infiltrate the germinating spore, where it halts transcription and subsequent protein synthesis by binding to RNAP preventing the expression of genes necessary for spore outgrowth.

This hypothesis is further reinforced by work conducted by Allen and colleagues, who performed similar germination and outgrowth experiments with two strains, UK-1 and CD196, both belonging to the PCR-ribotype 027 group. It is reported in this work that the antibiotic vancomycin (VAN) also inhibits the outgrowth of *C. difficile* spores, although not before an increase in OD_{600} of the culture (Allen *et al.*, 2013). This is not observed when spores are incubated with FDX and is undoubtedly due to the different modes of action of these two drugs. VAN exerts its bactericidal activity by binding to the D-alanine dipeptide during cell wall synthesis, preventing the cross-linking and polymerisation of peptidoglycan in the growing cell wall of Gram-positive bacteria. Since the spore germ cell wall provides peptidoglycan for the outgrowing cell, some spore outgrowth is observed. This suggests that FDX acts upon the germinating spore at an earlier stage than VAN. Although it should be stressed that *C. difficile* strains of different PCR-ribotypes have been shown to display a varied and diverse germination profile, therefore a wide repertoire of strains is crucial when studying *C. difficile* spore germination (Heeg & Burns, 2012).

3.3.2 The Effect of rpoB Mutations on Susceptbility to Fidaxomicin

The three *rpoB* SNPs chosen for this study were selected as they had been shown to confer the most elevated MICs of the previously identified spontaneous mutants isolated in the work conducted by Seddon and colleagues in 2011 (Table 3.2).

The strain selected for this study, designated R20291, was isolated from the Stoke Mandeville hospital outbreaks, between 2003 and 2006. These cases were the first known incidences of the so called "hypervirulent" PCR-ribotype 027 strains, also referred to as restriction endonuclease analysis BI (REA BI) and North American pulsed-field type 1 (NAP-1), in the UK (Health Care Commission, 2007).

PCR-ribotype 027 strains are often implicated in outbreaks with increased disease severity and as previously mentioned, 027 strains are amongst the most prevalent in Europe, accounting for 41.3 % of isolated ribotypes in the UK for the years 2007 to 2008, (Brazier *et al.,* 2008) and 19 % of clinical isolates in Europe in 2013 (Davies *et al.,* 2016). The strain R20291 therefore represents a relevant and important strain of study.

Through the re-creation of these previously identified SNPs in the *C. difficile* chromosome, a reduction in the susceptibility to FDX was observed. This observation coupled with the whole genome sequencing of the mutant strains provides a definitive confirmation of the association between these genotypes and FDX susceptibility. The elevated MICs of the mutants created in this work consistently follow the trend of the previously identified random mutations in *rpoB* as reported by Seddon and colleagues in 2011, with the T3428A substitution conferring an MIC of >32 μg/mL, followed by T3428G with 8 μg/mL and G3427T with 3 μg/mL (Table 3.4).

Consequently, it is hypothesised that RNAP target modification is the cause of the observed reduced susceptibility to FDX. This is as a result of the substitution of the amino acid valine at residue 1143 in the β subunit. This mutational hotspot is spatially located

in the β switch-3 region, which along with the β' switch-2, connects the clamp to the body of the RNAP core enzyme (Cramer *et al*., 2001; Murakami *et al*., 2002). Mutations in the *rpoC* gene which code for the β' subunit have also been shown to confer reduced susceptibility to FDX (Seddon *et al*., 2011).

During the initiation of transcription, the clamp must be in an open conformation to permit entry of DNA in to the active site cleft, and closed during elongation to retain the DNA in the catalytic core of the enzyme. This action is dependent upon the conformational state of the switch region at the base of the clamp, which is affected by the interaction of switch residues with the nucleic acids of promoter DNA (Vassylyev *et al*., 2002). These interactions mediate the melting of the promoter strands and formation of the transcription bubble.

One proposed model of inhibition is that FDX sterically blocks these interactions by binding to the switch elements in the core enzyme, preventing conformational changes in the switches and the initiation of transcription. This model has been discounted through foot printing analysis, revealing protection of RNAP bound DNA from cleavage in the presence of FDX (Artsimovich *et al*., 2012).

By probing RNAP-DNA complexes with potassium permanganate ($KMnO_4$), which reacts with single stranded thymine nucleotides, it has been observed that promoter DNA was inaccessible to $KMnO_4$ in the presence of FDX, suggesting that the melting of the promoter and initiation of the transcription bubble is inhibited by FDX (Tupin *et al*., 2010).

The amino acid substitution conferring reduced susceptibility to FDX may result in the modification of the conformational structure of the core enzyme, altering the response of the enzyme to the binding of FDX in an allosteric manner, or by directly reducing the affinity of FDX for the target site allowing transcription to proceed. As well as definitively confirming the link between the substitution in the β subunit and reduced FDX

susceptibility in *C. difficile*, these data also validates previous work hypothesising the binding site of FDX to encompass the C-terminus of the β switch-3 region in RNAP.

3.3.3 The Impact of rpoB Substitutions on the Fitness of C. difficile

Although the created mutants exhibited elevated FDX MICs, phenotypic characterisation assays indicate that the strains carrying these mutations also incur associated fitness costs. When compared to the parental, each of the mutant strains were significantly attenuated in terms of growth rate and competitive growth advantage whilst the sporulation and germination phenotypes appear to be marginally affected. Sporulation and germination are important characteristics of *C. difficile* in the clinical setting, with spores being the main vehicle of transmission for *C. difficile* infection (Burns & Minton, 2011). Toxin production and cytotoxicity of the mutants were reduced when compared to the parental strain. These characterisation data suggest that *C. difficile* strains carrying these mutations in the *rpoB* gene could be less virulent, and therefore, less able to cause disease.

To study the effects of these mutations on the virulence of *C. difficile in vivo*, the Golden Syrian hamster model of infection was employed. The T3428A, T3428G and G3427T mutants were each re-created in an erythromycin resistant derivative of the R20291 strain, R20291::*ermB.* This was essential due to the sensitivity of the parental R20291 strain to the antibiotic clindamycin, which is used to induce dysbiosis in the hamster model prior to infection. It has been shown that clindamycin can remain for up to 11 days post administration in the hamster gut at inhibitory concentrations (Larson & Borriello, 1990). The strain R20291::*ermB* produces a more reliable infection in the hamster model of infection than the erythromycin sensitive parental strain (Kelly *et al.*, 2016).

These *in vivo* data suggest that mutants with reduced susceptibility to FDX are less infectiously fit, with each mutant taking longer to initiate the clinical end-point when compared the parental strain. This is despite each of the mutant strains colonising the hamster cecum to levels comparable to that of the parental. These data somewhat mirror the results of the *in vitro* toxin production and cytotoxicity characterisation assays, and it is conceivable that this reduced toxicity is a significant contributing factor for the reduced virulence *in vivo* of the mutant strains. This would require further evaluation by quantifying hamster faecal toxin levels.

It can be hypothesised that the observed attenuations are due to the altering of the conformational state of the RNAP core enzyme, impairing transcriptional competency. Utilising data from mutational studies carried out over the last two decades and the more recent resolution of crystal structures of bacterial RNAP, it has become apparent that amino acid substitutions in close proximity to the active core of the enzyme results in reduced polymerisation rates *in vitro* (Trinh *et al*., 2006). It is probable that such a reduction is accountable for the observed attenuation and fitness burden of the RNAP mutants in this study. The investigated substitution in the β subunit could therefore offer some insight in to the structure-function relationships in the components of *C. difficile* RNAP. The Val1143 substitution lies in the switch-3 region which has been shown to interact with the DNA-RNA hybrid (Gnatt *et al*., 2001). It is therefore plausible that the substitution of residues in this region affects protein-nucleic acid interactions, attenuating strand separation and RNA/DNA translocation during elongation (Figure 3.18)

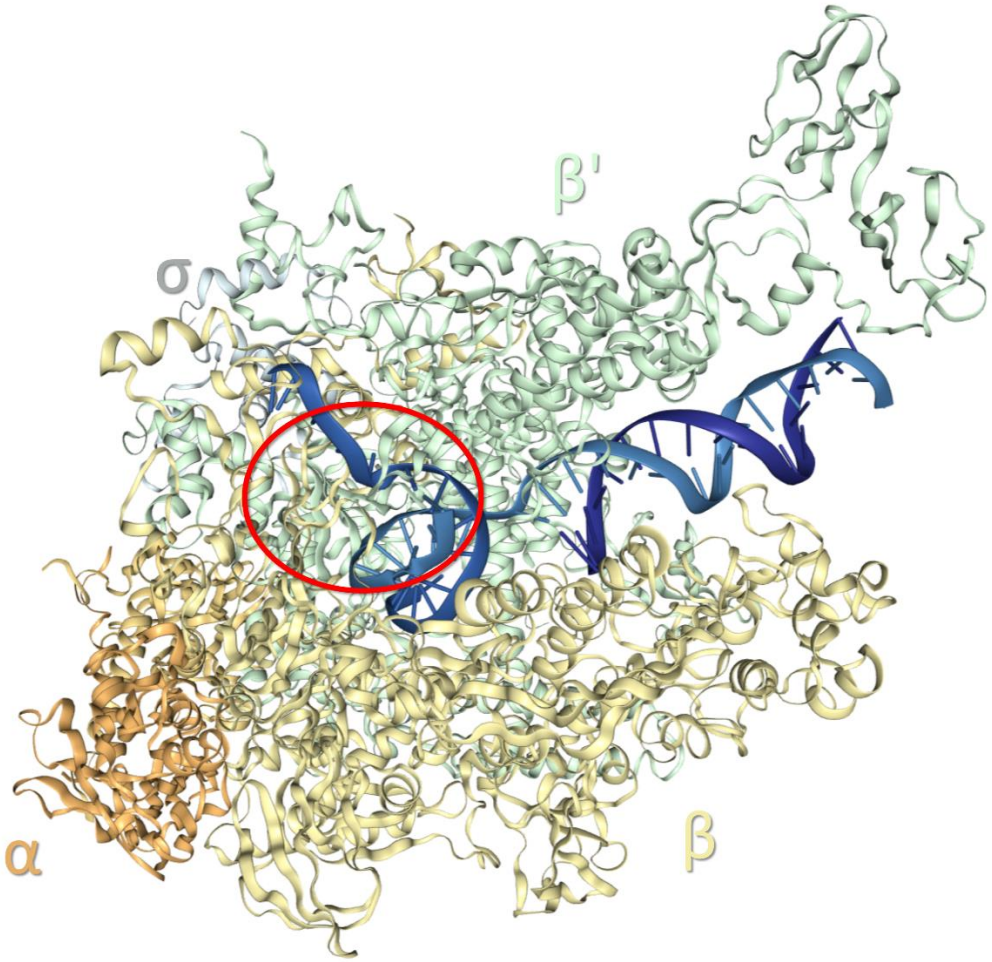


Figure 3.18 Cartoon Crystal Structure of Thermus thermophilus RNAP holoenzyme: α subunit in orange, β subunit in yellow β' subunit in green and σ subunit in grey. The switch-3 region highlighted by the red circle, showing protein-nucleic acid interactions and RNA-DNA hybrid strand separation. PDB reference: 4WQS.

The disparity in the levels of phenotypic attenuation observed across the different characterisation assays could be due to the structural differences of the RNAP core enzyme-sigma complexes during different phases of growth. Substitutions in the β subunit of *E. coli* RNAP, although relatively benign for the host bacterium, confer defects in the MotA and AsiA sigma factor mediated transcription of T4 phage (James *et al.*, 2010). These phage factors bind to and structurally remodel the host primary sigma factor, σ^{70}, which then associates with the core RNAP, directing the enzyme to T4 promoters. This provides basis for the hypothesis that the attenuation and reduction of polymerisation rates due to a single mutation could be sigma factor dependent. The

binding of different sigma factors may stabilise the conformational structure of the impaired RNAP, or the mutation may affect the efficiency of binding by certain sigma factors. In this study, toxin production appears to be more significantly attenuated than sporulation, this could be explained by the fact that both of these characteristics are regulated by complex networks of different alternative sigma factors, TcdR and SigH respectively (Girinathan *et al.,* 2017; Saujet *et al*., 2011).

Clinically relevant RNAP mutations have been observed in other organisms, such as rifamycin resistant *Mycobacterium tuberculosis*. However, the majority of these clinically isolated strains carry multiple mutations in the *rpoA*, *rpoB* and *rpoC* genes, suggesting that secondary mutations can provide fitness compensation (Brandis *et al*., 2012).

Conversely, rifamycin resistance in *C. difficile*, is not generally associated with a fitness burden (Dang *et al*., 2016; Stevenson *et al*., 2016). The most commonly isolated mutation in the clinic, an arginine to lysine substitution at position 505 in the β subunit, was found to confer no fitness burden *in vitro*. No cross resistance between rifamycin antibiotics and FDX has been observed (Babakhani *et al*., 2014). This further supports evidence for the distinct mechanisms of action between rifamycin antibiotics and FDX and the relative importance of the structure-function of these two regions of the β subunit.

FDX resistance in *C. difficile* is rarely observed clinically, with only one reported case in clinical trials, post therapy (Goldstein *et al*., 2011). It should also be noted that this lack of prevalence of FDX resistance is also highly likely due to the high therapeutic concentrations achieved by FDX in the gut due to favourable pharmacokinetics, enabling persistence of the drug at the site of infection. FDX reaches faecal concentrations of >1000 μg/g, far in excess of the observed elevated *in vitro* MICs in this study (Sears *et al*., 2013).

Due to the fitness burden imposed by the amino acid substitutions in the subunits of RNAP, it is unlikely that these mutations will result in the rapid spread of FDX resistant *C. difficile*. Without selective pressure, mutants would likely be outcompeted by their more competitively fit, FDX susceptible counterparts and therefore do not represent a significant clinical threat, in terms of disease severity and FDX treatment.

Chapter 4 The Bile Acid Metabolism of *Clostridium scindens*

4.1 Introduction

4.1.1 Primary Bile Acid Synthesis and The Enterohepatic Circulation

A major part of mammalian cholesterol catabolism is the formation of primary bile acids. Bile acids are physiological important molecules, derived from dietary cholesterol and synthesised in the pericentral hepatocytes of the liver. Primary bile acids are formed through the sequential, multistep hydroxylation of cholesterol via the classical or acidic pathways. The majority of primary bile acids are the result of the classical pathway, where cholesterol 7α-hydroxylase catalyses the conversion of cholesterol to 7-hydroxycholesterol, preceding epimerisation by steroid oxidoreductases. Before the ensuing steps of the classical pathway, these sterol intermediates are either acted upon by sterol 12α-hydroxylase resulting in the subsequent formation of cholic acid (CA) or escape the action of this enzyme forming chenodeoxycholic acid (CDCA), the two most abundant bile acids in the human bile pool (Figure 4.1)

Primary bile acids are further metabolised in the hepatocytes by conjugation to the amino acids glycine or taurine, via the formation of an amide bond between amino group and the bile acid terminal carboxyl group. These conjugations result in the conjugated bile acids, glycocholic acid (GC) and taurocholic acid (TC) from CA and glycochenodeoxycholic acid (GCDCA) and taurochenodeoxycholic acid (TCDCA) from CDCA (Figure 4.1). This process functions to decrease the cytotoxicity and increase amphipathicity of primary bile acids, easing secretion across the canalicular membrane of the hepatocytes via the ATP dependent bile salt export pump. Once transported from the hepatocytes, conjugated bile acids join the fluid known as the bile (Figure 4.1). At the physiological pHs of the bowel (pH 6 – 6.7), conjugated bile acids are entirely ionised and referred to as bile salts (Hoffman, 1999).

Bile salts are a significant constituent of the bile, a complex fluid also comprised of water, electrolytes and organic molecules such as phospholipids and cholesterol. Bile is stored and concentrated in the gall bladder, where it remains during the interdigestive period. After the ingestion of food, the hormone cholecystokinin is secreted from endocrine cells, induced by protein and fatty acid secretagogues. The physiological effects of cholecystokinin result in the stimulation of pancreatic enzyme secretion, gastric emptying and gallbladder contraction, effectively coordinating the ingestion, digestion and absorption of nutrients (Liddle, 1997).

The bile leaves the gallbladder and flows in to the small intestine via the biliary tract, where it serves two main functions; elimination of waste products from the body and aiding in the digestion and reabsorption of fats. The majority of the bile is actively reabsorbed in to the liver from the ileum via the sinusoidal membrane, completing the enterohepatic cycle. However, around 600 to 800 milligrams escape enterohepatic circulation in to the large bowl, per day (Ridlon *et al.,* 2006). Here, members of the gut microbiota metabolize conjugated primary bile acids through pathways of deconjugation, oxidation and reduction, generating secondary bile acids.

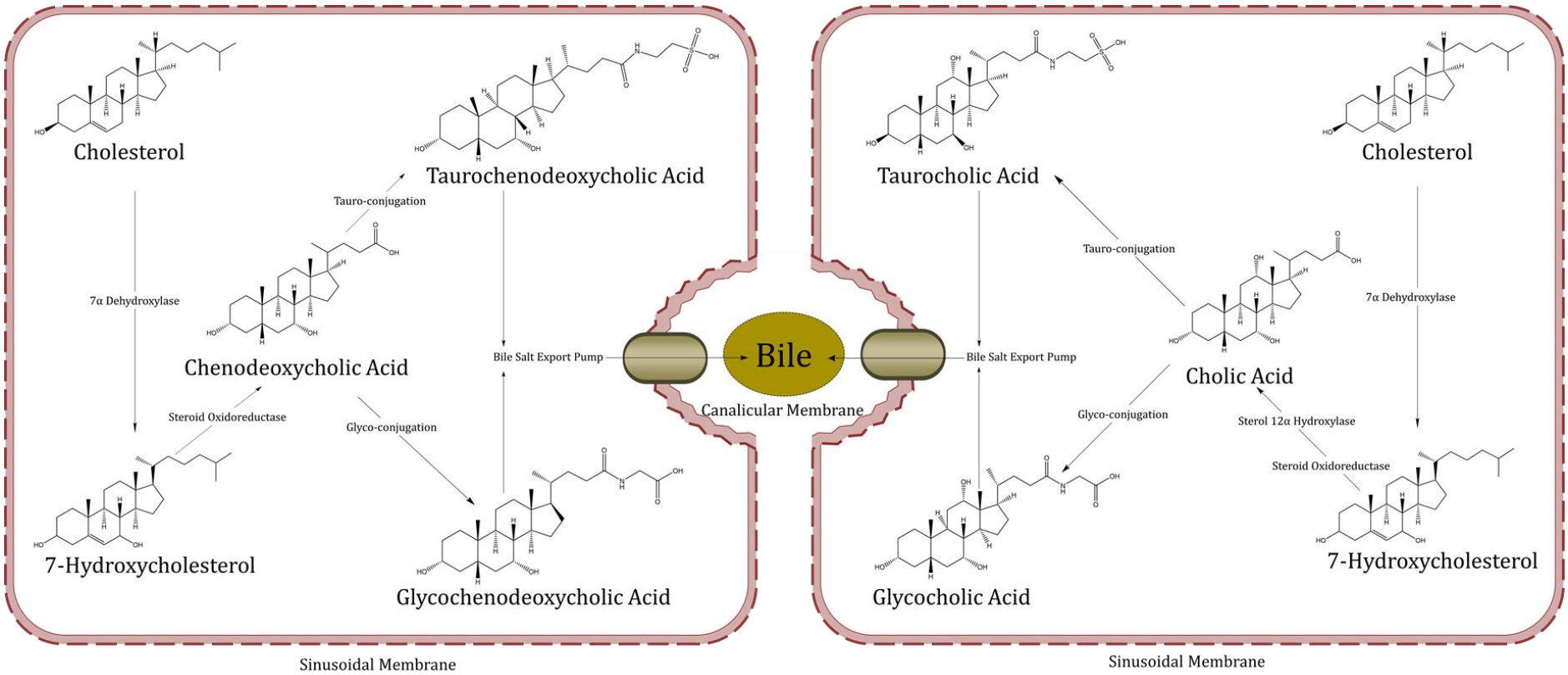


Figure 4.1 Primary Bile Acid Synthesis and Conjugation in the Hepatocytes of the Liver: The formation of cholic acid and chenodeoxycholic acid from cholesterol via the classical pathway of bile acid synthesis and amino conjugation forming of the primary bile acids cholic acid and chenodeoxycholic acid before export across the canalicular membrane by the bile sale export pump.

4.1.2 Bile Acid Deconjugation

Deconjugation involves the hydrolytic cleavage of the amide bond between the primary bile acids and their amino acid counterparts, catalysed by a class of microbial enzymes collectively termed bile salt hydrolases (BSH). Genes encoding BSHs have been observed in a wide variety of Gram-positive and Gram-negative bacteria residing in both the small and large bowels from genera including, but not limited to, *Bacteroides*, *Bifidobacterium*, *Clostridium*, *Enterococcus*, *Lactobacillus* and *Listeria* (Jones *et al*., 2008). It has been hypothesised that the hydrolysis of bile salts is a mechanism which aids colonisation, potentially as a means of detoxification or provision of cellular nutrients (De Smet *et al.*, 1995).

4.1.3 Bile Acid 7-Dehydroxylation

The major transformation of bile acids in the colon, is the 7α-dehydroxylation of CA and CDCA to the secondary bile acids deoxycholic acid (DCA) and lithocholic acid (LCA) respectively. The process of 7β-dehydroxylation occurs to a much lesser extent and involves the transformation of the secondary bile acid, ursodeoxycholic acid (UDCA) to LCA. UDCA is formed by the epimerisation of the 7α-hydroxyl group of CDCA by microbial hydroxysteroid dehydrogenase (HSDH) enzymes.

Conversely to the widely observed BSH enzymes, the genes required or the 7-dehydroxylation of primary bile acids are limited to a select few anaerobic species of the colonic microbiota. The predominant 7α-dehydroxylating organism in the human gut is *C. scindens*.

The first step in the biotransformation of bile acids is the proton motive force dependent transport of free, unconjugated primary bile acids in to the cell. This occurs via the H+

dependant bile acid transporter, BaiG and is reliant upon the availability of unconjugated bile acids, the result of microbial deconjugation in the large bowel. Once in the cytosol, co-enzyme A, (CoA) is conjugated to the free terminal carboxyl group on the primary bile acids. This step is performed by BaiB, a 56kDa bile acid CoA ligase. The ATP dependent thioesterification of bile acid to CoA is a two-step process. Firstly, the ATP dependent adenylation of the bile acid followed by CoA exchange, releasing AMP. This ligation serves to commit bile acids to 7-dehydroxylation and to trap these CoA conjugates inside the cell (Mallonee *et al*., 1992).

The following step in the pathway involves BaiA, which catalyses the oxidation of the C3 hydroxyl group of the CoA conjugated bile acids. There are three separate *baiA* genes present in *C. scindens*, the monocistronic *baiA1* and *baiA3*, and *baiA2* found on the polycistronic bai operon. All three of these genes encode for 3α-HSDH enzymes, which exclusively recognise the CoA conjugated forms of primary bile acids and utilise NAD+ and NADP+ as electron acceptors, (Bhowmik *et al.*, 2014).

The ensuing step involves the gene products of *baiCD* and *baiH*, both of which have been shown to encode steroid oxidoreductases of 70 and 72 kDa respectively (Kang *et al.*, 2008). These enzymes act stereospecifically upon 7α-hydroxy bile acids. BaiH has been shown to act preferentially upon CoA conjugates of 3-oxo-epicholic acid and 3-oxo-UDCA, whereas BaiCD acts upon CoA conjugates of 3-oxo-CDCA and 3-oxo-CA. The purpose of this further oxidisation step is to increase the chemical lability of the primary bile acids for the next step in the pathway, 7α/β-dehydration by the gene products of *baiE* and *baiI* respectively.

The *baiF* gene encodes for a 47.5kDa protein for which CoA hydrolase activity has been observed. It has subsequently been shown that BaiF also has CoA transferase activity (Ridlon & Hylemon 2012). BaiF has been determined to belong to the type III family of CoA transferases by comparison of amino acid sequences. This CoA transferase activity

would be metabolically favourable to the cell as the initial rounds of 7-dehydroxylation would require the hydrolysis of ATP. Transfer of CoA from the BaiH and BaiCD mediated oxidisation products back to the CoA conjugation step of the pathway results in a substantial conservation of energy for the cell (Ridlon & Hylemon 2012).

The following dehydration steps are catalysed by the gene products of *baiE* and *baiI*. BaiE is a 19.5kDa 7α-dehydratase and as a result of homology between amino acid sequences, it is hypothesised that *baiI* encodes for a 7β-dehydratase enzyme (Ridlon *et al.*, 2006). These enzymes catalyse the removal of oxygen and hydrogen as water, preparing the intermediates for the following reductive steps in the pathway.

The genes involved in the subsequent reduction of the products of 7α/β-dehydration remain unknown, but it is hypothesised that these genes will encode oxidoreductases. These enzymes catalyse the reduction of the dehydrated compounds to secondary bile acids, which are then exported from the cell by an as yet unidentified transporter, (Table 4.1, Figure 4.2).

Table 4.1 The Steps of 7α-Dehydroxylation of Cholic Acid in *C. scindens*

Gene	Activity	Intermediate	Product
Oxidative Arm			
baiG	Transport	CA	CA
baiB	CoA Ligation	CA	CA-CoA
baiA	3α oxidation	CA-CoA	3-oxo-CA-CoA
baiCD	Δ^4 oxidation	3-oxo-CA-CoA	3-oxo-Δ^4-CA-CoA
baiF	CoA Hydrolase/Transferase	3-oxo-Δ^4-CA-CoA	3-oxo-Δ^4-CA
baiE	7α-dehydration	3-oxo-Δ^4-CA	3-dehydro-$\Delta^{4,6}$-DCA
Reductive Arm			
?	Δ^6 reduction	3-dehydro-$\Delta^{4,6}$-DCA	3-dehydro-Δ^4-DCA
?	Δ^4 reduction	3-dehydro-Δ^4-DCA	3-dehydro-DCA
?	3α reduction	3-dehydro-DCA	DCA
?	Transport	DCA	DCA

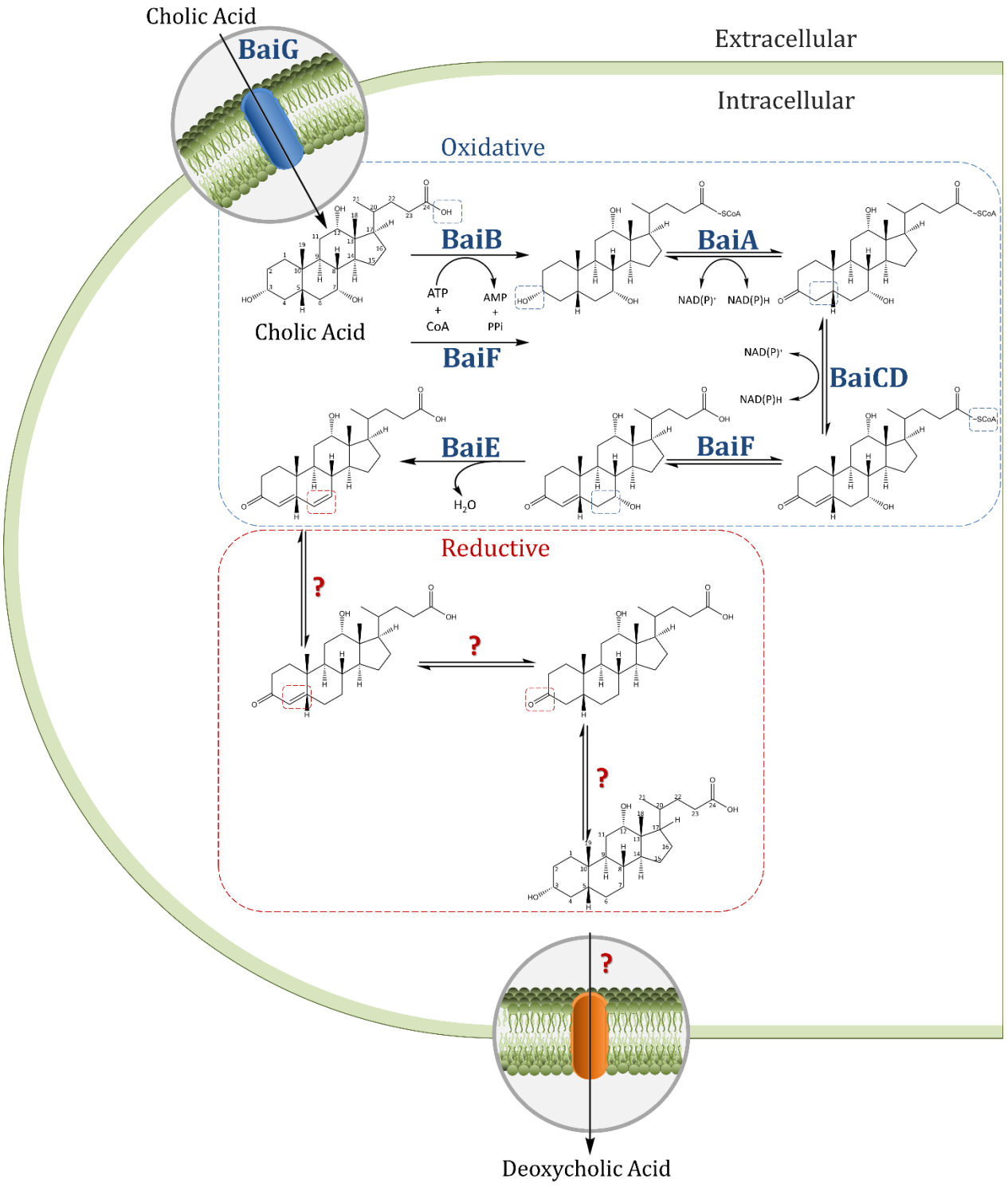


Figure 4.2 The Bile Acid Transformation of *C. scindens* strain ATCC 35704. The oxidative and unknown reductive arms of the 7α-dehydroxylation pathway of *C. scindens* strain ATCC 35704, converting the primary bile acid CA to the secondary DCA by the removal of the 7α-hydroxyl group.

4.1.4 Aims and Hypotheses

The first aim of this study is to investigate the response of *C. difficile* spores to *C. scindens* derived secondary bile acids. This will be achieved by challenging the spores of six *C. difficile* clinical isolates with the cell free supernatants (CFS) obtained from *C. scindens* cultures, supplemented with and without the primary bile acid, CA (CFS +CA, CFS -CA). The response of spores of *C. difficile* will be assessed in terms of the initiation of spore germination and outgrowth. This will aid in supporting the hypothesis that *C. scindens* derived secondary bile acids inhibit *C. difficile* spore germination or outgrowth, as well as providing evidence for a link between *C. difficile* inhibition and the bile acid metabolism of *C. scindens*.

The second aim of this study is to elucidate the genes of the bile acid inducible operon of *C. scindens* strain ATCC 35704. This will be achieved by RNA sequencing (RNA-seq) and analysis of RNA extracted from *C. scindens* strain ATCC 35704 cultures in CA induced and non-induced conditions. The hypothesis is that the unknown genes of the reductive arm of the pathway will be upregulated in CA induced culture conditions. These gene expression data would provide insight in to the transcriptomic states of *C. scindens* during the biotransformation of bile acids and would potentially aid in the identification of genes involved in the reductive arm of the 7α-dehydroxylating pathway of *C. scindens* ATCC 35704.

For RNA-seq, high quality RNA is required, with an RNA integrity number (RIN) of at least 8.0. To achieve this, an RNA extraction protocol was developed to extract high yield and quality RNA for *C. scindens.*

4.2 Results

4.2.1 The Effect of C. scindens *Derived Secondary Bile Acids on the Germination of* C. difficile *Spores*

To investigate the effect of *C. scindens* derived secondary bile acids on the initiation of *C. difficile* spore germination, spore preparations of at least 95 % purity were prepared for each strain (as outlined in 2.3.4). Spores were heat treated at 65 ^{o}C for 30 minutes and incubated in BHIS broth supplemented with 1 % TC, 1 % TC and 2 mM DCA, 1 % TC and CFSs obtained from *C. scindens* cultures supplemented with and without CA. BHIS without germinant supplementation was used as a negative germination control. OD_{600} measurements were taken every minute for 10 minutes and then every 10 minutes, over an 80-minute period (2.3.9. Germination initiation was detected as a decrease in OD_{600} as the spores rehydrate, becoming less optically dense and losing spore specific characteristics.

Germination initiation was observed in all cases, except for the negative control. This indicates that DCA and the secondary bile acids generated by *C. scindens* from CA does not inhibit the germination of the *C. difficile* strains of study (Figures 4.3.1, 4.3.2).

DH1916 spores showed an average drop in OD_{600} of 15.8 %, with an average final OD_{600} measurement of 0.84. The most significant average decrease of 14 % was observed between 6 and 20 minutes (Figure 4.3.1).

R12801 spores showed average drop of 22.9 % in OD_{600}, with an average final OD_{600} measurement of 0.77. The most significant average decrease of 18.4 % was observed between 0 and 3 minutes (Figure 4.3.1).

Serogroup F spores showed an average drop in OD_{600} of 20.5 %, with an average final OD_{600} measurement of 0.8. The most significant average decrease of 15.5 % was observed between 0 and 3 minutes (Figure 4.3.1).

L26 spores showed an average drop in OD_{600} of 20.5 %, with an average final OD_{600} measurement of 0.79. The most significant average decrease of 17.3 % was observed between 0 and 3 minutes (Figure 4.3.2).

EK15 spores showed an average drop in OD_{600} of 20.3 %, with an average final OD_{600} measurement of 0.8. A decrease of 19.9 % was observed between 0 and 7 minutes (Figure 4.3.2).

EK28 spores showed an average drop in OD_{600} of 14.9 %, with an average final OD_{600} measurement of 0.85. A decrease of 8.9 % was observed between 0 and 5 minutes (Figure 4.3.2).

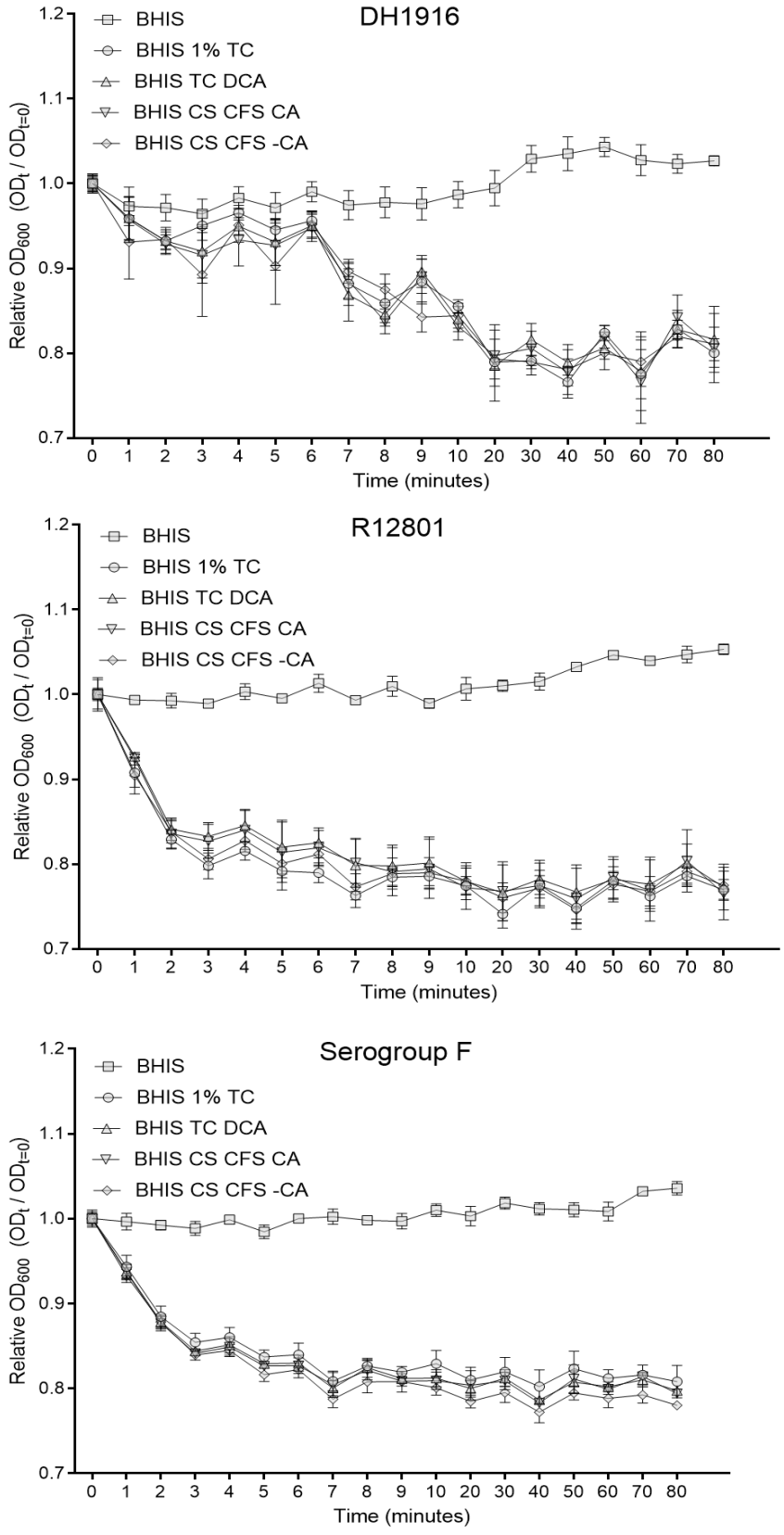


Figure 4.3.1 Initiation of Germination: Spores of *C. difficile* isolates were challenged with DCA and *C. scindens* cell free supernatants (+/- CA) in the presence of the germinant TC. The initiation of germination proceeded in all test cases with observed decreases in OD_{600}, and no drop in OD_{600} observed for the negative no germinant controls. Plotted values are mean ± SD normalised to the starting OD_{600} value at $T_{(0)}$.

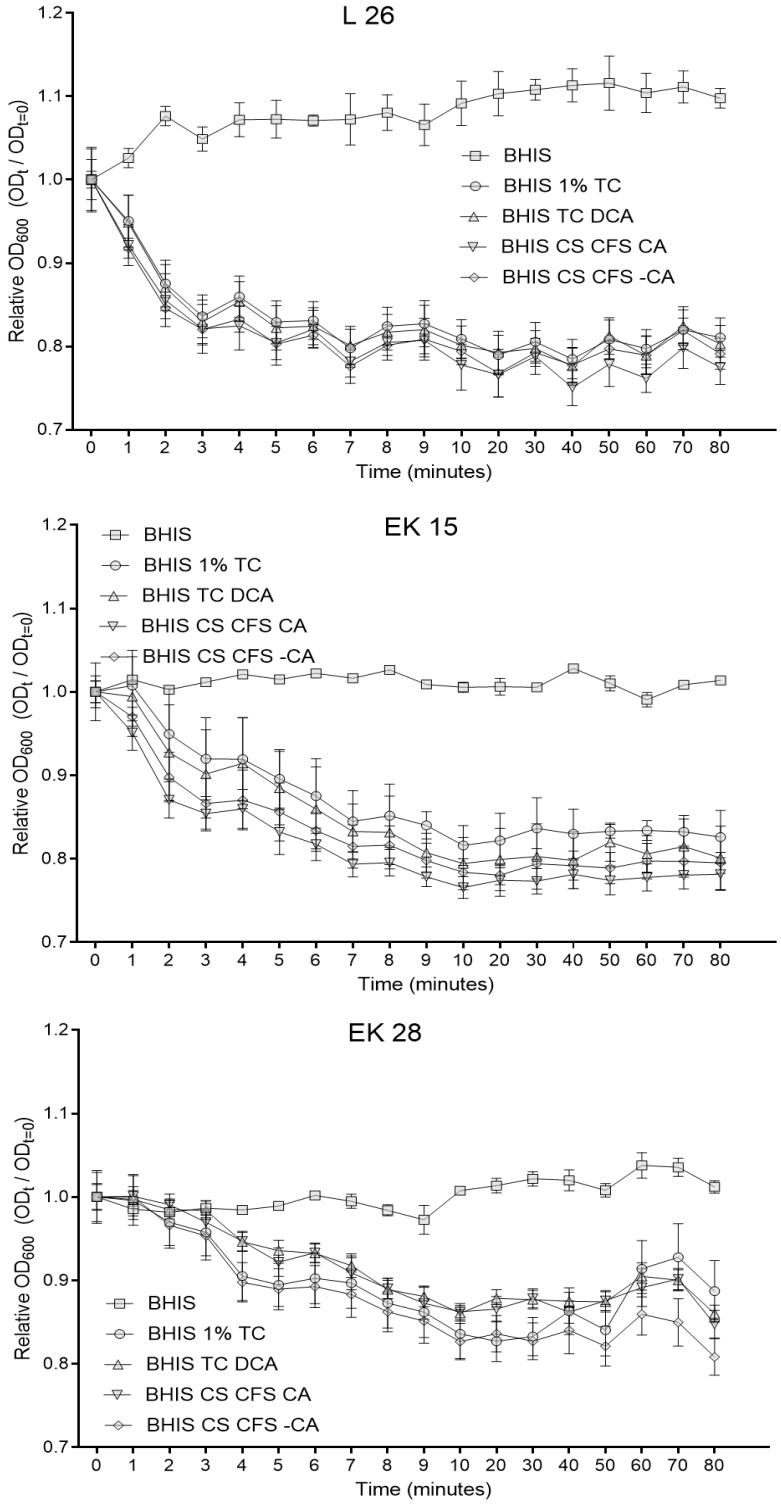


Figure 4.3.2 Initiation of Germination: Spores of *C. difficile* isolates were challenged with DCA and *C. scindens* cell free supernatants (+/- CA) in the presence of the germinant TC. The initiation of germination proceeded in all test cases with observed decreases in OD_{600}, and no drop in OD_{600} observed for the negative no germinant controls. Plotted values are mean ± SD normalised to the starting OD_{600} value at $T_{(0)}$.

4.2.2 The Effect of C. scindens *Derived Secondary Bile Acids on the Outgrowth of* C. difficile *Spores*

To investigate the effect of *C. scindens* derived secondary bile acids on the outgrowth of *C. difficile* spores, spore preparations of at least 95 % purity were prepared for each strain (2.3.4). Spores were heat treated at 65 ^{o}C for 30 minutes and incubated in BHIS broth, supplemented with 1 % TC, 1 % TC and 2mM DCA and 1 % TC with CFSs obtained from *C. scindens* cultures supplemented with and without CA. OD_{600} measurements were taken every hour over a 15 h period (2.3.10).

DH1916 spores incubated with TC and CFS -CA both germinated and grew out. The TC incubated spores lagged until 5 h, exiting exponential phase at 10 h, reaching an average peak OD_{600} of 0.95 at 12 h. The spores incubated with TC CFS -CA lagged until 5 h, exiting exponential phase at 11 h, reaching an average peak OD_{600} of 0.85. The growth profiles for both TC and CDS -CA were found to be significantly different statistically (P = 0.0159). Spores incubated with TC DCA and TC CFS +CA were all inhibited in terms of outgrowth (Figure 4.4).

R12801 Spores incubated with TC and CFS -CA both germinated and grew out. The TC incubated spores lagged until 6 h, exiting exponential phase at 11 h and reached an average peak OD_{600} of 0.88 at 14 h. Spores incubated with TC CFS -CA lagged until 7 h, exiting exponential phase at 12 h, reaching an average peak OD_{600} of 0.81 at 14 h. The growth profiles for both TC and CDS -CA were found to be significantly different statistically (P = 0.0056). R12801 spores incubated with TC DCA and TC CFS +CA were all inhibited in terms of outgrowth (Figure 4.4).

Serogroup F Spores incubated with TC and CFS -CA both germinated and grew out. The TC incubated spores lagged until 6 h, exiting exponential phase at 11 h and reached an average peak OD_{600} of 0.95 at 14 h. Spores incubated with TC CFS -CA lagged until 7 h, exiting exponential phase at 12 h, reaching an average peak OD_{600} of 0.83 at 12 h. The

growth profiles for both TC and CDS -CA were found to be significantly different statistically ($P < 0.0001$). Serogroup F spores incubated with TC DCA and TC CFS +CA were all inhibited in terms of outgrowth (Figure 4.4).

L26 Spores incubated with TC and CFS -CA both germinated and grew out. The TC incubated spores lagged until 4 h, exiting exponential phase at 9 h and reached an average peak OD_{600} of 0.9 at 12 h. Spores incubated with TC CFS -CA lagged until 4 h, exiting exponential phase at 9 h, reaching an average peak OD_{600} of 0.79 at 12 h. The growth profiles for both TC and CDS -CA were found to be significantly different statistically ($P = 0.0089$). Serogroup F spores incubated with TC DCA and TC CFS +CA were all inhibited in terms of outgrowth (Figure 4.4).

EK15 Spores incubated with TC and CFS -CA both germinated and grew out. The TC incubated spores lagged until 5 h, exiting exponential phase at 8 h and reached an average peak OD_{600} of 0.87 at 12 h. Spores incubated with TC CFS -CA lagged until 6 h, exiting exponential phase at 8 h, reaching an average peak OD_{600} of 0.79 at 12 h. The growth profiles for both TC and CDS -CA were found to be significantly different statistically ($P = 0.0010$). EK15 spores incubated with TC DCA and TC CFS +CA were all inhibited in terms of outgrowth (Figure 4.4).

EK28 Spores incubated with TC and CFS -CA both germinated and grew out. The TC incubated spores lagged until 4 h, exiting exponential phase at 8 h and reached an average peak OD_{600} of 0.87 at 15 h. Spores incubated with TC CFS -CA lagged until 4 h, exiting exponential phase at 9 h, reaching an average peak OD_{600} of 0.71 at 12 h. The growth profiles for both TC and CDS -CA were found to be significantly different statistically ($P < 0.0001$). EK28 spores incubated with TC DCA and TC CFS +CA were all inhibited in terms of outgrowth (Figure 4.4).

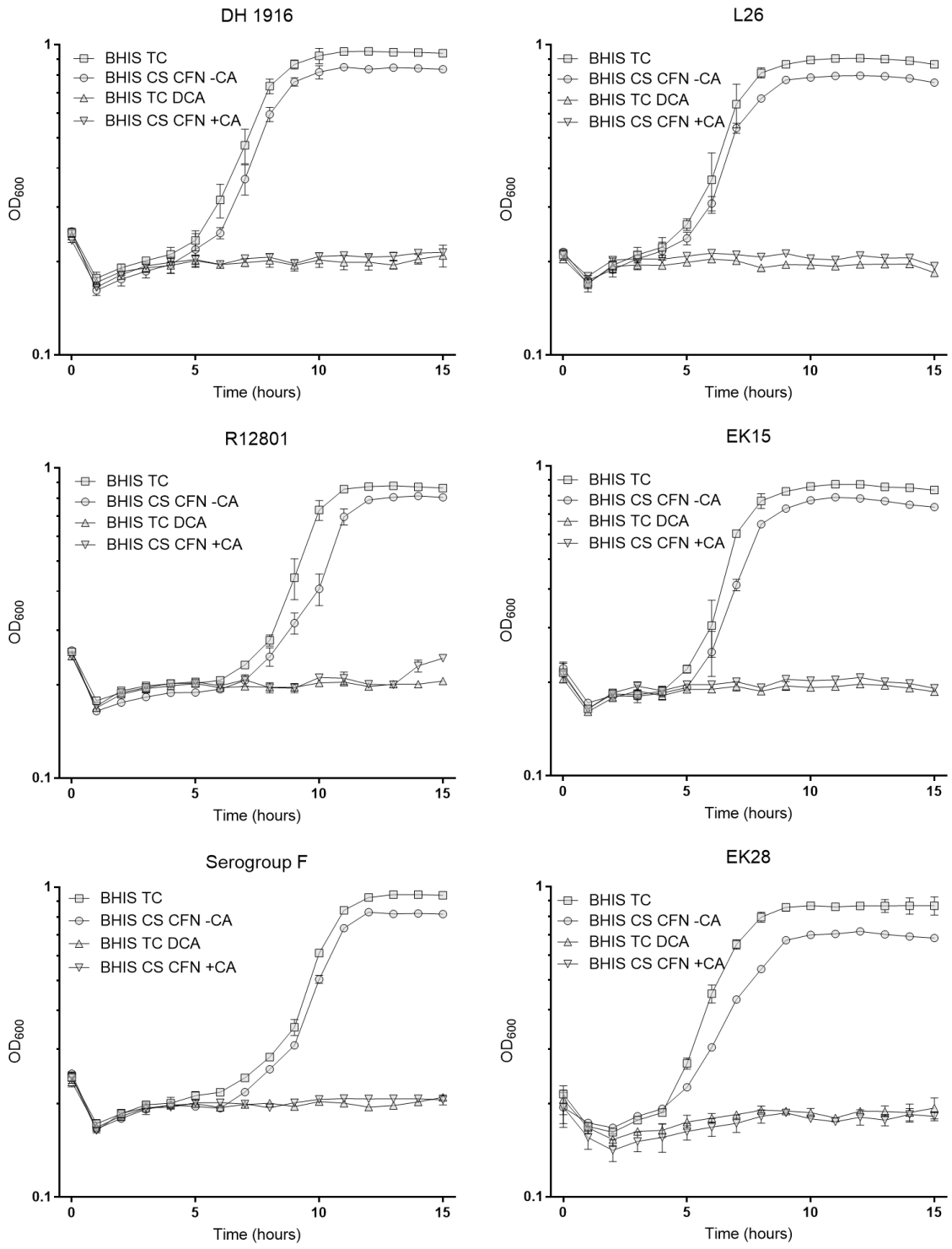


Figure 4.4 Spore Outgrowth: Spores of *C. difficile* isolates were challenged with DCA and *C. scindens* cell free supernatents (+/- CA) in the presence of the germinant TC. Outgrowth of spores was inhibited for all strains when incubated with CFS+CA and proceeded for spores incubated with CFS -CA, with an initial drop in OD_{600} for all strains. Plotted values are mean ± SD.

4.2.1 C. scindens *Cholic Acid Growth*

To study gene expression between both CA and non-CA induced conditions, it was necessary to find a concentration of CA at which the growth between cultures was comparable, to reduce the introduction of variables extraneous to bile acid metabolism.

To establish the concentration of CA to be used in subsequent RNA extraction experiments, *C. scindens* growth rate profiles in a range of CA concentrations were established by inoculating overnight cultures of *C. scindens* ATCC 35704 in to pre-reduced BHIS broth supplemented with 0 - 500 μM CA and incubated as detailed in 2.3.1.

Growth in cultures supplemented with 300 – 500 μM CA was significantly reduced compared to the BHIS control, with P values > 0.0001. The growth profiles of the 100 and 200 μM cultures were found to be not significantly different when compared to the BHIS control, with P values 0.8951 and 0.5254 respectively (Figure 4.5).

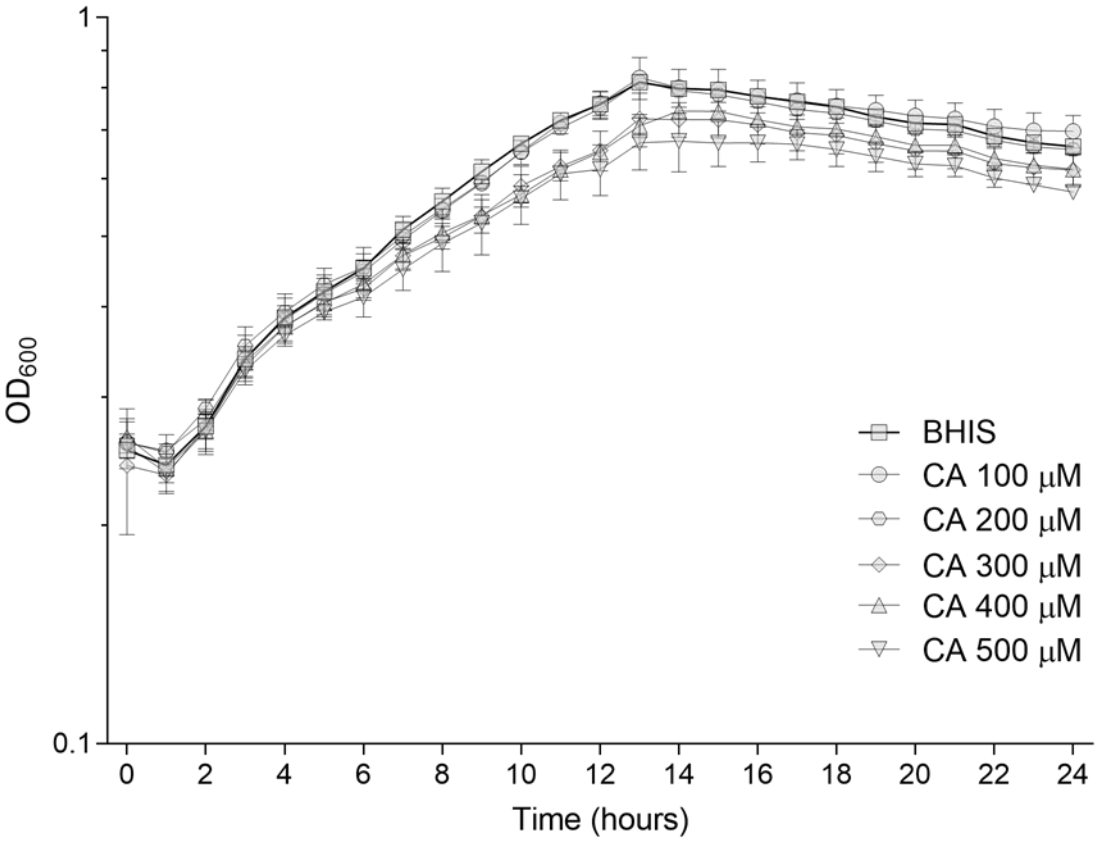


Figure 4.5 Cholic Acid Growth Curves for *C. scindens* strain ATCC 35704: Growth rates for *C. scindens* cultured in BHIS supplemented with a range of Cholic Acid (CA) concentrations, from 100 – 500 μM, as well as BHIS control. Plotted values are mean ± SD.

4.2.2 C. scindens *Growth Curve and In-Vitro Bile Acid Transformations*

To ascertain the time points for RNA extraction, *C. scindens* ATCC 35704 growth rate was established by inoculating overnight cultures in to pre-reduced BHIS broth supplemented with 200 μM CA and incubated as detailed in 2.3.1.

Cultures appeared to lag until the 3 h time point, when exponential phase is reached. Cultures began to leave exponential growth phase at 10 h, reaching stationary phase at 12 h, where a peak OD_{600} of 0.4 was reached. The culture then declined to an OD_{600} of 0.32 (Figure 4.6).

Samples were also taken for bile acid analysis to establish the time points at which the *bai* operon and associated genes were being expressed. Samples were taken at 4, 8 and 24 h in triplicate.

At 4 h, averages of 108.89 µM CA, 0.215 µM DCA and 0.41 µM of 12-keto-Deoxychocholic acid (12-keto-DCA) was measured. At 8 h, 4.86 µM CA, 4.396 µM DCA and 91.63 µM of 12-keto-LCA was measured. At 24 h, 0 µM CA, 1.51 µM DCA and 128.375 µM of 12-keto-LCA was measured (Figure 4.6).

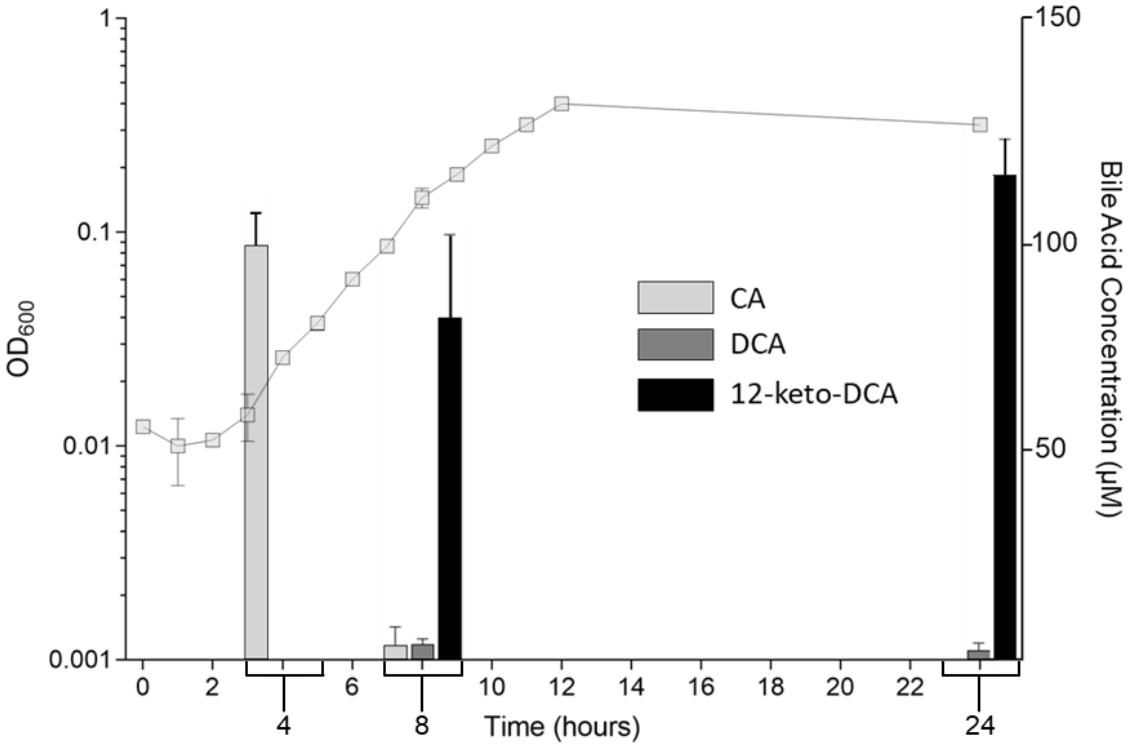


Figure 4.6 *C. scindens* Growth Rate and In-vitro Bile Acid Transformations:
The Growth of *C. scindens* over 24 hours with analysis of bile acids at **4**, **8** and **24** hours, showing transformation of cholic acid (CA) to deoxycholic acid (DCA) and to 12-keto-DCA. Plotted values are mean ± SD.

