Evolution of *Clostridium difficile*.

Michelle Dawn Cairns
Student number: 1047773

A thesis presented to the University College London
for the degree of

Doctor of Philosophy

University College London
Division of Infection and Immunity

Supervised by

Professor Timothy McHugh
University College London
&
Professor Brendan Wren
London School of Hygiene and Tropical Medicine

2018

Declaration

I, Michelle Dawn Cairns confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

None of the material presented herein has been submitted previously for the purpose of obtaining another degree. This thesis does not exceed 100,000 words as required by University College London.

Signed:

Abstract

Clostridium difficile continues to be a leading cause of healthcare-associated infections in the developed world. Increased detection of *C. difficile* infection (CDI) and development of typing schemes to differentiate between strains is primarily due to the recognition of global outbreaks of a single strain, BI/NAP1/027 which is characterised by three common typing techniques; restriction endonuclease analysis (REA), pulsed-field gel electrophoresis (PFGE) and PCR ribotyping.

Phylogenetic analysis using multilocus sequence typing (MLST) divides *C. difficile* into five phylogenetic lineages which align the well-known PCR ribotypes; 027, 023, 017, 078 and a lineage containing diverse PCR ribotypes. MLST data in this thesis confirmed the five phylogenetic lineages were maintained after testing a larger collection of isolates from varied sources with further micro-diversity within the individual lineages. MLST investigation did not identify a lineage exclusive to non-human strains or any correlation between sequence type and geographical location. Data in this thesis also supports the notion that PCR ribotyping and REA do not correspond as well as previously considered. This may result in phylogenetically similar strains being designated as a different type or variant.

The toxin A-B+ PCR ribotype 017 strain that forms a predominant lineage is little investigated. Through whole genome sequencing (WGS) and single nucleotide polymorphism (SNP) analysis, a historical clone of PCR ribotype 017 was identified from a London hospital ward. Although no phenotype exclusive to the clonal strain

was characterised, this is the first report in the UK investigating the phylohistory of isolates from hospitalised patients with CDI due to PCR ribotype 017.

Further investigation of PCR ribotype 017 with a larger and global collection of strains revealed two distinct sub-lineages containing multiple independent clonal expansions, antimicrobial resistant SNP determinants, deletions and insertions which were well distributed geographically and temporally. The data suggests transmission between humans and animals and findings support a USA origin with multiple, global transmission events.

The key findings of this thesis are that *C. difficile* as a species is continually evolving with the appearance of divergent sub-lineages. WGS is superior to routine typing methodologies for tracking this evolution and will have significant impacts for outbreak investigation, understanding the phylohistory and phylogeography of *C. difficile* and other pathogens that are a threat to human health.

Impact Statement

The Gram positive bacterium *Clostridium difficile* causes gastrointestinal disease in humans and animals and continues to be a leading cause of healthcare-associated infections in the developed world. Prior to 2003, there were few reports of *C. difficile* infection (CDI) yet in 2003 multiple outbreaks of CDI were reported in Europe, United States of America (USA) and Canada affecting humans. Typing schemes were introduced to differentiate between strains for investigation of outbreaks; restriction endonuclease analysis (REA), pulsed-field gel electrophoresis (PFGE) and PCR ribotyping and a clonal strain of *C. difficile* was identified (BI/NAP1/027). Multilocus sequence typing (MLST) is a different typing technique useful for studying the population and evolutionary genetics of bacteria by exploiting the nucleotide sequences of housekeeping gene fragments. By testing a larger collection of isolates from varied sources using MLST, this thesis confirms the findings of previous studies; *C. difficile* is made up of five phylogenetic lineages of which four align with the well-known PCR ribotypes; 027, 023, 017, 078 and a fifth lineage containing diverse PCR ribotypes. Further micro-diversity within the individual lineages was also revealed suggesting continued evolution of this species. MLST investigation did not identify a lineage exclusive to non-human strains or any correlation between sequence type and geographical location suggesting transmission of strains between humans and animals and lack of global spread of a clonal strain according to sequence type. Furthermore, data in this thesis shows that PCR ribotyping and REA do not correspond as well as previously considered. This inconsistency may result in phylogenetically similar strains being designated as a different type or variant which has significant implications to infection control and

management of outbreaks. This thesis further investigated toxin A-B+ PCR ribotype 017 strains in greater detail. Through whole genome sequencing (WGS) and single nucleotide polymorphism (SNP) analysis, a historical clone of PCR ribotype 017 was identified from a London hospital ward. Although no phenotype exclusive to the clonal strain was characterised, this is the first report in the UK investigating the phylohistory of isolates from hospitalised patients with CDI due to the PCR ribotype 017. This contributes to our understanding of an interesting strain which is toxin A negative and reported to be prevalent in Asia. Further investigation of PCR ribotype 017 by testing a larger and global collection of strains using WGS revealed two distinct sub-lineages containing multiple independent clonal expansions, antimicrobial resistant SNP determinants, deletions and insertions which were well distributed geographically and temporally. The data suggests transmission between humans and animals and findings support a USA origin with multiple, global transmission events.

The three studies described above are published as papers (available in the public domain) and either presented as posters or presented as an oral presentation at an international conference. The findings are significant to academia worldwide whereby the continual monitoring of the expansion of *C. difficile* improves our understanding of the evolution of this species. This in turn has implications to the non-academic world where the data is translational to healthcare and its patients; this is imperative. WGS allows us to predict and/or detect a more virulent and/or clonal strain early enabling improved diagnostics and prompt outbreak interventions. This can only result in improvements to patient morbidity, mortality and associated costs.

Acknowledgements

Firstly, I would like to thank **Professor Brendan Wren, Professor Stephen Gillespie, Dr Clare Ling** and **Dr Nandini Shetty** for support in seeking funding for a PhD, obtaining the PhD position and facilitating this to work alongside my role as a Clinical Scientist. I thank the **National Institute for Health Research (NIHR)** for the award of a *NIHR & Chief Scientific Officer Healthcare Scientist Doctoral Research Fellowship* and I thank **contributors** who have shared study isolates (Page 27). Without the above, this PhD would not have been possible.

Professor Timothy McHugh I thank you for all-encompassing support and wisdom with combining life and a PhD; the years have included some trying times but you have *always* been there as both a supervisor and friend.

Professor Brendan Wren I thank you for your unremitting encouragement throughout my PhD, this you provide tirelessly to all of your staff. You continually exude an exciting and enthusing scientific research vision, this I feel privileged to have experienced.

I have progressed throughout my PhD with the support and experience of **Dr Richard Stabler** and **Dr Esmeralda Valiente**. Richard, I thank you for help with data analysis and patience with my many versions of 'custom Cairns scripts'. Esmeralda, your kindness and encouraging nature have been much appreciated.

I thank **Dr Mark Preston** for your support with SNP data analysis and members of the **LSHTM 'Wren-Crew',** past and present for chatter and assistance. It has been a pleasure working with you all; I will always have fond memories of my time at the LSHTM. I thank scientific colleagues at **Public Health Laboratory (PHL), London** for your understanding with my divided time as a Clinical Scientist and PhD student. I especially thank **Catherine Ryan**; you have been a wonderful sidekick whilst we both worked part-time supporting the PHL London services.

I thank my '**London Clinical Scientist troop'** for everlasting friendship and support. I thank my **family and friends** for constant love and praise. Despite not quite understanding what my PhD was about, you have kept me on the road to completion.

I finally thank **Rebecca Gorton**. We have supported each other from beginning to end each having ups and downs with lab work, thesis writing and life in general. A Christmas or birthday card will no longer include reference to '*we will get through it*' or '*next year it will be over*'. We can now enjoy life. **We made it.**

In loving memory of Nana
Mrs Vivien Cairns

"Keep your pecker up".

Contents

List of Figures

List of Tables

List of communications arising from this thesis

Peer-reviewed publications (shown at the back of this thesis)

- Stabler, R. A., Dawson, L. F., Valiente, E., **Cairns, M. D.,** Martin, M. J., Donahue, E. H., Riley, T. V., Songer, J. G., Kuijper, E. J., Dingle, K. E., and Wren, B. W. (2012). *Macro and Micro Diversity of Clostridium difficile Isolates from Diverse Sources and Geographical Locations.* PLoS One, 7 (3).

- Valiente, E., Dawson, L. F., **Cairns, M. D.,** Stabler R. A., and Wren B. W. (2012). *Emergence of new PCR ribotypes from the hypervirulent Clostridium difficile 027 lineage.* Journal of Medical Microbiology. 61, 49-56.

- **Cairns, M. D.,** Stabler, R. A., Shetty, N., and Wren, B. W. (2012). *The continually evolving Clostridium difficile species.* Future Microbiology 7(8), 1-13.

- **Cairns, M. D.,** Preston, M. D., Lawley, T. D., Clark, T. G., Stabler, R. A., and Wren, B.W. (2015). *Genomic epidemiology of a protracted hospital outbreak caused by a toxin A negative Clostridium difficile sublineage PCR Ribotype 017 strain in London, England.* Journal of Clinical Microbiology. Vol: 53, Issue 10.

- **Cairns, M. D.,** Preston, M. D., Hall, C. L., Gerding, D. N., Hawkey, P. M., Kato, H., Kim, H., Kuijper, E. J., Lawley, T. D., Pituch, H., Reid, S., Kullin, B., Riley, T. V., Solomon, K., Tsai, P. J., Weese, J. S., Stabler, R. A. and Wren, B. W. (2017). *Comparative genome analysis and global phylogeny of the toxin variant Clostridium difficile PCR Ribotype 017 reveals the evolution of two independent sub-lineages.* Journal of Clinical Microbiology. Vol: 55, Issue 3.

- Shaw, H., Preston, M. D., Vendrik, K., **Cairns, M. D.,** Browne, H., Stabler, R,. A., Crobach, M., Cover, J., Pituch, H., Ingebretsen, I., Pirmohamed, M.,

Valiente, E., Lawley, T., Fairweather, N., Kuijper., E. and Wren, B. W. (2018). *Analysis of Clostridium difficile Ribotype 023 strains reveals the recent emergence of a highly related virulent clade with unique characteristics.* Nature Microbiology. Manuscript submitted.

Poster presentations (shown at the back of this thesis)

- **Cairns, M. D.,** Stabler, R. A., and Wren, B. W. (2013). *The global emergence, evolution and transmission of the Clostridium difficile toxin A-B+ PCR Ribotype 017.* Public Health England Annual Conference, Warwick, UK.

- **Cairns, M. D.,** Valiente, E., and Wren, B. W. (2016). *Phenotypic characterisation of toxin A negative, Clostridium difficile PCR Ribotype 017 strains of different sub-lineages in London, England.* European Congress of Clinical Microbiology and Infectious Diseases. Amsterdam, The Netherland's.

Oral presentations

- **Cairns. M. D.** (2015). *The phylogeny of Clostridium difficile toxin A-B+ PCR Ribotype 017.* International *Clostridium difficile* Symposium. Lake Bled. Slovenia.

Abbreviations

ABACAS	Algorithm based automatic contiguation of assembled sequences
ACT	Artemis comparison tool
AFLP	Amplified fragment length polymorphism
ANOVA	Analysis of variance
BEAST	Bayesian evolutionary analysis sampling tree
BHIS	Brain heart infusion
BLAST	Basic local alignment search tool
bp	base pair
BWA	Burrows-wheeler aligner
CDI	*Clostridium difficile* infection
CDRN	*Clostridium difficile* ribotyping network
cfu	Colony forming unit
°C	degree Celcius
DNA	Deoxyribonucleic acid
FQR	Fluoroquinolone resistance
gDNA	Genomic DNA
HPV	Hydrogen peroxide vapour
IMS	Industrialised methylated spirit
LSHTM	London School of Hygiene and Tropical Medicine
MAF	Minor allele frequency
MAFFT	Multiple alignment using fast fourier transform
MIC	Minimum inhibitory concentration
MLST	Multilocus sequence typing
MLVA	Multilocus variable-number tandem-repeat analysis
NT	Nontypeable
NAP1	USA Pulseotype 1
OD	Optical density
PaLoc	Pathogenicity locus
PBS	Phosphate buffer saline
PBS-T	PBS plus 0.5% tween solution
PCR	Polymerase chain reaction

PFGE	Pulsed-field gel electrophoresis
PLG	Phase lock gel light
PMC	Pseudomembraneous colitis
RaXML	Randomized axelerated maximum likelihood
RCM	Robertson’s cooked meat medium
REA	Restriction endonuclease analysis
RNA	Ribonucleic acid
S-layer	Surface layer
SAM	Sequence Alignment Map
SLP	Surface layer protein
SNP	Single nucleotide polymorphism
TcdA	Toxin A
TcdB	Toxin B
A+B+	Toxin A positive toxin B positive
A-B+	Toxin A negative toxin B positive
UHL	University Hospital Lewisham
UK	United Kingdom
UPGMA	Unweighted pair group method with arithmetic mean
USA	United States of America
WGS	Whole genome sequencing
WTSI	Wellcome Trust Sanger Institute

Collaborators

The following collaborators are kindly thanked and acknowledged for provision of isolates:

Brazier, J.: Anaerobe reference unit, Public Health Wales Microbiology, University Hospital of Wales, Heath Park, Cardiff, CF14 4XW.

***Clostridium difficile* Ribotyping Network:** Public Health Laboratory London, Division of Infection, The Royal London Hospital, London, E1 2ES. UK.

Dingle, K. E.: Nuffield Department of Clinical Medicine, Oxford University, John Radcliffe Hospital, United Kingdom.

Drudy, D.: Dublin Institute of Technology, Aungier Street, Dublin 2, Ireland.

Gerding, D. N.: Edward Hines Jr. Veterans Affairs Hospital, Hines, Illinois and Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, United States.

Hawkey, P. M.: Institute of Microbiology and Infection, University of Birmingham, Edgbaston Campus, Birmingham, UK. B15 2TT.

Holland, M.: Maidstone Hospital, Maidstone, Kent, United Kingdom.

Kato, H.: Department of Bacteriology II, National Institute of Infectious Diseases, Tokyo 208-0011, Japan.

Kim, H.: Department of Laboratory Medicine and Research Institute of Bacterial Resistance, Yonsei University College of Medicine, Seoul, Korea.

Kuijper, E. J.: National Reference Laboratory for *Clostridium difficile*, Leiden University Medical Centre and RIVM, Bilthoven. The Netherlands.

Lawley, T. D.: Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK. CB10 1SA.

Pituch, H.: Department of Medical Microbiology, Medical University of Warsaw, Warsaw, Poland.

Reid, S.: Department of Molecular and Cell Biology, University Avenue, Upper Campus, University of Cape Town, Rondebosch, 7701, South Africa.

Riley, T. V.: The University of Western Australia, School of Pathology and Laboratory Medicine, Crawley 6009. Australia.

Solomon, K.: School of Medicine and Medical Science, UCD Veterinary Sciences Centre, University College Dublin, Belfield, Dublin 4, Ireland.

Songer, J. G.: The University of Arizona, Department of Veterinary Sciences and Microbiology, Building 90, Room 212, 1117 East Lowell St., Tucson, AZ 85721, USA.

Tsai, P. J.: Department of Medical Laboratory Science and Biotechnology, National Cheng Kung University, Medical College, Tainan, Taiwan.

Weese, J. S.: Department of Pathobiology, University of Guelph, Guelph, Ontario, N1G2W1. Canada.

Chapter 1

Introduction

1 Introduction

The Gram positive bacterium *Clostridium difficile* causes gastrointestinal disease in humans and animals. Prior to 2003, there were few reports of *C. difficile* infection (CDI) yet in 2003, multiple outbreaks of CDI were reported in Europe, United States of America (USA) and Canada (Loo et al., 2005, McDonald et al., 2005, Warny et al., 2005, Labbe et al., 2008, Healthcare Commission, 2007). Outbreaks are primarily a result of patients with CDI excreting spores; these are highly infectious and allow for transmission and spread of infection (Riggs et al., 2007, Buggy et al., 1983, Sorg and Sonenshein, 2008, Fimlaid et al., 2013). Since the outbreaks between 2003 and 2007, the incidence of CDI has increased or decreased dependent upon geographical location. Europe has seen a decrease in incidence where counts in the UK have dropped from ~17,000 in 2007 to ~3,000 in 2017 (Public Health England, 2017). Canada and the USA have seen a steady rise and CDI has now been reported globally (Hawkey et al., 2013, Freeman et al., 2010, Public Health England, 2016, Rupnik et al., 2009, European Centre for Disease Prevention and Control, 2013).

The main risk factors associated with the development of CDI are consumption of broad-spectrum antibiotics, age ≥ 65 years and duration of stay within healthcare facilities (Bartlett, 2002, Kelly and LaMont, 2008). CDI is routinely treated with either metronidazole or vancomycin which are generally effective but other therapies such as faecal transplantation have also shown to be successful (Shahinas et al., 2012, Kassam et al., 2012, Kassam et al., 2013, van Nood et al., 2013, Kelly et al., 2014, Aroniadis et al., 2016). The development of infection and onward transmission

is prevented by antimicrobial stewardship and infection control practices (Gerding et al., 2008a, McNulty et al., 1997, Dingle et al., 2017).

1.1 The organism *Clostridium difficile*

The genus *Clostridium* is a member of the bacterial family *Peptostreptococcaceae* (formally *Clostridiaceae*). The type strain for the genus *Clostridium* is *Clostridium butyricum* (Yutin and Galperin, 2013, Lawson and Rainey, 2015) and due to differences recently observed on 16S rRNA sequence analysis between *C. difficile* and *C. butyricum, Clostridium perfringens* and *Clostridium tetani,* it has been proposed that *C. difficile* be re-named *Clostridioides difficile* (Lawson et al., 2016). Currently, *Clostridium difficile* remains the widely used terminology and is therefore used in this thesis.

The prokaryotic genus *Clostridium* are anaerobic, Gram-positive, rod-shaped bacteria and has achieved its status as a consequence of five species that are pathogenic to humans and animals; *C. botulinum* (botulism), *C. perfringens* (gas gangrene), *C. tetani* (tetanus), *Clostridium sordellii* (multi-organ disease) and *C. difficile* (gastrointestinal infection). The capability of *C. difficile* to form spores enables its survival in multiple environmental niches such as soil and water (al Saif and Brazier, 1996, Janezic et al., 2016). *C. difficile* has been found in different water sources such as the sea, rivers, lakes, inland drainage and swimming pools (al Saif and Brazier, 1996, Zidaric et al., 2010, Pasquale et al., 2011, Romano et al., 2012). *C. difficile* was initially named *Bacillus difficilis* and first described in 1935 as part of the neonatal microflora (Hall and O'Toole, 1935) and further characterised by Snyder in 1937 (Snyder, 1937). There were no other documented reports of *B.*

difficilis until 1960, when the renamed organism, *C. difficile* was cultured from the intestinal contents of a seal (McBee, 1960).

1.2 *C. difficile* infection

C. difficile was first suggested as a human pathogen when it was isolated from bodily samples of patients (Smith and King, 1962). In 1978 the first confirmed case of CDI was reported (Larson et al., 1978) and in 1999, the use of clindamycin, which was a commonly used antibiotic at the time, was associated with the occurrence of CDI (Johnson et al., 1999). Typical clinical features of CDI include; watery diarrhoea, abdominal pain and cramps, lower quadrant tenderness, fever, leucocytosis and hypoalbuminaemia (Mylonakis et al., 2001). CDI is highly variable, ranging from uncomplicated mild diarrhoea to life threatening toxic megacolon and pseudomembranous colitis (PMC) requiring surgical intervention or leading to death. CDI is the most frequent cause of healthcare-acquired infectious diarrhoea in developed countries and one of the major problems with CDI is reoccurrence of disease which is common with a rate of 15% to 35% (Garey et al., 2008, Johnson, 2009). These contribute to the significant morbidity and financial burden associated with CDI.

1.3 Risk factors for *C. difficile* infection

The surface area of a healthy human gastrointestinal tract is colonised with bacterial species: the gut microbiota. The dense gut microbiota has a symbiotic relationship with the host providing protection against exposure to dietary antigens, viable pathogens and bacterial products (Kato et al., 2001, Ozaki et al., 2004, Miyajima et al., 2011). Changes to the microbiota that have a negative effect on the host is termed

‘dysbiosis’ (Hawrelak and Myers, 2004). *C. difficile* is considered a member of the normal gut microbiota in less than 5% of healthy adults (Viscidi et al., 1981) and is recognised as a gut coloniser in domestic animals and livestock (Songer, 2004, Arroyo et al., 2005, Rodriguez-Palacios et al., 2009). In states of dysbiosis, *C. difficile* is able to flourish and invade the mucosal cells allowing toxins produced by *C. difficile* to cause similar infection in both humans and animals (Limaye et al., 2000). The two major risk factors for gut dysbiosis permitting CDI are the consumption of broad-spectrum antibiotics and age ≥ 65 years (Bignardi, 1998). The most frequently implicated antibiotics; clindamycin, penicillins, cephalosporin and the fluoroquinolones disrupt the microbiota of the gut and advancing age is thought to result in immune system failure known as ‘senescence’ of the immune response, resulting from a combination of comorbidities, immune related changes in the faecal flora and normal age related changes (Ginaldi et al., 2001).

1.4 Virulence factors

Strains of *C. difficile* implicated in CDI have been demonstrated to possess a multitude of virulence factors, these include toxins, surface associated proteins, sporulation and germination.

1.4.1 Toxins A and B

The pathogenesis of *C. difficile* in cases of CDI is hypothesised to primarily result from the production of two glucosylating enterotoxins, TcdA and TcdB which both cause damage to epithelial cells (Voth and Ballard, 2005, Chumbler et al., 2016) and are considered a major virulence factor in CDI. Their genes *tcdA* and *tcdB* are chromosomally located along with three accessory genes forming the 19.6 kb

Pathogenicity locus (PaLoc [Figure 1.1 A]). This is made up of*: tdcR* which encodes an alternative ribonucleic acid (RNA) polymerase sigma factor, a positive regulator of toxin production (Mani and Dupuy, 2001), *tcdC*, a negative regulator of toxin production which interferes with the RNA polymerase formed with *tcdR* (Matamouros et al., 2007), and *tcdE*, a gene thought to encode an holin like protein (Dupuy et al., 2008).

Strains of *C. difficile* do not always possess *tcdA* and *tcdB* genes and therefore may demonstrate variation in toxin production. This variation was exploited and a toxinotyping scheme was developed in 1998 (Rupnik et al., 1998) which was performed against 'VPI 10463', the first strain of *C. difficile* to have its toxin genes sequenced (von Eichel-Streiber et al., 1992). Toxinotyping is a PCR restriction fragment length poloymorphism based typing method assigning strains as toxinotypes. The reference strain VPI 10463 is designated as toxinotype 0 and strains similar are defined as 'nonvariant' strains of toxinotype 0. All other strains with changes in genes *tcdA* and/or *tcdB* are defined as 'variant strains'. Changes in the PaLoc range from minimal deletions limited to only *tcdA* and a few point mutations in *tcdB* to significant changes in *tcdB* and large deletions leaving only a residue of the PaLoc. Currently, there are 34 known toxinotypes which individually are designated with roman numerals (I to XXXIV). The genomic organisation of the PaLoc region of *C. difficile* and different toxinotypes are depicted in Figure 1.1.

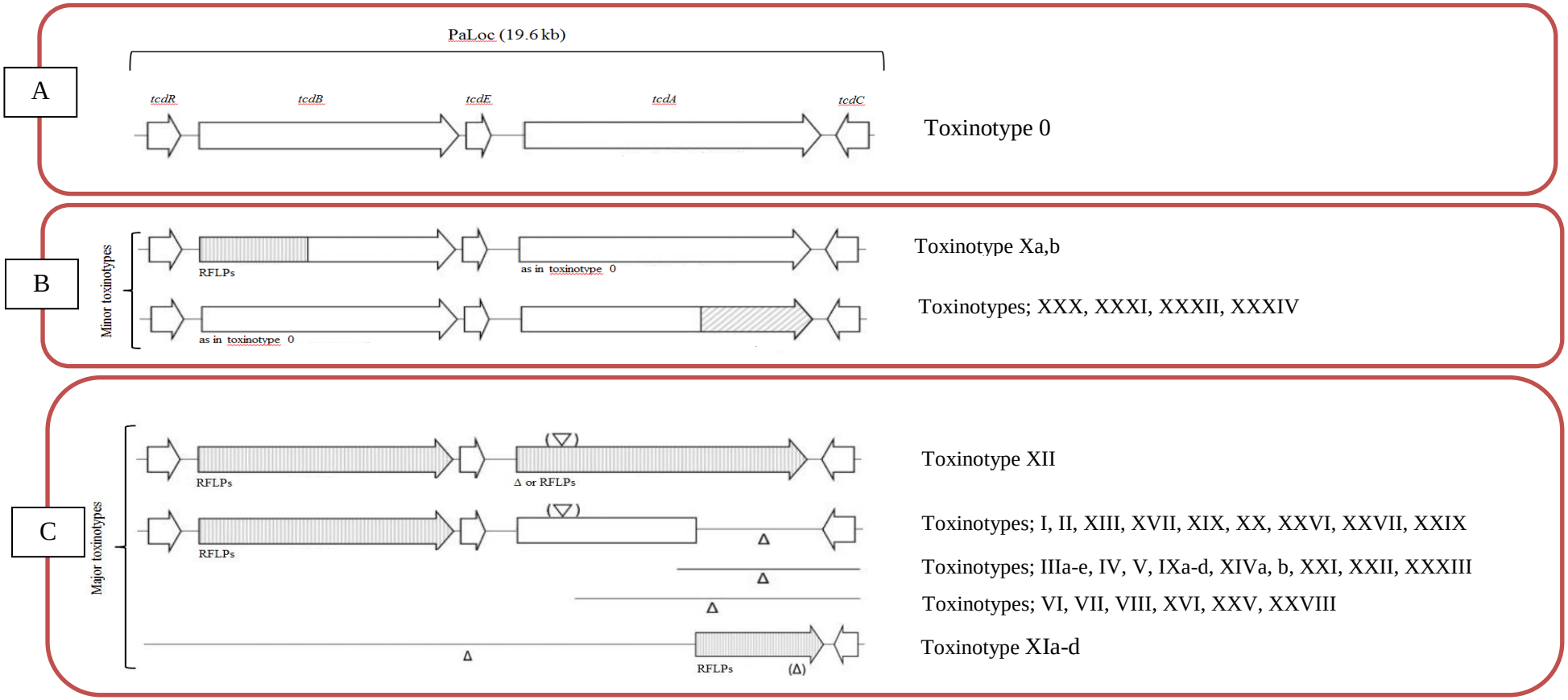


Figure 1.1: Genomic organisation of the PaLoc region of *C. difficile*
A figure adapted from a publication by Rupnik and Janezic which depicts the genetic variation in toxin types of *C. difficile* (Rupnik and Janezic, 2016). The genomic organisation of A) the PaLoc of toxinotype 0 B) Minor toxinotypes and C) Major toxinotypes. Changes in variant toxinotypes can be single SNPs (seen as RFLPs), deletions, or insertions. RFLPs (vertical stripes) are more common in the tcdB gene, while deletions and insertions (shown with diagonal stripes or graphical symbols) are more common in the tcdA gene.

The PaLoc is absent from non-toxigenic strains and in strains lacking this locus, the region is replaced by non-coding deoxyribonucleic acid (DNA) (Braun et al., 1996). Experiments by Brouwer *et al.*, have demonstrated that the PaLoc is mobile and is able to transfer from a toxigenic strain to a non-toxigenic strain (Brouwer et al., 2013). Strains that do not produce either toxin are non-pathogenic (Kuehne et al., 2010) and although the majority of pathogenic strains produce both toxins (A+B+), clinically relevant toxin A negative toxin B positive (A-B+) strains have been frequently described (Alfa et al., 2000) and are discussed further in Section 1.13.

1.4.2 Other toxins

Approximately 20% of strains of *C. difficile* have been observed to produce an actin modifying ADP-ribosylating binary toxin (Gerding et al., 2014), *C. difficile* transferase (CDT) (Schwan et al., 2009). It has been suggested that CDT has a role in adherence and colonisation of *C. difficile* to gut epithelial cells by stimulating microtubule-based protrusions from host cells (Schwan et al., 2009).

1.4.3 Sporulation and germination

C. difficile has the capacity to form spores; a dormant state that allows the bacterium to survive in conditions otherwise detrimental to their vegetative cell existence (Lawley et al., 2009, Paredes et al., 2005). Spores of *C. difficile* are highly resistant to adverse chemical and physical stress and can persist and contaminate an environment for up to several months (Setlow, 2007). The spore structure is made up of multiple layers including an exosporium, coat, cortex, membrane and a DNA core (Lawley et al., 2009). *C. difficile* spores are excreted with faeces from colonised and infected patients and can be ingested by a susceptible host (Deakin et al., 2012). The

spore is initially protected from the acids and enzymes in the stomach by their multiple layer structure until it reaches the intestinal tract where it is exposed to germinants such as bile salts (Giel et al., 2010). Here, the spore commences germination losing its spore-specific properties resulting in the upregulation of several genes and the reversal to a vegetative state.

In *Bacillus subtilis*, the gene *Spo0A* is a key transcriptional regulator that is required in the early stages of sporulation (Molle et al., 2003). *C. difficile* possesses a homolog of this gene and its role in CDI has been studied in a murine model of disease. These studies demonstrated that *C. difficile Spo0A* mutant derivatives can cause intestinal disease but are unable to persist within and effectively transmit between mice indicating that the *C. difficile Spo0A* gene plays a key role in persistent infection and host-to-host transmission (Deakin et al., 2012).

1.4.4 Surface associated proteins

Cell surface proteins are essential to the bacterial cells interaction within its environment. They may be responsible for motility, adhesion and invasion of host cells as well as defending against host responses (Lin et al., 2002, Niemann et al., 2004). *C. difficile* expresses numerous cell surface associated molecules including surface layer proteins (SLPs) (Spigaglia et al., 2013), pili (Maldarelli et al., 2014) and flagella (Dingle et al., 2011b, Aubry et al., 2012, Baban et al., 2013, Barketi-Klai et al., 2014).

The bacterial surface layer (S-layer) of *C. difficile* is proteinaceous, two-dimensional and coats the entire outer surface of the vegetative cell (Fagan and Fairweather,

2014). The S-layer is composed of the high molecular weight SLP and low molecular weight SLP. SLPs have been demonstrated to facilitate the initial colonisation of the gut by *C. difficile* (Sambol et al., 2000, Johnson et al., 2001) and low molecular weight SLPs play an antigenic role in immune evasion (Bianco et al., 2011, Ryan et al., 2011). The two SLPs are generated by post-translational cleavage of a pre-protein (SlpA) by the cell wall cysteine protease, Cwp84. SlpA is the most abundant protein of the S-layer and may act as an important colonisation factor (Ni Eidhin et al., 2008, Ryan et al., 2011, Sambol et al., 2000).

Pili (Latin for hair) are filamentous surface appendage structures initially identified in Gram-negative organisms (Anderson, 1949). Pili are associated with bacterial adhesion to host cells (Kline et al., 2009) and have been implicated in urinary, genital and gastrointestinal infections (Mulvey et al., 1998, Swanson, 1973, Boudeau et al., 2001). In Gram-positive bacteria, pili were first observed in *Corynebacterium renale* in 1968 (Yanagawa et al., 1968) and more recently have been observed in members of the Clostridia Class (Piepenbrink et al., 2015, Purcell et al., 2015). Pili in *C. difficile* have been associated with phenotypic variation and biofilm formation (Maldarelli et al., 2016) potentially contributing to the species virulence.

Flagella are whip-like protein appendages on a bacterial cell surface that drive motility of many bacterial species (Berg, 2003). Flagella have been shown to enable bacteria to colonise host cells in *Campylobacter jejuni* and *Vibrio chloerae* (Nachamkin et al., 1993, Richardson, 1991) and play a role in adherence in *V. chloerae* (Postnova et al., 1996).

Strains of *C. difficile* that exhibit peritrichous flagella are implicated in host cell colonisation (Tasteyre et al., 2001) and more recently, a direct link between flagella regulation and toxin production has been made (Dingle et al., 2011b, Aubry et al., 2012, Barketi-Klai et al., 2014). *C. difficile* flagella are genetically encoded by three distinct operons; F1, F2 and F3 (Stabler et al., 2009). Through the mutagenesis and complementation assays, Valiente *et al.*, investigated the function of the flagella associated genes and demonstrated their importance in motility, cell aggregation, biofilm formation, epithelial cell adhesion and recognition by the human immune system protein, toll-like receptor 5 (Valiente et al., 2016).

1.5 Diagnosis of *C. difficile* infection

The gold standard for the laboratory diagnosis of CDI is the direct detection of *C. difficile* toxin in faeces and the most recent guidance from the Department of Health states that a combination of two tests should be used (HPA, December 2008). These should be a Polymerase Chain Reaction (PCR) assay for the detection of *tcdB* or a glutamate dehydrogenase enzyme immunoassay followed by a sensitive enzyme immunoassay test for the detection of toxin in a stool sample. The laboratory diagnosis of CDI is often inconclusive due to the fact that the symptoms of CDI are similar to other intestinal disease and use of laxatives and side effects of some antimicrobials often result in diarrhoea. Generally, CDI is diagnosed clinically and the laboratory test results help confirm or rule out CDI.

