

**Understanding *Clostridium difficile*
Infection Outcomes, through Host
Clinical Variables, and Bacterial Whole
Genome and Phenotypic Analysis**

A thesis submitted by Emma Butt to the University of Exeter as a thesis for the
degree of Doctor of Philosophy in Biological Sciences

In February 2013

This thesis is available for Library use, on the understanding that it is copyright
material and that no quotation from the thesis may be published without proper
acknowledgement.

I certify that all material in this thesis which is not my own work has been
identified and that no material has previously been submitted and approved for
the award of a degree by this or any other University

Abstract

Clostridium difficile is a clinically problematic pathogen and continues to persist within the healthcare system. Presentation of disease symptoms ranges from mild to severe diarrhoea, through to fulminant pseudomembranous colitis. Approximately 20% of patients will suffer from recurrent episodes and of all patients who die from *C. difficile* related causes, approximately 41% of death certificates mention *C. difficile* as an underlying cause of death, and this poses a significant burden on healthcare facilities.

Three methods of investigation were employed to develop a more comprehensive understanding of both the host and isolate association with the outcomes of *C. difficile* infection; mortality and recurrence. These methods were; analysing patient clinical data to try and identify host markers of infection outcomes, evaluating *C. difficile* type association with infection outcomes, and genetically and phenotypically characterising the clinically relevant *C. difficile* isolates associated to these outcomes.

During this study statistical analysis of clinical data revealed that there were four variables; white cell count, serum albumin, C-reactive protein and respiratory rate, which were prognostic of mortality in patients with *C. difficile* infection. Threshold levels of these variables were used to create a clinical prediction rule to classify patients with *C. difficile* infection who were more 'at risk' from mortality, with statistical significance in both a derivation and validation cohort. However, analysis was unable to determine variables prognostic of recurrent infection.

Due to small sample sizes of some groups of isolates, no groups of *C. difficile* isolates were significantly associated with increased recurrent infection or mortality during this study. Some groups of isolates were associated with primary only infection and/or low mortality. There was a non-significant trend in particular *C. difficile* isolate groups being associated with infection outcomes; a panel of representative isolates was therefore chosen to be characterised in more detail.

Phenotypic and genetic analysis of a panel of sixteen *C. difficile* isolates revealed isolate specific differences in toxin production, conservation of transposable elements and SNP abundances, which may have played a role in infection outcomes. Isolate motility and antibiotic resistance profiles were not statistically significantly different between isolates within a particular group of *C. difficile* types.

One hypothesis from the collective results obtained during this study suggests that the phenotypic and genotypic changes in isolates may have facilitated differences in their interaction with the host. In turn, the host specific inflammatory response to the infecting *C. difficile* isolate may have played a role in host outcomes.

Research conducted during this study has begun to assess which host specific responses may be important in determining the outcome of *C. difficile* infection, and which *C. difficile* isolate characteristics may in part also contribute. However, the assessment of both host and isolate association to infection outcomes would benefit from further investigation in a larger cohort, in order to prove or refute conclusively any hypotheses generated in this study.

Acknowledgements

I would first and foremost like to dedicate this PhD thesis to my granddad, Alec (John) Ireland; he never got to see me finish the research on which this PhD thesis is based. As a great chemist, his enthusiasm for science was bestowed upon me, and without which, I am sure I would not have pursued a career in science with such passion as I now do.

You are always in my thoughts granddad.....

To all those who made this PhD possible I wish to sincerely thank you. I am sincerely grateful to my principle supervisor Dr Steve Michell, your continued support has been invaluable to me. It has allowed me to develop as scientist and as a result of this, have found a research area which I now wish to pursue to a higher level.

Thanks to my second supervisor Dr Ed Keedwell, who has provided much needed statistical support and whose patience was greatly appreciated.

To everyone on the 4th floor of Geoffrey Pope, I am extremely grateful for your help and guidance in the lab and for making a great working environment at the University of Exeter. I have thoroughly enjoyed working with you all.

Finally, I would like to thank my parents.... You let me move back home to do this PhD and for that I am eternally grateful. I have loved every minute of living back home with you! My fiancée Jamie..... You are the light when the moments are dark, you have been patient, kind and have accepted all that has come with this PhD. I love you so much and am very lucky to have you in my life.