4.2.3 C. scindens *Growth Curves for RNA Extraction*

For RNA extractions, *C. scindens* ATCC 35704 was grown in BHIS and BHIS supplemented with 200 μM CA. Samples for RNA extraction were taken at 4, 6, 8 and 10 hours, covering early, mid and late exponential. Samples were also taken for OD_{600} measurement, to ensure the growth profiles for each culture were consistent.

Both cultures entered the exponential phase of growth after a short lag period at 2 - 3 h. Exponential growth continued until 9 – 10 h, reaching peak average OD_{600} readings of 0.55 and 0.50 for the BHIS and BHIS-CA cultures, respectively. At 24 h, the BHIS and BHIS-CA cultures declined to average OD_{600} readings of 0.45 and 044 respectively (Figure 4.7).

The growth profiles for each condition were found to be not statistically dissimilar ($P = 0.09621$).

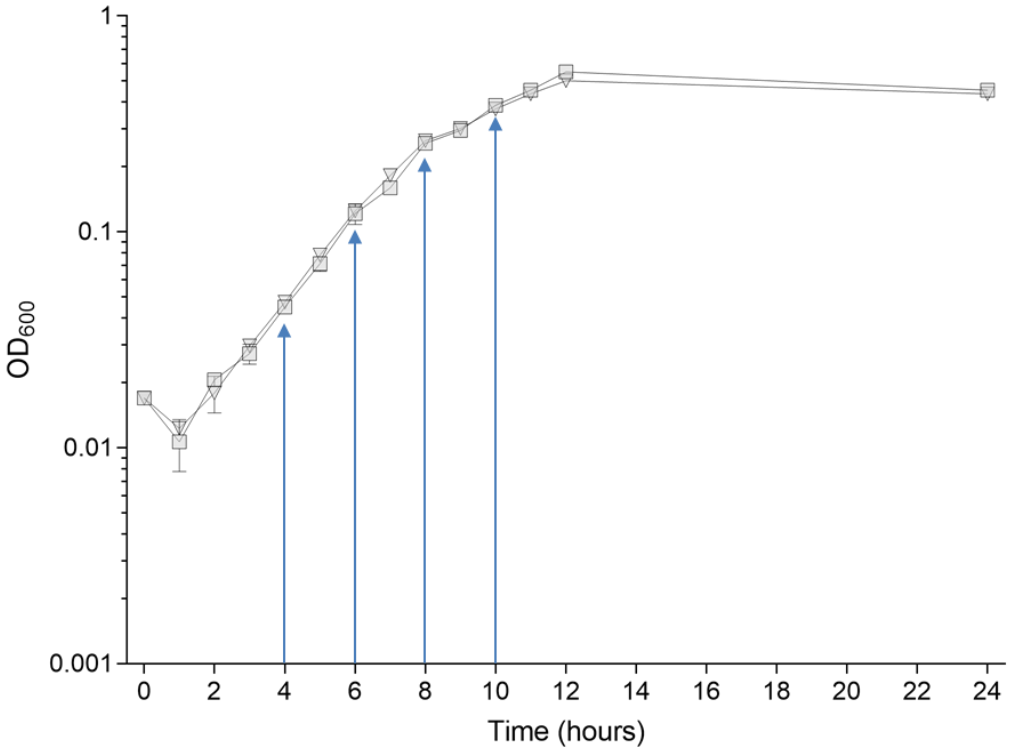


Figure 4.7 Growth Rates of *C. scindens* Cultures for RNA Extractions: *C. scindens* strain ATCC 35704 was cultured in BHIS and BHIS supplemented with 200 µM cholic acid (CA). Samples for RNA extractions were taken at 4, 6, 8 and 10 h. Growth curves for both conditions were statistically comparable. Plotted values are mean ± SD.

5.2.2 Assembly and Annotation of the C. scindens ATCC 35704 Genome

A previously released genome for *C. scindens* strain ATCC 35704 was made available by the Washington University Genome Sequencing Centre as a part of the Human Microbiome Project in 2007 (NCBI accession number: NZ_ABFY00000000). The whole genome shotgun assembly at a level of 41 scaffolds was deemed inappropriate for RNAseq analysis and it was therefore decided to assemble a more complete genome sequence utilising next generation sequencing methods.

The genome of *C. scindens* strain ATCC 35704 was determined using a combination of PacBio SMRT and illumina MiSeq sequencing technologies. The sequence was assembled by mapping MiSeq paired end reads to a single contig generated by SMRT sequencing, as a reference assembly and a consensus sequence extracted in CLC Workbench 9.5.3 (2.8.2). The resulting 3,645,103 bp nucleotide sequence was annotated using the RAST pipeline and manual curation with the number of annotated features as follows, genes; 3242, coding sequences (CDS); 3182, rRNA; 12, tRNA; 58. The genome guanine plus cytosine (G+C) content is 46.2 %, notably high for a member of the genus *Clostridium*, but is consistent with the initial reported value of 45 % for the type strain, determined by the thermal melting point methodology (Morris *et al.,* 1985) (Figure 4.8).

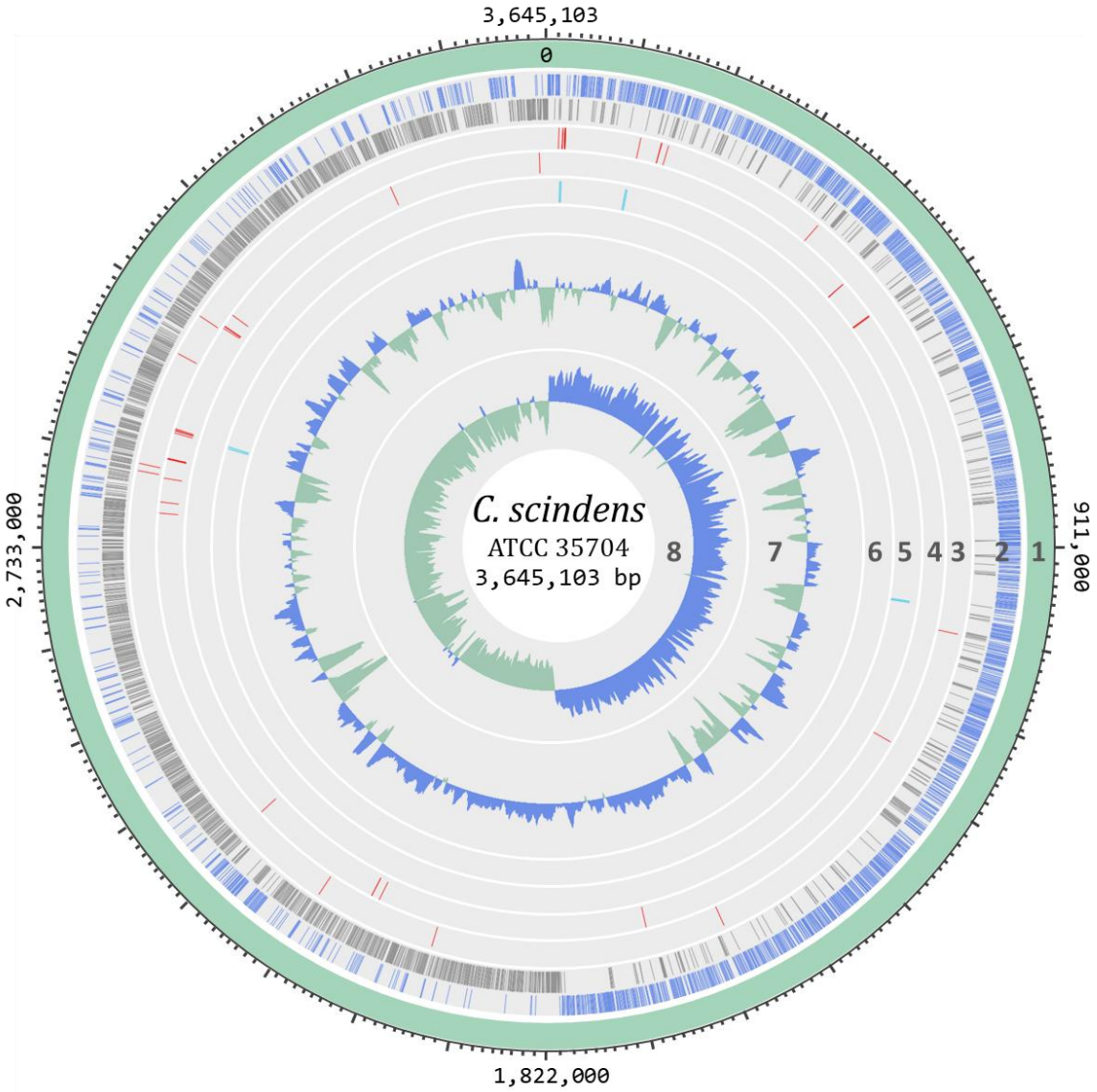


Figure 4.8 The Genome of *C. scindens* ATCC 35704: Track **1**: source sequence; Track **2**: Coding sequences on the **positive** and **negative** strand; Track **3**: positive strand tRNA CDS; Track **4**: negative strand tRNA CDS; Track **5**: positive strand rRNA CDS; Track **6**: negative strand rRNA CDS; Tracks **7** and **8**: G+C content and G+C skew, (**blue** above average GC, **green** below average GC).

4.2.4 Development and Optimisation of RNA Extraction from C. scindens

The initial RNA extraction protocol was based on previously detailed extraction protocols utilised in the group for isolation of total RNA from *C. difficile* (2.5.19).

Samples were quantified using a NanoDrop Lite spectrophotometer (2.5.17) and diluted appropriately for RNA quality assessment using an Agilent 2100 bioanalyser (2.5.18). The instrument detects RNA degradation by running a micro-gel and computing the RIN, on a scale of 1 – 10. Electropherograms are outputted, with fluorescence units (FU) plotted against time in seconds, showing rRNA peaks and other dominating types of RNA. Signal between the marker peak and the 16 S rRNA peak is an indicator of RNA degradation (Appendix 8.1)

The initial extractions revealed that these methods were not appropriate for *C. scindens* and/or the conditions of these experiments, with RNA extractions found to be of quality not sufficient for RNA-seq. Degradation was observed in all samples, but particularly evident in the early exponential phase samples.

The large volumes of culture required at early time points increased the sample processing time and made the use of RNA stabilisation solutions at this step non-viable. After further experiments, it was subsequently decided to use a combination of liquid nitrogen snap freezing, and RNAprotect to combine cell pellets prior to extraction (2.5.2). These optimisations yielded higher quality RNA at all time points sampled, suitable for RNA-seq (Tables 4.2, 4.3, Appendix 8.1).

Table 4.2 RNA Sample Details for 4 and 6 Hour Timepoints

Sample	Concentration (ng/µL)	A260/280	RIN
4_1	284.3	2.14	10
4_2	296.2	2.14	10
4_3	374.9	2.14	10
4_4	655.3	2.07	8.9
4_5	468.2	2.05	8.0
4_6	873.7	2.09	9.8
6_1	506.4	2.15	9.7
6_2	563.4	2.17	10
6_3	532.3	2.14	10
6_4	535.4	2.14	8.3
6_5	579.7	2.14	8.1
6_6	514.3	2.14	8.0

The sample names, concentrations, ratio of absorbance at 260/280 nm and the RNA Integrity Number for the 4 and 6 hour time points.

Table 4.3 RNA Sample Details for 4 and 6 Hour Timepoints

Sample	Concentration (ng/µL)	A260/280	RIN
8_1	232.3	2.17	10
8_2	248.2	2.18	10
8_3	269	2.18	10
8_4	407.5	2.15	10
8_5	545.7	2.17	9.9
8_6	385.4	2.17	10
10_1	264.7	2.16	10
10_2	297.1	2.17	10
10_3	235.4	2.16	10
10_4	242.8	2.17	10
10_5	398.5	2.17	10
10_6	220.9	2.17	10

The sample names, concentrations, ratio of absorbance at 260/280 nm and the RNA Integrity Number for the 8 and 10 h time points.

Following appropriate quality controls, RNA samples were sent for sequencing at the University of Nottingham Deepseq sequencing facility. RNA-seq was carried out in duplicate, with the two best samples from each time point and condition selected.

RNA samples were quantified using a QuBit fluorometer as in 2.5.17. Sequencing was carried out as detailed in 2.7.4 and analysis as described in 2.8.3.

4.2.5 Preliminary Identification of Uncharacterised Bile Acid Metabolism Genes

Prior to RNA-seq analysis, a search of the annotated genome of *C. scindens* ATCC 35704 was carried out for potential gene candidates for the reductive arm of the 7α-dehydroxylation pathway. The search criteria were based upon annotations indicating reductase activity and NAD/FAD binding domains. Fifteen gene candidates were identified, including 7 genes belonging to the short chain dehydrogenase/reductases (SDR) family, along with the characterised SDRs, *baiA1* and *baiA2*. There were 3 genes belonging to the oxidoreductase family, with Rossman fold nucleotide binding domains. Of particular interest were two genes, encoding for a FAD binding flavoprotein and an NAD/FAD dependent oxidoreductase. Both genes code for proteins containing FAD binding domains and appear to exhibit oxidoreductase activity (Table 4.4). The characterised flavin oxidoreductases *baiCD* and *baiH* genes were also identified by these search parameters but were excluded from the table.

Table 4.4 Identified Gene Candidates for the Reductive Arm of 7α-Dehydroxylation

Gene ID	Annotation	Predicted Mr (kDa)	Domains	Designation
Gene 152	Oxidoreductase Family Protein	48.9	Oxidoreductase Family, NAD binding Rossman Fold	152
Gene 222	Oxidoreductase Family Protein	37.6	Oxidoreductase Family, NAD binding Rossman Fold	222
Gene 2401	Oxidoreductase Family Protein	48.2	Oxidoreductase Family, NAD binding Rossman Fold	2401
Gene 1319	Oxidoreductase, aldo/keto reductase family	43.5	Aldo/keto reductase family	1319
Gene 622	flavin reductase domain protein	22.5	Flavin reductase like domain	622
Gene 719	FAD binding flavoprotein	60	FAD dependent oxidoreductase	719
Gene 1190	NAD(P)/FAD-dependent oxidoreductase	44.4	flavoprotein, HI0933 family	1190
Gene 3872	Oxidoreductase Zinc Binding Dehydrogenase	39.2	L-threonine dehydrogenase	3872
Gene 3967	3-oxoacyl- reductase	25.1	Enoyl-(Acyl carrier protein) reductase, SDR Family	3967
Gene 3818	3-oxoacyl- reductase	28.2	Enoyl-(Acyl carrier protein) reductase, SDR Family	3818
Gene 3593	3-oxoacyl- reductase	26.3	Enoyl-(Acyl carrier protein) reductase, SDR Family	3593
Gene 3221	3-oxoacyl- reductase	28.5	Enoyl-(Acyl carrier protein) reductase, SDR Family	32212
Gene 4157	3-oxoacyl- reductase	25.4	Enoyl-(Acyl carrier protein) reductase, SDR Family	4157
Gene 2756	3-oxoacyl- reductase	28.6	Enoyl-(Acyl carrier protein) reductase, SDR Family	2756
Gene 136	3-oxoacyl- reductase	24.4	Enoyl-(Acyl carrier protein) reductase, SDR Family	136

4.2.6 Differential Gene Expression of the Bile Acid Inducible Operon

Across all time points, the genes associated with the bile acid inducible operon were significantly upregulated in CA induced conditions, when compared to non-CA induced conditions with P values < 0.005 (Figures 4.9 and 4.10). The regulatory *bar* genes, *barA* and *barB* are constitutively expressed (Figure 4.9), with no significant fold changes between the induced and non-induced conditions at all time points (Tables 4.5, 4.6, 4.7 and 4.8).

During mid exponential phase, the polycistronic *bai* operon is most highly expressed (Figure 4.10). The normalised counts at the 6 h time point are the highest, followed by the 8 h, 10 h and 4 h time points, with the differences in expression levels being statistically significant with P values of 0.003 - 0.001 (Tables 4.5, 4.6, 4.7 and 4.8).

The expression of the monocistronic *baiA1* gene also follows this transcriptomic pattern, being more highly expressed at 6 and 8 h, when compared to the 4 and 10 h time points (P values of 0.0027 – 0.0078) (Tables 4.5, 4.6, 4.7 and 4.8).

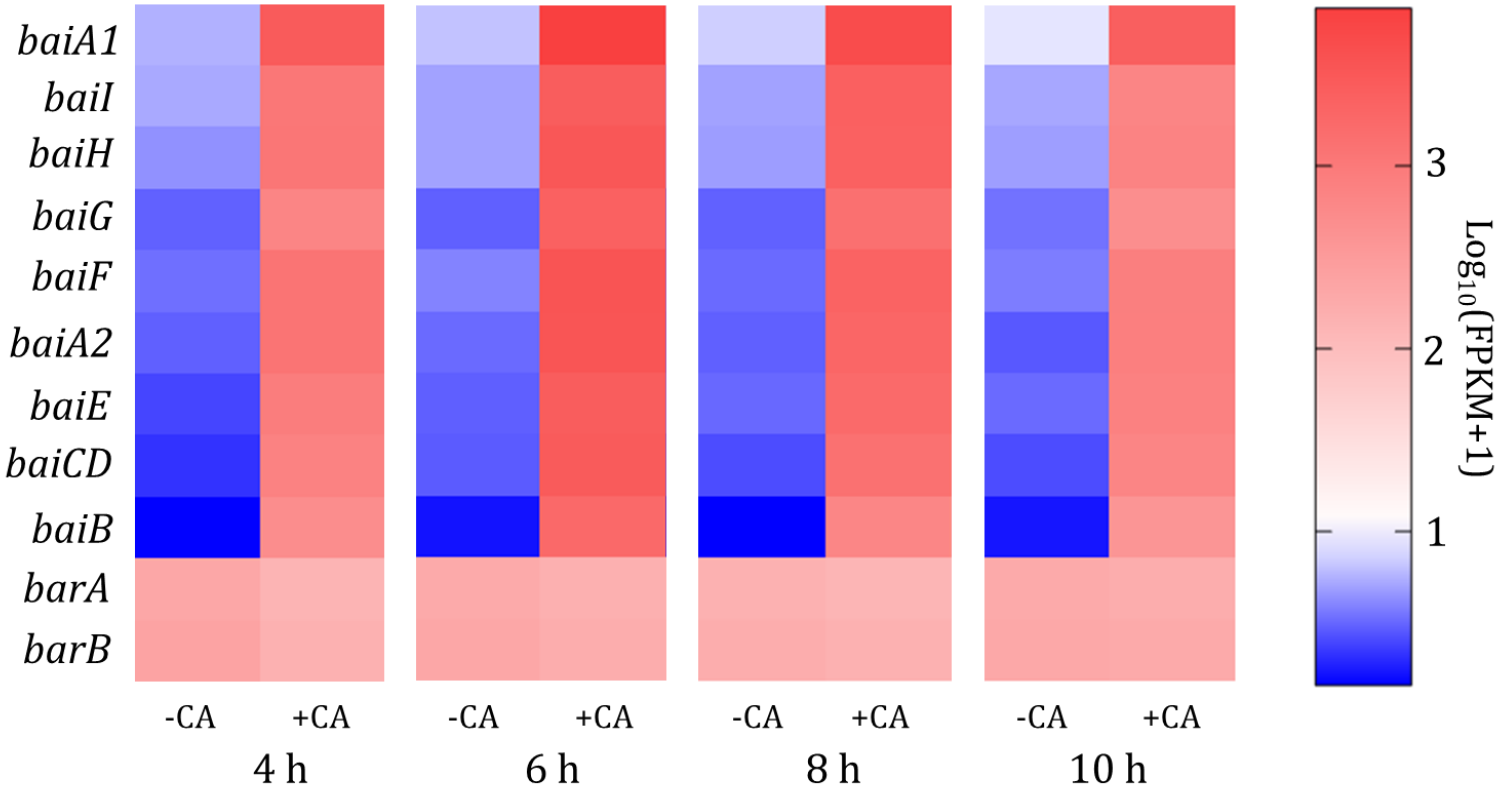


Figure 4.9 Gene expression of The Bile Acid Inducible Operon at 4, 6, 8 and 10 Hours: Heatmaps comparing the Log_{10} (FPKM+1) of the *bai* operon and *bar* regulator genes in cholic acid (CA) induced and non-induced conditions. At each time point the *bai* genes are significantly upregulated in the CA induced conditions, when compared to non-CA induced conditions ($P < 0.005$), the regulatory *bar* genes are constitutively expressed at all timepoints.

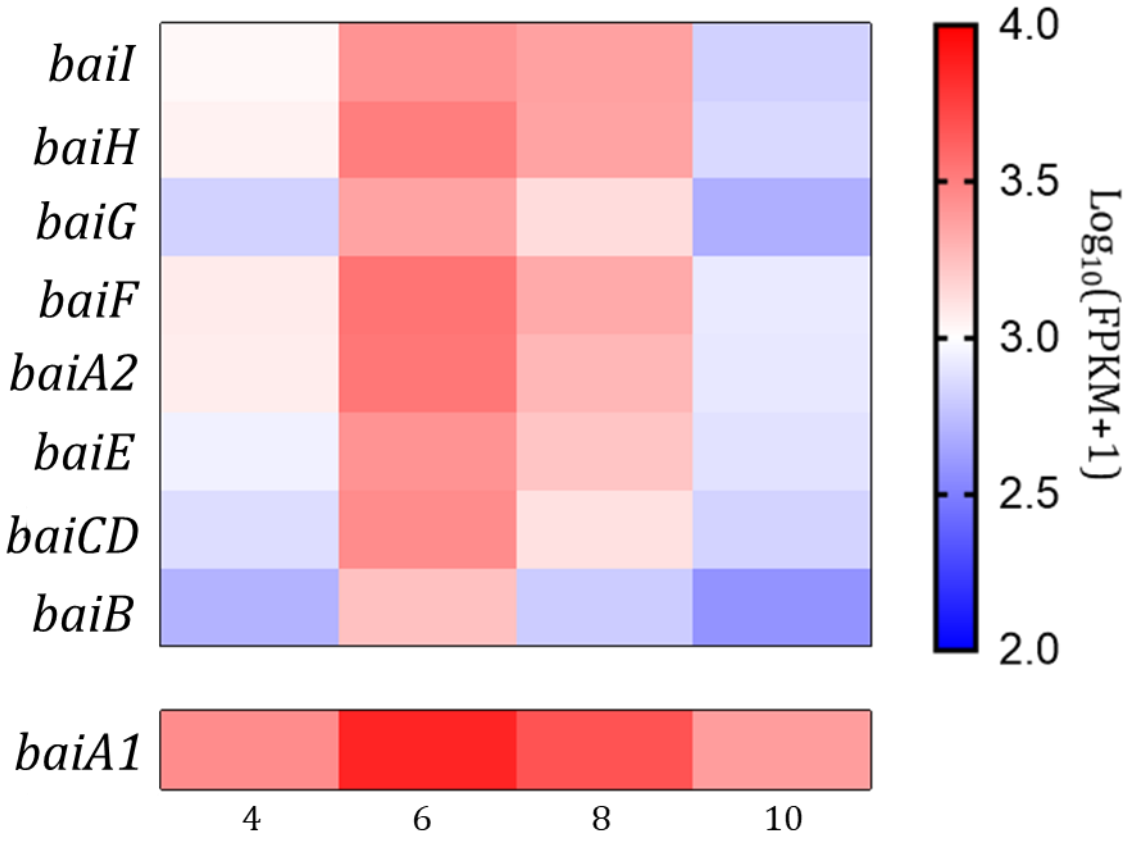


Figure 4.10 Expression Levels of the Bile Acid Inducible Operon in CA Induced Conditions: Heatmap comparing gene expression, as Log_{10} (FPKM+1), of the genes of the polycistronic *bai* operon and the monocistronic *baiA1* gene over 4, 6, 8 and 10 h time points. The genes of the *bai* operon were the most highly expressed at the 6 and 8 h time points, when compared to the 4 and 10 h time points (P = 0.003 - 0.001)

Table 4.5 Differential Expression of the *bai* genes in CA Induced and Non Induced Conditions at 4 Hours

Gene ID	Gene Annotation	Mean Uninduced FPKM	Mean CA Induced FPKM	Log_2 Fold Change	P Value	Significant
Gene 3019	***baiI***	4.26	1124.70	6.82	4.20E-72	Yes
Gene 3020	***baiH***	3.36	1174.15	7.40	1.32E-97	Yes
Gene 3021	***baiG***	2.02	680.29	7.32	6.32E-94	Yes
Gene 3022	***baiF***	2.34	1220.81	7.95	1.30E-117	Yes
Gene 3023	***baiA2***	1.99	1195.75	8.09	1.50E-122	Yes
Gene 3024	***baiE***	1.42	891.93	7.93	1.75E-107	Yes
Gene 3025	***baiCD***	1.33	875.24	8.19	1.87E-125	Yes
Gene 3027	***baiB***	0.44	501.47	8.82	1.84E-149	Yes
Gene 1825	***baiA1***	4.07	2846.49	8.48	1.08E-153	Yes
Gene 3029	***barA***	2.32	2.12	-0.13	N/A	N/A
Gene 3030	***barB***	2.38	2.17	-0.13	N/A	N/A

Table 4.6 Differential Expression of the *bai* genes in CA Induced and Non Induced Conditions at 6 Hours

Gene ID	Gene Annotation	Mean Uninduced FPKM	Mean CA Induced FPKM	Log_2 Fold Change	P Value	Significant
Gene 3019	***baiI***	4.09	2648.41	7.33	2.27E-163	Yes
Gene 3020	***baiH***	4.02	3183.21	8.65	0	Yes
Gene 3021	***baiG***	1.98	2295.69	9.77	0	Yes
Gene 3022	***baiF***	2.92	3518.84	9.91	0	Yes
Gene 3023	***baiA2***	2.22	3428.72	9.04	0	Yes
Gene 3024	***baiE***	1.96	2680.55	7.76	1.64E-173	Yes
Gene 3025	***baiCD***	2.18	2995.42	8.96	0	Yes
Gene 3027	***baiB***	0.64	1737.10	10.51	0	Yes
Gene 1825	***baiA1***	7.25	6944.13	9.57	0	Yes
Gene 3029	***barA***	2.27	2.17	-0.06	N/A	N/A
Gene 3030	***barB***	2.32	2.22	-0.06	N/A	N/A

Table 4.7 Differential Expression of the *bai* genes in CA Induced and Non Induced Conditions at 8 Hours

Gene ID	Gene Annotation	Mean Uninduced FPKM	Mean CA Induced FPKM	Log_2 Fold Change	P Value	Significant
Gene 3019	***baiI***	3.95	2341.80	8.22	0	Yes
Gene 3020	***baiH***	3.77	2291.92	8.66	0	Yes
Gene 3021	***baiG***	2.01	1370.30	8.53	0	Yes
Gene 3022	***baiF***	2.23	2151.36	8.98	0	Yes
Gene 3023	***baiA2***	2.06	1915.30	8.50	0	Yes
Gene 3024	***baiE***	2.18	1694.08	8.12	0	Yes
Gene 3025	***baiCD***	1.60	1406.37	8.47	0	Yes
Gene 3027	***baiB***	0.44	626.19	8.76	0	Yes
Gene 1825	***baiA1***	6.01	4424.13	8.87	0	Yes
Gene 3029	***barA***	2.16	2.11	-0.04	N/A	N/A
Gene 3030	***barB***	2.23	2.16	-0.04	N/A	N/A

Table 4.8 Differential Expression of the *bai* genes in CA Induced and Non Induced Conditions at 10 Hours

Gene ID	Gene Annotation	Mean Uninduced FPKM	Mean CA Induced FPKM	Log_2 Fold Change	P Value	Significant
Gene 3019	***baiI***	4.16	672.91	6.31	4.51E-127	Yes
Gene 3020	***baiH***	3.82	731.86	6.62	4.27E-148	Yes
Gene 3021	***baiG***	2.52	512.51	6.20	1.25E-95	Yes
Gene 3022	***baiF***	2.74	884.84	6.89	1.64E-126	Yes
Gene 3023	***baiA2***	1.85	862.51	7.03	2.34E-120	Yes
Gene 3024	***baiE***	2.22	810.19	6.79	4.19E-115	Yes
Gene 3025	***baiCD***	1.63	717.60	7.37	3.15E-157	Yes
Gene 3027	***baiB***	0.69	374.82	7.70	1.67E-186	Yes
Gene 1825	***baiA1***	8.03	2303.87	7.31	1.24E-214	Yes
Gene 3029	***barA***	2.26	2.23	-0.02	N/A	N/A
Gene 3030	***barB***	2.30	2.27	-0.02	N/A	N/A

4.2.7 Whole Genome Differential Gene Expression

MA scatter plots were produced, plotting the $\log_2$ fold change in gene expression against the average normalised counts for each gene, showing transcriptional changes occurring throughout the exponential growth phase.

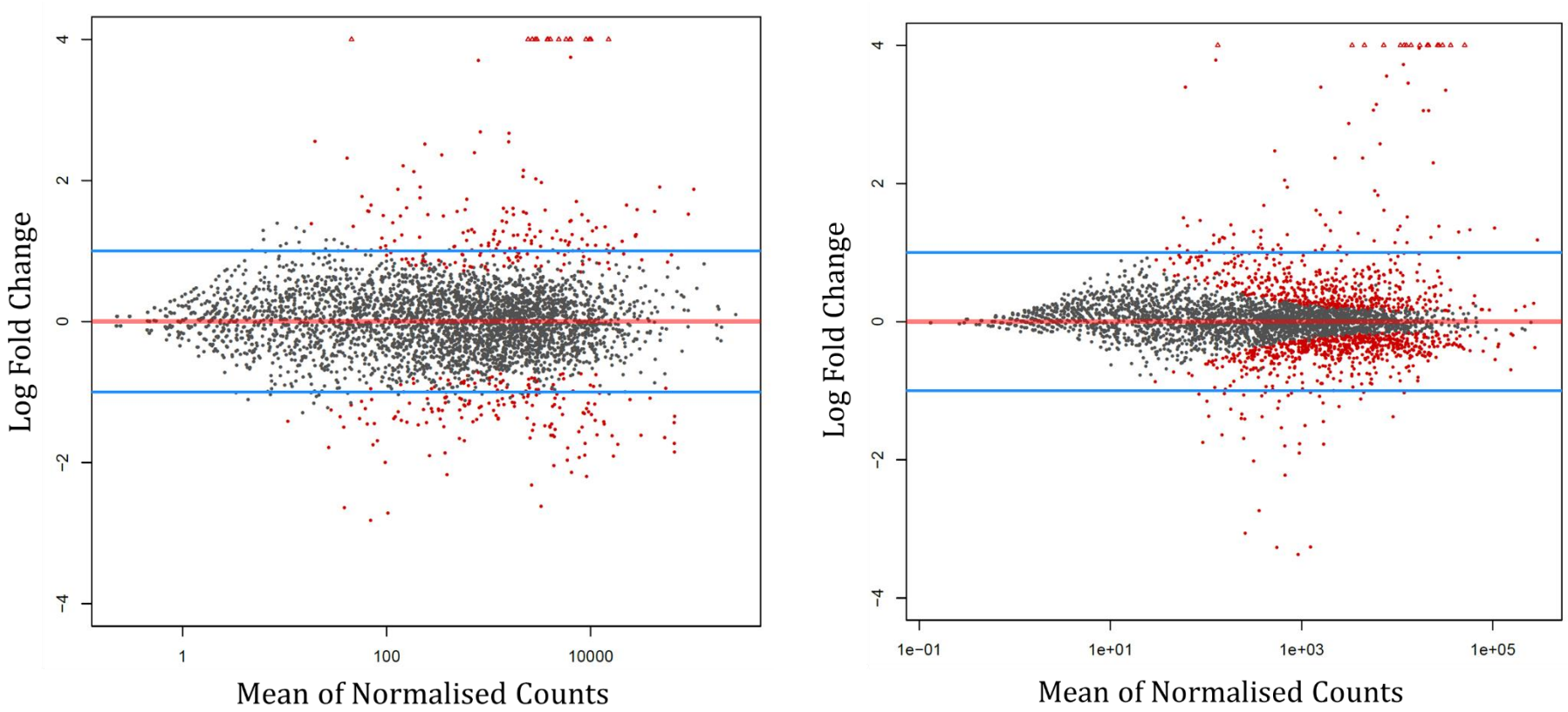
During early exponential phase, at 4 h, 147 genes were significantly upregulated and 129 significantly downregulated between the two conditions (Figure 4.11 A; Appendix 8.2.1). At 4 h, the genes with the highest levels of differential expression were the *bai* genes, an FAD binding reductase and genes involved in sialic acid metabolism, with $\log_2$ fold changes of > 4 over the non-induced condition.

The changes in gene expression become more pronounced during the mid-exponential time points. At 6 h, 482 genes were significantly upregulated and 580 significantly downregulated (Figure 4.11 B; Appendix 8.2.2). The genes with the highest levels of differential expression were the *bai* genes, FAD binding flavoprotein, nuclear transport protein and components of membrane transporters belonging the major facilitator superfamily (MFS) all with $\log_2$ fold changes of > 4.

At 8 h, 255 genes were significantly upregulated and 187 significantly downregulated between the two conditions (Figure 4.12 A; Appendix 8.2.3). The genes with the highest levels of differential expression were the *bai* genes, FAD binding flavoprotein, nuclear transport protein and MFS membrane transporter proteins, with $\log_2$ fold changes of > 4.

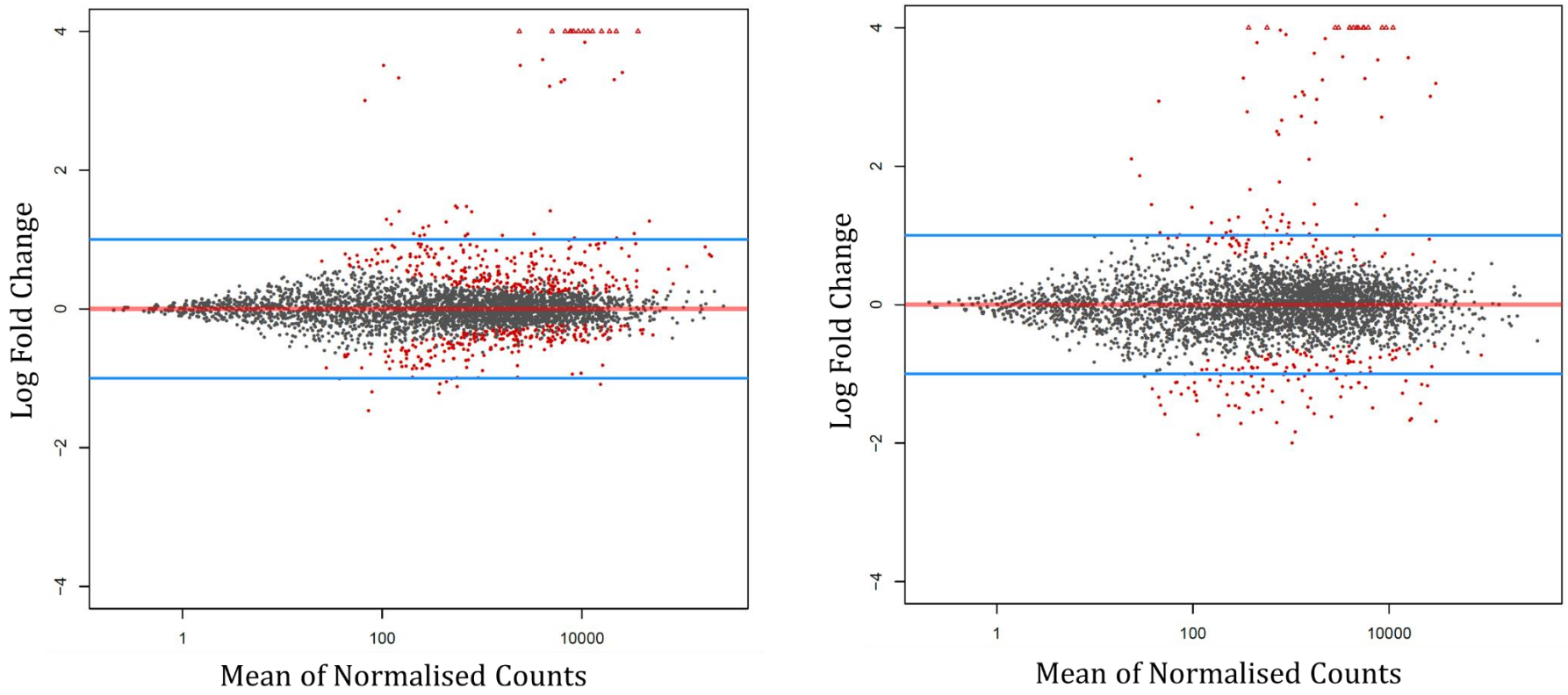
As the cultures begin to exit exponential phase, at 10 h, changes in gene expression become less pronounced with 104 genes significantly upregulated and 114 significantly downregulated between the two conditions (Figure 4.12 B; Appendix 8.2.4). The genes with the highest levels of differential expression, with $\log_2$ fold changes of > 4, at 10 h were the *bai* genes, FAD binding flavoprotein, an MFS membrane transporter and a

substrate binding component of a membrane transporter belonging to the tripartite ATP-independent periplasmic (TRAP) family.



A B

Figure 4.11 A; B Differential gene expression between CA induced and non-induced conditions at 4 (A) and 6 (B) hours: A positive fold change indicates upregulated genes and negative fold change indicates downregulated genes, between the two conditions. **Red** dots represent significant differentially expressed genes between the two conditions. **Grey** dots represent non-significant differential expression. The horizontal **blue** lines represent 1 log fold change. **A**: 4 h, 147 genes significantly upregulate, 129 genes significantly downregulated; **B**: 6 h, 482 genes significantly upregulated, 580 significantly downregulated.



A B

Figure 4.12 A; B Differential gene expression between CA induced and non-induced conditions at 8 (A) and 10 (B) hours: A positive fold change indicates upregulated genes and negative fold change indicates downregulated genes, between the two conditions. **Red** dots represent significant differentially expressed genes between the two conditions. **Grey** dots represent non-significant differential expression. The horizontal **blue** lines represent 1 log fold change. **A**: 8 h, 255 genes significantly upregulate, 187 genes significantly downregulated; **B**: 10 h, 104 genes significantly upregulated, 114 significantly downregulated.

4.2.8 Hierarchical Clustering

To group genes with similar expression profiles, hierarchical clustering analysis was carried out and heatmaps and dendrograms produced (2.8.3). Hierarchical clustering was employed to offer insight in to genes with potentially similar functions (Figure 4.13)

At the 4 h time point, all of the genes of the *bai* operon cluster together in to the same dendrogram (Table 4.9).

At the 6 h time point, the genes of the *bai* operon cluster together, along with a FAD binding flavoprotein, nuclear transport protein, three genes belonging to the oxidoreductase family and MFS membrane transporters (Table 4.10).

At the 8 h time point, the genes of the *bai* operon cluster together along with the gene encoding for an FAD binding flavoprotein (Table 4.10).

At the 10 h time point, the genes of the all of the genes of the *bai* operon cluster together exclusively in to the same dendrogram (Table 4.11).

The previously identified FAD binding flavoprotein listed in Table 4.4 has been highlighted.

Table 4.9 Hierarchical Clustering Dendrogram 3 at 4 Hours

Annotation	4 h $\log_{10}$ (FPKM + 1)	
	-CA	**+CA**
baiB	0.16	2.70
baiCD	0.29	2.76
baiG	0.48	2.83
baiE	0.38	2.95
baiA1	0.70	3.45
baiA2	0.48	3.08
baiF	0.52	3.09
baiH	0.64	3.07
baiI	0.72	3.05

Table 4.10 Hierarchical Clustering Dendrogram 3 at 6 Hours

Annotation	6 h $\log_{10}$ (FPKM + 1)	
	-CA	**+CA**
baiA1	0.64	3.87
baiB	0.22	3.24
baiCD	0.41	3.43
baiG	0.47	3.36
baiE	0.47	3.43
baiI	0.71	3.42
baiH	0.70	3.50
baiF	0.59	3.55
baiA2	0.51	3.54
Threonine dehydrogenase	0.87	2.43
TRAP transporter periplasmic	1.17	2.84
FAD binding flavoprotein	0.95	3.01
MFS Transporter	1.97	2.99
MFS Transporter	1.75	2.80
Glucokinase	1.81	2.79
6-phosphogluconate dehydrogenase	1.80	3.45
D-3-phosphoglycerate dehydrogenase	1.31	3.34
5-keto-D-gluconate 5-reductase	1.40	3.32
TRAP transporter permease	1.48	3.04
4-Hydroxy-2-oxoglutarate aldolase	1.61	3.11
MFS Transporter	1.44	3.01
MFS Transporter	1.59	3.00
Nuclear Transport Family Protein	1.58	3.01

Table 4.11 Hierarchical Clustering Dendrogram 2 at 8 Hours

Annotation	8 h $\log_{10}$ (FPKM + 1)	
	-CA	**+CA**
baiA1	0.85	3.65
baiB	0.16	2.50
FAD binding flavoprotein	1.00	2.46
baiCD	0.42	2.89
baiG	0.48	2.87
baiH	0.68	3.10
baiI	0.69	3.11
baiF	0.51	3.07
baiE	0.50	2.97
baiA2	0.49	3.02

Table 4.12 Hierarchical Clustering Dendrogram 3 at 10 Hours

Annotation	10 h $\log_{10}$ (FPKM + 1)	
	-CA	**+CA**
baiA1	0.87	3.40
baiB	0.23	2.27
baiG	0.55	2.24
baiI	0.71	2.43
baiH	0.68	2.44
baiCD	0.40	2.42
baiF	0.57	2.45
baiA2	0.45	2.44
baiE	0.51	2.43

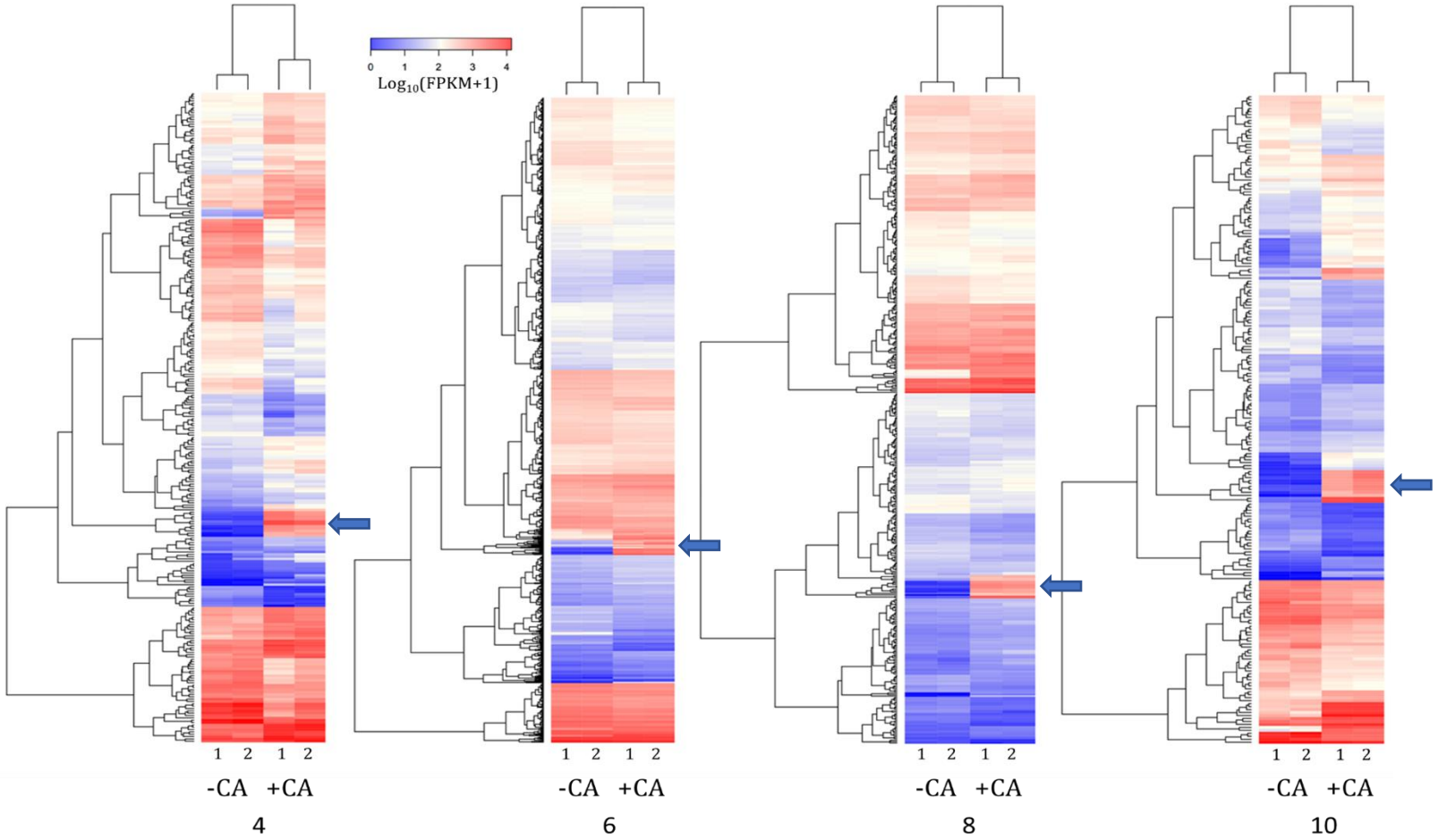


Figure 4.13 Whole Transcriptome Heatmap and Hierarchical Clustering for 4, 6, 8 and 10 hour Time Points: The heatmaps represent gene expression levels as Log_{10} (FPKM+1), with genes grouped in to dendrograms using hierarchical clustering. The *bai* operon genes cluster into the dendrograms highlighted with blue arrows at each time point.

4.3 Discussion

4.3.1 The Reponse of C. difficile *Spores to* C. scindens *Derived Secondary Bile Acids*

The *C. difficile* strains in this study were selected for the reasons detailed in section 3.3.1 and for consistency.

The cell free supernatants (CFS) resulting from *C. scindens* cultured with the primary bile acid CA were hypothesised to yield CFS containing the secondary bile acid DCA, as a product of *C. scindens* mediated 7α-dehydroxylation. The CFS obtained were diluted 1:1 in 2 x concentrated BHIS broth, to yield BHIS broth supplemented with and without the products of CA metabolism, (CFS +CA and CFS -CA). Prior to incubation of *C. difficile* spores, BHIS CFS -CA/+CA were supplemented with the germinant TC.

The initiation of germination assays confirms that the secondary bile acid DCA does not inhibit the initiation of germination for the *C. difficile* strains of study (Figures 4.3.1/2). This was also observed for spores incubated with CFS +CA and CFS -CA, with germination proceeding indicated by the decrease in OD_{600}.

The outgrowth assays show that DCA is inhibitory for *C. difficile* spore outgrowth for the strains of study. Outgrowth was also inhibited for *C. difficile* spores incubated with CFS +CA, demonstrating that the products of *C. scindens* mediated transformation of CA inhibit the outgrowth of the spores of *C. difficile* strains of study. Outgrowth proceeded for spores incubated with CFS -CA, further implicating bile acid metabolism as an inhibitory factor in spores challenged with CFS +CA. The outgrowth of CFS -CA incubated spores was attenuated, with statistical significance, when compared to the positive control of BHIS TC. This is likely due to the sub-optimal conditions of the CFS BHIS media mixture in the test cases when compared to the optimal conditions provided by fresh BHIS broth in the BHIS TC positive control (Figure 4.4).

Subsequent bile acid analysis of *C. scindens* cultured with BHIS supplemented with CA, revealed the unexpected formation of 12-keto-DCA at the 8 and 24 h time points, which may be responsible for the observed inhibition of outgrowth in the CFS +CA incubated spores.

Due to the minute difference between DCA and 12-keto-DCA at the 12α-hydroxyl group, it is possible that the observed inhibition of outgrowth in the CFS +CA conditions is due to 12-keto-DCA (Figure 4.4). The 12α-hydroxyl group is thought the be important in germinant receptor recognition and initiation of germination, with it being present in DCA and TC which have both been shown to initiate germination of *C. difficile* spores *in vitro* (Sorg & Sonenshein, 2008). This 12α-hydroxyl group is lacking in the primary bile acid CDCA and secondary bile acid LCA, both of which have been shown to inhibit the germination of *C. difficile* spores (Sorg & Sonenshein, 2009), although the inhibition of germination by CDCA is not consistently observed for all *C. difficile* isolates (Heeg & Burns *et al.,* 2012).

It should be noted that the CFSs used in these germination and outgrowth assays were obtained from cultures independent from the *in vitro* 7α-dehydroxylation and bile acid analysis experiments and so do not necessarily represent the constituents of the CFSs tested.

4.3.2 In-Vitro 7α-Dehydroxylation and Bile Acid Analysis

Bile acid analysis of samples taken at 4, 8 and 24 h showed that *in vitro*, *C. scindens* appears to metabolise CA to 12-keto-DCA, via DCA as an intermediate. These data suggest that the bile acid transformations occur during the early to mid-exponential growth phase, which is consistent with subsequent RNA-seq analysis, showing the highest levels of expression of the genes involved in 7α-dehydroxylation at the 6 h time point (Figure 4.10).

The secondary bile acid 12-keto-DCA is formed via the epimerisation of the 12α-hydroxyl group on the DCA steroid skeleton (Figure 4.14). This transformation occurs through the action of 12α-hydroxy steroid dehydrogenase (12α-HSDH) enzymes, and indicates the presence of a gene in the *C. scindens* ATCC 35704 genome encoding an enzyme with 12α-HSDH activity, although this gene remains unidentified. The *in vivo* epimerisation of the 12α hydroxyl group of bile acids by the gut microbiota is influenced by the oxidation/reduction potential of the gut environment. The formation of oxo-bile acids is more prominent in bacteria located nearer to the higher redox potential environment of the mucosal surface, compared to the lower redox potential conditions of the intestinal lumen (Ridlon *et al*., 2006).

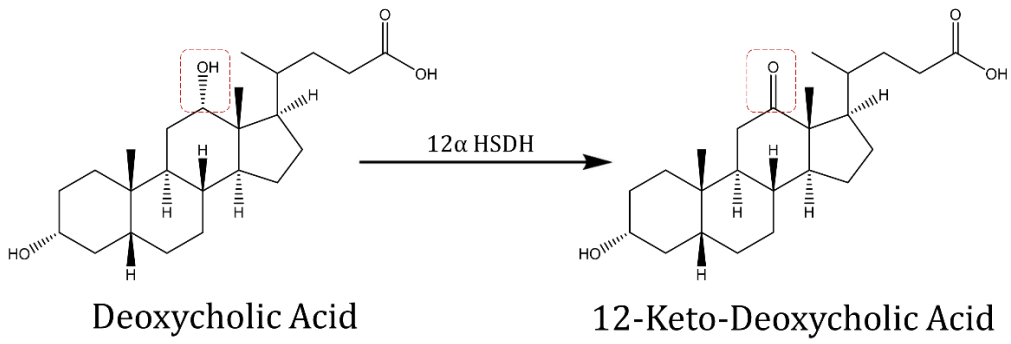


Figure 4.14 The 12α HSDH mediated transformation of DCA to 12-keto-DCA through the epimerisation of the 12α-hydroxyl group.

This transformation has been observed *in vivo* in a rat model, with 12-keto-DCA being recovered from the faeces of rats fed with either CA or DCA. It was hypothesised that 12-keto-DCA was formed directly from DCA in DCA fed rats and from CA, with DCA as an intermediate, in CA fed rats (Sakai *et al*., 1978).

Cultures for *in vitro* 7α-dehydroxylation were supplemented with 200 μM CA, which was determined to be an appropriate concentration for a consistent *C. scindens* growth profile (Figure 4.5). The discrepancies between the intended media concentrations of CA and the measurements are possibly due to the matrix effect, with the constituents of the media samples affecting bile acid quantification which can result in analyte signal enhancement, or in this case, suppression (Cappiello *et al*., 2008).

4.3.3 Expression of the Bile Acid Inducible Operon

Across all time points, the genes of the *bai* operon were significantly upregulated in CA supplemented conditions when compared to cultures without CA supplementation. These data confirm that the expression of the *bai* operon is induced by the primary bile acid, CA (Figure 4.9). Furthermore, it was observed that the genes of the *bai* operon were most highly expressed at the 6 h time point, concomitant with bile acid analyses of CA cultures, which indicates the majority of bile acid transformation occurring in early to mid-exponential phase (Figure 4.6).

Expression of the *bai* operon is under the control of two putative transcriptional regulators *barA* and *barB* (bile acid regulatory proteins A and B), located upstream of the promoter region and the *baiB* gene (Ridlon *et al.,* 2006). The complete mechanism of transcriptional regulation of the *bai* operon is currently unknown, but *barA* contains a helix-turn-helix DNA binding domain, characteristic of the Arac/XylS family of transcriptional regulators. It can therefore be hypothesised that in the absence of the

appropriate ligand, CA or CDCA, *barA* represses the transcription of the *bai* operon by directly binding to the promoter region or by mediating DNA looping, obstructing the promoter and preventing transcription. In the presence of CA or CDCA, it can be hypothesised that the bile acid ligand binds to BarA, reducing its affinity for the bound DNA sequence, releasing the DNA loop or promoter, allowing access to the promoter, permitting transcription (Gallegos *et al*., 1997; Egan, 2002).

The monocistronic *baiA1* gene was also significantly upregulated in CA induced conditions, when compared to non-induced conditions (Figure 4.9). The transcription of the *baiA1* gene appears to be regulated by the response regulator and sensor histidine kinase of a two-component system (TCS) located upstream of the promoter region. This TCS likely uses the primary bile acids CA and CDCA as an extracellular signal, which is transduced in to the cell through the autophosphorylation of the integral membrane sensor kinase. The sensor kinase in turn modifies the phosphorylation state of the response regulator, enabling binding to the *baiA1* promoter sequence, initiating transcription.

Although regulated by two separate systems, the expression of the polycistronic *bai* genes and the monocistronic *baiA1* are comparable, with the expression of *baiA1* clustering indiscriminately with the genes of the *bai* operon. This observation provides strong evidence that both systems of regulation respond to the same bile acid ligand.

4.3.5 Identification of a Novel Bile Acid Dehydratase

The *baiE* gene encodes a 19.5 kDa 7α-dehydratase and the *baiI* gene encodes a 22.3 kDa protein with putative 7β-dehydratase activity. These enzymes catalyse the removal of 7α/β hydroxyl group of 3-oxo-Δ4-bileacid CoA conjugates through the release of water; which is the fourth, rate limiting and irreversible step in the 7α/β-dehydroxylation of

primary bile acids. Both genes belong to the Nuclear Transport Factor 2 superfamily (NTF2), with the solved crystal structure of *baiE* exhibiting a twisted α+β barrel fold canonical to the NTF2 family (Figure 4.15) (Bhowmik *et al.,* 2016).

RNA-seq and differential expression analysis revealed a gene encoding a 20.7 kDa protein of the Nuclear Transport Family, significantly upregulated at all CA-induced time points, with expression levels comparable to that of the *bai* operon genes. Hierarchical clustering also revealed the gene encoding this hypothetical protein clustered in to the same dendrogram as the *bai* genes at the 6 h time point (Table 4.10). Amino acid homology indicates this protein belongs to the NTF2 superfamily. Additionally, this hypothetical protein shares the conserved SnoaL fold domain with BaiE and BaiI. The SnoaL domain, which takes its name from the polketide cyclase SnoaL protein, is characterised by a twisted α+β barrel SnoaL fold, encompassing active site (Sultana *et al*., 2004). The CA-induced expression profile and the shared domains suggests the identification of a novel 7-dehydratase in *C. scindens* ATCC 35704 involved in the rate limiting step of the oxidative arm of the pathway.

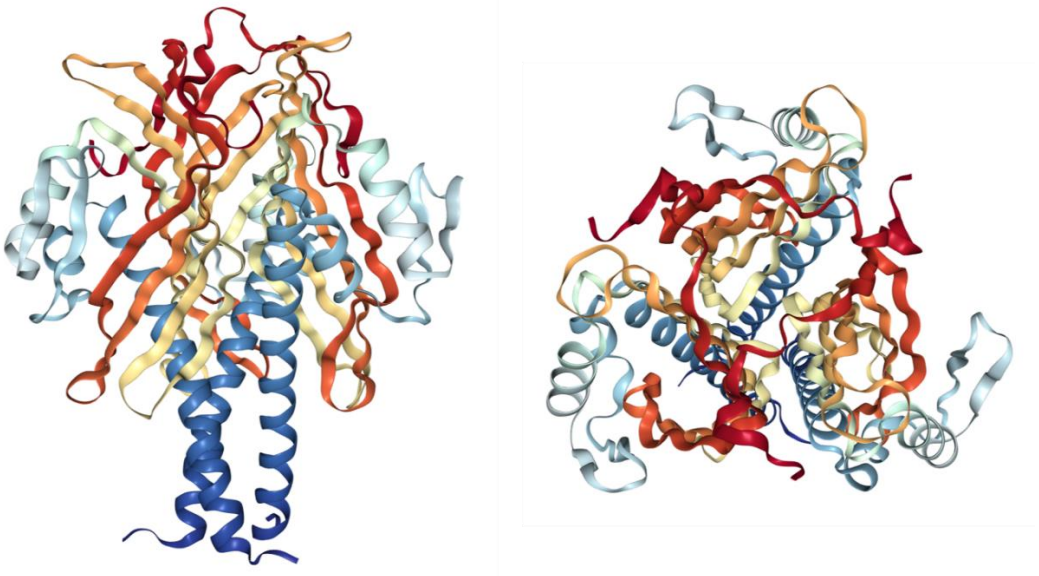


Figure 4.15 Side and Top Views of the BaiE 7α-Dehydratase of *C. scindens* ATCC 3704. The crystal structure of the BaiE homo-trimer exhibiting the active site comprised of the canonical twisted α+β barrel fold, characteristic of the NTF 2 family. The α helices shown in **blue** and the β sheets in **orange** (Bhowmik *et al.,* 2016) PDB ref: 4LEH.