1.6 Clinical management and therapeutic options

Once a patient has been diagnosed as suffering CDI, current antimicrobial therapy should be discontinued or altered. First line treatment is limited to only a few agents;

oral metronidazole and vancomycin with the latter being used for severe or recurrent cases of CDI and doses of vancomycin can be tapered or pulsed as a way of improving response and recovery (Gerding et al., 2008b). A more recently licensed antimicrobial, fidaxomicin is the first of a new class of narrow spectrum macrocyclic antibiotic drugs (Louie et al., 2011, Golan and Epstein, 2012). Fidaxomicin is bactericidal, has a prolonged post-antibiotic effect allowing for reduced dosing and has fewer side effects and a lower rate of reoccurrence when compared with metronidazole and vancomycin (Babakhani et al., 2011).

Therapies other than the use of antimicrobial drugs are available as last resort or are in development. The concept of 'faecal microbiota transplantation', also known as 'faecal biotherapy' and 'bacteriotherapy', was first described as a treatment of food poisoning and severe diarrhoea in Chinese literature from the 4th to the 16th century (Zhang et al., 2012) and referred to as 'yellow soup'. Faecal transplantation is practiced with the aim of restoring normal faecal microbiota where it has re-emerged as a treatment option for severe and recurrent CDI and many trials have investigated its efficacy with promising results (Kassam et al., 2013, Shahinas et al., 2012, van Nood et al., 2013, Kelly et al., 2014, Aroniadis et al., 2016). Studies investigating the optimal combination of gut bacteria as a treatment of severe or recurrent CDI are aiming to produce a customised microbiota pill to replace the faecal bacteria which potentially contain pathogens with normal healthy bacterial flora (Reeves et al., 2011, Reeves et al., 2012, Lawley et al., 2012, Petrof et al., 2013).

Bile salt analogues such as cholate metabenzene sulfonic acid that targets the interaction between *C. difficile* spores and taurocholate inhibiting germination have

been investigated as a potential therapeutic option (Howerton et al., 2013). Toxin binding agents such as Cholestyramine, a bile salt ion-exchange resin which has been shown to form a complex with *C. difficile* toxins and Tolevamer, a non-antibiotic polymer binds to *C. difficile* toxins have been examined in clinical studies, however, neither has been as efficacious as antimicrobial therapy (Weisman et al., 2015).

Immunotherapy such as vaccination brings sight of improving the current treatment options especially for recurrent CDI. A toxoid vaccine candidate produced by Sanofi Pasteur is currently being tested in phase III clinical trials and a recombinant vaccine consisting of two truncated toxins A and B by Valneva is in pre-clinical stages of development (Anosova et al., 2015, Greenberg et al., 2012, Foglia et al., 2012, Karczewski et al., 2014). Toxin-specific monoclonal antibody therapy as a treatment option is under investigation by Merck, and Sanofi Pasteur are developing a pair of monoclonal antibodies against toxins A and B (Lowy et al., 2010, Babcock et al., 2006, Anosova et al., 2015).

1.7 Management of outbreaks and infection control

Spores excreted with faeces from colonised and infected patients are easily transmitted via persons, fomites and air and are thought to be the main factors of environmental persistence and host transmission (Deakin et al., 2012). Spores of *C. difficile* demonstrate capacity to survive, persist and spread in hostile environments and cause outbreaks of CDI amongst susceptible hosts in wards, hospitals and other healthcare facilities (Kim et al., 1981). Infection control measures are required to limit the spread of *C. difficile* between patients and patients suspected of having CDI should be isolated to limit the spread of spores to other patients. Daily cleaning of the

environment and equipment that has been in contact with a patient with CDI with a chlorine agent is required, as is good hand hygiene by patients and staff, most notable with soap and water since alcohol gels are ineffective against *C. difficile* spores and can induce sporulation (Jabbar et al., 2010).

The consumption of antibiotics is a risk factor for the development of CDI, it has been demonstrated that by prescribing narrow-spectrum antibiotics as opposed to broad-spectrum antibiotics (where appropriate) can reduce the incidence of CDI. Antimicrobial stewardship, the monitoring and controlled use of antimicrobials has shown to reduce the burden of CDI (Fowler et al., 2007, Muto et al., 2007, Dingle et al., 2017, Pear et al., 1994).

All National Health Service trusts are required to participate in the Department of Health's mandatory CDI monitoring system whereby all cases of CDI in patients over the age of two years must be reported. Each acute care provider in the United Kingdom (UK) has a target number of CDI cases that they must not breach per financial year. Each CDI case should be assessed by performing a root cause analysis to determine if it was related to a lapse in the quality of care provided to patients. As part of a root cause analysis, samples and isolates maybe referred to their local Public Health England (PHE) *C. difficile* Ribotyping Network (CDRN) laboratory. The seven CDRN laboratories in the UK offer *C. difficile* PCR ribotyping (Section 1.8.1); these results are valuable for inferring and excluding transmission of *C. difficile*.

1.8 Genotyping schemes and routine surveillance

Typing methodologies enable bacterial isolates of the same species to be categorised as either related or distinct (Sabat et al., 2013) which has utility for the management of outbreaks by infection control teams. Identifying similarities or differences between strains of the same species and combining this with time and place of infection enables clinicians and epidemiologists to identify cross-transmission and/or outbreaks of infection. This is valuable for local, national and global surveillance whereby the number of infections and pathogen evolution can be monitored for changing prevalence's and/or pathogen virulence respectively.

Historically, typing methodologies have been based on exploiting phenotypic variations but molecular methods are now routinely used (Sabat et al., 2013).

Molecular typing approaches have been developed to study *C. difficile* and transmission including; multilocus sequence typing (MLST), multilocus variable-number tandem-repeat analysis (MLVA), Amplified fragment length polymorphism (AFLP), Pulsed-field gel electrophoresis (PFGE), Restriction endonuclease analysis (REA) and PCR ribotyping (Table 1.1). They vary in their discriminatory power (ability to differentiate between two unrelated strains) and therefore their application. Whole genome sequencing (WGS) is increasingly used to investigate outbreaks and epidemiology with finer resolution though this technique is not yet part of routine investigation of outbreaks of CDI in the UK (WGS is further discussed in Section 1.16, Page 58).

All typing methods apart from MLST are applied to the investigation of cross-transmission and/or outbreaks of infection (MLST is discussed further in Section 1.12, Page 53). PCR ribotyping is the most widely adopted scheme in Europe for genotyping strains of *C. difficile* with the Lead CDRN laboratory in Leeds, UK holding the collection of strains and assigning new PCR ribotypes (Public Health England., 2016); 877 PCR ribotypes have been identified (Dr Warren Fawley, personal communication 12/04/2018). In contrast, PFGE is the method of genotyping *C. difficile* in the USA. PCR ribotyping and PFGE are used for identifying cross-transmission, outbreaks of CDI in healthcare facilities and surveillance of the species.

Table 1.1: Molecular typing methods used for *C. difficile*

Application	Methodology	Reference
Outbreaks	AFLP	(Bowman et al., 1991)
	REA	(Clabots et al., 1993)
	PFGE	(Kato et al., 1999)
	PCR Ribotyping	(Stubbs et al., 1999)
	MLVA	(Marsh et al., 2010)
Evolutionary Studies & Surveillance	MLST	(Lemee et al., 2004)

AFLP - Amplified fragment length polymorphism
REA - Restriction endonuclease analysis
PFGE - Pulsed-field gel electrophoresis
PCR Ribotyping - Polymerase chain reaction Ribotyping
MLVA - Multilocus variable-number tandem-repeat analysis
MLST - Multilocus sequence typing

1.8.1 PCR ribotyping

PCR ribotyping was initially described in 1986 (Kuijper et al., 2006) and has since been applied to multiple bacterial genera (Eldar et al., 1997, Chesneau et al., 2000, Joung and Cote, 2002, Miteva et al., 2001, Yurlova et al., 1996, Salmenlinna and Vuopio-Varkila, 2001). The original method utilised for *C. difficile* was published in 1999 (Stubbs et al., 1999) and involves PCR amplification of the 16S to 23S intergenic spacer region of DNA using conserved primers (Figure 1.2). This region of DNA is both variable in length in base pairs (bp) and copy number between PCR ribotypes of the same species. It is this variation that following PCR amplification generates multiple PCR products which vary in size. While the ribosomal genes are quite conserved, the intergenic spacer regions vary considerably between *C. difficile* strains. Electrophoretic resolution of these DNA fragments generates profiles which are interpreted as a fingerprint, unique to that PCR ribotype, and can be used to distinguish *C. difficile* strains belonging to different PCR ribotypes (Figure 1.3).

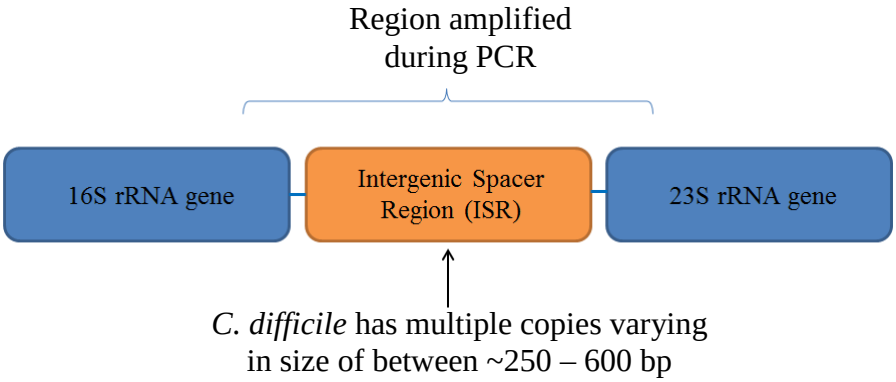


Figure 1.2: PCR ribotyping amplification of *C. difficile*

This figure illustrates the region of the *C. difficile* genome that is amplified during PCR ribotyping.

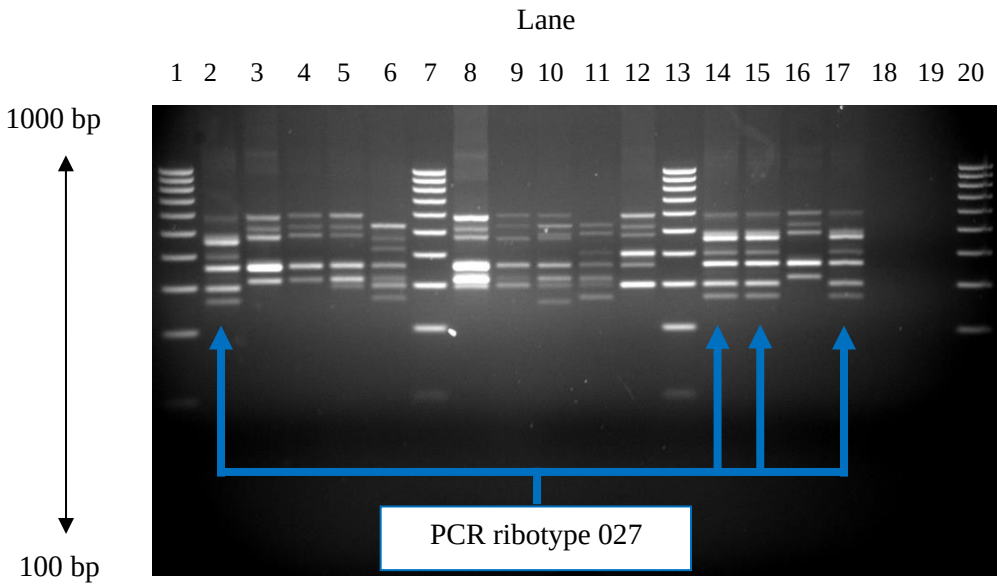


Figure 1.3: PCR ribotyping agarose gel electrophoresis of *C. difficile*

A figure of a typical PCR ribotyping agarose gel electrophoresis image of *C. difficile*. Lanes 1, 7, 13 and 20 are a 100 bp reference ladder, lanes 2 to 6, 8 to 12 and 14 to 17 are ribotyping profiles for different isolates and lanes 18 and 19 are negative extract and negative PCR controls respectively. Lanes 2, 14, 15 and 17 are also labelled to indicate a particular PCR ribotype; 027. The number of bands per lane indicates the copy number of the intergenic spacer region for that isolate and the size of the band indicates the length of the intergenic spacer region.

1.9 Global emergence of PCR ribotype 027

Increased recognition of CDI and the development of a PCR ribotyping scheme to differentiate between *C. difficile* strains was primarily due to the recognition of global outbreaks of a particular PCR ribotype; 027.

The earliest record of PCR ribotype 027 was in 1985 when the strain CD196 was isolated from a patient with CDI in a Parisian hospital (Popoff et al., 1988). PCR ribotype 027 was again isolated from a patient with CDI in a Minneapolis hospital in 1988 and designated BI-1 according to REA typing (Razaq et al., 2007). In 2003-2004, outbreaks in hospitals with patients experiencing notably severe CDIs were reported in Canada (Loo et al., 2005, McDonald et al., 2005) and between August 2004 and July 2007, more than 20,000 nosocomial cases of BI-1 were reported in the Quebec region (Labbe et al., 2008). Concurrently, BI-1 had also been isolated in several states in the USA (Warny et al., 2005) and in March 2004 a report was made by the Stoke Mandeville Hospital, UK describing a major outbreak including 334 CDI cases and 38 deaths (Healthcare Commission, 2007). Outbreak strains from the UK, USA and Canada were found to be due to the dissemination of a strain identified as BI by REA, NAP1 by PFGE and PCR ribotype 027 by PCR ribotyping and referred to as BI/NAP1/027 (Killgore et al., 2008). This was so named an ‘epidemic’ strain due to its propensity to cause outbreaks and the subsequent global spread is depicted in Figure 1.4. Reasons for the global emergence and spread of the epidemic PCR ribotype 027 strain of *C. difficile* are discussed in Section 1.10, Page 50.

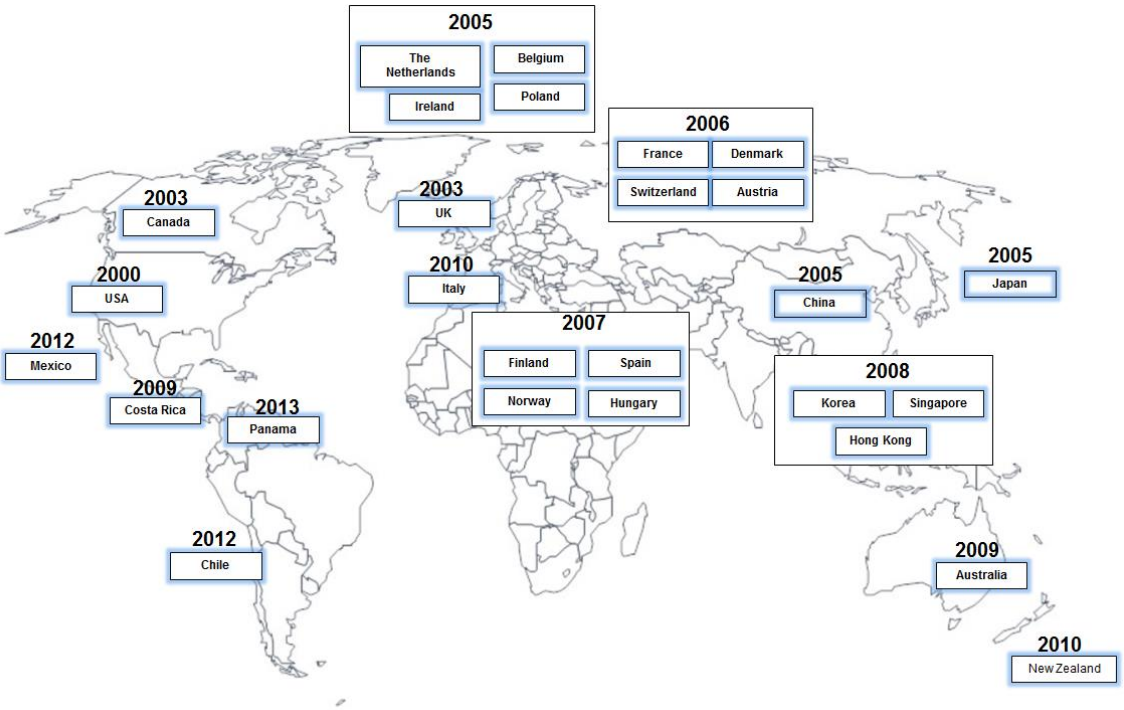


Figure 1.4: Map of isolation dates and global geographical spread of *C. difficile* PCR ribotype 027

Figure produced by M. Cairns and adapted from a publication by Valiente *et al.*, which depicts the dates and global spread of the epidemic *C. difficile* PCR ribotype 027 strain (Valiente et al., 2014).

1.10 Virulence and transmissibility of PCR ribotype 027

The establishment of PCR ribotype 027 and its apparent propensity to cause more severe diarrhoea, higher mortality and more re-occurrences of symptoms (Loo et al., 2005, Warny et al., 2005, Vohra and Poxton, 2011, Redelings et al., 2007) gave rise to this strain being described as 'hypervirulent' as well as epidemic when compared to other PCR ribotypes. A variety of phenotypic characteristics have been postulated to contribute to this ability to cause outbreaks and more severe CDI such as; increased toxin production (Warny et al., 2005, Curry et al., 2007, Freeman et al., 2007), presence of an 18 bp deletion and a frameshift mutation due to a single bp deletion in the *tcdC* gene affecting toxin expression (Dupuy et al., 2008, MacCannell et al., 2006). Some PCR ribotypes including PCR ribotype 027 produce a binary toxin (Stubbs et al., 2000, Sundriyal et al., 2010) encoded by two genes, *cdtA* and *cdtB* (Carter et al., 2007) which has been linked with increased severity of disease (Barbut et al., 2005, McEllistrem et al., 2005). Isolates of PCR ribotype 027 are also reported to sporulate earlier in the bacterial growth cycle and produce more spores in total compared to non-PCR ribotype 027 isolates and this was postulated to increase the rate of transmission (Merrigan et al., 2010, Fawley et al., 2007). Germination rates for isolates of PCR ribotype 027 have been shown to be higher (Burns et al., 2010a) and exhibit higher germination efficiencies compared to other PCR ribotypes in the presence of 0.1 % (w/v) sodium taurocholate (Moore et al., 2013).

However, these observations remain contentious due to other evidence not supporting these suggestions (Dupuy et al., 2008, Burns et al., 2011) nor that these characteristic are PCR ribotype specific (Burns et al., 2010b). *C. difficile* spore germination in response to sodium taurocholate varied significantly even amongst

isolates of PCR ribotype 027 (Heeg et al., 2012). Therefore, no definitive virulence factors have convincingly been attributed to isolates of PCR ribotype 027.

1.11 Distribution of *C. difficile* PCR ribotypes

Since the emergence and explosion of outbreaks of CDI associated with PCR ribotype 027, there have been significant changes to the diversity of *C. difficile*. In the UK, rates of CDI, most notably PCR ribotype 027 have declined as depicted in Figure 1.5. This has occurred simultaneously with an increase in a variety of other PCR ribotypes, and in general, the pattern of PCR ribotypes in the UK has become more heterogeneous (Public Health England., 2016). This is likely due to multiple factors including; increased testing and PCR ribotyping, changing antibiotic usage patterns and emergence of new strains. Europe has also seen an increase in PCR ribotype diversity. Previously, 65 different PCR ribotypes were identified across 26 countries with PCR ribotype 027 the most commonly isolated (18.4%) (Bauer et al., 2011). More recently, a European-wide study found 125 PCR ribotypes with no dominant PCR ribotype (Davies et al., 2016). Recent data from the USA suggests a different PCR ribotype diversity compared to Europe; although diversity has increased, PCR ribotype 027 remains the leading cause of CDI in the USA (Rusk, 2011). Although CDI is a major problem in Europe, Canada and USA, less is known about its epidemiology for other continents. The few studies that have been performed in other geographical locations, for example Asia, demonstrate that CDI is a significant cause of nosocomial disease (Hawkey et al., 2013). There is a pressing need to understand the sources and routes of transmission of *C. difficile* beyond PCR ribotyping.

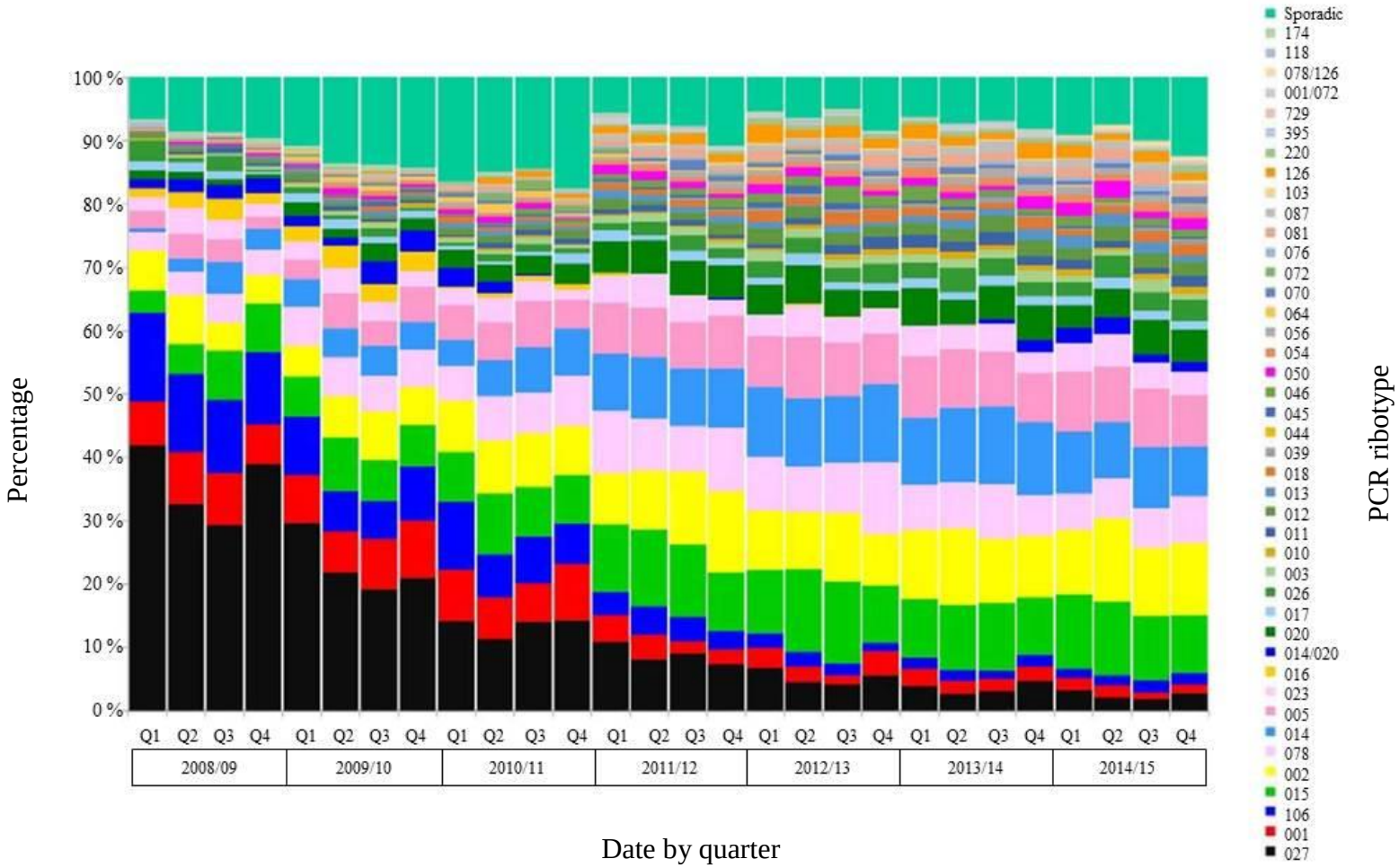


Figure 1.5: Prevalence and diversity of *C. difficile* PCR ribotypes in England by quarter (April 2008 to March 2015)

Courtesy of the 2016 CDRN report (Public Health England., 2016).

1.12 Phylogenetic studies

Multilocus sequence typing (MLST) is a useful approach for studying population and evolutionary genetics and the epidemiology of bacteria. It is a molecular epidemiological typing technique that permits microbial isolate characterisation by exploiting the nucleotide sequences of housekeeping gene fragments. The method utilises the DNA sequences of multiple (usually seven) housekeeping genes (approximately 450-500 bp). Each housekeeping gene is defined as distinct sequences (alleles), and for each isolate, the combination of sequences define the allelic profile or sequence type (Griffiths et al., 2010). The rationale is that a single genetic event resulting in a new allele can occur by point mutation (altering only a single nucleotide site), or by a recombinational replacement (will often change multiple sites).

MLST was first developed by Brian Spratt's research group for *Neisseria meningitides* (Spratt, 1999) and has since successfully been used to study the population genetics of other organisms including; *Streptococcus pneumoniae* (Rayner et al., 2015), *Staphylococci* species, *Streptococcus pyogenes* (Enright et al., 2001), *Haemophilus influenzae* (Meats et al., 2003), *Campylobacter jejuni* (Dingle et al., 2011a), *Enterococcus faecium* (Homan et al., 2002, Burgos et al., 2014, de Been et al., 2015) among others.

There are many terms used to describe and classify the phylogenetic relatedness between bacterial strains; Table 1.2 describes some of the common terms used in the literature and in this thesis.

Table 1.2: Phylogenetic relatedness terminology

Terminology	Definition
Lineage	A single, temporal phylogenetic division of a continual line of ancestral decent
Sub-lineage	A sub-ordinate descent of a lineage
Clade	A phylogenetic division where strains share a common ancestor (not necessarily linear in terms of ancestral decent)
Clone	Strains that are indistinguishable in genotype

In this thesis and associated publications where I am first author, I refer to lineage and sub-lineage and not clade.

MLST studies on diverse collections of *C. difficile* strains suggest that the species can be divided into five ancestral lineages. These five MLST lineages of *C. difficile* can be divided by PCR ribotype; 027, 017, 023, 078 and a large group including the rest of the PCR ribotypes with PCR ribotype 078 representing an interestingly highly divergent lineage (Lemee et al., 2004, Lemee et al., 2005, Griffiths et al., 2010, Dingle et al., 2011a). These lineages are depicted in Figure 1.6.

The rapid spread of PCR ribotype 027 has detracted attention from other virulent PCR ribotypes of *C. difficile* (PCR ribotype 017 [Section 1.13, Page 56], PCR ribotype 078 [Section 1.14, Page 57] and PCR ribotype 023 [Section 1.15, Page 58]) and coupled with the identification of PCR ribotype specific lineages other than PCR ribotype 027, the study of other PCR ribotypes and lineages has come to prominence.

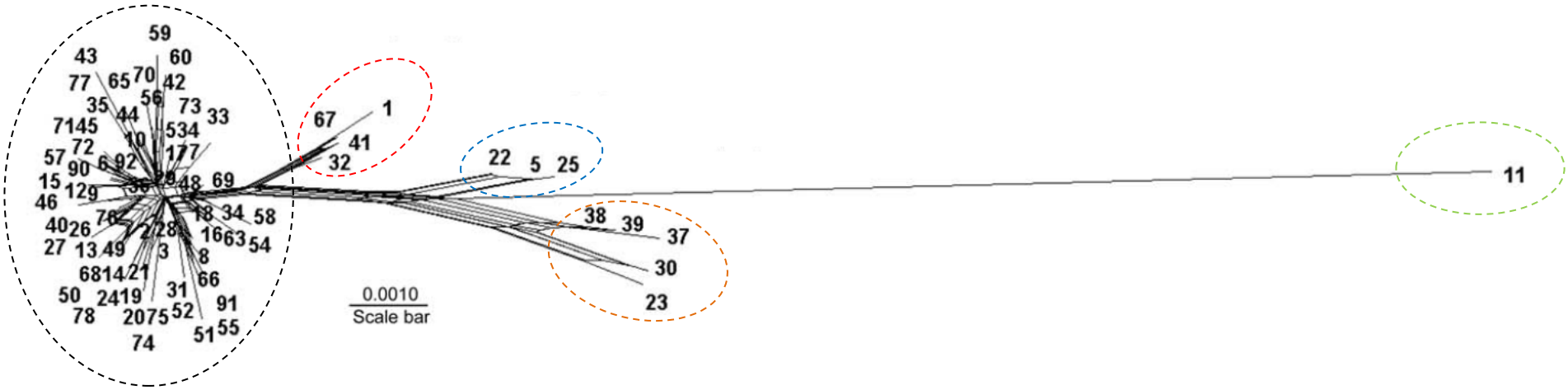


Figure 1.6: The five phylogenetic lineages of *C. difficile* by MLST analysis

A figure adapted from a publication by Dingle *et al.*, illustrates the phylogenetic relationships among 78 sequence types comprising the five lineages (Dingle et al., 2011a). Branch colouring; black = lineage 1 (mixed PCR ribotypes), red = lineage 2 (PCR ribotype 027), blue = lineage 3 (PCR ribotype 023), orange = lineage 4 (PCR ribotype 017), and green = lineage 5 (PCR ribotype 078). The numerical figures depicted within the encircled lineages are designated sequence types.

1.13 The emergence of toxin A-B+ strains and PCR ribotype 017

The first documentation of a toxin A-B+ strain of *C. difficile* was by Haslam *et al*., in 1986 who identified a clinical isolate (8864) of *C. difficile* that did not produce toxin A using an enzyme-linked immunoabsorbent assay (Haslam et al., 1986). This was confirmed using either an enzyme-linked immunoabsorbent assay or probes specific to toxin A in 1991 (Torres, 1991), 1992 (Borriello et al., 1992, Lyerly et al., 1992) and 1993 (Depitre et al., 1993). It was later shown that the toxin A negative samples did possess the toxin A gene using PCR (Pituch et al., 1998) and that the loss of toxin A expression was due to a nonsense point mutation (von Eichel-Streiber et al., 1999).

The earliest reports of outbreaks of CDI due to toxin A-B+ *C. difficile* strains originated in Canada in 1998 (Alfa et al., 2000) and the Netherlands between 1997 and 1998 (Kuijper et al., 2001). This was followed with further isolation in both Japan (Komatsu et al., 2003, Kato et al., 1999), the UK, Belgium and USA (Johnson et al., 2003) and Ireland (Drudy et al., 2007b). Outbreak strains of toxin A-B+ were compared using three typing methodologies and 20/23 isolates were found to be PCR ribotype 017 (others being PCR ribotypes; 110, 036 and 047) (Johnson et al., 2003).

PCR ribotype 017 is toxinotype VIII and it has since been reported in Poland (Pituch et al., 2001, Pituch et al., 2006), Israel (Samra et al., 2002), China (Huang et al., 2010), Korea (Kim et al., 2010a, Kim et al., 2010b, Shin et al., 2008) (where PCR ribotype 017 was the predominant PCR ribotype), Argentina (Goorhuis et al., 2009), Israel (Samra et al., 2002), Australia (Elliott et al., 2011) and Thailand (Ngamskulrungroj et al., 2015). Given that some diagnostic laboratories rely on

detecting toxin A only, the incidence of toxin A-B+ PCR ribotype 017 is likely to be significantly under reported.

1.14 The emergence of PCR ribotype 078

PCR ribotype 078 is the predominant PCR ribotype isolated from animal species with CDI, most notably pigs, calves and horses (Keel et al., 2007, Jhung et al., 2008, Goorhuis et al., 2008, Rupnik et al., 2008). *C. difficile* has also been found in contaminated food for human consumption with PCR ribotype 078 being the most frequently implicated (Rodriguez-Palacios et al., 2009, Broda et al., 1996, Songer et al., 2009, Simango and Mwakurudza, 2008). It was found in 4.8% (5/119) of seafood and fish samples from a grocery store; all toxin positive isolates were found to be PCR ribotype 078 (Metcalf et al., 2011).

Although CDI is primarily associated with exposure to a healthcare associated environment, patients from the community can develop CDI; these numbers also appear to be on the rise. With the increased recognition of *C. difficile* from food products for human consumption, there are concerns regarding transmission between animals and humans (or possibly humans to animals). Although *C. difficile* is not a proven food-borne pathogen, there is evidence that the same strain can cause symptomatic disease in both pigs and humans (Debast et al., 2009) and studies from several countries have found certain strains to be indistinguishable between human, animal and food origin (Jhung et al., 2008, Goorhuis et al., 2008, Debast et al., 2009, Gould and Limbago, 2010).

1.15 The potential emergence of other PCR ribotypes

MLST analyses revealed a distinct lineage of *C. difficile* that predominantly contains strains of PCR ribotype 023 (Figure 1.6). Although this PCR ribotype is not well described like PCR ribotypes 027, 078 and 017, two studies have reported its prevalence in Europe (Bauer et al., 2011, Barbut et al., 2007) and retrospective analysis in the UK associated PCR ribotype 023 with severe CDI (Wren et al., 2009).