Table of Contents

Content	Page Number
Title Page.....	i
Abstract.....	ii
Acknowledgments.....	iii
Table of contents.....	iv
List of figures and tables.....	xii
List of Manuscripts Submitted/Prepared as a Result of Research Conducted during this PhD.....	xxii
Abbreviations.....	xxiii
 Chapter One. General Introduction	
1.1 <i>Clostridium difficile</i>	1
1.2 History of <i>C. difficile</i> and Clinical Disease Presentation.....	1
1.3 Pathogenicity of <i>C. difficile</i>	3
1.4 TcdA and TcdB: Sequence Structure and Cell Binding Properties.....	4
1.5 Genetic Analysis of <i>C. difficile</i> to Understand Pathogenicity.....	6
1.6 Diagnosis of <i>C. difficile</i> Infection.....	6
1.7 <i>C. difficile</i> Strain Typing.....	8
1.8 Risk Factors for <i>C. difficile</i> Infection.....	9
1.9 Summary.....	11
1.10 Aim.....	11

PART ONE- Identifying Host Markers of *C. difficile* Infection Outcomes and Development of Clinical Prediction Rules.....13

Chapter Two. A Review of Clinical Risk Factors and Prediction Rules for Mortality and Recurrent *C. difficile* Infection.....14

2.1 Introduction.....	14
2.2 <i>C. difficile</i> and Mortality.....	15
2.3 <i>C. difficile</i> and Recurrent Infection.....	17
2.4 Clinical Prediction of <i>C. difficile</i> Infection Outcomes.....	18
2.5 Clinical Risk Factors and Prediction of Mortality.....	18
2.6 Clinical Risk Factors and Prediction of Recurrent <i>C. difficile</i> Infection.....	19
2.7 Faecal Calprotectin as a Biomarker for <i>C. difficile</i> Infection Outcomes.....	21
2.8 Study Aim.....	22

Chapter Three. Clinical Outcomes of *C. difficile* Infection- Mortality and Recurrence; Risk factors and Clinical Prediction Rules.....23

3.1 Introduction and Aim.....	23
3.2 Study Cohort.....	23
3.3 Patient Demographics.....	24
3.4 Mortality Outcome Statistics.....	25
3.5 Severity of Infection Outcome Statistics.....	26
3.6 Site of Infection Acquisition Statistics.....	26
3.7 Univariate analysis.....	28
3.7.1. Variables Independently Associated with All-Cause Mortality.....	29
3.7.1. i. Parametric tests.....	29
3.7.1. ii. Non- Parametric Tests.....	29
3.7. 2 Variables Independently Associated with Recurrence.....	30
3.8 Multivariate Analysis of Clinical Variables Independently Associated with All-cause Mortality.....	31
3.9 Model Derivation.....	32
3.10 Decision Tree Classification Results for All Cause Mortality.....	33

	Page Number
3.11 Prediction Rule for Mortality in Patients with <i>C. difficile</i> Infection.....	36
3.12 Prediction Rule Validation.....	38
3.13 Discussion.....	41
3.13.1 Discussion of Mortality Data Analysis.....	43
3.13.2 Prediction Rule Generation.....	41
3.13.3 Predicting Mortality in Patients with <i>C. difficile</i> Infection.....	43
3.13.4 Clinical Relevance of Variables.....	45
3.13.5 Discussion of Recurrent <i>C. difficile</i> Infection Data Analysis.....	48
3.14 Summary.....	50
Chapter Four. Investigating the use of Calprotectin as a Novel Host Marker of <i>C. difficile</i> Outcomes.....	51
4.1 Introduction, Aim and Method.....	51
4.2 Results.....	52
4.2.1 Patient Demographics.....	52
4.2.2 Investigating Calprotectin levels in Patients with Primary only and Recurrent <i>C. difficile</i> Infection	53
4.2.3 Investigating Calprotectin levels in Patients with Varying Severity of <i>C. difficile</i> Infection.....	54
4.2.4 Investigating Calprotectin Levels in Patients with <i>C. difficile</i> Infection who Died versus those who Survived.....	56
4.2.5 Results for Analysis of Stool Calprotectin Levels with Respect to Department of Health Scorecard Measures	57
4.2.6 Testing Calprotectin levels against stool storage duration.....	57
4.3 Discussion.....	58
4.3.1 Conclusion.....	60
4.4 Part One Summary.....	60