4.3.4 Identification of Potential Reductive Arm Candidates

Whole transcriptomic differential expression and hierarchical clustering was utilised to aid in the discovery of the genes involved in the reductive arm of the 7α-dehydroxylation pathway. The reductive arm of the pathway is comprised of 3 reduction steps and the export of the secondary bile acid from the intracellular environment.

Potential genes for the reductive arm of the pathway were preliminarily identified by searching the annotated genome of *C. scindens* ATCC 35704 for genes with predicted NAD/FAD binding domains and reductase activity, with 15 candidates identified (Table 4.4).

A potential candidate identified in the genome annotation search, gene 1190, encoding for a 44 kDa NAD(P)/FAD-dependent oxidoreductase, was found to be marginally upregulated at the 6 hour time point, with a $\log^2$ fold change of 0.25, without statistical significance (P = 0.020895569). As opposed to CA or intermediate induction, FPKM values suggest a relatively moderate level of constitutive expression for this putative flavoprotein. Interestingly, since this study was initiated, a recently published PhD thesis identifies this same gene as a potential reductive arm candidate. The *In vitro* characterisation of the recombinant protein shows the loss of four atomic mass units from the substrate, possibly indicating the reduction of two carbon-carbon double bonds. These data have led to the hypothesis that this putative flavoprotein could be responsible for catalysing the reduction of the $\Delta^{4,6}$ double bonds on the 3-dehydro-$\Delta^{4,6}$-bile acid intermediates in the reductive arm of the pathway (Harris, 2017). Further work is underway to confirm this activity, the data in this study neither refutes nor substantiates this hypothesis.

Another of the preliminarily identified genes (FAD binding flavoprotein) was significantly upregulated in CA-induced conditions. The FAD binding flavoprotein, was significantly upregulated in response to CA-induction at all time points, with $\log_2$ fold changes of > 4. The normalised expression levels of this gene were second only to the genes of the *bai* operon at each time point sampled. Hierarchical clustering sorted the FAD binding flavoprotein in to the same dendrogram as the *bai* genes at 6 and 8 hours, the time points at which the *bai* genes appear to be the most highly expressed under CA-induction (Figure 4.10; Tables 4.10, 4.11). Amino acid sequence homology and structural alignment revealed homology to the enzyme, 3-ketosteroid dehydrogenase (3-KSD), belonging to the EC 1.3 class of oxidoreductases. Enzymes in this class specifically act upon the CH-CH groups of substrates, forming single or double bonds between two adjacent carbon atoms through dehydrogenation or reduction, utilising FAD as a hydrogen donor or acceptor. The 3-KSD enzyme is involved in steroid biotransformations and catalyses the conversion of 4-androstene-3,17-dione to 1,4-androstadiene-3,17-dione by the FAD dependent dehydrogenation of the C1 and C2 atoms of the A ring, forming a double bond (Figure 4.16) (Rhoman *et al*., 2013).

RNA-seq analysis and the homology of the amino acid sequence to a member of the EC 1.3 family of oxidoreductases suggests that this flavoprotein is a potential candidate to code for an enzyme involved in the reductive arm of the 7-dehydroxylation pathway. It can be hypothesised that the product of this gene, is a CA (or one of the pathway intermediates) induced, FAD dependent reductase enzyme, which catalyses one or both initial reduction steps in the reductive arm of the pathway, prior to 3α-reduction and export from the cell (Figure 4.17).

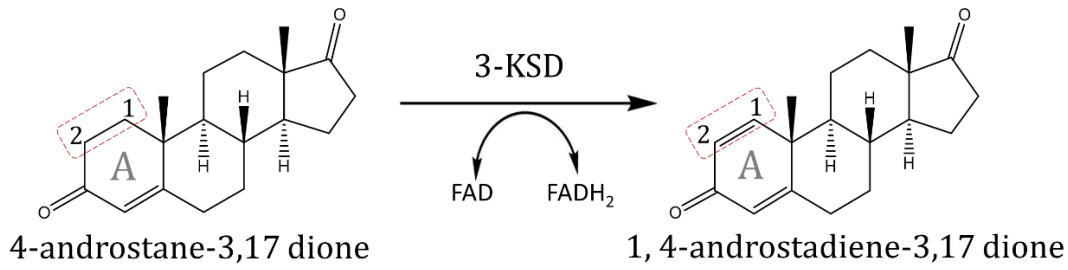


Figure 4.16 The Reaction Catalysed by 3-Ketosteroid Dehydrogenase: 3-KSD catalysing the FAD dependent removal of hydrogen atoms at the C1 and C2 atoms of the A ring, resulting in the formation of a double bond.

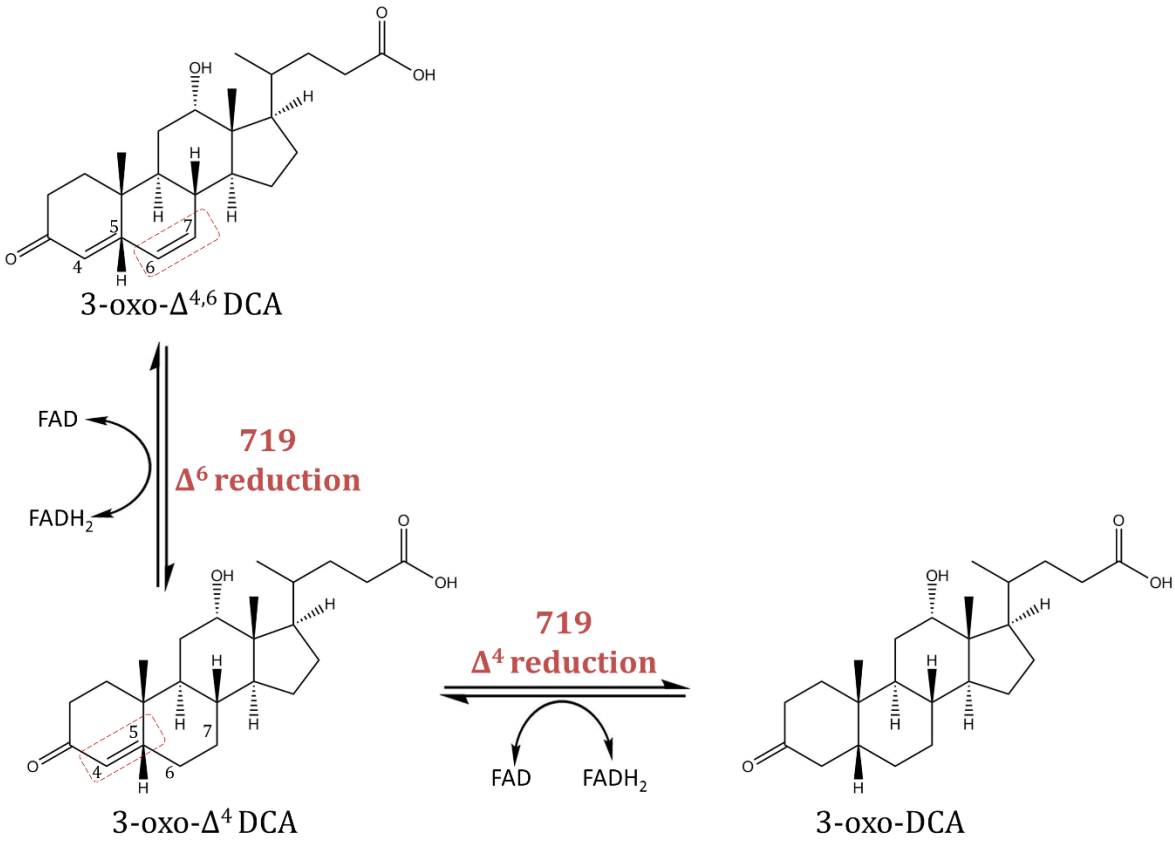


Figure 4.17 The Hypothesised Activity of the FAD binding reductase in the reductive steps of 7α-Dehydroxylation: The FAD dependent reduction of the C4,5 and C6,7 double bonds in the reduction of 3-oxo-$\Delta^{4,6}$ DCA to 3-oxo-Deoxycholic Acid.

The mechanism by which secondary bile acids are exported from the intracellular environment is currently unknown. Early work to characterise the cofactor requirements for 7α-dehydroxylation noted that neither CA nor the Δ^6 intermediate accumulated in CA-induced and non-induced cells, indicating a constitutively expressed bile acid exporter (White *et al*., 1981). Converse to this observation, the RNA-seq analysis performed in this study reveals a number of membrane transporters significantly upregulated in response to CA induction. The most significantly upregulated transporters represent 3 membrane transporters which share sequence homology with the large Major Facilitator Superfamily (MFS) with conserved domains belonging to the EmrB/QacA subfamily of drug efflux proteins. The H^+ dependent primary bile acid transporter BaiG also belongs to the MFS and shares homology with the EmrB/QacA subfamily. These genes could therefore be hypothesised to code for the unknown secondary bile acid exporter.

The RNA-seq analysis performed in this study offers an insight in to potential unidentified genes involved in the 7α-dehydroxylation of CA by *C. scindens* ATCC 35704. A novel putative 7-dehydratase enzyme and an enzyme belonging to the EC 1.3 oxidoreductase family were significantly upregulated under CA-induced conditions. There were also 3 membrane transporters belonging to the MFS group of transporters which were significantly upregulated in response to CA. These identified genes represent potential for further study to categorise the specific activity of the proteins coded for to elucidate the genes responsible for the reductive arm of the 7α-dehydroxylation pathway.

Chapter 5 *Clostridium scindens* Molecular Characterisation and DNA Transfer

5.1 Introduction

5.1.1 Bacterial Transformation

In 1928, English bacteriologist Frederick Griffith set out to investigate the virulence of two Group IV *Streptococcus pneumoniae* strains in a mouse model. The two strains of study were morphologically and pathologically distinguishable. The strain designated "strain R" formed colonies with rough edges on solid media and did not cause disease in the mouse, whereas the strain designated "strain S" exhibited a smooth colony morphology and was lethal in the mouse model. Griffith discovered that a mixture of heat killed virulent strain S and live culture of avirulent strain R was lethal to the mouse, and bacteria recovered from the mouse exhibited the previously observed smooth colony morphology. Griffith termed this phenomena transformation, concluding that the transference of a transforming principle is responsible for the transformation of the R strain in to the virulent S strain (Griffith, 1928).

Since the discovery of bacterial transformation and subsequent the elucidation of transformation by exogenous DNA, bacterial transformation has become a widely utilised tool in the fields of molecular biology and biotechnology, aiding in the study and exploitation of microorganisms for production of beneficial and industrial relevant compounds.

5.1.2 Mechanisms of Horizontal Gene Transfer

In nature, horizontal gene transfer may occur through natural transformation, bacteriophage transduction or conjugal transfer. Each of these processes involves the transfer of DNA from a donor to a recipient cell, the heritable chromosomal integration

(or plasmid maintenance) and functional expression of acquired genes. There are three non-mutually exclusive benefits of horizontal DNA transfer in bacteria; genetic diversity, nutrition and DNA repair:

1. Transferred DNA can be used to increase a host's genetic diversity through the acquisition of genes conferring antibiotic resistance, virulence factors or novel metabolic functions.

2. Exogenous DNA can also be used as a source of carbon, phosphorous and nitrogen, with these DNA degradation products made available for cellular metabolism, or nucleotides for DNA synthesis.

3. Incoming homologous DNA from closely related bacteria may be utilised for the recombinational repair of genotoxic stress lesions in host DNA.

The probabilities of the occurrence of each of these gene transfer events in their natural environments is based upon the specific conditions and requirements for each mechanism.

5.1.3 Horizontal Gene Transfer by Natural Transformation

Natural transformation is the most frequent form of horizontal gene transfer in nature, as it does not require a live donor and the persistence and dissemination of extracellular DNA enables spatial and temporal separation between the donor and host cells. Griffiths Streptococcal virulence experiments are an example of natural transformation.

The recipient cell must be naturally competent, possessing a system for exogenous DNA uptake and importantly the recipient and donor needn't be genetically related. Natural competence is dependent upon the expression of 20 – 30 associated genes, in a time limited manner of competence in response to nutrient availability, population dynamics

quorum sensing or starvation (Thomas & Nielsen, 2005). Extracellular DNA, released from lysed or disrupted cells, or viral particles, is transported in to the cytosol of the host bacterium in a complex process (Figure 5.1 c).

The process of DNA uptake is defined as the change in state of unprotected extracellular DNA, to protected by crossing the outer membrane of Gram-negative bacteria into the periplasm or crossing the cytoplasmic membrane and in to the cytosol of Gram-positive bacteria. Both Gram-positive and Gram-negative bacteria use type IV pili and type II secretion system related proteins of a common competence system. These proteins form a cell envelope spanning pseudopilus assembly, to which extracellular DNA binds in sequence specific and non-specific manners, which is then transported in to the cytosol by the retraction of the pseudopilus (Chen & Dubnau, 2003). Only single stranded DNA enters the cytosol, with the non-transforming strand degraded and released in to the extracellular environment (Dubnau, 1971). Once in the cytosol, internalised single-stranded DNA is integrated in to the chromosome in a homology dependent process by the bacterial recombination protein, RecA. Reconstitution and stable maintenance of linearized single stranded plasmids can also occur (Chen & Dubnau, 2004).

5.1.4 Horizontal Gene Transfer by Transduction

Horizontal gene transfer by transduction requires a metabolically active donor cell, within which bacteriophage particles are propagated and released in to the extracellular environment. Bacteriophage transduction can occur with spatial and temporal separation between the donor and recipient cells due to the persistence of genetic material within the viral particle. The main requirement for transduction by bacteriophage particles is for the donor and recipient cells to share a common sensitivity to the transducing bacteriophage (Figure 5.1 a).

Bacteriophages are important vectors of horizontal gene transfer between bacterial species, having a significant impact on microbial evolution. Successful bacteriophage infection of a bacterium may result in two outcomes, lysis or lysogeny. The phage lytic lifecycle involves the multiplication and subsequent lysis of the bacterial host, releasing progeny bacteriophage in to the extracellular environment. The lysogenic lifecycle results in the integration of prophage DNA in to the bacterial genome, which can often result in the host bacterium acquiring extra genes, termed lysogenic conversion genes (LCG). The tailed bacteriophages of the *Siphoviridae* (λ), *Myoviridae* (T4) and *Podovididae* (T7) are characterised by relatively large double stranded DNA genomes and a propensity to initiate the lysogenic lifecycle, facilitating evolutionarily successful horizontal gene transfer events (Rokyta *et al.,* 2005).

Horizontal gene transfer by bacteriophages may either by generalised or specialised. Generalised transduction is the erroneous process by which a bacteriophage particle is packaged with bacterial DNA and little or no phage DNA, for example the *Salmonella* phage P22, which will then integrate in to the chromosome of the ensuing bacterial host, (Ebel-Tsipis *et al.,* 1972). Specialised transduction involves the imprecise excision of prophage DNA from the bacterial chromosome which can result in flanking bacterial DNA being packaged in to the virion, in a process analogous to Hfr (high frequency of recombination) strain chromosomal gene transfer. This process is frequently associated with the acquisition of virulence factors, such as toxin encoding genes, for example the *Staphylococcal* φSa3 (Baba *et al.,* 2002) and the shiga toxin encoding prophages of *E. coli* O157:H7, the expression and export of which depends upon prophage induction (Gamage *et al.,* 2004).

5.1.5 Horizontal Gene Transfer by Conjugation

Conjugation is the least probable horizontal gene transfer event occurring in nature, with this process requiring the presence of plasmid or transposon conjugative elements and the establishment of a sufficiently stable physical contact between the donor and recipient cells. Although plasmids may be acquired through the DNA uptake processes associated with natural transformation, plasmid acquisition is more specifically associated with conjugal transfer (Figure 5.1 b).

The donor cell contains a fertility factor (F) and is termed F^+, which is transferred to the F^- recipient cell. The F factor is usually contained on a self-replicating episome, or plasmid, but can also be integrated in to the chromosome of donor cells, designated high frequency of recombination (Hfr) strains. The integrated F factor of Hfr strains may excise from the chromosome in an aberrant recombination event, mobilizing small sections of the adjacent chromosomes.

The F factor comprises all the functional elements required for conjugal transfer, including the *tra* operon, which encodes for more than 20 genes responsible for the transfer for DNA and pili assembly involved in conjugation. The F plasmid, also possess a *cis*-acting *oriT* locus, where a strand specific nick is introduced enabling the 5' to 3' transfer of DNA to the recipient bacterium at the interbacterial junction. The F plasmid has been established as the paradigm for conjugative gene transfer, with the closely related R plasmids, more specifically associated with the transfer of resistance determinant genes (Fu *et al.*, 1991).

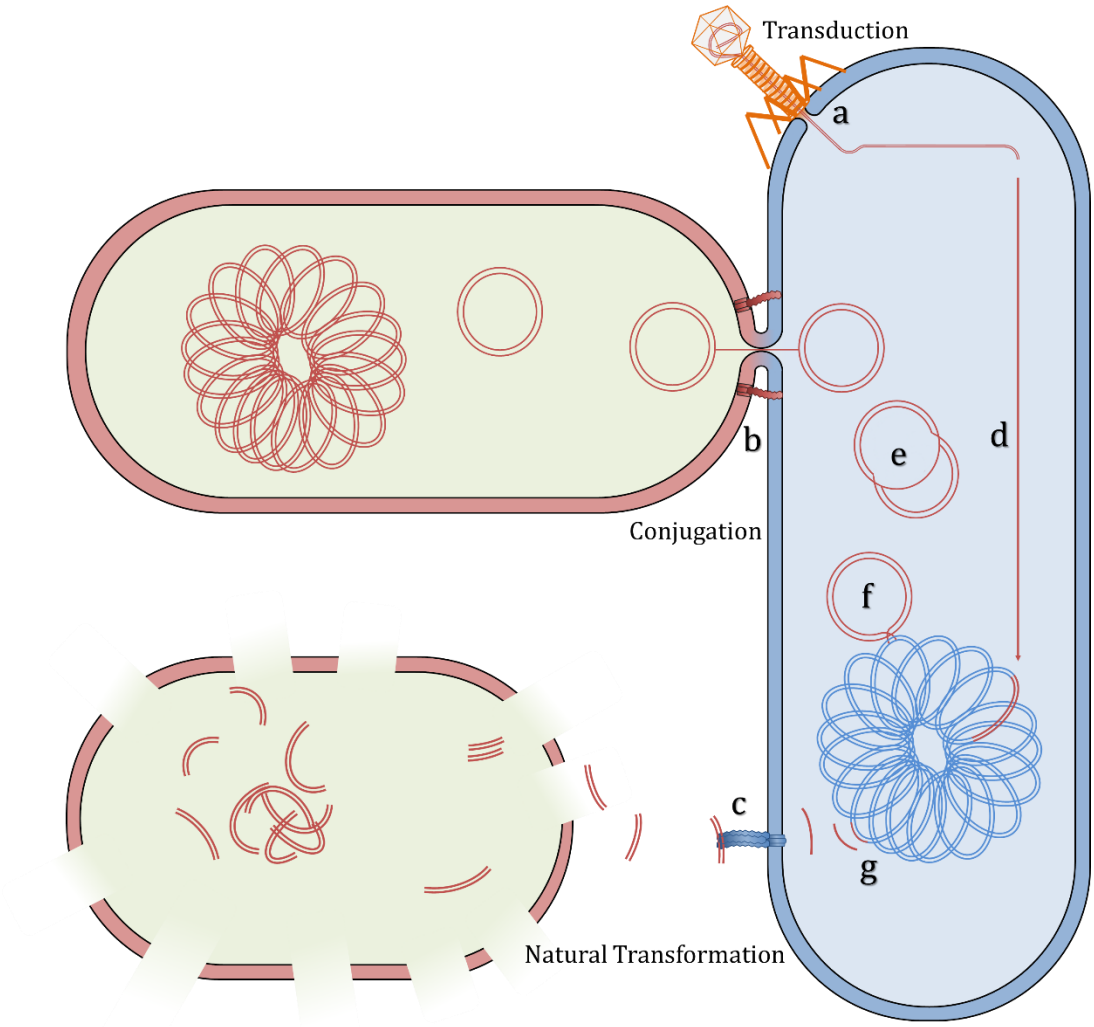


Figure 5.1 Mechanisms of Horizontal Gene Transfer: **a**: Bacteriophage adsorption and DNA injection; **b**: Conjugal plasmid transfer between donor and recipient bacteria by rolling-circle replication at the mating-pair apparatus; **c**: Pili mediated uptake of extracellular DNA; **d**: Integration of bacteriophage DNA into the host chromosome **e**: Replication and expression of exogenous plasmid DNA; **f**: Homologous recombination between foreign plasmid and host chromosomal DNA; **g**: Integration of exogenous DNA in to the host chromosome.

5.1.6 Barriers To Gene Transfer

The major barriers to horizontal gene transfer by natural transformation are related to the availability of adaptive DNA in the environment, specificity of host DNA uptake systems, lack of foreign-host DNA homology and bacterial competence. For horizontal transfer by transduction and conjugation, there are a number of strain and species-specific mechanisms that act as barriers to the transfer, uptake and stabilisation of foreign DNA (Figure 5.2)

The *traS* and *traT* genes located on the *tra* operon of conjugative plasmids form an effective barrier to conjugation known as surface exclusion (Figure 5.2 d). The *traT* gene encodes an outer membrane lipoprotein which associates with peptidoglycan and alters the outer surface of the cell, reducing receptivity to the F pilus and blocking conjugal transfer by limiting the stability of the interbacterial junction. The *traS* gene encodes an inner membrane protein, which is thought to inhibit conjugal DNA metabolism (Jalajakumari *et al.,* 1987). It is thought that surface exclusion ensures that conjugal transfer of closely related plasmids does not occur but could also serve as a mechanism to promote the breakdown of mating pairs once successful transfer has occurred.

The discovery of clustered regularly interspaced palindromic repeat (CRISPR) loci in the vicinity of CRISPR associated (Cas) genes, lead to the elucidation of an adaptive immune system which protects bacteria from the foreign DNA of conjugative plasmids and bacteriophages (Mojica *et al.,* 2005; Pourcel *et al.,* 2005). The CRISPR loci, comprised of roughly palindromic, direct repeats separated by spacer sequences, is separated from the *cas* operon by an AT rich leader sequence and encodes the instructions for this adaptive immunity. Spacers are derived from exogenous DNA, acquired in a polarised manner at the anterior end of the CRISPR loci, the transcription of which and post processing results in individual CRISPR RNA (crRNA) containing sequences specific for

invading DNA. The crRNAs guide effector complexes, comprised of Cas proteins, to cleave the complementary DNA downstream of a protospacer adjacent motif (PAM) on the invading DNA. The nucleolytic activity of the Cas complex inactivates the exogenous DNA, promoting recombinational isolation and preventing cell lysis by bacteriophage attack (Figure 5.2 a)

Restriction modification (RM) systems also represent another form of a primitive immunity against foreign DNA. Just as CRISPR-Cas systems can be thought of as adaptive immunity, RM systems are analogous to an innate immune system. RM systems recognise specific chemical methylation signatures enabling the discrimination between "self" and "non-self" DNA, which is cleaved by nuclease activity upon recognition of specific sites by restriction endonucleases (REase). There are two types of methylation in bacterial RM systems, N6-methyladenine (6mA) and C5-methylcytosine (5mC), the methylation of adenine and cytosine residues, respectively.

Plasmids transferred by conjugation without the correct "self" methylation pattern are restricted by host REs at specific sequences on the re-constituted double stranded DNA (dsDNA) plasmid. Single stranded DNA (ssDNA) internalised through natural transformation is effectively immune to restriction, as REs only restrict dsDNA. Homologous ssDNA can be integrated in to the chromosome through RecA mediated homologous recombination events, with the resolution of the heteroduplex resulting in hemimethylated chromosomal DNA, resistant to restriction. This process becomes problematic for the cell when heterologous DNA flanked by regions of homology is incorporated in to the chromosome. The replication dependent resolution of this heteroduplex results in a segment of unmethylated chromosomal DNA sensitive to restriction by host REs, resulting in chromosome restriction and cell death. In this manner, classical RM systems limit the acquisition of exogenous DNA by natural transformation, preserving the genetic integrity of the species.

There are four different types of RM systems in bacteria, with their classification based upon the subunit composition, sequence recognition, cleavage position and substrate specificity (Figure 5.3) (Roberts *et al.,* 2003).

Type I systems are the most sophisticated RM systems, comprised of a hetero-oligomeric protein complex capable of both methylation and restriction of DNA. The type I complex is a molecular machine, consisting of proteins encoded for by host specificity determinant (*hsd*) genes. The *hsdR*, *hsdM* and *hsdS* genes encode for restriction (R), modification (M) and specificity (S) subunits, which form the enzyme complex, which is comprised of two M, two R and one S subunit. The M subunit contains a methyltransferase catalytic domain, responsible for the modification of target, hemimethylated or newly replicated host DNA. The S subunit has two variable regions which recognise bipartite target sequences, separated by a number of non-specific nucleotides, for example the 5'-AACNNNNNNGTGC-3' recognition sequences of EcoKI (Studier & Bandyopadhyay, 1988). Cleavage of DNA is carried out by the R subunit which is comprised of a nucleolytic and ATP hydrolysing translocation domain. The translocation domain of the R subunit is a molecular motor, which enables the bidirectional reeling in of DNA base pairs towards the enzyme complex, at speeds of up to 1000 bp per second (Seidel *et al.,* 2004). This translocation of DNA results in the extrusion of DNA loops, with cleavage occurring indiscriminately, away from the target recognition sequence, when DNA translocation stalls by colliding with another translocating enzyme complex or due to DNA topology (Kennaway *et al.,* 2012)

The type II enzymes are a broad group of unrelated enzymes, classified by enzymatic behaviour as opposed to evolutionary relatedness, divided in to the 11 subgroups; A, B, C, E, F, G, H, M, P, S and T signifying a specific property (Roberts *et al.,* 2003). These systems are the most prevalent and simplest of the RM systems, comprised of separate and independently acting MTase and REase components, which modify and cleave DNA in to discrete fragments at a fixed position relative to the recognition site. Type II

restriction enzymes generally cleave DNA in a homodimer enzyme complex, with the MTase enzyme monomer acting independently. Due to their ability to restrict discrete fragments of DNA, type II restriction enzymes are frequently used for cloning applications in molecular biology.

Type III systems can be thought of as an intermediate between the type I and II systems, comprised of MTase and REase components, forming a heterodimeric complex. Enzyme complexes work in pairs, each recognising specific sites on the DNA molecule. Translocation and the formation of DNA loops occurs as with the type I enzymes with cleavage occurring as a result of enzyme collisions, but unlike the type I systems, type III cleavage is at a defined point, close to the recognition sequence (Crampton *et al.,* 2007). Cleavage by type III enzymes is not complete, due to competition between the methylation and restriction activities of the heterodimer.

The novel type IV class was recently proposed when the structure and enzymatic properties of the enzyme RM.Eco571 were found to distinguish it from all other described classes of RM system (Janulaitis *et al.,* 1992). Type IV systems are not classical RM systems, as they are only comprised of a REase component which recognises DNA modified by methylation, hydroxymethylation or glucosyl-hydroxymethylation. Type IV system serve to maintain the methylation status of the host genome by cleaving the chromosome, initiating cell death when a foreign MTase is expressed in the cell (Fukuda *et al.,* 2008). However, cleavage of self chromosomal DNA has also been observed under stress conditions, suggesting a role in addition to the RM cellular defence paradigm (Aertsen & Michiels, 2005)

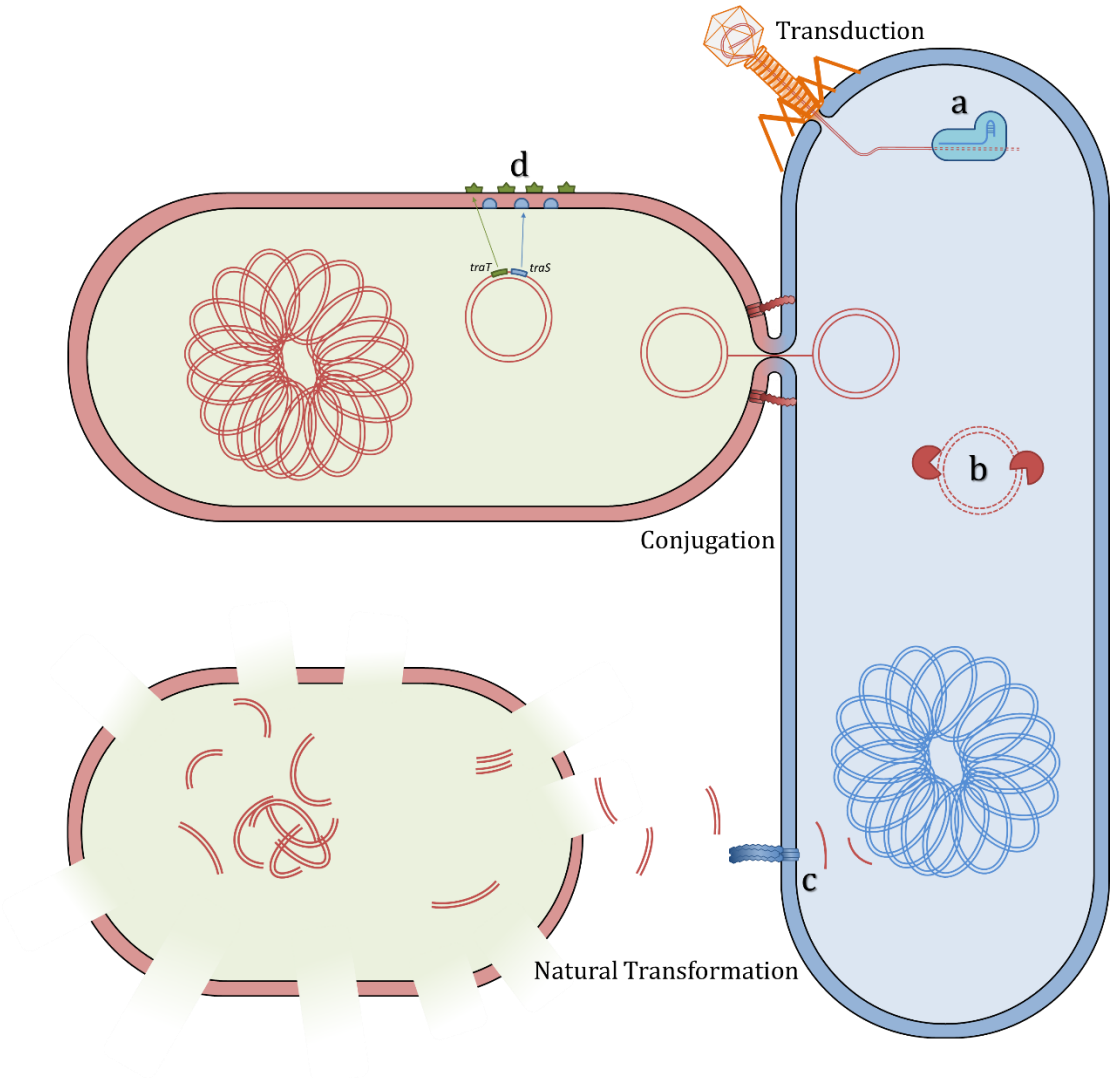


Figure 5.2 Barriers to Gene Transfer: **a**: Inactivation of foreign DNA by CRISPR Cas complex; **b**: Digestion of non-host methylated DNA by restriction endonucleases; **c**: Specificity of uptake apparatus and lack of homology between foreign DNA and the host chromosome; **d**: Surface exclusion mediated by *traS* and *traT*.

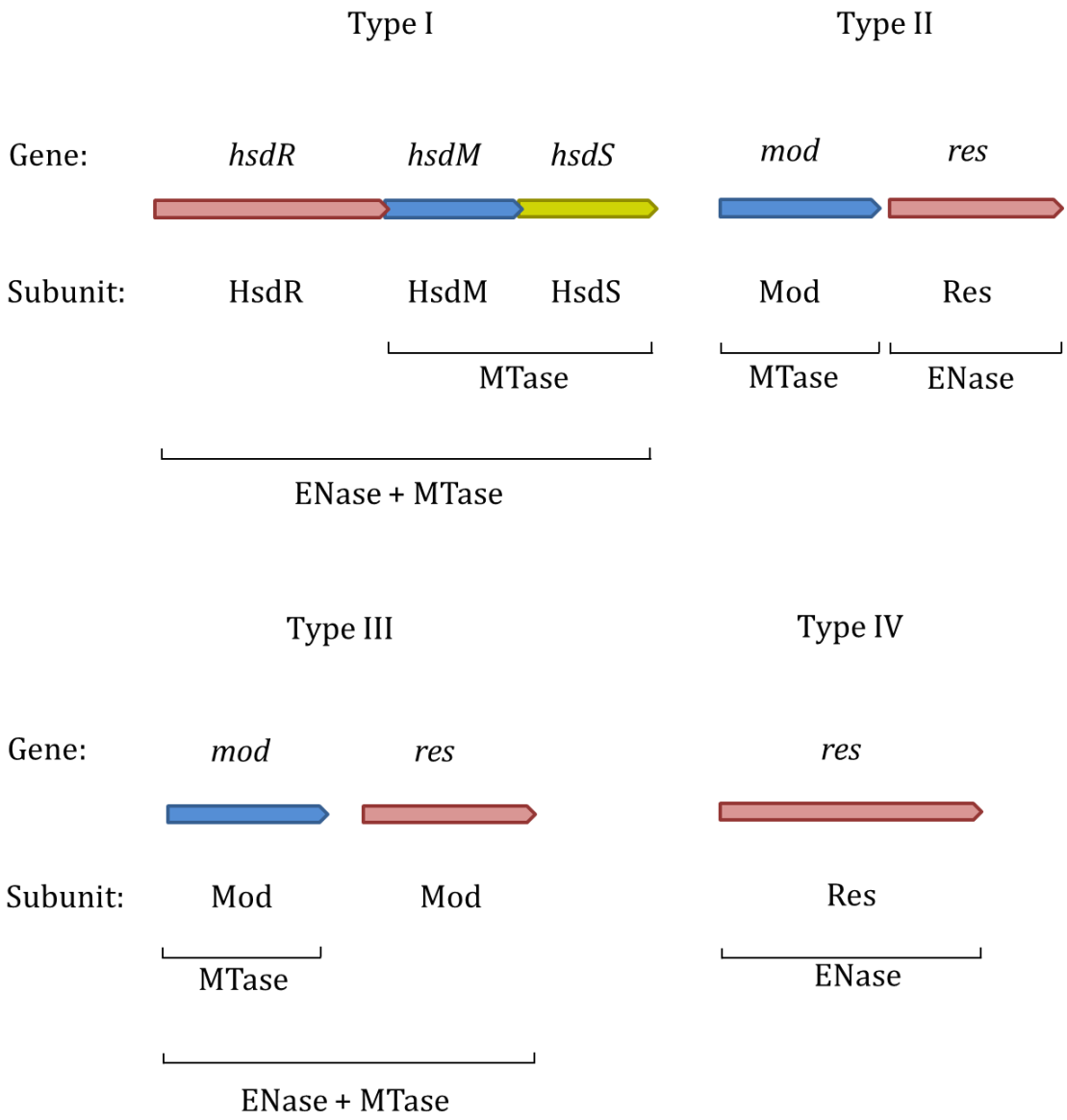


Figure 5.3 The Four Types of Bacterial Restriction Modification Systems: The hetero-oligomeric type I system consisting of restriction, methyltransferase and specificity subunits; the monomeric type II system, the heterodimer type III systems and the non-classical monomeric type IV system.

5.1.7 Aims and Hypotheses

The aim of this study is to establish DNA transfer in *C. scindens* through transformation and conjugation methodologies. Potential barriers will be identified by characterisation of the methylation status of the *C. scindens* ATCC 35704 strain through SMRT sequencing base modification and genome sequence analysis. A range of strains and conditions will be tested in order to establish the optimal conditions for DNA transfer.

It is hypothesised that the barriers to DNA transfer are strain specific and therefore a variety of *C. scindens* strains will be tested, with the thought that one of these strains may be more amenable to accepting foreign DNA, enabling transfer. The shuttle vectors used in these studies employ a number of different Gram-positive replicons to enable the replication and maintenance of transferred plasmids in the recipient *C. scindens* strains.

Conjugation will be attempted using a variety of *E. coli* donor strains, with different host methylation profiles, with the hypothesis that one of these methylation profiles may enable the circumvention of the endogenous RM systems of *C. scindens* ATCC 35704.

Transformation will be attempted utilising *in vitro* methylation with a GpC methyltransferase (NEB), which may provide protection against the *C. scindens* host RM systems.

The base modification data and genome sequence analysis will allow the construction of conjugal *E. coli* donors expressing methyltransferase genes from *C. scindens* strain ATCC 35704. It is hypothesised that this will enable the mimicry of the *C. scindens* methylation status and thus enable DNA transfer.

5.2 Results

5.2.1 Initial Attempts to Establish DNA Transfer by Conjugation

To establish gene transfer in *C. scindens*, initial attempts were based upon conjugation protocols developed for *C. difficile* (2.3.8). Three *E. coli* conjugal donor strains were tested; CA434, S17-1 and SExpress. CA434 is a HB101 derivative carrying the R702 conjugal R factor plasmid, S17-1 is an Hfr strain and SExpress is a derivative of the NEB Express cloning strain harbouring the R702 conjugal R factor plasmid (Woods, 2017 unpublished). As it is unclear which Gram-positive replicon would be suitable for plasmid stability in *C. scindens*, separate conjugation attempts were carried with shuttle vectors carrying four different modular Gram-positive replicons (Heap *et al.,* 2009)

Preliminary antibiotic testing indicated that *C. scindens* is susceptible to the antibiotic thiamphenicol at a concentration of 15 μg/ml and therefore the chloramphenicol acetyltransferase gene, *catP*, was used as a marker on the plasmids to be transferred. Conjugal mixtures were plated on to BHIS agar supplemented with D-cycloserine at 250 μg/mL and chloramphenicol at 15 μg/mL.

A selection of eight *C. scindens* strains were tested for DNA transfer; the type strain, ATCC 35704, the VPI 12708 strain and six clinical isolates provided by collaborators at Virginia Commonwealth University. Heat treatment of the recipient strains at 44 ^{o}C for 15 minutes prior to mating was also attempted for the ATCC 35704 and VPI 12708 strains to temporarily inactivate any RM systems which may be providing a barrier to DNA transfer (Kirk *et al.,* 2016).

All observed putative transconjugants were subjected to a *C. scindens* PCR screen using primers specific for the *baiCD* gene (2.10).

All attempts to establish DNA transfer in to *C. scindens* by conjugation under these conditions proved to be fruitless. Initial conjugation assays with the ATCC 35704 and VPI 12708 strains were repeated 3 times, with all subsequent assays repeated twice. The conjugation assays attempted are summarised in table 5.1.

Table 5.1 *C. scindens* Conjugation Assay Conditions

	Conjugation Attempts							
***E. coli* Donor Strains**	CA434 S17-1 sExpress	CA434 S17-1 sExpress	CA434 S17-1	CA434 S17-1	CA434 S17-1	CA434 S17-1	CA434 S17-1	CA434 S17-1
***C. scindens* Recipient Strain**	ATCC 35704	VPI 12708	SA14	S076	S077	KS11	G11	C132
Donor Heat Shock	✓	✓						
Gram +ve Replicon	pBP1 pCB102 pCD6 pIM13	pBP1 pCB102 pCD6 pIM13	pBP1 pCB102 pCD6 pIM13	pBP1 pCB102 pCD6 pIM13	pBP1 pCB102 pCD6 pIM13	pBP1 pCB102 pCD6 pIM13	pBP1 pCB102 pCD6 pIM13	pBP1 pCB102 pCD6 pIM13
Selective Marker	*catP*	*catP*	*catP*	*catP*	*catP*	*catP*	*catP*	*catP*
Mating Times (h)	4 8 12 24	4 8 12 24	4 8 12 24	4 8 12 24	4 8 12 24	4 8 12 24	4 8 12 24	4 8 12 24

5.2.2 Initial Attempts to Establish DNA Transfer by Transformation

Due to the repeated failure of conjugation, transformation by electroporation was attempted. Electrocompetent *C. scindens* cells were prepared as described for *E. coli* (2.3.6), except for under anaerobic conditions, using an electroporation wash buffer optimised for Gram-positive organisms (2.3.7). Cells were assessed for viability prior to electroporation by streaking competent cell preparations on to non-selective media.

A cocktail of plasmids with the four Gram-positive replicons was electroporated in to electrocompetent *C. scindens* ATCC 35704 cells using a BioRad MicroPulser at a range of settings, detailed in table 5.2. Due to the failure in attaining DNA transfer through exhaustive conjugation experiments, it was suspected that one or more RM systems may be a barrier to transfer, and therefore plasmids to be transformed by electroporation were also subjected to *in vitro* methylation using GpC MTase, conferring protection to any cytosine-based RM system (NEB) (2.5.13)

As with the conjugation, all electroporation experiments were also unsuccessful.

5.2.3 Features of the C. scindens *ATCC 35704 Genome*

To aid in establishing gene transfer, the genome of *C. scindens* strain ATCC 35704 was analysed for features and any potential barriers to gene transfer (Figure 5.4)

The polycistronic bile acid inducible operon is located on the positive strand, with the monocistronic *baiA1* gene located on the same strand downstream. The gene encoding the putative FAD binding reductase identified in RNA-seq experiments, is located on the negative strand and the gene encoding the putative novel 7-dehydratase is located further downstream on the negative strand.

The cortisol inducible *desABCD* operon is located on the positive strand , comprised of 4 genes encoding the transketolases DesA and DesB, the 20α-HSDH DesC and the cortisol transporter DesD.

A Type IC CRISPR associated (Cas) operon is located on the positive strand consisting of a *cas3* helicase, *cas5c* endonuclease, *cas8c* PAM recognition factor, *cas7*, *cas4* exonuclease and the ubiquitous *cas1* and *cas2* nucleases.

A 33.2 Kb prophage, designated ϕClsc35704, is located on the negative strand. The phage genome encompasses a predicted 49 coding sequences, with a GC content of 44.92 %. The synteny of the head and tail genes, with the head located 5' to the tail genes, indicates that ϕClsc35704 has a Siphoviral morphotype. This prophage shares the most sequence homology with the phage ϕCT453B of *Clostridium botulinum*, which belongs to the dsDNA Caudovirales order of the Siphoviridae family of bacteriophages.

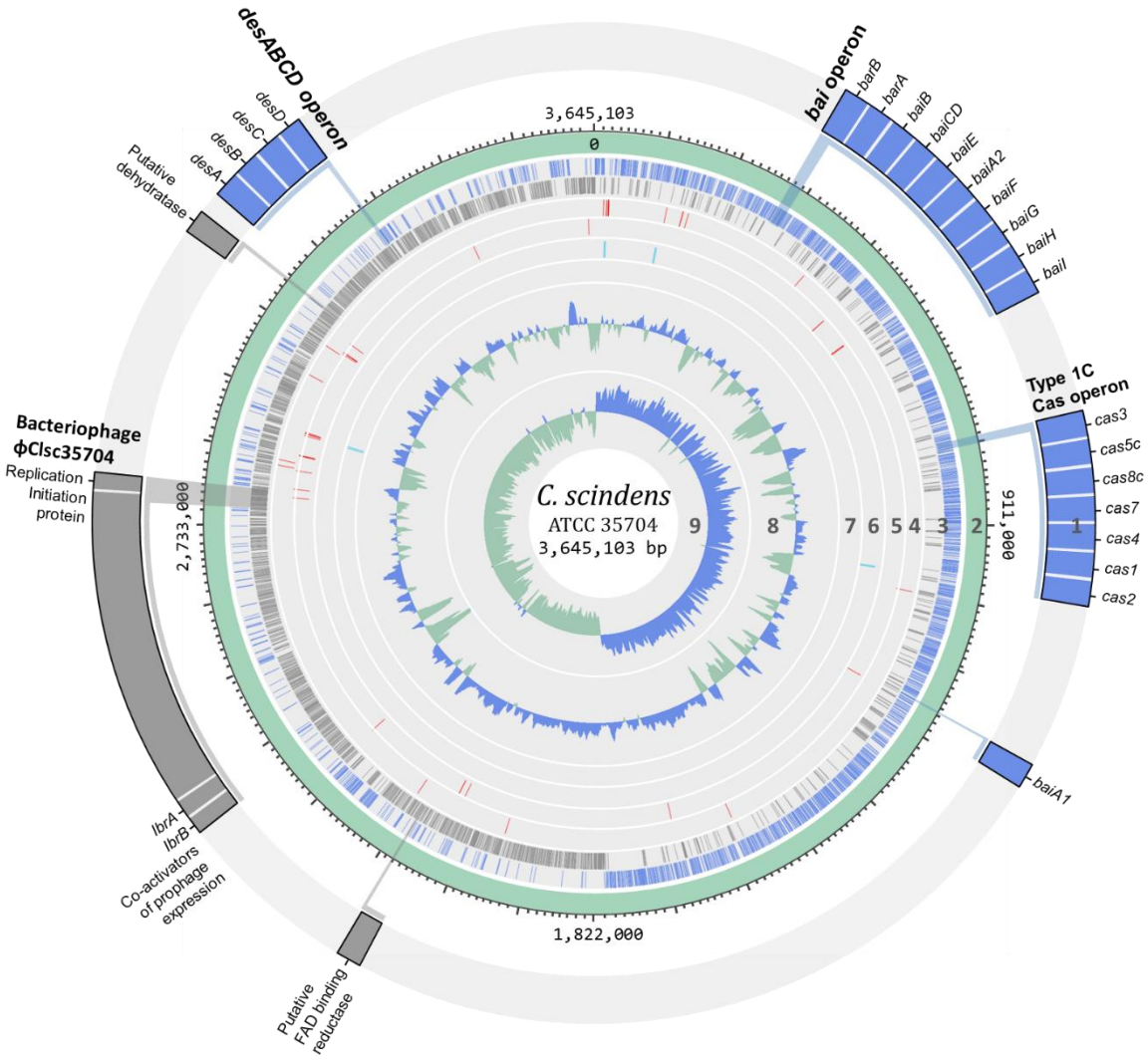


Figure 5.4 The Genome Features of *C. scindens* ATCC 35704: Track **1**: Genome features of interest; the *bai* genes, putative FAD binding reductase, putative dehydratase, the *desABCD* operon, CRISPR operon and bacteriophage φCsc35704; Track **2**: source sequence; Track **3**: Coding sequences on the **positive** and **negative** strand; Track **4**: positive strand tRNA CDS; Track **5**: negative strand tRNA CDS; Track **6**: positive strand rRNA CDS; Track **7**: negative strand rRNA CDS; Tracks **8** and **9**: G+C content and G+C skew, (**blue** above average GC, **green** below average GC).

5.2.4 The CRISPR Cas System of C. scindens *ATCC 35704*

The genome of *C. scindens* ATCC 35704 was searched for clustered regularly interspaced short palindromic repeat (CRISPR) sequences, with three confirmed and one questionable CRISPR loci identified using the CRISPR finder webtool (available at http://crispr.i2bc.paris-saclay.fr/) (Figure 5.5).

CRISPR locus number 1, is comprised of 9 spacers separated by the 33 bp direct repeat (DR) sequence 5'-GTCGTTCCCTGCAATGGGAACGTGGATTGAAAT-3', with a total size of 636 bp.

CRISPR locus number 3 is comprised of 13 spacers, separated by the 32 bp DR sequence 5'-CTTTCAATCCACTCCACCACACAGGTGGAGAC-3', with a total size of 869 bp.

CRISPR locus number 2 is comprised of 39 spacers, separated by the 33 bp DR sequence 5'-ATTTCAATCCACTCTCCCTATACAGGGAGAGAC-3'.. Locus 2 is adjacent to the cas operon with a size of 2,653 bp, separated by a ~185 bp AT rich leader sequence.

There is also a questionable CRISPR locus comprised of one spacer, separated by the 28 bp DR sequence 5'-AGAAGCAGGATGATCCAGATACAGCTGT-3'. This locus is determined to be questionable due to the relatively low number of DR sequences, flanking one spacer sequence.

To ascertain whether the CRISPR Cas system of *C. scindens* ATCC 35704 is a potential barrier to DNA transfer, homology between the identified spacers and the sequences of the clostridial shuttle vectors utilised in this study was assessed using the NCBI BLAST utility. No homology was found to the pMTL series vectors used in attempted gene transfer assays. Spacer homology was observed for bacteriophage elements and segments of DNA from bacterial genera inhabiting the gut.

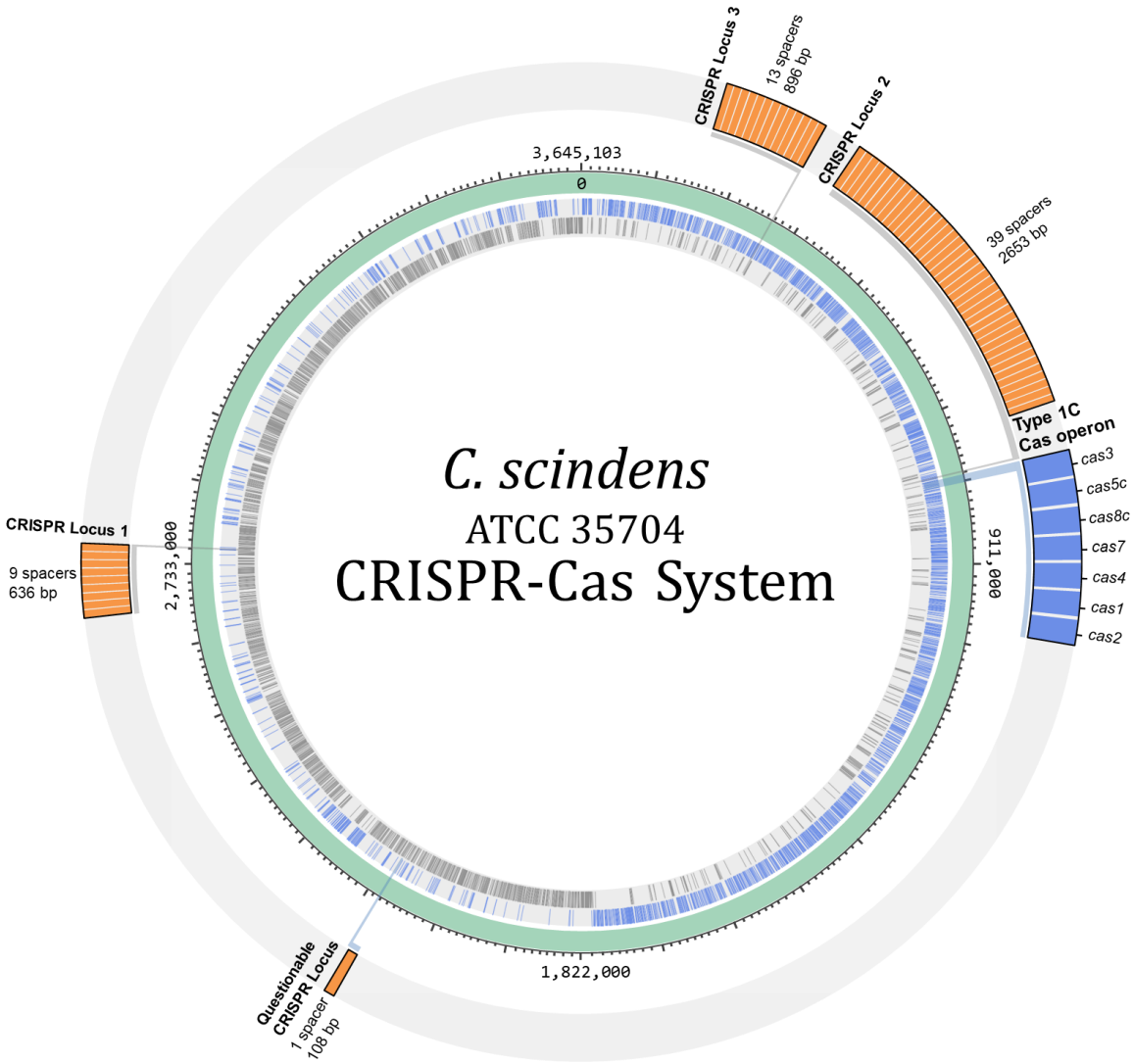


Figure 5.5 The CRISPR Loci and Cas operon of *C. scindens* ATCC 35704: The CRISPR loci are shown on the outer track, along with the *cas* operon. **Locus 1** comprised of 9 spacers, **locus 2**; 39 spacers and **locus 3**; 13 spacers. The questionable CRISPR locus is comprised of 1 spacer.

5.2.5 Methylome Analysis of C. scindens *ATCC 35704*

A total of 17 loci predicted to encode two type I RM systems, three complete and 11 incomplete type II RM systems, one type IV RM system and three apparent orphan MTase genes were identified on the genome sequence of *C. scindens* ATCC 35704 (Figures 5.6 and 5.7). Out of the 17 systems identified, only the complete RM systems with cognate REase components were characterised further, as potential barriers to DNA transfer, highlighted in figure 5.7.

To characterise putative RM systems, sequences were subjected to non-redundant BLAST searches of the NCBI database and New England Biolabs REBASE database (Roberts *et al.,* 2015).

PacBio SMRT sequencing was utilised to identify base modifications and to determine if the identified MTase genes were responsible for the observed methylation. The detected methylation patterns are summarised in table 5.1, along with the number detected and the number expected, based upon the genome sequence. The objective score is based upon the motif finding algorithm, with higher scores indicative of a higher confidence in the detected motif (Table 5.3).

Table 5.3 Methylation Patterns Detected by SMRT Sequencing

Sequence Motif	Type	N Detected	N Genome	Objective Score
5′-CNRT^{6mA}ANNNNNNNT^{6mA}CNNC-3′	6mA	1124	1124	123023
5′-GA^{6mA}AGCTT^{6mA}C-3′	6mA	24	146	186.48
5′-GGAT^{6mA}C-3′	6mA	6462	6468	749611.6
5′-CA^{6mA}T^{6mA}CGC-3′	6mA	470	1028	182510.22

The target recognition domain (TRD) of the specificity subunit of the type I RM system Csc35704ORFIP shares a 47 % identity to the specificity subunit of the type I system Smy11118I of *Sporocytophaga myxococcoides*, which has a recognition sequence motif of AAGANNNNNNNTAYNG. The detected 5'-CNRT^{6mA}ANNNNNNNT^{6mA}CNNC-3' motif was therefore assigned to Csc35704ORFIP. The other type I RM system, Csc35704ORFM, contains a frameshift mutation in the MTase component and with no other type I-like base modification motifs detected, is therefore hypothesised to be inactive.

The MTase component of the type II RM system Csc35704ORFRP (M.Csc35704ORFRP) was found to have a 72 % sequence identity to the MTase M.EacI of *Eubacterium acidaminophilum* (*E. acidaminophilum*), which has been shown to protect DNA from digestion by the restriction enzyme EacI and isoschizomeric enzymes, such as AlwI. The detected recognition sequence of 5'-GGAT^{6mA}C-3' was therefore assigned to M.Csc35704ORFRP. The MTase Csc35704ORFSP, lies 100 bp upstream of M.Csc35704ORFRP and could also be associated with this system.

Based upon the REBASE database, the type II system Csc35704ORFL is predicted to recognise the sequence AGCT with 5mC methylation and could therefore be an isoschizomer of the AluI REase. This MTase of the Csc35704ORFL RM system could also be responsible for the detected base modification 5'-GA^{6mA}AGCTT^{6mA}C-3' and would therefore indicate a novel 8 bp restriction enzyme.

The REBASE database characterises the type II system Csc35704ORFP/Q as containing a hybrid REase and MTase gene with a 6mA methylation type. By the process of elimination, the methylation pattern 5'-CA6mAT6mACGC-3' was assigned.

The type IV RM system, Csc35704McrBCP is a 5mC methylation dependent restriction system, which cleaves DNA containing a methyl cytosine preceded by a purine nucleotide.

The putatively characterised RM systems are summarised in table 5.4.