MLST analyses more recently identified a novel 6th lineage that formed a well-separated branch in the phylogenetic tree and contained sequence type 122 and PCR ribotype 131 (Knetsch et al., 2012). Other common PCR ribotypes including 001, 002, 014, 015 and 106, do not form distinct lineages by MLST analysis like seen with PCR ribotypes 027, 078, 017 and 023, but instead, occur within the same heterogeneous group (Figure 1.6).

1.16 Genome-based analysis

Application of both PCR ribotyping and MLST techniques has demonstrated the utility of these methods in describing the epidemiology and population structure of *C. difficile*. Typing methods lack the sensitivity to precisely discriminate between isolates so as to infer or exclude transmission events. Genotyping methodologies also only detect variation in specific regions of a genome. In order to address this it is necessary to describe the whole genome of a strain under investigation.

WGS methodologies enable the genomic DNA sequence of an organism or population to be determined (Heather and Chain, 2016). In comparison to other typing methods which target specific and limited regions of the genome, WGS

techniques provide greater resolution than conventional typing methodologies with the ability to distinguish strains that differ at only a single nucleotide per genome (Roetzer et al., 2013). The capacity of WGS to describe genome content means that it can readily identify DNA acquisition and loss events (Xu et al., 2016); current typing methodologies may or may not reveal the horizontal transfer of genes amongst strains.

WGS has been demonstrated to be superior to conventional genotyping for phylogenetic, evolutionary and outbreak analyses, the methodology has become the 'gold-standard' for the description of bacterial phylogenetic relatedness of multiple species of bacterial pathogens (examples are provided in Table 1.3). Since the advent of 'first generation' DNA sequencing technologies in 1977 with Sanger's 'chain-termination' technique (Sanger et al., 1977), sequencing technology has significantly evolved. The expansion of 'next-generation' high throughput sequencing technologies have the advantages of; a requirement for less input DNA, increased turn-around-time of data, reduced cost and improved accuracy. The common next-generation technologies are; Illumina (Solexa) sequencing, Roche 454 sequencing, Ion Torrent sequencing and SOLiD sequencing. This has been followed by 'third-generation' sequencing technologies which emerged in 2010 (McCarthy, 2010). These include Pacific BioSciences (PacBio) and Oxford Nanopore Technologies and generate significantly increased read lengths. Table 1.4 summarises a comparison of sequencing technologies.

Table 1.3: Selected studies using WGS for pathogens

Pathogen	Number of isolates	Focus of study	Reference
Mycobacterium tuberculosis	86	Outbreak related transmission	(Roetzer et al., 2013)
	2,099	Drug resistance	(Walker et al., 2015)
	24	WGS direct from clinical samples	(Brown et al., 2015)
	13	Comparison of genetic variation to phenotypic characteristics	(Satta et al., 2016)
Vibrio cholerae	5	Identification of geographic source of outbreak	(Chin et al., 2011)
Staphylococcus aureus	63	Micro-evolution, geographical clustering, intercontinental spread and transmission events	(Harris et al., 2010)
	181	Transmission between residents of healthcare facilities and environmental contamination	(Harrison et al., 2016)
	1,013	Macro-epidemiology of Meticillin resistant *S. aureus* (MRSA)	(Reuter et al., 2016)
	308	Identification of high-risk clones	(Aanensen et al., 2016)
Vancomycin Resistant Enterococci	45	Evidence for genetic relatedness between long term carriage and bloodstream infections	(Brodrick et al., 2016)
Measles virus	27	Outbreak related transmission	(Gardy et al., 2015)
Salmonella spp.	6,887	Comparison with conventional typing methodology	(Ashton et al., 2016)
	29	Comparison with conventional typing methodology	(Bale et al., 2016)
Campylobacter spp.	1,713	Evaluation of molecular diagnostic assays	(Jansen van Rensburg et al., 2016)
Candida auris	41	Identification of clonal populations on three continents	(Lockhart et al., 2017)

Table 1.4: WGS technologies

Generation	Technology (Company)	Reference	Chemistry	Read length	Strengths	Weaknesses
First	**Sanger Sequencing** (Applied Biosystems Inc, California, USA)	(Fleischmann et al., 1995)	Chain termination	Up to 900 bp	Improved on the mainstay approach of multiple PCRs	Low throughput
Second (next)	**Illumina** (Illumina Ltd, Cambridge, UK)	(Bentley et al., 2008)	Synthesis	Up to 600 bp	Versatile for a variety of applications	High cost compared with other technologies
	Roche 454 (Roche Diagnostics Ltd, Basel, Switzerland)	(Margulies et al., 2005)	Synthesis	Up to 1 kb	Long reads	High cost compared with other technologies
	Ion Torrent (Ion Torrent Systems Inc, Gilford, USA)	(Rusk, 2010)	Synthesis	Up to 400 bp	Short run time and low cost compared with other technologies	Low throughput and short read length
	SOLiD sequencing (ThermoFisher Scientific Inc, Massachusetts, USA)	(Shendure et al., 2005)	Ligation	Up to 100 bp	High throughput and high accuracy	Short read length
Third	**PacBio** (Pacific Biosciences Inc, California, USA)	(Chin et al., 2011)	Single-molecule, real-time long reads	Up to 20,000 bp	Long reads, short run time and ability to sequence regions of high G/C content	Poor accuracy
	MinION (Oxford Nanopore Technologies, Ltd, Oxford, UK)	(Madoui et al., 2015)	Nanopore	Theoretically, no instrument restriction on read length	Long reads, sequencing in real-time and portable platform	Poor accuracy

1.17 Genome sequencing of *C. difficile*

Following on from the identification of PCR ribotype specific lineages through MLST analysis and the heightened interest associated with PCR ribotype 027, using DNA sequencing and phenotypic assays Stabler *et al.,* performed a three-way, whole-genome comparison between the original PCR ribotype 027 Paris strain (CD196), the index PCR ribotype 027 strain from the Stoke Mandeville outbreak in the UK (R20291) and the first published sequence strain of *C. difficile,* strain 630, PCR ribotype 012 which was isolated from a patient with severe PMC and caused an outbreak of diarrheal disease in a Swiss hospital (Sebaihia et al., 2006, Wust et al., 1982, Stabler et al., 2009). Phenotypic differences were observed with motility, antibiotic resistance and toxicity, lineage specific genes were present in the two PCR ribotype 027 genomes and the modern epidemic PCR ribotype 027 strains had five unique genetic regions absent from both the non-epidemic PCR ribotype 027 and the PCR ribotype 012 strains. Subsequently, a SNP based study of 21 PCR ribotype 027 strains of *C. difficile* revealed chronological microevolution and confirmed PCR ribotypes; 017, 027 and 078 strains to form individual lineages with the latter appearing to be highly divergent (He et al., 2010).

Using illumina WGS technology to sequence the genomes of a global collection (n=151) of *C. difficile* BI/NAP1/027 isolates primarily from hospital patients between 1985 and 2010, He *et al.,* identified 3,686 SNPs which inferred the presence of two phylogenetic sub-lineages; fluoroquinolone resistance (FQR) 1 and FQR2 (He et al., 2013). The authors suggest both FQR1 and FQR2 evolved independently and acquired an identical mutation in the DNA gyrase gene leading to high level fluoroquinolone resistance. FQR1 and FQR2 are depicted in Figure 1.7.

The authors imply that the FQR1 sub-lineage originated in Pennsylvania, USA in 2001 whereas the FQR2 sub-lineage emerged in USA with the first documented isolate from Montreal in 2003. The FQR2 sub-lineage was shown to disseminate across the Atlantic with multiple introductions into the UK causing hospital outbreaks between 2004 and 2006. Isolates from the base of the phylogeny and outside of both epidemic sub-lineages are from various geographical locations and are not thought to be associated with major hospital outbreaks suggesting that these represent pre-epidemic isolates from which the two epidemic sub-lineages emerged. Additionally, fluoroquinolone antibiotics were one of the most commonly prescribed antibiotic classes in the late 1990s and early 2000 in the USA (Linder et al., 2005) so it is plausible to suggest that during this time, this created a selective pressure for the acquisition and maintenance of fluoroquinolone resistance and provided the environment for the emergence of both sub-lineages.

Only two and seven SNPs defined the branches leading to the two epidemic sub-lineages respectively. However, besides the DNA gyrase gene mutation there were no SNPs that were shared by both sub-lineages and little evidence that a substantial change in phenotype could result from any of the SNPs that define the two epidemic sub-lineages.

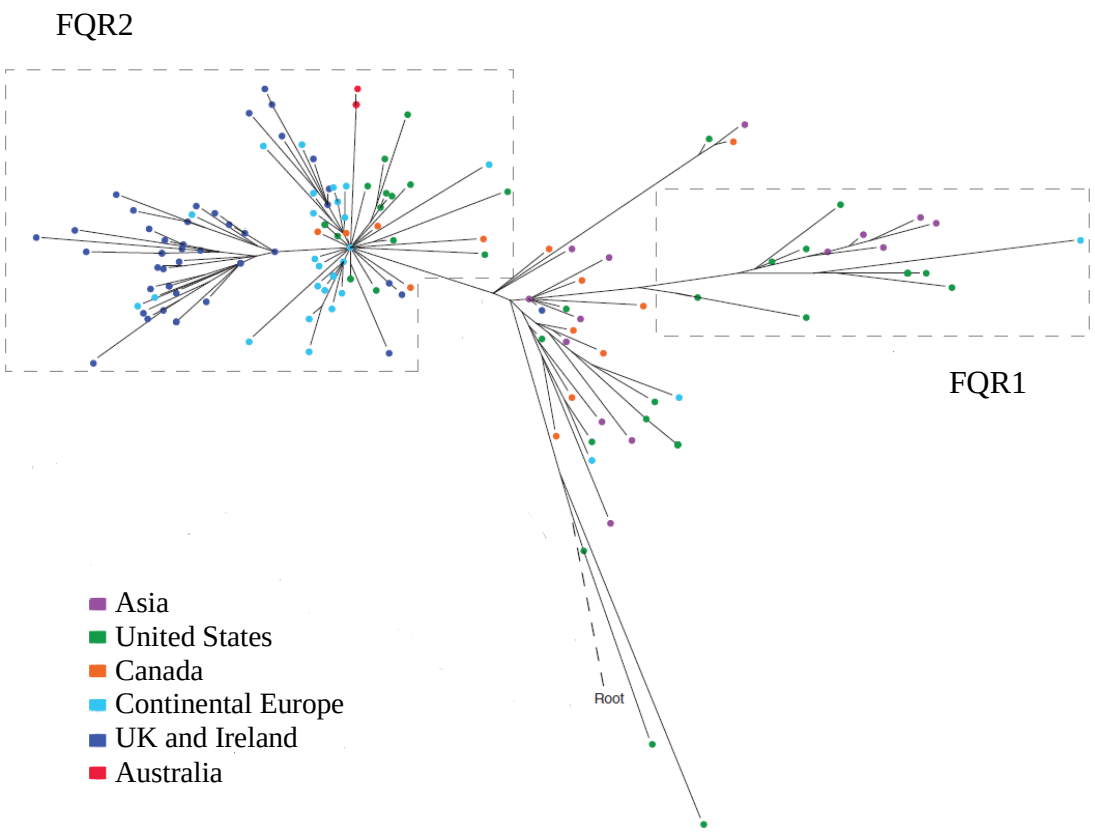


Figure 1.7: Global phylogeny of *C. difficile* PCR ribotype 027

A figure adapted from a publication by He *et al.*, that depicts the global phylogeny of 151 *C. difficile* 027/BI/NAP1 isolates based on core genome SNPs (He et al., 2013). Coloured nodes indicate the geographic source of the isolates. The position of the inferred root is indicated by a dashed line, and dashed outlines enclose the isolates with the mutation associated with fluoroquinolone resistance.

1.18 Aims of thesis

Given the poor understanding of the phylohistory and phylogeography of *C. difficile*, the focus of this thesis is to enhance the understanding of the phylogeny of this species with an overarching aim to investigate and understand the changing evolution of *C. difficile* by:

- Investigating the population structure of a collection of *C. difficile* strains isolated from diverse human, animal and food sources, continents using different typing methodologies and comparing the correlation of typing methodologies used for inter-laboratory comparison (chapter 3, Page 93).

- Investigating the genetic relatedness of a collection of *C. difficile* PCR ribotype 017 strains isolated from London hospitals using WGS and phenotypic assays (chapter 4, Page 135).

- Investigating the global population structure of *C. difficile* PCR ribotype 017 using WGS and antimicrobial susceptibility assays (chapter 5, Page 177).

Chapter 2

Materials and Methods

2 Materials and Methods

2.1 Materials

Unless otherwise indicated, chemicals and biochemicals were supplied by Sigma-Aldrich, Gillingham, UK and bacterial growth media and commercially prepared agar plates were supplied by Oxoid, Basingstoke, UK.

2.1.1 Bacterial study isolates

All bacterial isolates used in this thesis are shown in Appendix 1, Page 229, Table 4.1, Page 143 and Appendix 2, Page 249 for Chapters 3, 4 and 5 respectively and their sources are described on Page 27. Isolates were received from collaborators on agar plates, swabs or in broth suspensions and were recovered by inoculating to blood agar (Section 2.2.2, Page 70).The maximum passage number for the isolates provided by the London CDRN Laboratory was three.

2.1.2 Bacterial control isolates

The *C. difficile* PCR ribotype 017 reference strain M68 and its reference genome; GenBank accession number FN668375 (He et al., 2010) were used as controls for assays, sequencing and analysis. This strain was selected as the control since it is the fully sequenced annotated PCR ribotype 017 strain and is widely used and well described. The *C. difficile* PCR ribotype 027 reference strain R20291 was also used as a control for PCR ribotyping assays since this was part of the standard operating procedure for PCR ribotyping at the CDRN regional London laboratory.

2.1.3 Data collection

Meta data for isolates (source, origin, sample date and geographical location) were collated by contacting the source (Page 27) and by examination of patient request forms for isolates provided by the CDRN regional London laboratory.

2.2 Microbiological methods

2.2.1 Sterilisation

All reagents (where indicated), media and solutions were sterilised by autoclaving at 121 degree Celcius (°C) for 20 minutes in a wet autoclave (LTE Scientific, Oldham, UK) or dry steam sterilisation was performed at 134 °C for 15 minutes in an AAJ autoclave (Astell Scientific, Sidcup, UK). Filter sterilisation was performed using a 10 ml syringe (BD Plastipak, Oxford, UK) and a 0.2 μm (32 mm) Acrodisc® syringe filter with Stupor® membrane (Pall Life Sciences, UK).

2.2.2 Bacterial growth media

Isolates were routinely grown on blood agar which was either prepared in-house or commercially purchased. In-house blood agar plates were prepared by dissolving 16 g of Columbia blood agar base in 500 ml Milli-Q grade water (Millipore, Billerica, USA) in a 500 ml bottle. The suspension was inverted to mix and then autoclaved at 121 °C for 20 minutes and allowed to cool. Once cooled to handheld temperature, 7 % (vol/vol) defribrinated horse blood (Scientific Laboratory Supplies, UK) was added, the bottle gently inverted to mix and the suspension then poured to petri dishes (Thermo Scientific, USA) to make 20 ml plates. All other growth media made in-house are described in Table 2.1.

2.2.3 Bacterial growth conditions

Agar plates were routinely incubated at 37 °C under anaerobic conditions using a Don Whitley A85 anaerobic workstation filled with a mixture of 80 % nitrogen, 10 % hydrogen and 10 % carbon dioxide (Don Whitley Scientific, West Yorkshire, United Kingdom).

Table 2.1: Details of growth media made in-house

Media	Acronym	Composition	Supplemented
Brain heart infusion (BHI) broth	BHIS	18.5 g Brain heart infusion 2.5 g Yeast extract 500 ml Milli-Q grade water	**C. difficile* supplement 25 % L-cysteine
Yeast peptone sporulation medium	SM	8 g Peptone 4 g Yeast extract 2.5 g NaCL 500 ml Milli-Q grade water	**C. difficile* supplement 1 ml Tween 80 (0.2 % vol/vol)
Blood agar with sodium taurocholate	BA_ST	100 µl 0.1 % filter sterilised sodium taurocholate spread over plate and allowed to air dry around a flame	Nil

*contains the antibiotics D-cycloserine and Cefoxitin to inhibit the growth of other bacterial organisms (OXOID).

2.2.4 Storage of isolates

All *C. difficile* isolates were stored at -80 °C in duplicate. Using a loop, biomass from 72 hour growth on blood agar was swept and inoculated into 1.5 cryovials containing 20 % glycerol, 20 % SM medium and 60 % molecular grade water. These were stored at -80 °C (New Brunswick Scientific, St Albans, UK).

2.2.5 Hospital ward environmental screen

The infection control nurses at University Hospital Lewisham (UHL) performed an environmental screen by dampening Amies swabs (Medical Wire, UK) with sterile water prior to swabbing the following environmental surfaces; nurses station keyboard and telephone, mobile computer, patient bay door handle, sink tap, table ledge, window ledge, toilet arm rest, housekeeping room door handle, dirty utility rooms slipper-pan, wash bowl, apron holder, door handle, store room hoist, drugs trolley, medical notes trolley, pantry door handle, side-room door handle, side-room bed frame, side-room window ledge, side-room patient locker, side-room floor and side-room toilet.

All swabs were referred to the CDRN regional London laboratory. To recover *C. difficile*, swabs were inoculated directly onto Brazier agar medium and incubated anaerobically for 48 hours. Brazier agar medium contains cholic acid to promote spore germination, and cycloserine and cefoxitin to inhibit the growth of other bacterial organisms. Growth from Brazier agar medium was subbed to blood agar and incubated anaerobically for 48 hours. Pure growth was subjected to PCR ribotyping (Section 2.3.1, Page 77).

To enhance the recovery of *C. difficile* that maybe in small numbers, swabs were also inoculated into Robertson's cooked meat medium (RCM) and incubated for five days at room temperature (RCM aids the recovery of bacterial organisms as it contains glucose for rapid, heavy growth of anaerobic bacteria). Following incubation, RCM broths were sub-cultured to Brazier agar medium and incubated anaerobically for 48

hours. This was followed by subbing to blood agar and PCR ribotyping (Section 2.3.1, Page 77).

To again enhance recovery of *C. difficile* by killing all vegetative cells from other genera and species that are likely to also be present on the swab samples, alcohol shock was performed on the RCM broths. Briefly, 1.5 ml RCM broth was added to a bijoux containing 0.5 ml saline. Equal volume, 2 ml of Industrialised Methylated Spirit (IMS) [VWR, USA]) was added to the bijoux which was vortexed and left to stand at room temperature for 45 – 60 minutes. After incubation, three drops from the lower layer of the bijoux was subbed to Brazier agar medium. This was followed by sub-culture to blood agar and subjected to PCR ribotyping (Section 2.3.1, Page 77).

2.2.6 Growth kinetic assays

To measure the growth rates of *C. difficile*, 10 ml pre-equilibrated BHIS broth were inoculated with three colonies of blood agar culture and incubated at 37 °C anaerobically on a shaking platform (Orbital Shaker SSL1, Stuart^{TM} Staffordshire, UK) at 60 rpm for 24 hours. New cultures were made by inoculating fresh, pre-equilibrated BHIS broths with 1/100 of the initial 24 hour culture. Every hour for the first 9 hours and then again at 24 hours, 1 ml aliquots of each new culture were removed and their optical density (OD) at λ590 (OD_{600}) was measured in a spectrophotometer (WPA CO8000, Biochrom, UK).

2.2.7 MIC assays

Minimum Inhibitory Concentrations (MICs) of a panel of antibiotics were determined using the broth dilution method as previously described (Andrews,

2001). Briefly, isolates were grown on blood agar anaerobically for 48 hours. Ten ml aliquots of BHI broth with *C. difficile* supplement and cysteine were made in 20 ml universal containers (VWR, USA) and placed in the anaerobic cabinet and allowed to equilibrate anaerobically overnight.

Antibiotics were reconstituted according to manufacturer's guidelines. Stock antibiotic solutions were made by making 20 ml of 10 mg/ml stock in sterile water (0.2 g in 20 ml). A 1/10 dilution of the 10 mg/ml stock (1 ml / 9 ml sterile water) was then made. Ten antibiotic concentrations were made in a universal of 25 ml BHI broth; 256 µg/ml, 128 µg/ml, 64 µg/ml, 32 µg/ml, 16 µg/ml, 8 µg/ml, 4 µg/ml, 2 µg/ml, 1 µg/ml and negative antibiotic control (broth only). Using a 24-well plate, 990 µl of the antibiotic dilutions were aliquoted to appropriate wells and incubated anaerobically at 37 °C for 3 to 4 hours (including a well for the broth without antibiotic as a negative organism control). The 10 ml broths in the cabinet were inoculated with three colonies and incubated anaerobically (including an un-inoculated broth as a negative control) on a shaking platform (Orbital Shaker SSL1, Stuart^{TM} Staffordshire, UK) at 60 rpm for 3 to 4 hours. Once an OD_{600} = 0.3 was reached, 10 µl of the 10 ml broth was aliquoted to the appropriate wells (including the un-inoculated broth). The plate was incubated anaerobically at 37 °C on a shaking platform at 60 rpm for 16 hours. Following this, the OD_{600} of each well were measured in a micro titre reader (Gen 5, BioTek, Vermont, USA) to ascertain the MIC.

The following antimicrobials were tested for the range 1 to 256 µg/ml 1) chapter 4: metronidazole, vancomycin, erythromycin, lincomycin, fuscidic acid, naladixic acid,

rifampicin and rifaximin and 2) chapter 5: chloramphenicol, rifampicin, tetracycline, erythromycin, naladixic acid, gentamicin, teicoplanin and ampicillin. Breakpoints described by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) and the Clinical and Laboratory Standards Institute (CLSI) were used to determine MICs (The European Committee on Antimicrobial Susceptibility Testing., 2017, Clinical and Laboratory Standards Institute., 2013, Clinical and Laboratory Standards Institute., 2012). A well of broth with no organism or antibiotic was used as a negative control and the *C. difficile* PCR ribotype 017 reference strain M68 was used as a positive control (Section 2.1.2, Page 69).

2.2.8 Vegetative cell and spore count assays

Cultures were made by inoculating 10 ml pre-equilibrated BHIS broths with three colonies of 48 hours blood agar culture. These were incubated anaerobically at 37 °C on a shaking platform (Orbital Shaker SSL1, Stuart™ Staffordshire, UK) at 60 rpm for 24 hours. New cultures were made by inoculating fresh, pre-equilibrated SM broths with a 1/20 dilution of the initial 24 hour culture. The ODs were measured until each new culture reached OD_{600} = 0.5. Fresh SM broths were inoculated with 1/20 of the new culture and incubated at 37 °C anaerobically on a shaking platform at 60 rpm. After 24, 72 and 144 hours, for vegetative cell counts; 500 µl of culture were aliquoted to 1.5 ml microcentrifuge tubes, serially diluted in sterile 1 x phosphate buffered saline (PBS) and plated onto BA_ST. For spore counts, 500 µl of culture were aliquoted to 1.5 ml microcentrifuge tubes, incubated at 65 °C for 30 minutes to heat inactivate the vegetative cells. The aliquots were serially diluted in sterile 1 x sterile PBS and plated onto BA_ST. All plates were incubated at 37 °C anaerobically. After 24 hours, colony counts were enumerated on plates and calculations performed to give colony forming units (cfu)/ml.

2.2.9 Disinfectant assays

Cultures of pre-equilibrated, 10 ml, SM broths plus 0.2 % (vol/vol) Tween 80 were inoculated with three colonies of 48 hours blood agar culture. These were incubated at 37 °C anaerobically on a shaking platform at 60 rpm for 16 hours. New cultures were made by inoculating fresh, pre-equilibrated SM broths plus 0.2 % (vol/vol) Tween 80 with 1/20 of the initial 16 hour culture and incubated at 37 °C anaerobically on a shaking platform at 60 rpm. After 24 hours, for each new culture, two 1 ml aliquots were removed to 1.5 ml microcentrifuge tubes; one was treated with 1 ml Achtichlor Plus disinfectant (EcoLab, UK) for 30 minutes and the other with 1 ml 1 x sterile PBS for 30 minutes. Each aliquot was centrifuged at 8,000 x *g* for 5 minutes, washed with 1 ml 1 x sterile PBS and centrifuged at 8,000 x g for 5 minutes. The supernatant was removed and the pellet re-suspended in 1 ml sterile 1 x PBS. Each aliquot was serially diluted 1/10 in sterile 1 x PBS and 10 µl were inoculated onto BA_SD and incubated at 37 °C anaerobically. After 24 hours, colony counts were enumerated on plates and calculations performed to give cfu/ml.

2.2.10 Phenotypic data analysis

All statistical analyses of phenotypic data were carried out in GraphPad Prism software (California, USA), using One-Way and Two-Way analysis of variance (ANOVA) for individual comparisons and Tukey's for multiple comparisons. A p value of less than 0.05 was reported as statistically significant (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

2.3 Molecular methods

2.3.1 PCR ribotyping

For DNA extraction, biomass from 48 hour blood agar growth were swept and re-suspended into a 1.5 ml microcentrifuge tube containing 5 % (wt/vol) solution of Chelex-100 resin (BIO-RAD, USA) and heated to 100 ºC for 20 minutes. An un-inoculated aliquot of Chelex-100 resin solution was also used as a negative extraction control. The suspension was separated by centrifugation at 13,000 x *g* for 12 minutes and the supernatant aliquoted to a fresh 1.5 ml microcentrifuge tube.

For PCR, 5 µl of the DNA extract was added to a 45 µl PCR mixture containing 25 mM of each primer (Table 2.2), 2.5 U HotStar Taq DNA polymerase (Qiagen, UK), 0.4 mM dNTPs (Fisher Scientific, UK) and 3.75 mM $MgCl_2$ (Qiagen, UK) per reaction.

Table 2.2: Primers used for PCR ribotyping

Locus	Primer Name	Primer Sequence (5'-3')	Amplicon Size
16S gene	P3	CTGGGGTGAAGTCGTAACAAGG	564 bps
23S gene	P5	GCGCCCTTTGTAGCTTGACC	

Purchased from Sigma Aldrich, UK.

The reaction mixture was subjected to a PCR using a GeneAmp 9600 thermal cycler (Perkin-Elmer, Beaconsfield, UK). The PCR thermocycling conditions were 30 cycles at 95 ºC for 1 minute, 92 ºC for 1 minute, 55 ºC for 1 minute and 72 ºC for 1.5 minutes. This was followed by hold steps of 95 ºC for 1 minute, 55 ºC for 45 seconds

and 72 ºC for 5 minutes. The PCR ribotyping products were concentrated to 20 µl by heating at 75 ºC for 40 minutes. The PCR ribotyping products were separated by agarose gel electrophoresis using Ready Agarose precast 0.5 % Tris-acetate-Ethylenediaminetetraacetic acid (TAE) agarose gels containing ethidium bromide (BIO-RAD, USA). This was performed in 3 % TAE buffer (BIO-RAD, USA) at 100 mA for 3.5 hours alongside a 100 bp ladder (Invitrogen, UK). Banding patterns were analysed using GelCompar software (Applied Maths, Belgium).

2.3.2 Multilocus sequence typing

Genomic DNA (gDNA) was extracted either by cell lysis, phenol chloroform and ethanol precipitation or Tris-EDTA (TE) boilate methodology.

Phenol chloroform methodology

For cell lysis, overnight growth in BHIS was centrifuged at 4,000 x *g* for 5-10 minutes at 4 ºC. The supernatant was discarded and the pellet lysed by resuspending in 3 ml ETDA. To make a lysate, 750 µl lysozyme (20 mg/ml 100 µl mutanolysin (10 KU/ml), 100 µl lysostaphin (5 mg/ml) and 20 µl RNase (100 mg/ml [Invitrogen, UK]) were added to the cell suspension and the lysate was incubated at 37 ºC for 1 hour. This was followed by adding 90 µl Proteinase K (25 mg/ml) and 90 µl of 20 % Sodium Dodecyl Sulphate (SDS) to the lysate and this was incubated at 50 ºC for 1 hour.

Genomic DNA (gDNA) was extracted from the lysate by performing phenol:chloroform:isoamyl alcohol (25:24:1) and chloroform (24:1) washes in 15 ml phase lock gel light (PLG) tubes (5 PRIME, Scientific Laboratory Supplies, UK). All

PLG tubes were centrifuged prior to use at 4,000 x *g* for 2 minutes to prepare the gel in the tube as advised in the manufacturer's instructions.

The lysate was mixed with 3 ml of phenol:chloroform:Isoamyl Alcohol (25:24:1) in a PLG tube and vortexed for 10 seconds. The PLG was centrifuged at 4,000 x *g* for 5 minutes at 4 ºC to separate the phases and the upper-phase supernatant from above the gel was transferred to a fresh PLG tube with 3 ml of phenol:chloroform:Isoamyl Alcohol (25:24:1) and the above step was repeated. The upper-phase lysate was transferred to a fresh PLG tube with 3 ml of chloroform, vortexed for 10 seconds and centrifuged at 4,000 x *g* for 5 minutes at 4 ºC and the above step repeated. The upper-phase supernatant was transferred to a 15 ml solvent safe tube containing 8 ml of chilled 100 % ethanol and incubated overnight at -20 °C.

The gDNA from the phenol:chloroform:Isoamyl Alcohol (25:24:1) extraction was precipitated by centrifuged at 4,000 x *g* for 15 minutes at 4 ºC. The supernatant was discarded and the gDNA purified by adding 1 ml of 70 % ethanol and centrifuged at 4,000 x *g* for 15 minutes at 4 ºC. The supernatant was again discarded and the pellet was washed by adding another 1 ml of 70 % ethanol and centrifuged at 4,000 x *g* for 15 minutes at 4 ºC. The supernatant was discarded and the solvent safe tube was pulse-centrifuged and left to air-dry for 5 minutes. The DNA pellet was resuspended in 50-100 µl 1x TE buffer and left overnight at 4 °C.

Tris-EDTA (TE) boilate

Biomass from 48 hour blood agar culture were swept and re-suspended into a 1.5 ml microcentrifuge tube containing 1.5 ml TE and heated at 100 °C for 10 minutes. The

suspension was separated by centrifugation at 13,000 x *g* for 2 minutes and the supernatant aliquoted to a fresh 1.5 ml microcentrifuge tube.

The quality of the gDNA was confirmed using agarose gel electrophoresis; products were visualised on a 1 % (wt/vol) TAE agarose gel resolved at 100 mV for 1 hour and stained with 1.5 µl of 0.5 µg/µl ethidium bromide solution (Promega, UK) in TAE buffer alongside a 100 bp ladder (Invitrogen, UK).

MLST was performed using the scheme described by Griffiths *et al.,* (Griffiths et al., 2010). The primers used to detect seven housekeeping genes (*adk, atpA, dxr, glyA, recA, sodA* and *tpi)* are shown in Table 2.3.

For each of the seven MLST loci, 50 µl PCR reactions were performed in 96-well plates. The PCR mixture contained 39.75 µl of molecular grade water, 5 µl of 10 x PCR buffer (Qiagen, UK), 1 µl of a 10 µM concentration of each forward and reverse primer, 1 µl of 10 mM deoxynucleoside triphosphate (dNTP) mix (Invitrogen, UK), 0.25 µl of HotStart Taq DNA polymerase (Qiagen, UK) and 2 µl of *C. difficile* gDNA. The MLST PCR was performed using a GeneAmp 9600 thermocycler (Perkin-Elmer, UK) and thermocycling conditions were; 35 cycles at 94 ºC for 15 seconds, 50 ºC for 1 minute, 72 ºC for 1 minute. This was followed by a hold step of 72 ºC for 7 minutes.

Table 2.3: Primers used for MLST

Locus	Primer Name	Primer Sequence (5'-3')	Amplicon Size
adk	adk1F	TTACTTGGACCTCCAGGTGC	635 bps
	adk1R	TTTCCACTTCCTAAGGCTGC	
atpA	atpA1F	TGATGATTTAAGTAAACAAGCTG	674 bps
	atpA1R	AATCATGAGTGAAGTCTTCTCC	
dxr	dxr3F	GCTACTTTCCATTCTATCTG	525 bps
	dxr3R	CCAACTCTTTGTGCTATAAA	
glyA	glyA1F	ATAGCTGATGAGGTTGGAGC	625 bps
	glyA1R	TTCTAGCCTTAGATTCTTCATC	
recA	recA2F	CAGTAATGAAATTGGGAGAAGC	705 bps
	recA2R	ATTCAGCTTGCTTAAATGGTG	
sodA	sodA5F	CCAGTTGTCAATGTATTCATTTC	585 bps
	sodA5R	ATAACTTCATTTGCTTTTACACC	
tpi	tpi2F	ATGAGAAAACCTATAATTGCAG	640 bps
	tpi2R	TTGAAGGTTTAACACTTCCACC	

Purchased from Invitrogen, UK.