PART TWO- Assessing <i>C. difficile</i> Ribotype Epidemiology and Association with Outcomes of Infection.....	62
Chapter Five. A Review of <i>C.difficile</i> Epidemiology and Emerging <i>C.difficile</i> Ribotypes.....	63
5.1 Introduction.....	63
5.2 Global Epidemiology of <i>C. difficile</i>	63
5.3 Distribution of Ribotypes in the South West.....	66
5.4 Clinical Relevance of Emerging Ribotypes.....	67
5.5 Aim.....	69
Chapter Six. Cohort Ribotype Distribution and Association with the Infection Outcomes; Recurrence and Mortality.....	70
6.1 Introduction and Aim.....	70
6.2 Results	
6.2.1 Ribotype Distribution for RD&E Cohort.....	70
6.2.2 Distribution of Cases of Recurrent Infection within Ribotypes.....	75
6.2.2. i. Binomial Analysis Results.....	76
6.2.3 Distribution of Cases of Mortality within Ribotypes.....	78
6.2.3. i. Binomial Analysis Results.....	79
6.3 Discussion.....	81
6.3.1 Ribotype Distribution.....	81
6.3.2 Ribotype Association with Infection Outcomes.....	83
6.4 Conclusion.....	87
6.5 Summary.....	87
PART THREE- Phenotypic and Genetic Characterisation of <i>C. difficile</i> Isolates Associated with Outcomes of Infection.....	89
Chapter Seven. Molecular Basis of <i>C. difficile</i> Virulence.....	90
7.1 Introduction.....	90

	Page Number
7.2 Whole Genome Analysis to Identify Bacterial Virulence Determinants.....	90
7.3 Whole Genome Sequencing of <i>C.difficile</i>; Linking Phenotypic and Genotypic	
Data.....	91
7.4 Role of <i>C.difficile</i> Virulence Factors in Disease.....	92
7.4.1 <i>C. difficile</i> Virulence Characteristics.....	92
7.4.2 Toxin Mediated Virulence.....	93
7.5 Role of Toxins in Disease.....	94
7.6 Transcriptional Regulation and Translation of the Pathogenicity Locus; Role in Disease	
Outcomes.....	95
7.7 Non- toxin Mediated Virulence Factors; Antibiotic Resistance.....	98
7.8 Non- toxin Mediated Virulence Factors; Sporulation.....	100
7.9 Non- toxin Mediated virulence Factors; Bacterial Motility.....	101
7.10 Summary.....	102
7.11 Aim.....	102
Chapter Eight. Whole Genome Sequence Analysis of Clinically Relevant	
Isolates.....	104
8.1 Introduction.....	104
8.1.1 Isolate Selection.....	104
8.1.2 Aim.....	107
8.1.3 Method.....	108
8.2 SNP Detection for Ribotype 027 Isolates.....	109
8.3 Whole Genome Sequence Alignment of Ribotype 027 Isolates.....	111
8.4 SNP Detection for Ribotype 078 Isolates.....	113
8.4.1 SNP Detection in Ribotype 078 Sequential Isolates.....	113
8.4.2 SNP Detection in Ribotype 078 Isolates from Primary Only Infections.....	114
8.5 Whole Genome Sequence Alignment of Ribotype 078 isolates.....	115
8.6 SNP Detection for Ribotype 023 Isolates.....	118
8.6.1 SNP Detection in Ribotype 023 Sequential Isolates.....	118
8.6.2 SNP Detection in Ribotype 023 Isolates from Primary Only Infections.....	119
8.7 Whole Genome Sequence Alignment of Ribotype 023 Isolates.....	120

	Page Number
8.8 SNP Detection for Ribotype 015 Isolates.....	123
8.9 Whole Genome Sequence Alignment of Ribotype 015 Isolates.....	124
8.10 SNP Detection for Ribotype 014 Isolates.....	125
8.11 Whole Genome Alignment of Ribotype 014 Isolates.....	126
8.12 SNP Detection for Ribotype 013 Isolates.....	127
8.13 Whole Genome Alignment of Ribotype 013 Isolates.....	129
8.14 Sequence Analysis of Toxin Genes.....	129
8.15 Sequence Analysis of Antibiotic Resistance Genes.....	130
8.16 Discussion.....	134
8.16.1 General SNP Inferences.....	134
8.16.2 Inferences from Whole Genome Sequence Alignment.....	139
8.16.3 Inferences from Sequence Analysis of Toxin and Antibiotic Resistance Genes	143
8.16.4 Conclusion.....	146
8.17 Summary.....	148
 Chapter Nine. Phenotypic Characterisation of Outcome Isolates.....	 149
9.1. Introduction and Aim.....	149
9.2 Results.....	151
9.2.1 <i>C. difficile</i> growth.....	151
9.2.2 Characterisation of Isolate Swimming Motility.....	151
9.2.3 Characterisation of <i>in vitro</i> TcdA and TcdB Production.....	154
9.2.3. i Detection of TcdA.....	155
9.2.3. ii Detection of TcdB.....	159
9.2.4 Characterisation of Isolate Antibiotic Resistance Profiles.....	159
9.2.5 Characterisation of Isolate Spore Germination Profiles.....	164
9.3 Discussion.....	169
9.3.1 Bacterial Motility.....	169
9.3.2 <i>In vitro</i> detection of TcdA and TcdB.....	170
9.3.3 Antibiotics MIC profiles.....	173
9.3.4 Sporulation Characteristics.....	176
9.4 Conclusion.....	178