Table 5.4 Summary of Putative Characterised Methyltransferase Genes

Methyltransferase Name	Designation	Predicted recognition sequence	Type	Predicted Family
M.Csc35704ORFIP	1A	CNRTANNNNNNNTCNNC	6mA	Unknown
M.Csc35704ORFMP	1B	Unknown	6mA	Unknown
M.Csc35704ORFLP	2A	AGCT	5mC	M.AluI like
M.Csc35704ORFQP	2B	CATCGC	6mA	Unknown
M.Csc35704ORFSP	2C	GGATC	6mA	Unknown
M.Csc35704ORFRP	2D	GGATC	6mA	M.EacI/M.AlwI like

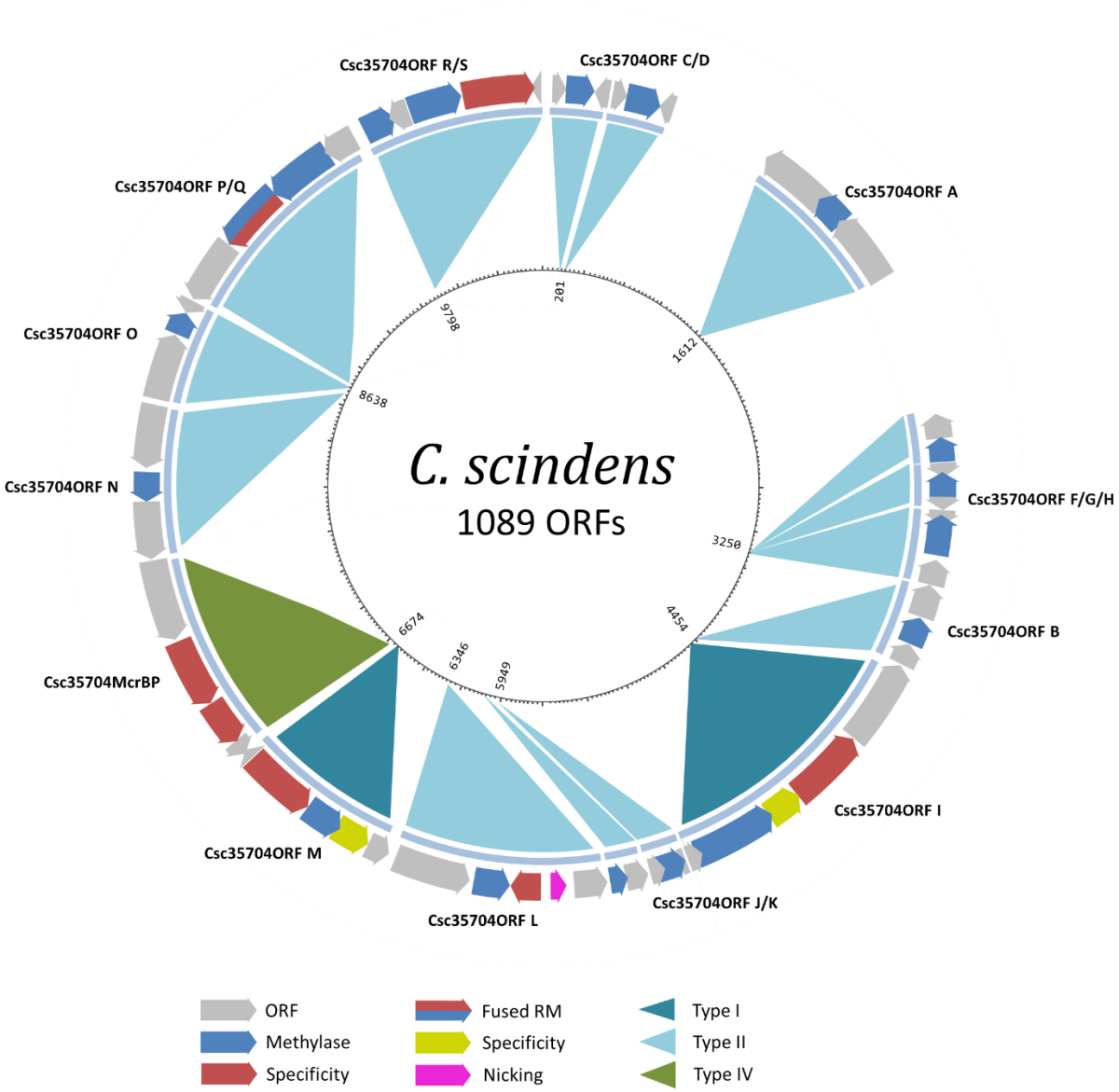


Figure 5.6 The Restriction Modification Systems of *C. scindens* ATCC 35704: The genome of *C. scindens* ATCC 35704 contains two type I, eleven type II and one type IV RM systems with an additional three orphan methyltransferases. The **red** arrows signify REases components, the **blue** MTase, the **yellow** specificity and the **pink** nicking enzymes.

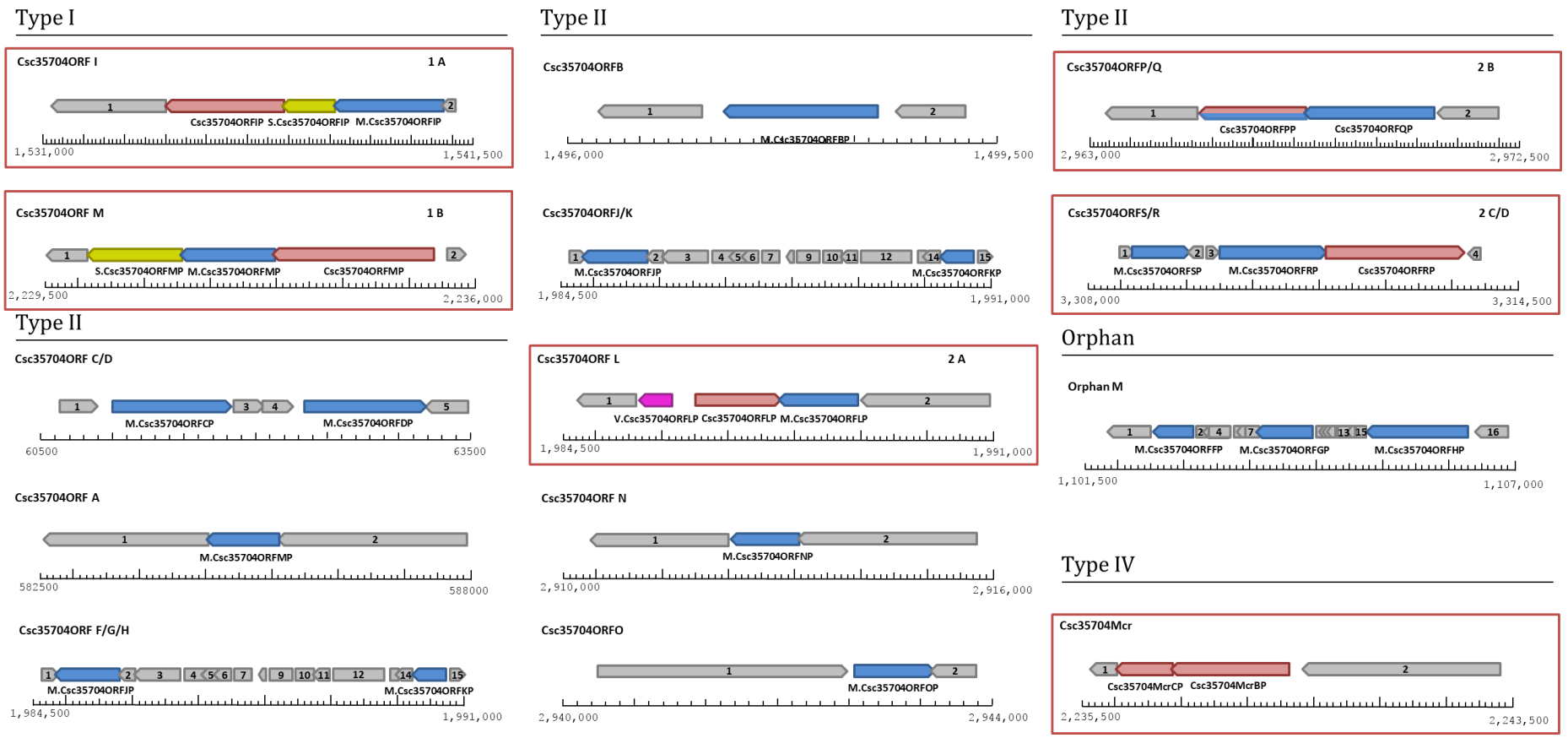


Figure 5.7 The Restriction Modification Systems of *C. scindens* ATCC 35704: The two type I, 11 type II and one type IV RM systems of *C. scindens* ATCC 35704. The **red** arrows signify REases components, the **blue** MTase, the **yellow** specificity and the **pink** nicking enzymes. The RM systems hypothesised to represent barriers to DNA transfer are highlighted in the red boxes.

5.2.6 Restriction Analysis of C. scindens *ATCC 35704 Total DNA*

Based upon sequence homology to characterised RM systems and SMRT base modification data, it was predicted that *C. scindens* ATCC 35704 encodes two distinct MTase genes, M.Csc35704ORFLP (2A) and M.Csc35704ORFRP (2D) that protect DNA against restriction by AluI and AlwI respectively. To verify these putative MTase activities, genomic DNA was extracted from *C. scindens* ATCC 35704 and subjected to restriction by AluI and AlwI (2.5.5). The digests were then analysed by agarose gel electrophoresis, confirming that the *C. scindens* ATCC 35704 genome is protected against restriction by AluI and AlwI. This suggests that the predicted cytosine MTase 2A is responsible for methylation at the predicted recognition site 5'-AGC^{5mC}T-3' and the adenine MTase 2D is responsible for methylation at the predicted recognition site 5'-GGAT^{6mA}C-3'. (Figure 5.8)

Total DNA extracted from *C. scindens* ATCC 35704 was also challenged with restriction by the enzyme DpnI. This was to ascertain whether any of the orphan MTases sensitised host DNA to restriction by this methylation dependent REase. Restriction analysis showed that *C. scindens* ATCC 35704 genomic DNA is sensitive to restriction by the enzyme DpnI, which indicates that one of the MTases present in *C. scindens* ATCC 35704 is responsible for the sensitising methylation of the adenine nucleotide in the GATC DpnI recognition sequence. The apparent orphan MTase M.Csc35704ORFFP (Figure 5.6; Orphan M) is the most likely candidate for this, with BLAST and REBASE search results indicating that ORFFP is an adenine MTase with a GATC recognition domain.

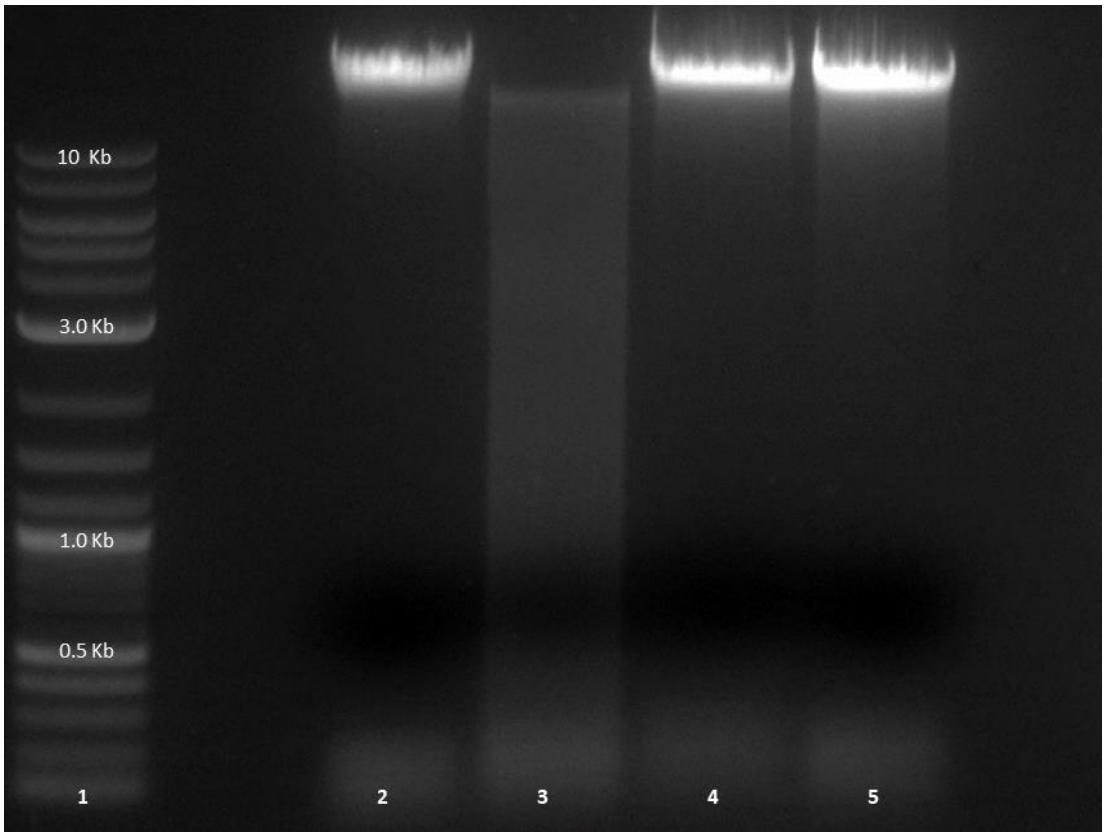


Figure 5.8 Restriction Analysis of Genomic DNA from *C. scindens* ATCC 35704: **Lane 1**: 2 log DNA marker (NEB), **Lane 2**: Unrestricted gDNA, **Lane 3**: gDNA restricted with DpnI, **Lane 4**: gDNA restricted with AluI, **Lane 5**: gDNA restricted with AlwI. Restriction analysis indicates the RM systems of *C. scindens* ATCC 35704 protect self DNA against AluI and AlwI restriction. The gDNA of *C. scindens* ATCC 35704 is vulnerable to cleavage by DpnI.

5.2.7 C. scindens *Lysate Restriction Analysis*

To investigate the nucleolytic activity of the RM systems of *C. scindens* against the shuttle vectors utilised in this study, plasmids were incubated with cell lysates and analysed for DNA cleavage activity.

Cell lysates were prepared from overnight cultures or *C. scindens* strain ATCC 35704 by multiple freeze thaws in CutSmart buffer (NEB), centrifuged and lysate removed. Restriction digest reactions were set up for each of the modular vectors, harbouring the four different Gram-positive replicons as in 2.5.5, using 1 µL of cell lysate in place of the restriction enzyme, and incubated overnight at 37 ^{o}C. Plasmids were also incubated in CutSmart buffer without lysate as a control.

After overnight incubation, reaction mixtures were analysed by agarose gel electrophoresis (Figure 5.9). The gel profiles show that under the *in vitro* conditions of the lysate digest, each of the modular vectors is sensitive to digestion by cell lysates obtained from *C. scindens* ATCC 35704, indicating the presence of active RM systems, likely to be preventing successful DNA transfer in to the organism.

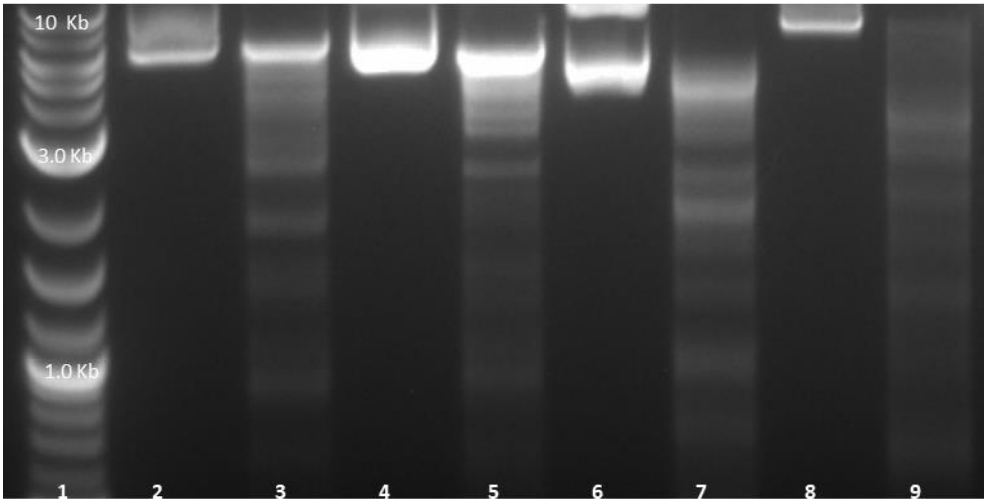


Figure 5.9 Lysate Restriction Analysis: **Lane 1**; 2 log DNA marker (NEB), **Lane 2**; pMTL82151 undigested **Lane 3**; pMTL82151 lysate incubated **Lane 4**; pMTL83151 undigested **Lane 5**; pMTL83151 lysate incubated **Lane 6**; pMTL84151 undigested, **Lane 7** pMTL84151 lysate incubated **Lane 8**; pMTL85151 undigested, **Lane 9**; pMTL85151 lysate incubated. Cell lysates effectively cleaved all of the shuttle vectors tested.

5.2.8 Transcriptomic Analysis of the Restriction Modification Systems of C. scindens *ATCC 35704*

The expression levels of the six RM systems hypothesised to be barriers to DNA transfer in *C. scindens* ATCC 35704 were investigated through the previously acquired RNA-seq data set. The expression profiles were compared with the average expression level of the housekeeping gene *rpoD* (Figure 5.10)

The type I RM systems, 1 A and 1 B, were the most highly expressed RM systems, with $\log_{10}$ (FPKM+1) values of greater than 2 at each time point. Both type I systems were constitutively expressed throughout the exponential phase with levels of expression significantly lower than that of the *rpoD* gene, with P values ranging from 0.0001 – 0.0017 (Figure 5.10; 1 A, 1 B).

The type II RM system 2A was expressed constitutively across all time points, the MTase gene was expressed significantly higher than the REase gene (P values < 0.0001). Both the REase and MTase components were significantly less expressed than *rpoD* with P values of 0.0001 and 0.0006 respectively (Figure 5.10; 2A).

The type II RM system 2B was constitutively expressed throughout exponential period, with each component showing comparable levels of expression (P values > 0.05). The genes of the 2B RM system were significantly less expressed (P = 0.0018 and 0.0021) than the housekeeping gene *rpoD* (Figure 5.10; 2 B)

The type II RM system 2C/D was expressed constitutively over the exponential period. The MTase gene designated 2C was significantly less expressed than the upstream 2D MTase gene (P = 0.005 – 0.0017). All the genes of the 2C/D RM system were significantly less expressed than *rpoD* with P values = 0.0001 (Figure 5.10; 2 C/D).

The type IV RM system was constitutively expressed across all the time points of exponential phase, exhibiting the lowest levels of expression when compared to the type

I and type II RM systems, with average $\log_{10}$ (FPKM+1) values of < 1.5. The system was less highly expressed than *rpoD* with P values < 0.0001 (Figure 5.10; 4 A).

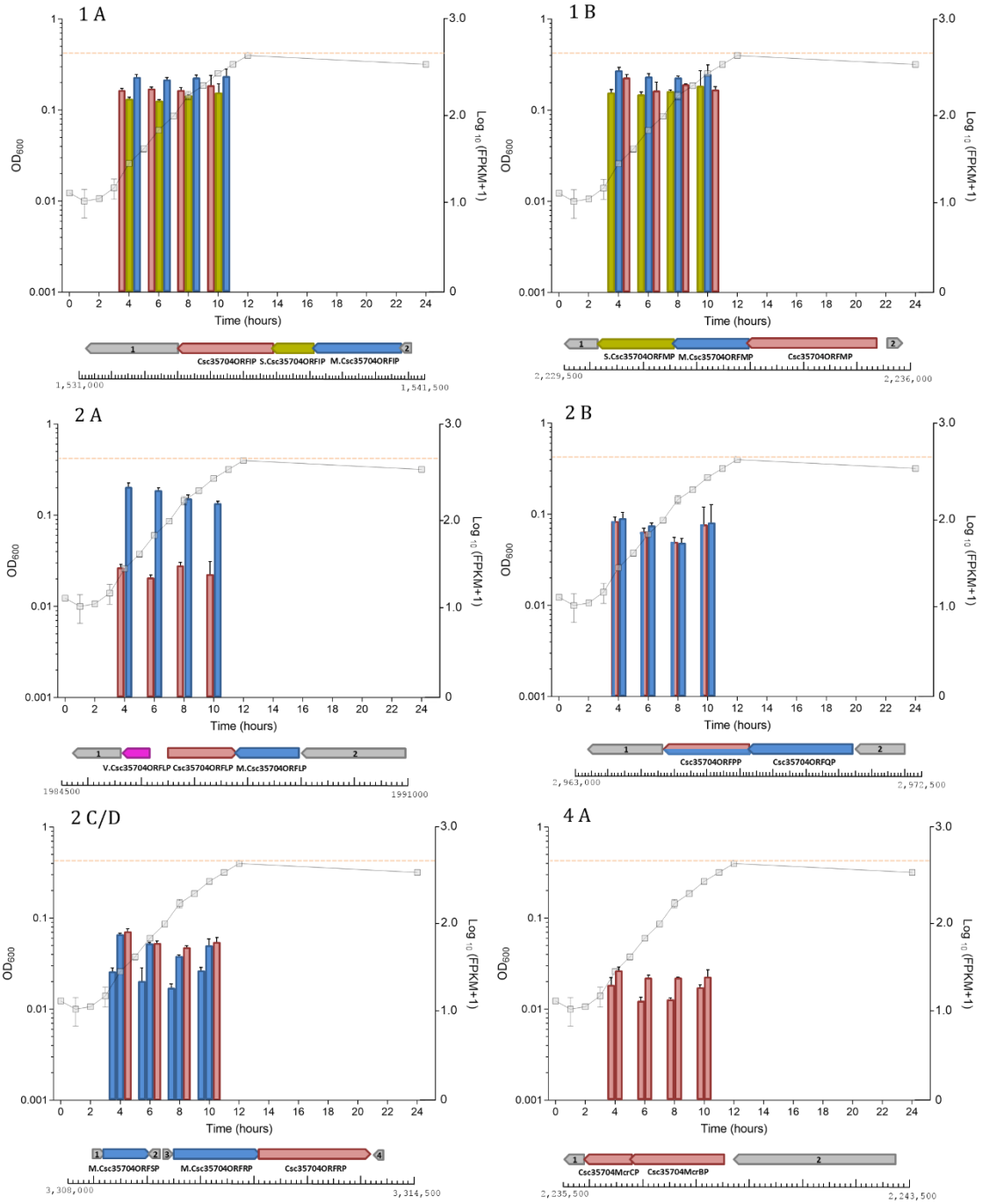


Figure 5.10 Expression Profiles of the RM Systems During Exponential Growth: The Log_{10} (FPKM+1) values for each gene associated with the previously identified RM systems plotted at 4, 6, 8 and 10 h during exponential growth. Blue bars represent the **MTase** genes, red bars the **REase** genes and yellow bars the **specificity** genes. The dashed **orange line** representing the average expression of the housekeeping gene *rpoD*. All of the RM systems were constitutively expressed at all time points throughout the time points sampled in exponential phase.

5.2.9 Cloning of Methyltransferase Genes

Each of the six identified MTase genes were PCR amplified from the genome of *C. scindens* ATCC 35704, with putative ribosome binding sites in the 5' untranslated region (UTR). The type I MTase genes were amplified along with their cognate specificity components (Figure 5.11).

Each of the six identified MTase genes were cloned in to the plasmid pRBV1 (Figures 5.11, 5.12) and electroporated in to *E. coli* TOP10. Each of the MTase genes are placed under the control of the P_{BAD} arabinose inducible promoter, to protect against any potential toxic effects of the MTases to the *E. coli* host.

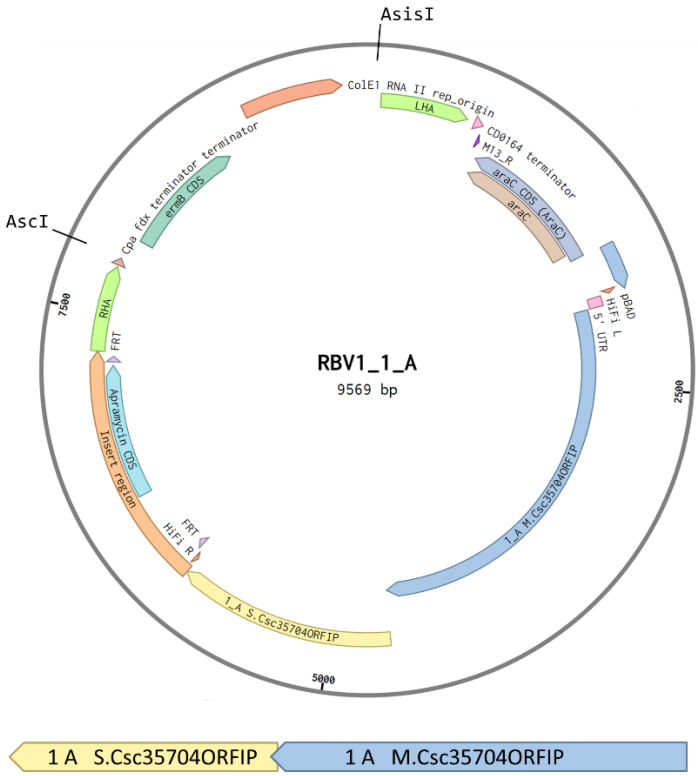


Figure 5.11 The Constructed pRBV1-1A Vector: The 1 A methyltransferase and specificity components cloned in to the RBV vector under the control of the P_{BAD} arabinose inducible promoter and apramycin resistance determinant, *aac*(3)-IV, flanked by *aad9* homology arms.

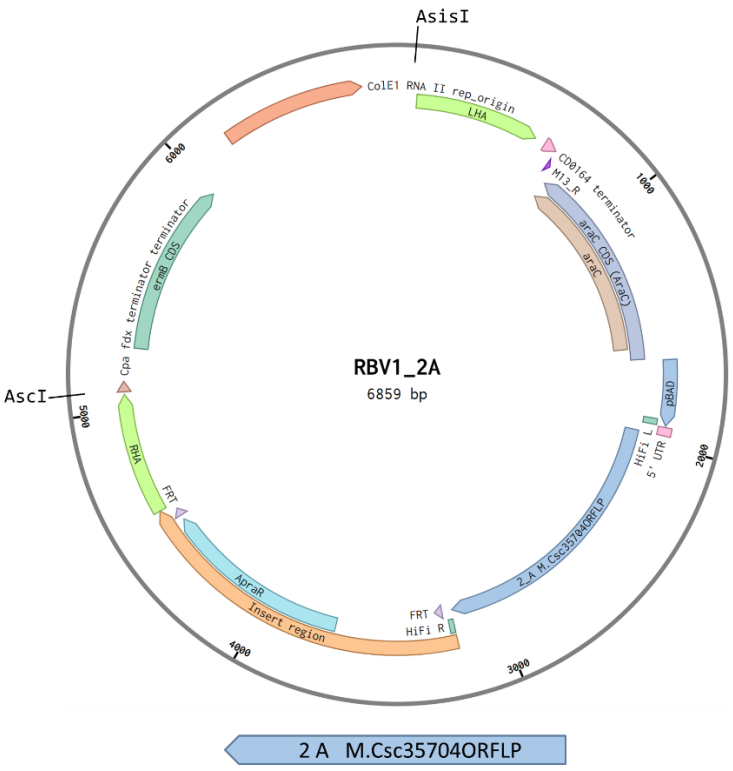


Figure 5.12 The Constructed pRBV1-2A Vector: The methyltransferase gene of 2A cloned in to the RBV vector under the control of the P_{BAD} arabinose inducible promoter and apramycin resistance determinant, *aac*(3)-IV, flanked by *aad9* homology arms.

5.2.10 Restriction Analysis of pRBV1_2A and pRBV1_2D

To further investigate the hypothesised AluI and AlwI-like MTase activities of the type II RM systems, 2A and 2D, the constructed pRBV1 2A/2D vectors were subjected to restriction by the appropriate enzyme. Restriction by AlwI is protected by the deoxyadenosine MTase (*dam*) of the host *E. coli* TOP10 strain, and so to test the pRBV1_2 D construct, all of the constructs were also transformed in to the *dam- E. coli* stellar strain.

Overnight cultures were set up of the appropriate *E. coli* hosts harbouring the pRBV1_2A and pRBV1_2D constructs, as well as all other constructs to act as positive restriction controls, in LB supplemented with erythromycin and apramycin at working concentrations of 50 μg/mL (Table 2.4). Cultures were set up in duplicate, with one culture containing arabinose at 10 mM to induce the expression of the MTase genes. After overnight incubation, a sample was removed from the non-induced culture, from which the plasmid was extracted. Arabinose was then added to the non-induced culture 1 h prior to plasmid extraction, to the concentration of 10 mM. Plasmids were extracted from all three culture conditions to yield plasmids from non-induced, 1 h induction and overnight induction culture conditions.

Digests were then set up for each vector and conditions, with AluI restriction for pRBV1_2A and AlwI restriction for pRBV1_2D, as well as positive controls (2.5.5). After 1 h x, the digests were analysed by agarose gel electrophoresis along with an undigested control.

The AluI restriction analyses show that the MTase 2A confers protection against restriction by AluI, whilst the other five constructs are sensitive to AluI restriction. The RBV backbone contains 23 AluI restriction sites as well as a variable number of internal AluI sites in each of the cloned MTase genes, resulting in the observed restriction profiles (Figure 5.15).

The AlwI restriction analyses show that the MTase 2D confers protection against restriction by AlwI, whilst the other five constructs are sensitive to AlwI restriction. The RBV backbone contains 17 AlwI restriction sites as well as a variable number of internal AlwI sites in each of the cloned MTase genes, resulting in the observed restriction profiles (Figure 5.16).

Interestingly, the uninduced conditions for both the 2A and 2D constructs also conferred protection against AluI and AlwI restriction, although not completely as in the induced conditions. (Figure 5.15; gel 3, Figure 5.16 gel 6).

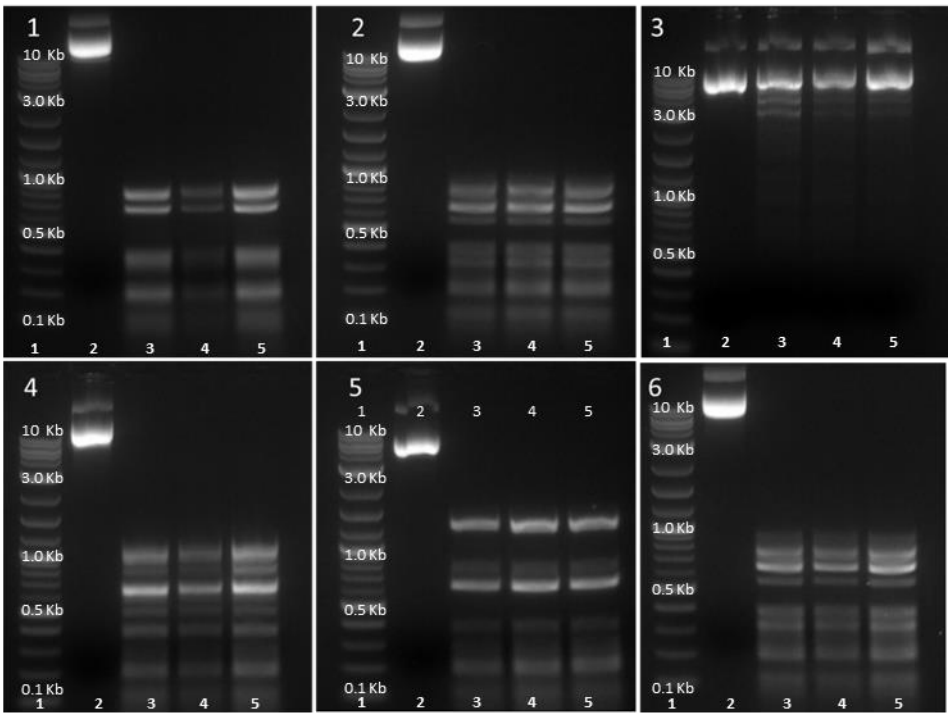


Figure 5.15 AluI Restriction Analysis: Gel 1; 1A, Gel 2; 1B, Gel 3; 2A, Gel 4; 2B, Gel 5; 2C, Gel 6; 2D. For each gel: **Lane 1**; 2 log DNA marker (NEB), **Lane 2**; Unrestricted, **Lane 3**; Restricted Non-induced, **Lane 4**; Restricted 1 hr induced, **Lane 5**; Restricted overnight induction. The MTase 2A appears to confer protection against AluI cleavage.

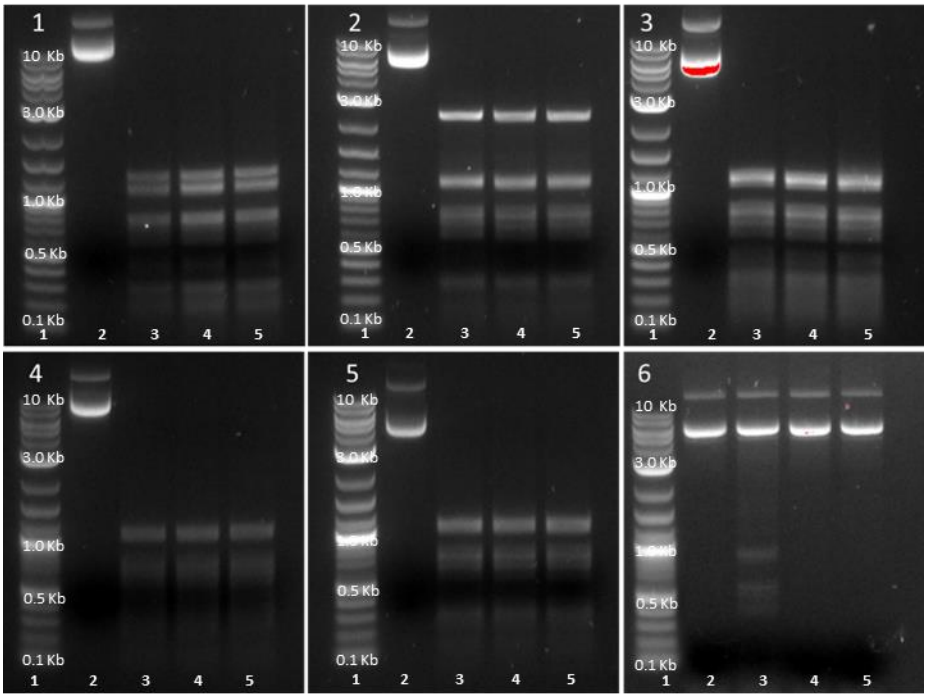


Figure 5.16 AlwI Restriction Analysis: Gel 1; 1A, Gel 2; 1B, Gel 3; 2A, Gel 4; 2B, Gel 5; 2C, Gel 6; 2D. For each gel: Lane 1; 2 log DNA marker (NEB), Lane 2; Unrestricted, **Lane 3**; Restricted Non-induced, **Lane 4**; Restricted 1 hr induced, **Lane 5**; Restricted overnight induction. The MTase 2D appears to confer protection against AlwI cleavage.

5.2.11 Recombineering of the 2A Methyltransferase

A strategy was implemented in the group to express MTase genes from the *E. coli* R702 conjugal R factor was utilised to insert the 2A MTase in to the *aad9* spectinomycin resistance gene, located on R702 (Mansfield, 2017 unpublished).

The entire 2A MTase recombineering cassette was amplified from the pRBV1_2A construct, with flanking *aad9* homology arms, arabinose inducible promoter and apramycin resistance gene and transformed in to the *E. coli* RE-InterStellar host harbouring both the R702 R factor plasmid and pKD46 lambda red recombineering plasmid.

The 2A cassette was inserted in to the *aad9* gene located on the R702 conjugal plasmid using the lambda red recombineering system, as described in 2.5.16 (Figure 5.18). Correct recombinants were subsequently screened for by using colony PCR (2.5.3) with a primer pair which binds internally on the MTase gene and at an external site on the R702 plasmid (Figure 5.19). The R702-2A constructs were confirmed with Sanger sequencing (2.7.1).

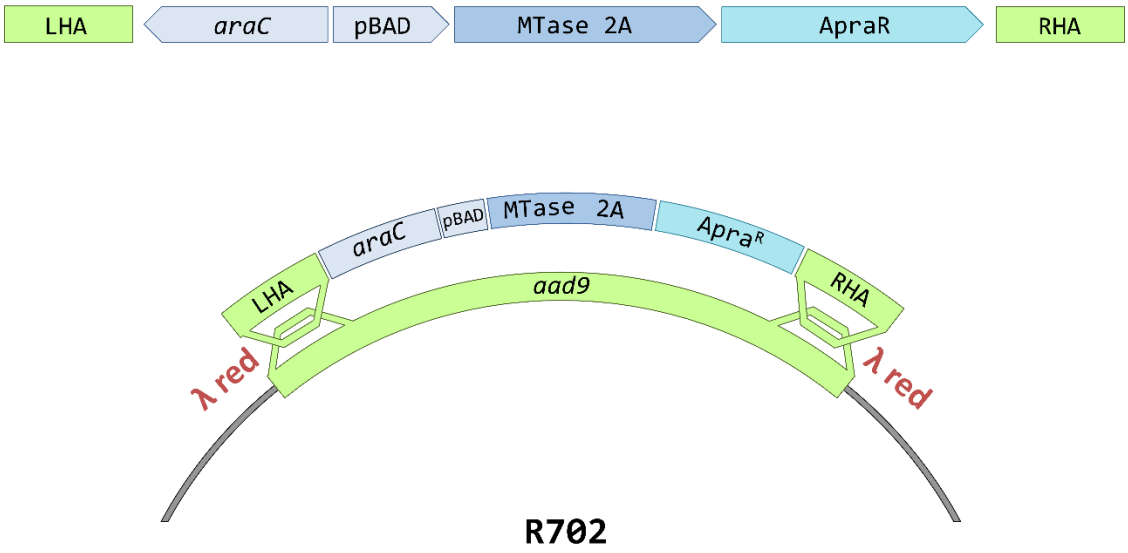


Figure 5.18 Recombineering of the 2A Methyltransferase: The 2A MTase was amplified from the pRBV1_2A vector with *aad9* homology arms, arabinose inducible promoter and an apramycin resistance gene. The entire cassette was inserted in to the *aad9* gene located on the R702 R factor plasmid using the lambda red recombineering system.

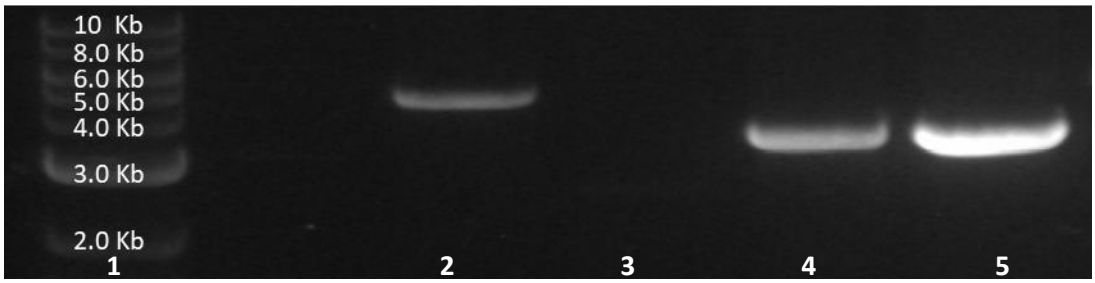


Figure 5.19 Screening of Recombinant R702-2A Plasmids: **Lane 1**: 2 log DNA marker (NEB), **Lanes 2** and **3**; Unsuccessful, **Lanes 4** and **5**; 3.8 Kb bands indicating successful recombineered insertion of 2A in to the *aad9* gene of the R702 conjugal plasmid.

5.2.12 Final Conjugation Attempts

Competent cells were prepared of *E. coli* InterStellar strain harbouring the R702-2A conjugal plasmid (2.3.6). Shuttle vectors carrying four different Gram-positive replicons were transferred in to the cells by electroporation (2.5.11).

Conjugation assays were set up as described in 2.3.8, with conjugal mixtures plated on to BHIS agar supplemented with D-cycloserine at 250 µg/mL and chloramphenicol at 15

µg/mL to select for potential *C. scindens* transconjugants. The assays are detailed below in table 5.5.

No successful *C. scindens* transconjugants were observed.

Table 5.5 Final *C. scindens* Conjugation Assays

	Conjugation Attempts
E.coli **Donor Strains**	InterStellar R702-2A
C. scindens **Recipient Strain**	ATCC 35704
Gram +ve Replicon	pBP1 pCB102 pCD6 pIM13
Selective Marker	*catP*
Mating Times (h)	8 24

5.3 Discussion

5.3.1 Initial DNA Transfer Attempts

DNA transfer was initially attempted by conjugation via the *E. coli* conjugal donor strain CA434. This *E. coli* strain is a derivative of the *E. coli* HB101 strain harbouring the IncPβ conjugative R factor plasmid, R702. The full sequence of R702 has yet to be determined,

but it is known to carry resistance genes for spectinomycin, tetracycline, kanamycin, sulphonamides and mercury salts, as well as the *tra* genes necessary for conjugal transfer. The plasmid R702 was first isolated from a naturally occurring strain of *Proteus mirabilis* with a size estimated to be 70-90 Kb (Hedges and Jacob, 1974). The donor strain CA434 has been utilised widely in conjugation assays with *C. difficile* and was selected to form the ground work for conjugal DNA transfer attempts in *C. scindens* strains ATCC 35704 and VPI 12708 (Purdy *et al.,* 2002). Heat shock of the recipient *C. scindens* strains was also attempted, which has been proved to increase conjugal transfer frequencies in a panel of *C. difficile* strains by the temporal inactivation of RM systems (Kirk *et al.,* 2016).

Due to the failure of early DNA transfer attempts by conjugation with the two aforementioned strains of *C. scindens*, a wider repertoire of *C. scindens* strains was acquired from researchers at Virginia Commonwealth University in the USA. The six additional clinical isolate strains (SA14, SO76, SO77, KS11, G11 and C132) were tested for conjugation, with the possibility that these strains may be more amenable to accepting exogenous DNA, although these additional strains also proved to be unsuccessful for conjugation attempts.

Next, further *E. coli* donor stains were tested, the Hfr strain S17-1 and SExpress. The S17-1 strain was selected as it was an additional conjugal donor strain available in the laboratory and is an Hfr strain, where the conjugal transfer functions of the RP4 plasmid have been integrated in to the chromosome (Strand *et al.,* 2014). The SExpress strain is a derivative of the NEB Express cloning host, which harbours the R702 conjugative R factor plasmid (Woods, 2016 unpublished). The SExpress strain has been shown to increase transfer frequency in *Clostridium autoethanogenum* where the type IV RM system is a barrier to reliable transfer rates (Woods 2017 unpublished). The SExpress strain is dam^{+}/dcm^{-} and therefore yields plasmids protected from type IV systems which recognise foreign cytosine methylated DNA.

The absolute failure of establishing DNA transfer by conjugation indicated the presence of significant barriers to exogenous DNA transfer in *C. scindens*. One potential barrier was the occlusion of the conjugative pilus by the fimbriated surface of *C. scindens,* a form of surface exclusion. Another potential barrier was the possible presence of RM systems in *C. scindens.* A strategy to circumvent both barriers by utilising *in vitro* methylation and transfer by electroporation was attempted. A cocktail of plasmids was modified *in vitro* by GpC MTase (NEB), by methylation of all cytosine residues in the double stranded sequence 5'-GC'-3' (2.5.13). This cocktail was then electroporated in to competent *C. scindens* cells (2.5.12).

After electroporation attempts also failed, it was decided to employ PacBio SMRT sequencing to obtain a reliable genome sequence and base modification data to analyse the methylome of *C. scindens* strain ATCC 35704.

5.3.2 Genome and Methylome Analysis

The ATCC 35704 *C. scindens* strain was selected for further analysis as this strain represents the *C. scindens* type strain and has proved to be easier to work with in the laboratory.

The genome sequence of *C. scindens* ATCC 35704 revealed a type IC CRISPR-Cas system and associated CRISPR loci (Figure 5.5). This could potentially provide a barrier to conjugal transfer, as DNA transfer by conjugation has been shown to be limited by the CRISPR-Cas system of *Staphylococcus epidermidis*. A clinical isolate of *S. epidermidis* was found to harbour a spacer in its CRISPR locus homologous to the *nes* nickase gene, found on all of the sequenced staphylococcal conjugative plasmids (Marraffini & Sontheimer, 2008). However, it has also been shown that CRISPR-Cas systems do not effectively inhibit the acquisition of beneficial genes, such as antibiotic resistance, through the loss

of CRISPR function where a defective or lost CRISPR spacer is maintained due to selective pressure (Jiang *et al.,* 2013). None of the spacers in the CRISPR loci of *C. scindens* ATCC 35704 exhibited homology to the shuttle vectors utilised in this study and therefore, the CRISPR-Cas system was determined to offer no threat to attempted DNA transfer efforts.

The genome sequence also revealed the presence of multiple RM systems in *C. scindens* ATCC 35704, (Figures 5.6 and 5.7), six of which had cognate REase components presenting a potentially significant barrier to DNA transfer. Transcriptomic analysis showed that these systems were expressed constitutively during exponential growth phase, indicating a persistent threat to exogenous DNA in the intracellular environment (Figure 5.10).

Further characterisation and sequence analysis using the NCBI and REBASE databases revealed putative identities and recognition sequences for the RM systems of interest. Two of the type II RM systems were identified as AluI and EacI/AlwI-like systems.

The PacBio base modification data detected the motif 5'-GGAT^{6mA}C-3' which was assigned to the AlwI-like M.Csc35704ORFRP (designated 2D). PacBio base modification data detected only N6-methyladenine (6mA) and no C5-methylcytosine (5mC) (Tables 5.3 and 5.4). It was subsequently deduced that the lack of detected 5mC methylation patterns was due to insufficient sequencing depth of the SMRT sequencing. More SMRT cell sequencing runs are required to achieve the coverage depth to detect cytosine base modification.

The SMRT base modification data showed the presence of the 6mA methylation pattern of 5'-GA^{6mA}AGCTT^{6mA}C-3', which was assigned to M.Csc35704ORFL (designated 2A) although sequence homology and the predicted recognition sequence of 2A indicated a relation to the AluI type II RM system, with the recognition sequence 5'-AGCT-3'. The detected 8 base 6mA motif would represent a novel 8 bp cutter type II restriction enzyme discovered in the *C. scindens* ATCC 35704 strain. However, due to the relatively low

objective score of 186.48 for this motif compared to scores > 100,000 for the other detected motifs in the base modification data (Table 5.3) and the insufficient sequencing depth to detect the 5mC methylation pattern of 5'-AGC^{5mC}T-3', it was decided that the assignment of AGCT to 2A and putative characterisation of an AluI-like system was more likely.

The characterisation of 2A and 2D as AluI and AlwI-like systems was further confirmed by restriction analysis of gDNA extracted from *C. scindens* and recombinant 2A and 2D MTase plasmids challenged with AluI and AlwI restriction respectively (Figures 5.8, 5.15, 5.16). Protection of non-induced plasmids expressing the 2A and 2D MTases was also observed, and it was hypothesised that this was due to the SNP in the operator region of the arabinose inducible promoter.

The tetranucleotide sequence AGCT appears ubiquitously in DNA sequences and therefore, the 2A RM system of *C. scindens* ATCC 35704 represents a significant threat to unprotected exogenous DNA and was identified as a key barrier to establishing DNA transfer in *C. scindens* ATCC 35704.

A number of incomplete and orphan MTases without cognate REase components were also identified in the genome sequence of *C. scindens* ATCC 35704. Orphan MTases are solitary genes, which are not associated with a REase gene with a possible RM system ancestor, whereas incomplete MTases are those which have lost their REase partner relatively recently. These incomplete MTases may exist through degradation of the REase component, with the MTase maintained through the selective pressure of post segregational lethality, due to residual REases in subsequent daughter cells in a toxin-antitoxin-like manner (Handa & Kobayashi, 1999).

Orphan MTase are thought to have roles in coordination of cellular functions and epigenetics. The Dam N6-methyladenine orphan MTase is responsible for the regulation of DNA replication and repair, transposition of insertional elements and gene

transcription in γ-proteobacteria and is required for virulence in certain *Haemophilus, Vibrio, Salmonella* and *Yersinia* species (Wion & Casadesús, 2006). Orphan MTases may also persist in the genome to protect against parasitism by selfish RM systems on mobile genetic elements entering the cell, for example the Dcm C5-methylcytosine MTase of *E. coli* protects against EcoRII RM system found on the *E. coli* plasmid N3 (Takahashi *et al.,* 2002).

Utilising the PacBio methylome data (Table 5.2) and the confirmed activities of the 2A and 2D RM systems (Figures 5.15 and 5.16), it is possible to assess the sensitivity of plasmid DNA to be transferred in to *C. scindens* ATCC 35704 and develop an approach to circumvent the RM systems through the mimicking of the protective host methylation pattern and by measures to reduce the recognition sites on the plasmid (Figure 5.20).

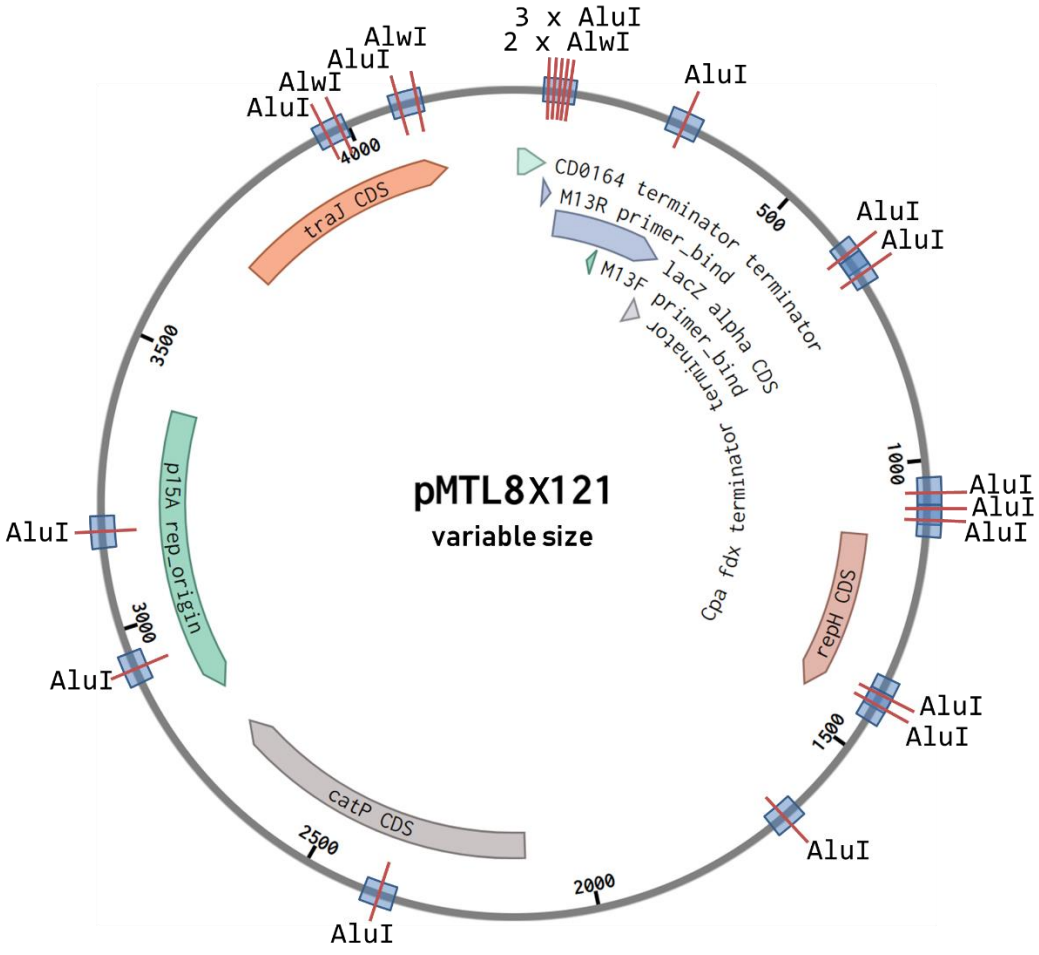


Figure 5.20 The Modular Vector pMTL8X121: A suitable vector for DNA transfer in *C. scindens* ATCC 35704 when transferred from a *dam*$^{+}$/*dcm*$^{-}$ *E. coli* conjugal donor expressing the 2A methyltransferase, protecting the 18 AGCT AluI sites from cleavage by the 2A REase of *C. scindens*.

The type I RM system 1A, with the putative recognition sequence of 5'-CNRTANNNNNNNTCNNC-3' can be predicted to cut once in pMTL shuttle vectors which utilise the ColE1 Gram-negative replicon. The ColE1 replicon contains the sequence motif 3'-CAGTATTTGGTATCTGC-5' which would result in the binding and potential cleavage by the type I RM system, 1A. Using the bi-directional translocation model of DNA cleavage by type I RM systems, DNA cleavage of circular DNA with only one recognition sequence is hypothesised to occur when the single bound enzyme complex stalls during translocation once it has navigated the entirety of the plasmid, whereas linear DNA with one recognition sequence cannot be cleaved (Janscak *et al.,* 1999). The type I RM system 1B, as previously mentioned, is hypothesised to be inactive due to a frameshift mutation in the MTase subunit.

The type II RM system 2A, characterised as an AluI-like enzyme with the recognition sequence of 5'-AGCT-3' represents a significant restriction threat to the shuttle vectors utilised in this study, with 18 - 21 GATC sites present in the plasmid sequence, depending upon the Gram-positive replicon present. Protection against the 2A RM systems could be provided by a recombinant 2A MTase expressed in the *E. coli* conjugal donor host (Figure 5.20).

The type II RM system 2B, with a putative recognition sequence of 5'-CATCGC-3' is not a barrier to DNA transfer with the vectors utilised in this study, as there are none of these sequence motifs present.

Following restriction analysis, the type II RM system 2C is hypothesised to encode and orphan or incomplete MTase, un-related to the activities of its close neighbour 2D. The type II RM system 2D, characterised as an AlwI-like enzyme closely related to the EacI

RM system of *E. acidaminophilum*, with the recognition sequence of 5'-GGATC-3' is predicted to restrict the pMTL vectors at three to six cut sites, depending upon the Gram-positive replicon. However, the restriction activity of AlwI and it's isoschizomers is abolished by Dam methylation and could therefore be neutralised by using a *dam*$^{+}$ *E. coli* conjugal donor.

The type IV RM system 4A is an McrBC-like type IV system, cleaving DNA in a methyl-dependent manner. The presence of Type IV systems, such as McrBC, can result in a reduction of transformation efficiency, if the DNA to be transformed originates from a strain with an *mcr* or *mcc* genotype (Bickle & Kruger, 1993; Panne *et al.,* 2001). The 4A system can therefore be circumvented by using an *E. coli* host with an *mcr*$^{-}$/*mcc*$^{-}$ genotype, such as the SExpress donor strain.

The RM systems of *C. scindens* ATCC 35704 present a considerable barrier to the transfer of exogenous DNA and provide an innate immunity against threats to the genome, such as bacteriophages. It is surprising therefore to find a latent bacteriophage residing in the *C. scindens* ATCC 35704 genome, designated ϕClsc35704. The 33.2 Kb nucleotide sequence for ϕClsc35704 contains 50 AlwI and seven AluI recognition sites, making it extremely sensitive to nucleolytic cleavage by just the 2A and 2D type II RM systems alone (Figure 5.21).

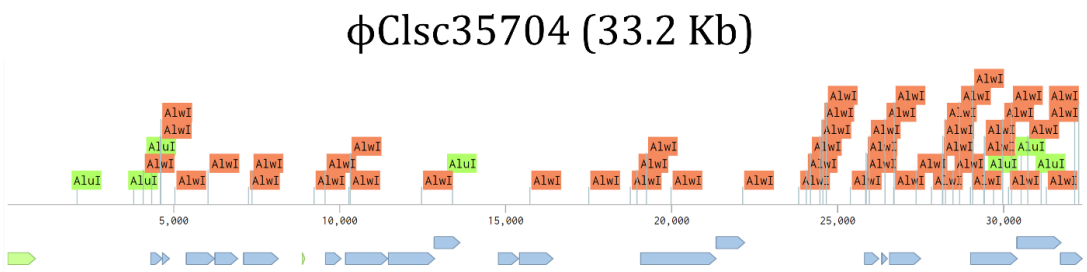


Figure 5.21 The Linear Prophage Sequence ϕClsc35704 of *C. scindens* ATCC 35704: The nucleotide sequence of ϕClsc35704 showing 50 **AlwI** and seven **AluI** recognition

sites, indicating the vulnerability of the foreign bacteriophage DNA to the type II REase systems of *C. scindens*.

Study of type II RM systems based upon statistical atypical codon usage of RM genes and the host bacteria, suggests that horizontal gene transfer is likely to be responsible for the wide distribution of such systems in bacteria. The mechanism by which type II RM systems work proposes the sequential acquisition of the protective MTase, followed by the REase component, (Jeltsch & Pingoud, 1996). The mobilisation of RM systems can occur through mobile genetic elements such as plasmids, insertion sequence elements, transposons and prophages. This suggests an evolutionary temporal separation between the transduction of *C. scindens* ATCC 35704 by ϕClsc35704 and the acquisition of the 2A and 2D RM systems, with phage infection occurring prior HGT of these type II systems.

Another hypothesis is that ϕClsc35704 originated from a bacterium with a common methylation pattern to that of *C. scindens* ATCC 35704, so that any resulting bacteriophage progeny would be able to infect with DNA protected against cleavage by the type II systems of *C. scindens* ATCC 35704. Interestingly, further inspection of the ϕClsc35704 prophage genome revealed the presence of two apparent orphan MTase genes, a 565 bp C5 cytosine MTase and an 846 bp N6 adenine MTase. These MTase genes are yet to be characterised, but it can be hypothesised that these bacteriophage encoded genes are able to confer protection against host RM systems, such as those present in *C. scindens* ATCC 35704 (Murphy *et al.*, 2013). The maintenance of these genes on the relatively small genome of ϕClsc35704 indicates a survival benefit and the integration of orphan MTase encoding genes in to bacteriophage genomes may be a contributing factor to extending the host range of bacteriophages. These data highlight the constant and evolving arms race between bacteria and bacteriophages, and the drive of these viruses to evolve mechanisms to overcome the resistance mechanisms of their hosts.

Perhaps lessons can be learned from these ancient adversaries of bacteria; characterisation of the orphan MTases of ϕClsc35704 could provide the key to establishing DNA transfer in *C. scindens* ATCC 35704.

Further work to implement the strategies discussed to protect shuttle vectors from the RM systems of *C. s cindens* strain ATCC 35704 offers the promise to establish DNA transfer in this organism. This would enable the development of tools to manipulate the genome of *C. scindens* to aid in the study of this important organism.