MLST PCR products were confirmed using agarose gel electrophoresis; these were visualised on a 1 % (wt/vol) agarose gel resolved at 100 mV for 1 hour and stained with 1.5 µl of 0.5 µg/µl ethidium bromide solution (Promega, UK).

Once confirmed, the PCR products were purified by precipitation with 20 % polyethylene glycol (molecular weight, 8,000) and 2.5 M NaCl. The nucleotide sequences were determined for each DNA strand using the amplification primers in Table 2.3 and BigDye Ready Reaction Mix (Applied Biosystems, UK) as follows: each 10 µl sequencing reaction mixture comprised 2 µl of PCR amplicon, 4 µl of a 1:15 dilution of either forward or reverse PCR primer (0.66 µM), 0.25 µl of BigDye Ready Reaction Mix, 1.875 µl of 5 x sequencing buffer (20 ml of stock solution

comprised 200 µl of 1 M $MgCl_2$, 8 ml of 1 M Tris-HCl, pH 9 and 11.8 ml of molecular grade water and 1.875 µl of 5 x sequencing buffer (20 ml of stock solution comprised 200 µl of 1 M $MgCl_2$, 8 ml of 1 M Tris-HCl, pH 9, and 11.8 ml of molecular biology-grade water) and 1.875 µl of molecular grade water.

The reaction conditions were 30 cycles of 96 ºC for 10 seconds, 50 ºC for 5 seconds and 60 ºC for 2 minutes. Unincorporated dye terminators were removed by precipitation of the termination products with two volumes of ethanol and 0.1 volume of sodium acetate (pH 5.2 [3M]) followed by vortexing and incubation at room temperature for 1 hour. The plate was centrifuged at 2,750 x *g* for 10 minutes, the resulting pellet was washed with 150 µl of 70 % ethanol and the plate was centrifuged at 2,750 x *g* for 10 minutes. The supernatant was carefully pipetted away and the pellet was allowed to air dry for 5 minutes. The pellets were resuspended with 10 µl Hi-Di Formamide (Thermo Scientific, USA).

The precipitated PCR products were analysed on a ABI 3730 DNA analyser (Applied Biosystems, USA) and nucleotide sequences were extracted using Chromas v1.61 (Queensland, Australia). Allele designations were obtained by parsing forward and reverse sequencing reads through the *C. difficile* pubMLST batch profile query page which contains a database of all *C. difficile* MLST profiles (http://pubmlst.org/cdifficile/). Novel sequence types (n=18) and existing *C. difficile* pubMLST database sequence types (n=123) [82 previously published and 41 unpublished and used with submitters permission]) were analysed using Multiple Alignment using Fast Fourier Transform (MAFFT) (http://mafft.cbrc.jp/alignment/software/) and Archeopteryx

(www.phylosoft.org/archeopteryx). Phylogenies were calculated by MAFFT and MrBayes (Ronquist and Huelsenbeck, 2003, Huelsenbeck and Ronquist, 2001). MAFFT used the neighbour-joining method (Saitou and Nei, 1987) using all ungapped sites with 1000 boot strap resamplings.

2.3.3 Whole genome sequencing

2.3.3.1 Genomic DNA extraction

The gDNA extraction performed for WGS was a modification of the phenol:chloroform:Isoamyl Alcohol (25:24:1) methodology described in section 2.3.2 for MLST. Ten millilitre BHIS broths were inoculated with three colonies from 48 hour blood agar culture. An un-inoculated broth was also used as a negative extraction control. Broths were incubated anaerobically for 18 hours. The biomass was harvested by centrifugation of the broths at 4,000 x *g* for 10 minutes and the pellet was re-suspended in 10 ml sterile 1 x PBS. Centrifugation was repeated and the pellet was stored overnight at -20 ºC. The pellet was re-suspended in 800 µl fresh 25 % sucrose in TE buffer and 100 µl of 20 mg/ml Lysozyme; this suspension was incubated at 37 ºC for 45 minutes. Forty microliters of 20 mg/ml Proteinase K, 1.2 µl of 100 mg/ml RNase A, 160 µl of 0.5 M EDTA pH8.0 (Na salt) and 100 µl of fresh 10 % Sarkosyl (10 mg/ml) were added to the suspension which was left on ice for 1 to 2 hours and then incubated in a 50 ºC water bath overnight. The lysate was made to 2 ml with TE buffer. In a fume hood, the 2 ml lysate was mixed with 2.5 ml phenol:chloroform:Isoamyl Alcohol (25:24:1) in a PLG tube and vortexed for 10 seconds. The PLG tube was centrifuged at 4,000 x *g* for 15 minutes at 4 °C to separate the phases and the upper-phase supernatant from above the gel was

transferred a fresh PLG tube with 2.5 ml phenol:chloroform:Isoamyl Alcohol (25:24:1) and the above step repeated but with centrifugation for 10 minutes.

The upper-phase supernatant was transferred to a fresh PLG tube containing equal volume of chloroform, vortexed for 10 seconds and centrifuged at 4,000 rpm for 5 minutes. This step was repeated and the upper-phase supernatant was transferred to a 15 ml solvent safe tube containing 2.5 volume of chilled ethanol and incubated at -20 ºC for 15 minutes to precipitate the DNA. The precipitate was centrifuged at 4,000 x *g* for 10 minutes, supernatant gently pipetted away and the pellet washed with 1 ml chilled 70 % ethanol and centrifuged at 4,000 x *g* for 15 minutes again. The supernatant was gently pipetted away and the solvent safe tube allowed to air dry for 5 minutes. The pellets were re-suspended in 120-300 µl of 1 x TE buffer and allowed to re-suspend at 4 ºC overnight. gDNA was confirmed using agarose gel electrophoresis; the presence of a thick, smeary high molecular weight band was visualised on 0.5 % (w/v) agarose gel resolved at 110 V for 50 minutes and stained with 1.5 µl of 0.5 µg/µl ethidium bromide solution.

2.3.3.2 DNA quantification using the Qubit Fluorometer System

To obtain an accurate quantification of the gDNA, the Qubit Fluorometer System (Invitrogen, UK) was used according to manufacturer's instructions. gDNA samples were quantified using the Qubit dsDNA broad range and high sensitivity assay kits (Invitrogen, UK) as required. Two standards were used for each batch of samples tested and for each DNA sample or standard, a total volume of 200 µl was prepared by diluting the sample or provided standard in Qubit working solution. Qubit working solution was made for each standard and gDNA sample by diluting the Qubit reagent 1:200 in Qubit buffer. The assay tubes were then prepared by adding

10 µl of standard or 2 µl gDNA sample 190 µl or 198 µl of working solution respectively. These were vortexed for 2-3 seconds and incubated at room temperature for 2 minutes. Assay tubes were loaded onto the Qubit Fluorometer, their absorbance was read and gDNA concentration in the original suspension was calculated.

2.3.3.3 Library construction and sequencing

Library construction and sequencing was performed using an Illumina HiSeq 2000 Sequencing System (Illumina, California, USA) by the Pathogen Genomics Sequencing Team at the Wellcome Trust Sanger Institute (Cambridge, UK) and by M. Cairns using a Illumina MiSeq Sequencing System (Illumina, California, USA) at the London School of Hygiene and Tropical Medicine (LSHTM). To ensure the data from both platforms and different runs was consistent, the control strain M68 was processed with each and every sequencing run. Libraries were created as previously described (Harris et al., 2010) or using the Nextera XT kit (Illumina, California, USA) according to the manufacturer's instructions. The Nextera XT DNA library preparation consisted of five main stages. The main stages are described and illustrated in Figures 2.1 and 2.2. In order to check the final size and concentration of DNA libraries made after clean-up, the libraries were analysed using the Agilent 2100 Bioanalyzer (Agilent Technologies, UK [Section 2.3.3.4, Page 88]).

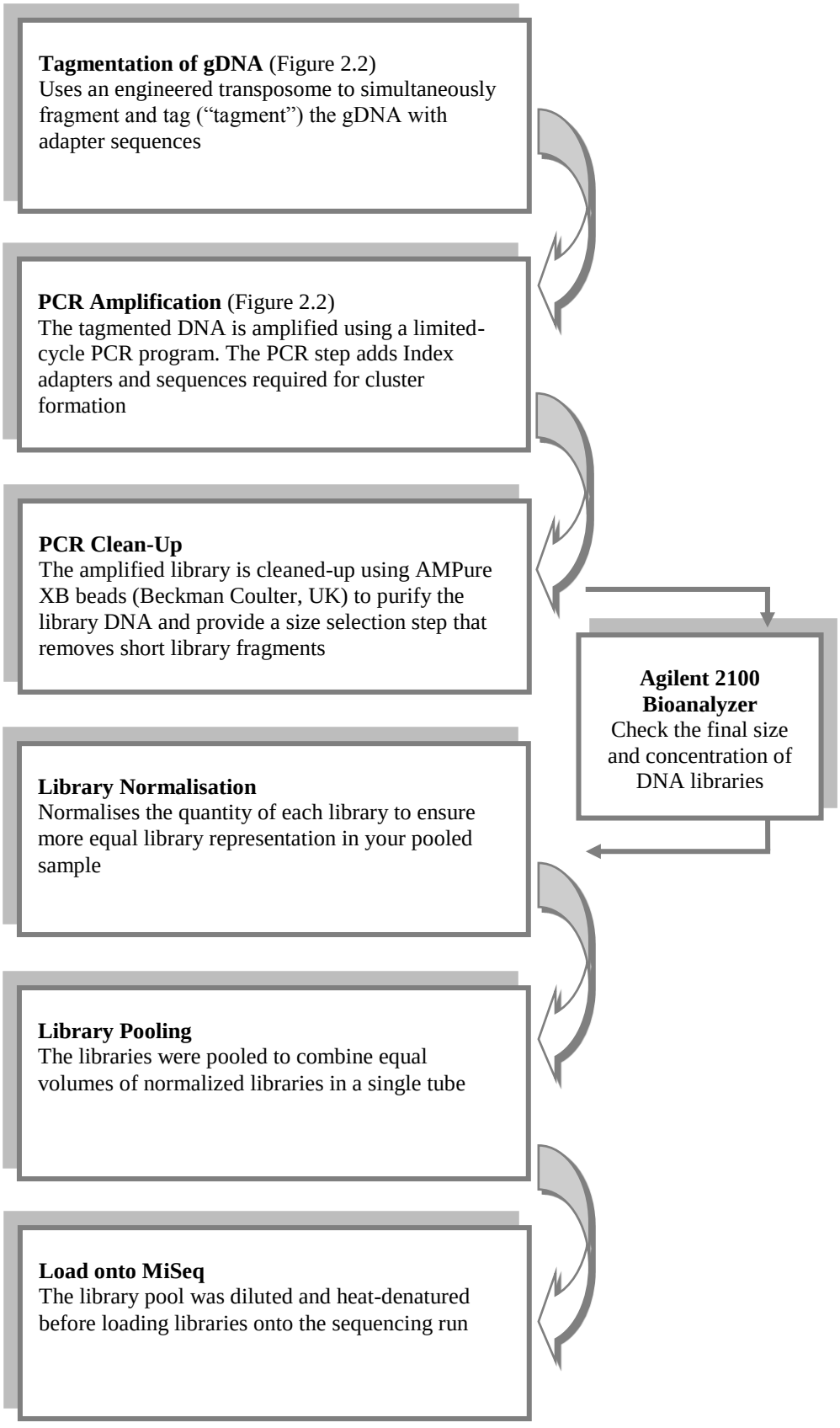
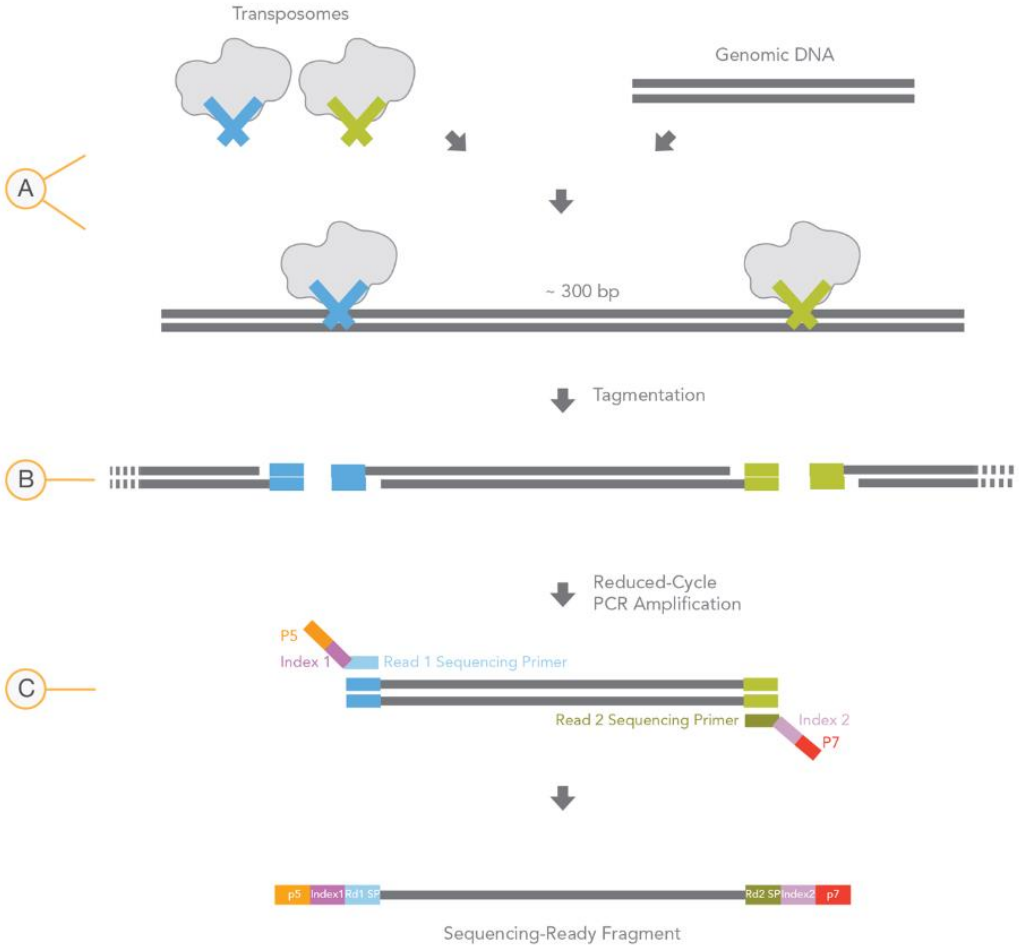


Figure 2.1: Flowchart of the Nextera XT assay for WGS

Courtesy of illumina (IIIumina ®. 2016).



A = Nextera XT transposome with adapters combined with template DNA
B = Tagmentation to fragment and add adapters
C = Limited cycle PCR to add sequencing primer sequences and indices

Figure 2.2: Illustration of the Nextera XT assay tagmentation and PCR amplification steps

A figure that provides an illustration of the Nextera XT assay tagmentation and PCR amplification steps for WGS. Courtesy of illumina (IIIumina ®. 2016).

2.3.3.4 Validation of libraries using the Agilent Bioanalyzer 2100

The Agilent 2100 Bioanalyzer (Agilent Technologies, UK) and Agilent High Sensitivity Kit (Agilent Technologies, UK) were used according to the manufacturer's instructions. Briefly, gel-dye mix was prepared by equilibrating DNA dye and DNA gel matrix to room temperature for 30 minutes, vortexing for 10 seconds and pulse centrifugation. Twenty-five microliters of DNA dye concentrate was added to the DNA gel matrix vial and vortexed for 10 seconds. The solution was applied to a fresh spin filter and centrifuged at 2,240 x g for 15 minutes. Following centrifugation, the spin filter was discarded and 9 µl of gel-dye mix was added to the appropriate well of a DNA chip which was placed on the chip priming station. The syringe was set to 1 ml and then pressure was slowly applied to the syringe to distribute the gel around the chip. Following priming, DNA marker, DNA ladder and sample/s were added to the appropriate wells. The chip was vortexed for 60 seconds at 2,400 rpm, loaded onto and analysed using the Bioanalyzer. This illustrated traces of successfully sequenced libraries whereby a typical library would show a broad size distribution of between ~250 and 1000 bps. Libraries were then normalised, pooled and loaded onto the MiSeq according to the manufacturer's instructions (Figure 2.1).

2.3.3.5 Whole genome sequence data mapping and assembly

A variety of freely available software packages were used to process the sequence data, these are either referenced or a Uniform Resource Locator (URL) is provided for reference (Table 2.4).

Table 2.4: Software packages used in this thesis for processing WGS data

Software	Function	URL or Reference
Trimmomatic	Trims and crops sequences	http://www.usadellab.org/cms/index.php?page=trimmomatic
BWA MEM	Maps reads	(Li et al., 2009)
SAMtools	Reading/writing/editing/indexing/viewing SAM/BAM/CRAM format	http://www.htslib.org/
VCFtools	Identifies SNPs	http://vcftools.sourceforge.net/downloads.html
Velvet	Assembles short reads	(Zerbino, 2010)
Velvet Optimiser	Multi-threaded Perl script	http://bioinformatics.net.au/software.velvetoptimiser.shtml
BLAST	Search for sequence and amino acid similarity	http://www.ncbi.nlm.nih.gov
MAFFT	Multiple sequence alignment	http://mafft.cbrc.jp/alignment/software/
Perl	Programming language	https://www.perl.org/get.html
R	Statistical computing and graphics	R development core team, https://www.r-project.org/
ABACAS	Contiguation of assembled sequences	(Assefa et al., 2009)
Prokka	Annotation of prokaryotic genomes	(Seemann, 2014)
RaXML	Maximum-likelihood based phylogenetic inference	(Stamatakis, 2014)
BEAST	Bayesian analysis of molecular sequences	(Drummond et al., 2012)
mclust	Cluster analysis	(Fraley et al., 2012)
Artemis	Displays sequence features and results of analysis	http://www.sanger.ac.uk/science/tools/
ACT	Displays pairwise comparisons between two or more sequences	http://www.sanger.ac.uk/science/tools/

ABACAS - Algorithm Based Automatic Contiguation of Assembled Sequences
ACT - Artemis Comparison Tool
BEAST - Bayesian Evolutionary Analysis Sampling Tree
BLAST - Basic Local Alignment Search Tool
BWA MEM - Burrow-Wheeler Aligner
MAFFT - Multiple Alignment using Fast Fourier Transform
mclust - Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation
RaXML - Randomised Accelerated Maximum Likelihood

The sequence data were processed and quality controlled according to a standard pipeline as previously described (Preston et al., 2014). Briefly, FASTQ-formatted sequencing reads were quality controlled with a minimum-quality Phred score of 30 (as a rolling average over 4 bases) using Trimmomatic; a fast, multithreaded command line tool that can trim and crop FASTQ data. The resulting reads were mapped, using BWA-MEM, a software package for mapping sequences against a reference genome, in this case, against the control strain M68 reference genome. SNP mutations in the samples that had read depths of 60 and had 70 % of reads identified with the same allele (99.8 % of SNPs were supported by ~90 % of contributing reads) and a minimum quality score of 30 were identified using Sequence Alignment Map (SAM) tools and VCFtools software.

The methodology used in this thesis for SNP detection applies a core-genome methodology; the core-genome refers to genes common to all strains in a species. SNPs in 'core genes' were determined by aligning sequences against the reference control genome; GenBank accession number FN668375 (He et al., 2010). If there were regions of sequence/s present in the genome of the 'test' strain/s but absent in the reference control genome, then these regions were not analysed for SNPs. The methodology used in this thesis for SNP detection also does not read regions of repetitive DNA (sequence patterns that occur in multiple copies throughout the genome). Therefore, any SNPs in repetitive regions are not detected.

Velvet and Velvet Optimiser were used to *de novo* assemble the trimmed reads into contiguous pieces of DNA sequence. A multi-threaded Perl script was used for the

optimisation of Velvet. Perl is a programming language that facilitates analysis, manipulation and graphical presentation of experimental data.

2.3.3.6 BLAST

Searches for sequence and amino acid similarity were routinely performed using BLAST, specifically the blastn and blastp algorithms.

2.3.3.7 Sequence alignment

Sequence alignments were performed using the multiple sequence alignment program 'Multiple alignment program for amino acid of nucleotide sequences' (MAFFT).

2.3.3.8 Nucleotide sequence accession numbers

All Illumina sequence data from this thesis were submitted to the European Nucleotide Archive under accession numbers ERP009770 and PRJEB11868.

2.3.3.9 Phylogenetic analysis

Pipeline, post-analyses, genetic, phylogenetic and phylogeographic analysis were carried out using Perl and R. Algorithm Based Automatic Contiguation of Assembled Sequences (ABACAS) software was used to contiguate assembled contigs, and based on the control strain M68 reference sequence, Prokka, a software tool was used for rapid genome annotation.

Maximum-likelihood phylogenetic trees were generated using Randomised Accelerated Maximum Likelihood (RaXML) and BEAST software to produce a SNP phylogeny from the SNPs as well as geographical and temporal data combined in phylogeographic analysis and mclust software for maximum likelihood cluster

analysis. A minor allele frequency (MAF) of less than 1 % were used to remove any SNPs that may be associated with recombination and which would mask the true phylogeny, and SNPs within 1 bp distance of an insertion or deletion (absent from the reference strain M68) site were excluded from further analysis.

2.3.3.10 Identification of genomic deletions and insertions

To investigate genomes for indication of horizontal gene transfer, a key mechanism driving *C. difficile* evolution, programmatic and visual inspection of the comparisons was performed using the Artemis Comparison Tool (ACT), a JAVA application that displays pairwise comparisons between two or more DNA sequences which can be used to analyse regions of similarity and differences between genomes. Artemis software was also used to browse genomes and visualise sequence features. Insertions were regions of DNA that were present in a test strain but absent from the reference control M68 whereas deletions were regions of DNA that were present in the control strain M68 but absent from a test strain.

Chapter 3

Characterisation of *C. difficile* strains isolated from diverse sources and geographical locations using MLST and PCR ribotyping

3 Characterisation of *C. difficile* strains isolated from diverse sources and geographical locations using MLST and PCR ribotyping

3.1 Statement of contribution

Genomic DNA extraction and PCR ribotyping was performed by M. Cairns, MLST PCR assays and MLST designations were performed by Professor Wren's Laboratory Research Group at the LSHTM collectively (five individuals) including M. Cairns (20%) and collation of data was performed by M. Cairns.

3.2 Introduction

PCR ribotyping, PFGE, REA and MLVA are valuable molecular tools for outbreak investigation where rapid genotyping of *C. difficile* strains is necessary for infection control purposes. However, the genome targets used in these methods are too discriminatory to resolve the phylogenetics that would facilitate tracing the origins of *C. difficile*. For example, with PCR ribotyping, recombination of repeats present in the intergenic spacer region between the 16S and 23S rRNA might lead to the formation of a novel PCR ribotype without a clear phylogenetic link to the PCR ribotype prior to recombination. However, recombination in a MLST housekeeping gene target would change the allelic profile on a single locus only; a novel sequence type would still be closely related to the original sequence type maintaining the phylogenetic link making MLST useful for phylogenetic studies.

Few studies have investigated *C. difficile* using MLST. Lemee, *et al.*, developed the first MLST scheme for *C. difficile* and investigated a collection of 72 isolates from various hosts, geographic sources and toxigenic types (Lemee et al., 2004). They demonstrated that *C. difficile* had a predominantly clonal population structure consisting of stable subpopulations that are globally disseminated. They found no correlation between sequence type and geographic origin nor were they able to characterise any host specificity but they did find toxin A-B+ isolates shared the same sequence type. The Lemee group further utilised MLST and examined a collection of 29 isolates selected as representative of the main clusters defined by MLST in their previous study (Lemee et al., 2005) and confirmed the finding from their previous study that toxin A-B+ strains form an individual lineage. Another MLST investigation by Griffiths, *et al.*, confirmed toxin A-B+ strains to form an individual lineage and additionally identified a further four different phylogenetic lineages of *C. difficile* that contained common PCR ribotypes; 027, 078, 023 and a lineage containing multiple mixed PCR ribotypes (Griffiths et al., 2010). A larger study tested a collection of 1290 isolates from humans with CDI between September 2006 and December 2009 (both hospital and community) and confirmed the clonal population structure and presence of five phylogenetic lineages (Dingle et al., 2011a) depicted in Figure 1.6, Page 55.

During this time, comparative phylogeny using whole genome microarray analysis also identified four clonal lineages made up of PCR ribotypes; 017, 027 and 078 and a heterogeneous grouping of mixed PCR ribotypes (Stabler et al., 2006). A WGS study based on SNPs in conserved core genes also confirmed the existence of these four phylogenetic lineages (He et al., 2010) and all three studies found PCR ribotype

078 to be highly divergent from all other PCR ribotypes (Griffiths et al., 2010, Stabler et al., 2006, He et al., 2010). MLST, whole genome microarray analysis and WGS SNP analysis, although different molecular approaches, concur that there are at least four phylogenetic lineages of *C. difficile*.

PCR ribotype 078 which forms a highly divergent lineage through MLST analysis has been associated with strains isolated from non-human sources including animals and food for human consumption (Jhung et al., 2008, Gould and Limbago, 2010, Songer et al., 2009). However, only few studies have shown that strains of PCR ribotype 078 isolated from human and animal sources are similar in genotype using typing methodologies other than PCR ribotyping; strains from humans and animals have been found related by MLVA (Goorhuis et al., 2008) and MLST (Debast et al., 2009). There are no reports detailing the relatedness of strains of PCR ribotype 078 from food sources using typing methodologies other than PCR ribotyping.

Outbreak strains from the UK, USA and Canada were found to be due to the dissemination of a strain of PCR ribotype 027, sequence type 1 (Killgore et al., 2008) and a global collection of PCR ribotype 027 strains were found to be clonal (He et al., 2013). However, little is known about the phylogeny of other PCR ribotypes and sequence types when comparing strains from different geographical locations.

Inter-laboratory comparisons between sequence types, PCR ribotypes, NAP types and REA types would be useful however the adequacy of their correlation is unknown. Recent to our study Tenover *et al.*, showed that that there is poor correlation, with only 84/92 PCR ribotype 027 isolates were NAP1 by PFGE and BI

by REA (Tenover et al., 2011). MLST studies have found correlation between individual lineages, PCR ribotypes and sequence types (i.e. lineage 2, PCR ribotype 027 and sequence type 1). However, similarities in PCR ribotyping banding patterns have been reported between PCR ribotypes 126 and 078 (Spigaglia et al., 2010) and 078 with 033, 066 and 045 (Rupnik et al., 2001). It is unknown if these PCR ribotypes are similar in their ancestry and share the same MLST sequence type which would suggest the occurrence of micro-diversity in the individual lineages.

3.3 Hypotheses of the research described in this chapter

Prior to this investigation, MLST studies on *C. difficile* had focused on isolates from either; various hosts and geographic sources, or hospital and community and various PCR ribotypes. No MLST study had yet tested a collection of isolates encompassing all of these variables nor had a study compared the *C. difficile* sequence types of strains isolated between human, animal and food or varying geographical origins. By testing a larger and more varied collection of strains would provide robust evidence to support the number of phylogenetic lineages of *C. difficile* previously identified. By performing MLST on strains of *C. difficile* isolated from animals and relating to strains isolated from humans and food sources will enable more in-depth associations to be made and may concur with PCR ribotyping data that has shown strains isolated from humans, animals and food sources to be related. This has significant implications with regards to transmission and zoonosis. Performing MLST and PCR ribotyping on strains isolated from various geographical locations will help identify clonal clusters associated with global spread which may infer how the species is evolving and if selective pressure has played a role. Finally, by examining the relatedness of PCR ribotypes, MLST sequence types and REA types will

demonstrate how well these typing techniques correlate which will have implications for inter-laboratory comparisons globally. This study was designed to test the following hypotheses:

Chapter 3: hypothesis 1

Application of MLST to a large collection of *C. difficile* isolates from diverse sources, geographical origin and PCR ribotype will retain the five distinct phylogenetic lineages observed in previous studies (Lemee et al., 2004, Lemee et al., 2005, Dingle et al., 2011a).

Chapter 3: hypothesis 2

C. difficile strains isolated from human, animal and food origin are not phylogenetically distinct by MLST.

Chapter 3: hypothesis 3

C. difficile strains isolated from different geographical origin are phylogenetically distinct by MLST.

Chapter 3: hypothesis 4

Molecular characterisation of *C. difficile* using MLST, PCR ribotyping and REA provides adequate correlation for inter-laboratory comparison.

3.4 Results

A total of 385 *C. difficile* isolates were collated from collaborators (Page 27) and breakdowns of the isolates are illustrated in the following sections with details of alleles, provider, source, geographical location and isolation date shown in Appendix 1, Page 229. All isolates were propagated on appropriate culture media (Sections 2.2.2, Page 70 and 2.2.3, Page 71) from which genomic DNA was extracted using methods described for PCR ribotyping and MLST (Sections 2.3.1, Page 77 and 2.3.2, Page 78 respectively). PCR ribotyping was performed by PCR amplification and agarose gel electrophoresis (Section 2.3.1, Page 77) and MLST was performed by PCR amplification, agarose gel electrophoresis and sequencing (Section 2.3.2, Page 78). Sequence types were constructed from this data and used to elucidate the phylogenetic relatedness of the strain panel and answer the hypotheses under analysis.

3.4.1 Analysis of *C. difficile* isolated from various sources, geographical origin using PCR ribotyping

PCR ribotyping and agarose gel electrophoresis was performed on all 385 isolates with reference strains R20291 (PCR ribotype 027) and M68 (PCR ribotype 017) used as controls. PCR analysis identified 68 different PCR ribotypes (319/385) and found 66 isolates (66/385) to be non-typeable (Table 3.1).

.

Table 3.1: Number and proportion of isolates by PCR ribotype

PCR ribotype	Number of isolates	Percentage of isolates
078	107	27.8%
027	53	13.8%
001	16	4.2%
002	16	4.2%
106	10	2.6%
126	10	2.6%
017	9	2.3%
015	7	1.8%
014	5	1.3%
023	5	1.3%
050	4	1.0%
237	4	1.0%
005	3	0.8%
010	3	0.8%
012	3	0.8%
087	3	0.8%
127	3	0.8%
262	3	0.8%
020	2	0.5%
054	2	0.5%
081	2	0.5%
094	2	0.5%
176	2	0.5%
*Other	45	11.7%
Nontypeable	66	17.1%
TOTAL	**385**	**100%**

*Forty-five isolates were of other PCR ribotypes with only one isolate; 003, 009, 011, 013, 018, 021, 022, 026, 029, 030, 031, 033, 036, 042, 046, 052, 053, 059, 062, 064, 070, 085, 097, 107, 111, 116, 118, 135, 139, 140, 186, 196, 212, 216, 239, 243, 259, 264, 268, 271, 274, 280, 281, 282 and 291.

3.4.2 Analysis of *C. difficile* isolated from various sources, geographical origin using MLST

MLST was performed on all 385 isolates with reference strains R20291 (PCR ribotype 027) and M68 (PCR ribotype 017) used as controls. MLST analysis identified 48 known sequence types (365/385) and found 18 sequence types (20/385) to be novel (Table 3.2).

Table 3.2: Number and proportion of isolates by sequence type

Sequence type	Number of isolates	Percentage of isolates
11	155	40.3%
1	61	15.8%
3	25	6.5%
8	12	3.1%
42	10	2.6%
37	9	2.3%
35	6	1.6%
2	6	1.6%
5	6	1.6%
6	5	1.3%
44	5	1.3%
10	4	1.0%
15	4	1.0%
55	4	1.0%
63	4	1.0%
48	3	0.8%
54	3	0.8%
61	3	0.8%
12	2	0.5%
13	2	0.5%
17	2	0.5%
18	2	0.5%
26	2	0.5%
33	2	0.5%
41	2	0.5%
43	2	0.5%
46	2	0.5%
67	2	0.5%
*Other	20	5.2%
**Novel	20	5.2%
TOTAL	**385**	**100%**

*Twenty isolates were of other sequence types with only one isolate; 7, 79, 716, 721, 722, 729, 732, 734, 739, 745, 749, 751, 752, 753, 756, 757, 762, 64, 66 and 86.

**Eighteen isolates were novel sequence types with one isolate; 131, 132, 133, 134, 135, 136, 137, 138, 140, 141, 142, 143, 144, 145, 146, 147, 148 and 139 with three isolates.

3.4.3 Analysis of PCR ribotype and MLST sequence type associations

To confirm known associations between sequence type and PCR ribotype profiles already reported and to identify novel associations, PCR ribotyping and MLST analysis data were collated and compared. Of those isolates with a PCR ribotype designated (319/385), 39 PCR ribotype/sequence type associations previously reported were found (Table 3.3) and 46 novel PCR ribotype/sequence type associations were revealed (Table 3.4).