	Page Number
9.5 Concluding Summary and Discussion	180
9.5.1 Summary	180
9.5.1. i Variables Prognostic of <i>C. difficile</i> Infection Outcomes	180
9.5.1. ii <i>In vitro</i> Isolate Virulence Characteristics	181
9.5.1. iii Whole Genome Sequence Analysis	185
9.5.2 Conclusion	186
9.6 Future work	182
9.6.1 Expanding the Panel of Isolates Chosen for phenotypic and Genetic Analysis	188
9.6.2 Further Characterisation of Isolates Studied During this Research	188
9.6.1. i Toxin Gene Expression, Protein Purification and Cytotoxicity Assays	188
9.6.1. ii Sporulation Studies	189
9.6.1. iii Antibiotic Resistance to Levofloxacin, Moxifloxacin and Gatifloxacin	189
9.6.1. iv Whole Genome Sequence Analysis	189
9.6.1. v Adhesion Assays	189
9.6.1. vi Transcriptomics	189
9.6.1. vii <i>In vivo</i> Infection of Mice and Hamsters	189
Chapter Ten. Materials and Methods	190
10.1 Culture Media	190
10.2 Buffers	190
10.3 Bacterial Isolates	191
10.4 Statistical Analysis Methods	192
10.4.1 Study Cohort	192
10.4.2 Univariate analysis	195
10.4.3 Multivariate Analysis	195
10.4.4 Classification Models	195

	Page Number
10.4.5 Binomial Analysis.....	196
10.5 Culturing <i>Clostridium difficile</i> from Frozen Faecal Samples.....	196
10.6 <i>Clostridium difficile</i> Growth Curves.....	196
10.7 Ribotyping.....	197
10.8 Calprotectin ELISA.....	198
10.9 Antibiotic Minimum Inhibitory Concentration (MIC) Assay.....	199
10.10 Motility Assay.....	199
10.11 Sporulation Assay.....	199
10.12 <i>In vitro</i> Toxin A and B Immunodot Blots.....	200
10.12.1 Bacterial Growth.....	200
10.12.2 Immunoblot.....	201
10.13 Genomic DNA extraction.....	202
10.14 Whole Genome Sequencing.....	203
References	205
Appendices	220
Appendix One. Clinical Data Associated with Characterised Isolates.....	220
Appendix Two (compact disc). Command line syntax, Perl scripts, Bowtie assembly statistics, SNP data and velvet assembly data for unaligned sequence reads	
Appendix Three. <i>C. difficile</i> Growth Curve and Motility Assay Results.....	223

Figures

Chapter One

Figure 1.1- Genetic map of the 19.6 Kbp PaLoc of *C. difficile*, showing the deletion of the PaLoc in non-toxigenic strains.....**4**

Figure 1.2- Representative diagram of the structure of LCTs.....**5**

Chapter Two

Figure 2.1- Number of death certificates mentioning *C. difficile* by underlying cause and other mentions, from 2006-2010 in England and Wales. Source: Office for National Statistics licensed under the Open Government Licence v.1.0.205.....**16**

Chapter Three

Figure 3.1- Normalised importance of the variables in the mortality outcome model.....**34**

Figure 3.2- CRT diagram for test data for the outcome all-cause mortality.....**35**

Figure 3.3- Summary of the prediction rule developed during this research.....**36**

Figure 3.4 - ROC curve for the prediction rule, to classify those at risk of mortality, as tested in the derivation cohort.....**38**

Figure 3.5- ROC curve for the prediction rule (excluding respiratory rate), to classify those at risk of mortality, as tested in the derivation cohort.....**40**

Figure 3.6 - ROC curve for the prediction rule (excluding respiratory rate), to classify those at risk of mortality, as tested in the validation cohort.....**40**

Chapter Four

Figure 4.1- Levels of calprotectin within different infection outcome groups. Bars represent median calprotectin levels.....**54**

Figure 4.2 - Levels of calprotectin within different severity groups. Bars represent median calprotectin levels.....**55**

Figure 4.3- Levels of calprotectin within different mortality outcome groups. Bars represent median calprotectin levels.....**56**

Chapter Five

Figure 5.1- Geographic distribution of *Clostridium difficile* PCR ribotypes in European countries with more than five typable isolates, as of November 2008. Pie charts show proportion of most frequent PCR ribotypes per country. The number in the centre of the pie chart is the number of typed isolates in this country²⁸**64**