Chapter 6 Final Conclusions and Perspectives

6.1 Conclusions and Outcomes

The human gut microbiota is a diverse microbial ecosystem inhabiting the gastrointestinal tract, with bacteria representing a dominant and diverse component of the human gut microbiome. The bacterial population of the human gut numbers in the tens of trillions, significantly contributing to the optimal health and wellbeing of an individual through complex microbiota-host homeostatic interactions. Pathological imbalances in the gut microbiota, as a result of changes in dietary habits and pharmaceutical treatment regimens, predispose an individual to certain health conditions, including carcinogenesis and colonisation by pathogenic bacteria. Broad spectrum antimicrobial therapy may potentially render a patient susceptible to symptomatic colonisation by the gut pathogen *C. difficle*, with disease ranging from mild diarrhoea to potentially life-threatening conditions.

C. difficile is an important pathogen of the gut, associated with a massive economic burden on worldwide healthcare. The shifting landscape of *C. difficile* epidemiology has seen clear changes in risk factors, clinical presentation, responsiveness to treatment and disease outcome. The increases in disease severity appear to be correlated with the emergence of so called hypervirulent strains, producing elevated levels of toxin and exhibiting increased resistance to antibiotics, resulting in higher morbidity and mortality rates. The frontline antibiotic treatments for *C. difficile* infection, metronidazole and vancomycin, are associated with disruptions to the normal gut microbiota and may therefore contribute to relapse and reinfection rates (Lewis *et al.,* 2015). It is therefore socially and economically important to develop new treatments for *C. difficile* with less deleterious effects on the gut microbiota.

Fidaxomicin, a novel first in class macrocyclic antibiotic is the first new drug to be approved for *C. difficile* treatment in over three decades. The drug exhibits favourable

pharmacokinetics, persisting at the site of infection and is reported to have a narrow spectrum of activity, with reduced perturbation of the Gram-negative organisms inhabiting the gut. In a panel of six clinically relevant *C. difficile* PCR ribotypes, it was found that FDX inhibits spores in the early stages of outgrowth, preventing the recrudescence of dormant spores to toxin producing vegetative cells at concentrations of 0.03 – 0.05 μg/mL.

Using allelic exchange technology and whole genome re-sequencing, it was definitively established that SNPs in the *rpoB* gene confer a reduced susceptibility to FDX. The re-created SNPs result in amino acid changes in the switch region of the β subunit of *C. difficile* RNA polymerase, affecting the ability of FDX to bind and inhibit RNA transcription. Furthermore, the mutations in *rpoB* which give rise to reduced susceptibility to FDX also resulted in associated fitness costs *in vitro* and *in vivo*, hypothesised to be due to inefficient association with specific sigma factors involved in the initiation of the transcription of certain genes. When compared to the parental strain, the isogenic *rpoB* mutants exhibited significant attenuation in terms of growth rates and competitive growth advantage, whilst sporulation and germination were marginally affected. Toxin production and cytotoxicity of the mutants was also reduced when compared to the parental strain. These data suggest that the *rpoB* mutants are less infectiously fit than the FDX susceptible parental strain, a hypothesis that was backed up by *in vivo* studies utilising the Golden Syrian hamster model of infection. Each of the mutant strains took significantly longer to reach the clinical end-point when compared to the parental strain despite comparable caeca colonisation levels between the parental and mutant strains.

Patients treated with FDX have been reported to have lower relapse rates than those treated with vancomycin (Cornerly *et al*., 2012). This is hypothesised to be an effect of the narrow spectrum of the drug, particularly correlated to the sparing of the bile acid metabolising members of the gut microbiota. Crucially, FDX is reported to be inactive

against rRNA cluster XIVa clostridia, which include the 7α/β-dehydroxylating bacteria *C. scindens* and *C. hylemonae*, unlike vancomycin and metronidazole (Tannock *et al*., 2010). Secondary bile acids have been reported to be inhibitory to *C. difficile* vegetative cell growth, (Sorg & Sonenshein, 2008), and significantly reduce germination, outgrowth and toxin production of clinically relevant strains (Thanissery *et al*., 2017).

The cell free supernatants obtained from *C. scindens* cultures supplemented with the primary bile acid, CA, inhibited the outgrowth of a panel of six clinically relevant strains of *C. difficile* in this study. The lack of inhibition by CFS from cultures without CA supplementation suggests a link between the inhibition of *C. difficile* vegetative cell growth and the bile acid metabolism of *C. scindens*. Although this effect could also be due to other secondary metabolites in the media and does not provide a definitive link between the inhibition of *C. difficile* and *C. scindens* derived secondary bile acids.

Based upon the evidence in the literature and in this study, it can be hypothesised that the colonisation resistance against *C. difficile* provided by the normal, healthy gut microbiota is associated with the bacterial bio-transformation of bile acids. The bile acid inducible pathway which results in the formation of the hypothesised inhibitory secondary bile acids has been studied extensively in the organisms *C. scindens* and *C. hylemonae*, with the oxidative arm of the pathway completely elucidated, whilst the reductive arm remains unknown. To further the understanding of the bile acid transformation of *C. scindens* and related organisms, it is important to gain a complete understanding of the pathway. This study aimed to focus the current understanding of the genes involved in the reductive arm of the 7α/β-dehydroxylation pathway, by performing transcriptomic analysis of *C. scindens* cultured with and without primary bile acid induction. The transcriptomic data provided further evidence of the inducible nature of the pathway and identified novel genes, which may be involved in bile acid metabolism in *C. scindens*.

Firstly, a gene encoding a novel bile acid dehydratase was identified, with proposed activity analogous to the rate determining dehydration step performed by BaiE in the oxidative arm of the pathway (Bhowmik *et al.,* 2016). The gene was found to encode a protein belonging to the NTF2 family, with structural similarity to the 7α-dehydratase, BaiE, exhibiting the characteristic SnoaL fold domain, twisted α+β barrel active site. The identification of a second 7α-dehydratase enzyme in *C. scindens* ATCC 35704 has potential implications for this rate-limiting dehydration step of the pathway.

A gene annotated as coding for a FAD binding flavoprotein was the most highly upregulated uncharacterised gene, placed in to the same dendrogram as the *bai* genes at the 6 and 8 hour time points, by hierarchical clustering. The protein encoded by this gene was found to share structural homology to the steroid modifying enzyme 3-ketosteroid dehydrogenase, belonging to the EC 1.3 class of oxidoreductases. This indicates the potential for the identified gene to code for an enzyme capable of reducing the double CH-CH bonds of the 3-oxo-$\Delta^{4,6}$ DCA intermediate, using FAD as a hydrogen donor. The proposed gene may code for an enzyme involved in the antepenultimate steps of the reductive arm of the pathway, preceding 3α reduction of the 3-dehydro-DCA intermediate.

To further the understanding of the bile acid metabolism of *C. scindens* and its potential involvement in the competitive exclusion of *C. difficile*, it is of vital importance to develop tools to create precise modifications in the *C. scindens* genome. The first step in the development of gene systems is the establishment of DNA transfer, which to date has not been achieved in *C. scindens*. After exhaustive attempts to establish DNA transfer in *C. scindens* by conjugation were unsuccessful, it was decided to employ PacBio SMRT sequencing to identify potential barriers to DNA transfer. The complete genome sequence for *C. scindens* strain ATCC 35704 was determined using PacBio SMRT sequencing and Illumina MiSeq technologies. The annotated 3.6 Mbp sequence revealed the presence of significant barriers to DNA transfer in the presence of numerous

endogenous RM systems in the ATCC 35704 genome, six of which had cognate REase genes. Base modification data from SMRT sequencing allowed the putative characterisation of the active systems, reinforced by data from restriction analysis performed on total DNA extracted from *C. scindens* ATCC 35704 and plasmids expressing recombinant *C. scindens* MTases. Although DNA transfer is still yet to be established, the key barriers have been identified in this study allowing the development of a promising strategy to circumvent the RM barrier.

6.2 Future Perspectives

The rise in antimicrobial resistance, coupled with a lack of newly developed drugs to combat resistant pathogens could see society heading for a perfect storm scenario, with disastrous social and economic consequences. A concerning trend in the emergence of multi drug resistant *C. difficile* strains of the hypervirulent ribotype 027 group has been observed amongst North American isolates (Tenover *et al*., 2011). The novel antimicrobial FDX offers the promise to turn the tide in the battle to treat obstinate *C. difficile* infection, with resistance to FDX rarely observed; only one case isolated in the clinic to date (Goldstein *et al*., 2011). The unique target site of FDX on the RNAP core enzyme ensures no cross resistance between FDX and other RNAP inhibitors such as the rifamycin class (Babakhani *et al*., 2014). It can be concluded from the data presented in this study that the clinical relevance of mutant strains resistant to FDX is low, with such mutants likely to be outcompeted by the more fit, susceptible strains when the selective pressure of FDX is removed.

Another promising alternative to treat CDI is therapy to reinstate the natural colonisation resistance provided by members of the gut microbiota. Faecal microbiota

transplant (FMT) has been shown to be an effective means of treatment for recurrent CDI by reconstitution of the normal, healthy microbiota (Gough *et al.,* 2011). However, problems arise with this approach when considering how to regulate such a therapy, especially given the risks of transmissible stool borne diseases (Hamilton, *et al*., 2012)

A model of *C. difficile* colonisation resistance based upon the microbial transformations of bile acids in the gut has been proposed, with *C. scindens* hypothesised to be at the centre, as the most active 7α/β-dehydroxylating member of the gut microbiota. Influential work conducted by Buffie and colleagues showed that the administration of *C. scindens* specifically conferred protection in the mouse model (Buffie *et al*., 2015). The addition of *C. scindens* to a defined consortium of microbes in the mouse significantly reduced *C. difficile* colonisation, associated with the microbial transformation of bile acids and the resultant alterations to the murine bile acid pool (Studer et al., 2016).

To definitively prove the link between the protection conferred by *C. scindens* and bile acid metabolism, the ability to manipulate the *C. scindens* genome is essential. Once DNA transfer has been achieved, the clostridial roadmap for gene systems development will be implemented, utilising the well-established ClosTron, ACE (Allele Coupled Exchange) and *mariner* transposon technologies (Minton *et al*., 2016). Perhaps the most promising technology is the newly developed CRISPR/*cas9* system for creating gene knockouts, using the *Streptococcus pyogenes* Cas9 protein to select for double cross over mutants by chromosomal DNA cleavage of the wildtype gene by the CRISPR-Cas complex (Cong *et al*., 2013). Another possibility is the repurposing of the identified native CRISPR-*cas* system of *C. scindens*, which could be harnessed to generate mutants reducing the relatively large cargo requirements of CRISPR knock out vectors.

Utilising *C. scindens* as a probiotic therapy for CDI is an exciting prospect, although it should be considered that *C. scindens* is just one member of a syndicate of bacteria involved in the transformation of bile acids. To permit the 7α/β-dehydroxylation of primary bile acids by *C. scindens*, the conjugated primary bile acids exported from the liver must first be liberated from their taurine or glycine conjugates by microbial bile salt hydrolysis. The genes required for bile salt hydrolysis are not present in *C. scindens* but have been observed in a number of Gram-positive genera inhabiting the gut including; *Clostridium*, *Lactobacillus*, *Bifidobacterium* and *Enterococcus* (Ridlon *et al.,* 2016). Taurine conjugates of CA are widely acknowledged to initiate the germination of *C. difficile* spores *in vitro*, further highlighting the importance of the hydrolysis of bile salts in colonisation resistance, to reduce the levels of TC germinant (Wilson *et al*., 1982). The tauro-conjugation of primary bile acids appears to be heavily influenced by diet, with the high animal protein "Western Diet" resulting in increased levels of taurine conjugates, when compared to vegetarian diets, which favour the formation of glycine conjugates (Hardison, 1978).

It is therefore important when designing a probiotic to treat CDI, to consider all bacteria involved in bile acid metabolism, although some caution is advised with this approach. The secondary bile acids, DCA and LCA, produced by *C. scindens* have been differentially implicated in colon carcinogenesis (Ridlon *et al*., 2016). It has also been observed that obesity induced alterations in the gut microbiota of mice increases the levels of the DCA, resulting in DNA damage and the development of the senescence-associated secretory phenotype (SASP) of hepatic cells, promoting hepatocellular carcinoma (Yoshimoto *et al*., 2013).

This all serves to reinforce the importance of the intricate bacteria-bacteria and bacteria-host relationships in the gut. The gut microbiota of a healthy individual is a finely tuned organ performing extensive complex metabolic functions involving the degradation of indigestible compounds such as polysaccharides and polyphenols and the synthesis of

vitamins. The gut microbiota is also thought to play roles in the aetiology of gastrointestinal diseases such as colon cancer, inflammatory bowel disease, irritable bowel syndrome and colonisation by pathogenic bacteria. It has also been hypothesised that the gut microbiota may have a role in obesity and diabetes. Therefore, any probiotic therapies to treat and prevent CDI must be tightly regulated and take in to account these complex interactions.

The importance of the human microbiota cannot be overstated. Targeted therapies which limit the disruption of the gut microbiota are highly desirable and the harnessing of the beneficial bacteria of the gut could yield novel, personalised probiotic therapies to prevent and treat an array of diseases.

Looking to the future, the further study and characterisation of the microorganisms inhabiting our bodies, such as *C. scindens*, will unlock the vast untapped potential of our microscopic allies.

7. Bibliography

Aertsen, A., & Michiels, C. W. (2005). Mrr instigates the SOS response after high pressure stress in Escherichia coli. *Molecular Microbiology*, *58*(5), 1381–1391. https://doi.org/10.1111/j.1365-2958.2005.04903.x

Alander, M., Satokari, R., Korpela, R., Saxelin, M., Vilpponen-Salmela, T., Mattila-Sandholm, T., & von Wright, a. (1999). Persistence of colonization of human colonic mucosa by a probiotic strain, Lactobacillus rhamnosus GG, after oral consumption. *Applied and Environmental Microbiology*, *65*(1), 351–4. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=91031&tool=pmcentrez&rendertype=abstract

Allen, C. A., Babakhani, F., Sears, P., Nguyen, L., & Sorg, J. A. (2013). Both fidaxomicin and vancomycin inhibit outgrowth of *Clostridium difficile* spores. *Antimicrobial Agents and Chemotherapy*, *57*(1), 664–667. https://doi.org/10.1128/AAC.01611-12

Al-Nassir, W. N., Sethi, A. K., Li, Y., Pultz, M. J., Riggs, M. M., & Donskey, C. J. (2008). Both oral metronidazole and oral vancomycin promote persistent overgrowth of vancomycin-resistant enterococci during treatment of *Clostridium difficile*-associated disease. *Antimicrobial Agents and Chemotherapy*, *52*(7), 2403–6. https://doi.org/10.1128/AAC.00090-08

Al-Nassir, W. N., Sethi, A. K., Nerandzic, M. M., Bobulsky, G. S., Jump, R. L. P., & Donskey, C. J. (2008). Comparison of clinical and microbiological response to treatment of *Clostridium difficile*-associated disease with metronidazole and vancomycin. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *47*(1), 56–62. https://doi.org/10.1086/588293

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(R106), 1–12. https://doi.org/10.1186/gb-2010-11-10-r106

Avery, O. T. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation By a Deoxyribonucleic Acid Fraction Isolated From Pneumococcus Type Iii. *Journal of Experimental Medicine*, *79*(2), 137–158. https://doi.org/10.1084/jem.79.2.137

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K. I., Oguchi, A., … Hiramatsu, K. (2002). Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet*, *359*(9320), 1819–1827. https://doi.org/10.1016/S0140-6736(02)08713-5

Babakhani, F., Gomez, A., Robert, N., & Sears, P. (2011). Killing kinetics of fidaxomicin and its major metabolite, OP-1118, against *Clostridium difficile*. *Journal of Medical Microbiology*, *60*(8), 1213–1217. https://doi.org/10.1099/jmm.0.029470-0

Babakhani, F., Seddon, J., & Sears, P. (2014). Comparative microbiological studies of transcription inhibitors fidaxomicin and the rifamycins in *Clostridium difficile*. *Antimicrobial Agents and Chemotherapy*, *58*(5), 2934–2937. https://doi.org/10.1128/AAC.02572-13

Bartlett, J. G. (2008). Historical perspectives on studies of *Clostridium difficile* and *C. difficile* infection. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *46 Suppl 1*(Suppl 1), S4-11. https://doi.org/10.1086/521865

Bauer, M. P., Notermans, D. W., Van Benthem, B. H., Brazier, J. S., Wilcox, M. H., Rupnik, M., … Kuijper, E. J. (2011). *Clostridium difficile* infection in Europe: A hospital-based survey. *The Lancet*, *377*(9759), 63–73. https://doi.org/10.1016/S0140-6736(10)61266-4

Begley, M., Gahan, C. G. M., & Hill, C. (2005). The interaction between bacteria and bile. *FEMS Microbiology Reviews*, *29*(4), 625–51. https://doi.org/10.1016/j.femsre.2004.09.003

Bengmark, S. (1998). Ecological control of the gastrointestinal tract. The role of probiotic flora. *Gut*, *42*(1), 2–7. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1726957&tool=pmcentrez&rendertype=abstract

Bhowmik, S., Jones, D. H., Chiu, H.-P., Park, I.-H., Chiu, H.-J., Axelrod, H. L., … Lesley, S. a. (2014). Structural and functional characterization of BaiA, an enzyme involved in secondary bile acid synthesis in human gut microbe. *Proteins*, *82*(2), 216–29. https://doi.org/10.1002/prot.24353

Bhowmik, S., Chiu, H. P., Jones, D. H., Chiu, H. J., Miller, M. D., Xu, Q., … Lesley, S. A. (2016). Structure and functional characterization of a bile acid 7α dehydratase BaiE in secondary bile acid synthesis. *Proteins: Structure, Function and Bioinformatics*, *84*(3), 316–331. https://doi.org/10.1002/prot.24971

Bickle, T. A., & Kruger, D. H. (1993). Biology of DNA restriction. *Microbiol Rev*, *57*(2), 434–450. https://doi.org/10.1111/j.1365-2958.1995.tb02255.x

Bidet, P., Barbut, F., Lalande, V., Burghoffer, B., & Petit, J.-C. (1999). Development of a new PCR-ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. *FEMS Microbiology Letters*, *175*(2), 261–266. https://doi.org/10.1111/j.1574-6968.1999.tb13629.x

Black, E. P., Koziol-dube, K., Guan, D., Setlow, B., Cortezzo, D. E., Dallas, G., … Hoover, D. G. (2005). Factors Influencing Germination of Bacillus subtilis Spores via Activation of Nutrient Receptors by High Pressure Factors Influencing Germination of Bacillus subtilis Spores via Activation of Nutrient Receptors by High Pressure. *Applied and Environmental Microbiology*, *71*(10), 5879–5887. https://doi.org/10.1128/AEM.71.10.5879

Bocci, V. (1992). The neglected organ: bacterial flora has a crucial immunostimulatory role. *Perspectives in Biology and Medicine*, *35*(2), 251–260. https://doi.org/10.1353/pbm.1992.0004

Brandis, G., Wrande, M., Liljas, L., & Hughes, D. (2012). Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. *Molecular Microbiology*, *85*(1), 142–151. https://doi.org/10.1111/j.1365-2958.2012.08099.x

Brazier J S, Patel B, Pearson A. (2007). Distribution of *Clostridium difficile* PCR ribotype 027 in British hospitals. Euro Surveill. 2007; 12(17):pii=3182. https://doi.org/10.2807/esw.12.17.03182-en

Brazier, J. S., Raybould, R., Patel, B., Duckworth, G., Pearson, A., Charlett, A., & Duerden, B. I. (2008). Distribution and antimicrobial susceptibility patterns of Clostridium difficle PCR ribotypes in english hospitals, 2007-08. *Surveilance and Outbreak Reports*, 1–5.

Buendgens, L., Bruensing, J., Matthes, M., Dückers, H., Luedde, T., Trautwein, C., … Koch, A. (2014). Administration of proton pump inhibitors in critically ill medical patients is associated with increased risk of developing *Clostridium difficile*-associated diarrhea. *Journal of Critical Care*. https://doi.org/10.1016/j.jcrc.2014.03.002

Buffie, C. G., Bucci, V., Stein, R. R., McKenney, P. T., Ling, L., Gobourne, A., … Pamer, E. G. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature*, *517*(7533), 205–208. https://doi.org/10.1038/nature13828

Burns, D. A., & Minton, N. P. (2011). Sporulation studies in *Clostridium difficile*. *Journal of Microbiological Methods*, *87*(2), 133–138. https://doi.org/10.1016/j.mimet.2011.07.017

Burns, D. a., Heap, J. T., & Minton, N. P. (2010). *Clostridium difficile* spore germination: An update. *Research in Microbiology*, *161*(9), 730–734. https://doi.org/10.1016/j.resmic.2010.09.007

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., & Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*, *104*(6), 901–912. https://doi.org/10.1016/S0092-8674(01)00286-0

Cappiello, A., Famiglini, G., Palma, P., & Trufelli, H. (2010). Matrix effects in liquid chromatography - mass spectrometry. *Journal of Liquid Chromatography & Related Technologies*, *33*(9–12), 1067–1081. https://doi.org/10.1080/10826076.2010.484314

Carter, G. P., Douce, G. R., Govind, R., Howarth, P. M., Mackin, K. E., Spencer, J., … Lyras, D. (2011). The anti-sigma factor tcdc modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of *Clostridium difficile*. *PLoS Pathogens*, *7*(10). https://doi.org/10.1371/journal.ppat.1002317

Cartman, S. T., Heap, J. T., Kuehne, S. a, Cockayne, A., & Minton, N. P. (2010). The emergence of "hypervirulence" in *Clostridium difficile*. *International Journal of Medical Microbiology : IJMM*, *300*(6), 387–95. https://doi.org/10.1016/j.ijmm.2010.04.008

Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T., & Minton, N. P. (2012). Precise Manipulation of the *Clostridium difficile* Chromosome Reveals a Lack of Association between the *tcdC* Genotype and Toxin Production. *Applied and Environmental Microbiology*, *78*(13), 4683–4690. https://doi.org/10.1128/AEM.00249-12

Cato, E. P., Stackebrandt, E (2012). Taxonomy and Phylogeny. In: Biotechnology Handbooks, Vol. 3: Clostridia, pp. 1-26. (Ed.: Minton, N. P. and Clarke, D. J.). Plenum Press, New York (1989)

Cerone-McLernon, A. M., Winter, J., Mosbach, E. H., & Bokkenheuser, V. D. (1981). Side-chain cleavage of cortisol by fecal flora. *Biochimica et Biophysica Acta (BBA)/Lipids and Lipid Metabolism*, *666*(3), 341–347. https://doi.org/10.1016/0005-2760(81)90292-7

Chen, I., & Dubnau, D. (2003). DNA transport during transformation. *Frontiers in Bioscience*, *8*(August), 544–556. https://doi.org/10.2741/1047

Chen, I., & Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nature Reviews Microbiology*, *2*(3), 241–249. https://doi.org/10.1038/nrmicro844

Chopra I (2007). Bacterial RNA polymerase: a promising target for the discovery of new antimicrobial agents. Curr. Opin. Investig. Drugs. 2007;8:600–607. https://www.ncbi.nlm.nih.gov/pubmed/17668362

Cocanour, C. S. (2011). Best Strategies in Recurrent or Persistent *Clostridium difficile* Infection. *Surgical Infections*, *12*(3), 235–239. https://doi.org/10.1089/sur.2010.080

Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., … Farrow, J. A. (1994). The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol*, *44*(4), 812–826. https://doi.org/10.1099/00207713-44-4-812

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Hsu, P. D., … Marraffini, L. A. (2013). Multiplex Genome Engineering Using CRISPR/VCas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143.Multiplex

Cornely, O. a, Miller, M. a, Louie, T. J., Crook, D. W., & Gorbach, S. L. (2012). Treatment of first recurrence of *Clostridium difficile* infection: fidaxomicin versus vancomycin. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *55 Suppl 2*(Suppl 2), S154-61. https://doi.org/10.1093/cid/cis462

Cornely, O. A., Crook, D. W., Esposito, R., Poirier, A., Somero, M. S., Weiss, K., … Gorbach, S. (2012). Fidaxomicin versus vancomycin for infection with *Clostridium difficile* in Europe, Canada, and the USA: A double-blind, non-inferiority, randomised controlled trial. *The Lancet Infectious Diseases*, *12*(4), 281–289. https://doi.org/10.1016/S1473-3099(11)70374-7

Coronelli, C., White, R., Lancini, G., & Parenti, F. (1975). Lipiarmycin, a new antibiotic from Actinoplanes. II. Isolation, chemical, biological and biochemical characterization. *J. Antibiot.*, *XXVIII*(4), 253–259. https://doi.org/10.7164/antibiotics.28.253

Corver, J., Bakker, D., Brouwer, M. S. M., Harmanus, C., Hensgens, M. P., Roberts, A. P., … van Leeuwen, H. C. (2012). Analysis of a *Clostridium difficile* PCR ribotype 078 100 kilobase island reveals the presence of a novel transposon, Tn6164. *BMC Microbiology*, *12*(1), 130. https://doi.org/10.1186/1471-2180-12-130

Cramer, P., Bushnell, D. A., & Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 ångstrom resolution. *Science*, *292*(5523), 1863–1876. https://doi.org/10.1126/science.1059493

Crampton, N., Roes, S., Dryden, D. T. F., Rao, D. N., Edwardson, J. M., & Henderson, R. M. (2007). DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. *EMBO Journal*, *26*(16), 3815–3825. https://doi.org/10.1038/sj.emboj.7601807

Credito, K. L., & Appelbaum, P. C. (2004). Activity of OPT-80, a novel macrocycle, compared with those of eight other agents against selected anaerobic species. *Antimicrobial Agents and Chemotherapy*, *48*(11), 4430–4434. https://doi.org/10.1128/AAC.48.11.4430-4434.2004

Dang, U. T., Zamora, I., Hevener, K. E., Adhikari, S., Wu, X., & Hurdle, J. G. (2016). Rifamycin resistance in *Clostridium difficile* is generally associated with a low fitness burden. *Antimicrobial Agents and Chemotherapy*, *60*(9), 5604–5607. https://doi.org/10.1128/AAC.01137-16

Davies, K. A., Longshaw, C. M., Davis, G. L., Bouza, E., Barbut, F., Barna, Z., … Wilcox, M. H. (2014). Underdiagnosis of *Clostridium difficile* across Europe:

The European, multicentre, prospective, biannual, point-prevalence study of *Clostridium difficile* infection in hospitalised patients with diarrhoea (EUCLID). *The Lancet Infectious Diseases*, *14*(12), 1208–1219. https://doi.org/10.1016/S1473-3099(14)70991-0

Davies, K. A., Ashwin, H., Longshaw, C. M., Burns, D. A., Davis, G. L., Wilcox, M. H., & group, on behalf of the E. study. (2016). Diversity of *Clostridium difficile* PCR ribotypes in Europe: results from the European, multicentre, prospective, biannual, point-prevalence study of *Clostridium difficile* infection in hospitalised patients with diarrhoea (EUCLID), 2012 and 2013. *Eurosurveillance*, *21*(29), 30294. https://doi.org/10.2807/1560-7917.ES.2016.21.29.30294

De Smet, I., Van Hoorde, L., Vande Woestyne, M., Christiaens, H., & Verstraete, W. (1995). Significance of bile salt hydrolytic activities of lactobacilli. *The Journal of Applied Bacteriology*, *79*(3), 292–301. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7592123

Deneve, C., Delomenie, C., Barc, M., Collignon, A., and Janoir, C. (2008). Antibiotics involved in Clostridium difficile - associated disease increase colonization factor gene expression. Journal of Medical Microbiology, 732–738. https://doi.org/10.1099/jmm.0.47676-0

Dimitri M. Drekonja; Mary Butler; Roderick MacDonald; Donna Bliss; Gregory A. Filice; Thomas S. Rector; and Timothy J. Wilt (2011). Comparative Effectiveness of *Clostridium difficile* Treatments. *Annals of Internal Medicine*, *155*, 839–847. https://www.ncbi.nlm.nih.gov/pubmed/22184691

Dubnau, D., & Cirgiliano, C. (1971). Fate of transforming DNA following uptake by competent Bacillus subtilis - VI. Non-covalent association of donor and recipient DNA. *MGG Molecular & General Genetics*, *120*(2), 101–106. https://doi.org/10.1007/BF00267237

Dupuy, B., Govind, R., Antunes, a, & Matamouros, S. (2008). *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. *Journal of Medical Microbiology*, *57*(Pt 6), 685–9. https://doi.org/10.1099/jmm.0.47775-0

Ebel-Tsipis, J., Botstein, D., & Fox, M. S. (1972). Generalized transduction by phage P22 in Salmonella typhimurium. I. Molecular origin of transducing DNA. *Journal of Molecular Biology*, *71*(2), 433–448. https://doi.org/10.1016/0022-2836(72)90361-0

Egan, S. M. (2002). Growing Repertoire of AraC / XylS Activators. *Journal of Bacteriology*, *184*(20),5529–5532. https://doi.org/10.1128/JB.184.20.5529

Eiseman B, Silen W, Bascom GS, Kauvar AJ. 1958. Fecal Enema as an Adjunct in the Treatment of Pseudomembranous Enterocolitis. Surgery 44:854–859.

Elsayed, S. and Zhang, K. (2004). Human Infection Caused *by Clostridium hathewayi*. Emerging Infectious Diseases, 10(11), 10–12. http://doi.org/10.3201/eid1011.040006

Eriksson, H., & Gustafsson, J.-A. (1971). Excretion of Steroid Hormones in Adults. *European Journal of Biochemistry*, *18*, 146–150.

Fimlaid, K. a, Bond, J. P., Schutz, K. C., Putnam, E. E., Leung, J. M., Lawley, T. D., & Shen, A. (2013). Global analysis of the sporulation pathway of *Clostridium difficile*. *PLoS Genetics*, *9*(8), e1003660. https://doi.org/10.1371/journal.pgen.1003660

Finegold, S. M., Molitoris, D., Vaisanen, M., Song, Y., Liu, C., Bolaños, M., & Bolan, M. (2004). In Vitro Activities of OPT-80 and Comparator Drugs against Intestinal Bacteria In Vitro Activities of OPT-80 and Comparator Drugs against Intestinal Bacteria. *Antimicrobial Agents and Chemotherapy*, *48*(12), 4898–4902. https://doi.org/10.1128/AAC.48.12.4898

Francis, M. B., Allen, C. a, Shrestha, R., & Sorg, J. a. (2013). Bile acid recognition by the *Clostridium difficile* germinant receptor, CspC, is important for establishing infection. *PLoS Pathogens*, *9*(5), e1003356. https://doi.org/10.1371/journal.ppat.1003356

Freedberg, D. E., Salmasian, H., Friedman, C., & Abrams, J. a. (2013). Proton pump inhibitors and risk for recurrent *Clostridium difficile* infection among inpatients. *The American Journal of Gastroenterology*, *108*(11), 1794–801. https://doi.org/10.1038/ajg.2013.333

Fu, Y. H. F., Tsai, M. M., Luo, Y., & Deonier, R. C. (1991). Deletion analysis of the F plasmid oriT locus. *Journal of Bacteriology*, *173*(3), 1012–1020.

Fujitani, S., George, W. L., Morgan, M. a, Nichols, S., & Murthy, a R. (2011). Implications for vancomycin-resistant Enterococcus colonization associated with *Clostridium difficile* infections. *American Journal of Infection Control*, *39*(3), 188–93. https://doi.org/10.1016/j.ajic.2010.10.024

Fukuda, E., Kaminska, K. H., Bujnicki, J. M., & Kobayashi, I. (2008). Cell death upon epigenetic genome methylation: a novel function of methyl-specific deoxyribonucleases. *Genome Biology*, *9*(11), R163. https://doi.org/10.1186/gb-2008-9-11-r163

Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K., & Ramos, J. L. (1997). Arac/XylS family of transcriptional regulators. *Microbiology and Molecular Biology Reviews : MMBR*, *61*(4), 393–410. https://doi.org/9409145

Galperin, M. Y., Mekhedov, S. L., Puigbo, P., Smirnov, S., Wolf, Y. I., & Rigden, D. J. (2012). Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environmental*

Microbiology, *14*(11), 2870–90. https://doi.org/10.1111/j.1462-2920.2012.02841.x

Gamage, S., Patton, A. K., Hanson, J. F., & Weiss, A. A. (2004). Diversity and Host Range of Shiga Diversity and Host Range of Shiga Toxin-Encoding Phage. *Society*, *72*(12), 7131–7139. https://doi.org/10.1128/IAI.72.12.7131

Gentry, D. R., & Burgess, R. R. (1993). Cross-Linking of Escherichia coli RNA Polymerase Subunits: Identification of B' as the Binding Site of gamma. *Biochemistry*, *32*, 11224–11227.

Gerding, D. N., Johnson, S., Rupnik, M., & Aktories, K. (2014). *Clostridium difficile* binary toxin CDT: mechanism, epidemiology, and potential clinical importance. *Gut Microbes*, *5*(1), 15–27. https://doi.org/10.4161/gmic.26854

Girinathan, B. P., Monot, M., Boyle, D., McAllister, K. N., Sorg, J. A., Dupuy, B., & Govind, R. (2017). Effect of *tcdR* Mutation on Sporulation in the Epidemic *Clostridium difficile* Strain R20291. *mSphere*, *2*(1), e00383-16. https://doi.org/10.1128/mSphere.00383-16

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., & Kornberg, R. D. (2001). Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 ?? resolution. *Science*, *292*(5523), 1876–1882. https://doi.org/10.1126/science.1059495

Goldenberg, S. D., Brown, S., Edwards, L., Gnanarajah, D., Howard, P., Jenkins, D., … Whitney, L. (2016). The impact of the introduction of fidaxomicin on the management of *Clostridium difficile* infection in seven NHS secondary care hospitals in England: a series of local service evaluations. *European Journal of Clinical Microbiology and Infectious Diseases*, *35*(2), 251–259. https://doi.org/10.1007/s10096-015-2538-z

Goldstein, E. J. C., Citron, D. M., Sears, P., Babakhani, F., Sambol, S. P., & Gerding, D. N. (2011). Comparative susceptibilities to fidaxomicin (OPT-80) of isolates collected at baseline, recurrence,and failure from patients in two phase III trials of fidaxomicin against *Clostridium difficile* infection. *Antimicrobial Agents and Chemotherapy*, *55*(11), 5194–5199. https://doi.org/10.1128/AAC.00625-11

Goorhuis, A., Bakker, D., Corver, J., Debast, S. B., Harmanus, C., Notermans, D. W., … Kuijper, E. J. (2008). Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *47*(9), 1162–70. https://doi.org/10.1086/592257

Goorhuis, A., Debast, S. B., van Leengoed, L. a M. G., Harmanus, C., Notermans, D. W., Bergwerff, A. a, & Kuijper, E. J. (2008). *Clostridium difficile* PCR ribotype 078: an emerging strain in humans and in pigs? *Journal of Clinical*

Microbiology, *46*(3), 1157; author reply 1158. https://doi.org/10.1128/JCM.01536-07

Gough, E., Shaikh, H., & Manges, A. R. (2011). Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent *Clostridium difficile* infection. *Clinical Infectious Diseases*, *53*(10), 994–1002. https://doi.org/10.1093/cid/cir632

Govind, R., & Dupuy, B. (2012). Secretion of Clostridium difficile toxins A and B requires the holin-like protein TcdE. *PLoS Pathogens*, *8*(6), 1–14. https://doi.org/10.1371/journal.ppat.1002727

Govind, R., Fitzwater, L., & Nichols, R. (2015). Observations on the role of TcdE isoforms in *Clostridium difficile* toxin secretion. *Journal of Bacteriology*, *197*(15), 2600–2609. https://doi.org/10.1128/JB.00224-15

Griffith, F. (1928). The Significance of Pneumococcal Types. *Journal of Hygiene*, *27*(2), 113–159. https://doi.org/10.1017/S0022172400031879

Gruber, T. M., & Gross, C. A. (2003). Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Review of Microbiology*, *57*(1), 441–466. https://doi.org/10.1146/annurev.micro.57.030502.090913

Gualtieri, M., Villain-guillot, P., Leonetti, J., & Bastide, L. (2006). Mutation in the Bacillus subtilis RNA Polymerase β ′ Subunit Confers Resistance to Lipiarmycin Mutation in the Bacillus subtilis RNA Polymerase ̃Subunit Confers Resistance to Lipiarmycin. *Antimicrobial Agents and Chemotherapy*, *50*(1), 401–403. https://doi.org/10.1128/AAC.50.1.401

Gualtieri, M., Tupin, A., Brodolin, K., & Leonetti, J. P. (2009). Frequency and characterisation of spontaneous lipiarmycin-resistant Enterococcus faecalis mutants selected in vitro. *International Journal of Antimicrobial Agents*, *34*(6), 605–606. https://doi.org/10.1016/j.ijantimicag.2009.07.003

Gurtler, V. (1993). Typing of *Clostridium difficile* strains by PCR-amplification of variable length 16 S-23 S rDNA spacer regions. *Journal of General Microbiology*, *139*(12), 3089–3097. https://doi.org/10.1099/00221287-139-12-3089

Hamilton, M. J., Weingarden, A. R., Sadowsky, M. J., & Khoruts, A. (2012). Standardized frozen preparation for transplantation of fecal microbiota for recurrent *Clostridium difficile* infection. *American Journal of Gastroenterology*, *107*(5), 761–767. https://doi.org/10.1038/ajg.2011.482

Handa, N., & Kobayashi, I. (1999). Post-segregational killing by restriction modification gene complexes: Observations of individual cell deaths. *Biochimie*, *81*(8–9), 931–938. https://doi.org/10.1016/S0300-9084(99)00201-1

Hardison, W. G. M. (1978). Hepatic Taurine Concentration and Dietary Taurine as Regulators of Bile Acid Conjugation with Taurine. *Gastroenterology*, *75*(1), 71–75. https://doi.org/10.1016/0016-5085(78)93767-8

Harris, S. C. (2017). Discovery and Characterization of Bile Acid and Steroid Metabolism Pathways in Gut-Associated Microbes, 1–240. Retrieved from http://scholarscompass.vcu.edu/cgi/viewcontent.cgi?article=5771&context=etd

Healthcare Commission (2007) available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/250569/0913.pdf

Hedges, R. W., & Jacob, A. E. (1974). Transposition of ampicillin resistance from RP4 to other replicons. *MGG Molecular & General Genetics*, *132*(1), 31–40. https://doi.org/10.1007/BF00268228

Hedges, R. W. (1975). R factors from Proteus mirabilis and P. vulgaris. *Journal of General Microbiology*, *87*(2), 301–311. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=1095684&retmode=ref&cmd=prlinks

Heeg, D., Burns, D. a., Cartman, S. T., & Minton, N. P. (2012). Spores of *Clostridium difficile* clinical isolates display a diverse germination response to bile salts. *PLoS ONE*, *7*(2), e32381. https://doi.org/10.1371/journal.pone.0032381

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

Henriques, A. O., & Moran, C. P. (2007). Structure, assembly, and function of the spore surface layers. *Annual Review of Microbiology*, *61*, 555–88. https://doi.org/10.1146/annurev.micro.61.080706.093224

Hensgens, M. P. M., Keessen, E. C., Squire, M. M., Riley, T. V., Koene, M. G. J., de Boer, E., … Kuijper, E. J. (2012). *Clostridium difficile* infection in the community: a zoonotic disease? *Clinical Microbiology and Infection*, *18*(7), 635–645. https://doi.org/10.1111/j.1469-0691.2012.03853.x

Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of General Physiology*, *36*(1), 39–56. https://doi.org/10.1085/jgp.36.1.39

Higgins, D., & Dworkin, J. (2012). Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiology Reviews*, *36*(1), 131–148. https://doi.org/10.1111/j.1574-6976.2011.00310.x

Ho, M. X., Hudson, B. P., Das, K., Arnold, E., & H, R. (2010). NIH Public Access. *October*, *19*(6), 715–723. https://doi.org/10.1016/j.sbi.2009.10.010.Structures

Hofmann, A. F., AF, H., DM, H., AF, H., (eds.), H. A., (eds.), H. A., … C, G.-K. (1999). The Continuing Importance of Bile Acids in Liver and Intestinal Disease. *Archives of Internal Medicine*, *159*(22), 2647. https://doi.org/10.1001/archinte.159.22.2647

Hooper, L. V. (2001). Commensal Host-Bacterial Relationships in the Gut. *Science*, *292*(5519), 1115–1118. https://doi.org/10.1126/science.1058709

Hudson, K. D., Corfe, B. M., Kemp, E. H., Feavers, I. M., Coote, P. J., & Moir, a. (2001). Localization of GerAA and GerAC germination proteins in the Bacillus subtilis spore. *Journal of Bacteriology*, *183*(14), 4317–4322. https://doi.org/10.1128/JB.183.14.4317

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., … White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234

Ishihama, A. (1992). Role of the RNA polymerase alpha subunit in transcription activation. *Mol Microbiol*, *6*(22), 3283–3288. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1484484

Jalajakumari, M. B., Guidolin, A., Buhk, H. J., Manning, P. A., Ham, L. M., Hodgson, A. L. M., … Skurray, R. A. (1987). Surface exclusion genes traS and traT of the F sex factor of Escherichia coli K-12. Determination of the nucleotide sequence and promoter and terminator activities. *Journal of Molecular Biology*, *198*(1), 1–11. https://doi.org/10.1016/0022-2836(87)90452-9

James, T. D., Cashel, M., & Hinton, D. M. (2010). A mutation within the β subunit of Escherichia coli RNA polymerase impairs transcription from bacteriophage T4 middle promoters. *Journal of Bacteriology*, *192*(21), 5580–5587. https://doi.org/10.1128/JB.00338-10

Janscak, P., MacWilliams, M. P., Sandmeier, U., Nagaraja, V., & Bickle, T. A. (1999). DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO Journal*, *18*(9), 2638–2647. https://doi.org/10.1093/emboj/18.9.2638

Janulaitis, A., Petrusyte, M., Maneliene, Z., Klimasauskas, S., & Butkus, V. (1992). Purification and properties of the Eco57I restriction endonuclease and methylase--prototypes of a new class (type IV). *Nucleic Acids Research*, *20*(22), 6043–9. https://doi.org/10.1093/nar/20.22.6043

Jeltsch, A., & Pingoud, A. (1996). Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *Journal of Molecular Evolution*, *42*(2), 91–96. https://doi.org/10.1007/BF02198833

Jiang, W., Maniv, I., Arain, F., Wang, Y., Levin, B. R., & Marraffini, L. A. (2013). Dealing with the Evolutionary Downside of CRISPR Immunity: Bacteria and Beneficial Plasmids. *PLoS Genetics*, *9*(9). https://doi.org/10.1371/journal.pgen.1003844

Jiang, H., Lei, R., Ding, S. W., & Zhu, S. (2014). Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, *15*(1), 1–12. https://doi.org/10.1186/1471-2105-15-182

Joce RE, Bruce J, Kiely D, Noah ND, Dempster WB, Stalker R, Gumsley P, Chapman PA, Norman P, Watkins J, Smith HV, Price TJ,. & Watts D (1995). An outbreak of cryptosporidiosis associated with a resort swimming pool. Epidemiology and Infection, 115(3), 545–553. https://doi.org/10.1017/S0950268800058714

Johnson, J. L., & Francis, B. S. (1975). Taxonomy of the Clostridia: Ribosomal Ribonucleic Acid Homologies among the Species. *Journal of General Microbiology*, *88*(2), 229–244. https://doi.org/10.1099/00221287-88-2-229

Jones, B. V., Begley, M., Hill, C., Gahan, C. G. M., & Marchesi, J. R. (2008). Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(36), 13580–5. https://doi.org/10.1073/pnas.0804437105

Just, I., & Gerhard, R. (2004). Large clostridial cytotoxins. *Reviews of Physiology, Biochemistry and Pharmacology*, *152*(September), 23–47. https://doi.org/10.1007/s10254-004-0033-5

Kang, D.-J., Ridlon, J. M., Moore, D. R., Barnes, S., & Hylemon, P. B. (2008). *Clostridium scindens baiCD* and *baiH* genes encode stereo-specific 7alpha/7beta-hydroxy-3-oxo-delta4-cholenoic acid oxidoreductases. *Biochimica et Biophysica Acta*, *1781*(1–2), 16–25. https://doi.org/10.1016/j.bbalip.2007.10.008

Kelly, M. L., Ng, Y. K., Cartman, S. T., Collery, M. M., Cockayne, A., & Minton, N. P. (2016). Improving the reproducibility of the NAP1/B1/027 epidemic strain R20291 in the hamster model of infection. *Anaerobe*, *39*, 51–53. https://doi.org/10.1016/j.anaerobe.2016.02.011

Kennaway, C. K., Taylor, J. E., Song, C. F., Potrzebowski, W., Nicholson, W., White, J. H., … Dryden, D. T. F. (2012). Structure and operation of the DNA-

translocating type i DNA restriction enzymes. *Genes and Development*, *26*(1), 92–104. https://doi.org/10.1101/gad.179085.111

Kent, T., Kashkina, E., Anikin, M., & Temiakov, D. (2009). Maintenance of RNA-DNA hybrid length in bacterial RNA polymerases. *Journal of Biological Chemistry*, *284*(20), 13497–13504. https://doi.org/10.1074/jbc.M901898200

Khanna S, & DS, P. (2010). The growing incidence and severity of *Clostridium difficile* infection in inpatient and outpatient settings. *Expert Review of Gastroenterology and Hepatology*, *4*(4), 409–416. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed9&NEWS=N&AN=2010420687

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kirk, J. A., & Fagan, R. P. (2016). Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe*, *42*, 1–5. https://doi.org/10.1016/j.anaerobe.2016.06.009

Kitahara, M., Takamine, F., Imamura, T., & Benno, Y. (2000). Assignment of Eubacterium sp. VPI 12708 and related strains with high bile acid 7alpha-dehydroxylating activity to *Clostridium scindens* and proposal of Clostridium hylemonae sp. nov., isolated from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, *50 Pt 3*, 971–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10843034

Kitahara, M., Sakata, S., Sakamoto, M., & Benno, Y. (2004). Comparison among fecal secondary bile acid levels, fecal microbiota and *Clostridium scindens* cell numbers in Japanese. *Microbiology and Immunology*, *48*(5), 367–75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15215624

Kuehne, S. A., Dempster, A. W., Collery, M. M., Joshi, N., Jowett, J., Kelly, M. L., … Minton, N. P. (2017). Characterization of the impact of *rpoB* mutations on the in vitro and in vivo competitive fitness of *Clostridium difficile* and susceptibility to fidaxomicin. Journal of Antimicrobial Chemotherapy, 973–980. https://doi.org/10.1093/jac/dkx486

Kuehne, S. a, Cartman, S. T., Heap, J. T., Kelly, M. L., Cockayne, A., & Minton, N. P. (2010). The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature*, *467*(7316), 711–713. https://doi.org/10.1038/nature09397

Kuehne, S. A., Collery, M. M., Kelly, M. L., Cartman, S. T., Cockayne, A., & Minton, N. P. (2013). The importance of toxin A in virulence of an epidemic *Clostridium difficile* strain. *Journal of Infectious Diseases*, *1*(0), 1–16.

Kurabachew, M., Lu, S. H. J., Krastel, P., Schmitt, E. K., Suresh, B. L., Goh, A., … Sambandamurthy, V. K. (2008). Lipiarmycin targets RNA polymerase and has good activity against multidrug-resistant strains of Mycobacterium tuberculosis. *Journal of Antimicrobial Chemotherapy*, *62*(4), 713–719. https://doi.org/10.1093/jac/dkn269

Lane, W. J., & Darst, S. A. (2010). Molecular Evolution of Multisubunit RNA Polymerases: Structural Analysis. *Journal of Molecular Biology*, *395*(4), 686–704. https://doi.org/10.1016/j.jmb.2009.10.063

Lanis, J. M., Heinlen, L. D., James, J. a., & Ballard, J. D. (2013). *Clostridium difficile* 027/BI/NAP1 Encodes a Hypertoxic and Antigenically Variable Form of TcdB. *PLoS Pathogens*, *9*(8), e1003523. https://doi.org/10.1371/journal.ppat.1003523

Larson, HE; Borriello, S. P. (1990). Quantitative Study of Antibiotic-Induced Susceptibility. *Antimicrobial Agents and Chemotherapy*, *34*(7), 1348–1353.

Lawley, T. D., Clare, S., Walker, A. W., Stares, M. D., Connor, T. R., Raisen, C., Dougan, G. (2012). Targeted Restoration of the Intestinal Microbiota with a Simple, Defined Bacteriotherapy Resolves Relapsing *Clostridium difficile* Disease in Mice. PLoS Pathogens, 8(10). https://doi.org/10.1371/journal.ppat.1002995

Lawson, P. A., Citron, D. M., Tyrrell, K. L., & Finegold, S. M. (2016). Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe*, *40*, 95–99. https://doi.org/10.1016/j.anaerobe.2016.06.008

Leeds, J. A., Sachdeva, M., Mullin, S., Whitney Barnes, S., & Ruzin, A. (2014). In vitro selection, via serial passage, of *Clostridium difficile* mutants with reduced susceptibility to fidaxomicin or vancomycin. *Journal of Antimicrobial Chemotherapy*, *69*(1), 41–44. https://doi.org/10.1093/jac/dkt302

Lewis, B. B., Buffie, C. G., Carter, R. A., Leiner, I., Toussaint, N. C., Miller, L. C., … Pamer, E. G. (2015). Loss of Microbiota-Mediated Colonization Resistance to *Clostridium difficile* Infection with Oral Vancomycin Compared with Metronidazole. *Journal of Infectious Diseases*, *212*(10), 1656–1665. https://doi.org/10.1093/infdis/jiv256

Liddle, R. A. (1997). Cholecystokinin Cells. *Annual Review of Physiology*, *59*(1), 221–242. https://doi.org/10.1146/annurev.physiol.59.1.221

Louie, T. J., Emery, J., Krulicki, W., Byrne, B., & Mah, M. (2009). OPT-80 eliminates *Clostridium difficile* and is sparing of Bacteroides species during treatment of *C. difficile* infection. *Antimicrobial Agents and Chemotherapy*, *53*(1), 261–263. https://doi.org/10.1128/AAC.01443-07

Louie, T. J., Miller, M. A., Mullane, K. M., Weiss, K., Lentnek, A., Golan, Y., … Shue, Y.-K. (2011). Fidaxomicin versus Vancomycin for *Clostridium difficile* Infection. N Engl J Med, 364, 422–431. https://doi.org/10.1056/NEJMoa0910812

Luckey, T. D. (1972). Introduction to Intestinal Microecology. The American Journal of Clinical Nutrition, 25(22), 1292–1294.

Luckey, T. D. (2008). Sir samurai T. D. luckey, PhD. Dose-Response, 6(1), 97–112. https://doi.org/10.2203/dose-response.06-019.Luckey

Lyras, D., O'Connor, J. R., Howarth, P. M., Sambol, S. P., Carter, G. P., Phumoonna, T., … Rood, J. I. (2009). Toxin B is essential for virulence of *Clostridium difficile*. Nature, 458(7242), 1176–1179. https://doi.org/10.1038/nature07822

Mallonee, D. H., & Hylemon, P. B. (1996). Sequencing and expression of a gene encoding a bile acid transporter from Eubacterium sp. strain VPI 12708. Journal of Bacteriology, 178(24), 7053–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8955384

Mallonee, D. H., Adams, J. L., & Hylemon, P. B. (1992). The bile acid-inducible baiB gene from Eubacterium sp. strain VPI 12708 encodes a bile acid-coenzyme A ligase. Journal of Bacteriology, 174(7), 2065–2071. https://doi.org/10.1128/jb.174.7.2065-2071.1992

Mani, N., & Dupuy, B. (2001). Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. Proceedings of the National Academy of Sciences of the United States of America, 98(10), 5844–5849. https://doi.org/10.1073/pnas.101126598

EJ, Marraffini & LA and Sontheimer (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. Science. 2008 December 19; 322(5909): 1843–1845. doi:10.1126/science.1165771

Marsh, J. W., Arora, R., Schlackman, J. L., Shutt, K. A., Curry, S. R., & Harrison, L. H. (2012). Association of relapse of *Clostridium difficile* disease with BI/NAP1/027. *Journal of Clinical Microbiology*, *50*(12), 4078–4082. https://doi.org/10.1128/JCM.02291-12

McCune, V. L., Struthers, J. K., & Hawkey, P. M. (2014). Faecal transplantation for the treatment of *Clostridium difficile* infection: a review. International Journal of Antimicrobial Agents, 43(3), 201–6. https://doi.org/10.1016/j.ijantimicag.2013.10.009

Miller, B. A., Chen, L. F., Sexton, D. J., & Anderson, D. J. (2011). Comparison of the Burdens of Hospital-Onset, Healthcare Facility-Associated *Clostridium difficile* Infection and of Healthcare-Associated Infection Due to Methicillin-Resistant Staphylococcus aureus in Community Hospitals. Infection Control

and Hospital Epidemiology: The Official Journal of the Society of Hospital Epidemiologists of America, 32(4), 387–390. https://doi.org/10.1086/659156

Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M., … Zhang, Y. (2016). A roadmap for gene system development in Clostridium. Anaerobe, 41, 104–112. https://doi.org/10.1016/j.anaerobe.2016.05.011

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of Molecular Evolution, 60(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3

Morris, G. N., Winter, J., Cato, E. P., Ritchie, A. E., & Bokkenheuser, V. D. (1985). *Clostridium scindens* sp. nov., a Human Intestinal Bacterium with Desmolytic Activity on Corticoids. International Journal of Systematic Bacteriology, 35(4), 478–481. https://doi.org/10.1099/00207713-35-4-478

Moser, S. A., & Savage, D. C. (2001). Bile Salt Hydrolase Activity and Resistance to Toxicity of Conjugated Bile Salts are Unrelated Properties in Lactobacilli. Applied and Environmental Microbiology, 67(8), 3476–3480. https://doi.org/10.1128/AEM.67.8.3476-3480.2001

Mukhopadhyay, J., Das, K., Ismail, S., Koppstein, D., Jang, M., Hudson, B., … Ebright, R. H. (2008). The RNA Polymerase "Switch Region" Is a Target for Inhibitors. Cell, 135(2), 295–307. https://doi.org/10.1016/j.cell.2008.09.033

Murakami, K., Kimura, M., Owens, J. T., Meares, C. F., & Ishihama, A. (1997). The two alpha subunits of Escherichia coli RNA polymerase are asymmetrically arranged and contact different halves of the DNA upstream element. Proceedings of the National Academy of Sciences, 94(5), 1709–1714. https://doi.org/10.1073/pnas.94.5.1709

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., & Darst, S. A. (2002). Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. Science, 296(5571), 1285–1290. https://doi.org/10.1126/science.1069595

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. https://doi.org/10.1128/AEM.02229-13

Nabarro JDN, Moxham A, Walker G, S. J. (1957). Rectal Hydrocortisone. *British Medical Journal*, *1957 Aug 3; 2(5039): 272–274.*

Nüsslein, C., & Heyden, B. (1972). Chromatography of RNA Polymerase from Escherichia Coli on Single Stranded DNA-Agarose Columns. Biochemical and Biophysical Research Communications, 47(1), 282–289.

Olling, A., Seehase, S., Minton, N. P., Tatge, H., Schröter, S., Kohlscheen, S., … Gerhard, R. (2012). Release of TcdA and TcdB from *Clostridium difficile* cdi 630 is not affected by functional inactivation of the tcdE gene. *Microbial Pathogenesis*,*52*(1),92–100. https://doi.org/10.1016/j.micpath.2011.10.009

Omura, S., Imamura, N., Oiwa, R., Kuga, H., Iwata, R., Masuma, R., & Iwai, Y. (1986). Clostomicins, new antibiotics produced by Micromonospora echinospora subsp. armeniaca subsp. nov. I. Production, isolation, and physico-chemical and biological properties. *Journal of Antibiotics*, *39*(10), 1407–1412. https://doi.org/10.7164/antibiotics.39.1407

Paidhungat, M., & Setlow, P. (2000). Role of Ger proteins in nutrient and nonnutrient triggering of spore germination in Bacillus subtilis. *Journal of Bacteriology*, *182*(9), 2513–2519. https://doi.org/10.1128/JB.182.9.2513-2519.2000

Paidhungat, M., Setlow, B., Daniels, W. B., Hoover, D., Papafragkou, E., & Setlow, P. (2002). Mechanisms of Induction of Germination of Bacillus subtilis Spores by High Pressure Mechanisms of Induction of Germination of Bacillus subtilis Spores by High Pressure, *68*(6), 1–5. https://doi.org/10.1128/AEM.68.6.3172

Panne, D., Müller, S. A., Wirtz, S., Engel, A., & Bickle, T. A. (2001). The McrBCx restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *EMBO Journal*, *20*(12), 3210–3217. https://doi.org/10.1093/emboj/20.12.3210

Papatheodorou, P., Carette, J. E., Bell, G. W., Schwan, C., Guttenberg, G., Brummelkamp, T. R., & Aktories, K. (2011). Lipolysis-stimulated lipoprotein receptor (LSR) is the host receptor for the binary toxin *Clostridium difficile* transferase (CDT). *Proceedings of the National Academy of Sciences of the United States of America*, *108*(39), 16422–7. https://doi.org/10.1073/pnas.1109772108

Paredes-Sabja, D., Shen, A., & Sorg, J. a. (2014). *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. *Trends in Microbiology*, 1–11. https://doi.org/10.1016/j.tim.2014.04.003

Parenti, F., Pagani, H., Beretta, G., & Durando, V. (1972). Lipiarmycin, A New Antibiotic from Actinoplanes. *The Journal of Antibiotics*, *XXVIII No.4.*, 247–252.