Table 3.3: PCR riboype and sequence type associations previously identified

PCR ribotype	Sequence type	Number of isolates
078	11	106
027	1	53
001	3	15
002	8	10
106	42	9
017	37	8
015	44	4
023	5	4
010	15	3
012	54	3
015	10	3
005	6	2
014	2	2
020	2	2
050	18	2
054	43	2
262	3	2
003	57	1
009	3	1
013	45	1
014	49	1
018	17	1
021	56	1
022	66	1
023	22	1
026	7	1
046	35	1
050	16	1
053	63	1
064	33	1
070	55	1
081	9	1
085	39	1
097	21	1
118	42	1
139	52	1
140	26	1
186	51	1
216	33	1

Table 3.4: Novel PCR riboype and sequence type associations

PCR ribotype	Sequence type	Number of isolates
126	11	10
237	11	5
002	35	4
127	11	3
087	46	2
094	12	2
176	1	2
002	2	1
002	48	1
002	146 (novel)	1
005	131 (novel)	1
011	138 (novel)	1
014	13	1
014	132 (novel)	1
017	86	1
029	137 (novel)	1
030	48	1
031	29	1
033	11	1
036	62	1
042	6	1
050	6	1
052	136 (novel)	1
059	53	1
062	44	1
081	139 (novel)	1
087	145 (novel)	1
106	135 (novel)	1
107	139 (novel)	1
111	140 (novel)	1
116	10	1
135	41	1
196	144 (novel)	1
212	5	1
239	147 (novel)	1
243	139 (novel)	1
259	141 (novel)	1
262	143 (novel)	1
264	142 (novel)	1
268	3	1
271	6	1
274	133 (novel)	1
280	11	1
281	11	1
283	134 (novel)	1
291	148 (novel)	1

Some PCR ribotypes were associated with multiple sequence types and some sequence types were associated with multiple PCR ribotypes. These are listed in Tables 3.5 and 3.6 respectively.

Table 3.5: PCR riboype and associated sequence type/s

PCR ribotype	Associated sequence types
001	3
002	2, 8, 35, 48, **146**
003	12, 57
005	6, 72, 89, 90, **131**
009	3
010	15
011	36, 77, **138**
012	54
013	45, 71, 78
014	2, 13, 14, 49, 50, **132**
015	10, 44
017	37, 86
018	17
019	67
020	2, 28, 68
021	56, 70
022	66
023	5, 22, 25
026	7
027	1
029	**137**
030	48
031	29
033	11
035	40
036	1, 62
038	48
039	26
042	6
046	35
050	6, 16, 18
052	**136**
053	63
054	43
056	34, 58
059	53
060	38
062	44, 75
063	5
064	33
066	11

PCR ribotype	Associated sequence types
067	27
070	15, 55
072	3, 69
076	2
078	11
081	9, **139**
085	39
087	46, **145**
094	12
097	21
103	53, 73, 76
104	48
106	41, 42, **135**
107	**139**
110	10
111	**140**
115	3
116	10
118	42
126	11
127	11
129	13
135	41
137	4
138	23
139	52
140	26
153	32
159	8
174	42
176	1
186	51
191	46
194	41
196	**144**
202	20, 24
212	5
216	33
220	2
224	65
225	12
228	92
237	11
239	**147**
243	**139**
259	51, **141**
262	3, **143**
264	**142**
268	3
271	6
274	**133**
280	11
281	11

PCR ribotype	Associated sequence types
283	**134**
291	**148**
305	3
316	59
319	74
320	46
321	41
323	31
326	91
336	60

Those underlined are novel PCR ribotype/sequence type associations and those in bold are novel sequence types found in our study.

Table 3.6: Sequence types and associated PCR ribotype/s

Sequence type	Associated PCR ribotypes
1	027, 176
2	002, 014, 020
3	001, 009, 262, 268
5	023, 212
6	005, 042, 050, 271
7	026
8	002
9	081
10	015, 116
11	066, 078, 126, 127, 237, 280, 281
12	094
13	014
15	010
16	050
17	018
18	050
21	097
22	023
26	140
29	031
33	064, 216
35	002, 046
37	017
39	085
41	135
42	106, 118
43	054
44	015, 062
45	013
46	087
48	002, 030
49	014
51	186
52	139
53	059
54	012
55	070
56	021
57	003
62	036
63	053
66	022
86	017
131	005
132	014
133	274
134	283
135	106
136	052
137	029
138	011

Sequence type	Associated PCR ribotypes
139	081, 107, 243
140	111
141	259
142	264
143	262
144	196
145	087
146	002
147	239
148	291

Those underlined are novel PCR ribotype/sequence type associations and those in bold are novel sequence types found in our study.

3.4.4 Phylogeny of *C. difficile* isolated from various sources, geographical origin, PCR ribotype and MLST sequence type

Previous MLST studies identified five distinct phylogenetic lineages of *C. difficile* that contained common PCR ribotypes; 027, 078, 017, 023 and a lineage containing multiple mixed PCR ribotypes (Lemee et al., 2004, Lemee et al., 2005, Griffiths et al., 2010, Dingle et al., 2011a). To determine if the five distinct phylogenetic lineages observed in previous studies would be maintained when applying MLST to a large, diverse collection of *C. difficile* isolates, phylogenetic analysis was performed using the pubMLST database (http://pubmlst.org/cdifficile/) and calculating phylogenies using MAFFT and MrBayes. Analysis comprised of; 82 previously published sequence types, 41 unpublished sequence types (used with the submitter's permission), and the 18 novel sequence types found in our study (Table 3.7). Phylogenetic analysis using the sequence types listed in Table 3.7 was used to investigate the relative evolutionary relatedness of sequence types and the resultant phylogeny is depicted in Figures 3.1 and 3.2.

Table 3.7: Sequence types used for MLST analysis

Sequence types learnt from:	Sequence Type
Sequence types already in the MLST database	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 63, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 89, 90, 91, 92, 102, 103, 104, 105 and 106
Sequence types used with submitters permission	29, 47, 61, 62, 64, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 93, 94, 95, 96, 97, 98, 99, 100, 101, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122 and 123
Sequence types novel to this study	131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147 and 148

(A)

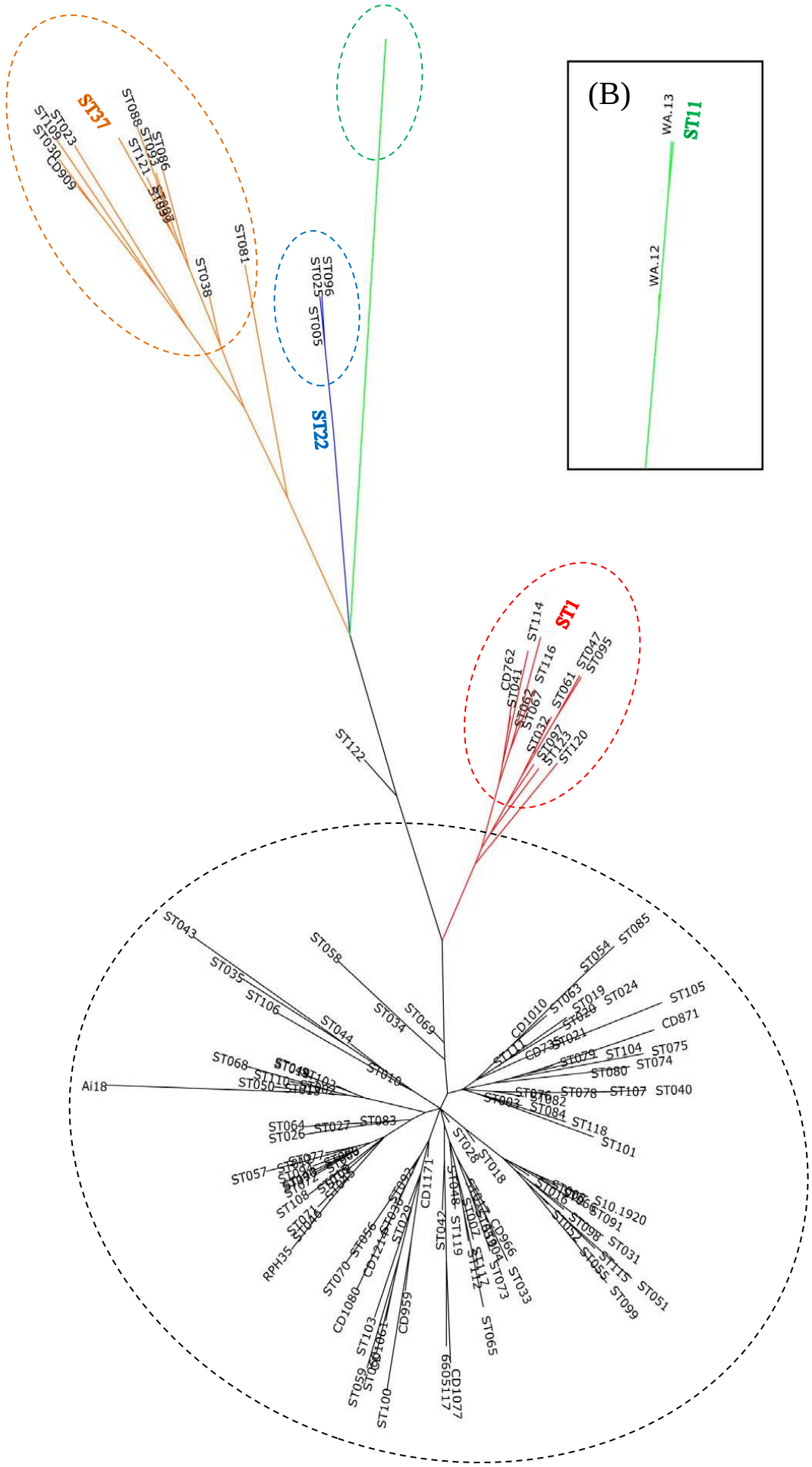

ST37
ST023
ST1082
ST030
CD909
ST088
ST093
ST086
ST121
ST089
ST038
ST081
ST096
ST025
ST005
ST22
ST122
ST1
ST114
CD762
ST041
ST116
ST062
ST067
ST061
ST047
ST095
ST032
ST097
ST123
ST120
ST043
ST035
ST106
ST058
ST034
ST069
ST044
ST068
ST049
ST103
ST110
ST050
ST010
Ai18
ST064
ST026
ST027
ST083
ST057
ST108
ST071
ST070
ST056
RPH35
CD1080
CD1214
ST029
CD1171
ST042
ST048
ST119
ST002
ST112
ST073
CD966
ST033
ST065
ST103
ST059
ST100
CD959
CD1077
ST054
ST085
ST063
ST019
ST024
ST105
CD1010
ST111
ST021
CD871
ST079
ST104
ST075
ST080
ST074
ST076
ST078
ST107
ST040
ST003
ST084
ST118
ST101
ST028
ST018
ST016
ST066
S10.1920
ST091
ST098
ST031
ST055
ST115
ST051
ST099
(B)
WA.13
ST11
WA.12

Figure 3.1: Relative evolutionary relatedness of the five main lineages of *C. difficile* based on MLST analyses

Data for figure collated by M. Cairns, figure produced by Dr R. Stabler (LSHTM) and adapted from a publication by Stabler *et al*., (Stabler et al., 2012).

(A) Overview of phylogeny and (B) details of lineage 5.

Branch colouring; black = lineage 1 (mixed PCR ribotypes & sequence types), red = lineage 2 (sequence type 1/PCR ribotype 027), blue = lineage 3 (sequence type 22/PCR ribotype 023), orange = lineage 4 (sequence type 37/PCR ribotype 017) and green = lineage 5 (sequence type 11/PCR ribotype 078). The key sequence types for each lineage are in bold.

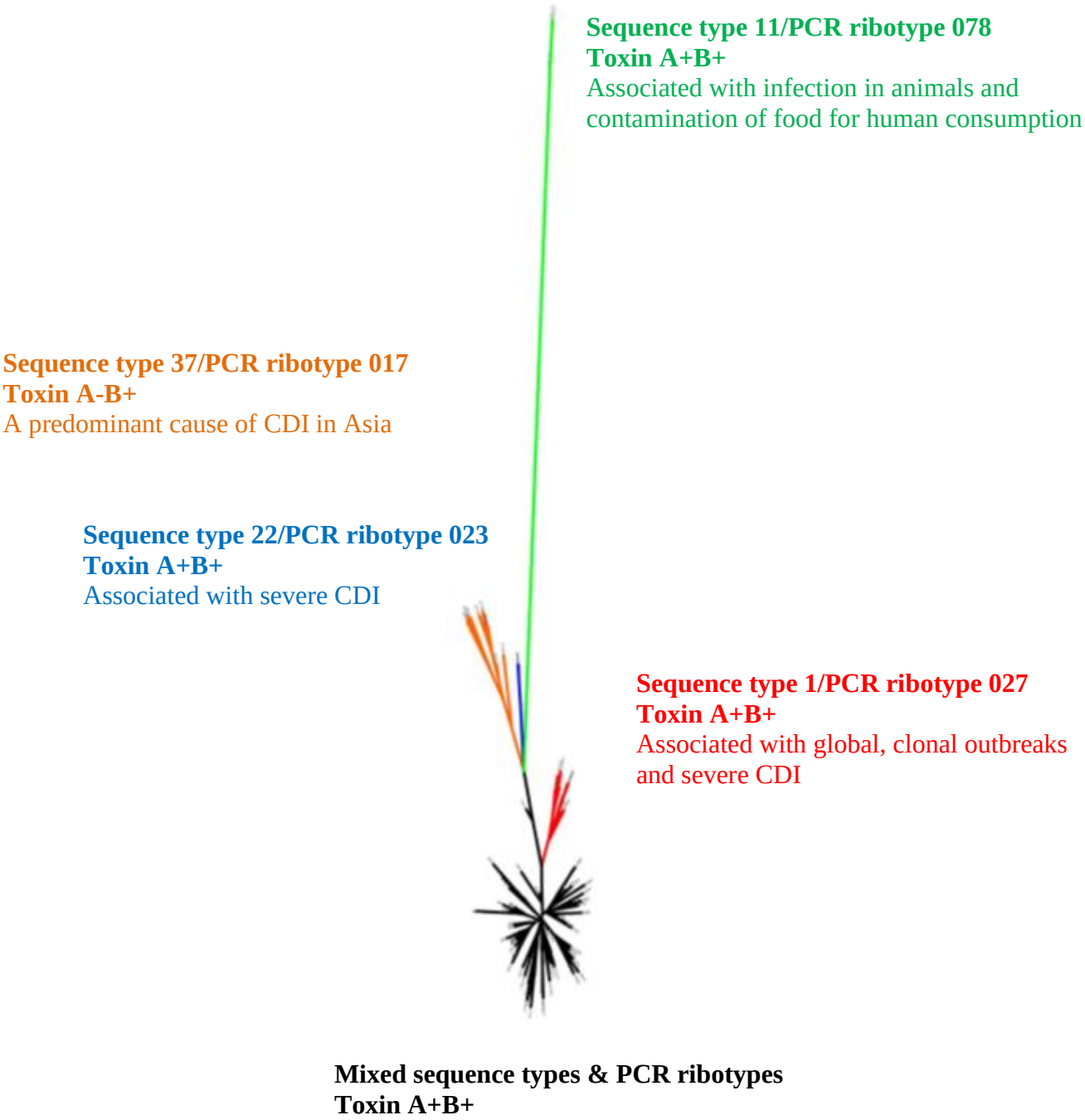


Figure 3.2: Overview of phylogeny of *C. difficile* based on MLST analyses

An overview of the relative evolutionary relatedness of the five main lineages of C. *difficile* based on MLST analysis. Data for figure collated by M. Cairns, figure produced by Dr R. Stabler (LSHTM) and adapted from a publication by Stabler *et al*., (Stabler et al., 2012). The figure includes significant genotypic and/or phenotypic characteristics for the key PCR ribotypes for the given lineage where known.

Our data indicate that the five distinct phylogenetic lineages of *C. difficile* observed in previous studies are maintained when applying MLST to a larger collection of isolates from diverse sources, geographical origin and PCR ribotype (Figures 3.1 and 3.2).

Lineage 1 (139/365) consisted of 51 sequence types which were heterogeneous in terms of PCR ribotypes. Lineage 2 consisted of 14 sequence types which can be sub-divided into two sublineages; a sequence type 1 sublineage (2i) and a sequence type 32 sublineage (2ii). Sublineage 2i contained sequence type 1/PCR ribotype 027, sequence type 67/PCR ribotype 019 and sequence type 41 (PCR ribotypes 106/164/321). Sublineage 2i also contained sequence type 62/PCR ribotype 036 and sequence type 140/PCR ribotype 111 which is a novel combination to our study. Sublineage 2ii contained six sequence types (32, 47, 61, 95, 97 and 123). Lineage 3 included sequence type 22/PCR ribotype 023, with three additional sequence types; 5, 25 and 96. Lineage 4 included sequence type 37/PCR ribotype 017 and an additional 12 sequence types. Lineage 4 can be subdivided into two sublineages; a sequence type 37 sublineage (4i) and a sequence type 23 sublineage (4ii). Sublineage 4i contains four sequence types, including both PCR ribotype 017 sequence types (37 and 86) as well as sequence type 39/PCR ribotype 085. Sublineage 4ii contains eight sequence types including sequence type 142/PCR ribotype 264 and sequence type 23/PCR ribotype 138. Sublineage 4ii has significant divergence from sublineage 4i and may represent a sixth lineage. Lineage 5 included sequence type 11/PCR ribotype 078 and an additional two sequence types which were both novel; 132 and 147.

3.4.5 MLST of *C. difficile* isolates by source

To determine if *C. difficile* strains isolated from human, animal and food origins are phylogenetically distinct by MLST, the sequence type profiles of the collection of isolates from a variety of sources were compared. Table 3.8 lists the number and percentage of isolates in this study by source. Table 3.9 lists the human isolates by lineage and sequence type. Table 3.10 lists in individual sub-tables, the animal, food and household isolates by lineage and sequence type. Human isolates were of multiple sequence types and fell into either lineage 1, 2, 3, 4 or 5 (Table 3.9). Although the number of animal and food source isolates in this study is small, our data does show a range of sequence types and lineages associated with strains isolated from animals and food (Table 3.10). Our study identified the following sequence types to be associated with strains isolated from animals and food; 1, 2, 3, 8, 11, 35, 48, 61 and 132. Most of the animal isolates (85.2% [92/108]) were sequence type 11. All sequence types for the animal isolates were of sequence types associated with human isolates (excluding ST132 which was novel to our study). These data indicate that strains isolated from humans, animals and food origins are not phylogenetically distinct by MLST.

Table 3.8: Number and percentage of isolates by source

Source	Number of isolates	Percentage of isolates
Human	211	54.8%
Bovine	80	20.8%
Porcine	17	4.4%
Food	16	4.1%
Murine	5	1.3%
Equine	4	1%
Canine	1	0.3%
Kangaroo	1	0.3%
Household	1	0.3%
Unknown	49	12.7%
TOTAL	**385**	**100%**

Table 3.9: Phylogeny of isolates of human origin

Lineage	Number of isolates	Sequence type	Number of isolates
1	103	*novel	13
		139 (novel)	2
		2	4
		3	11
		6	4
		8	9
		9	1
		010	4
		12	2
		13	1
		15	4
		16	1
		17	2
		18	1
		21	1
		26	2
		29	1
		33	2
		34	1
		35	1
		42	9
		43	2
		44	4
		45	1
		46	2
		48	1
		49	1
		51	1
		52	1
		53	1
		54	3
		55	4
		56	1
		57	1
		63	3
		66	1
2	50	140 (novel)	1
		1	42
		67	2
		62	1
		32	1
		41	1
		41	1
		61	1
3	5	5	4
		22	1
4	11	142 (novel)	8
		37	1
		039	1
		86	1
5	42	11	40
		147 (novel)	1
		148 (novel)	1
Total number of isolates			**211**

*Thirteen isolates were novel sequence types with only one isolate in lineage 1; sequence types; 131, 133, 134, 135, 136, 137, 138, 141, 143, 144, 145 and 146.

Table 3.10: Phylogeny of isolates of animal, food and household origin

Bovine

Lineage	Number of isolates	Sequence type	Number of isolates
1	1	3	1
2	6	1	6
5	73	11	73
Total number of isolates			**80**

Canine

Lineage	Number of isolates	Sequence type	Number of isolates
1	1	2	1
Total number of isolates			**1**

Equine

Lineage	Number of isolates	Sequence type	Number of isolates
1	2	3	1
		8	1
2	1	1	1
5	1	11	1
Total number of isolates			**4**

Murine

Lineage	Number of isolates	Sequence type	Number of isolates
1	5	35	5
Total number of isolates			**5**

Porcine

Lineage	Number of isolates	Sequence type	Number of isolates
1	4	ST8	1
		ST48	2
		ST132 (novel)	1
5	14	*ST11	14
Total number of isolates			**18**

*One isolate was from a Kangaroo.

Food and household

Lineage	Number of isolates	Sequence type	Number of isolates
2	9	*ST1	7
		ST61	2
5	8	ST11	8
Total number of isolates			**17**

*One isolate was from a household.

3.4.6 MLST of *C. difficile* isolates by geographical origin

The study by Lemee *et al.,* investigated 72 strains of *C. difficile* isolated from various global locations and found no correlation between sequence type and geographical location (Lemee et al., 2004). This study aimed to determine if *C. difficile* strains isolated from different geographical origins are phylogenetically distinct by MLST using a larger collection of strains from various geographical locations. Table 3.11 lists the number and percentage of isolates in this study by geographical origin (where provided). The majority of isolates were from the USA and UK and so these have been broken down further; Tables 3.12 and 3.13 list the number and percentage of isolates by sequence type for the USA and UK respectively. Our data shows that multiple lineages are associated with different geographical locations and multiple sequence types are associated with isolates from the UK and USA.

Table 3.11: Number and percentage of isolates by geographical origin and lineage

Geographical Location	Number of isolates	Percentage of isolates	Lineage/s
USA	168	43.6%	1, 2 and 5
UK	120	31.2%	1, 2, 3, 4 and 5
Australia	36	9.4%	1,4 and 5
Belgium	2	0.5%	4
Canada	2	0.5%	2
Ireland	2	0.5%	4 and 5
Italy	2	0.5%	5
Spain	2	0.5%	5
Switzerland	2	0.5%	1 and 5
France	1	0.3%	2
Germany	1	0.3%	1
Unknown	47	12.2%	1, 2, 3, 4 and 5
TOTAL	**385**	**100%**	**1, 2, 3, 4 and 5**

Table 3.12: Number and percentage of isolates from the USA by sequence type

Sequence type	Number of isolates	Percentage of isolates
11	108	64.3%
1	36	21.4%
3	6	3.6%
35	5	3.0%
61	3	1.8%
8	2	1.2%
48	2	1.2%
67	2	1.2%
***Other**	4	2.4%
TOTAL	**168**	**100%**

* Four isolates were of other sequence types with only one isolate; sequence types 2, 32, 41 and 46.

Table 3.13: Number and percentage of isolates from the UK by sequence type

Sequence type	Number of isolates	Percentage of isolates
1	19	15.8%
3	10	8.3%
42	8	6.7%
11	7	5.8%
8	6	5.0%
37	5	4.2%
10	4	3.3%
44	4	3.3%
6	3	2.5%
2	3	2.5%
33	2	1.7%
5	2	1.7%
12	2	1.7%
17	2	1.7%
55	2	1.7%
63	2	1.7%
***Other**	23	19.2%
****Novel**	16	13.3%
TOTAL	**120**	**100%**

*Twenty-three isolates were of other sequence types with only one isolate, sequence types; 13, 15, 16, 18, 21, 22, 26, 35, 39, 41, 43, 45, 46, 48, 49, 51, 52, 53, 54, 56, 57, 62 and 66.

**Sixteen isolates were novel sequence types; 131, 133, 134, 135, 136, 137, 138, 140, 141, 143, 144, 146 and 139 for three isolates.

3.4.7 Molecular characterisation of *C. difficile* using the different typing techniques; MLST, PCR ribotyping and REA

Outbreak strains are often referred to as PCR ribotype 027/NAP1/BI, suggesting that the three typing methods correlate however the adequacy of inter-laboratory correlation between PCR ribotypes, PFGE types and REA types is unknown. To investigate the notion that not all REA type BI strains are PCR ribotype 027 by PCR ribotyping, 16 assumed PCR ribotype 027 strains identified as type BI by REA in the USA were PCR ribotyped (Table 3.14).

Table 3.14: PCR ribotyping profiles of REA type BI strains

Strain	PCR ribotype
BI-1	027
BI-2	027
BI-3	027
BI-4	027
BI-5	027
BI-6	176
BI-6p	027
BI-6p2	027
BI-7	027
BI-8	027
BI-10	027
BI-11	198
BI-12	027
BI-13	027
BI-14	244
BI-15	027

We found 3 (3/16) strains were not PCR ribotype 027 but instead were 176, 198 and 244 (Table 3.14). Although different PCR ribotypes, it was observed that the PCR ribotyping banding patterns for 176 and 198 are extremely similar to that of 027 (plus or minus one band [Figure 3.3]). PCR ribotype 244 was observed to be two bands different so not as similar.

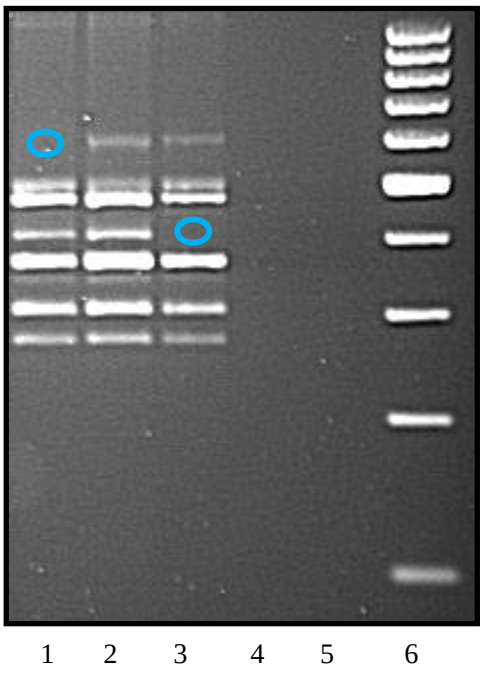


Figure 3.3: PCR ribotyping agarose gel electrophoresis of strains BI-6 (PCR ribotype 176), R20291 (PCR ribotype 027) and BI-11 (PCR ribotype 198)

DNA extraction, PCR ribotyping and agarose gel electrophoresis performed by M. Cairns, figure produced by M. Cairns and adapted from a publication by Valiente *et al.*, (Valiente et al., 2012). PCR ribotyping amplicon banding patterns as observed by agarose gel electrophoresis.

Lane 1= BI-11 (PCR ribotype 198/sequence type 1)
Lane 2 = R20291 (PCR ribotype 027/sequence type 1 [positive control])
Lane 3 = BI-6 (PCR ribotype 176/sequence type 1)
Lane 4 = Negative extract control
Lane 5 = Negative PCR control
Lane 6 = 100 bp ladder

This figure depicts the plus/minus band similarity between PCR ribotyping patterns (locations of missing bands are encircled in blue).

The MLST lineage 2 most often contains PCR ribotype 027/sequence type 1. To investigate the diversity of strains found in MLST lineage 2, the sequence types and PCR ribotypes of 71 isolates found in MLST lineage 2 were compared. Results are listed in Table 3.15.

Table 3.15: MLST lineage 2 isolates

Lineage	Sequence type	PCR ribotype	Number of isolates
2	1	027	53
		Nontypeable	6
		176	2
	61	Nontypeable	3
	67	262	2
	32	Nontypeable	1
	62	036	1
	41	Nontypeable	1
		135	1

MLST lineage 2 contained sequence types other than 1 and PCR ribotypes other than 027 (Table 3.15). With our larger collection of strains, our data has found further sequence types, 61 and 62 associated with lineage 2.

3.5 Discussion

To date, MLST studies on *C. difficile* have focused on human isolates from limited geographical regions. In our study MLST profiles of 385 *C. difficile* isolates were investigated, including previously undocumented PCR ribotypes taken from human, animal and food sources and from multiple geographical locations. Despite the diverse origins of the isolates in our study, the five phylogenetic lineages of *C. difficile* previously reported (Griffiths et al., 2010, Dingle et al., 2011a, Stabler et al., 2006, He et al., 2010) were confirmed (Figure 3.1).

PCR analysis identified 68 different PCR ribotypes (319/385) and found 66 isolates (66/385) to be non-typeable (Table 3.1). This is an increase in diversity of PCR ribotypes compared to previous studies; n=62 (Lemee et al., 2004), n=49 (Griffiths et al., 2010) and n=61 (Dingle et al., 2011a) strengthening our ability to investigate the pylohistory of *C. difficile*. Unsurprisingly the majority of isolates in our study were PCR ribotype 078 (27.8%) and PCR ribotype 027 (13.8%) as listed in Table 3.1; these were the commonest cause of CDI in animals (Keel et al., 2007, Jhung et al., 2008, Goorhuis et al., 2008, Rupnik et al., 2008) and humans respectively (Loo et al., 2005, Warny et al., 2005, Redelings et al., 2007).

MLST analysis identified 48 known sequence types (365/385) and found 18 sequence types (20/385) to be novel (Table 3.2). This is an increase in diversity of sequence types compared to the Lemee *et al.,* (Lemee et al., 2004) and Griffiths *et al.,* (Griffiths et al., 2010) studies with n=34 and n=40 sequence types respectively but less than the Dingle *et al.,* study with n=69 sequence types (Dingle et al., 2011a). The study by Dingle *et al.,* was based on 1290 isolates, our study was significantly

less with 385 isolates, however, our study included isolates of more diverse PCR ribotypes; 14 of the 20 novel sequence types in our study were of PCR ribotypes not included in the study by Dingle *et al.,*. This likely explains why our study identified a further 18 novel sequence types compared with the previous study by Dingle *et al.*, (Appendix 1, Page 229).

Previous studies have investigated 45 (Griffiths et al., 2010) and 29 (Dingle et al., 2011a) PCR ribotypes and reported 74 PCR ribotype/sequence type associations i.e. PCR ribotype 027 isolates are sequence type 1 and vice versa. Using a larger collection of PCR ribotypes (n=67), we identified 39 previously described and 46 novel PCR ribotype/sequence type associations (including 20 novel sequence types [Table 3.4]). The finding of 20 novel sequence types indicates further micro-diversity with sub-lineages within lineages which suggests the continued evolution of *C. difficile* strains and the potential emergence of hypervirulent strains.

In the current study, it was found that lineage 1 (139/365) consisted of 51 sequence types which were heterogeneous in terms of PCR ribotypes. This is consistent with that found by previous MLST studies on *C. difficile* (Griffiths et al., 2010, Dingle et al., 2011a).

Lineage 2 consisted of 14 sequence types which can be sub-divided into two sublineages; a sequence type 1 sublineage (2i) and a sequence type 32 sublineage (2ii). Sublineage 2i contained sequence type 1/PCR ribotype 027, sequence type 67/PCR ribotype 019 which concurs with the study by Griffiths *et al.,* (Griffiths et al., 2010) and sequence type 41 (PCR ribotypes 106/164/321) which is consistent

with the study by Dingle *et al.,* (Dingle et al., 2011a). Sublineage 2i also contained sequence type 62/PCR ribotype 036 and sequence type 140/PCR ribotype 111 which is a novel combination to our study. Sublineage 2ii contained six sequence types (32, 47, 61, 95, 97 and 123) with only sequence type 32/PCR ribotype 153 known previously (Griffiths et al., 2010). We found lineage 3 included sequence type 22/PCR ribotype 023, with three additional sequence types; 5 and 25 as previously reported (Griffiths et al., 2010, Dingle et al., 2011a) plus sequence type 96 which has not been reported before in this lineage. We found lineage 4 included sequence type 37/PCR ribotype 017 which was previously reported (Lemee et al., 2004, Griffiths et al., 2010, Dingle et al., 2011a) and an additional 12 sequence types. Lineage 4 can be subdivided into two sublineages; a sequence type 37 sublineage (4i) and a sequence type 23 sublineage (4ii). Sublineage 4i contains four sequence types, including both PCR ribotype 017 sequence types (37 and 86) as well as sequence type 39/PCR ribotype 085. Sublineage 4ii contains eight sequence types including sequence type 142/PCR ribotype 264 and sequence type 23/PCR ribotype 138. Sublineage 4ii has significant divergence from sublineage 4i and may represent a sixth lineage. Lineage 5 included sequence type 11/PCR ribotype 078 as previously reported (Griffiths et al., 2010, Dingle et al., 2011a) and an additional two sequence types which were both novel; 132 and 147. This all suggests continued evolution of *C. difficile* species.

Strains isolated from humans, animals and food origins were not phylogenetically distinct by MLST (Tables 3.8, 3.9 and 3.10). As previously shown (Griffiths et al., 2010, Dingle et al., 2011a), human isolates are of multiple sequence types and can fall into either lineage 1, 2, 3, 4 or 5 (Table 3.9).