Figure 5.2 - Distribution of *C. difficile* ribotypes in England from 2007-2010⁶.....**65**

Figure 5.3 - Distribution of *C. difficile* ribotypes in the South West of England, from 2007-2010⁶.....**66**

Chapter Six

Figure 6.1- Representative gel image of the ribotype patterns that were produced for all isolates in the study cohort. 1) SupperLadder Low (100 bp), 2) Ribotype pattern of *C. difficile* 630Δerm (012), 3) Ribotype pattern of *C. difficile* R20291 (027).....**71**

Figure 6.2- Histogram of ribotype distribution within CDI cases from the Royal Devon and Exeter Hospital from 2007-2009.....**72**

Figure 6.3- Relative frequency distribution of *C. difficile* ribotypes in the Royal Devon and Exeter Hospital from 2007-2009. Fifteen ribotypes made the remaining relative frequency data and were present ≤ 4% over the study period.....**73**

Figure 6.4– Distribution of percent recurrent and primary CDI cases for specific ribotypes. N= number of patients. Some ribotypes are missing as they were present at ≤ 3 times in the population.....**76**

Figure 6.5- Distribution of percent mortality for specific ribotypes. N= number of cases. Some ribotypes are missing as they were present at ≤ 3 times in the population. Grey hatched boxes represent survival, red hatched boxes represent CDI related mortality and green hatched boxed represent non-CDI related mortality.....**78**

Figure 6.6- Percent residual mortality rates for clinical ribotypes⁶84

Chapter Seven

Figure 7.1- Representation of the distribution, by functional categories, of the shared and unique *C. difficile* CDSs relative to the sequenced clostridial genomes. Blue: *C. difficile* CDSs, purple: *C. difficile* CDSs shared with the sequenced clostridia, yellow: CDSs unique to *C. difficile* relative to the sequenced clostridia¹⁸⁶92

Figure 7.2- Genetic map of the 19.6 Kbp PaLoc of *C. difficile*. Direction of the block arrows indicates the direction of transcriptional regulation of the toxin genes during early log stages of growth. The small arrows above TcdA and TcdB represent independent transcription from their own promoters, during late log and stationary growth phases.....95

Chapter Eight

Figure 8.1- A small region of whole genome alignment of two *C. difficile* genomes from a patient with recurrent CDI. Green box= magnified region of genetic variability; no sequence coverage 16 Kbp region in clinical isolates.....111

Figure 8.2- Part of whole genome alignment of 078 *C. difficile* genomes. Red hatched box represents the 106 Kbp region in M120 that is not mapped in three clinical genomes (However some genes from region are found in the unaligned contig files). The green box represents the 46 Kbp region contained only in FA08006656 and the reference M120 genome. The orange boxes represent a 37 Kbp region (annotated as CTn2 in *C. difficile* 630) found in both *C. difficile* 630 and M120.....116

Figure 8.3- A small region of the whole genome alignment of 078 *C. difficile* genomes depicting the loss of the F3 flagellar operon (highlighted in red) from the 078 isolates as compared to the reference isolates. Solid red box indicates the transposase gene which has replaced the F3 operon in 078 genomes.....117

Figure 8.4- Mauve alignment view of all four 023 isolate genomes in comparison to the 630 and R20291 reference genomes. The region highlighted in the red represents the probable insertion of a 140 Kbp region comprised of phage and hypothetical protein related genes.....**121**

Figure 8.5- Magnified region of the mauve alignment view of all four 023 isolate genomes in comparison to the 630 and R20291 reference genomes. Red hatched lines represent the probable insertion of a 6 Kbp region between *tcdA* and *tcdB* of 023 isolates.....**122**

Figure 8.6- Mauve alignment view of ribotype 015 isolate genomes in comparison to the 630 and R20291 reference genomes. Red hatched lines represent the probable insertion of a 150 Kbp region comprised of phage and hypothetical protein related genes.....**125**

Figure 8.7- Mauve alignment view of ribotype 014 isolate genomes in comparison to the 630 and R20291 reference genomes. Red hatched lines represent the probable insertion of a 190 Kbp region comprised of phage and hypothetical protein related genes.....**127**

Chapter Nine

Figure 9.1- Graphical representation of mean *C. difficile* isolate motility. Motility was scored as <10 mm (non-motile), 10-20 mm (motile) and >20 mm (highly motile). Green bars= 078 ribotypes, Red bars= 023 ribotypes, Yellow bars= 027 ribotypes, Blue bars= 015 ribotypes, Orange bars= 014 Ribotypes, Beige bars= 013 ribotypes.**153**