Pepin, J., Alary, M.-E., Valiquette, L., Raiche, E., Ruel, J., Fulop, K., … Bourassa, C. (2005). Increasing Risk of Relapse after Treatment of *Clostridium difficile*

Colitis in Quebec, Canada. *Clinical Infectious Diseases*, *40*(11), 1591–1597. https://doi.org/10.1086/430315

Pereira, F. C., Saujet, L., Tomé, A. R., Serrano, M., Monot, M., Couture-Tosi, E., … Henriques, A. O. (2013). The spore differentiation pathway in the enteric pathogen *Clostridium difficile*. *PLoS Genetics*, *9*(10), e1003782. https://doi.org/10.1371/journal.pgen.1003782

Permpoonpattana, P., Phetcharaburanin, J., Mikelsone, A., Dembek, M., Tan, S., Brisson, M.-C., … Cutting, S. M. (2013). Functional characterization of *Clostridium difficile* spore coat proteins. *Journal of Bacteriology*, *195*(7), 1492–503. https://doi.org/10.1128/JB.02104-12

Permpoonpattana, P., Tolls, E. H., Nadem, R., Tan, S., Brisson, A., & Cutting, S. M. (2011). Surface layers of *Clostridium difficile* endospores. *Journal of Bacteriology*, *193*(23), 6461–70. https://doi.org/10.1128/JB.05182-11

Pizarro-Guajardo, M., Olguín-Araneda, V., Barra-Carrasco, J., Brito-Silva, C., Sarker, M. R., & Paredes-Sabja, D. (2014). Characterization of the collagen-like exosporium protein, BclA1, of *Clostridium difficile* spores. *Anaerobe*, *25*, 18–30. https://doi.org/10.1016/j.anaerobe.2013.11.003

Popoff, M. R., Rubin, E. J., Gill, D. M., & Boquet, P. (1988). Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infection and Immunity*, *56*(9), 2299–2306.

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653–663. https://doi.org/10.1099/mic.0.27437-0

Purdy, D., O'Keeffe, T. A., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., … Minton, N. P. (2002). Conjugative transfer of clostridial shuttle vectors from Escherichia coli to *Clostridium difficile* through circumvention of the restriction barrier. *Mol Microbiol*, *46*(2), 439–452. https://doi.org/3134 [pii]

Rainey FA1, Stackebrandt E (1993). 16 S rDNA analysis reveals phylogenetic diversity among the Polysaccharolytic Clostridia. *FEMS Microbiol Lett. 1993 Oct 15;113(2):125-8*. https://www.ncbi.nlm.nih.gov/pubmed/7505248

Raju, D., Waters, M., Setlow, P., & Sarker, M. R. (2006). Investigating the role of small, acid-soluble spore proteins (SASPs) in the resistance of Clostridium perfringens spores to heat. BMC Microbiology, 6, 50. https://doi.org/10.1186/1471-2180-6-50

Rohman, A., Van Oosterwijk, N., Thunnissen, A. M. W. H., & Dijkstra, B. W. (2013). Crystal structure and site-directed mutagenesis of 3-ketosteroid δ1-dehydrogenase from Rhodococcus erythropolis SQ1 explain its catalytic

mechanism. *Journal of Biological Chemistry*, *288*(49), 35559–35568. https://doi.org/10.1074/jbc.M113.522771

Ridlon, J. M., & Hylemon, P. B. (2012). Identification and characterization of two bile acid coenzyme A transferases from *Clostridium scindens* , a bile acid 7α-dehydroxylating intestinal bacterium. *Journal of Lipid Research*, *53*(1), 66–76. https://doi.org/10.1194/jlr.M020313

Ridlon, J. M., Kang, D.-J., & Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research*, *47*(2), 241–259. https://doi.org/10.1194/jlr.R500013-JLR200

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., … Xu, S. Y. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research*, *31*(7), 1805–1812. https://doi.org/10.1093/nar/gkg274

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2015). REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, *43*(D1), D298–D299. https://doi.org/10.1093/nar/gku1046

Rokyta, D. R., Burch, C. L., Caudle, S. B., & Wichman, H. a. (2006). Horizontal Gene Transfer and the Evolution of Microvirid Coliphage Genomes Horizontal Gene Transfer and the Evolution of Microvirid Coliphage Genomes. *Journal of Bacteriology*, *188*(3), 1134–1142. https://doi.org/10.1128/JB.188.3.1134

Rupnik, M., Wilcox, M. H., & Gerding, D. N. (2009). *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nature Reviews. Microbiology*, *7*(7), 526–36. https://doi.org/10.1038/nrmicro2164

Rupnik, M., Pabst, S., Rupnik, M., von Eichel-Streiber, C., Urlaub, H., & Söling, H.-D. (2005). Characterization of the cleavage site and function of resulting cleavage fragments after limited proteolysis of *Clostridium difficile* toxin B (TcdB) by host cells. *Microbiology*, *151*(Pt 1), 199–208. https://doi.org/10.1099/mic.0.27474-0

Sakai, K., Makino, T., Kawai, Y., & Mutai, M. (1980). Intestinal microflora and bile acids. Effect of bile acids on the distribution of microflora and bile acid in the digestive tract of the rat. *Microbiology and Immunology*, *24*(3), 187–196.

Saujet, L., Monot, M., Dupuy, B., Soutourina, O., & Martin-Verstraete, I. (2011). The key sigma factor of transition phase, sigh, controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*. *Journal of Bacteriology*, *193*(13), 3186–3196. https://doi.org/10.1128/JB.00272-11

Sears, P., Crook, D. W., Louie, T. J., Miller, M. A., & Weiss, K. (2012). Fidaxomicin attains high fecal concentrations with minimal plasma concentrations following oral administration in patients with *Clostridium difficile* infection. *Clinical Infectious Diseases*, *55*(SUPPL.2), 116–120. https://doi.org/10.1093/cid/cis337

Sears, P., Ichikawa, Y., Ruiz, N., & Gorbach, S. (2013). Advances in the treatment of *Clostridium difficile* with fidaxomicin: A narrow spectrum antibiotic. *Annals of the New York Academy of Sciences*, *1291*(1), 33–41. https://doi.org/10.1111/nyas.12135

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., … Parkhill, J. (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genetics*, *38*(7), 779–86. https://doi.org/10.1038/ng1830

Seddon, J., Babakhani, F., Gomez, A., Artsimovitch, I., & Sears, P. (2011). RNA Polymerase Target Modification in *Clostridium difficile* with Reduced Susceptibility to Fidaxomicin. *Poster Presented at: 51st Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC); September 17-20, 2011; Chicago, IL, USA*, Abstract C1-631.

Seidel, R., Van Noort, J., Van Der Scheer, C., Bloom, J. G. P., Dekker, N. H., Dutta, C. F., … Dekker, C. (2004). Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I. *Nature Structural and Molecular Biology*, *11*(9), 838–843. https://doi.org/10.1038/nsmb816

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8), 1–14. https://doi.org/10.1371/journal.pbio.1002533

Setlow, P. (2003). Spore germination. *Current Opinion in Microbiology*, *6*(6), 550–556. https://doi.org/10.1016/j.mib.2003.10.001

Shue, Y. K., Sears, P. S., Shangle, S., Walsh, R. B., Lee, C., Gorbach, S. L., … Preston, R. A. (2008). Safety, tolerance, and pharmacokinetic studies of OPT-80 in healthy volunteers following single and multiple oral doses. *Antimicrobial Agents and Chemotherapy*, *52*(4), 1391–1395. https://doi.org/10.1128/AAC.01045-07

Smith, L. D., & King, E. O. (1962). Occurrence of *Clostridium difficile* in infections of man. *Journal of Bacteriology*, *84*, 65–67.

Snyder, M. L. (1937). Further studies on bacillus difficilis (Hall and O'Toole). *Journal of Infectious Diseases*, *60*(2), 223–231. https://doi.org/10.1093/infdis/60.2.223

Sorg, J. A., & Sonenshein, A. L. (2008). Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *Journal of Bacteriology*, *190*(7), 2505–2512. https://doi.org/10.1128/JB.01765-07

Sorg, J. A. (2014). Microbial bile acid metabolic clusters: The bouncers at the bar. *Cell Host and Microbe*, *16*(5), 551–552. https://doi.org/10.1016/j.chom.2014.10.015

Sorg, J. A., & Sonenshein, A. L. (2010). Inhibiting the initiation of *Clostridium difficile* spore germination using analogs of chenodeoxycholic acid, a bile acid. *Journal of Bacteriology*, *192*(19), 4983–4990. https://doi.org/10.1128/JB.00610-10

Sorg, J. A., & Sonenshein, A. L. (2009). Chenodeoxycholate is an inhibitor of *Clostridium difficile* spore germination. *Journal of Bacteriology*, *191*(3), 1115–1117. https://doi.org/10.1128/JB.01260-08

Srivastava, A., Talaue, M., Liu, S., Degen, D., Ebright, R. Y., Sineva, E., … Ebright, R. H. (2011). New target for inhibition of bacterial RNA polymerase:'switch region'. *Current Opinion in Microbiology*, *14*(5), 532–543. https://doi.org/10.1016/j.mib.2011.07.030.NEW

Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., … Wren, B. W. (2009). Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biology*, *10*(9), R102. https://doi.org/10.1186/gb-2009-10-9-r102

Stabler, R. A., Valiente, E., Dawson, L. F., He, M., Parkhill, J., Wren, B. W., & Brendan, E. (2010). In-depth genetic analysis of *Clostridium difficile* PCR-ribotype 027 strains reveals high genome fluidity including point mutations and inversions. *DNA Sequence*, (August), 269–276. https://doi.org/10.1186/gb-2009-10-9-r102.reviously

Stackebrandt, E., Kramer, I., Swiderski, J., & Hippe, H. (1999). Phylogenetic basis for a taxonomic dissection of the genus Clostridium. *FEMS Immunology and Medical Microbiology*, *24*(3), 253–258. https://doi.org/10.1016/S0928-8244(99)00039-5

Stalons, D. R., & Thornsberry, C. (1975). Broth-Dilution Method for Determining the Antibiotic Susceptibility of Anaerobic Bacteria. *Antimicrobial Agents and Chemotherapy*, *7*(1), 15–21. https://doi.org/10.1128/AAC.7.1.15

Strand, T. A., Lale, R., Degnes, K. F., Lando, M., & Valla, S. (2014). A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS ONE*, *9*(3), 1–6. https://doi.org/10.1371/journal.pone.0090372

Studer, N., Desharnais, L., Beutler, M., Brugiroux, S., Terrazos, M. A., Menin, L., … Hapfelmeier, S. (2016). Functional Intestinal Bile Acid 7α-Dehydroxylation by *Clostridium scindens* Associated with Protection from *Clostridium difficile* Infection in a Gnotobiotic Mouse Model. *Frontiers in Cellular and Infection Microbiology*, *6*(December), 1–15. https://doi.org/10.3389/fcimb.2016.00191

Studier, F. W., & Bandyopadhyay, P. K. (1988). Model for how type I restriction enzymes select cleavage sites in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(13), 4677–4681.

Sung, J. Y., Shaffer, E. A., & Costerton, J. W. (1993). Antibacterial activity of bile salts against common biliary pathogens. Effects of hydrophobicity of the molecule and in the presence of phospholipids. *Digestive Diseases and Sciences*, *38*(11), 2104–2112. https://doi.org/10.1007/BF01297092

Surawicz, C. M., Brandt, L. J., Binion, D. G., Ananthakrishnan, A. N., Curry, S. R., Gilligan, P. H., … Zuckerbraun, B. S. (2013). Guidelines for diagnosis, treatment, and prevention of *Clostridium difficile* infections. *American Journal of Gastroenterology*, *108*(4), 478–498. https://doi.org/10.1038/ajg.2013.4

Takahashi, N., Naito, Y., Handa, N., & Kobayashi, I. (2002). A DNA methyltransferase can protect the genome from postdisturbance attack by a restriction-modification gene complex. *Journal of Bacteriology*, *184*(22), 6100–6108. https://doi.org/10.1128/JB.184.22.6100-6108.2002

Tanner, R. S., Stackebrandt, E., Fox, G. E., & Woesell, C. R. (1981). A Phylogenetic Analysis of Acetobacterium woodii, Clostridium barkeri, Clostridium butyricum, Clostridium lituseburense, Eubacterium limosum, and Eubacterium tenue. *Current Microbiology*, *5*, 35–38. https://doi.org/10.1007/BF01566595

Tannock, G. W., Munro, K., Taylor, C., Lawley, B., Young, W., Byrne, B., … Louie, T. (2010). A new macrocyclic antibiotic, fidaxomicin (OPT-80), causes less alteration to the bowel microbiota of *Clostridium difficile*-infected patients than does vancomycin. *Microbiology*, *156*(11), 3354–3359. https://doi.org/10.1099/mic.0.042010-0

Tenover, F. C., Tickler, I. A., & Persing, D. H. (2012). Antimicrobial-resistant strains of *Clostridium difficile* from North America. *Antimicrobial Agents and Chemotherapy*, *56*(6), 2929–2932. https://doi.org/10.1128/AAC.00220-12

Thanissery, R., Winston, J. A., & Theriot, C. M. (2017). Inhibition of spore germination, growth, and toxin activity of clinically relevant *C. difficile* strains by gut microbiota derived secondary bile acids. *Anaerobe*, *45*, 86–100. https://doi.org/10.1016/j.anaerobe.2017.03.004

Tenover, F. C., Akerlund, T., Gerding, D. N., Goering, R. V, Boström, T., Jonsson, A.-M., … Persing, D. H. (2011). Comparison of strain typing results for *Clostridium difficile* isolates from North America. *Journal of Clinical Microbiology*, *49*(5), 1831–1837. https://doi.org/10.1128/JCM.02446-10

Theriault, R. J., Karwowski, J. P., Jackson, M., Girolami, R. L., Sunga, G. N., Vojtko, C. M., & Coen, L. J. (1987). Tiacumicins, a novel complex of 18-membered macrolide antibiotics. I. Taxonomy, fermentation and antibacterial activity. *The Journal of Antibiotics*, *40*(5), 567–574. https://doi.org/10.7164/antibiotics.40.567

Theriot, C. M., Bowman, A. a, & Young, V. B. (2015). Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for *Clostridium difficile* Spore Germination and Outgrowth in the Large Intestine. *American Society for Microbiology*, *1*(1), e00045-15. https://doi.org/10.1128/mSphere.00045-15.Editor

Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, *3*(9), 711–721. https://doi.org/10.1038/nrmicro1234

Underwood, S., Guan, S., Vijayasubhash, V., Baines, S. D., Graham, L., Lewis, R. J., … Stephenson, K. (2009). Characterization of the sporulation initiation pathway of *Clostridium difficile* and its role in toxin production. *Journal of Bacteriology*, *191*(23), 7296–305. https://doi.org/10.1128/JB.00882-09

Valiente, E., Cairns, M. D., & Wren, B. W. (2014). The *Clostridium difficile* PCR ribotype 027 lineage: a pathogen on the move. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 1–9. https://doi.org/10.1111/1469-0691.12619

Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., & Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature*, *417*(6890), 712–719. https://doi.org/10.1038/nature752

Voth, D. E., & Ballard, J. D. (2005). *Clostridium difficile* toxins: mechanism of action and role in disease. *Clinical Microbiology Reviews*, *18*(2), 247–263. https://doi.org/10.1128/CMR.18.2.247

Wade, A. P., Slater, J. D. H., Kellie, A. E., Holliday, M. E., (1959). Urinary Excretion of 17-Ketosteroids Following Rectal Infusion of Cortisol *The Journal of Clinical Endocrinology & Metabolism, Volume 19, Issue 4, 1 April 1959, Pages 444–453*. http://doi.org.https.sci-hub.tw/10.1210/jcem-19-4-444

Warny, M., Pepin, J., Fang, A., Killgore, G., Thompson, A., Brazier, J., … McDonald, L. C. (2005). Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe.

Lancet, *366*(9491), 1079–1084. https://doi.org/10.1016/S0140-6736(05)67420-X

Weingarden, A., González, A., Vázquez-Baeza, Y., Weiss, S., Humphry, G., Berg-Lyons, D., … Sadowsky, M. J. (2015). Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent *Clostridium difficile* infection. *Microbiome*, *3*(1), 1–8. https://doi.org/10.1186/s40168-015-0070-0

White, B. A., Cacciapuoti, A. F., Fricke, R. J., Whitehead, T. R., Mosbach, E. H., & Hylemon, P. B. (1981). Cofactor requirements for 7a-dehydroxylation e of cholic and chenodeoxycholic acid in cell extracts of the intestinal anaerobic bacterium , Eubacterium species V.P.I. 12708. *Journal of Lipid Research*, *22*, 891–898.

Wilson, K. H., Kennedy, M. J., & Fekety, F. R. (1982). Use of sodium taurocholate to enhance spore recovery on a medium selective for *Clostridium difficile*. *Journal of Clinical Microbiology*, *15*(3), 443–446.

Wion, D., & Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nature Reviews. Microbiology*, *4*(3), 183–192. https://doi.org/10.1038/nrmicro1350

Woese, C. R. (1988). Bacterial Evolution. *Canadian Journal of Microbiology*, *34*(4), 547–551. https://doi.org/10.1139/m88-093

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., … Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16 S rRNA gene sequences. *Nature Reviews Microbiology*, *12*(9), 635–645. https://doi.org/10.1038/nrmicro3330

Yoshimoto, S., Loo, T. M., Atarashi, K., Kanda, H., Sato, S., Oyadomari, S., … Ohtani, N. (2013). Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature*, *499*(7456), 97–101. https://doi.org/10.1038/nature12347

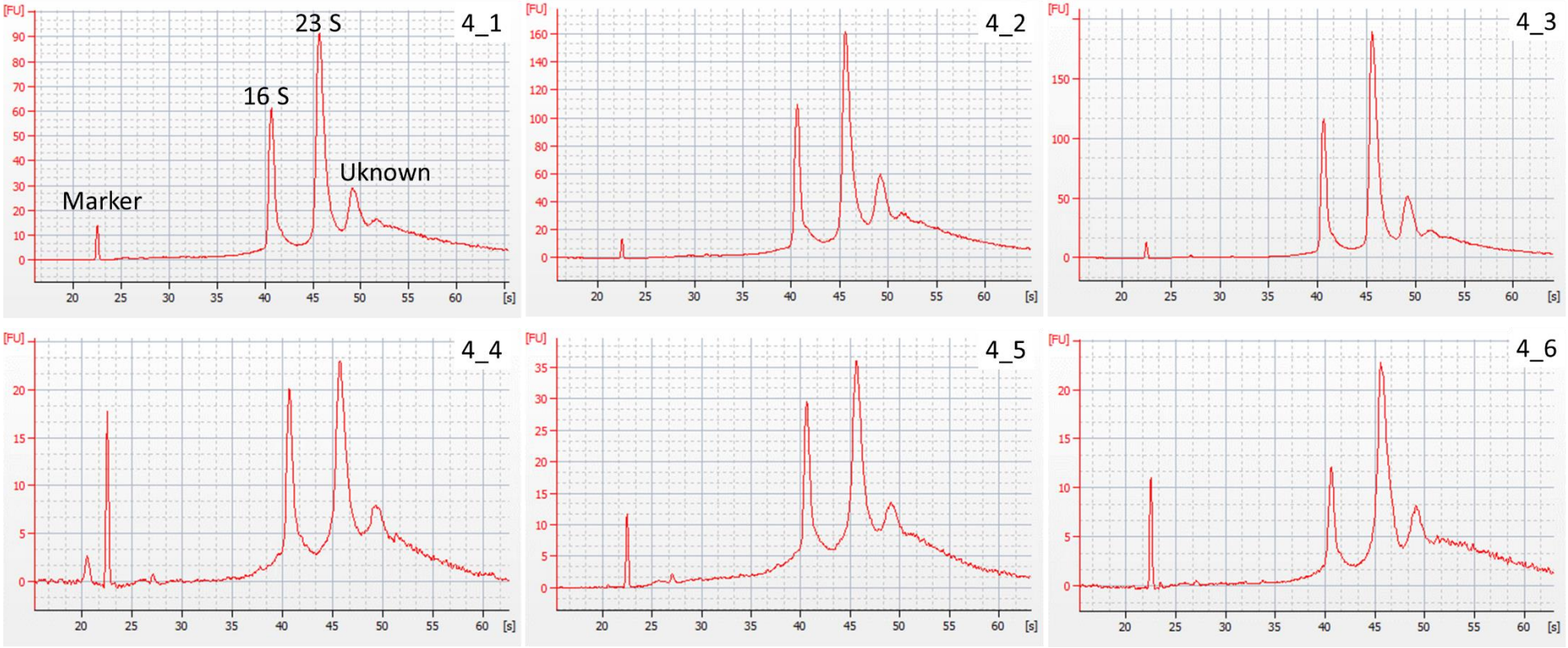
Youngster, I., Sauk, J., Pindar, C., Wilson, R. G., Kaplan, J. L., Smith, M. B., … Hohmann, E. L. (2014). Fecal Microbiota Transplant for Relapsing Clostridium difficile Infection Using a Frozen Inoculum From Unrelated Donors: A Randomized, Open-Label, Controlled Pilot Study. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, 1–8. https://doi.org/10.1093/cid/ciu135

Yutin, N., Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore-formers and other misplaced clostridia. *Environmental Microbiology*, *15*(10), 2631–2641. https://doi.org/10.1111/1462-2920.12173.A

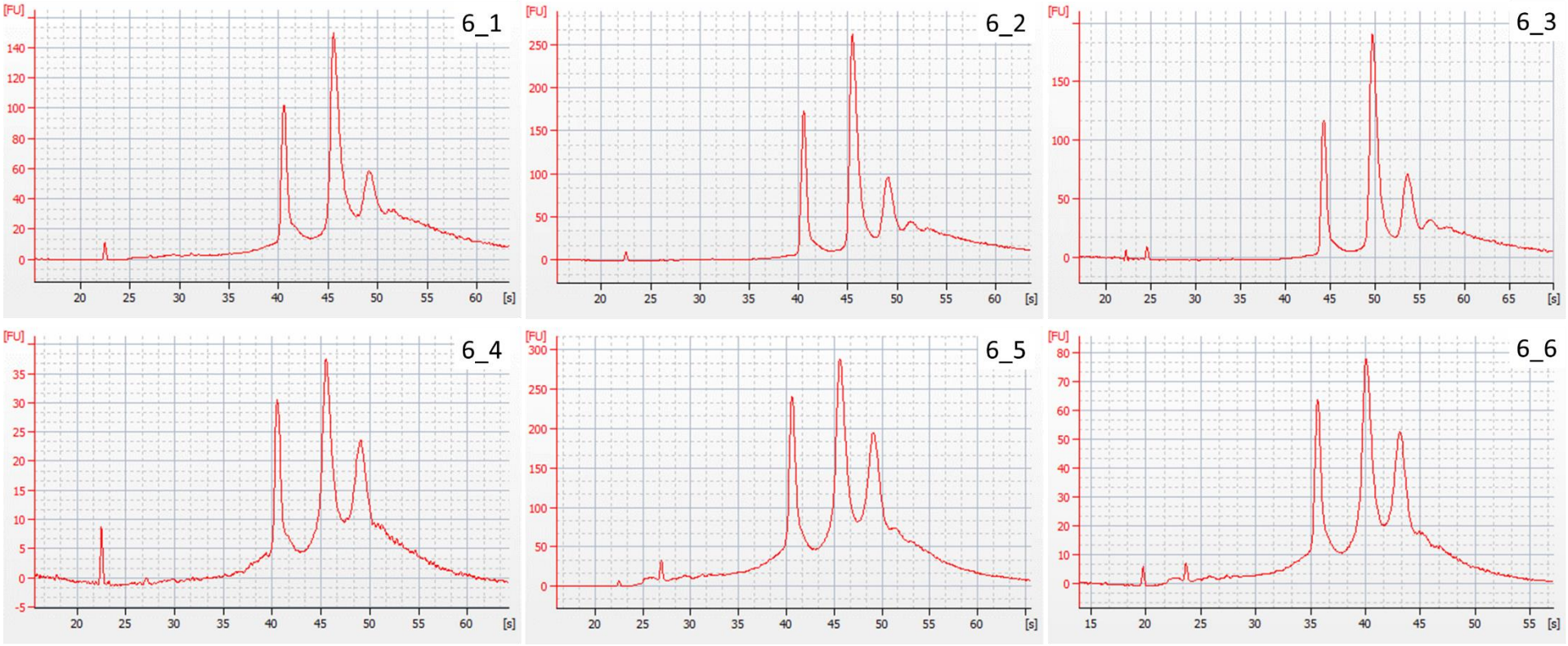
8 Appendices

8.1 Bioanalyser Electropherograms

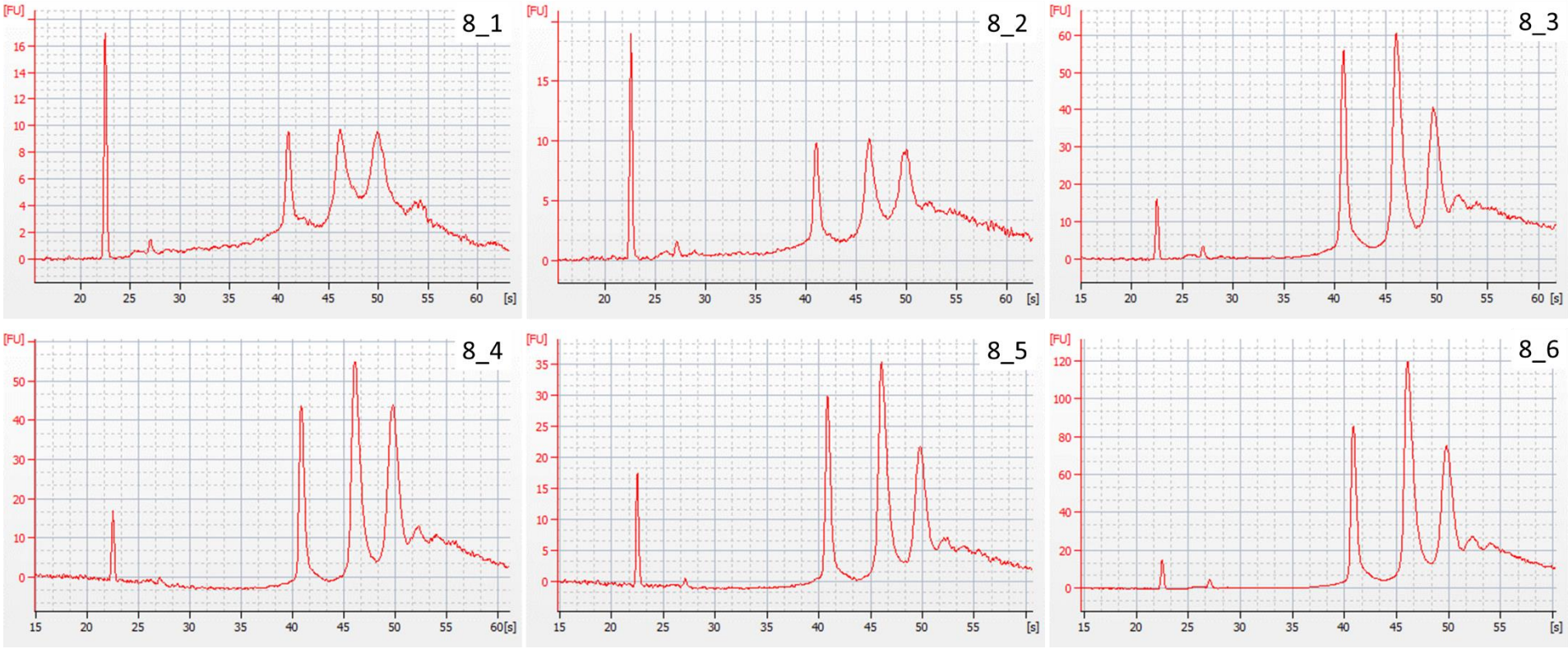
8.1.1 Electropherograms for RNA Extrations at 4 Hours



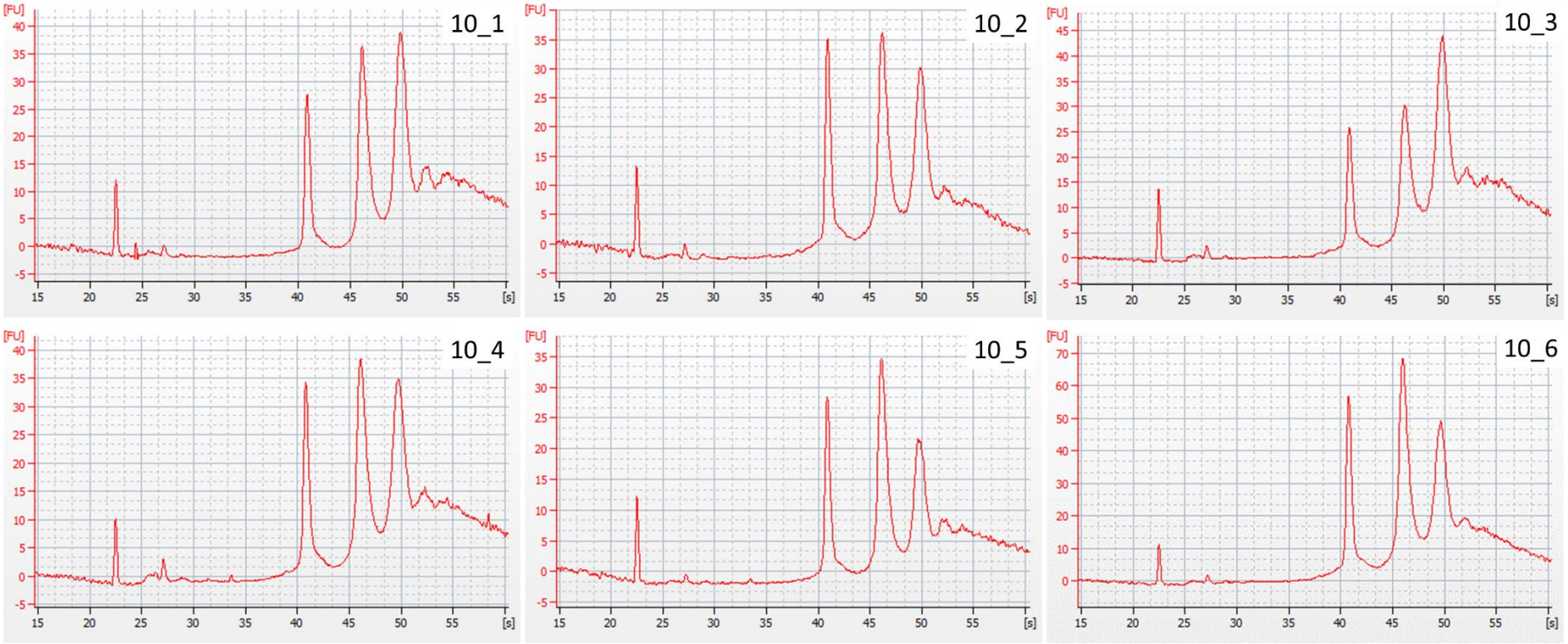
8.1.2 Electropherograms for RNA Extrations at 6 Hours



8.1.3 Electropherograms for RNA Extrations at 8 Hours



8.1.4 Electropherograms for RNA Extrations at 10 Hours



8.2 Cholic Acid Induced Differenial Gene Expression

8.2.1 Differential Gene Expression: 4 Hours

GeneID	GeneName	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Gene 3027	baiB	4881.139829	8.818449206	0.338676545	2.60E+01	1.84E-149	3.24E-146
Gene 1825	baiA1	9756.489561	8.477766862	0.321018154	2.64E+01	1.08E-153	3.80E-150
Gene 3025	baiCD	6365.700284	8.192443776	0.34386972	2.38E+01	1.87E-125	2.20E-122
Gene 3023	baiA2	5763.300364	8.088847423	0.343587448	2.35E+01	1.50E-122	1.32E-119
Gene 3022	baiF	10061.73764	7.945267048	0.344617102	2.31E+01	1.30E-117	9.15E-115
Gene 3024	baiE	2869.893999	7.925116054	0.359864295	2.20E+01	1.75E-107	1.03E-104
Gene 1824	baiA1	3773.833515	7.847933831	0.357748714	2.19E+01	1.15E-106	5.80E-104
Gene 3026	baiCD	2980.311425	7.806117577	0.367709072	2.12E+01	5.15E-100	2.27E-97
Gene 3020	baiH	15035.0227	7.403324088	0.353096617	2.10E+01	1.32E-97	5.15E-95
Gene 3021	baiG	6283.799125	7.315443128	0.355816735	2.06E+01	6.32E-94	2.22E-91

Gene 3019	bail	4031.388297	6.82208578	0.379902492	1.80E+01	4.20E-72	1.23E-69
Gene 719	FAD binding flavoprotein	3832.249667	5.144973141	0.273923706	1.88E+01	1.05E-78	3.36E-76
Gene 2325	Predicted sialic acid transporter	9031.944938	4.688600181	0.348286052	1.35E+01	2.62E-41	6.59E-39
Gene 2324	FIG00617520: hypothetical protein	2681.551799	4.681191417	0.397943363	1.18E+01	6.02E-32	1.41E-29
Gene 2327	N-acetylmannosamine-6-phosphate 2-epimerase (EC	2441.204726	4.579780117	0.327475046	1.40E+01	1.92E-44	5.20E-42
Gene 1822	hypothetical protein	45.2927309	4.199274302	0.505772284	8.30E+00	1.02E-16	1.99E-14
Gene 2326	N-acetylneuraminate lyase (EC 4.1.3.3)	6387.046562	3.745997004	0.334246749	1.12E+01	3.76E-29	8.26E-27
Gene 2323	FIG00516798: hypothetical protein	794.3124995	3.706356688	0.337788491	1.10E+01	5.19E-28	1.07E-25
Gene 3456	COG1272: Predicted membrane protein hemolysin III	829.7215995	2.690160768	0.412032098	6.53E+00	6.62E-11	8.04E-09
Gene 1297	Drug resistance transporter EmrB/QacA subfamily	1587.387871	2.669486492	0.333911663	7.99E+00	1.30E-15	2.41E-13
Gene 3017	hypothetical protein	19.81293411	2.555032989	0.517318747	4.94E+00	7.85E-07	4.69E-05
Gene 3738	NTF2 protein	1575.730992	2.548993831	0.349136601	7.30E+00	2.86E-13	4.57E-11
Gene 3659	Transmembrane component BL0694 of energizing	236.7954296	2.515611481	0.483666515	5.20E+00	1.98E-07	1.39E-05
Gene 1298	Permease of the drug/metabolite transporter (DMT)	726.5332704	2.398187221	0.31537528	7.60E+00	2.87E-14	4.80E-12
Gene 3660	Duplicated ATPase component BL0693 of energizing	346.9810156	2.364576807	0.495346676	4.77E+00	1.81E-06	9.95E-05

Gene 3016	FIG00518880: hypothetical protein	41.00900464	2.321699902	0.454178828	5.11E+00	3.19E-07	2.12E-05
Gene 3658	Substrate-specific component BL0695 of predicted	145.3208601	2.210922535	0.508525266	4.35E+00	1.38E-05	0.000590517
Gene 3749	Predicted erythritol ABC transporter 2,	2194.919419	2.146159557	0.316535765	6.78E+00	1.20E-11	1.63E-09
Gene 3662	ABC transporter ATP-binding protein	185.1455877	2.12654749	0.518824967	4.10E+00	4.15E-05	0.001353701
Gene 3745	Unknown pentose isomerase TM0951	2172.135171	2.057055696	0.339685373	6.06E+00	1.40E-09	1.49E-07
Gene 2071	Permease of the drug/metabolite transporter (DMT)	2905.024371	2.026804443	0.257351729	7.88E+00	3.39E-15	5.97E-13
Gene 3751	Ribose ABC transport system, permease protein	3294.481188	1.975798713	0.297368823	6.64E+00	3.05E-11	3.83E-09
Gene 3743	Ribose operon repressor	213.2090519	1.911796651	0.370101184	5.17E+00	2.40E-07	1.65E-05
Gene 1845	1-phosphofructokinase (EC 2.7.1.56)	47603.51212	1.909704766	0.28442434	6.71E+00	1.89E-11	2.46E-09
Gene 1844	PTS system, fructose-specific IIA component (EC	102807.6941	1.879909862	0.295638613	6.36E+00	2.03E-10	2.24E-08
Gene 3657	Substrate-specific component BL0695 of predicted	128.7797422	1.874068974	0.511251819	3.665647541	0.000246714	0.005591838
Gene 2561	hypothetical protein	57.27647064	1.773313962	0.406991681	4.36E+00	1.32E-05	0.000579839
Gene 4091	FIG00525551: hypothetical protein	212.4794133	1.752792047	0.394415446	4.44E+00	8.83E-06	0.000408929
Gene 2072	hypothetical protein	610.4801341	1.734080474	0.269730864	6.43E+00	1.29E-10	1.51E-08
Gene 3655	Tetracenomycin polyketide synthesis	7245.834057	1.703135048	0.497580492	3.422833241	0.000619721	0.011671301

Gene 3654	ABC transporter ATP-binding/transmembrane	22540.70483	1.650792848	0.398690944	4.14E+00	3.47E-05	0.001184157
Gene 3754	Ribokinase (EC 2.7.1.15)	70.49311135	1.650103774	0.393855236	4.19E+00	2.79E-05	0.001023521
Gene 492	hypothetical protein	157.1788949	1.614085974	0.384006134	4.20E+00	2.63E-05	0.000974754
Gene 3746	Transketolase, N-terminal section (EC 2.2.1.1)	1399.288316	1.610476187	0.423477475	3.80297 9576	0.000142966	0.003871083
Gene 3747	Transketolase, C-terminal section (EC 2.2.1.1)	1737.336604	1.605141945	0.383896566	4.18E+00	2.90E-05	0.001023521
Gene 1551	Substrate-specific component RibU of riboflavin	588.5568588	1.591138134	0.347695253	4.58E+00	4.73E-06	0.000238052
Gene 3652	FIG00519028: hypothetical protein	28573.97	1.590654477	0.24920574	6.38E+00	1.74E-10	1.97E-08
Gene 3750	Ribose ABC transport system, ATP-binding protein	3584.430074	1.583741026	0.374711658	4.23E+00	2.37E-05	0.000890409
Gene 2273	Phage portal protein	65.40011972	1.569083465	0.378482668	4.15E+00	3.39E-05	0.001169007
Gene 3663	ABC transporter ATP-binding protein	67.04498347	1.562209896	0.522188441	2.99165 9281	0.002774657	0.039022473
Gene 2760	hypothetical protein	42481.38207	1.561635391	0.312051683	5.00E+00	5.60E-07	3.50E-05
Gene 3753	Ribokinase (EC 2.7.1.15)	492.1368523	1.561181093	0.34834186	4.48E+00	7.40E-06	0.000352159
Gene 2065	hypothetical protein	2196.100557	1.556994794	0.311685247	5.00E+00	5.87E-07	3.56E-05
Gene 3100	hypothetical protein	1508.750169	1.55419743	0.288604533	5.39E+00	7.24E-08	5.79E-06
Gene 2067	Permease of the drug/metabolite transporter (DMT)	2341.713215	1.538279834	0.303628376	5.07E+00	4.06E-07	2.60E-05

Gene 1842	Phosphoenolpyruvate-protein phosphotransferase of	90930.95817	1.526232257	0.262164257	5.82E+00	5.83E-09	5.86E-07
Gene 2068	N-acetyl-L,L-diaminopimelate deacetylase (EC	8227.76253	1.518780187	0.272091657	5.58E+00	2.38E-08	2.15E-06
Gene 3744	lacI-family transcriptional regulator	254.0595815	1.517901706	0.408090819	3.719519371	0.000199602	0.004845516
Gene 2353	hypothetical protein	1757.250061	1.514982676	0.316984843	4.78E+00	1.76E-06	9.83E-05
Gene 2270	Phage capsid protein	92.60900374	1.5037379	0.374790612	4.01E+00	6.02E-05	0.00185737
Gene 2274	Phage terminase, large subunit	136.5972016	1.500452423	0.336994534	4.45E+00	8.49E-06	0.000398439
Gene 1550	FIG00261907: hypothetical protein	360.2710033	1.495941343	0.391606498	3.820011547	0.000133445	0.003727999
Gene 3667	no significant homology.	6432.394084	1.437015033	0.299364252	4.80E+00	1.58E-06	9.00E-05
Gene 3755	Glycerol kinase (EC 2.7.1.30)	1359.284693	1.426782623	0.372834636	3.826851064	0.000129793	0.003654971
Gene 3653	FIG00519028: hypothetical protein	9474.244499	1.426676536	0.310673366	4.59E+00	4.39E-06	0.00022374
Gene 3748	Ribose ABC transport system, periplasmic	6262.662421	1.421960901	0.382490331	3.717638816	0.000201093	0.004848282
Gene 599	FIG00518951: hypothetical protein	1499.592091	1.411292179	0.259502127	5.44E+00	5.37E-08	4.40E-06
Gene 3455	FIG01031768: hypothetical protein	114.7270463	1.404962756	0.336069263	4.18E+00	2.91E-05	0.001023521
Gene 2759	Deoxyribose operon repressor, DeoR family	2511.395174	1.375250459	0.335214395	4.10E+00	4.09E-05	0.001343962
Gene 2810	Osmosensitive K+ channel histidine kinase KdpD	47.12148635	1.351356129	0.413325676	3.269470559	0.001077489	0.018501283

Gene 2122	Alanine dehydrogenase (EC 1.4.1.1)	13918.18943	1.34702912	0.314044327	4.29E+00	1.79E-05	0.000717009
Gene 4092	FIG00525551: hypothetical protein	11414.74802	1.331588524	0.352263254	3.780094882	0.000156769	0.004118101
Gene 2752	Mucin 2 precursor	2529.130312	1.311251574	0.397204716	3.301198399	0.000962728	0.016859711
Gene 1391	hypothetical protein	684.6042656	1.305102464	0.35614727	3.664502225	0.00024782	0.005591838
Gene 2063	hypothetical protein	1156.042223	1.288776111	0.299657609	4.30E+00	1.70E-05	0.00069647
Gene 2753	hypothetical protein	629.5551846	1.288327437	0.418953973	3.075104952	0.002104285	0.031122197
Gene 2758	Deoxyribose operon repressor, DeoR family	4660.096778	1.284379435	0.33474257	3.836916934	0.000124589	0.003565462
Gene 1846	Transcriptional repressor of the fructose operon,	10447.72873	1.284175328	0.336590928	3.81524046	0.00013605	0.003747309
Gene 2756	3-oxoacyl-[acyl-carrier protein] reductase (EC	3214.895573	1.283194504	0.363126151	3.533743024	0.000409719	0.008483598
Gene 2748	hypothetical protein	723.7378628	1.267418658	0.361831031	3.502791493	0.00046041	0.009156171
Gene 4044	D-proline reductase, 45 kDa subunit (EC 1.21.4.1)	562.7993037	1.260750545	0.305656639	4.12E+00	3.71E-05	0.00124797
Gene 118	hypothetical protein	1126.286573	1.2402456	0.362359451	3.422694222	0.000620038	0.011671301
Gene 4042	D-proline reductase, 26 kDa subunit	28043.53906	1.236599962	0.346557535	3.568238566	0.000359389	0.007761043
Gene 2066	membrane transport protein	6595.402529	1.233189344	0.377699589	3.265000488	0.001094639	0.018704509
Gene 3878	ABC transporter, substrate-binding protein	1928.589676	1.229566285	0.264055354	4.66E+00	3.22E-06	0.000168999

Gene 2754	ABC-type transport system involved in	2370.556041	1.226871892	0.395234548	3.104161566	0.001908191	0.028827606
Gene 3881	Ribose ABC transport system, permease protein	140.7318649	1.225887918	0.321346313	3.814849802	0.000136266	0.003747309
Gene 2761	Deoxyribose-phosphate aldolase (EC 4.1.2.4)	27531.61891	1.222509826	0.282809138	4.32E+00	1.54E-05	0.000653552
Gene 1443	Glycine/sarcosine/betaine reductase protein A	27545.96836	1.211521757	0.395829804	3.060713835	0.0022081	0.032193024
Gene 2623	ABC transporter permease protein	2852.192061	1.200817308	0.3285305	3.655116669	0.000257065	0.005763497
Gene 428	hypothetical protein	210.4189218	1.198409933	0.323743015	3.701732174	0.000214133	0.005024978
Gene 2755	ABC-type multidrug transport system, ATPase	4219.847239	1.182180678	0.37938823	3.116018332	0.001833109	0.028054534
Gene 2276	FIG00513960: hypothetical protein	89.782006079	1.181255031	0.384946746	3.068619341	0.002150504	0.03165863
Gene 3084	extracellular solute-binding protein, family 1	8251.357637	1.170153066	0.2639653	4.43E+00	9.29E-06	0.000424865
Gene 3376	hypothetical protein	1114.328879	1.165341312	0.379898826	3.067504378	0.002158543	0.03165863
Gene 1843	Phosphocarrier protein of PTS system	5111.960157	1.158674419	0.369221306	3.13815698	0.001700138	0.026363373
Gene 3537	hypothetical protein	608.6614655	1.155073007	0.285559592	4.04E+00	5.23E-05	0.001674729
Gene 2320	N-acetylglucosamine-6-phosphate deacetylase (EC	2256.785718	1.149582203	0.254292089	4.52E+00	6.16E-06	0.000301307
Gene 2747	Ribose ABC transport system, ATP-binding protein	5654.461552	1.14812408	0.364445474	3.150331562	0.001630853	0.025489425
Gene 1777	FIG00530205: hypothetical protein	2632.474157	1.144747159	0.392092947	2.919581102	0.003505022	0.045262636

Gene 3795	Predicted erythritol ABC transporter 1,	119.8358614	1.143476728	0.343046506	3.333299442	0.000858225	0.015334786
Gene 146	Ribonucleotide reductase of class III	20445.41088	1.141469148	0.329034714	3.469145046	0.000522117	0.010098096
Gene 3536	Peptidase M23/M37	1279.935424	1.131578481	0.264482129	4.28E+00	1.88E-05	0.000744276
Gene 3651	hypothetical protein	6215.111018	1.127184431	0.319342333	3.529705634	0.000416022	0.008563733
Gene 2123	transcriptional regulator, RpiR family	1001.80117	1.114339457	0.266258469	4.19E+00	2.85E-05	0.001023521
Gene 3405	Transcriptional regulator, AraC family	2631.327139	1.110093264	0.284807384	3.90E+00	9.71E-05	0.002872536
Gene 2321	Sialic acid utilization regulator, RpiR family	1229.315414	1.095671609	0.259263244	4.23E+00	2.38E-05	0.000890409
Gene 1036	Iron-sulfur cluster assembly scaffold protein	9306.812339	1.088767104	0.338282806	3.218511511	0.001288578	0.021073445
Gene 2749	Ribose ABC transport system, periplasmic	6030.090777	1.085803912	0.335041861	3.240800744	0.001191945	0.019855054
Gene 1449	Glycine reductase component B gamma subunit (EC	11237.55249	1.080993193	0.336231833	3.215023342	0.001304339	0.021157944
Gene 2025	Transcriptional regulator, MarR family	845.8796372	1.0797049	0.288231433	3.745965134	0.000179702	0.004550716
Gene 2261	Phage tail length tape-measure protein	97.84660335	1.073319134	0.346570437	3.096972566	0.00195508	0.029409745
Gene 1158	FIG00512775: hypothetical protein	998.0447097	1.073241296	0.287439857	3.733794288	0.000188617	0.004668052
Gene 2062	Diacylglycerol kinase (EC 2.7.1.107)	465.4821326	1.070201606	0.279594191	3.827696145	0.000129348	0.003654971
Gene 3669	FIG00264004: hypothetical protein	3244.197755	1.06938106	0.282733017	3.782299893	0.000155386	0.004118101

Gene 3879	ABC-type sugar transport system, ATP-binding	458.0788887	1.064597487	0.302726437	3.516698102	0.00043695	0.008839458
Gene 2356	hypothetical protein	87.1294094	1.052357018	0.359433888	2.927818029	0.003413498	0.044337681
Gene 3260	Probable cation-transporting ATPase	1817.133088	1.04157658	0.308493124	3.376336454	0.00073458	0.013328467
Gene 1454	Na(+)-linked D-alanine glycine permease	16756.74517	1.035420841	0.302883489	3.418545009	0.000629569	0.011787676
Gene 4045	D-proline reductase, 45 kDa subunit (EC 1.21.4.1)	1620.059914	1.034709861	0.279687302	3.699523908	0.000216004	0.005035044
Gene 3756	Transcriptional regulator, XRE family	1185.089157	1.028517988	0.289044604	3.558336584	0.000373211	0.007913869
Gene 4036	FIG00530766: hypothetical protein	7659.589533	1.025212544	0.306837487	3.341223241	0.000834101	0.014979778
Gene 3883	Ribose ABC transport system, permease protein	99.41750829	1.019792856	0.347799833	2.932125782	0.003366504	0.043889233
Gene 1577	FIG00527654: hypothetical protein	685.8500078	1.014426502	0.275183153	3.686368481	0.000227477	0.005233456
Gene 4133	Substrate-specific component RibU of riboflavin	1040.078959	1.007285817	0.342649893	2.939693946	0.003285366	0.043151074
Gene 1447	Thioredoxin reductase (EC 1.8.1.9)	6087.147225	1.00134085	0.310391673	3.226055779	0.001255089	0.020693148
Gene 1451	Glycine reductase component B beta subunit (EC	8371.000007	0.995960066	0.305930666	3.255509095	0.001131893	0.018972678
Gene 3506	hypothetical protein	838.8259081	0.980744616	0.321311942	3.052312992	0.002270852	0.032894643
Gene 3463	Negative regulator of genetic competence MecA	15252.37696	0.979695102	0.317768674	3.083044937	0.002048942	0.030431545
Gene 1445	Thioredoxin	4802.353591	0.979014784	0.32782307	2.986412099	0.00282272	0.039136988

Gene 2322	N-acetylmannosamine kinase (EC 2.7.1.60)	1438.419451	0.97675927	0.256705006	3.804987227	0.000141812	0.003869585
Gene 2442	hypothetical protein	1980.498514	0.971665153	0.260330277	3.732432376	0.00018964	0.004668052
Gene 4153	Na+/H+ antiporter NhaC	548.9418006	0.955735569	0.294991378	3.239876278	0.001195816	0.019855054
Gene 4037	Proline racemase (EC 5.1.1.4)	62094.38952	0.946913468	0.278989158	3.394086972	0.000688578	0.012623934
Gene 2384	CRISPR-associated helicase Cas3	5572.232683	0.944650182	0.264667643	3.569194076	0.000358081	0.007761043
Gene 3728	hypothetical protein	1111.816871	0.933027488	0.273272128	3.414279735	0.000639509	0.011910428
Gene 1452	GrdX protein	1909.383164	0.92854225	0.31829167	2.917268459	0.003531117	0.045363256
Gene 598	GTP-sensing transcriptional pleiotropic repressor	8536.351931	0.924324264	0.29118916	3.17430863	0.001501939	0.023707736
Gene 42	Lysine 2,3-aminomutase (EC 5.4.3.2)	1865.291594	0.916371087	0.272210154	3.366410379	0.000761533	0.013746651
Gene 3082	N-Acetyl-D-glucosamine ABC transport system,	1556.673598	0.901234209	0.254470007	3.541612698	0.000397689	0.008283225
Gene 3541	Sortase A, LPXTG specific	324.9206304	0.872667857	0.302550286	2.884372936	0.00392194	0.049659096
Gene 858	hypothetical protein	1942.668701	0.862673762	0.286724725	3.00871772	0.002623527	0.037237164
Gene 385	Ribosomal subunit interface protein	14793.56487	0.838628568	0.286008616	2.93217939	0.003365923	0.043889233
Gene 3628	DNA-binding response regulator	1725.570391	0.833634609	0.252965157	3.295452304	0.000982634	0.017123124
Gene 1575	FIG00518389: hypothetical protein	1417.802458	0.789548586	0.272767286	2.894586798	0.003796579	0.04824534

Gene 2700	Transcriptional regulator, AraC family	3282.183149	0.787125975	0.263290218	2.989575458	0.002793654	0.039022473
Gene 1579	hypothetical protein	2894.308985	0.786129168	0.256903826	3.060013469	0.00221327	0.032193024

8.2.2 Differential Gene Expression: 6 Hours

GeneID	GeneName	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Gene 3027	baiB	21277.85163	10.50809838	0.172009559	61.09020021	0	0
Gene 1824	baiA1	12462.85354	9.960758962	0.188362686	52.88074402	0	0
Gene 3022	biaiF	36259.5105	9.912646448	0.136692007	72.51811346	0	0
Gene 3021	baiG	26472.53143	9.773914827	0.135458199	72.154472	0	0
Gene 1825	baiA1	29838.51252	9.566384274	0.138431032	69.10577889	0	0
Gene 3026	baiCD	17222.01403	9.178767355	0.224126884	40.95344205	0	0
Gene 3023	baiA2	20687.59745	9.03806203	0.231068 05	39.11428702	0	0
Gene 3025	baiCD	27327.7833	8.963828283	0.230440522	38.89866334	0	0
Gene 3020	baiH	50842.2482	8.645320538	0.207872266	41.589558149	0	0
Gene 3024	baiE	10809.46853	7.757380859	0.27624262	2.81E+01	1.64E-173	2.47E-171
Gene 3019	baiI	11825.87419	7.334610624	0.269273061	2.72E+01	2.27E-163	3.27E-161
Gene 719	FAD binding flavoprotein	13961.1462	6.900036206	0.10837871	63.665977475	0	0
Gene 1297	Drug resistance transporter EmrB/QacA subfamily	7228.741293	5.27930829	0.102662156	51.424093352	0	0
Gene 1822	hypothetical protein	133.0396777	4.718538069	0.307363785	1.54E+01	3.45E-53	3.82E-51

Gene 3738	NTF2 protein	4542.322369	4.716982575	0.122629987	38.46516403	0	0
Gene 1298	Permease of the drug/metabolite transporter (DMT)	3375.147308	4.649416532	0.110044164	42.25045968	0	0
Gene 3856	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	16980.83426	3.96019823	0.329491736	1.20E+01	2.82E-33	1.87E-31
Gene 3016	FIG00518880: hypothetical protein	126.8885936	3.786387161	0.28108572	1.35E+01	2.33E-41	1.93E-39
Gene 3858	Ribose 5-phosphate isomerase B (EC 5.3.1.6)	11626.38344	3.72077661	0.10419706	3.57E+01	2.86E-279	5.00E-277
Gene 2071	Permease of the drug/metabolite transporter (DMT)	7741.344844	3.554876775	0.079649956	44.63124569	0	0
Gene 3854	5-keto-D-gluconate 5-reductase (EC 1.1.1.69)	13034.16661	3.453484136	0.335883474	1.03E+01	8.51E-25	4.03E-23
Gene 2072	hypothetical protein	1587.95142	3.3990228	0.113394235	3.00E+01	2.06E-197	3.42E-195
Gene 3017	hypothetical protein	60.9364732	3.39428371	0.326140757	1.04E+01	2.29E-25	1.19E-23
Gene 3855	6-phosphogluconate dehydrogenase, decarboxylating	32161.53981	3.350578572	0.329191385	1.02E+01	2.48E-24	1.16E-22
Gene 2067	Permease of the drug/metabolite transporter (DMT)	6054.51674	3.148111391	0.080707345	39.00650443	0	0
Gene 2065	hypothetical protein	5644.873071	3.059752787	0.085139934	3.59E+01	7.81E-283	1.44E-280
Gene 2066	membrane transport protein	18817.45544	3.057400857	0.073241827	41.74391868	0	0
Gene 2068	N-acetyl-L,L-diaminopimelate deacetylase (EC	21338.75171	3.055114526	0.075785191	40.31281684	0	0
Gene 2063	hypothetical protein	3121.550584	2.870498257	0.100981354	2.84E+01	9.64E-178	1.52E-175

Gene 3853	4-Hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16)	6624.281057	2.577874429	0.340862476	7.56E+00	3.94E-14	9.21E-13
Gene 3456	COG1272: Predicted membrane protein hemolysin III	523.8051569	2.473089868	0.176994389	1.40E+01	2.29E-44	2.05E-42
Gene 3847	Threonine dehydrogenase and related Zn-dependent	2236.082584	2.372566072	0.343317319	6.91E+00	4.82E-12	9.58E-11
Gene 3851	TRAP-type C4-dicarboxylate transport system,	4360.457376	2.367289017	0.343174955	6.90E+00	5.27E-12	1.04E-10
Gene 3852	TRAP-type C4-dicarboxylate transport system,	23888.95466	2.300970908	0.335483627	6.86E+00	6.95E-12	1.36E-10
Gene 720	Pyruvate formate-lyase (EC 2.3.1.54)	666.1141311	2.050281025	0.144755051	1.42E+01	1.53E-45	1.41E-43
Gene 2062	Diacylglycerol kinase (EC 2.7.1.107)	710.3718413	1.950097276	0.13704863	1.42E+01	6.03E-46	5.72E-44
Gene 3850	TRAP-type C4-dicarboxylate transport system,	5833.327694	1.897988549	0.340116147	5.58E+00	2.40E-08	2.96E-07
Gene 3848	Gluconokinase (EC 2.7.1.12)	6269.639347	1.831374252	0.340237828	5.38E+00	7.34E-08	8.48E-07
Gene 2457	Iron-sulfur cluster-binding protein	402.3950474	1.684922364	0.177403563	9.50E+00	2.15E-21	8.79E-20
Gene 2759	Deoxyribose operon repressor, DeoR family	7264.488385	1.611883499	0.301234449	5.35E+00	8.75E-08	9.90E-07
Gene 2456	Tungsten-containing aldehyde:ferredoxin	1420.842013	1.611656035	0.10458565	1.54E+01	1.40E-53	1.61E-51
Gene 2117	hypothetical protein	2533.155065	1.583590635	0.288974362	5.48E+00	4.25E-08	5.13E-07
Gene 925	PdxA-like	1570.30675	1.551693074	0.113277245	1.37E+01	1.04E-42	9.08E-41
Gene 2758	Deoxyribose operon repressor, DeoR family	12741.21475	1.517154495	0.311831504	4.87E+00	1.14E-06	1.05E-05