Although the number of animal and food source isolates in this study is small, our data does show a range of sequence types and lineages associated with strains isolated from animals and food (Table 3.10). Our study identified the following sequence types to be associated with strains isolated from animals and food; 1, 2, 3, 8, 11, 35, 48, 61 and 132. Some of these were previously identified by Lemee *et al.*, (sequence types; 1, 3, 8) (Lemee et al., 2004). Strains of *C. difficile* isolated from animals have been shown to be sequence type 11 (Griffiths et al., 2010) as is found in this study with 85.2% (92/108) of strains isolated from animals being sequence type 11. Non-human isolates were found in lineages 1, 2 and 5 amongst strains isolated from humans. All sequence types for the animal isolates were also associated with human isolates (excluding sequence type 132). Sequence type 132 was novel; whether it is a sequence type exclusively associated with animals cannot be determined in this study due to only one isolate of this sequence type. Our data did not find a lineage exclusive to non-human strains indicating that the strains in our study were able to colonise or infect non-human sources equal to humans.

No correlation between sequence type and geographical location using MLST was found (Table 3.11). This lack of geographical association indicates that there has not been a rapid spread of a particular strain. A limitation to this study is that the majority of isolates were from the UK and USA; however, representative strains from other geographical locations were included (Australia, Belgium, Canada, France, Germany, Ireland, Italy, Spain and Switzerland).

Our data indicate that the molecular characterisation of *C. difficile* using MLST, PCR ribotyping and REA does not provide adequate correlation for inter-laboratory

comparison. Sixteen assumed sequence type 1/PCR ribotype 027 strains identified as BI by REA in USA were PCR ribotyped in this study and three were found not to be PCR ribotype 027 but instead, 176, 198 and 244. This is similar to data shown elsewhere which have found that not all PCR ribotype 027 strains are type BI by REA (Tenover et al., 2011). The close phylogeny of PCR ribotype 027/sequence type 1, PCR ribotype 176/sequence type 1 and PCR ribotype 198/sequence type 1 is supported by data from a previous study where comparative genomic hybridisation and Bayesian phylogeny was performed and suggested these strains may share a common ancestor (Stabler et al., 2006). Our data also found the PCR ribotyping banding patterns to be highly similar for PCR ribotypes 027, 176 and 198 (Figure 3.3).

Reported elsewhere, a strain of PCR ribotype 176 was responsible for two outbreaks in 2008 and 2009 in Poland (Nyc et al., 2011) and a study in the Czech Republic analysed 624 *C. difficile* strains from eleven hospitals in 2013 and found 40% of isolates were PCR ribotype 176. Interestingly, other studies have found PCR ribotype 176 to be incorrectly identified as PCR ribotype 027 using the Xpert® *C. difficile* assay (Krutova et al., 2014). The Xpert® *C. difficile* assay detects the presence of *C. difficile* (and also specifically PCR ribotype 027) in stool samples using PCR (Sloan et al., 2008). It is a multiplex PCR for three targets present in the toxigenic *C. difficile* genome: the *tcdB* gene for toxin B production, the *cdtB* gene for binary toxin production and the single nucleotide deletion at position 117 in the regulatory *tcdC* gene. Similar to the study by Krutova *et al.,* twelve strains of PCR ribotype 244 were misidentified as PCR ribotype 027 using the Xpert® *C. difficile* assay in Australia (Lim et al., 2014). Like PCR ribotype 027, they found PCR ribotype 244 was

associated with more severe disease and a higher mortality rate. Their WGS SNP analysis also found PCR ribotype 244 and 027 to be distinct but closely related. Previous to this, it was recognised in New Zealand that ten patients with CDI due to PCR ribotype 244 had severe infection, 50% of which were community acquired (De Almeida et al., 2013). A further study in Australia identified a clonal outbreak due to PCR ribotype 244; again these patients had severe infection, the majority were community acquired and patients were younger in age (Eyre et al., 2015). *C. difficile* PCR ribotype 176 and 244 strains produce binary toxin and possess the single nucleotide deletion in the *tcdC* gene at position 117 like that of PCR ribotype 027 (Krutova et al., 2017, Lim et al., 2014) which explains why they were detected by the the Xpert® *C. difficile* assay in the reports just mentioned. This means that these targets are not specific to PCR ribotype 027 as previously thought. The comparative genomic hybridisation and Bayesian phylogeny study performed by Stabler *et al.,* also found PCR ribotype 244 (strain BI-14) in lineage 2 but, as an outlier within the lineage (Stabler et al., 2006). The PCR ribotyping banding pattern for PCR ribotype 244 also differed by two bands to PCR ribotype 027 instead of only one band like with PCR ribotypes 176 and 198.

Additional sequence types and PCR ribotypes were also found associated with lineage 2 in this study (Table 3.15). This is similar to that found by other studies. *Griffiths et al.,* found sequence types 32 and 67 (Griffiths et al., 2010) and *Dingle et al.,* found sequence type 67 (Dingle et al., 2011a) within lineage 2. *Griffiths et al.,* also found an additional PCR ribotype, 036 associated with sequence type 1 which was similar in PCR ribotyping banding pattern like what our study has found with PCR ribotypes 176 and 198. With our larger collection of strains, our data has found

further sequence types, 61 and 62 associated with lineage 2. This suggests that lineage 2 has evolved further with new sequence types and supports the notion that PCR ribotyping and REA analysis do not correspond as well as previously thought. Both typing methods rely on the examination of DNA migration through a matrix, usually an agarose gel and comparison to a standard. Simple changes in DNA can have distinct changes in resulting profiles. For example, a change in the repeat copy number, inter-chromosomal homologous recombination for PCR ribotyping, primer binding sites or SNPs within restriction sites for REA can produce differences in banding patterns, which may result in the strain being designated as a different type or variant even though they are phylogenetically similar. In light of the phylogenetic similarities between multiple PCR ribotypes i.e. 027, 198 and 076, it would be plausible to investigate as future work the phenotypic similarities between PCR ribotypes that are different PCR ribotypes but highly similar in PCR ribotyping profile and phylogeny by MLST.

3.6 Conclusion

With a large and diverse collection of strains (n=385), our study confirmed the five lineages of *C. difficile* found by other studies with evidence of further micro-diversity of the *C. difficile* species. By including isolates from a variety of sources and geographical origin, our data found no lineage exclusive to *C. difficile* strains isolated from non-human sources or a geographical origin. Our study has revealed new links between closely related PCR ribotypes and sequence types, providing insights into the microevolution of *C. difficile.*

Our study confirmed that not all *C. difficile* REA type BI strains are PCR ribotype 027 but other PCR ribotypes such as 176, 198 and 244. Considering the similarity in phylogeny of these PCR ribotypes in this study and other studies (Stabler et al., 2006, Krutova et al., 2017, Lim et al., 2014), suggests *C. difficile* REA type BI, PCR ribotype 027 and non-PCR ribotype 027 strains have co-evolved or possibly evolved from each other. Given that PCR ribotyping is the most widely used typing method worldwide, and that MLST studies confirm the close grouping of different PCR ribotypes such as PCR ribotypes 176, 198, 244 with 027, suggests that although PCR ribotyping is useful, it must be considered that there should be heightened awareness of the clinical significance of these closely related PCR ribotypes.

MLST is an appropriate method for studying the phylogeny of *C. difficile* and the continued collection and genotyping of diverse *C. difficile* strains from all sources is vital for monitoring the emergence and disappearance of evolving virulent clones. Although five lineages of *C. difficile* have been confirmed using MLST, these should be investigated in more detail using WGS which offers superior resolution compared to MLST. Therefore, I have further investigated an individual lineage in more detail in chapters 4 and 5; *C. difficile* PCR ribotype 017.

Chapter 4

Genetic and phenotypic characterisation

of a hospital outbreak of

C. difficile PCR ribotype 017

4 Characterisation of a hospital outbreak of *C. difficile* PCR ribotype 017 using WGS and phenotypic assays

4.1 Statement of contribution

Isolate selection, PCR ribotyping and genomic DNA extractions were performed by M. Cairns. WGS were performed both at the WTSI by D. Harris and at the LSHTM by M. Cairns. The collection of epidemiology data were performed by M. Cairns with help from the infection control team at UHL. SNP calling were performed by Dr Mark Preston. *De novo* assembly were performed by M. Cairns with help from Dr Richard Stabler. Phenotypic assays, statistical analysis and data interpretation were performed by M. Cairns.

4.2 Introduction

MLST and WGS studies have confirmed the existence of at least five clonal lineages of *C. difficile* (Dingle et al., 2011a, Stabler et al., 2006) and this was confirmed in chapter 3 using MLST (Stabler et al., 2012).

C. difficile PCR ribotype 027 has been well studied, but less is known about other PCR ribotypes that make up the five lineages identified by MLST and WGS such as PCR ribotype 017. It is known that pathogenic strains of PCR ribotype 027 produce both toxins A and B (A+B+) and an unrelated CDT that has been implicated in virulence. By contrast, PCR ribotype 017 strains lack most of the *tcdA* gene (A-B+) and completely lack the CDT gene yet have emerged worldwide causing significant disease (Johnson et al., 2003, Pituch et al., 2006, Goorhuis et al., 2009, Hawkey et al., 2013). The reasons for the emergence of a less toxigenic lineage remain unclear.

The epidemiology of *C. difficile* in the Asia/Pacific regions and eastern parts of Europe also appears to differ from elsewhere where the prevalence of toxin A-B+ strains is higher in these locations compared to other PCR ribotypes including 027 which is toxin A+B+ (Hawkey et al., 2013). Like PCR ribotype 027, *C. difficile* toxin A-B+ strains have been associated with increased disease severity where they have been related to PMC in patients (Shin et al., 2008, Elliott et al., 2009, Limaye et al., 2000, Sambol et al., 2000, Johnson et al., 2001).

The Department of Health guidelines recommend use of a chlorine-containing cleaning agent for routine environmental disinfection in hospitals; however, studies have suggested that PCR ribotypes have differing susceptibilities to disinfectants. A study by Dawson *et al.*, compared the efficacy of nine commonly used hospital disinfectants (including chlorine-containing cleaning agents) against a selection of difference PCR ribotypes and found their efficacy was either dependent on the concentration, PCR ribotype or both (Dawson et al., 2011). The study found the percentage spore count for PCR ribotype 017 was significantly higher than that for PCR ribotypes 027 (Dawson et al., 2011).

UHL in South London experienced multiple clusters of non-severe CDI caused by PCR ribotype 017 that occurred at different times between 2009 and 2013; 21 isolates were recovered from 20 patients, 17 of which were taken from patients whilst on one elderly care ward during this time period (Table 4.1, Page 143). Due to these clusters, an environmental screen was performed on the ward (30/08/2010) followed by hydrogen peroxide vapour (HPV) decontamination (30/09/2010).

However, subsequent to this, isolates of PCR ribotype 017 were again recovered from patients with CDI from the same ward.

WGS offers considerable advantages over traditional phenotypic and genotypic typing methods where it has superior resolution and can perform a fine-grain analysis that facilitates the accurate tracing of the sources and routes of transmission (Parkhill and Wren, 2011). As well as SNP analysis, WGS can identify regions of a genome suggestive of horizontal gene transfer and horizontal transfer of genes on mobile genetic elements; conventional typing methodologies are unable to do this. Mobile genetic elements refer to DNA that can move around within and between genomes and include; conjugative transposons, plasmids, bacteriophages and introns (Frost et al., 2005). Mobile genetic elements are important since they play a major role in bacterial evolution and virulence whereby they often carry genes associated with virulence (Wiedenbeck and Cohan, 2011). Approximately, 11% of the *C. difficile* genome is made up of mobile genetic elements and a significant proportion of these elements are conjugative transposons (Sebaihia et al., 2006). Studied conjugative transposons in *C. difficile* include; Tn*916* (Roberts and Mullany, 2009), Tn*5397* (Mullany et al., 1996), Tn*4453a* and Tn*4453b* (Wren et al., 1988) and Tn*1549* (Sebaihia et al., 2006, Garnier et al., 2000) each with varying accessory regions and putative virulence factors.

4.3 Hypotheses of the research described in this chapter

PCR ribotype 017 is interesting; it forms its own phylogenetic lineage by MLST and WGS analysis, it is toxin A negative but is still pathogenic, it has shown to produce more spores than the hypervirulent PCR ribotype 027 and a London hospital has experienced multiple clusters of CDI caused by this PCR ribotype. These data warrant further investigation of this PCR ribotype using WGS in order to investigate its transmission and virulence. Detection of transmission has an implication for infection control and surveillance and identification of a virulent phenotype would justify heightened awareness, surveillance and investigation of improvements to laboratory diagnostics like that with PCR ribotype 027 and the Xpert® *C. difficile* assay described in chapter 3. This study was designed to test the following hypotheses:

Chapter 4: hypothesis 1

The *C. difficile* isolates from UHL are clonal and represent a phylogenetically distinct cluster when compared to *C. difficile* isolates from other London hospitals.

Chapter 4: hypothesis 2

The *C. difficile* isolates from UHL are phenotypically different to *C. difficile* isolates from other London hospitals.

4.4 Results

With my role as lead CDRN scientist at the London laboratory, I noticed that there were multiple isolates of PCR ribotype 017 being reported to UHL. I discussed this finding with the infection control nurses at UHL and this study commenced.

A total of 37 *C. difficile* isolates were used in this study and their details are listed in Table 4.1. These were; 22 strains isolated from humans who at different times were on the same hospital ward at UHL and two strains isolated from the same hospital ward. An additional 13 contemporaneous strains isolated from humans from other London hospitals were also included for comparison.

Twenty-three of the UHL isolates were part of an ongoing cluster, one of which (H-UHL-1) was a historical isolate from 2005 which predates the other 22 isolates in the cluster which were isolated between 2009 and 2013, and even the existing building where the elderly care ward was located. Two of these cluster isolates were environmental (E-UHL-19 and E-UHL-20) and recovered from the toilet and floor of the elderly care ward side-room at UHL when an environmental screen was performed.

All isolates in this study were propagated on appropriate culture media (Sections 2.2.2, Page 70 and 2.2.3, Page 71) from which genomic DNA was extracted using methods described for PCR ribotyping and WGS (Sections 2.3.1, Page 77 and 2.3.3.1, Page 83 respectively). PCR ribotyping was performed by PCR amplification and agarose gel electrophoresis (Section 2.3.1, Page 77). In preparation for WGS, DNA was quantified (Section 2.3.3.2, Page 84) and WGS was performed using

Illumina Sequencing Technology (Sections 2.3.3.3, Page 85 and 2.3.3.4, Page 88). SNPs and haplotypes were identified by sequence mapping and assembly (Sections 2.3.3.5, Page 85 to 2.3.3.9, Page 92 inclusive), and *de novo* genome assembly analysis and visual inspection using ACT was performed to identify regions of DNA that were; present in a test strain but absent from the reference control M68 (insertions) and present in the control strain M68 but absent from a test strain (deletions) [Section 2.3.3.10, Page 92]). Software tools were used to produce phylogenetic trees to enable visual comparisons of SNP differences between UHL and non-UHL *C. difficile* isolates (Section 2.3.9, Page 91) and phenotypic assays and statistical analysis were performed on select isolates (UHL-1, UHL-3, UHL-19, CX-32) to identify a phenotype unique to the UHL *C. difficile* isolates (Sections 2.2.6, Page 73 to 2.2.10, Page 76 inclusive).

Table 4.1: Bacterial isolates used in chapter 4

Isolate	Provider	Sample Date	Hospital (date/s patient on elderly care ward at UHL if not the same as the date the sample was taken)
***H-UHL-1**	CDRN	2005	UHL (ward non-existent)
UHL-2	CDRN	11/03/09	UHL (ward)
UHL-3	CDRN	27/03/09	UHL (24/02/09-05/03/09)
UHL-4	CDRN	17/04/09	UHL (ward)
UHL-5	CDRN	16/04/09	UHL (ward)
UHL-6	CDRN	28/09/09	UHL (ward)
UHL-7	CDRN	20/09/09	UHL (ward)
UHL-8	CDRN	16/10/09	UHL (ward)
UHL-9	CDRN	29/10/09	UHL (ward)
UHL-10	CDRN	28/01/10	UHL (ward)
UHL-11	CDRN	08/02/10	UHL (ward)
UHL-12	CDRN	17/02/10	UHL (ward)
UHL-13	CDRN	01/04/10	UHL (ward)
UHL-14	CDRN	26/04/10	UHL (ward)
UHL-15	CDRN	17/07/10	UHL (ward)
UHL-16	CDRN	19/07/10	UHL (ward)
UHL-17	CDRN	06/08/10	UHL (never)
UHL-18	CDRN	10/08/10	UHL (ward)
****E-UHL-19**	CDRN	13/08/10	UHL (ward: side-room toilet)
****E-UHL-20**	CDRN	13/08/10	UHL (ward: side-room floor)
UHL-21	CDRN	04/10/10	UHL (never)
UHL-22	CDRN	07/10/10	UHL (04/06/10-12/07/10)
UHL-23	CDRN	26/04/11	UHL (15/02/1-14/04/11)
****C-UHL-24**	CDRN	08/03/13	UHL (26/12/12-28/12/12, 04/02/13-07/02/13 and 08/03/13-20/02/13)
NP-25	CDRN	13/05/08	Northwick Park Hospital
B-26	CDRN	27/02/09	Barnet Hospital
NM-27	CDRN	2005	North Middlesex Hospital
GOSH-28	CDRN	22/03/10	Great Ormond Street Hospital
GOSH-29	CDRN	24/03/10	Great Ormond Street Hospital
GOSH-30	CDRN	27/03/10	Great Ormond Street Hospital
RF-31	CDRN	09/12/10	Royal Free Hospital
CX-32	CDRN	25/01/11	Charing Cross Hospital
B-33	CDRN	11/05/11	Barnet Hospital
QM-34	CDRN	16/07/11	Queen Mary's Hospital
WX-35	CDRN	08/12/11	Whipp's Cross Hospital
WX-36	CDRN	16/01/12	Whipp's Cross Hospital
GOSH-37	CDRN	30/11/13	Great Ormond Street Hospital

Table adapted from a publication by Cairns *et al.*, (Cairns et al., 2015).

* = historical isolate pre-dating the build of the hospital ward, ** = environmental isolate recovered from the ward and *** = community-acquired infection isolate.

4.4.1 Genotypic comparison of isolates from London hospitals

4.4.1.1 PCR ribotyping

The 34 patient and two environmental isolates (Table 4.1) were confirmed to be PCR ribotype 017 by PCR ribotyping and agarose gel electrophoresis (Section 2.3.1). Reference strains R20291 (PCR ribotype 027) and M68 (PCR ribotype 017) were used as controls and all isolates were assumed to be toxinotype VIII, A-B+ based on their being PCR ribotype 017. There are no known reports of PCR ribotype 017 having any other toxin profile other than toxinotype VIII, A-B+.

4.4.1.2 Inference of phylogeny using SNP data

To investigate the phylogeny and clonality of the isolates in this study, after sequence quality control and mapping to the control strain M68 reference genome, SNP analysis revealed 162 bi-allelic SNP loci in the samples within the 4,308,325 bp of the control strain M68 with the majority 79.0% (128/162) exhibiting a MAF of less than 10% including 54.3% (88/162) loci being identified in one sample. Only 17 SNP loci (10.5%) had a non-reference allele frequency above 50% and each isolate contained up to 46/162 (28.4%) mutations, with 64.9% (24/37) isolates containing between 17 and 19 (10.5-11.7%) SNPs.

The complete dataset of 36 isolates revealed 23 different haplotypes designated A to W of between three and 47 SNPs (Table 4.2). Nine of the 23 haplotypes were only found in 23 isolates; H-UHL-1 and UHL-2 to UHL-23 (labelled A to I). Twenty-four SNP loci were unique to these 23 UHL samples; 16 non-synonymous, five synonymous and three non-genic (Table 4.3). Haplotype A was the ‘core’ haplotype consisting of 11 SNPs common to all, and unique to the UHL isolates. Haplotypes B

to H contained one or two extra SNPs to the haplotype A pattern. Haplotype I was distinguishable from haplotype A by five SNPs.

Table 4.2: Haplotypes for the London isolates

Haplotype	Number of isolates	Number of SNPs	Cluster
A	10	17	Cluster 1-UHL
B	4	18	Cluster 1-UHL
C	3	19	Cluster 1-UHL
D	1	19	Cluster 1-UHL
E	1	18	Cluster 1-UHL
F	1	18	Cluster 1-UHL
G	1	19	Cluster 1-UHL
H	1	18	Cluster 1-UHL
I	1	22	Cluster 1-UHL
J	1	3	Cluster 2-M68
K	1	7	Cluster 2-M68
L	1	18	Cluster 2-M68
M	1	19	Cluster 2-M68
N	1	23	Cluster 2-M68
O	1	9	Cluster 2-M68
P	1	7	Cluster 2-M68
Q	1	19	Cluster 2-M68
R	1	32	Cluster 3
S	1	38	Cluster 3
T	1	38	Cluster 3
U	1	38	Cluster 3
V	1	47	Cluster 3
W	1	38	Cluster 3

Table 4.3: SNPs unique to cluster 1-UHL isolates

Position in genome	Reference	Alternative	Type	Product/Putative Function	Haplotype
345335	S	R	Non-synonymous	Protein-tyrosine phosphatase reductase	A to I
433205	S	S	Synonymous	Formate/nitrite transporter	A to I
578215	P	S	Non-synonymous	Iron hydrogenase	A to I
707105	F	L	Non-synonymous	Multidrug family ABC transporter permease	A to I
1123155	G	S	Non-synonymous	Putative membrane protein	A to I
1241002	L	L	Synonymous	NhaC family Na+/H+ antiporter	A to I
1316457	A	A	Synonymous	3-hydroxybutyrate dehydrogenase	A to I
2764775	P	L	Non-synonymous	Diguanylate kinase signaling protein	A to I
3072208	G	A	Non-synonymous	Maf-like protein	A to I
3202066	L	F	Non-synonymous	Multidrug family ABC transporter	A to I
4025381	.	.	Intergenic	Unknown	A to I
1491685	.	.	Intergenic	Unknown	B, C and D
584197	C	R	Non-synonymous	Response regulator (quorum-sensing system)	E
1245898	H	N	Non-synonymous	Copper-sensing transcriptional repressor CsoR	H
1395682	I	L	Non-synonymous	Hypothetical protein	H
583796	R	L	Non-synonymous	Response regulator (quorum-sensing system)	F
1932695	I	V	Non-synonymous	Putative membrane protein	C
3698806	.	.	Intergenic	Unknown	G
3056134	L	I	Non-synonymous	Ribonuclease G (RNase G)	D
34552	S	Y	Non-synonymous	*rpoB* gene	I
2744067	E	E	Synonymous	Putative TPR repeat-containing protein	I
2813984	E	E	Synonymous	GntR family transcriptional regulator	I
3289962	S	Y	Non-synonymous	ABC transporter substrate-binding protein	I
3766047	K	Stop codon	Non-synonymous	Two-component response regulator	I

Table adapted from a publication by Cairns *et al.,* (Cairns et al., 2015).

A heat map of genetic Manhattan distance (Figure 4.1) depicts the three related groups of samples designated cluster 1-UHL, cluster 2-M68 and cluster 3. A heat map acts as a graphical illustration of the relationship between data points in a data set. Individual pairwise comparisons are represented by a value between 0 and 1 in a matrix. These are represented by colour in a heat map; black is completely identical data points (0) and white is completely different (1). By using Manhattan distance between two data points of SNP data, normalised between 0 and 1, a heat map can easily be made. Figure 4.1 illustrates genetic similarity as darker colours based on SNPs; the cluster 1-UHL isolates are darker colour indicating their clonality.

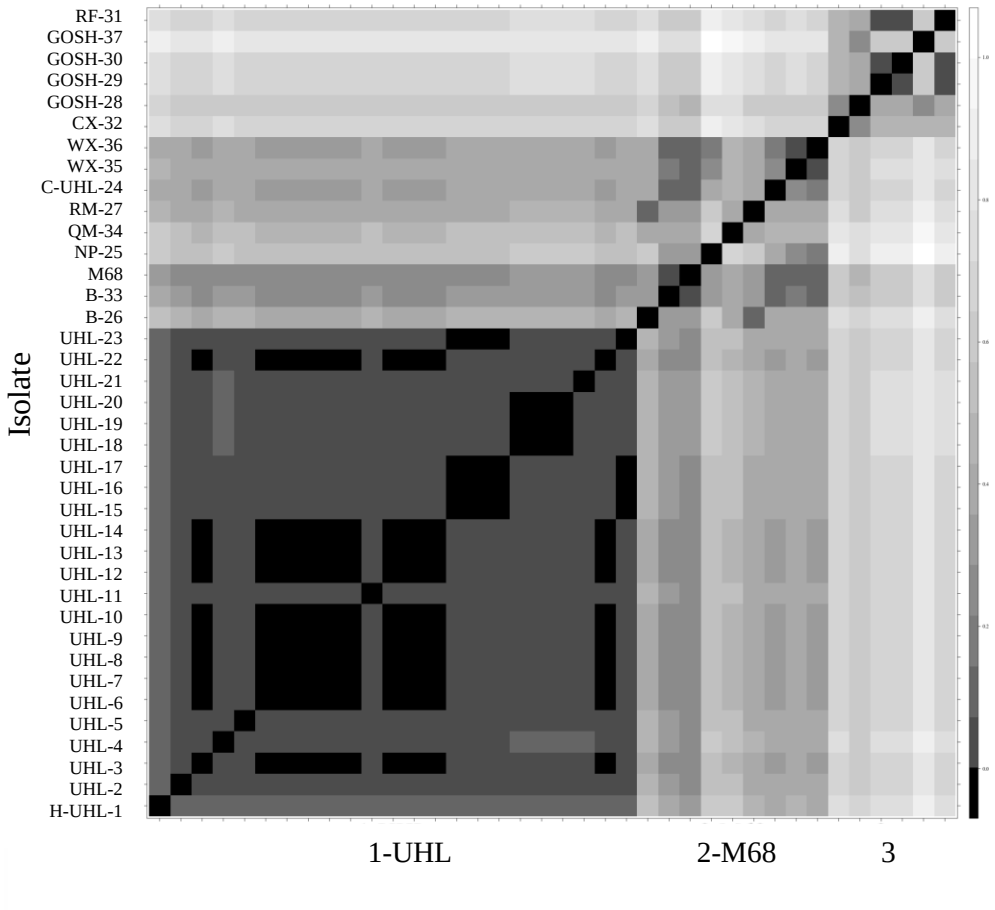


Figure 4.1: Heat map of the inter- and intra-cluster relatedness of the *C. difficile* PCR ribotype 017 isolates from London hospitals

Heat map taken from a publication by Cairns *et al*., (Cairns et al., 2015) to show the inter- and intra-cluster relatedness of the London *C. difficile* isolates. A normalised Manhattan distance between samples was used, with darker colours indicating a closer genetic identity between samples. The clonality of the cluster 1-UHL isolates is illustrated by the darker colour.

A maximum-likelihood phylogenetic tree (Figure 4.2) also depicts the three related groups of isolates designated cluster 1-UHL, cluster 2-M68 and cluster 3. Figure 4.2 also depicts the genetic relatedness specifically between the cluster 1-UHL isolates (a maximum of 5 SNPs).

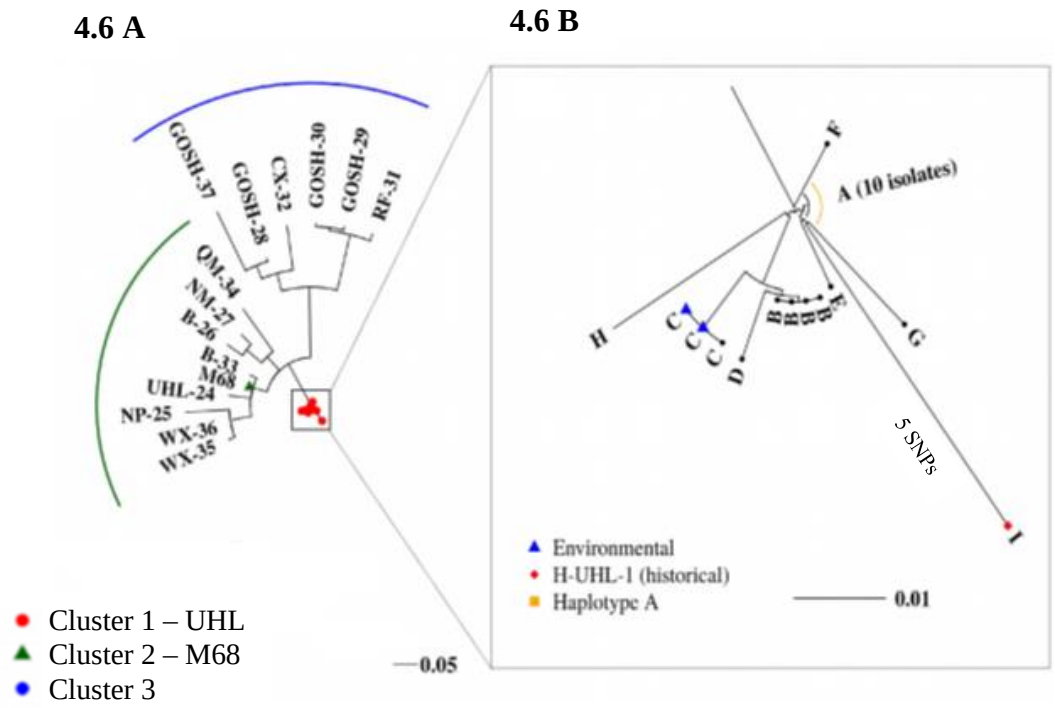


Figure 4.2: Maximum-likelihood phylogenetic analysis of the *C. difficile* PCR ribotype 017 isolates from London hospitals

Figure adapted from a publication by Cairns *et al.*, (Cairns et al., 2015). Maximum-likelihood phylogenetic analysis trees of all 37 London *C. difficile* isolates. A) based on core-genome SNPs against the control strain M68 and B) cluster 1-UHL haplotypes.

The cluster 1-UHL was composed exclusively of haplotypes A to I, containing 23/24 isolates from UHL; 96.3% of SNP loci were found exclusively in this cluster. One isolate in this cluster, H-UHL-1 (haplotype I) is an historical isolate from 2005 which predates the other 22 isolates in the cluster, and even the existing building where the elderly care ward was located. The H-UHL-1 sample/haplotype I was distinguishable from the UHL core haplotype (A) by only five SNPs. The four year isolation gap between the historical isolate in 2005 and the first cluster 1-UHL isolate in 2009 is consistent with a mutation rate of 0.74SNPs/year (Eyre et al., 2013), one to two SNPs/year (He et al., 2013) and 1.4 SNPs/year (Didelot and Maiden, 2010) suggesting that the historical H-UHL-1 isolate shares a common ancestor with cluster 1-UHL or is itself the ancestral strain. Haplotypes A to H were indistinguishable based on a ≤ 2 SNP difference suggesting they are the same strain and transmission has occurred between patients and/or via ward contamination. Additionally, the two environmental isolates (E-UHL-19 and E-UHL-20) recovered from the toilet and floor of the elderly care ward side-room were indistinguishable from the cluster 1-UHL; the environment was therefore contaminated with this strain. Two UHL isolates (UHL-17 and UHL-21) from patients who were never admitted onto the elderly care ward were found to be part of the cluster 1-UHL, strongly suggesting inter-ward transmission. One isolate (C-UHL-24) recovered from a patient taken on the day of admittance to the elderly care ward was distinguishable from the cluster 1-UHL; however, this CDI was defined as community-acquired based on the patient developing CDI < 48 hours of hospital admission. Although the patient had spent time on the elderly care ward in the past, this CDI episode was likely picked up in the community or visit/s to other healthcare institution/s. Subsequent to the hydrogen peroxide ward decontamination that was

performed in September 2010, two isolates of *C. difficile* PCR ribotype 017 indistinguishable from the cluster 1-UHL were recovered from patients in October 2010 and April 2011 (UHL-22 and UHL-23). This strain has persisted at UHL with a possible internal reservoir that was never eliminated during the HPV decontamination. With the historical isolate from 2005 being indistinguishable from the more recent clonal isolates, the possibility of an external source that has re-introduced this strain to the elderly care ward or a patients exit and re-entry to the ward cannot be excluded.

All isolates in cluster 1-UHL (haplotypes A to I) contained a gene putatively associated with virulence; a multidrug family ABC transporter (Table 4.3). These have been associated with contributing to antimicrobial resistance where they play a role in active efflux of antimicrobial agents from the bacterial cell rendering it ineffective against the bacteria (Lubelski et al., 2007). Cluster 2-M68 (containing the control strain M68) encompassed the outer London hospitals and a UHL patient (UHL-24) who had spent time on the elderly care ward (26/12/12 to 28/12/12, 04/02/13 to 07/02/13 and 08/03/13 to 20/02/13), though their CDI was defined as community acquired due to their symptoms beginning < 48 hours post hospital admission. Cluster 3 contains all of the isolates from the three inner London hospitals (Great Ormond Street, Royal Free and Charing Cross). All 162 SNPs were the same mutations and the same positions across all isolates, and only six SNP loci (3.8%) had mutations in more than one cluster. Cluster 3 contained isolates that were the most distinguishable from the control strain M68 with between 32 and 47 SNPs.