Figure 9.2- Immuno-dot blot for *in vitro* TcdA production *C. difficile* isolates. Culture supernatant from the toxin negative and CF5 strain was blotted with all isolates. No toxin was detected with the mAb specific for TcdA. Ref= Reference isolate. T= Growth time in hrs..**157**

Figure 9.3- Immuno-dot blot for *in vitro* TcdA production for *C. difficile* isolates. Red boxes represent toxin dots from recurrent infection (same patient). Culture supernatant from the toxin negative and CF5 isolates was blotted with all isolates. No toxin was detected with the mAb specific for TcdA. Ref= Reference isolate. T= Growth time in hrs.....**158**

Figure 9.4- Immuno-dot blot for TcdB for *C. difficile* isolates. Culture supernatant from the toxin negative strain was blotted with all isolates and no toxin was detected. Ref= Reference isolate. T= Growth time in hrs.....**159**

Figure 9.5a- Resistance of *C. difficile* isolates to ciprofloxacin. Values of 32 on the scale represent a value of > 32 µg/ml on the E-test strip. Error bars= SEM. MIC for control strain (*S. aureus*) for ciprofloxacin falls between 0.125–0.5 µg/ml according to manufacturer specification. Green bars= 078 Ribotypes, Red bars= 023 Ribotypes, Yellow bars= 027 ribotypes, Blue bars= 015 ribotypes, Orange bars= 014 Ribotypes, Beige bars= 013 ribotypes.**163**

Figure 9.5b- Resistance of *C. difficile* isolates to metronidazole. Error bars= SEM. MIC for control strain (*B. fragilis*) for metronidazole falls between 0.25–1 µg/ml according to manufacturer specification. Green bars= 078 Ribotypes, Red bars= 023 Ribotypes, Yellow bars= 027 ribotypes, Blue bars= 015 ribotypes, Orange bars= 014 Ribotypes, Beige bars= 013 ribotypes.....**163**

Figure 9.5c- Resistance of *C. difficile* isolates to vancomycin. Error bars= SEM. MIC for control strain (*S. aureus*) for vancomycin falls between 0.5–2 µg/ml, according to manufacturers specification. Green bars= 078 Ribotypes, Red bars= 023 Ribotypes, Yellow bars= 027 ribotypes, Blue bars= 015 ribotypes, Orange bars= 014 Ribotypes, Beige bars= 013 ribotypes.....**164**

Figure 9.6a- Heat resistant cfu at 24 hrs for clinical and reference isolates under investigation. Error bars= SEM of a minimum of two independent experiments. Dashed line represents mean heat resistant cfu for this time point. Red bars= 023 ribotypes, green bars= 078 ribotypes, yellow bars= 027 ribotypes, blue bars= 015 ribotypes and orange bars= 014 ribotypes.....**166**

Figure 9.6b- Heat resistant cfu at 48 hrs for clinical and reference isolates under investigation. Error bars= SEM of a minimum of two independent experiments. Dashed line represents mean heat resistant cfu for this time point. Red bars= 023 ribotypes, green bars= 078 ribotypes, yellow bars= 027 ribotypes, blue bars= 015 ribotypes, orange bars= 014 ribotypes and beige bars= 013 ribotypes.167

Figure 9.6c- Heat resistant cfu at 120 hrs for clinical and reference isolates under investigation. Error bars= SEM of a minimum of two independent experiments. Red bars= 023 ribotypes, green bars= 078 ribotypes, yellow bars= 027 ribotypes, blue bars= 015 ribotypes, orange bars= 014 ribotypes and beige bars= 013 ribotypes. #= Isolate which was significantly different from most other clinical isolates.....168

Tables	Page Number
Chapter One	
Table 1.1- List of diagnostic testing methods for CDI ²⁶	7
Table 1.2– Summary of <i>C. difficile</i> typing techniques ⁵⁵	9
Chapter Two	
Table 2.1- Number of deaths in England where <i>C. difficile</i> was mentioned on the death certificate as the underlying cause of death, 2007-2010.....	16
Chapter Three	
Table 3.1- Baseline clinical characteristics of the RD&E cohort.....	27
Table 3.2- Statistically significant predictor variables left in the multinomial logistic regression equation after block entry.....	31
Table 3.3- Parameter estimates for the all-cause outcome measure; CDI related mortality and non-CDI related mortality.....	32
Table 3.4- Classification table for multinomial regression model for all-cause mortality outcome.....	32
Table 3.5- Classification of test data by the CRT model.....	34
Table 3.6a- Cross tabulation of score for risk of mortality against the actual mortality outcome.....	37
Table 3.6b- Summary of mortality risk with increasing prediction rule score.....	37
Table 3.7- Prediction rule scoring system (excluding respiratory rate) tested on derivation cohort and validation cohort versus the prediction rule derived by Bhangu <i>et al.</i> ³⁰	39
Table 3.8- Techniques for the development of clinical prediction rules with corresponding advantages and disadvantages.....	42
Chapter Four	
Table 4.1- Laboratory parameters recorded.....	52
Table 4.2- Patient demographics for cases available for statistical analysis.....	53
Table 4.3- Descriptive statistics for variables.....	53