Gene 383	membrane protein, putative	58.12267425	1.505657951	0.30866738	4.88E+00	1.07E-06	9.96E-06
Gene 384	hypothetical protein	86.89896934	1.464837264	0.27843114	5.27E+00	1.35E-07	1.48E-06
Gene 926	Gluconate permease	1780.480273	1.403982369	0.12118 3351	1.16E+01	4.88E-31	3.11E-29
Gene 382	Catalyzes the cleavage of	181.9333461	1.402203425	0.21524 6895	6.51E+00	7.30E-11	1.27E-09
Gene 1753	DNA polymerase I (EC 2.7.7.7)	64.08741908	1.387591632	0.30063 0108	4.62E+00	3.92E-06	3.20E-05
Gene 2757	Ribose/xylose/arabinose/galactoside ABC-type	10697.40931	1.381012437	0.30728 1426	4.49E+00	6.98E-06	5.42E-05
Gene 3652	FIG00519028: hypothetical protein	27269.02918	1.3804409	0.10230 0185	1.35E+01	1.70E-41	1.44E-39
Gene 2760	hypothetical protein	104495.5759	1.355806719	0.28931 6974	4.69E+00	2.78E-06	2.34E-05
Gene 2761	Deoxyribose-phosphate aldolase (EC 4.1.2.4)	57734.28609	1.334614625	0.22714 5067	5.88E+00	4.21E-09	5.66E-08
Gene 3653	FIG00519028: hypothetical protein	10146.26802	1.329799588	0.20551 5393	6.47E+00	9.76E-11	1.67E-09
Gene 787	FIG00516488: hypothetical protein	378.3163545	1.32736421	0.16720 3734	7.94E+00	2.04E-15	5.51E-14
Gene 3146	hypothetical protein	2001.357902	1.318951214	0.30987 2384	4.26E+00	2.08E-05	0.00014 2604
Gene 924	Domain of unknown function DUF1537	315.3424421	1.304205251	0.17894 4563	7.29E+00	3.14E-13	6.98E-12
Gene 2122	Alanine dehydrogenase (EC 1.4.1.1)	43920.12336	1.296773633	0.08380 5799	1.55E+01	5.23E-54	6.20E-52
Gene 3669	FIG00264004: hypothetical protein	3901.693894	1.279017344	0.22487 0308	5.69E+00	1.29E-08	1.65E-07

Gene 3849	Gluconokinase (EC 2.7.1.12)	195.5299306	1.265814708	0.318756484	3.97E+00	7.15E-05	0.000437368
Gene 923	Domain of unknown function DUF1537	709.5966978	1.258369947	0.129923305	9.69E+00	3.48E-22	1.50E-20
Gene 2455	hypothetical protein	120.5135745	1.254094963	0.252906401	4.96E+00	7.10E-07	6.87E-06
Gene 1749	rRNA biogenesis protein rrp5, putative	149.1581342	1.221971345	0.229937726	5.31E+00	1.07E-07	1.20E-06
Gene 1750	Phage protein	371.9497754	1.221735615	0.184948842	6.61E+00	3.95E-11	7.16E-10
Gene 3654	ABC transporter ATP-binding/transmembrane	22510.03369	1.219873144	0.282353613	4.32E+00	1.56E-05	0.000110603
Gene 1751	Phage protein	156.4600133	1.207154376	0.231386083	5.22E+00	1.82E-07	1.95E-06
Gene 3330	FIG00617218: hypothetical protein	138.5450875	1.187919435	0.236989225	5.01E+00	5.37E-07	5.40E-06
Gene 3773	Ferrous iron transport protein B	293000.6594	1.183149712	0.081754784	1.45E+01	1.82E-47	1.95E-45
Gene 3667	no significant homology.	5742.392693	1.169930329	0.115380778	1.01E+01	3.68E-24	1.70E-22
Gene 2454	[Fe] hydrogenase, HymA subunit, putative	137.5718164	1.156719612	0.243764678	4.75E+00	2.08E-06	1.82E-05
Gene 777	no significant homology.	12647.64207	1.14547539	0.07926674	1.45E+01	2.47E-47	2.56E-45
Gene 2756	3-oxoacyl-[acyl-carrier protein] reductase (EC	9480.893919	1.144792644	0.335323614	3.413993513	0.000640181	0.00307658
Gene 69	Phage replication initiation protein	120.3029323	1.135660398	0.251804095	4.51E+00	6.48E-06	5.07E-05
Gene 99	FIG00528782: hypothetical protein	374.4605152	1.117377939	0.166544243	6.71E+00	1.96E-11	3.73E-10

Gene 104	hypothetical protein	569.3283008	1.115323069	0.152807679	7.30E+00	2.90E-13	6.50E-12
Gene 102	Phage tail length tape-measure protein	1727.43829	1.103752715	0.101565082	1.09E+01	1.65E-27	9.42E-26
Gene 78	Methyltransferase	52.856279 42	1.102643874	0.312917712	3.52375 0274	0.00042 5485	0.00214 4238
Gene 3668	Flavodoxin	2525.05 0849	1.089332 202	0.09995 4965	1.09E+01	1.18E-27	6.84E-26
Gene 65	FIG00530366: hypothetical protein	353.916 8957	1.076509 701	0.16694 5239	6.45E+00	1.13E-10	1.89E-09
Gene 98	hypothetical protein	190.527 0001	1.069160 957	0.21109 1689	5.06E+00	4.09E-07	4.22E-06
Gene 106	hypothetical protein	608.544 0203	1.064964 091	0.13940 1321	7.64E+00	2.18E-14	5.31E-13
Gene 1754	FIG00520233: hypothetical protein	39.0193 2352	1.063882 239	0.33697 2577	3.15717 7504	0.00159 3043	0.00696 9039
Gene 94	Scaffold protein	387.826 3453	1.059967 069	0.17684 9538	5.99E+00	2.05E-09	2.90E-08
Gene 2381	CRISPR-associated protein, Csd1 family	1643.90 1597	1.048359 876	0.13482 0054	7.78E+00	7.49E-15	1.94E-13
Gene 103	Phage tail length tape-measure protein	502.721 3223	1.044264 713	0.14782 6505	7.06E+00	1.62E-12	3.31E-11
Gene 95	Phage capsid protein	1394.65 0896	1.027441 166	0.11999 6765	8.56E+00	1.11E-17	3.71E-16
Gene 3331	FIG00517640: hypothetical protein	63.9936 1799	1.027261 205	0.30600 4127	3.35701 7479	0.00078 7881	0.00370 5836
Gene 3405	Transcriptional regulator, AraC family	2392.76 946	1.025628 619	0.09602 4679	1.07E+01	1.25E-26	7.03E-25
Gene 83	Phage-related protein	100.849 2246	1.021751 808	0.26720 5284	3.82384 5823	0.00013 1386	0.00074 9873

Gene 3455	FIG01031768: hypothetical protein	85.06632363	1.006137799	0.291351953	3.453341524	0.000553688	0.002712006
Gene 91	hypothetical protein	154.8231622	0.996338278	0.228610708	4.36E+00	1.31E-05	9.51E-05
Gene 2060	Ubiquinone biosynthesis monooxygenase UbiB	8152.018665	0.994952592	0.076901617	1.29E+01	2.75E-38	2.02E-36
Gene 107	hypothetical protein	283.8835216	0.9946074	0.194716868	5.11E+00	3.26E-07	3.39E-06
Gene 2755	ABC-type multidrug transport system, ATPase	11785.28294	0.993089664	0.333477627	2.977979885	0.00290155	0.011762275
Gene 2383	CRISPR-associated helicase Cas3	196.0555295	0.991442864	0.208671622	4.75E+00	2.02E-06	1.77E-05
Gene 1391	hypothetical protein	598.2889215	0.990739717	0.150494851	6.58E+00	4.60E-11	8.25E-10
Gene 3774	hypothetical protein	16908.05452	0.986957518	0.086291171	1.14E+01	2.72E-30	1.64E-28
Gene 97	Phage capsid and scaffold	236.1374977	0.979457539	0.195761299	5.00E+00	5.63E-07	5.61E-06
Gene 778	FIG00533233: hypothetical protein	2963.770909	0.979391803	0.098178349	9.98E+00	1.95E-23	8.61E-22
Gene 2061	hypothetical protein	436.7598067	0.977667577	0.154299524	6.34E+00	2.36E-10	3.79E-09
Gene 111	FIG01032338: hypothetical protein	612.3211206	0.977393862	0.136850414	7.14E+00	9.19E-13	1.92E-11
Gene 75	hypothetical protein	52.00471782	0.968401223	0.3252329	2.9775623	0.00290506	0.011763929
Gene 89	Portal protein, phage associated	1040.492182	0.960607441	0.114230795	8.41E+00	4.12E-17	1.29E-15
Gene 3795	Predicted erythritol ABC transporter 1,	216.5451588	0.954599369	0.209519955	4.56E+00	5.21E-06	4.18E-05

Gene 68	hypothetical protein	89.76696112	0.9500085	0.269970048	3.518940372	0.000433274	0.002180177
Gene 1036	Iron-sulfur cluster assembly scaffold protein	10197.77181	0.944908097	0.109173888	8.66E+00	4.93E-18	1.74E-16
Gene 70	hypothetical protein	84.67597012	0.937463198	0.293885285	3.189894988	0.001423245	0.006326381
Gene 2751	hypothetical protein	78.88159399	0.937170933	0.303681677	3.086030556	0.002028 48	0.008612598
Gene 3454	Stage II sporulation protein D (SpoIID)	69.46620517	0.9345371	0.293725424	3.18166908	0.00146429	0.006465495
Gene 3332	Chromosome (plasmid) partitioning protein ParB	53.81683164	0.931296428	0.310420009	3.000117263	0.002698757	0.011061901
Gene 1035	Cysteine desulfurase (EC 2.8.1.7)	44207.28571	0.929340276	0.104203996	8.92E+00	4.73E-19	1.72E-17
Gene 2445	Molybdopterin biosynthesis protein MoeA	428.7542047	0.918685835	0.158024939	5.81E+00	6.12E-09	8.08E-08
Gene 1752	hypothetical protein	42.41853138	0.917194982	0.322390283	2.844983338	0.004441375	0.017085381
Gene 3797	Maltose/maltodextrin ABC transporter, permease	198.7721706	0.91194406	0.209466488	4.35E+00	1.34E-05	9.67E-05
Gene 3796	Various polyols ABC transporter, periplasmic	513.2795207	0.900345954	0.151468669	5.94E+00	2.78E-09	3.89E-08
Gene 3771	hypothetical protein	8558.924818	0.897263028	0.074475589	1.20E+01	1.99E-33	1.35E-31
Gene 110	hypothetical protein	67.01617591	0.895750016	0.295408552	3.032241318	0.002427451	0.010125064
Gene 2453	NAD-reducing hydrogenase subunit HoxF (EC	30.19589484	0.894893387	0.33754093	2.651214437	0.00802029	0.028143156
Gene 82	Adenine-specific methyltransferase (EC 2.1.1.72)	82.59439416	0.891092549	0.276816703	3.219070743	0.001286068	0.005778591

Gene 100	FIG00533106: hypothetical protein	108.1203378	0.888413464	0.256144053	3.468413387	0.000523541	0.002587277
Gene 96	hypothetical protein	325.2783847	0.885915428	0.171886553	5.15E+00	2.55E-07	2.67E-06
Gene 2452	NAD-reducing hydrogenase subunit HoxF (EC	1010.072957	0.883268728	0.117364278	7.53E+00	5.24E-14	1.21E-12
Gene 2623	ABC transporter permease protein	2602.046002	0.880297281	0.103390636	8.51E+00	1.68E-17	5.50E-16
Gene 3802	Sorbitol dehydrogenase (EC 1.1.1.14)	242.3176735	0.873025097	0.19346773	4.51E+00	6.41E-06	5.02E-05
Gene 1882	Transcriptional regulator, TetR family	67.03788376	0.869746538	0.293395257	2.964419217	0.003032549	0.012218628
Gene 3801	Multiple sugar ABC transporter, ATP-binding	184.2816288	0.868863049	0.21901054	3.97E+00	7.27E-05	0.000443246
Gene 2382	CRISPR-associated protein, Cas5d family	786.5297969	0.855032394	0.129986595	6.58E+00	4.77E-11	8.51E-10
Gene 101	hypothetical protein	90.45283566	0.851685422	0.271599001	3.135819423	0.001713747	0.007428477
Gene 2451	Periplasmic [Fe] hydrogenase large subunit (EC	586.7976322	0.850275536	0.137727256	6.17E+00	6.67E-10	1.02E-08
Gene 3804	Putative dihydroxyacetone kinase (EC 2.7.1.29),	63.18990404	0.848349223	0.29760894	2.850550198	0.004364366	0.016847774
Gene 887	Sodium/glycine symporter GlyP	9515.067458	0.847048196	0.090497361	9.36E+00	7.98E-21	3.23E-19
Gene 66	hypothetical protein	157.8045011	0.844404856	0.226916707	3.7212106	0.00019827	0.001084923
Gene 61	hypothetical protein	46.00201163	0.843394922	0.317823502	2.653658138	0.007962445	0.027999436
Gene 3799	putative ATP-binding component of ABC	170.2868531	0.843260027	0.217568325	3.875840051	0.000106257	0.000625843

Gene 3798	Maltose/maltodextrin ABC transporter, permease	200.4221273	0.84107924	0.207842386	4.05E+00	5.19E-05	0.000328696
Gene 3655	Tetracenomycin polyketide synthesis	5684.042715	0.829826157	0.342693819	2.42147979	0.015457461	0.048584776
Gene 1037	tRNA-specific 2-thiouridylase MnmA	17113.83024	0.829187702	0.266142371	3.115579452	0.001835839	0.007895774
Gene 86	hypothetical protein	362.3925615	0.828134654	0.166950561	4.96E+00	7.04E-07	6.84E-06
Gene 90	putative phage protein	519.1349111	0.822769357	0.145863875	5.64E+00	1.69E-08	2.14E-07
Gene 88	Phage uncharacterized protein-like	1077.01685	0.821032029	0.114917052	7.14E+00	9.03E-13	1.89E-11
Gene 4154	Na+/H+ antiporter NhaC	523.6163759	0.818325663	0.152020151	5.38E+00	7.33E-08	8.48E-07
Gene 64	hypothetical protein	90.897110 89	0.815770454	0.273287431	2.98502734	0.00283553	0.011536953
Gene 2668	ABC transporter ATP-binding protein YvcR	232.0818931	0.811141496	0.224810982	3.60810441	0.000308442	0.001598117
Gene 2446	Methionine ABC transporter ATP-binding protein	197.0920494	0.794945044	0.221108224	3.595275781	0.000324048	0.001671141
Gene 3772	Ferrous iron transport protein A	4966.331599	0.793487204	0.092627979	8.57E+00	1.07E-17	3.61E-16
Gene 3817	Ribose 5-phosphate isomerase B (EC 5.3.1.6)	745.3914869	0.787566483	0.162412227	4.85E+00	1.24E-06	1.12E-05
Gene 3651	hypothetical protein	5080.686335	0.782987317	0.087344227	8.96E+00	3.12E-19	1.16E-17
Gene 876	Aminomethyltransferase (glycine cleavage system T	710.4978169	0.780739765	0.128852114	6.06E+00	1.37E-09	1.97E-08
Gene 67	hypothetical protein	77.00746146	0.776418882	0.283643108	2.737309178	0.006194405	0.022747116

Gene 105	hypothetical protein	267.931772	0.775983442	0.185605447	4.18E+00	2.90E-05	0.000195763
Gene 3964	Acyl carrier protein	2301.969851	0.76842555	0.101403229	7.58E+00	3.51E-14	8.38E-13
Gene 485	hypothetical protein	5097.446616	0.762416829	0.086189475	8.85E+00	9.09E-19	3.27E-17
Gene 2964	hypothetical protein	650.6891652	0.758828567	0.145617856	5.21E+00	1.88E-07	2.00E-06
Gene 2448	ABC-type tungstate transport system, periplasmic	354.0515111	0.756754255	0.168656062	4.49E+00	7.22E-06	5.60E-05
Gene 1937	DegV family protein	3520.936617	0.75544412	0.083352863	9.06E+00	1.27E-19	4.94E-18
Gene 3974	Enoyl-[acyl-carrier-protein] reductase [FMN] (EC	9969.620537	0.75101704	0.077252632	9.72E+00	2.44E-22	1.06E-20
Gene 1454	Na(+)-linked D-alanine glycine permease	8468.399426	0.750777388	0.096562261	7.78E+00	7.54E-15	1.94E-13
Gene 877	Aminomethyltransferase (glycine cleavage system T	1539.709658	0.748200573	0.107151935	6.98E+00	2.90E-12	5.82E-11
Gene 879	Glycine dehydrogenase [decarboxylating] (glycine	4358.343303	0.744659672	0.082489338	9.03E+00	1.76E-19	6.78E-18
Gene 3803	Putative dihydroxyacetone kinase (EC 2.7.1.29),	151.400205	0.741712837	0.232776309	3.18637 5965	0.00144 0672	0.00637 8197
Gene 3973	Acetyl-coenzyme A carboxyl transferase alpha	26557.49532	0.732909445	0.069068986	1.06E+01	2.64E-26	1.41E-24
Gene 1034	Cysteine desulfurase (EC 2.8.1.7)	1418.165727	0.720483409	0.108539932	6.64E+00	3.18E-11	5.93E-10
Gene 3977	NAD-dependent protein deacetylase of SIR2 family	8534.853979	0.716482641	0.084879007	8.44E+00	3.14E-17	1.00E-15
Gene 3966	Malonyl CoA-acyl carrier protein transacylase (EC	10862.77536	0.712378903	0.087316754	8.16E+00	3.39E-16	9.86E-15

Gene 3806	Sorbitol operon transcription regulator	130.3116608	0.711167819	0.244463848	2.909091982	0.003624802	0.014275348
Gene 2123	transcriptional regulator, RpiR family	1873.352695	0.70751808	0.115443715	6.13E+00	8.86E-10	1.31E-08
Gene 478	hypothetical protein	739.7400277	0.706057285	0.143812169	4.91E+00	9.13E-07	8.60E-06
Gene 295	Cobalt-precorrin-6 synthase, anaerobic	456.6121282	0.703980852	0.17631824	3.99E+00	6.53E-05	0.000404876
Gene 1869	O-acetylhomoserine sulfhydrylase (EC 2.5.1.49)	2278.906599	0.696355414	0.262142046	2.656404896	0.007897872	0.027861005
Gene 1557	diguanylate cyclase (GGDEF domain) with PAS/PAC	171.6299364	0.696012991	0.22051088	3.156365756	0.001597484	0.006977755
Gene 2270	Phage capsid protein	124.8264538	0.695613071	0.242464817	2.868923753	0.004118711	0.016048936
Gene 2450	Molybdenum cofactor biosynthesis protein MoaA	663.2178622	0.69487243	0.133304298	5.21E+00	1.86E-07	1.98E-06
Gene 3975	Transcriptional regulator of fatty acid	3788.886791	0.694596081	0.086042595	8.07E+00	6.88E-16	1.97E-14
Gene 3784	Ribulokinase (EC 2.7.1.16)	1562.07354	0.694097795	0.108344063	6.41E+00	1.49E-10	2.46E-09
Gene 4155	hypothetical protein	330.7943865	0.693781347	0.170245473	4.08E+00	4.60E-05	0.000295096
Gene 3972	Biotin carboxylase of acetyl-CoA carboxylase (EC	1564.817598	0.690443002	0.104288008	6.62E+00	3.57E-11	6.61E-10
Gene 484	FIG00617386: hypothetical protein	4201.948162	0.689774459	0.084195398	8.19E+00	2.56E-16	7.51E-15
Gene 1451	Glycine reductase component B beta subunit (EC	7216.597886	0.687302998	0.088412192	7.77E+00	7.61E-15	1.94E-13
Gene 109	Holin, toxin secretion/phage lysis	138.2694707	0.683804816	0.233956728	2.9227833	0.003469179	0.013711319

Gene 3650	Transcriptional regulator, TetR family	7284.242551	0.679756223	0.075454217	9.01E+00	2.08E-19	7.94E-18
Gene 1868	Cysteine synthase (EC 2.5.1.47)	1428.5951	0.674294556	0.142446931	4.73E+00	2.21E-06	1.90E-05
Gene 813	Chaperone protein DnaJ	11437.86511	0.67166325	0.07504129	8.95E+00	3.54E-19	1.30E-17
Gene 3976	Putative hydrolase	6504.932746	0.66406883	0.076821934	8.64E+00	5.42E-18	1.89E-16
Gene 482	hypothetical protein	3615.446097	0.656320398	0.085767793	7.65E+00	1.97E-14	4.85E-13
Gene 4157	3-oxoacyl-[acyl-carrier protein] reductase (EC	285.5706253	0.653144207	0.187816344	3.477568522	0.000505984	0.002507985
Gene 4106	Putative ADP-ribosylglycohydrolase	5246.244445	0.653110073	0.101167833	6.46E+00	1.08E-10	1.81E-09
Gene 878	Glycine cleavage system H protein	773.8200286	0.652077491	0.135381829	4.82E+00	1.46E-06	1.31E-05
Gene 4112	Methylcobalamin:coenzyme M methyltransferase,	446.9789481	0.651475374	0.162793274	4.00E+00	6.28E-05	0.000391733
Gene 2384	CRISPR-associated helicase Cas3	4369.643717	0.651219684	0.085075211	7.65E+00	1.94E-14	4.80E-13
Gene 1284	Alanine racemase (EC 5.1.1.1)	5331.238539	0.649655243	0.080221641	8.10E+00	5.58E-16	1.61E-14
Gene 880	Glycine dehydrogenase [decarboxylating] (glycine	2883.398641	0.649590431	0.090423637	7.18E+00	6.78E-13	1.46E-11
Gene 87	Phage protein	383.7575864	0.648084631	0.161738508	4.01E+00	6.15E-05	0.000384765
Gene 4153	Na+/H+ antiporter NhaC	699.7608649	0.646495753	0.134616982	4.80E+00	1.57E-06	1.40E-05
Gene 812	Chaperone protein DnaK	27828.71601	0.644836186	0.070554056	9.14E+00	6.27E-20	2.47E-18

Gene 1407	ABC transporter ATP-binding protein	102.7202307	0.640422973	0.260392563	2.459451862	0.013914936	0.044581572
Gene 3807	Xylulose kinase (EC 2.7.1.17)	532.1513878	0.639647914	0.144374051	4.43E+00	9.40E-06	7.09E-05
Gene 2831	Arsenic efflux pump protein	326.6866204	0.638084944	0.176382878	3.617612726	0.000297333	0.001552686
Gene 2369	DNA-binding domain of ModE	278.9802645	0.637299784	0.187154361	3.405209364	0.000661133	0.003163519
Gene 2380	CRISPR-associated protein, Csd2 family	2089.297406	0.633662801	0.111034427	5.71E+00	1.15E-08	1.48E-07
Gene 481	hypothetical protein	2708.489566	0.629363668	0.097812398	6.43E+00	1.24E-10	2.06E-09
Gene 483	hypothetical protein	5113.396497	0.62930895	0.07868457	8.00E+00	1.27E-15	3.47E-14
Gene 1277	Translocation-enhancing protein TepA	460.1881416	0.625473493	0.152906778	4.09E+00	4.30E-05	0.00027783
Gene 1285	NAD(P)HX epimerase / NAD(P)HX dehydratase	6446.909134	0.6240457	0.077792037	8.02E+00	1.04E-15	2.88E-14
Gene 2449	Molybdenum cofactor biosynthesis protein MoaB	194.2252368	0.62392866	0.210188399	2.968425776	0.002993294	0.012089845
Gene 2175	hypothetical protein	536.2164177	0.620910451	0.176872877	3.510489915	0.000447282	0.002237084
Gene 486	hypothetical protein	448.3732061	0.615524633	0.159799997	3.851843840	0.000117232	0.000685609
Gene 1558	hypothetical protein	333.9463543	0.613927909	0.169596154	3.619940054	0.000294671	0.001541214
Gene 2118	DNA repair protein RadA	1062.784588	0.610497331	0.113813405	5.36E+00	8.14E-08	9.27E-07
Gene 2529	hypothetical protein	2320.339033	0.610233935	0.09424694	6.47E+00	9.49E-11	1.63E-09

Gene 2530	Streptococcal hemagglutinin protein	1884.093399	0.607570604	0.101018241	6.01E+00	1.80E-09	2.58E-08
Gene 4111	Methylcobalamin methyltransferase MMP0830	458.9692135	0.606701035	0.159373839	3.806779317	0.000140788	0.000796679
Gene 3906	Galactitol-1-phosphate 5-dehydrogenase (EC	550.5428072	0.601983051	0.144820825	4.16E+00	3.23E-05	0.000215499
Gene 3087	FIG00529733: hypothetical protein	1756.786678	0.601105448	0.128777244	4.67E+00	3.04E-06	2.52E-05
Gene 3783	L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4)	690.0305969	0.599719839	0.136756244	4.39E+00	1.16E-05	8.53E-05
Gene 3260	Probable cation-transporting ATPase	2179.645651	0.59680901	0.101031757	5.91E+00	3.48E-09	4.73E-08
Gene 4104	Ribose ABC transport system, periplasmic	16358.85156	0.593122127	0.071588352	8.29E+00	1.18E-16	3.59E-15
Gene 2444	Molybdopterin biosynthesis protein MoeA /	762.9267117	0.592899633	0.126642261	4.68E+00	2.85E-06	2.38E-05
Gene 3816	Ribose 5-phosphate isomerase B (EC 5.3.1.6)	422.1270995	0.592142668	0.163737601	3.616412271	0.000298714	0.001555004
Gene 2622	ABC transporter, ATP-binding protein	299.8111542	0.591561889	0.187178737	3.160411797	0.001575463	0.006901235
Gene 3971	Biotin carboxylase of acetyl-CoA carboxylase (EC	19087.51927	0.587493915	0.074393826	7.90E+00	2.86E-15	7.64E-14
Gene 1124	peptidase, M23/M37 family	146.6477479	0.584509062	0.232759622	2.511213318	0.012031697	0.039580463
Gene 881	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	2521.912323	0.583413476	0.091574261	6.37E+00	1.88E-10	3.05E-09
Gene 298	Cobalamin biosynthesis protein CbiG	707.4958502	0.583009272	0.140260382	4.16E+00	3.23E-05	0.000215499
Gene 1449	Glycine reductase component B gamma subunit (EC	7458.589965	0.582059898	0.080842401	7.20E+00	6.02E-13	1.31E-11

Gene 1422	Deoxyguanosinetriphosphate triphosphohydrolase	554.0784396	0.58015903	0.152869149	3.795134826	0.000147563	0.000830563
Gene 222	oxidoreductase family protein	652.0288771	0.578882332	0.173081859	3.344558102	0.000824138	0.003859944
Gene 3969	Biotin carboxyl carrier protein of acetyl-CoA	11518.32139	0.576129216	0.078745991	7.32E+00	2.55E-13	5.75E-12
Gene 757	Single-stranded-DNA-specific exonuclease RecJ (EC	1252.317592	0.57447335	0.113943087	5.04E+00	4.61E-07	4.68E-06
Gene 2379	CRISPR-associated RecB family exonuclease Cas4	1618.863215	0.573830417	0.124122407	4.62E+00	3.78E-06	3.10E-05
Gene 84	Integrase/recombinase (XerC/CodV family)	161.5049131	0.571132168	0.225541054	2.532275865	0.011332479	0.037805334
Gene 1283	Programmed cell death toxin YdcE	1061.597825	0.570943111	0.118329701	4.83E+00	1.40E-06	1.26E-05
Gene 2447	ABC-type tungstate transport system, permease	203.8125089	0.570769461	0.206977113	2.757645285	0.005821934	0.021522332
Gene 2965	hypothetical protein	306.5649261	0.569947743	0.189783629	3.003144924	0.002672052	0.010965994
Gene 3843	predicted xylose isomerase	907.6009223	0.566611838	0.144348778	3.93E+00	8.66E-05	0.000520364
Gene 598	GTP-sensing transcriptional pleiotropic repressor	8349.976059	0.561336377	0.074187261	7.57E+00	3.83E-14	9.08E-13
Gene 4113	Dimethylamine corrinoid protein 2	389.9503011	0.560786957	0.169981862	3.299098795	0.000969958	0.004467194
Gene 892	Stage V sporulation protein AF (SpoVAF)	218.8925682	0.560590741	0.202880642	2.763155394	0.005724551	0.021233344
Gene 2700	Transcriptional regulator, AraC family	3644.092829	0.560554682	0.085257676	6.57E+00	4.87E-11	8.59E-10
Gene 3789	L-xylulose 5-phosphate 3-epimerase (EC 5.1.3.-)	373.7633512	0.56038547	0.176032349	3.183423241	0.001455447	0.006435016

Gene 252	Xaa-Pro aminopeptidase (EC 3.4.11.9)	28974.07296	0.55795163	0.069239078	8.06E+00	7.73E-16	2.17E-14
Gene 4110	Dimethylamine corrinoid protein 2	239.0860386	0.556929976	0.194961377	2.856616964	0.004281822	0.016625904
Gene 1122	hypothetical protein	164.5397148	0.55386937	0.223400606	2.479265296	0.013165334	0.042531195
Gene 3965	Enoyl-[acyl-carrier-protein] reductase [FMN] (EC	6423.320627	0.553035273	0.077262057	7.16E+00	8.19E-13	1.74E-11
Gene 3674	FIG00253762: hypothetical protein	959.1833224	0.552582575	0.163549154	3.378694188	0.00072831	0.003445186
Gene 299	Cobalt-precorrin-3b C17-methyltransferase	931.575878	0.550754432	0.122170793	4.51E+00	6.54E-06	5.10E-05
Gene 1502	N-Acetyl-D-glucosamine ABC transport system,	4188.523324	0.550338415	0.089551008	6.15E+00	7.97E-10	1.21E-08
Gene 599	FIG00518951: hypothetical protein	1331.592082	0.545871655	0.109799134	4.97E+00	6.64E-07	6.52E-06
Gene 1425	FIG00515392: hypothetical protein	1225.169809	0.542633696	0.10752624	5.05E+00	4.50E-07	4.62E-06
Gene 1501	hypothetical protein	190.2869499	0.537846635	0.209925587	2.562082324	0.010404665	0.035134287
Gene 867	FIG01032623: hypothetical protein	1818.396868	0.530635269	0.119395262	4.44E+00	8.82E-06	6.70E-05
Gene 296	Cobalt-precorrin-2 C20-methyltransferase (EC	490.8138515	0.530314668	0.163269846	3.24808702	0.001161837	0.005270387
Gene 866	Protein export cytoplasm protein SecA ATPase RNA	2038.259275	0.529864322	0.12009681	4.41E+00	1.02E-05	7.68E-05
Gene 1033	Iron-sulfur cluster regulator IscR	9724.961322	0.525332996	0.079506976	6.61E+00	3.91E-11	7.13E-10
Gene 4108	Indigoidine synthase A-like protein,	5379.569683	0.524086335	0.086280555	6.07E+00	1.25E-09	1.81E-08

Gene 3057	Xanthine/uracil/thiamine/ascorbate permease	25239.01077	0.519759695	0.072317537	7.19E+00	6.61E-13	1.43E-11
Gene 4129	hypothetical protein	2101.661504	0.51867382	0.102719256	5.05E+00	4.43E-07	4.56E-06
Gene 3499	Guanylate kinase (EC 2.7.4.8)	352.7626038	0.515672396	0.166059787	3.10534 1793	0.00190 0593	0.00811 3162
Gene 820	FIG00524296: hypothetical protein	2305.553333	0.513171 89	0.11589 647	4.43E+00	9.52E-06	7.16E-05
Gene 3610	V-type ATP synthase subunit I (EC 3.6.3.14)	24013.82642	0.507359 405	0.08546 7776	5.94E+00	2.92E-09	4.05E-08
Gene 1452	GrdX protein	1650.29 3408	0.506128 188	0.11121 5798	4.55E+00	5.34E-06	4.27E-05
Gene 487	Phage tape measure	10423.7 3541	0.505859 158	0.07262 8635	6.97E+00	3.28E-12	6.56E-11
Gene 4107	ADP-ribosylglycohydrolase (EC 3.2.-.-)	3386.27 666	0.505464 259	0.08428 1476	6.00E+00	2.01E-09	2.84E-08
Gene 3967	3-oxoacyl-[acyl-carrier protein] reductase (EC	10677.8 673	0.500823 03	0.07495 689	6.68E+00	2.37E-11	4.46E-10
Gene 1948	Nitroreductase family protein	281.732 1757	0.498374 062	0.18326 4484	2.71942 5239	0.00653 9548	0.02377 7566
Gene 1022	RecA protein	605.596 1065	0.496540 451	0.13822 9849	3.59213 6239	0.00032 7978	0.00168 6164
Gene 23	Glycerol kinase (EC 2.7.1.30)	1406.69 7296	0.495711 777	0.11060 1432	4.48E+00	7.40E-06	5.70E-05
Gene 297	Cobalt-precorrin-4 C11-methyltransferase (EC	641.858 2251	0.493438 217	0.14171 7824	3.48183 598	0.00049 7989	0.00247 2051
Gene 1023	RecA protein	793.029 7967	0.493385 066	0.12378 6302	3.99E+00	6.73E-05	0.00041 5325
Gene 3968	3-oxoacyl-[acyl-carrier-protein] synthase, KASII	21680.9 1292	0.491614 292	0.07678 4175	6.40E+00	1.53E-10	2.51E-09

Gene 4135	FIG00520517: hypothetical protein	923.7145878	0.491489481	0.12851433	3.824394373	0.000131094	0.000749496
Gene 118	hypothetical protein	430.9159485	0.490443718	0.154292482	3.17866244	0.001479563	0.006524243
Gene 489	hypothetical protein	5167.189548	0.48910474	0.079779748	6.13E+00	8.75E-10	1.30E-08
Gene 4103	Ribokinase (EC 2.7.1.15)	5505.963408	0.486913204	0.07833616	6.22E+00	5.11E-10	7.84E-09
Gene 1548	Lead, cadmium, zinc and mercury transporting	5477.81735	0.486011097	0.098672829	4.93E+00	8.42E-07	8.02E-06
Gene 1453	hypothetical protein	2712.790784	0.485918781	0.089897754	5.41E+00	6.47E-08	7.61E-07
Gene 3995	hypothetical protein	1726.244437	0.485522206	0.105379657	4.61E+00	4.08E-06	3.32E-05
Gene 2353	hypothetical protein	362.8383353	0.484781278	0.180199295	2.690250693	0.007139836	0.025650809
Gene 3785	Rhamnulokinase (EC 2.7.1.5)	1967.407091	0.481039326	0.101572209	4.74E+00	2.18E-06	1.88E-05
Gene 496	Glutamate synthase [NADPH] large chain (EC	2249.211559	0.48103672	0.105207197	4.57E+00	4.82E-06	3.88E-05
Gene 3160	6-phosphofructokinase (EC 2.7.1.11)	3446.495229	0.480725045	0.090199059	5.33E+00	9.84E-08	1.11E-06
Gene 3500	Guanylate kinase (EC 2.7.4.8)	653.6053654	0.479472638	0.133395953	3.594356711	0.000325194	0.001674446
Gene 1971	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate	1643.952188	0.474720905	0.1016861	4.67E+00	3.03E-06	2.52E-05
Gene 497	N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28)	10702.43516	0.474660614	0.097580157	4.86E+00	1.15E-06	1.05E-05
Gene 819	Glycogen debranching enzyme (EC 3.2.1.-)	3699.465013	0.472372462	0.093696922	5.04E+00	4.62E-07	4.68E-06

Gene 3811	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	970.3525163	0.469771629	0.117182121	4.01E+00	6.10E-05	0.000382385
Gene 586	FIG011945: O-methyltransferase family protein	1830.885248	0.46477257	0.101167863	4.59E+00	4.35E-06	3.53E-05
Gene 2549	ClpB protein	2548.381055	0.464088415	0.088994194	5.21E+00	1.84E-07	1.97E-06
Gene 4115	probable electron transfer protein	1034.65572	0.462277892	0.119045844	3.88319218	0.000103094	0.000609376
Gene 722	FIG00617256: hypothetical protein	6229.840046	0.461688935	0.089138203	5.18E+00	2.23E-07	2.35E-06
Gene 2846	Phosphate regulon transcriptional regulatory	834.3099957	0.461429389	0.124506874	3.706055532	0.000210512	0.001140618
Gene 1579	hypothetical protein	2177.631871	0.461307985	0.098239355	4.70E+00	2.66E-06	2.24E-05
Gene 2528	hypothetical protein	1174.342654	0.461105526	0.11647135	3.96E+00	7.53E-05	0.000457174
Gene 4114	Methylcobalamin:coenzyme M methyltransferase,	1076.968967	0.461070778	0.115158263	4.00E+00	6.23E-05	0.000389257
Gene 376	HPr kinase/phosphorylase (EC 2.7.1.-) (EC	1326.932422	0.460782117	0.121742212	3.784900156	0.00015377	0.000860584
Gene 488	hypothetical protein	4141.279175	0.458915567	0.084709739	5.42E+00	6.04E-08	7.18E-07
Gene 1578	FIG00527654: hypothetical protein	458.7584737	0.457852393	0.15152083	3.021712555	0.002513491	0.01043146
Gene 3627	sensor histidine kinase	4629.33627	0.456951225	0.082253561	5.56E+00	2.77E-08	3.39E-07
Gene 3058	Adenine deaminase (EC 3.5.4.2)	36144.09926	0.456904317	0.071968709	6.35E+00	2.17E-10	3.51E-09
Gene 24	Glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	1957.421806	0.452828807	0.105013496	4.31E+00	1.62E-05	0.000113129

Gene 491	integrase/recombinase XerD	5482.166425	0.450227975	0.082488334	5.46E+00	4.81E-08	5.76E-07
Gene 1500	sugar ABC transporter permease	2765.976581	0.446567014	0.095675095	4.67E+00	3.05E-06	2.52E-05
Gene 1406	ABC transporter permease protein	444.334263	0.442630305	0.155086079	2.854094373	0.004315971	0.016719345
Gene 2378	CRISPR-associated protein Cas1	2289.691893	0.433871933	0.101861835	4.26E+00	2.05E-05	0.000141007
Gene 300	Cobalt-precorrin-6y C5-methyltransferase (EC	2151.111415	0.431784947	0.103112664	4.19E+00	2.82E-05	0.000191254
Gene 2233	DegV family protein	1645.557044	0.429574888	0.105416395	4.08E+00	4.60E-05	0.000295096
Gene 3790	Xylulose kinase (EC 2.7.1.17)	659.7385858	0.42935747	0.164455732	2.610778377	0.009033642	0.031301521
Gene 2482	Thiosulfate sulfurtransferase, rhodanese (EC	2176.298717	0.428984471	0.115509863	3.713834127	0.000204143	0.001111555
Gene 4102	Ribose ABC transport system, permease protein	9567.750785	0.428020435	0.076056772	5.63E+00	1.83E-08	2.30E-07
Gene 3970	3-hydroxyacyl-[acyl-carrier-protein] dehydratase,	4174.870474	0.42729092	0.086854871	4.92E+00	8.67E-07	8.22E-06
Gene 4045	D-proline reductase, 45 kDa subunit (EC 1.21.4.1)	644.6583914	0.426850127	0.169034241	2.525228757	0.0115623	0.038340586
Gene 375	Excinuclease ABC subunit C	2269.201065	0.424134087	0.092391976	4.59E+00	4.42E-06	3.57E-05
Gene 3904	Xylulose kinase (EC 2.7.1.17)	1340.850503	0.422891399	0.114267837	3.700878644	0.000214854	0.001162246
Gene 1575	FIG00518389: hypothetical protein	1074.969818	0.422634894	0.111768989	3.781325194	0.000155996	0.000870845
Gene 3245	Two-component response regulator	856.2490193	0.419535959	0.121440411	3.454665166	0.000550977	0.00270272

Gene 315	Additional substrate-specific component CbiN of	5938.208357	0.413549195	0.099410844	4.16E+00	3.18E-05	0.000213193
Gene 3611	V-type ATP synthase subunit I (EC 3.6.3.14)	19826.57579	0.412617513	0.084014312	4.91E+00	9.05E-07	8.55E-06
Gene 2866	ATP-dependent DNA helicase RecQ	1661.13315	0.410338088	0.106562863	3.850666884	0.000117797	0.0006877
Gene 1158	FIG00512775: hypothetical protein	723.7704512	0.405812705	0.132888147	3.05379159	0.00225969	0.009472986
Gene 3901	Ribulose-5-phosphate 4-epimerase and related	613.8703412	0.405677853	0.13767384	2.946658946	0.003212273	0.012818167
Gene 3137	Dipeptide-binding ABC transporter, periplasmic	8247.263943	0.403292141	0.082557562	4.88E+00	1.03E-06	9.63E-06
Gene 3818	3-oxoacyl-[acyl-carrier protein] reductase (EC	3655.838967	0.401803464	0.091821828	4.38E+00	1.21E-05	8.83E-05
Gene 2847	Phosphate transport system regulatory protein	575.1060409	0.395871929	0.140429022	2.81901791	0.004817083	0.018192992
Gene 2698	Galactose-1-phosphate uridylyltransferase (EC	24865.90107	0.395374299	0.068620696	5.76E+00	8.33E-09	1.09E-07
Gene 4116	probable electron transfer protein	613.812783	0.394704661	0.142043814	2.778752906	0.005456802	0.020376976
Gene 1577	FIG00527654: hypothetical protein	487.9707903	0.392634282	0.150221603	2.613700517	0.008956751	0.03113269
Gene 3903	Tagatose 1,6-bisphosphate aldolase (EC 4.1.2.40)	701.6375305	0.391208201	0.129890616	3.01182806	0.002596796	0.010709496
Gene 2796	hypothetical protein	427.6169977	0.390688232	0.158582939	2.463620826	0.013754153	0.044109063
Gene 3486	hypothetical protein	487.9977152	0.385494079	0.152019636	2.535817674	0.011218513	0.037500592
Gene 622	flavin reductase domain protein, FMN-binding	28573.8369	0.384680642	0.08126635	4.73E+00	2.21E-06	1.90E-05

Gene 113	Superfamily II DNA/RNA helicase	868.2760525	0.384501056	0.120311767	3.1958724	0.001394087	0.006221796
Gene 3602	Fe-S oxidoreductase	1080.094147	0.38424383	0.118441723	3.244159409	0.001177978	0.005336305
Gene 2156	Diaminopimelate epimerase (EC 5.1.1.7)	2443.914853	0.381436964	0.09336304	4.09E+00	4.40E-05	0.00028163
Gene 3902	Transcriptional repressor of the fructose operon,	703.2673786	0.381295542	0.143066725	2.6651588	0.007695198	0.027320423
Gene 2696	DNA polymerase IV (EC 2.7.7.7)	431.7145255	0.380763636	0.156548838	2.432235473	0.015005948	0.04743539
Gene 3112	N-Acetyl-D-glucosamine ABC transport system,	64558.02367	0.378999059	0.070036169	5.41E+00	6.25E-08	7.40E-07
Gene 3893	Xylulose kinase (EC 2.7.1.17)	1029.998652	0.378958909	0.114838045	3.299942175	0.000967048	0.004466197
Gene 302	Cobyrinic acid A,C-diamide synthase	850.0004117	0.374383283	0.122210595	3.063427381	0.002188173	0.009231529
Gene 811	Heat shock protein GrpE	6139.813952	0.372043089	0.077712489	4.79E+00	1.69E-06	1.50E-05
Gene 147	Thioredoxin reductase (EC 1.8.1.9)	2079.983359	0.372038204	0.118888095	3.129314209	0.001752148	0.007554007
Gene 3498	Arginine decarboxylase (EC 4.1.1.19) / Lysine	6004.009086	0.371821825	0.082087147	4.53E+00	5.91E-06	4.67E-05
Gene 146	Ribonucleotide reductase of class III	24740.58747	0.371216832	0.077706816	4.78E+00	1.78E-06	1.57E-05
Gene 1495	Molybdopterin biosynthesis protein MoeA	4542.127435	0.370564395	0.090397679	4.10E+00	4.14E-05	0.000269478
Gene 35	FIG00531184: hypothetical protein	3577.425712	0.370288666	0.08519523	4.35E+00	1.38E-05	9.96E-05
Gene 4136	TsaE protein, required for	526.6304046	0.367458323	0.151915693	2.418830576	0.01557049	0.04875519

Gene 3844	Xylulose kinase (EC 2.7.1.17)	2818.672712	0.367235288	0.092385097	3.98E+00	7.04E-05	0.000432092
Gene 500	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)	1424.061343	0.365641438	0.117423209	3.113877062	0.001846464	0.007931185
Gene 2699	Galactokinase (EC 2.7.1.6)	17385.14024	0.365357275	0.071920614	5.08E+00	3.77E-07	3.91E-06
Gene 3501	DNA polymerase III delta prime subunit (EC	3171.754099	0.359201293	0.097294997	3.691878339	0.000222604	0.001196361
Gene 2070	Regulatory protein (induces abgABT, used to	916.1284555	0.355885989	0.12555375	2.834530939	0.004589301	0.017552622
Gene 3519	1,4-alpha-glucan (glycogen) branching enzyme,	4035.58789	0.354751897	0.090018176	3.94E+00	8.12E-05	0.000489437
Gene 3895	Rhamnulokinase (EC 2.7.1.5)	1101.494911	0.352513335	0.112343839	3.13780745	0.001702167	0.007387938
Gene 779	Thymidylate synthase (EC 2.1.1.45)	2122.18696	0.351352349	0.103978313	3.379092607	0.000727255	0.003445111
Gene 602	Heat shock protein 60 family chaperone GroEL	13162.36824	0.348976508	0.078070543	4.47E+00	7.82E-06	5.99E-05
Gene 490	hypothetical protein	4111.432425	0.346672651	0.090806365	3.817713116	0.000134694	0.000764806
Gene 1815	Ribonucleotide reductase of class Ia (aerobic),	5478.849022	0.345996417	0.087022822	3.98E+00	7.01E-05	0.000431299
Gene 810	Heat-inducible transcription repressor HrcA	6562.832564	0.344720111	0.076497966	4.51E+00	6.60E-06	5.14E-05
Gene 112	hypothetical protein	682.2575871	0.343716809	0.13781828	2.493985626	0.012631766	0.04098526
Gene 2666	FIG00513822: hypothetical protein	2159.260065	0.342692053	0.093332847	3.671719691	0.000240924	0.001282349
Gene 3086	Glucosamine-6-phosphate deaminase (EC 3.5.99.6)	6042.923684	0.341343725	0.077575519	4.40E+00	1.08E-05	8.03E-05

Gene 2547	hypothetical protein BH3604	4110.554808	0.339882186	0.081185743	4.19E+00	2.83E-05	0.000191731
Gene 1282	Magnesium and cobalt efflux protein CorC	1725.597056	0.339294745	0.108304898	3.132773777	0.001731628	0.007496188
Gene 3612	V-type ATP synthase subunit K (EC 3.6.3.14)	10281.4988	0.338268616	0.07813148	4.33E+00	1.49E-05	0.000106584
Gene 3142	hypothetical protein	3348.585911	0.336711037	0.097233136	3.462924801	0.000534338	0.002632784
Gene 378	Glucokinase (EC 2.7.1.2)	2454.931886	0.336129054	0.090662366	3.707481607	0.000209331	0.001136073
Gene 121	hypothetical protein	5630.682806	0.334862924	0.094455547	3.545190686	0.000392329	0.001989241
Gene 130	ABC transporter, ATP-binding protein	5396.956263	0.334593671	0.109480366	3.056197945	0.002241632	0.00940918
Gene 4128	Transcription accessory protein (S1 RNA-binding	8354.924448	0.334226966	0.077392415	4.32E+00	1.57E-05	0.00011102
Gene 2584	Mobile element protein	2037.67404	0.332822662	0.099241595	3.353660947	0.0007975	0.003745766
Gene 4087	Formate dehydrogenase-O, major subunit (EC	6969.029367	0.331450557	0.076973681	4.31E+00	1.66E-05	0.000116035
Gene 582	Putative Holliday junction resolvase YqgF	7820.353547	0.329780669	0.076456882	4.31E+00	1.61E-05	0.000112999
Gene 868	hypothetical protein	1094.567485	0.328979955	0.114568164	2.871477935	0.004085573	0.015957315
Gene 3136	Dipeptide transport system permease protein DppB	11285.84648	0.327608324	0.085595201	3.827414638	0.000129496	0.000742923
Gene 1489	Phosphatidate cytidylyltransferase (EC 2.7.7.41)	3771.277932	0.326097945	0.087802503	3.713993725	0.000204014	0.001111555
Gene 4131	Proton/glutamate symport protein	5309.312812	0.325701376	0.085072517	3.828514632	0.000128919	0.000740893

Gene 314	Substrate-specific component CbiM of cobalt ECF	16642.62032	0.325408428	0.075022797	4.34E+00	1.44E-05	0.000103455
Gene 2672	Transcriptional regulator, GntR family	1588.452844	0.325302967	0.100913334	3.223587544	0.001265956	0.005695943
Gene 1970	DNA polymerase III alpha subunit (EC 2.7.7.7)	9311.601412	0.32133266	0.074487933	4.31E+00	1.60E-05	0.000112932
Gene 1556	Inosine-5-monophosphate dehydrogenase (EC	3372.201566	0.319214274	0.085368823	3.739237158	0.00018458	0.00102011
Gene 1430	Ribose ABC transport system, periplasmic	5793.839232	0.316879759	0.078018577	4.06E+00	4.87E-05	0.000310804
Gene 3159	DNA polymerase III subunits gamma and tau (EC	6425.430424	0.316245235	0.090076894	3.510836366	0.000446699	0.002237084
Gene 316	Transmembrane component CbiQ of energizing module	19331.78238	0.315702962	0.091538868	3.448840574	0.000562999	0.002749491
Gene 3084	extracellular solute-binding protein, family 1	6428.36269	0.31428282	0.077251839	4.07E+00	4.74E-05	0.00030259
Gene 585	Ribonuclease J2 (endoribonuclease in RNA	21595.2716	0.314252035	0.07471899	4.21E+00	2.60E-05	0.000177155
Gene 1795	Adenine deaminase (EC 3.5.4.2)	2656.928536	0.312156502	0.108837387	2.868099942	0.004129451	0.0160719
Gene 584	Ferric uptake regulation protein FUR	5197.04705	0.312091516	0.083362683	3.743779654	0.000181273	0.001003506
Gene 3111	N-Acetyl-D-glucosamine ABC transport system,	13399.76278	0.309050174	0.078419041	3.94E+00	8.11E-05	0.000489437
Gene 1840	D-serine dehydratase (EC 4.3.1.18)	1173.833271	0.308451896	0.115691984	2.666147515	0.007672602	0.027269397
Gene 21	Tungsten-containing aldehyde:ferredoxin	828.2427201	0.308275353	0.12643741	2.43816568	0.014762006	0.046842884
Gene 3062	Aspartyl-tRNA synthetase (EC 6.1.1.12)	10849.5917	0.308269167	0.086016648	3.583831423	0.000338591	0.001729995

Gene 2845	Phosphate regulon sensor protein PhoR (SphS) (EC	3039.000236	0.308018098	0.086994016	3.540681448	0.000399095	0.002017377
Gene 3517	Permeases of the major facilitator superfamily	3816.110053	0.308007092	0.089411758	3.444816425	0.000571447	0.002782554
Gene 4117	Methylcobalamin methyltransferase MMP0830	1873.913832	0.307847036	0.106284131	2.896453424	0.003774067	0.014845557
Gene 2963	FIG00532938: hypothetical protein	1677.204061	0.303472831	0.106386989	2.852537079	0.004337175	0.016781882
Gene 1777	FIG00530205: hypothetical protein	2363.934448	0.302144874	0.117126331	2.579649443	0.009890066	0.033670901
Gene 1007	FIG00513736: hypothetical protein	944.6170852	0.301767256	0.124437294	2.425054782	0.015306077	0.048154602
Gene 3812	Xylulose kinase (EC 2.7.1.17)	807.5643984	0.301498951	0.123689616	2.437544557	0.014787392	0.046878576
Gene 4130	Inner membrane protein	2975.497839	0.301125065	0.090985244	3.309603307	0.000934283	0.004332616
Gene 4137	TsaB protein, required for	1600.559942	0.295251611	0.099820994	2.957810078	0.003098322	0.012408258
Gene 191	ATPase component of general energizing module of	3770.844445	0.294033582	0.086685211	3.391969399	0.000693922	0.003301356
Gene 3100	hypothetical protein	2130.869585	0.292002351	0.116644703	2.503348579	0.012302431	0.040231619
Gene 1106	5-methyltetrahydrofolate--homocysteine	7585.675389	0.291451242	0.075828453	3.843560438	0.000121262	0.000700532
Gene 1845	1-phosphofructokinase (EC 2.7.1.56)	22172.26776	0.291403527	0.081151842	3.59084303	0.00032961	0.001691931
Gene 601	Heat shock protein 60 family chaperone GroEL	25915.5733	0.290465912	0.075333228	3.855747606	0.000115376	0.000675951
Gene 791	33 kDa chaperonin	1224.890997	0.289325576	0.111050231	2.605357701	0.009177839	0.031725464

Gene 2022	HMP-PP hydrolase (pyridoxal phosphatase) Cof,	1054.043879	0.288489249	0.115645504	2.49459978	0.012609927	0.040954473
Gene 1488	1-deoxy-D-xylulose 5-phosphate reductoisomerase	5356.653491	0.286155947	0.080776613	3.542559352	0.000396264	0.002006126
Gene 462	Excinuclease ABC subunit A	4230.48982	0.285470732	0.080640587	3.540037897	0.00040007	0.002019225
Gene 661	Site-specific recombinase	1626.044754	0.285399007	0.099212683	2.876638342	0.00401936	0.01575437
Gene 3922	Transcriptional regulator, DeoR family	1228.565515	0.285137794	0.107545891	2.65131277	0.008017955	0.028143156
Gene 2544	Helicase PriA essential for oriC/DnaA-independent	4328.973358	0.284534832	0.081051029	3.510564087	0.000447157	0.002237084
Gene 1429	Glycogen debranching enzyme (EC 3.2.1.-)	12586.53592	0.284008342	0.073403448	3.869141719	0.000109219	0.00064101
Gene 1113	Cobalt-zinc-cadmium resistance protein	1413.09624	0.283489494	0.109514222	2.588608937	0.009636446	0.032908811
Gene 3907	1-phosphofructokinase (EC 2.7.1.56)	964.4434573	0.283189148	0.117023104	2.419942204	0.015522974	0.048698375
Gene 1090	Exodeoxyribonuclease VII large subunit (EC	2092.14672	0.279876627	0.112154320	2.495460061	0.012579392	0.040935488
Gene 4101	Fructose ABC transporter, ATP-binding component	7462.672734	0.277596337	0.075594653	3.672168949	0.000240501	0.001282154
Gene 1044	Two-component response regulator BceR	1376.574014	0.273479456	0.106349207	2.571523236	0.010125222	0.034365646
Gene 1006	FIG00520221: hypothetical protein	1504.930746	0.272427958	0.107292759	2.539108852	0.011113534	0.037224728
Gene 3497	Glucose-6-phosphate isomerase (EC 5.3.1.9)	16135.93736	0.271916088	0.072123826	3.770128448	0.000163164	0.000907803
Gene 3937	Rubrerythrin	268136.0032	0.27037505	0.085319932	3.168955283	0.001529879	0.006728223

Gene 1613	Mobile element protein	1727.578449	0.269253276	0.101463563	2.653694276	0.007961592	0.027999436
Gene 4034	FIG00518220: hypothetical protein	6017.735686	0.268168788	0.078380265	3.421381487	0.000623039	0.003011657
Gene 1975	Putative amidotransferase similar to cobyric acid	1922.208863	0.267900966	0.102906838	2.603334918	0.009232172	0.031856276
Gene 3256	DUF819 domain-containing protein	62465.69085	0.266566148	0.067300095	3.96E+00	7.47E-05	0.000454388
Gene 92	hypothetical protein	1297.096852	0.266468732	0.108653079	2.452472901	0.014187807	0.045368145
Gene 908	SENSOR KINASE OF TWO-COMPONENT REGULATORY SYSTEM	6398.258463	0.26477837	0.092390445	2.865863123	0.00415874	0.01616692
Gene 3551	Catalyzes the cleavage of	2858.443362	0.262808407	0.090128477	2.915930852	0.003546291	0.013999405
Gene 587	peptidase, U32 family large subunit [C1]	2478.628458	0.26237689	0.1038665	2.52609734	0.011533752	0.038284206
Gene 3155	Transaldolase (EC 2.2.1.2)	4250.812439	0.260651201	0.081568643	3.195482852	0.001395971	0.006221827
Gene 583	FIG00513284: hypothetical protein	1990.810327	0.259799132	0.100162068	2.593787622	0.00949251	0.032517732
Gene 2867	Lysophospholipase (EC 3.1.1.5); Monoglyceride	2147.525743	0.259743609	0.093682745	2.772587506	0.005561257	0.020720367
Gene 3225	Putative deoxyribonuclease YcfH	2413.053708	0.259432809	0.097451424	2.662175666	0.007763735	0.027475502
Gene 1497	Anaerobic dehydrogenases, typically	10258.36479	0.258719278	0.079603889	3.250083382	0.001153712	0.005240696
Gene 545	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)	3361.794929	0.257030579	0.088260257	2.912189318	0.00358905	0.014151357
Gene 1288	Hydrolase, alpha/beta hydrolase fold family	2256.048586	0.25265434	0.092693004	2.725710976	0.006416316	0.023458106