As well as SNPs unique to the cluster 1-UHL haplotypes, a noteworthy SNP was present in both cluster 1-UHL and cluster 3 isolates but absent from cluster 2-M68 isolates. This was a non-synonymous SNP in the *bioB* gene (Appendix 3, Page 256). This is one of four genes that play a major role in the biotin biosynthetic pathway (Satiaputra et al., 2016). Biotin mediates the transport of CO_2 in many vital metabolic reactions and is essential for life. Interestingly, studies have shown decreased available biotin results in increased toxin synthesis by *C. difficile* (Yamakawa et al., 1996). This SNP may affect biotin production, limiting its availability and in-turn increasing toxin synthesis.

4.4.1.3 *De novo* genome assembly analysis

The SNP analysis tells us about single base pair changes but doesn't tell us about larger deletions or insertions or other rearrangements within the genome, therefore I undertook *de novo* assembly of each isolate in comparison to the control strain M68 with programmatic and visual inspection of the comparisons using ACT software. This revealed a 49 kbp genetic region exclusive to the cluster 1-UHL (23/37) isolates (Figure 4.3). No other large structural variations between samples were revealed. Annotation of the region identified 30 coding sequences that had orthologues to known gene sequences and 15 coding sequences with no known orthologue (Table 4.4). These 45 predicted genes are highly conserved; 41/45 had 100% amino acid identity across all 23 samples in the 1-UHL cluster.

The genetic region contained genes suggestive for transposition; conjugative transfer of transposon-like mobile genetic element genes and an endonuclease relaxase MobA/VirD2 (Byrd and Matson, 1997, Silby et al., 2007). Transposons are known

to carry additional genes associated with virulence and/or transmissibility (Brouwer et al., 2011). The 49 kbp genetic region in this study does contain genes putatively associated with virulence and transmissibility (Table 4.4); sortase B gene (Spirig et al., 2011), the sporulation gene *spo0J* (Wu and Errington, 2003) and genes associated with antimicrobial resistance; tetracycline resistance transcriptional regulator and a multi antimicrobial extrusion protein (Nies et al., 2012).

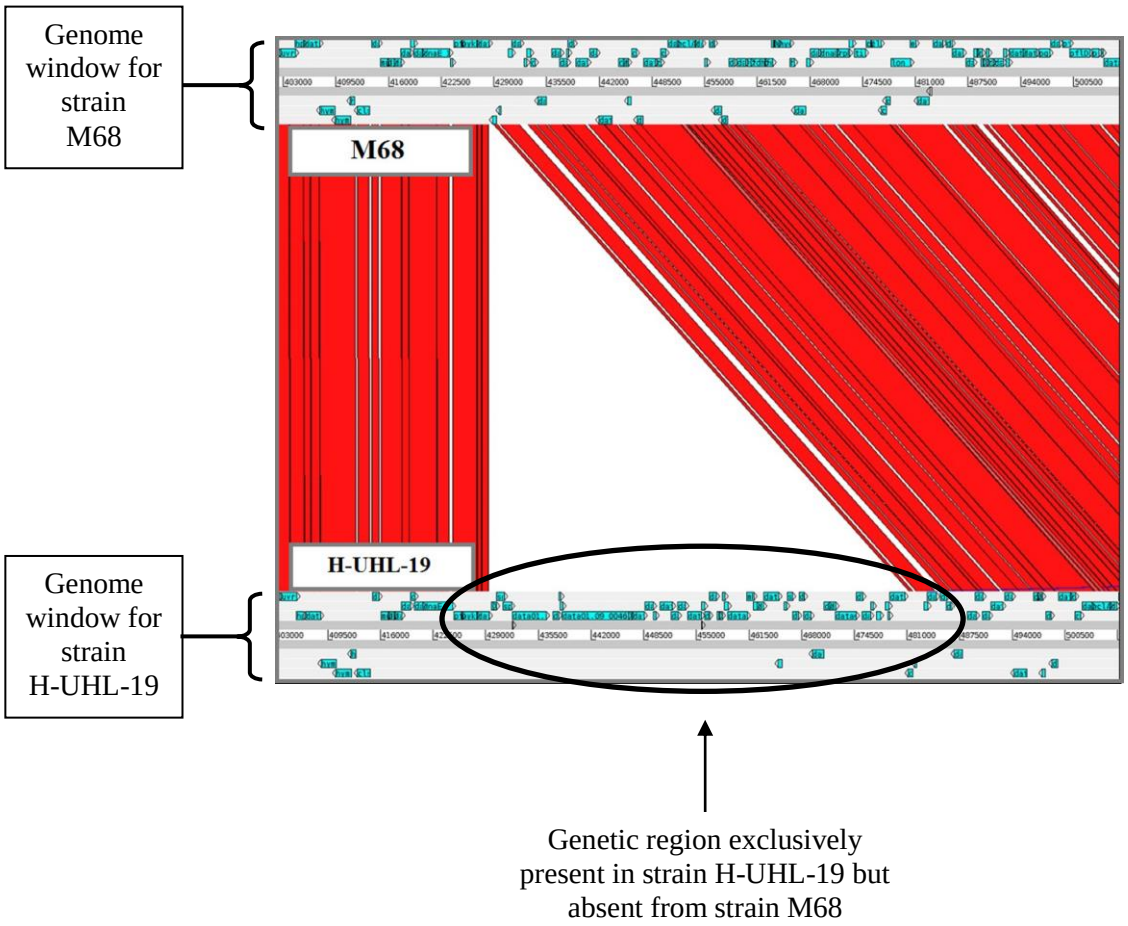


Figure 4.3: ACT illustration of the 49 kbp genetic region exclusive to the cluster 1-UHL isolates

Figure adapted from a publication by Cairns *et al*., (Cairns et al., 2015). Comparative genomic analyses were performed using ACT software with the *C. difficile* genomes for the reference control strain M68 and strain H-UHL-19. Red bars indicate sequence similarity. The ACT analysis revealed a genetic region exclusive to isolates in cluster 1-UHL (including H-UHL-19) which is encircled.

Table 4.4: Predicted coding sequences found in the 49 kbp genetic region exclusive to cluster 1-UHL isolates

Genetic region in the H-UHL-19 genome	Putative Product/Role/Function according to NCBI Blast and UniProt
Data01.09_00455	Uncharacterised protein
Data01.09_00456	Uncharacterised protein
Data01.09_00457	Soj (*parA*) protein (plasmid and chromosome partitioning)
Data01.09_00458	Spo0J (*parB*) protein (plasmid and chromosome partitioning)
Data01.09_00459	Collagen-binding Cna protein, TonB-dependent receptor
Data01.09_00460	Uncharacterised protein
Data01.09_00461	RadA - ATP-dependent protein (DNA repair and degradation of proteins) protease
Data01.09_00462	Uncharacterised protein
Data01.09_00463	Superfamily II DNA and RNA helicase (RNA and DNA metabolism)
Data01.09_00464	Uncharacterised protein
Data01.09_00465	None
Data01.09_00466	Uncharacterised protein
Data01.09_00467	Uncharacterised protein
Data01.09_00468	Endonuclease relaxase, MobA/VirD2
Data01.09_00469	Uncharacterised protein
Data01.09_00470	Uncharacterised protein
Data01.09_00471	Conjugative transfer of transposon-like mobile genetic elements
Data01.09_00472	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase
Data01.09_00473	Adenine-specific methyltransferase (macrolide resistance)
Data01.09_00474	Uncharacterised protein
Data01.09_00475	Phd_YefM (antitoxins in type II toxin-antitoxin systems)
Data01.09_00476	Plasmid stabilisation system protein
Data01.09_00477	Uncharacterised protein
Data01.09_00478	PrgI family protein (uncharacterised protein)
Data01.09_00479	AAA-like domain protein (uncharacterised protein)
Data01.09_00480	Modification methylase MboII (DNA methylation)
Data01.09_00481	Uncharacterised protein
Data01.09_00482	Sortase B (surface protein attachment)
Data01.09_00483	Transglycosylase (enzymatic activity)
Data01.09_00484	MerR family transcriptional regulator (DNA binding)
Data01.09_00485	Uncharacterised protein
Data01.09_00486	Uncharacterised protein
Data01.09_00487	Conjugative transposon protein (enzymatic activity)
Data01.09_00488	Tetracycline resistance, transcriptional regulator
Data01.09_00489	Iron-sulfur protein (4Fe-4S binding domain)
Data01.09_00490	Multi antimicrobial extrusion protein (drug transportation)
Data01.09_00491	DNA binding, transcriptional regulator
Data01.09_00492	ABC transporter family protein (membrane transport)
Data01.09_00493	ftsX-like permease family protein (membrane structure)
Data01.09_00494	Two-component response regulator (uncharacterised protein)
Data01.09_00495	Two-component sensor histidine kinase (environmental stimuli response)
Data01.09_00496	Cysteine-rich KTR (uncharacterised protein)
Data01.09_00497	RNA polymerase ECF-type sigma factor (RNA transcription)
Data01.09_00498	PemK family transcriptional regulator (cell growth regulation)
Data01.09_00499	Short C-terminal (SHOCT) domain (oligomerisation and nucleic acid binding)

Table adapted from a publication by Cairns *et al.*, (Cairns et al., 2015).

4.4.2 Phenotypic comparison of isolates from London hospitals

With the findings that isolates of cluster 1-UHL haplotypes contained a gene putatively associated with antimicrobial resistance (multidrug family ABC transporter permease) and a 49 kbp genetic region exclusive to the cluster 1-UHL isolates with genes potentially involved in sporulation and antimicrobial resistance (sortase B gene, the sporulation gene *spo0J* and genes associated with antimicrobial resistance; tetracycline resistance and a multi antimicrobial extrusion protein), phenotypic assays were performed to try and identify a phenotype unique to the UHL clonal isolates that may play a role in environmental persistence and/or transmission. Isolates were selected to identify characteristics that would contribute to the ability of the cluster 1-UHL isolates to persist and spread in a population and environment; a strain with increased speed of growth, sporulation, germination, resistant antibiogram and resistance to hospital disinfectant/s would be at advantage to persist and spread. Four London isolates were selected based on them having the most variation with regards to isolation date and geographical location; UHL-1, UHL-3, UHL-19 and CX-32.

4.4.2.1 Growth kinetics

In vitro growth curves were performed to identify any differences in growth kinetics between the UHL and non-UHL isolates. In order to measure growth rates, isolates were grown at 37 °C anaerobically in BHIS broth and cell densities were measured by monitoring the OD590 over time, this was performed in duplicate and with three independent biological experiments.

Overall the growth rates of strains UHL-3, UHL-19, CX-32 and the control strain M68 were similar to one another (Figure 4.4), all entering stationary phase at around six hours. However, the growth rate of UHL-1 was consistently slower entering stationary phase at around seven hours. At four hours, the cell density for UHL-19 was marginally higher compared to UHL-1 ($p \leq 0.05$) and at five hours, the cell density for UHL-19 and CX-32 was significantly higher ($p \leq 0.0001$) and for M68 was marginally higher ($p \leq 0.05$) compared to UHL-1. The cell density for UHL-3 was also marginally higher compared CX-32 ($p \leq 0.05$). The cell density for M68 and UHL-3 at six hours was significantly higher compared to UHL-1 ($p \leq 0.01$). At 24 hours, the cell density for CX-32 was significantly higher compared to M68, UHL-19 and UHL-1 ($p \leq 0.001$, $p \leq 0.0001$ and $p \leq 0.01$ respectively) and that for UHL-3 was marginally higher compared to M68 and significantly higher compared UHL-19 ($p \leq 0.05$ and $p \leq 0.001$ respectively).

These experiments show the historical isolate from UHL (UHL-1) to have a slower growth rate compared to the more recent isolates from UHL (UHL-3 and UHL-19), the reference strain M68 and the isolate from Charing Cross hospital (CX-32). However, no statistically significant differences in growth rate between UHL and non-UHL isolates were observed.

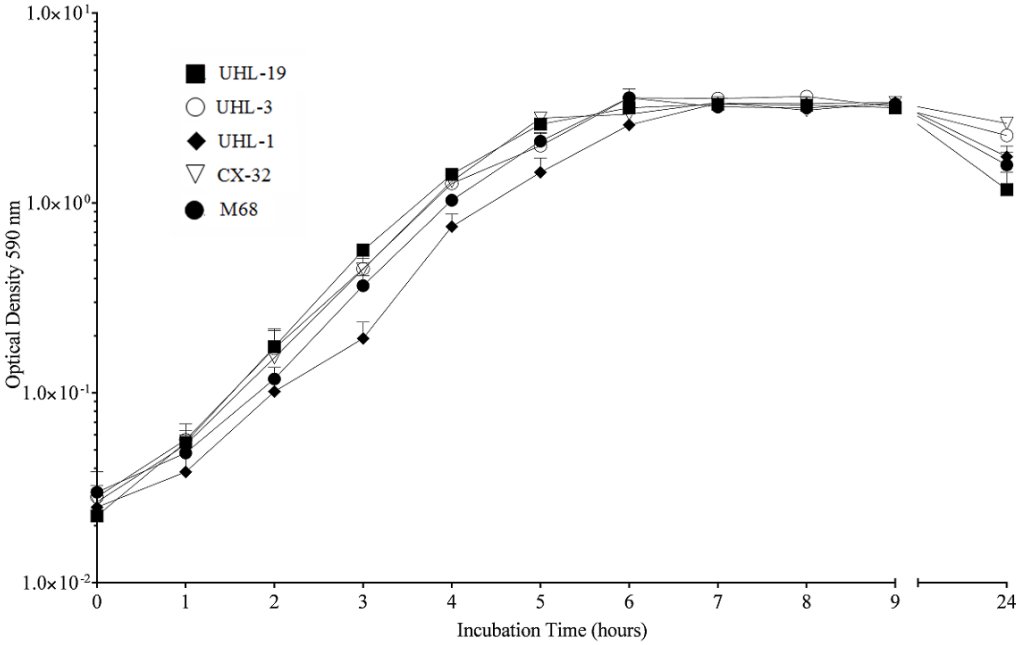


Figure 4.4: Growth kinetics for the London *C. difficile* phenotypic study isolates

The data represent the average of three independent experiments and duplicate technical replicates from separate cultures of five strains. Statistical significance was assessed using Two-Way ANOVA followed with Tukey's multiple comparison tests in GraphPad Prism. Although there was some variation between isolates, no statistically significant differences in growth rate between UHL and non-UHL isolates were observed that could explain the persistence of the cluster 1-UHL clone.

4.4.2.2 Sporulation

To investigate the rate at which the four London *C. difficile* phenotypic study isolates formed spores *in vitro*, the isolates were grown in medium supplemented with cysteine and yeast extract as the organism is known to sporulate well under these conditions (Sorg and Sonenshein, 2008). Cells were inoculated to fresh secondary cultures and grown anaerobically; samples were taken at 24 hours, 72 hours and 144 hours and were subjected to heat treatment at 60 ºC for 30 minutes to kill vegetative cells but not the spores. Colony forming units were enumerated on blood agar supplemented with sodium taurocholate which has been shown to induce germination of *C. difficile*.

The control strain M68 appears to have a slower sporulation rate than the other strains, but they all peak at a similar cfu by 144 hours (Figure 4.5). The sporulation rate of the control strain M68 increases from log 4 to log 5 between 24 hours and 72 hours then log 5 to log 6.5 at 72 hours to 144 hours. This is in comparison to the other strains that were already sporulating between log 5 to log 6.5 hours at 24 hours and 72 hours but then reached their peak sporulation and did not increase anymore between 72 hours and 144 hours. At 24 hours, CX-32 had a higher spore count compared to UHL-19 and the control strain M68 ($p \leq 0.001$ and $p \leq 0.0001$ respectively) and UHL-1 was significantly higher than control strain M68 ($p \leq 0.01$). At 72 hours, UHL-19 had a significantly higher spore count compared to control strain M68, UHL-3, CX-32 and a marginally higher spore count compared to UHL-1 ($p \leq 0.01$, $p \leq 0.01$, $p \leq 0.01$ and $p \leq 0.05$ respectively). UHL-19 had a higher spore count at 144 hours, though this was not statistically significant.

These spore assays showed variation in sporulation rates with the isolate from Charing Cross hospital (CX-32) producing more spores after 24 hours incubation compared with the other four isolates, and the environmental isolate from UHL produced more spores after 72 hours incubation. Four isolates were found to sporulate to higher cfus than the control strain M68 at 24 hours and 72 hours, however, no statistically significant differences in sporulation rate between cluster 1-UHL and non-cluster 1-UHL isolates were observed.

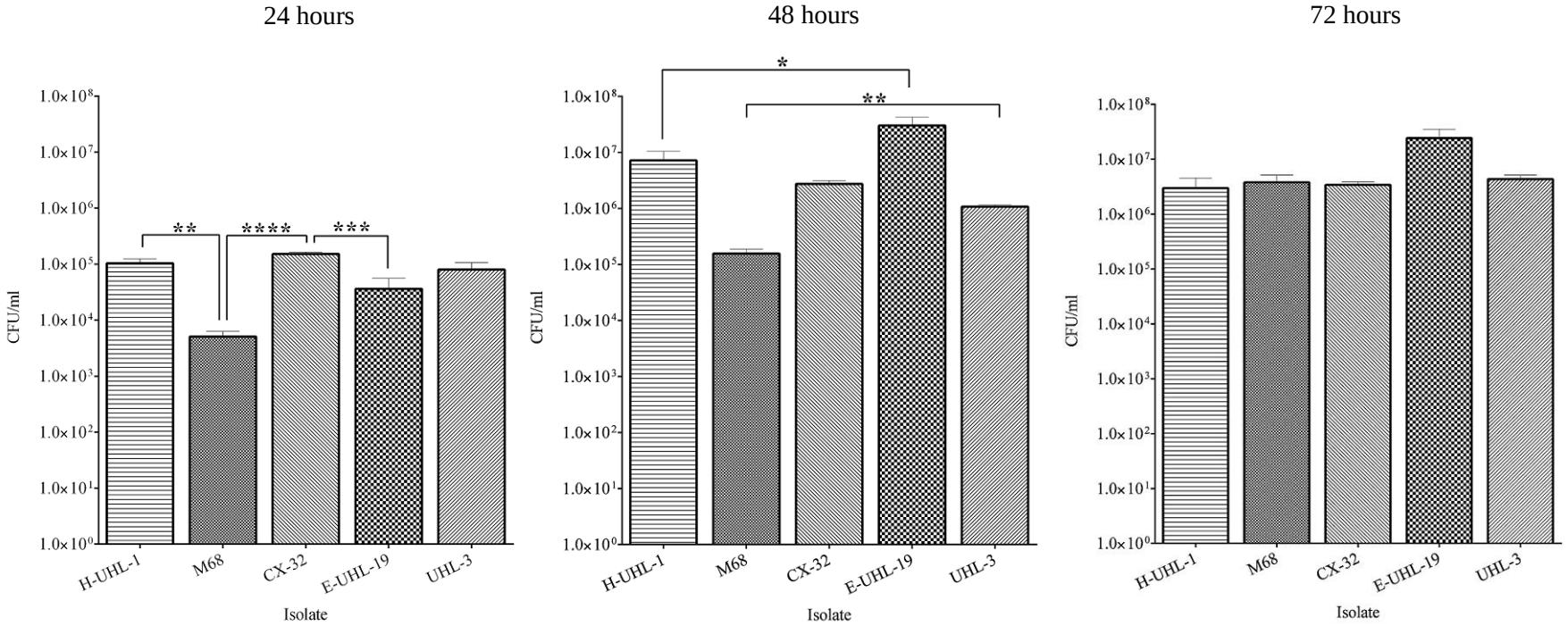


Figure 4.5: Total spores recovered at 24 hours, 72 hours and 144 hours for the London *C. difficile* phenotypic study isolates

The data represent the average of three independent experiments and duplicate technical replicates from separate cultures. Statistical significance was assessed using One-Way ANOVA followed with Tukey's multiple comparison tests in GraphPad Prism. Error bars indicate the standard error of the means and significant differences are marked with a bracket (* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$ and **** = $p \leq 0.0001$). Although there was some variation between isolates, no statistically significant differences in spore production between UHL and non-UHL isolates were observed that could explain the persistence of the cluster 1-UHL clone.

4.4.2.3 Percentage spore production

To estimate the rate of sporulation more accurately, and to take into account possible variation in cell growth rate, the sporulation rates were expressed as percentage of spores to vegetative cells and calculated as follows: [*N* spore count *100] / *N* all cell count. Percentage spore production for the study isolates are shown in Figure 4.6. The percentage spore production for the non-UHL isolate (CX-32) at 72 hours was significantly higher than the UHL isolate (UHL-3) but only marginally higher than the control strain M68 ($p \leq 0.01$ and $p \leq 0.05$ respectively). No significant difference in the percentage spore production was observed at 24 hours or 144 hours. No statistically significant differences in percentage spore production between cluster 1-UHL and non-cluster 1-UHL isolates were observed.

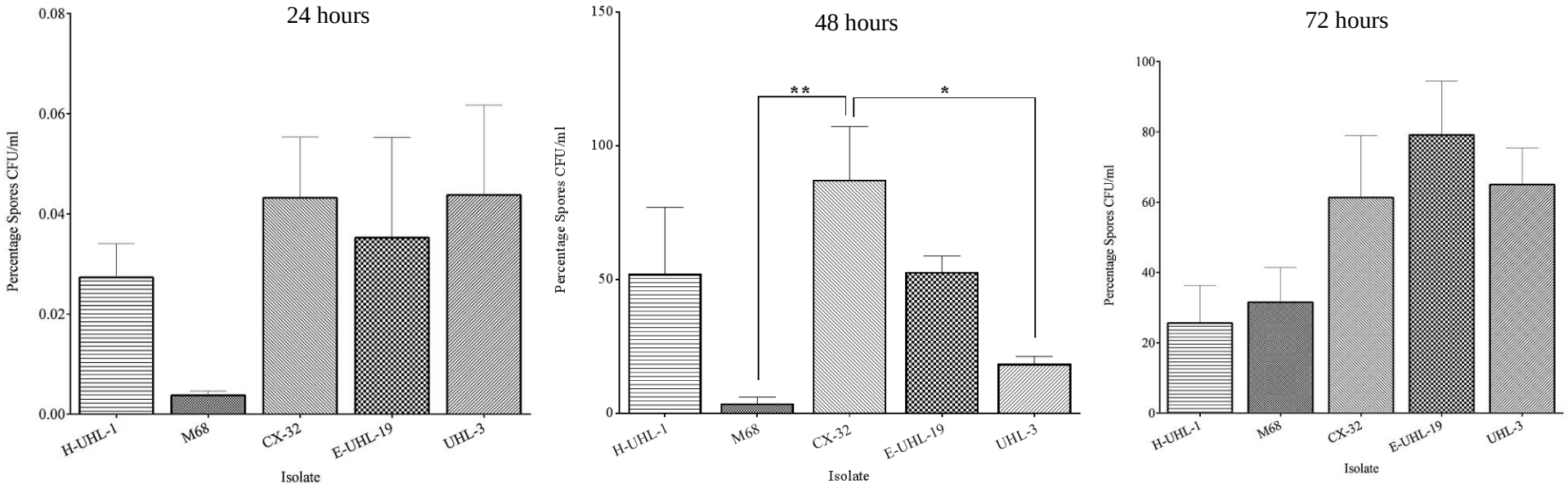


Figure 4.6: Percentage spores recovered at 24 hours, 72 hours and 144 hours for the London *C. difficile* phenotypic study isolates

The data represent the average of three independent experiments and duplicate technical replicates from separate cultures. Statistical significance was assessed using One-Way ANOVA followed with Tukey's multiple comparison tests in GraphPad Prism. Error bars indicate the standard error of the means and significant differences are marked with a bracket (* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$ and **** = $p \leq 0.0001$). Although there was some variation between isolates, no statistically significant differences in percentage spore production between UHL and non-UHL isolates were observed that could explain the persistence of the cluster 1-UHL clone.

4.4.2.4 Total cell recovery

To investigate the total cell recovery (vegetative cells and spores) characteristics, cfus were enumerated on blood agar supplemented with sodium taurocholate without heat treatment which allows for all cells, both vegetative cells and spores to be recovered. These counts are shown in Figure 4.7. No significant differences were observed in the cell counts at 24 hours and 144 hours, however, at 72 hours, the cell count for the control strain M68 were marginally higher than CX-32 ($p \leq 0.05$). No statistically significant differences in all cell recovery rates between cluster 1-UHL and non-cluster 1-UHL isolates were observed.

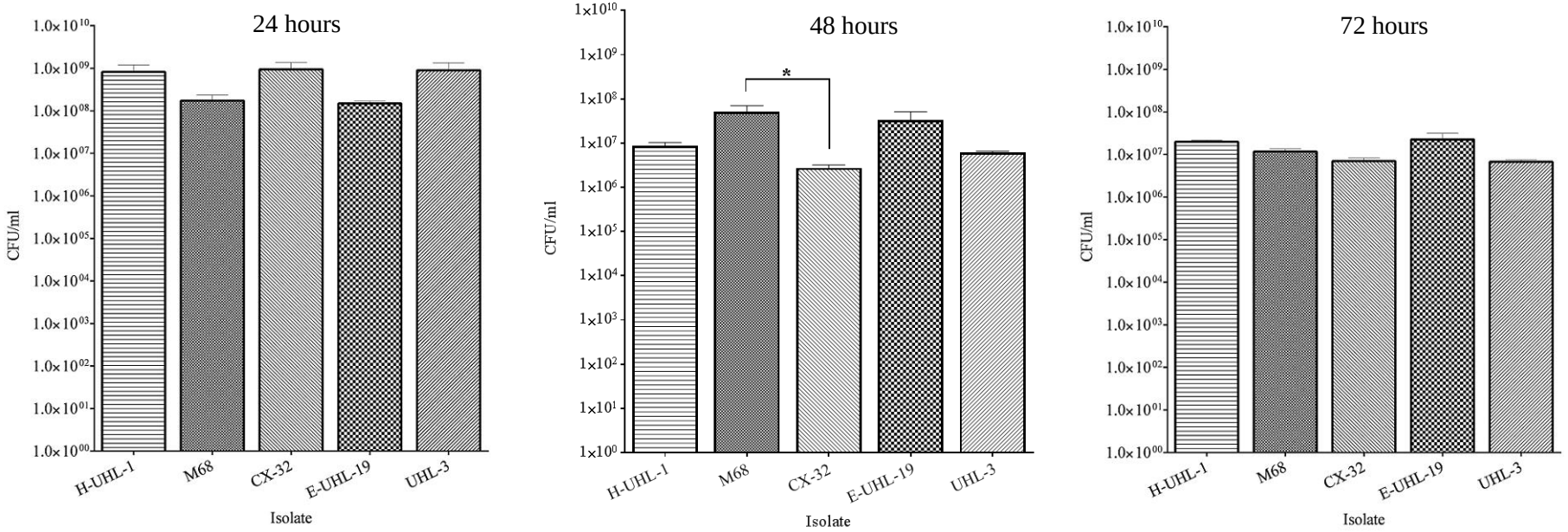


Figure 4.7: Total cells recovered at 24 hours, 72 hours and 144 hours for the London *C. difficile* phenotypic study isolates

The data represent the average of three independent experiments and duplicate technical replicates from separate cultures. Statistical significance was assessed using One-Way ANOVA followed with Tukey's multiple comparison tests in GraphPad Prism. Error bars indicate the standard error of the means and significant differences are marked with a bracket (* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$ and **** = $p \leq 0.0001$). Although there was some variation between isolates, no statistically significant differences in total cell production between UHL and non-UHL isolates were observed that could explain the persistence of the cluster 1-UHL clone.

4.4.2.5 Minimum inhibitory concentrations

MICs were determined using the broth dilution method against the antibiotics; metronidazole, vancomycin, erythromycin, lincomycin, fusidic acid, naladixic acid, rifampicin and rifamycin. MIC data are shown in Table 4.5.

Table 4.5: MIC assay data for the London phenotypic study isolates

Isolate	Cluster	Antibiotic and MIC value (µg/ml)							
		Metronidazole	Vancomycin	Erythromycin	Lincomycin	Fusidic Acid	Naladixic Acid	Rifampicin	Rifaximin
M68	2-M68	1.5	1	≥ 256	≥ 256	4	128	≤ 0.5	≤ 0.5
UHL-3	2-M68	1.5	1	≥ 256	≥ 256	4	128	≤ 0.5	≤ 0.5
UHL-1	1-UHL	1.5	1	≥ 256	≥ 256	4	128	≤ 0.5	2
UHL-19	1-UHL	1.5	1	≥ 256	≥ 256	4	128	≤ 0.5	≤ 0.5
CX-32	3	1.5	1.5	≤ 0.5	32	4	128	≤ 0.5	≤ 0.5

Five strains were tested using an antibiotic range of 0 (negative control) to 256µg/ml. The data represent the average of at least two independent experiments and duplicate technical replicates from separate cultures. Although there was variation in MICs between some isolates, no defined antibiogram was exclusive to the cluster 1-UHL isolates.

Although no antimicrobial susceptibility phenotype exclusive to the cluster 1-UHL isolates was identified, there were some notable variations. Compared with the other four isolates tested, the historic isolate UHL-1 had a higher MIC to the antibiotic rifaximin, one that is considered as intermediate resistance in the clinical setting.

This isolate (haplotype I) was shown in this study to have five additional SNPs compared to the core modern UHL sample haplotype A (Tables 4.2 and 4.3). Three SNPs were non-synonymous and found in; the *rpoB* gene, an ABC transporter and in a two-component response regulator resulting in a stop-codon. The remaining two SNPs were synonymous and found in a *gntR* gene (family transcriptional regulator) and the other a hypothetical protein of unknown function.

Isolate CX-32 from a patient at Charing Cross hospital had significantly lower MICs to the macrolide antibiotics; erythromycin and lincomycin compared with the other four isolates tested (the three isolates from cluster 1-UHL tested phenotypically were found resistant). *De novo* analysis in this study revealed all isolates in cluster 1-UHL and isolates M68, QM-34, NM-27, B-26 and B-33 to carry the well characterised transposon Tn*6194* which carries the *ermB* gene and is known to confer resistance to the macrolide class of antibiotics (Brouwer et al., 2013). Isolates WX-35, WX-36 and NP-25 and all isolates in cluster 3 which includes CX-32 were found not to carry or have lost this transposon.

4.4.2.6 Disinfectant susceptibility assays

Studies have found *C. difficile* spores are able to survive the recommended 1000 parts-per million strength for chlorine releasing agents (Dawson et al., 2011, Kulikovsky et al., 1975). This could explain the persistence of the cluster 1-UHL isolates. Actichlor Plus, a combined chlorine releasing agent plus a detergent was the commonly used disinfectant at UHL. To investigate the resistance to Actichlor Plus as a cause of the cluster 1-UHL isolates environmental persistence, the susceptibility to this disinfectant was determined for the four study isolates tested phenotypically

and the control strain M68. These data are shown in Figure 4.8. The disinfectant Actichlor Plus at a concentration of 1000 parts-per million demonstrated ≥ 99% efficacy with all five isolates tested. Furthermore, to take into consideration possible variation in germination and sporulation rates, the percentage survival was calculated which showed no significant variation (Figure 4.9). No statistically significant variation in the susceptibilities to Achtichlor Plus was observed between the isolates tested.

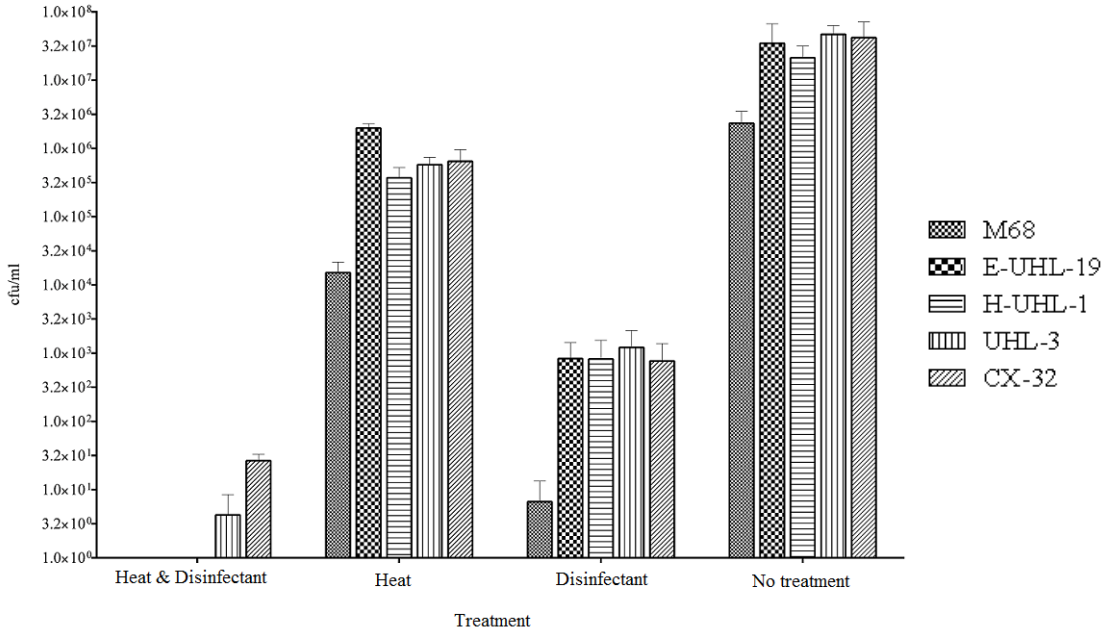


Figure 4.8: Total Achtichlor Plus assay data for the London *C. difficile* phenotypic study isolates

The data represent the average of three independent experiments and duplicate technical replicates from separate cultures. Statistical significance was assessed using One-Way ANOVA followed with Tukey's multiple comparison tests in GraphPad Prism. Error bars indicate the standard error of the means and significant differences are marked with a bracket (* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$ and **** = $p \leq 0.0001$). The disinfectant Actichlor Plus at a concentration of 1000 parts-per million demonstrated ≥ 99% efficacy with all five isolates tested though no statistically significant variation in susceptibility between isolates was observed.