Table 4.4- Number of cases and median stool calprotectin levels for groups of patients in the severity outcome.....**55**

Table 4.5- Approximate duration of storage of stool samples.....**57**

Table 4.6- Median calprotectin levels for storage periods.....**58**

Chapter Five

Table 5.1- Changing prevalence of most common (top 10) *C. difficile* ribotypes detected by CDRN in 2007/08, 2008/09 and 2009/10⁶.....**67**

Chapter Six

Table 6.1- Yearly change in prevalence of ribotypes in the RD&E cohort as compared to the national change in prevalence reported by the 2009/2010 CDRN report⁶.....**74**

Table 6.2- Ribotypes significantly associated with one episode of CDI.....**77**

Table 6.3- Ribotypes significantly associated with lower mortality.....**70**

Table 6.4- Clinical isolates selected for whole genome and phenotypic analysis.....**88**

Chapter Seven

Table 7.1- Positions and genes of the flagellar operons in the *C. difficile* 630 genome.....**101**

Chapter Eight

Table 8.1- Clinical isolates selected for whole genome sequence analysis.....**108**

Table 8.2- Isolates selected for use as reference genomes during whole genome sequence analysis.....**108**

Table 8.3- Non-synonymous changes found within two 027 *C. difficile* isolates from one patient with recurrent CDI.....**110**

Table 8.4- Summary of SNP site differences in four 078 isolates.....**112**

Table 8.5- Non-synonymous changes found within all 078 *C. difficile* isolates.....**113**

Table 8.6- Non-synonymous changes found within two 078 *C. difficile* isolates from one patient with recurrent CDI.....**114**

Table 8.7- Non-synonymous changes found within FA09004991.....**115**

Table 8.8- Summary of SNP site differences in four 023 isolates.....**118**

Table 8.9 - Non-synonymous changes found within FA08005864.....	119
Table 8.10 - Non-synonymous changes found within FA08006661.....	120
Table 8.11 - Summary of SNP site differences in the 015 isolates.....	123
Table 8.12 - Non-synonymous changes found within both 015 <i>C. difficile</i> isolates.....	124
Table 8.13 - Summary of SNP site differences in 014 isolates.....	126
Table 8.14 - Summary of SNP site differences in 013 isolates.....	128
Table 8.15 - Non-synonymous SNPs in putative sporulation related genes.....	128
Table 8.16 - Presence/absence of toxin genes and <i>tcdC</i> genotypes that were found in clinical isolates.....	130
Table 8.17 - Amino acid substitutions in GyrA and GyrB and presence of putative genes that could be involved in antibiotic resistance.....	133
Table 8.18 - Number of NS SNPs in isolates from different mortality outcomes.....	137
Table 8.19 - Conserved putative vancomycin/teicoplanin resistance genes and presence/absence of putative SNPs within them.....	144

Chapter Nine

Table 9.1 - Clinical isolates selected for analysis.....	150
Table 9.2 - Reference isolates selected as controls.....	150
Table 9.3 - Level of detection of TcdA in all clinical isolates after 8 hrs of growth in TY and BHI broth.....	156
Table 9.4 - Mean antibiotic MICs of all clinical cohort isolates.....	161
Table 9.5 - Range of MIC values from study isolates with the resistance breakpoints used. (Adapted from ^{34, 62, 166} and personal communication with Dr Val Hall).....	162
Table 9.6 - TcdC Genotypes and patient outcomes of Isolates that produced TcdA at 8 hours of growth in TY broth.....	171

Concluding Discussion

Table 9.7 - Overview of phenotypic and genetic characteristics of cohort clinical isolates.....	184
Table 9.8 - Overview of Antibiotic MICs of cohort clinical isolates.....	185

Chapter Ten	Page Number
Table 10.1- Culture Media.....	190
Table 10.2- Buffers.....	190
Table 10.3- Antibiotic susceptible positive control strains.....	191
Table 10.4- List of reference <i>C. difficile</i> isolates used in this study.....	192