Gene 1116	FIG042921: similarity to aminoacyl-tRNA editing	1794.94386	0.250972883	0.100654347	2.493413263	0.012652149	0.041011268
Gene 1190	Discovery of flavin-dependent squalene desaturase	5614.595471	0.248257169	0.089652177	2.769114778	0.005620883	0.020895569
Gene 1195	DNA mismatch repair protein MutS	2313.457729	0.247632816	0.092408585	2.679759835	0.0073675	0.026382969
Gene 1042	acetyl esterase family enzyme	19867.97569	0.245644791	0.071470524	3.437008402	0.000588177	0.002859818
Gene 3110	N-Acetyl-D-glucosamine ABC transport system,	11470.5976	0.242550725	0.073745379	3.289029468	0.001005335	0.004623704
Gene 910	metal dependent phosphohydrolase	2042.097028	0.239613258	0.095243038	2.51580865	0.011875964	0.039184773
Gene 1117	Isopentenyl-diphosphate delta-isomerase (EC	5573.137943	0.239131241	0.080664549	2.96451468	0.003031608	0.012218628
Gene 3763	Two-component sensor histidine kinase	2624.67997	0.238091702	0.094371201	2.522927532	0.011638238	0.038515367
Gene 1499	Neuraminidase NanP	8095.899075	0.237708273	0.07537656	3.153610008	0.001612644	0.007026976
Gene 3633	Putative stomatin/prohibitin-family membrane	4193.229573	0.237181789	0.086935991	2.728234735	0.006367428	0.023330819
Gene 3899	Xylulose kinase (EC 2.7.1.17)	3033.325351	0.236359661	0.087698422	2.695141542	0.007035874	0.025361498
Gene 1043	2,3-bisphosphoglycerate-independent	35021.96562	0.236211844	0.070072831	3.370947665	0.000749101	0.003538488
Gene 2023	ABC-type multidrug/protein/lipid transport	2867.173617	0.235624853	0.087908076	2.680355012	0.007354412	0.026364572
Gene 1976	proposed amino acid ligase found clustered with	4213.697725	0.232891591	0.089151127	2.612323574	0.008992909	0.031209721
Gene 3061	Aspartyl-tRNA(Asn) amidotransferase subunit A (EC	3872.673634	0.232873447	0.087291139	2.667778754	0.007635452	0.02716648

Gene 3175	Phosphoglucosamine mutase (EC 5.4.2.10)	12788.14822	0.232439228	0.071414669	3.254782681	0.001134792	0.005168914
Gene 1493	Multi antimicrobial extrusion protein (Na(+)/drug	5373.670469	0.232402957	0.08262821	2.812634526	0.004913746	0.01851589
Gene 3154	Transketolase, N-terminal section (EC 2.2.1.1)	6937.027039	0.231747243	0.085765942	2.702089412	0.006890523	0.024917093
Gene 1168	Cytidylate kinase (EC 2.7.4.25)	5462.717498	0.231445827	0.078516339	2.947740926	0.003201052	0.012788782
Gene 566	Branched-chain amino acid transport system	2832.39008	0.228947966	0.093721735	2.442848147	0.014571867	0.046328199
Gene 3609	hypothetical protein	2756.941597	0.228054612	0.090747184	2.513076467	0.011968339	0.039437944
Gene 1490	Undecaprenyl diphosphate synthase (EC 2.5.1.31)	4487.086884	0.227802064	0.079957129	2.849052573	0.004384963	0.016907603
Gene 2545	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-	7500.213544	0.226587329	0.075122543	3.016236144	0.002559339	0.010582003
Gene 3227	Methionyl-tRNA synthetase (EC 6.1.1.10)	6815.338905	0.225222709	0.079360862	2.837956938	0.004540331	0.017405476
Gene 781	Glycosyltransferase (EC 2.4.1.-)	4019.559423	0.224614882	0.086221115	2.605101893	0.009184694	0.031725464
Gene 1103	Glycogen synthase, ADP-glucose transglucosylase	6994.50192	0.224083476	0.075217153	2.979153927	0.002890455	0.011746017
Gene 192	Transmembrane component of general energizing	3931.95486	0.223541743	0.088848254	2.515994772	0.011869694	0.039184773
Gene 3459	Rubrerythrin	215753.683	0.223493682	0.074298104	3.008067105	0.002629151	0.010803301
Gene 2542	Methionyl-tRNA formyltransferase (EC 2.1.2.9)	3294.00597	0.223001221	0.085832884	2.598086074	0.009374499	0.032313762
Gene 3166	Cell division protein FtsH (EC 3.4.24.-)	6055.276863	0.221355128	0.088174492	2.510421359	0.012058718	0.039630038

Gene 3099	Inner membrane protein	2860.952584	0.221069413	0.091778582	2.408725532	0.016008332	0.049984585
Gene 220	Probable M18-family aminopeptidase 1 (EC	3440.949766	0.219218142	0.084587701	2.591607763	0.009552862	0.032657
Gene 3463	Negative regulator of genetic competence MecA	14024.51235	0.214866307	0.080238586	2.677842629	0.007409802	0.02650583
Gene 544	Glucose-1-phosphate thymidylyltransferase (EC	4912.499353	0.213693221	0.083311253	2.564998295	0.010317631	0.034911495
Gene 3575	Oligopeptide transport ATP-binding protein OppF	9205.154966	0.211813336	0.078584258	2.695365992	0.007031135	0.025361498
Gene 1276	Cell division protein FtsK	11194.60598	0.209030291	0.076106954	2.746533387	0.006022876	0.022166321
Gene 3109	Putative predicted metal-dependent hydrolase	10336.49245	0.208049211	0.072409966	2.873212403	0.004063208	0.015888678
Gene 3066	RNA methyltransferase, TrmA family	6896.663796	0.207560982	0.077952007	2.662676562	0.007752189	0.027463951
Gene 546	dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133)	4864.262959	0.204255779	0.079570552	2.566977015	0.010258941	0.034783895
Gene 3576	Oligopeptide transport ATP-binding protein OppD	18682.24077	0.204072949	0.078740675	2.591709404	0.009550004	0.032657
Gene 194	Pyrophosphate-energized proton pump (EC 3.6.1.1)	45903.58997	0.203987549	0.068130463	2.994072563	0.002752805	0.011255611
Gene 221	hypothetical protein	5528.672907	0.203013684	0.080240405	2.529953103	0.011407777	0.037980011
Gene 907	Peptide synthetase	26371.56884	0.199476291	0.070194577	2.841762137	0.004486495	0.017219002
Gene 199	Sodium-dependent transporter	14574.93828	0.180111344	0.072127707	2.497139934	0.012519975	0.040806926
Gene 3458	Rubrerythrin	78039.72233	0.17748512	0.073134508	2.426831385	0.015231333	0.048010552

Gene 4033	Glutamate synthase [NADPH] large chain (EC	75751.45384	0.174462113	0.067643065	2.579157419	0.009904164	0.033684316
Gene 1040	Phosphoglycerate kinase (EC 2.7.2.3)	92164.07524	0.17209329	0.069364453	2.481001166	0.013101394	0.042384609
Gene 251	Di-/tripeptide transporter	174225.6396	0.167388076	0.068180761	2.455063193	0.014085982	0.045086021

8.2.3 Differential Gene Expression: 8 Hours

GeneID	GeneName	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Gene 3022	baiF	22105.20827	8.97730109	0.158695093	56.56949394	0	0
Gene 1825	baiA1	18956.22842	8.873527262	0.14030314	63.24539323	0	0
Gene 3027	baiB	7652.656926	8.762963632	0.177842715	49.27367218	0	0
Gene 3020	baiH	36530.49465	8.661588175	0.141813731	61.07721801	0	0
Gene 3021	baiG	15767.66148	8.526515698	0.159949389	53.3075852	0	0
Gene 1824	baiA1	8252.180503	8.509581119	0.166597008	51.07883527	0	0
Gene 3023	baiA2	11522.45635	8.497932037	0.177884303	47.77224245	0	0
Gene 3025	baiCD	12793.32156	8.467676604	0.17769641	47.65249121	0	0
Gene 3026	baiCD	7790.932709	8.313767607	0.176370195	47.13816644	0	0
Gene 3019	baiI	10431.2133	8.202717357	0.167634505	48.93215376	0	0
Gene 3024	baiE	6809.310708	8.115640669	0.183241947	44.28920779	0	0
Gene 719	FAD binding flavoprotein	9231.556886	5.730736732	0.161749991	3.54E+01	5.98E-275	1.19E-272
Gene 1297	Drug resistance transporter EmrB/QacA subfamily	5034.389063	4.965027194	0.108527138	45.7491766	0	0
Gene 1298	Permease of the drug/metabolite transporter (DMT)	2369.172412	4.393280382	0.119971434	3.66E+01	1.41E-293	3.18E-291

Gene 2071	Permease of the drug/metabolite transporter (DMT)	10709.53091	3.843072345	0.101214879	37.96944076	0	0
Gene 3738	hypothetical protein	4050.244271	3.592108707	0.183071009	1.96E+01	1.02E-85	1.56E-83
Gene 2072	hypothetical protein	2416.860843	3.514220848	0.108441407	3.24E+01	2.21E-230	3.75E-228
Gene 1822	hypothetical protein	103.005551	3.512791721	0.258325754	1.36E+01	4.10E-42	5.79E-40
Gene 2068	N-acetyl-L,L-diaminopimelate deacetylase (EC	25431.94705	3.41015717	0.085854385	39.72024435	0	0
Gene 3016	FIG00518880: hypothetical protein	146.1599145	3.332704265	0.243445612	1.37E+01	1.17E-42	1.72E-40
Gene 2067	Permease of the drug/metabolite transporter (DMT)	6693.608419	3.308803999	0.097820825	3.38E+01	8.42E-251	1.59E-248
Gene 2066	membrane transport protein	21076.3992	3.30634826	0.09048682	3.65E+01	2.61E-292	5.54E-290
Gene 2065	hypothetical protein	6187.93805	3.274953026	0.099553054	3.29E+01	2.46E-237	4.39E-235
Gene 2063	hypothetical protein	4751.161771	3.211605836	0.104433901	3.08E+01	1.13E-207	1.83E-205
Gene 3017	hypothetical protein	67.21867167	3.002205434	0.269268456	1.11E+01	7.20E-29	9.04E-27
Gene 2935	N-Acetyl-D-glucosamine ABC transport system,	543.2080748	1.486689904	0.235758617	6.31E+00	2.86E-10	1.59E-08
Gene 2062	Diacylglycerol kinase (EC 2.7.1.107)	698.3930051	1.480600583	0.136633418	1.08E+01	2.32E-27	2.80E-25
Gene 2934	N-Acetyl-D-glucosamine ABC transport system,	563.5470821	1.460109573	0.240050769	6.08E+00	1.18E-09	6.27E-08
Gene 1937	DegV family protein	4821.305422	1.414612702	0.121562984	1.16E+01	2.68E-31	3.63E-29

Gene 2933	hypothetical protein	147.7829357	1.410260564	0.251743971	5.60E+00	2.12E-08	9.45E-07
Gene 2175	hypothetical protein	787.0595057	1.399615374	0.129261474	1.08E+01	2.54E-27	2.97E-25
Gene 1546	hypothetical protein	110.55321	1.29206959	0.224659557	5.75E+00	8.86E-09	4.17E-07
Gene 2122	Alanine dehydrogenase (EC 1.4.1.1)	47407.78888	1.267809794	0.197203222	6.43E+00	1.28E-10	7.64E-09
Gene 2928	SN-glycerol-3-phosphate transport ATP-binding	437.5789501	1.254896273	0.22875303	5.49E+00	4.12E-08	1.77E-06
Gene 1774	Probable Hsp20-family chaperone	123.3135358	1.225193097	0.211476821	5.79E+00	6.89E-09	3.34E-07
Gene 2929	probable sugar ABC transporter, permease protein	292.3390204	1.194111267	0.245206398	4.87E+00	1.12E-06	3.44E-05
Gene 2930	N-Acetyl-D-glucosamine ABC transport system,	256.4384034	1.173624761	0.257725655	4.55E+00	5.27E-06	0.00013434
Gene 812	Chaperone protein DnaK	33509.6263	1.091104952	0.110457159	9.88E+00	5.18E-23	5.67E-21
Gene 1545	FIG00647252: hypothetical protein	206.578318	1.08715029	0.187644766	5.79E+00	6.89E-09	3.34E-07
Gene 841	FIG00516872: hypothetical protein	719.7293649	1.08289247	0.230990606	4.69E+00	2.76E-06	7.79E-05
Gene 600	Heat shock protein 60 family co-chaperone GroES	3306.154275	1.082374168	0.116613448	9.28E+00	1.67E-20	1.77E-18
Gene 94	Scaffold protein	265.6296645	1.068280795	0.168710617	6.33E+00	2.42E-10	1.39E-08
Gene 3405	Transcriptional regulator, AraC family	1599.162293	1.063048734	0.186548385	5.70E+00	1.21E-08	5.61E-07
Gene 95	Phage capsid protein	681.4425152	1.059017228	0.126270429	8.39E+00	4.99E-17	4.57E-15

Gene 65	FIG00530366: hypothetical protein	238.1851092	1.057235357	0.17094094	6.18E+00	6.22E-10	3.35E-08
Gene 2060	Ubiquinone biosynthesis monooxygenase UbiB	8415.084365	1.023320658	0.08878479	1.15E+01	9.74E-31	1.27E-28
Gene 622	flavin reductase domain protein, FMN-binding	22350.39379	1.020834291	0.143630103	7.11E+00	1.18E-12	8.72E-11
Gene 811	Heat shock protein GrpE	7412.906021	0.994597701	0.125051485	7.95E+00	1.81E-15	1.58E-13
Gene 602	Heat shock protein 60 family chaperone GroEL	16945.70725	0.951608922	0.11175364	8.52E+00	1.66E-17	1.66E-15
Gene 601	Heat shock protein 60 family chaperone GroEL	34685.88035	0.942766964	0.112096821	8.41E+00	4.09E-17	3.95E-15
Gene 3460	Peroxide stress regulator PerR, FUR family	14693.70181	0.938554711	0.17751238	5.29E+00	1.24E-07	4.78E-06
Gene 2692	FIG00529669: hypothetical protein	230.9377258	0.937512994	0.176623056	5.31E+00	1.11E-07	4.32E-06
Gene 86	hypothetical protein	246.558698	0.930869895	0.178399397	5.22E+00	1.81E-07	6.67E-06
Gene 813	Chaperone protein DnaJ	13973.02875	0.930524536	0.110681565	8.41E+00	4.20E-17	3.95E-15
Gene 2061	hypothetical protein	760.3170541	0.929532736	0.136690186	6.80E+00	1.04E-11	6.68E-10
Gene 2663	Chaperone protein HtpG	12122.1732	0.921631636	0.120453582	7.65E+00	1.99E-14	1.69E-12
Gene 720	Pyruvate formate-lyase (EC 2.3.1.54)	613.0295264	0.920710946	0.154941369	5.94E+00	2.81E-09	1.44E-07
Gene 97	Phage capsid and scaffold	126.1043828	0.9155517	0.210305265	4.35E+00	1.34E-05	0.000298893
Gene 102	Phage tail length tape-measure protein	885.365741	0.90417158	0.125842228	7.18E+00	6.72E-13	5.18E-11

Gene 103	Phage tail length tape-measure protein	281.518532	0.902342986	0.168955957	5.34E+00	9.26E-08	3.69E-06
Gene 810	Heat-inducible transcription repressor HrcA	7754.585765	0.897875326	0.119681778	7.50E+00	6.28E-14	5.19E-12
Gene 99	FIG00528782: hypothetical protein	241.0528632	0.896312744	0.175105099	5.12E+00	3.08E-07	1.05E-05
Gene 3937	Rubrerythrin	171706.5118	0.89581737	0.272359093	3.28910395	0.001005069	0.01191323
Gene 107	hypothetical protein	133.1216964	0.886759256	0.208979337	4.24E+00	2.20E-05	0.000463798
Gene 2353	hypothetical protein	403.2983164	0.877171346	0.167071037	5.25E+00	1.52E-07	5.66E-06
Gene 3652	FIG00519028: hypothetical protein	25000.70029	0.876533903	0.173902379	5.04E+00	4.65E-07	1.54E-05
Gene 1544	Periplasmic [Fe] hydrogenase (EC 1.12.7.2)	390.8127188	0.872788541	0.173100801	5.04E+00	4.60E-07	1.54E-05
Gene 777	no significant homology.	7437.732075	0.870774068	0.172132857	5.06E+00	4.22E-07	1.43E-05
Gene 620	Transcriptional regulator, GntR family	3909.961605	0.868784434	0.100900949	8.61E+00	7.29E-18	7.49E-16
Gene 3774	hypothetical protein	14652.92173	0.867226539	0.147450079	5.88E+00	4.07E-09	2.03E-07
Gene 106	hypothetical protein	308.469244	0.864013956	0.158786841	5.44E+00	5.29E-08	2.21E-06
Gene 66	hypothetical protein	97.821690 82	0.863372892	0.224671681	3.84282 0289	0.00012 1629	0.00204 1192
Gene 1543	Periplasmic [Fe] hydrogenase (EC 1.12.7.2)	85.7505674	0.85923842	0.230674101	3.72490 1993	0.00019 5391	0.00308 0818
Gene 88	Phage uncharacterized protein-like	648.2663162	0.850907361	0.129450982	6.57E+00	4.92E-11	3.04E-09

Gene 90	putative phage protein	269.2706689	0.848203385	0.174971775	4.85E+00	1.25E-06	3.82E-05
Gene 2937	Vancomycin response regulator VanR	1392.543828	0.846147544	0.193927518	4.36E+00	1.28E-05	0.000287734
Gene 2938	D-alanyl-D-alanine carboxypeptidase (EC	1076.768355	0.83778512	0.276382336	3.031254207	0.002435401	0.023521396
Gene 3669	FIG00264004: hypothetical protein	3430.47569	0.831303225	0.138904368	5.98E+00	2.17E-09	1.13E-07
Gene 3653	FIG00519028: hypothetical protein	10715.448	0.824345039	0.187060922	4.41E+00	1.05E-05	0.00024356
Gene 89	Portal protein, phage associated	650.7944421	0.820440063	0.138454501	5.93E+00	3.11E-09	1.57E-07
Gene 2936	Sensor histidine kinase VanS (EC 2.7.3.-)	2512.995935	0.806500595	0.198015997	4.07E+00	4.64E-05	0.000903459
Gene 3667	no significant homology.	3094.5931	0.801635085	0.161675361	4.96E+00	7.11E-07	2.27E-05
Gene 91	hypothetical protein	90.53454344	0.801620771	0.237261854	3.378633181	0.000728471	0.009283902
Gene 104	hypothetical protein	282.6316813	0.79884836	0.17248141	4.63E+00	3.63E-06	9.77E-05
Gene 105	hypothetical protein	191.3784722	0.795621086	0.194261251	4.10E+00	4.21E-05	0.000825029
Gene 67	hypothetical protein	42.678281 87	0.793940117	0.267069609	2.972783463	0.002951126	0.027159193
Gene 111	FIG01032338: hypothetical protein	342.112652	0.791779044	0.169174078	4.68E+00	2.87E-06	8.03E-05
Gene 3773	Ferrous iron transport protein B	189819.6343	0.785032351	0.141946489	5.53E+00	3.19E-08	1.41E-06
Gene 64	hypothetical protein	55.99386917	0.776844326	0.255474333	3.040792075	0.002359567	0.023051679

Gene 2932	Probable Co/Zn/Cd efflux system membrane fusion	915.6367244	0.775900864	0.276498836	2.806163225	0.005013529	0.041374797
Gene 844	two-component sensor histidine kinase	4018.236356	0.764466714	0.19153692	3.99E+00	6.57E-05	0.001211063
Gene 879	Glycine dehydrogenase [decarboxylating] (glycine	4884.671426	0.76335127	0.190210078	4.01E+00	5.99E-05	0.001128133
Gene 3654	ABC transporter ATP-binding/transmembrane	29736.66848	0.760877199	0.168756895	4.51E+00	6.52E-06	0.000164998
Gene 2549	ClpB protein	3094.984734	0.760636899	0.1080077	7.04E+00	1.89E-12	1.28E-10
Gene 3459	Rubrerythrin	198419.0612	0.759850719	0.140776601	5.40E+00	6.76E-08	2.79E-06
Gene 96	hypothetical protein	191.1735236	0.755331757	0.183917934	4.11E+00	4.01E-05	0.000790367
Gene 2639	Mg(2+) transport ATPase protein C	236.2632187	0.754767002	0.206412496	3.656595481	0.000255587	0.003902886
Gene 87	Phage protein	293.1773829	0.752775439	0.161302383	4.67E+00	3.06E-06	8.50E-05
Gene 2454	[Fe] hydrogenase, HymA subunit, putative	216.9441735	0.74561821	0.235453954	3.166726223	0.001541654	0.016968203
Gene 2455	hypothetical protein	154.352091	0.744828524	0.249164589	2.989303287	0.002796144	0.026041014
Gene 2694	ABC-type multidrug transport system, ATPase	468.0671123	0.744128101	0.160866237	4.63E+00	3.73E-06	9.96E-05
Gene 2693	FIG00520859: hypothetical protein	330.7726502	0.743432769	0.16928309	4.39E+00	1.12E-05	0.000257667
Gene 778	FIG00533233: hypothetical protein	2115.750901	0.739462753	0.174567906	4.24E+00	2.28E-05	0.000473304
Gene 876	Aminomethyltransferase (glycine cleavage system T	757.6512597	0.734421243	0.177814924	4.13E+00	3.62E-05	0.000726862

Gene 3771	hypothetical protein	7312.590061	0.731740076	0.142693803	5.13E+00	2.93E-07	1.02E-05
Gene 84	Integrase/recombinase (XerC/CodV family)	146.3221661	0.724004373	0.198762299	3.642563886	0.000269936	0.004049039
Gene 878	Glycine cleavage system H protein	1014.281146	0.717419067	0.203137351	3.531694507	0.000412906	0.005881309
Gene 69	Phage replication initiation protein	78.7419971	0.708315753	0.235776169	3.004187214	0.002662914	0.025286495
Gene 100	FIG00533106: hypothetical protein	69.56912609	0.703368389	0.243337198	2.890509115	0.003846184	0.033866397
Gene 3772	Ferrous iron transport protein A	5280.945228	0.697490502	0.135755389	5.14E+00	2.78E-07	9.81E-06
Gene 877	Aminomethyltransferase (glycine cleavage system T	1961.174405	0.690367668	0.187665081	3.678722032	0.000234406	0.003645114
Gene 1878	diglucosyldiacylglycerol synthase (LTA membrane	545.7711889	0.689303875	0.223116507	3.089434691	0.002005378	0.0206007
Gene 109	Holin, toxin secretion/phage lysis	113.2117362	0.687757349	0.219315091	3.135932626	0.001713086	0.018246766
Gene 98	hypothetical protein	101.6747103	0.68378077	0.225033909	3.038567712	0.002377057	0.023155812
Gene 3260	Probable cation-transporting ATPase	3176.552595	0.670184837	0.157723197	4.25E+00	2.15E-05	0.000454708
Gene 1875	Integral membrane protein LafC, accessory	368.5901534	0.663227127	0.218792537	3.031305987	0.002434983	0.023521396
Gene 2691	FIG00529669: hypothetical protein	253.1324866	0.660169323	0.174256641	3.788488736	0.000151566	0.002482175
Gene 1036	Iron-sulfur cluster assembly scaffold protein	9459.918713	0.659657591	0.148744092	4.43E+00	9.21E-06	0.00021996
Gene 887	Sodium/glycine symporter GlyP	9943.262397	0.659347809	0.193772894	3.402683399	0.000667276	0.008601004

Gene 619	ABC-type multidrug transport system, ATPase	8621.266668	0.656405878	0.100538542	6.53E+00	6.63E-11	4.01E-09
Gene 2275	Phage terminase, small subunit	88.29206271	0.655098127	0.236643746	2.768288355	0.005635157	0.045347235
Gene 2700	Transcriptional regulator, AraC family	3499.083055	0.653628372	0.16539413	3.95E+00	7.75E-05	0.001368692
Gene 2456	Tungsten-containing aldehyde:ferredoxin	2144.077829	0.652008318	0.210072426	3.103731078	0.001910969	0.019810981
Gene 1877	glycosyl transferase, group 2 family	359.9578841	0.651720855	0.238688529	2.730423867	0.006325294	0.048733516
Gene 1033	Iron-sulfur cluster regulator IscR	9695.079871	0.645809739	0.153384951	4.21E+00	2.55E-05	0.000523777
Gene 3316	FIG01032298: hypothetical protein	371.4830737	0.639177634	0.161508928	3.96E+00	7.57E-05	0.001351119
Gene 880	Glycine dehydrogenase [decarboxylating] (glycine	3318.829978	0.638730738	0.192582608	3.316658474	0.000911009	0.010951492
Gene 3655	Tetracenomycin polyketide synthesis	10695.12184	0.628709678	0.141909005	4.43E+00	9.41E-06	0.000223007
Gene 2070	Regulatory protein (induces abgABT, used to	974.0937601	0.628341082	0.121458418	5.17E+00	2.30E-07	8.29E-06
Gene 3787	hypothetical protein	15378.54506	0.625971893	0.114926318	5.45E+00	5.13E-08	2.17E-06
Gene 4091	FIG00525551: hypothetical protein	190.2144398	0.623373835	0.199675832	3.121929314	0.001796701	0.018915576
Gene 2255	Retron-type RNA-directed DNA polymerase (EC	225.2462646	0.620825793	0.178155151	3.484747923	0.000492601	0.006706494
Gene 3668	Flavodoxin	2941.587349	0.615905738	0.130885653	4.71E+00	2.53E-06	7.21E-05
Gene 2123	transcriptional regulator, RpiR family	1470.359711	0.615175395	0.153476304	4.01E+00	6.12E-05	0.001139256

Gene 3461	hypothetical protein	1413.53762	0.613810213	0.116410296	5.27E+00	1.34E-07	5.12E-06
Gene 4089	Alcohol dehydrogenase (EC 1.1.1.1); Acetaldehyde	112371.4539	0.61123878	0.108848293	5.62E+00	1.96E-08	8.98E-07
Gene 3972	Biotin carboxylase of acetyl-CoA carboxylase (EC	1120.479578	0.605942491	0.138492276	4.38E+00	1.21E-05	0.000274085
Gene 952	Signal peptidase I (EC 3.4.21.89)	678.0227968	0.594192711	0.124702147	4.76E+00	1.89E-06	5.57E-05
Gene 2868	hypothetical protein	869.8885846	0.586070035	0.144744466	4.05E+00	5.14E-05	0.000979623
Gene 881	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	3287.230632	0.581144553	0.159018808	3.654581631	0.000257602	0.003916008
Gene 3100	hypothetical protein	4395.477571	0.578161635	0.134709748	4.29E+00	1.77E-05	0.000382499
Gene 3651	hypothetical protein	4949.477951	0.573203665	0.16575332	3.45817306	0.000543852	0.007201789
Gene 3458	Rubrerythrin	73772.11802	0.572153768	0.122721376	4.66E+00	3.13E-06	8.62E-05
Gene 3511	Substrate-specific component BioY of biotin ECF	220.0575118	0.563674747	0.184923578	3.048149683	0.002302552	0.022559686
Gene 3793	hypothetical protein	657.2210202	0.563554101	0.14549972	3.873231517	0.000107402	0.001820286
Gene 618	ABC transporter, ATP-binding protein	12387.76669	0.5594809	0.11577576	4.83E+00	1.35E-06	4.08E-05
Gene 4087	Formate dehydrogenase-O, major subunit (EC	5940.701985	0.558213029	0.106276835	5.25E+00	1.50E-07	5.65E-06
Gene 4088	Periplasmic [Fe] hydrogenase large subunit (EC	7925.355385	0.555532694	0.112693399	4.93E+00	8.24E-07	2.61E-05
Gene 1035	Cysteine desulfurase (EC 2.8.1.7)	35739.88801	0.555396224	0.145717503	3.811458567	0.000138149	0.002295715

Gene 3838	ABC transporter, ATP-binding protein	509.8542293	0.552111925	0.147331753	3.747406208	0.000178673	0.002845484
Gene 2016	Argininosuccinate synthase (EC 6.3.4.5)	2117.333222	0.550257707	0.1168748	4.71E+00	2.50E-06	7.18E-05
Gene 3259	FIG00267170: hypothetical protein	1466.834576	0.548350254	0.163763009	3.348437834	0.000812685	0.010091584
Gene 3892	Transcriptional regulator of sugar metabolism	2974.082886	0.542154592	0.145369687	3.729488601	0.000191869	0.003039417
Gene 181	Fibronectin-binding protein	683.7073833	0.539095338	0.130496533	4.13E+00	3.61E-05	0.000726862
Gene 256	FIG00533145: hypothetical protein	181.2730746	0.536556594	0.195619268	2.742861684	0.006090633	0.047684174
Gene 1034	Cysteine desulfurase (EC 2.8.1.7)	1453.735403	0.531968377	0.163277179	3.258069379	0.00112173	0.013112634
Gene 3973	Acetyl-coenzyme A carboxyl transferase alpha	18419.61249	0.531309394	0.119650245	4.44E+00	8.97E-06	0.000215761
Gene 3974	Enoyl-[acyl-carrier-protein] reductase [FMN] (EC	8136.166734	0.527572519	0.097778204	5.40E+00	6.83E-08	2.79E-06
Gene 1037	tRNA-specific 2-thiouridylase MnmA	18134.09285	0.52338273	0.121810306	4.30E+00	1.73E-05	0.000376718
Gene 3977	NAD-dependent protein deacetylase of SIR2 family	6689.173593	0.515785091	0.096079851	5.37E+00	7.95E-08	3.21E-06
Gene 1113	Cobalt-zinc-cadmium resistance protein	1051.487953	0.514497236	0.121333897	4.24E+00	2.23E-05	0.000467023
Gene 2690	FIG00518250: hypothetical protein	237.1986714	0.510825693	0.180051921	2.837102159	0.004552504	0.038582474
Gene 3650	Transcriptional regulator, TetR family	7374.671985	0.503071474	0.166949908	3.013307882	0.002584166	0.024607648
Gene 2766	C-5 cytosine-specific DNA methylase family	315.2958793	0.500281663	0.169944448	2.943795266	0.003242144	0.029248239

Gene 2022	HMP-PP hydrolase (pyridoxal phosphatase) Cof,	1767.354099	0.495831587	0.131725827	3.764118242	0.000167138	0.002698081
Gene 3839	hypothetical protein	975.9894324	0.489767992	0.122550274	4.00E+00	6.43E-05	0.001191041
Gene 3975	Transcriptional regulator of fatty acid	3183.315639	0.486087366	0.102834389	4.73E+00	2.28E-06	6.61E-05
Gene 2623	ABC transporter permease protein	3099.93307	0.480205456	0.173955637	2.760505287	0.005771202	0.046090895
Gene 2023	ABC-type multidrug/protein/lipid transport	4798.259894	0.479358545	0.114527921	4.19E+00	2.85E-05	0.000581034
Gene 3891	Transaldolase (EC 2.2.1.2)	1207.052698	0.473184838	0.150939096	3.134938865	0.001718899	0.018246766
Gene 2015	Argininosuccinate synthase (EC 6.3.4.5)	623.3541812	0.469141291	0.152512979	3.076074529	0.002097454	0.021224983
Gene 3895	Rhamnulokinase (EC 2.7.1.5)	1008.201672	0.462681684	0.143816123	3.217175344	0.001294595	0.014677844
Gene 830	Cys regulon transcriptional activator CysB	1817.824662	0.460402277	0.114805956	4.01E+00	6.07E-05	0.001135944
Gene 3976	Putative hydrolase	5499.547693	0.459103457	0.099412907	4.62E+00	3.87E-06	0.000102543
Gene 1038	Histidinol-phosphatase (EC 3.1.3.15)	20393.64979	0.459059674	0.118271136	3.881417648	0.000103849	0.001769091
Gene 599	FIG00518951: hypothetical protein	1812.023889	0.455583127	0.166217776	2.740881156	0.006127467	0.047751981
Gene 3971	Biotin carboxylase of acetyl-CoA carboxylase (EC	12779.49736	0.455204812	0.134857642	3.375446902	0.000736959	0.009321983
Gene 182	NPQTN specific sortase B	408.3316482	0.454318474	0.153602621	2.957752096	0.003098912	0.028164378
Gene 544	Glucose-1-phosphate thymidylyltransferase (EC	4643.05557	0.453732914	0.092827137	4.89E+00	1.02E-06	3.17E-05

Gene 3729	hypothetical protein	1059.310763	0.453506104	0.111375595	4.07E+00	4.66E-05	0.000903459
Gene 1576	FIG00518389: hypothetical protein	621.3919402	0.447895264	0.141588147	3.163366938	0.001559556	0.017109695
Gene 3688	hypothetical protein	487.5607098	0.441632506	0.136223181	3.241977637	0.001187033	0.013687221
Gene 2077	hypothetical protein	679.7432026	0.440071357	0.134580304	3.269953658	0.001075651	0.012617498
Gene 1621	FIG00669172: hypothetical protein	466.9936927	0.434258882	0.154135416	2.817385479	0.004841638	0.040327155
Gene 56	hypothetical protein	1185.715636	0.434184139	0.109254444	3.97E+00	7.07E-05	0.001280886
Gene 252	Xaa-Pro aminopeptidase (EC 3.4.11.9)	26971.11773	0.42678442	0.097507743	4.38E+00	1.20E-05	0.000273847
Gene 547	dTDP-4-dehydrorhamnose 3,5-epimerase (EC	4652.836429	0.423818934	0.120773502	3.509204649	0.000449449	0.00627009
Gene 3996	hypothetical protein	3935.462431	0.422815484	0.121260021	3.486844399	0.000488756	0.006698805
Gene 2926	membrane protein	974.8862859	0.420357596	0.137811198	3.050225364	0.002286697	0.022469284
Gene 3981	[FeFe]-hydrogenase maturation protein HydE	4690.740799	0.419918745	0.151323356	2.774977268	0.005520635	0.044792064
Gene 546	dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133)	5587.805539	0.418307316	0.115113449	3.633887007	0.000279202	0.004151286
Gene 545	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)	3420.504086	0.417996381	0.114180481	3.660839214	0.00025139	0.003868828
Gene 1986	LSU ribosomal protein L31p LSU ribosomal	22132.11275	0.41708012	0.120865929	3.450766666	0.000558997	0.007316597
Gene 3894	D-tagatose 3-epimerase (EC 5.3.1.-)	831.6316114	0.416913852	0.148465497	2.808153140	0.004982653	0.041298759

Gene 598	GTP-sensing transcriptional pleiotropic repressor	10558.81644	0.416171771	0.141553132	2.940039301	0.003281706	0.029509241
Gene 3728	hypothetical protein	1095.984673	0.414739591	0.110678521	3.747245516	0.000178787	0.002845484
Gene 3970	3-hydroxyacyl-[acyl-carrier-protein] dehydratase,	2837.737153	0.413577366	0.14759416	2.802125546	0.005076711	0.041771969
Gene 2621	FIG01116027: hypothetical protein	1413.2893	0.409393104	0.113588537	3.604176222	0.000313145	0.004595498
Gene 3893	Xylulose kinase (EC 2.7.1.17)	1147.859681	0.407992437	0.124682527	3.272250311	0.001066951	0.012558899
Gene 1957	Non-specific DNA-binding protein Dps /	10458.82417	0.404414561	0.102324274	3.95E+00	7.74E-05	0.001368692
Gene 92	hypothetical protein	1083.889813	0.402997429	0.120163459	3.353743584	0.000797262	0.009936465
Gene 253	Aminopeptidase YpdF (MP-, MA-, MS-, AP-, NP-	35067.04264	0.400594462	0.090018215	4.45E+00	8.58E-06	0.000209279
Gene 1618	Phage integrase (Site-specific recombinase)	529.4398345	0.397096153	0.137709099	2.883586884	0.003931742	0.034530063
Gene 2533	hypothetical protein	601.9767868	0.395909579	0.128327431	3.085151594	0.002034484	0.020723813
Gene 953	50S ribosomal subunit maturation GTPase RbgA (B.	1548.206708	0.39078528	0.116186374	3.363434678	0.000769791	0.009629483
Gene 21	Tungsten-containing aldehyde:ferredoxin	823.4864222	0.388435907	0.131958886	2.943613114	0.003244053	0.029248239
Gene 3656	Transcriptional regulator, AraC family	10923.80728	0.382706897	0.106298948	3.600288664	0.000317864	0.004644652
Gene 2024	Lipid A export ATP-binding/permease protein MsbA	5126.208165	0.379906584	0.120547928	3.151498238	0.001624351	0.017649203
Gene 3423	vrlL protein	1517.760041	0.378437991	0.124586697	3.03754734	0.002385119	0.023167778

Gene 3184	UDP-N-acetylmuramate--alanine ligase (EC	7868.977231	0.374441821	0.1136421	3.294917	0.000984491	0.011710262
Gene 3980	[FeFe]-hydrogenase maturation protein HydG	5755.34395	0.374250571	0.121514287	3.07988946	0.002070774	0.02101774
Gene 866	Protein export cytoplasm protein SecA ATPase RNA	1799.537788	0.373539185	0.135000418	2.766948356	0.005658372	0.045347235
Gene 3673	hypothetical protein	4633.553501	0.370370054	0.094251552	3.93E+00	8.51E-05	0.00148689
Gene 3255	hypothetical protein	1762.920876	0.369149536	0.107012327	3.449598247	0.000561421	0.007320071
Gene 280	Osmosensitive K+ channel histidine kinase KdpD	1321.315865	0.365767149	0.107953951	3.388177516	0.000703587	0.009034698
Gene 496	Glutamate synthase [NADPH] large chain (EC	2157.93326	0.363544658	0.103419763	3.515523391	0.000439367	0.006180305
Gene 2442	hypothetical protein	3823.32866	0.363225488	0.113078465	3.212154396	0.001317436	0.014887021
Gene 3672	Ferrous iron transport protein B	82375.86252	0.363156957	0.089333552	4.07E+00	4.80E-05	0.000924458
Gene 1275	Formate--tetrahydrofolate ligase (EC 6.3.4.3)	2303.321532	0.35717193	0.100860119	3.54126025	0.000398221	0.005720202
Gene 3675	FIG00645149: hypothetical protein	2417.150054	0.351303699	0.099362634	3.53557152	0.000406894	0.005820131
Gene 3463	Negative regulator of genetic competence MecA	17538.92584	0.342775017	0.116968489	2.930490239	0.003384276	0.030351049
Gene 3424	site-specific DNA-methyltransferase	2431.275836	0.339452113	0.114280312	2.970346389	0.002974641	0.0272123
Gene 3102	Chaperone protein DnaJ	9273.525616	0.336348503	0.121912088	2.758943	0.005798865	0.046145896
Gene 1777	FIG00530205: hypothetical protein	4881.07628	0.336310108	0.116849786	2.878140566	0.004000268	0.035041112

Gene 579	FIG00514008: hypothetical protein	13443.15231	0.333343922	0.094506967	3.527188857	0.000419997	0.005957281
Gene 1272	FIG01032388: hypothetical protein	5105.676757	0.326814582	0.100703509	3.245314732	0.001173209	0.013667278
Gene 1575	FIG00518389: hypothetical protein	1651.886362	0.317910884	0.105317355	3.018599188	0.002539463	0.024347428
Gene 257	FIG00533145: hypothetical protein	3542.391084	0.316937993	0.115028593	2.755297488	0.005863879	0.046336948
Gene 1043	2,3-bisphosphoglycerate-independent	35391.55846	0.316777123	0.086345765	3.668704481	0.000243782	0.003773617
Gene 1274	Formate--tetrahydrofolate ligase (EC 6.3.4.3)	18908.01623	0.315607637	0.111879535	2.820959513	0.004788025	0.040077539
Gene 3400	Transcriptional regulator, TetR family	5159.57045	0.312353282	0.093347504	3.346134426	0.000819467	0.010138662
Gene 3674	FIG00253762: hypothetical protein	1977.375959	0.310485599	0.107953606	2.876102149	0.004026194	0.035177313
Gene 991	Electron transport complex protein RnfA	2667.965432	0.31015003	0.106873487	2.902029675	0.003707534	0.032730575
Gene 282	FIG00530794: hypothetical protein	2185.80011	0.309647243	0.097669069	3.170371611	0.001522441	0.016811319
Gene 3763	Two-component sensor histidine kinase	2542.079217	0.309147192	0.113124052	2.73281575	0.006279545	0.048510451
Gene 987	Electron transport complex protein RnfC	8871.99667	0.29853592	0.108161247	2.760100567	0.005778357	0.046090895
Gene 2357	hypothetical protein	1346.180144	0.294765256	0.106228168	2.774831397	0.005523033	0.044792064
Gene 1042	acetyl esterase family enzyme	17431.01379	0.294258487	0.099316231	2.962843884	0.003048111	0.027777142
Gene 3623	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	6847.661965	0.293814087	0.095554635	3.074828209	0.002106238	0.021250044

Gene 1430	Ribose ABC transport system, periplasmic	7461.875574	0.291483776	0.102074349	2.855602593	0.004295524	0.036958952
Gene 3622	Shikimate kinase I (EC 2.7.1.71)	1486.341408	0.291482397	0.105194892	2.770879764	0.005590507	0.045220687
Gene 3101	FIG01032250: hypothetical protein	14060.49329	0.290274938	0.094947809	3.057205239	0.002234112	0.022341125
Gene 1046	ABC transporter ATP-binding protein YvcR	2167.343062	0.286789745	0.101699194	2.819980523	0.004802656	0.040100998
Gene 990	Electron transport complex protein RnfE	3633.639669	0.286572428	0.103121995	2.778965115	0.005453238	0.044545728
Gene 3491	UDP-N-acetylglucosamine--N-acetylmuramyl-	5858.388371	0.283949823	0.091075833	3.117729629	0.001822499	0.019127777
Gene 3060	Aspartyl-tRNA(Asn) amidotransferase subunit A (EC	3207.870891	0.283561031	0.098952106	2.865639163	0.004161683	0.03617463
Gene 3671	hypothetical protein	7385.323533	0.281856817	0.093973823	2.999312005	0.002705901	0.02555154
Gene 1168	Cytidylate kinase (EC 2.7.4.25)	5811.814683	0.278899945	0.102362001	2.724643346	0.006437098	0.049482454
Gene 929	tRNA pseudouridine synthase A (EC 4.2.1.70)	4698.757552	0.269214677	0.094489870	2.849137976	0.004383786	0.037527869
Gene 855	hypothetical protein	6392.873692	0.263561672	0.091870562	2.868837046	0.004119840	0.035902975
Gene 1041	Triosephosphate isomerase (EC 5.3.1.1)	52929.90152	0.260258772	0.093579936	2.781138593	0.005416861	0.044355453
Gene 1047	ABC transporter permease protein	4409.00331	0.256500313	0.093453299	2.744689767	0.006056813	0.047529154
Gene 352	FIG01032222: hypothetical protein	5981.125953	0.255012314	0.090210112	2.826870624	0.004700531	0.039442578
Gene 1408	Preprotein translocase subunit YajC (TC	17556.00831	0.247615651	0.089386181	2.770178209	0.005602563	0.045220687

Gene 393	Fructose-1,6-bisphosphatase, Bacillus type (EC	9345.048043	0.240495521	0.085707927	2.80598926	0.005016236	0.041374797

8.2.4 Differential Gene Expression: 10 Hours

GeneID	**GeneName**	**baseMean**	**log2FoldChange**	**lfcSE**	**stat**	**pvalue**	**padj**
Gene 1825	baiA1	9328.372142	7.31280968	0.233870102	3.13E+01	1.24E-214	4.39E-211
Gene 3027	baiB	4325.186052	7.696490628	0.264245155	2.91E+01	1.67E-186	2.94E-183
Gene 1824	baiA1	3953.493621	7.406630072	0.25752328	2.88E+01	6.60E-182	7.75E-179
Gene 3025	baiCD	6115.368147	7.37228925	0.275960135	2.67E+01	3.15E-157	2.77E-154

Gene 3026	baiCD	3986.444999	7.3074196	0.276127062	2.65E+01	2.52E-154	1.78E-151
Gene 3023	baiA2	4847.388517	7.033068102	0.301493744	2.33E+01	2.34E-120	9.16E-118
Gene 3022	baiF	8499.194394	6.893546781	0.288119156	2.39E+01	1.64E-126	7.24E-124
Gene 3024	baiE	3045.835396	6.792430969	0.297861535	2.28E+01	4.19E-115	1.48E-112
Gene 3020	baiH	10963.81135	6.624239145	0.255593469	2.59E+01	4.27E-148	2.51E-145
Gene 3019	baiI	2823.935908	6.313208543	0.263271385	2.40E+01	4.51E-127	2.27E-124
Gene 3021	baiG	5528.308968	6.204818788	0.299043638	2.07E+01	1.25E-95	3.69E-93
Gene 719	FAD binding flavoprotein	5445.042164	4.901961567	0.22091528	2.22E+01	4.35E-109	1.40E-106
Gene 3850	TRAP-type C4-dicarboxylate transport system,	569.3898372	4.208063561	0.300747337	1.40E+01	1.74E-44	3.42E-42
Gene 3847	Threonine dehydrogenase and related Zn-dependent	369.5728971	4.029166977	0.304993967	1.32E+01	7.62E-40	1.34E-37
Gene 1297	Drug resistance transporter EmrB/QacA subfamily	4660.227721	4.00134542	0.263747308	1.52E+01	5.49E-52	1.49E-49
Gene 3854	5-keto-D-gluconate 5-reductase (EC 1.1.1.69)	778.8569653	3.969858651	0.280335773	1.42E+01	1.60E-45	3.31E-43
Gene 3856	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	888.524659	3.903301639	0.309475491	1.26E+01	1.80E-36	2.54E-34
Gene 1298	Permease of the drug/metabolite transporter (DMT)	2235.810008	3.843385014	0.256379048	1.50E+01	8.40E-51	2.12E-48
Gene 3853	4-Hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16)	450.2873929	3.784462645	0.292202379	1.30E+01	2.30E-38	3.53E-36

Gene 3855	6-phosphogluconate dehydrogenase, decarboxylating	1727.815507	3.631121911	0.277144744	1.31E+01	3.21E-39	5.15E-37
Gene 2072	hypothetical protein	3366.593358	3.582544585	0.257148653	1.39E+01	4.06E-44	7.53E-42
Gene 2071	Permease of the drug/metabolite transporter (DMT)	15670.26422	3.570962674	0.272347608	1.31E+01	2.82E-39	4.73E-37
Gene 2065	hypothetical protein	7673.120131	3.534044448	0.248677523	1.42E+01	7.79E-46	1.72E-43
Gene 3851	TRAP-type C4-dicarboxylate transport system,	326.4406161	3.272407499	0.346175975	9.45E+00	3.29E-21	3.32E-19
Gene 2063	hypothetical protein	5665.034539	3.269746129	0.267731522	1.22E+01	2.66E-34	3.35E-32
Gene 2932	Probable Co/Zn/Cd efflux system membrane fusion	2087.264647	3.249553734	0.265607177	1.22E+01	2.04E-34	2.66E-32
Gene 2068	N-acetyl-L,L-diaminopimelate deacetylase (EC	29938.37181	3.195304236	0.214549812	1.49E+01	3.66E-50	8.59E-48
Gene 2934	N-Acetyl-D-glucosamine ABC transport system,	1308.710419	3.075501435	0.312619596	9.84E+00	7.74E-23	8.27E-21
Gene 2935	N-Acetyl-D-glucosamine ABC transport system,	1363.167448	3.02838718	0.304176849	9.96E+00	2.37E-23	2.62E-21
Gene 2066	membrane transport protein	26370.86913	3.012597865	0.233735942	1.29E+01	5.20E-38	7.64E-36
Gene 2928	SN-glycerol-3-phosphate transport ATP-binding	1102.057952	3.002386148	0.284389101	1.06E+01	4.70E-26	5.71E-24
Gene 3738	hypothetical protein	1824.456023	2.968005425	0.236491433	1.26E+01	3.97E-36	5.38E-34
Gene 1822	hypothetical protein	44.917458877	2.941038061	0.384586805	7.65E+00	2.05E-14	1.86E-12
Gene 2933	hypothetical protein	357.8628244	2.78904113	0.306920804	9.09E+00	1.02E-19	9.96E-18

Gene 2931	hypothetical protein	1275.957287	2.725222566	0.315194627	8.65E+00	5.33E-18	5.08E-16
Gene 2067	Permease of the drug/metabolite transporter (DMT)	8396.547003	2.707489446	0.266343903	1.02E+01	2.83E-24	3.22E-22
Gene 3848	Gluconokinase (EC 2.7.1.12)	804.4419827	2.666028562	0.257434526	1.04E+01	3.92E-25	4.61E-23
Gene 3852	TRAP-type C4-dicarboxylate transport system,	1782.363814	2.630416076	0.271291511	9.70E+00	3.14E-22	3.25E-20
Gene 2930	N-Acetyl-D-glucosamine ABC transport system,	717.3635291	2.505705197	0.332153752	7.54E+00	4.56E-14	3.88E-12
Gene 2929	probable sugar ABC transporter, permease protein	748.4099697	2.460707264	0.32544937	7.56E+00	4.00E-14	3.53E-12
Gene 3849	Gluconokinase (EC 2.7.1.12)	23.619883	2.107806152	0.398481055	5.29E+00	1.23E-07	6.43E-06
Gene 2938	D-alanyl-D-alanine carboxypeptidase (EC	1530.470813	2.098719647	0.278266494	7.54E+00	4.62E-14	3.88E-12
Gene 3016	FIG00518880: hypothetical protein	28.677870 71	1.862900753	0.39398472	4.73E+00	2.26E-06	9.28E-05
Gene 2062	Diacylglycerol kinase (EC 2.7.1.107)	761.9951233	1.776584725	0.236771049	7.50E+00	6.22E-14	5.10E-12
Gene 2927	Transcriptional regulator, PadR family	382.2840293	1.662440386	0.297894286	5.58E+00	2.40E-08	1.36E-06
Gene 1937	DegV family protein	4633.382379	1.453766886	0.232247816	6.26E+00	3.86E-10	2.72E-08
Gene 840	glycosyltransferase	1724.072864	1.453663346	0.222305454	6.54E+00	6.19E-11	4.65E-09
Gene 2055	Alcohol dehydrogenase (EC 1.1.1.1)	37.876884 94	1.446476779	0.37708857	3.835907248	0.000125102	0.003446159
Gene 2056	2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate	97.77561687	1.408595287	0.316966 51	4.44E+00	8.83E-06	0.000311369

Gene 839	alkaline phosphatase-like protein	571.3306354	1.369831952	0.23680531	5.78E+00	7.27E-09	4.42E-07
Gene 841	FIG00516872: hypothetical protein	816.6342645	1.305798316	0.24005051	5.44E+00	5.34E-08	2.99E-06
Gene 2060	Ubiquinone biosynthesis monooxygenase UbiB	9007.805207	1.286475262	0.22312 4931	5.77E+00	8.13E-09	4.86E-07
Gene 842	two-component response regulator	619.2859217	1.273640523	0.24087 6656	5.29E+00	1.24E-07	6.43E-06
Gene 2691	FIG00529669: hypothetical protein	290.026517	1.268200837	0.26198 4693	4.84E+00	1.29E-06	5.56E-05
Gene 2690	FIG00518250: hypothetical protein	234.0862328	1.221225315	0.26879004	4.54E+00	5.53E-06	0.00020 9852
Gene 2061	hypothetical protein	791.531263	1.208494372	0.28959907	4.17E+00	3.01E-05	0.00097 2486
Gene 2694	ABC-type multidrug transport system, ATPase	525.0353971	1.190670378	0.24456 7676	4.87E+00	1.12E-06	5.02E-05
Gene 97	Phage capsid and scaffold	154.7365342	1.186899289	0.29044 4932	4.09E+00	4.38E-05	0.00139 1198
Gene 3903	Tagatose 1,6-bisphosphate aldolase (EC 4.1.2.40)	953.2011015	1.183384129	0.24031 9156	4.92E+00	8.47E-07	3.83E-05
Gene 2926	membrane protein	1342.257485	1.177764901	0.23054 5158	5.11E+00	3.25E-07	1.59E-05
Gene 3858	Ribose 5-phosphate isomerase B (EC 5.3.1.6)	1825.402973	1.157391486	0.31285 6353	3.69943 4179	0.00021 6081	0.00532 7975
Gene 102	Phage tail length tape-measure protein	941.5233857	1.117137179	0.24838 6855	4.50E+00	6.87E-06	0.00024 9854
Gene 3316	FIG01032298: hypothetical protein	566.3725856	1.114197785	0.24120 7334	4.62E+00	3.85E-06	0.00015 2578
Gene 3787	hypothetical protein	7523.481428	1.088407517	0.22770 5198	4.78E+00	1.75E-06	7.36E-05

Gene 95	Phage capsid protein	720.0179728	1.072659772	0.235037706	4.56E+00	5.02E-06	0.000194671
Gene 99	FIG00528782: hypothetical protein	242.1619195	1.064884034	0.280125519	3.801453146	0.00014385	0.003842537
Gene 105	hypothetical protein	218.586107	1.045943928	0.272621615	3.83661409	0.000124742	0.003446159
Gene 3904	Xylulose kinase (EC 2.7.1.17)	1539.363094	1.022451438	0.233926471	4.37E+00	1.24E-05	0.000423731
Gene 3901	Ribulose-5-phosphate 4-epimerase and related	909.96771	1.016930533	0.262766678	3.870089393	0.000108795	0.003093651
Gene 87	Phage protein	263.5257762	1.006232199	0.273246693	3.682504575	0.000230954	0.005655159
Gene 90	putative phage protein	284.4396361	1.003417303	0.26298153	3.815542872	0.000135884	0.003657454
Gene 3839	hypothetical protein	537.4430424	1.001496278	0.249615873	4.01E+00	6.02E-05	0.001860995
Gene 844	two-component sensor histidine kinase	4368.235984	0.994979897	0.221100052	4.50E+00	6.79E-06	0.00024943
Gene 96	hypothetical protein	200.9891957	0.970696721	0.276013869	3.516840384	0.000436716	0.00974596
Gene 3838	ABC transporter, ATP-binding protein	257.7637297	0.966299644	0.280623303	3.443404851	0.000574439	0.012128568
Gene 3315	Chromosome (plasmid) partitioning protein ParB	145.8841989	0.959811705	0.294516222	3.258943427	0.00111828	0.021665132
Gene 2692	FIG00529669: hypothetical protein	247.9641801	0.959772529	0.294436755	3.259689944	0.001115341	0.021665132
Gene 3907	1-phosphofructokinase (EC 2.7.1.56)	1015.861894	0.947763199	0.243399274	3.89E+00	9.87E-05	0.002898982
Gene 3878	ABC transporter, substrate-binding protein	25733.51781	0.945470013	0.273980202	3.450869827	0.000558783	0.012013835

Gene 2937	Vancomycin response regulator VanR	1796.367987	0.941946835	0.258511981	3.643726031	0.000268719	0.006489759
Gene 111	FIG01032338: hypothetical protein	368.5189143	0.941697003	0.262801216	3.583305356	0.000339274	0.007975191
Gene 3322	Superfamily II DNA and RNA helicase	277.485657	0.930398231	0.260490582	3.571715427	0.000354651	0.008173191
Gene 107	hypothetical protein	141.1664057	0.916522185	0.294504706	3.112079935	0.001857742	0.033509162
Gene 2868	hypothetical protein	792.0158664	0.903798906	0.243354092	3.713925244	0.000204069	0.005128714
Gene 103	Phage tail length tape-measure protein	291.0838418	0.886220608	0.264748568	3.347404729	0.00081572	0.016722264
Gene 3905	Galactitol-1-phosphate 5-dehydrogenase (EC	352.7381482	0.884257258	0.260896737	3.389299805	0.000700714	0.014706642
Gene 1774	Probable Hsp20-family chaperone	170.4021998	0.883453718	0.297647986	2.968115897	0.002996313	0.04891204
Gene 88	Phage uncharacterized protein-like	649.1830115	0.883123973	0.236703181	3.730934113	0.000190771	0.004909919
Gene 3902	Transcriptional repressor of the fructose operon,	953.1147745	0.879780779	0.250661107	3.50984159	0.000448374	0.009943181
Gene 94	Scaffold protein	237.4446155	0.872616044	0.275567505	3.166614449	0.001542246	0.02877228
Gene 65	FIG00530366: hypothetical protein	297.6470962	0.85539807	0.256694648	3.33235647	0.000861139	0.017487902
Gene 2014	Transcriptional regulator, PadR family	1688.843347	0.843551722	0.225309295	3.743972134	0.000181134	0.004730945
Gene 86	hypothetical protein	247.401753	0.842078476	0.265105871	3.176385624	0.001491226	0.027968416
Gene 89	Portal protein, phage associated	647.4836624	0.806678189	0.234109011	3.445737456	0.000569503	0.01209 68

Gene 3899	Xylulose kinase (EC 2.7.1.17)	4276.453237	0.795160432	0.222026014	3.581384084	0.000341779	0.007980874
Gene 3885	ROK domain containing protein	2069.608341	0.768091487	0.220906168	3.477003357	0.000507051	0.010968488
Gene 4106	Putative ADP-ribosylglycohydrolase	3152.936593	0.760211481	0.249275229	3.049687218	0.002290798	0.039210456
Gene 2013	Cell division protein FtsW	9049.053618	0.741859949	0.229370563	3.23432937	0.001219288	0.023492939
Gene 1044	Two-component response regulator BceR	1398.285362	0.728523478	0.225804848	3.226341177	0.001253838	0.024026216
Gene 1046	ABC transporter ATP-binding protein YvcR	1848.170014	0.702521744	0.225554937	3.114636963	0.001841714	0.033473632
Gene 978	oxaloacetate decarboxylase	8550.783033	0.693943227	0.225291264	3.08020477	0.002068583	0.036469122
Gene 1045	sensor histidine kinase	2246.724204	0.691524228	0.220483018	3.136405854	0.001710323	0.031409376
Gene 1048	CAAX amino terminal protease family	1286.769469	0.684940863	0.222272324	3.081539126	0.002059334	0.036469122
Gene 3886	L-threonine 3-dehydrogenase (EC 1.1.1.103)	3218.379854	0.647793337	0.218701354	2.961999658	0.003056481	0.049594864