Appendices

Table A1.1- Infection outcome pattern for each clinical isolate.....	221
Figure A1.2- Patient demographics for clinical isolate panel.....	222
Figure A3.1- Growth curve of <i>C. difficile</i> ribotypes a) 027 and b) 015 in BHIS broth over 48 hrs. Black line= OD ₅₉₀ , Red line= cfu/ml. Error Bars= SEM; OD ₅₉₀ curve-SEM, N=4, cfu/ml curve-SEM, N=2.....	223
Figure A3.2- Growth curve of <i>C. difficile</i> ribotypes a) 014 and b) 013 in BHIS broth over 48 hrs. Black line= OD ₅₉₀ , Red line= cfu/ml. Error Bars= SEM; OD ₅₉₀ curve-SEM, N=4, cfu/ml curve-SEM, N=2.....	224
Figure A3.3- Growth curve of <i>C. difficile</i> ribotypes a) 023 and b) 078 in BHIS broth over 48 hrs. Black line= OD ₅₉₀ , Red line= cfu/ml. Error Bars= SEM; OD ₅₉₀ curve-SEM, N=4, cfu/ml curve-SEM, N=2.....	225
Figure A3.4- Representative motility of reference isolates. Bars indicate points from which motility was measured. A representative isolate was not available for ribotype 013.	
Figure A3.5- Representative motility of 078 and of 023 isolates. Bars indicate points from which motility was measured.....	226
Figure A3.6- Representative motility of 027,015 013 and 014 isolates. Bars indicate points from which motility was measured.....	226

List of Manuscripts Submitted/Prepared as a Result of Research Conducted during this PhD

Submitted Manuscripts

Derivation and Validation of a Simple, Accurate and Robust Prediction Rule for Risk of Mortality in Patients with *Clostridium difficile* Infection. Emma Butt¹, Jane AH Foster², Edward Keedwell³, Julia EA Bell⁴, Richard W Titball¹, Aneel Bhangu⁵, Stephen L Michell¹ and Ray Sheridan⁶. *BMC Infectious diseases*. *Under review*

***Clostridium difficile*: Evaluation of disease severity and the potential role of Faecal Calprotectin as a biomarker. Have the UK guidelines got it right?** Dr Jane AH Foster¹, Emma Butt², Dr Edward Keedwell³, Dr Stephen L Michell², Dr Tariq Ahmed¹, Dr Ray Sheridan¹. *In preparation*.

Characterisation of Clinically Relevant *Clostridium difficile* Isolates from different Infection Outcomes. ¹Emma Butt, ²Ed Keedwell, and ¹Stephen .L. Michell. *Journal of Medical Microbiology*. *In preparation*.

Abbreviations

ADP	Adenine tri phosphate
H ₁	Alternative hypothesis
AUC	Area under the curve
bp	Base pairs
bpm	Beats per minute
BHI	Brain heart infusion
mCL	Cells/micro litre
χ^2	Chi-Square value
CDI	<i>Clostridium difficile</i> infection
CDS	Coding Sequences
Cfu/s	Colony forming units
<i>df</i>	Degrees of freedom
DNA	Deoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
g	Grams
x g	Gravitational force
Gal β 1-4GlcNAc	N-Acetyllactosamine
GDH	Glutamate Dehydrogenase
GDP	Guanosine 5'-diphosphate
GTP	Guanosine 5'-Triphosphate
HPA	Health Protection Agency
HCAI	Healthcare acquired infection
hr/hrs	hour/hours
IL	Interleukin
Kbp	Kilo base pairs
kg	Kilo gram
L	Litres
MAP	Mitogen Activated Protein
μ l	Micro litres
μ M	Micro molar
mg	Milli grams
ml	Milli litres
mm	Milli metre
mM	Milli molar
MIC	Minimum inhibitory concentration
Mins	Minutes
MLST	Multi locus sequence typing
mAb/s	Monoclonal Antibody/ies
Nf- κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
ng	Nano grams
nm	Nano metre

NGS	Next generation sequencing
NS	Non-synonymous
H ₀	Null hypothesis
PaLoc	Pathogenicity locus
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
ROC	Receiver operating curves
REA	Restriction enzyme analysis
Rpm	Revolutions per minute
SEM	Standard error of the mean
S	Synonymous
Tcdx	<i>Clostridium difficile</i> Toxin (x)= A, B, C, R, or E
Tn	Transposon
TE	Tris- EDTA
TAE	Tris-acetate EDTA
TBS	Tris-buffered saline
TBST	Tris-buffered saline
TY	Tryptone yeast
TNF	Tumour necrosis factor
Yrs	Years