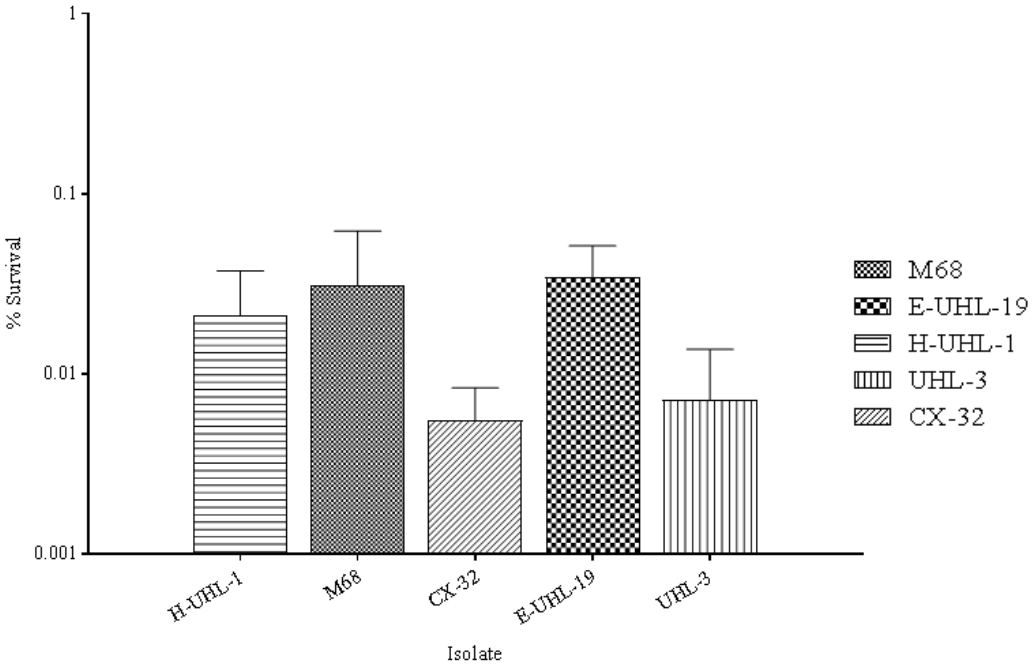


Figure 4.9: Percentage survival following Actichlor Plus treatment for the London *C. difficile* phenotypic study isolates

The data represent the average of three independent experiments and duplicate technical replicates from separate cultures. Statistical significance was assessed using One-Way ANOVA followed with Tukey's multiple comparison tests in GraphPad Prism. Error bars indicate the standard error of the means and significant differences are marked with a bracket (* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$ and **** = $p \leq 0.0001$). No statistically significant variation in the percentage survival following Achtichlor Plus treatment between the isolates tested was observed.

4.5 Discussion

Using PCR ribotyping, WGS and phenotyping assays, my study investigated a total of 37 *C. difficile* PCR ribotype 017 isolates (Table 4.1); 22 strains isolated from humans who at different times were present on the same hospital ward at UHL, two environmental strains isolated from the same hospital ward and 13 contemporaneous strains isolated from humans from other London hospitals.

My data shows that the *C. difficile* isolates from UHL are clonal and genotypically different to *C. difficile* isolates from other London hospitals included in my study. Phylogenetic analysis of SNPs revealed three diverging sub-lineages of *C. difficile* PCR ribotype 017 (cluster 1-UHL, cluster 2-M68 and cluster 3) with 96.3% of SNP loci found exclusively in only cluster 1-UHL. Through *de novo* assembly analysis, a 49 kbp genetic region that further differentiated cluster 1-UHL isolates from cluster 2-M68 and cluster 3 isolates was revealed. Although not all virulence genes are on mobile genetic elements, mobile genetic elements commonly carry virulence genes (Pallen and Wren, 2007) and the presence of putative virulence genes within this 49 kbp genetic region suggests that it may be a transposon-like putative mobile genetic element. To prove it is a transposon-like mobile genetic element and has been acquired by the clonal strains, the sequences spanning the insertion site of the 49 kbp genetic region could be sequenced using long read sequencing methodologies such as MinION or PacBio (Table 1.4, Page 61). This would confirm element and location and would help determine if the 49 kbp genetic region is similar to a known transposable element or meets the criteria to be assigned as a new transposable element, i.e. evidence of excision from the chromosome and presence of accessory genes not associated with transfer and which may encode functions that contribute to

a survival advantage, commonly antimicrobial resistance (Roberts et al., 2008, Campbell et al., 1979). Additionally, to meet the definition of a transposable element "specific DNA segments that can repeatedly insert into one or more sites in one or more genomes" (Roberts et al., 2008), bacterial mating experiments may also be attempted to confirm the genetic region is mobile and the phenotype is transferable (Brouwer et al., 2013, Hussain et al., 2005, Roberts et al., 2001).

A non-synonymous SNP in the *bioB* gene (Appendix 3, Page 256) was present in all isolates of cluster 1-UHL and cluster 3. This is one of four genes that play a major role in the biotin biosynthetic pathway (Satiaputra et al., 2016). Biotin mediates the transport of CO_2 in many vital metabolic reactions and is essential for life. Interestingly, studies have shown decreased available biotin results in increased toxin synthesis by *C. difficile* (Yamakawa et al., 1996). This SNP may affect biotin production, limiting its availability and in-turn increasing toxin synthesis. It would be worth investigating this phenomenon by repeating the assay reported in the study by Yamakawa *et al.,* which is dated. Here we could study the effect of biotin on toxin production between isolates from the different London clusters.

With the identification of a unique genotype and to identify a phenotype exclusive to the UHL outbreak isolates that may play a role in environmental persistence and/or transmission, phenotypic assays were performed on five isolates (UHL-1, UHL-3, UHL-19, CX-32 and the control strain M68) to place these strains into geographical and historical context. Variations in growth kinetics, sporulation, sporulation rates, vegetative cell counts and antimicrobial susceptibility were observed, though none

were statistically significant that would differentiate cluster 1-UHL and non-cluster 1-UHL isolates. There were no significant variations in susceptibility to the disinfectant Actichlor Plus between those isolates tested, indicating therefore that there was no potential survival advantage observed for the UHL isolates with regards to susceptibility to this disinfectant. A phenotype exclusive to the UHL outbreak isolates that may play a role in environmental persistence and/or transmission was not identified.

Although no statistically significant difference in antimicrobial susceptibility differentiating cluster 1-UHL and non-cluster 1-UHL isolates was found, there were some interesting observations. The three isolates from cluster 1-UHL (UHL-1, UHL-3 and UHL-19) were found phenotypically resistant and the isolate from Charing Cross hospital (CX-32) had significantly lower MICs to the macrolide antibiotics erythromycin and lincomycin. *De novo* analysis showed that isolate CX-32 did not carry or had lost the well characterised transposon Tn*6194* which carries the *ermB* gene of which is known to confer resistance to the macrolide class of antibiotics whereas all isolates in cluster 1-UHL and isolates M68 did carry this transposon. This may explain the variation in susceptibility of these isolates to erythromycin as the *ermB* gene is required for resistance to erythromycin and lincomycin (Wust and Hardegger, 1983, Hachler et al., 1987, Farrow et al., 2000). As previously described, CDI is associated with prior antimicrobial therapy and patients > 65 years of age. A common cause of hospitalisation in elderly patients is for the management of community acquired pneumonia; this is frequently treated with the macrolide class of antimicrobials to cover atypical pathogens and/or those patients who are penicillin allergic. Of note the reference strain M68 was also resistant to the macrolide

antibiotics and this strain was associated with outbreaks of CDI in Ireland (Drudy et al., 2007b, Drudy et al., 2007c). It may be hypothesised that this resistance may have contributed to the establishment of the outbreak at UHL. A clonal cluster of *C. difficile* that is exclusively resistant to an antimicrobial may have implications for antimicrobial stewardship. The occurrence of a clonal cluster of isolates with a specific antibiogram combined with time of sampling, patient location with regards to time and place and details of antimicrobial prescriptions will enable clinicians to decide if they need to alter local antimicrobial stewardship.

The historical isolate UHL-1 had a higher MIC to the antibiotic rifaximin, one that is considered as intermediate resistance in the clinical setting. This isolate (haplotype I) was also found to have five additional SNPs compared to the core modern UHL sample haplotype A (Tables 4.2 and 4.3). Three of these were non-synonymous and found in; the *rpoB* gene, an ABC transporter and in a two-component response regulator resulting in a stop-codon. Rifaximin resistance has been associated with SNPs in the *rpoB* gene of *C. difficile* (O'Connor et al., 2008, Curry et al., 2009) and so it is possible that the higher MIC to rifaximin in this historical isolate is a consequence of the non-synonymous SNP in the *rpoB* gene. Rifaximin is a treatment option for recurrent CDI, but rarely used for this indication, however, it is used for other bacterial infections in hospitalised patients. It would be worth investigating if the higher MIC to rifaximin is a consequence of the non-synonymous SNP in the *rpoB* gene since this resistance may have given the historical isolate a selective advantage allowing the persistence of the UHL cluster like that observed with the epidemic PCR ribotype 027 strain and its resistance to the fluoroquinolone class of antibiotics (He et al., 2013).

The transposon-like putative mobile genetic element identified in this study encodes a gene putatively associated with tetracycline resistance; a tetracycline resistance transcriptional regulator (Table 4.4). Transcriptional regulators are known to control the expression of genes associated with tetracycline resistance, for example the *tet* class of genes (Ramos et al., 2005). The presence of a putative tetracycline resistance transcriptional regulator may play an indirect role in resistance to tetracycline; however, the isolates selected for phenotypic study were not tested for susceptibility to tetracycline.

With the observations associated with the macrolides, rifaximin and tetracycline, it may be hypothesised that prior administration of these antimicrobials may have contributed to the cluster 1-UHL outbreak; it would be plausible to investigate if use of these antibiotics were a contributory cause of the outbreak at UHL. A timeline of CDI cases and prescriptions of these antibiotics would illustrate this and may alter local antimicrobial stewardship as well as susceptibility testing against tetracycline.

There were limitations in this study. Firstly, the choice of control isolate used for detection of SNPs and *de novo* analysis of insertions and deletions; obviously results will vary depending upon which control strain is used for comparison. This study compared the genomes of the study isolates with the reference genome of the *C. difficile* PCR ribotype 017 strain M68 GenBank accession number FN668375 (He et al., 2010). This was the most appropriate strain to use as a control for this study since only two reference genomes of *C. difficile* PCR ribotype 017 were available and there was a better understanding of strain M68 with it being involved in outbreaks in Ireland (Drudy et al., 2007a, Drudy et al., 2007b). Another limitation in

this study is the methodology applied for the detection of SNPs. The method obviously reads only the parts of the genome that match the control strain that it is assembled against; if there are regions of the genome that are not present in the control strain, then these regions are not analysed for SNPs. The methodology also does not read regions of repetitive DNA (sequences that are similar or identical to sequences elsewhere in the genome); therefore, any differences in repetitive regions between strains are not detected. Another limitation in this study is the methodology applied for the identification of unique genomic regions indicative of horizontal gene transfer; the method is programmatic and relies on visual inspection of comparisons using the ACT application. This has limitations in that it is manual and when analysing multiple strains, there is potential for user error.

4.6 Conclusion

This is the first study in the UK to investigate the phylohistory of isolates from hospitalised patients with CDI due to *C. difficile* PCR ribotype 017. It is a snapshot of isolates and suggests that a clonal strain of *C. difficile* toxin A-B+ PCR ribotype 017 has persisted in a ward at one London hospital for at least five years and which is different to *C. difficile* PCR ribotype 017 strains that are circulating in other London hospitals. The clonal strain was found to exclusively harbour a 49 kbp genetic region suggestive of being a transposon-like putative mobile genetic element however, no phenotype exclusive to this cluster was revealed. The *C. difficile* PCR ribotype 017 lineage, with its unique toxin profile and unusual global prevalence, is understudied. My study shows that there are existing questions about the population structure and epidemiology of *C. difficile* toxin A-B+ PCR ribotype 017 strains.

Chapter 5

The global phylogeny of

C. difficile PCR ribotype 017

5 The global phylogeny of *C. difficile* PCR ribotype 017

5.1 Statement of contribution

Isolate selection, PCR ribotyping and genomic DNA extractions were performed by M. Cairns. WGS were performed both at the WTSI by D. Harris and at the LSHTM by M. Cairns. The collection of epidemiology data were performed by M. Cairns. SNP calling were performed by Dr Mark Preston. *De novo* assembly were performed by M. Cairns with help from Dr Richard Stabler. MIC assays were performed by Catherine Hall (LSHTM). Data interpretation was performed by M. Cairns.

5.2 Introduction

With testing a larger collection of strains (n=385) of multiple PCR ribotypes isolated from multiple sources and geographical locations using MLST, chapter 3 confirmed that *C. difficile* is made up of five phylogenetic lineages; PCR ribotypes 017, 023, 027, 078 and a heterogeneous grouping of mixed PCR ribotypes. The MLST data found no association between sequence type and geographical location or source. The lineage associated with PCR ribotype 027 has been investigated further using WGS and identified the presence of two genetically diverse global sub-lineages (He et al., 2013), however, less is known about the in-depth phylogeny of the other four lineages of *C. difficile*.

Historically, *C. difficile* PCR ribotype 017 strains were identified in outbreaks of CDI in Asia; it is thought that PCR ribotype 017 spread from Asia to Europe and

other continents (Hawkey et al., 2013, Collins et al., 2013) however, this hypothesis has never been tested. In chapter 4, I investigated PCR ribotype 017 on a local scale where WGS identified a clonal strain from a single ward in a London hospital. WGS can also be applied to a larger collection of strains from multiple sources isolated over a large time scale to elucidate the phylogeny of a species like that applied to PCR ribotype 027 (He et al., 2013). No study had yet investigated the phylogeny of a large, global collection of PCR ribotype 017 strains using WGS, therefore, it is unknown if there are sub-lineages of PCR ribotype 017 like that seen with PCR ribotype 027. WGS would provide far superior discrimination compared with MLST.

The two lineages of PCR ribotype 027 identified through WGS had emerged in USA within a relatively short period of time after acquiring the same fluoroquinolone resistance conferring mutation (He et al., 2013). This implies that antimicrobial resistance played a significant role in driving the global spread of PCR ribotype 027. Reduced susceptibility and resistance to the first-line antimicrobial agents metronidazole and vancomycin have been described in *C. difficile* although the mechanisms of resistance are still not completely understood (Pelaez et al., 2008, Brazier et al., 2001, Chong et al., 2014, Goudarzi et al., 2013, Adler et al., 2015). It has been suggested that resistance to metronidazole is attributable to mutations within multiple genomic loci, including genes responsible for altered iron metabolism (Chong et al., 2014) whilst resistance to vancomycin may be due to amino acid changes in the peptidoglycan biosynthesis (Leeds et al., 2014). Reduced susceptibility to antibiotics commonly used for bacterial infections (other than *C. difficile*) not only contribute to the occurrence of CDI, but also in driving the evolution of *C. difficile* and emergence of new strains and lineages, like that

described with PCR ribotype 027 and fluoroquinolone resistance (He et al., 2013). Therefore, investigating other antimicrobials not routinely used in the treatment of CDI for resistance is useful for surveillance purposes. Resistance to the fluoroquinolone class of antimicrobials is attributable to mutations in the quinolone resistance determining region of the *gyrA* and/or *gyrB* genes (Ackermann et al., 2001, Drudy et al., 2006, Drudy et al., 2007d). Resistance to rifampicin has also been observed in *C. difficile* and this is due to alterations in the *rpoB* gene encoding the RNA polymerase (Freeman et al., 2015).

Antimicrobial resistance is confirmed phenotypically; an organism is grown in the presence of varying concentrations of an antimicrobial and the MIC is determined (Andrews, 2001). The most common phenotypic methodologies are the agar dilution and episilometer test (Etest) however; these methods are laborious and time-consuming. To overcome this, the molecular detection of known antimicrobial resistance determinants is employed. In the diagnostic setting, commercial target-specific PCR platforms for the molecular detection of organisms and antimicrobial genotypes such as the Xpert® platform mentioned in chapter 3 are available. This platform and similar have advanced the routine diagnostic microbiology laboratory improving the sensitivity of target detection and turn-around times of results. This is beneficial for organisms that possess fastidious growth requirements i.e. *Mycobacterium tuberculosis* and *C. difficile*. The molecular detection of a pathogen with a genotype associated with antimicrobial resistance in a diagnostic microbiology setting is also useful when prompt bed management is required. For example, the prompt detection of patients colonised with MRSA allowing infection control interventions such as cohorting and decolonisation (Jeyaratnam et al., 2008).

Additionally, the clinical management of a patient with sepsis is likely improved whereby reducing exposure to unnecessary broad-spectrum antibiotics would reduce the selective pressure for the development of antimicrobial resistance. However, such PCR platforms that are used in diagnostic microbiology laboratories are target specific and therefore limited to providing only the detection of that specifically sought.

As well as SNPs that confer antimicrobial resistance, mobile genetic elements such as transposons are also important whereby they often carry genes associated with antimicrobial resistance (Wiedenbeck and Cohan, 2011). Through *De novo* assembled genome analysis, the clonal strain of PCR ribotype 017 identified in chapter 4 was found to exclusively possess a 49 kbp genetic region suggestive of being a transposon-like putative mobile element which would not have been revealed by conventional typing methodologies.

WGS allows the unbiased analysis of the bacterial genome to detect putative antimicrobial resistance determinants, differentiate between varying mechanisms resulting in the same resistance pattern and identify new resistance determinants all in a single assay. Although WGS is not routinely used to detect a bacterium's antimicrobial phenotype, high concordance between genotypic and phenotypic resistance in multiple pathogens has been shown (Zankari et al., 2013, Read and Massey, 2014, Palmer and Kishony, 2013). Therefore WGS acts as a useful tool to predict phenotypic antimicrobial resistance which is useful diagnostically for a patient and also for surveillance and identification of clonal expansions that may be driven by antimicrobial resistance. PCR ribotype 017 has been associated with higher

levels of antimicrobial resistance compared with other PCR ribotypes (Drudy et al., 2007b), however, it is not known if there are sub-lineages of PCR ribotype 017 and if there are, if antimicrobial resistance played a role in their spread like with the sub-lineage of PCR ribotype 027 (He et al., 2013).

C. difficile is known to cause similar infection in both humans and animals and it is well documented that PCR ribotype 078 is the predominant PCR ribotype isolated from animal species with CDI (Keel et al., 2007, Jhung et al., 2008, Rupnik et al., 2008). Although *C. difficile* is not a proven food-borne pathogen, there is evidence that the same strain can cause symptomatic disease in both pigs and humans (Debast et al., 2009) and data from chapter 3 indicate that strains isolated from humans, animals and food origins are not phylogenetically distinct by MLST. With a global collection of *C. difficile* PCR ribotype 017 strains, it is therefore plausible to hypothesise that *C. difficile* PCR ribotype 017 isolated from human and non-human sources are not distinct by WGS.

5.3 Hypotheses of the research described in this chapter

The data in chapter 4 suggests a clonal strain of *C. difficile* toxin A-B+ PCR ribotype 017 has persisted in London for five years. This clonal strain was found to exclusively harbour a 49 kbp genetic region suggestive of being a transposon-like putative mobile genetic element. The PCR ribotype 017 lineage, with its unique toxin profile and unusual global prevalence, is understudied. There are existing questions about the population structure and epidemiology of toxin A-B+ PCR ribotype 017 strains; by resolving these may have implications on the awareness of this PCR

ribotype and whether this needs to be heightened. This study was designed to test the following hypotheses:

Chapter 5: hypothesis 1

One or more SNP/s and/or insertions or deletions (including those associated with antimicrobial resistance) are associated with the global spread of *C. difficile* PCR ribotype 017.

Chapter 5: hypothesis 2

C. difficile PCR ribotype 017 isolated from different geographical origins are phylogenetically distinct by WGS.

Chapter 5: hypothesis 3

C. difficile PCR ribotype 017 isolated from human and non-human sources are not phylogenetically distinct by WGS.

Chapter 5: hypothesis 4

C. difficile PCR ribotype 017 originated in Asia.

5.4 Results

In order to test my hypotheses, I aimed to collate the first global collection of *C. difficile* PCR ribotype 017 strains isolated from a broad range of hosts and geographical locations. In order to do this, I established a network of global collaborators which was initially started by networking with fellow researchers at international conferences and poster sessions. A total of 384 *C. difficile* isolates were collated from collaborators (Page 27).

Upon receipt of strains from collaborators all isolates were propagated on appropriate culture media (Sections 2.2.2, Page 70 and 2.2.3, Page 71). Before inclusion in the collection of strains, I first confirmed they were PCR ribotype 017 by PCR ribotyping and agarose gel electrophoresis (Section 2.3.1, Page 77). A total of 277 isolates were confirmed to be PCR ribotype 017 and were included for study. This collection of 277 strains were from; human (n=251), bovine (n=9), canine (n=11), equine (n=4) and hospital ward environment (n=2) with isolation dates between 1990 and 2013 and included isolates from six global continents (Appendix 2, Page 249). Once confirmed, in preparation for WGS, DNA was quantified (Section 2.3.3.2, Page 84) and WGS was performed using Illumina Sequencing Technology (Sections 2.3.3.3, Page 85 and 2.3.3.4, Page 88). SNPs and haplotypes were identified by sequence mapping and assembly (Sections 2.3.3.5, Page 85 to 2.3.3.9, Page 92 inclusive) and *de novo* genome assembly analysis and visual inspection using ACT was performed to identify genomic deletions and insertions (Section 2.3.3.10, Page 92). Software tools were used to produce phylogenetic trees to enable visual comparisons of SNP differences in the global collection of *C. difficile* isolates (Section 2.3.9, Page 91) and phenotypic antimicrobial susceptibility testing was

performed on select isolates (S- 017.72, WA 1514, S- 017.92, S- 017.27, S- 017.74, I 6 and 01-116) to identify an antibiogram that may be exclusive to a particular group of isolates and/or one which correspond with genotypic data (Section 2.2.7, Page 73). To investigate the phenomenon that *C. difficile* PCR ribotype 017 originated in Asia, global transmission events were inferred from bayesian evolutionary and geo-temporal analyses of the collection of 277 *C. difficile* PCR ribotype 017 isolates (Section 2.3.9, Page 91).

5.4.1 The phylogeny of a global collection of *C. difficile* PCR ribotype 017 strains using PCR ribotyping and WGS

5.4.1.1 PCR ribotyping

All 277 isolates in this study were confirmed to be PCR ribotype 017 by PCR ribotyping and agarose gel electrophoresis. Reference strains R20291 (PCR ribotype 027) and M68 (PCR ribotype 017) were used as controls and all isolates were assumed to be toxinotype VIII, A-B+ based on their being PCR ribotype 017. There are no known reports of PCR ribotype 017 having any other toxin profile other than toxinotype VIII, A-B+.

5.4.1.2 Inference of phylogeny using SNP data

To investigate the phylogeny and clonality of the isolates in this study, after sequence quality control and mapping to the control strain M68 reference genome, SNP analysis revealed 1288 bi-allelic SNPs with 311 present in greater than 1% of samples and greater than 1 bp from an insertion or deletion. Of these SNPs, 65.6%

(204/311) were non-synonymous (Appendix 3, Page 256), 17.7% (55/311) synonymous and 16.7% (52/311) were present in non-coding regions of the genome. Twelve SNPs affected stop-codons; eleven non-synonymous and one synonymous (Table 5.1). SNP data revealed 109 haplotypes containing between 0 and 52 SNPs with 76.5% (212/277) of isolates having between 10 and 35 SNPs compared to the reference genome.

Table 5.1: Stop-codon associated SNPs for the global isolates

Position in the M68 genome	M68 Reference Codon	Alternative Codon	Non-Synonymous / Synonymous	Gene	Predicted Function and/or Potential Impact	No. of isolates with SNP
1907433	TAA	GAA	NS	*msrAB*	Peptide methionine sulfoxide reductase	256
1204039	GGA	TGA	NS	M68_01144	Hydrolase	36
3304067	TCA*	GCA*	NS	*Sigma-54*	Controls expression of nitrogen related genes	29
132573	TGG	TGA	NS	M68_00168	Amino acid aminotransferase	16
3399853	TTG*	TAA*	NS	M68_03193	Ca2+/Na+ antiporter	13
3704987	CCA*	TGA*	NS	*sleB*	Spore-cortex-lytic protein	8
4157880	TTG*	TAA*	NS	M68_03851	PTS system, IIc component	6
557896	TTC*	TAA*	NS	*feoB3*	Ferrous iron transport protein B	3
1359584	GGA	TGA	NS	M68_01270	Extracellular solute-binding protein	3
1916756	AAT*	GAT*	S	M68_01782	Unknown	3
3402470	CAA	TAA	NS	*plfB*	Formate acetyltransferase	3
3784055	TTC*	TAA*	NS	M68_03513	Penicillin-binding protein	3

* = encoded on reverse strand

Table adapted from a publication by Cairns *et al.,* (Cairns et al., 2017).

A maximum-likelihood phylogenetic tree based on the 1288 SNPs was generated with geographical and temporal data combined in a phylogeographic analysis generated by mclust software (Figure 5.1). Of the 1288 SNPs, 76% (977/1288) had a MAF of ≤ 1% and/or were within 1 bp of an insertion or deletion. To control for false positive identification of SNPs (these SNPs may mask the true phylogeny of PCR ribotype 017) phylogenetic trees with and without these SNPs were generated. The inclusion of 977 SNPs only had a minor effect on the overall phylogenetic tree.

The mclust software divided the isolates into 20 clusters by cluster analysis (Figure 5.1). The combination of epidemiological, genetic and temporal data, cluster 2 (C2) was defined as the best fit and demonstrates the presence of two genetically diverse sub-lineages; SL1 and SL2.

By combining Bayesian evolutionary analysis with geo-temporal modelling of the 277 isolates, the evolution of *C. difficile* PCR ribotype 017 can be orientated though time (Figures 5.1 and 5.2). The analysis depicted in these figures indicate a split from SL1 (upper samples) into SL2 (lower samples) c1990. SL1 was more closely related to the control strain M68 of the two sub-lineages (Figure 5.2). Both Figures 5.1 and 5.2 also illustrate that isolates from different continents are amongst both SL1 and SL2 indicating that both sub-lineages are global in nature.

The two sub-lineages were differentiated by four SNPs; one present in a non-coding region and three non-synonymous SNPs (Table 5.2).

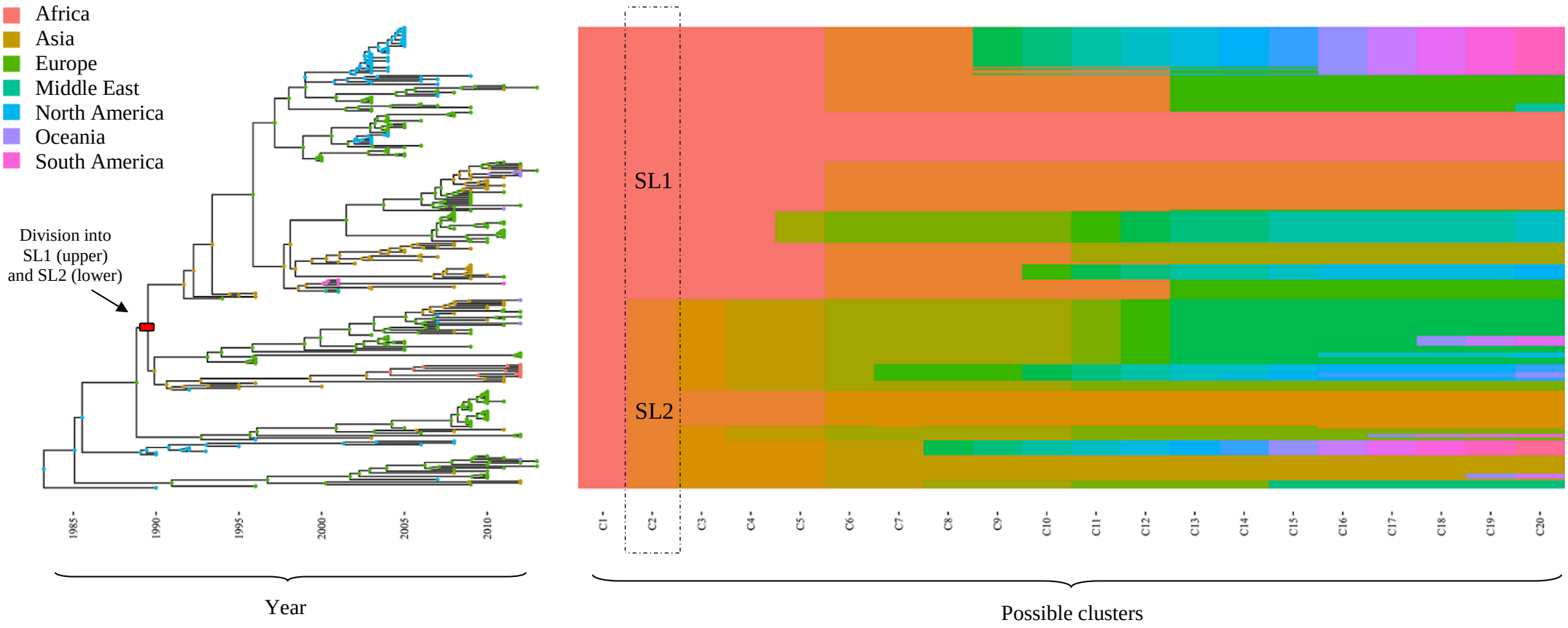


Figure 5.1: Temporal phylogeny and maximum likelihood clusters for the global *C. difficile* PCR ribotype 017 isolates

Figure adapted from a publication by Cairns *et al*., (Cairns et al., 2017). Temporal phylogeny and maximum likelihood clusters of the 277 *C. difficile* PCR ribotype 017 isolates. The phylogenetic tree illustrates how the collection of isolates has evolved over time with continents identified by colour and division between SL1 and SL2 identified. The heat map on this figure splits the collection of 277 isolates into 20 possible clusters (C1-C20) based on maximum likelihood generated by mclust software. Cluster 2 (C2) was selected as the best fit based on epidemiological, genetic and temporal data associated with the isolates.

Table 5.2: Lineage defining SNPs for the global isolates

Position	Amino Acid	Reference Base	Alternative Base	*NS/ S/NC	Gene	Predicted Function and/or Potential Impact
650374	19	A	G	NS	*merR*	Altered response to environmental stimuli
900866	.	C	T	NC	.	.
2914248	257	A	G	NS	*dacF*	Reduced resistance to heat or β-lactam antibiotics
3604289	329	C	A	NS	Hypothetical	Unknown

*Non-Synonymous (NS), Synonymous (S), Non-Coding (NC)

Table adapted from a publication by Cairns *et al.,* (Cairns et al., 2017).

5.4.1.3 *De novo* genome assembly analysis

In chapter 4, *de novo* assembled genomes revealed a 49 kbp transposon-like putative mobile genetic element exclusive to a hospital clonal cluster of PCR ribotype 017 strains. To investigate the global collection of strains in this study for this and/or other insertions or deletions, programmatic and visual inspection of the comparisons was performed. *De novo* assembly of each isolate and comparison to the control strain M68 revealed 56 regions of DNA between ~4 and ~61.5 kb that were absent in the control strain M68. These had 34 different insertion sites. Additionally, regions of DNA of between ~8 and ~29 kb at six sites were found absent from multiple strains but present in the control M68 strain. Details of genes of interest found in either an insertion or deletion are depicted in Figure 5.2 and listed in Table 5.3.

Bayesian evolutionary analysis and geo-temporal modelling was combined with a heat map to also depict presence/absence of insertions and antimicrobial resistance associated SNPs in relation to the isolates and continent (Figure 5.2). The deletions

and insertions are well distributed geographically and temporally and the 49 kb transposon-like putative mobile genetic element identified in chapter 4 was also found to insert at a different site in single isolates from Canada, the USA and the UK with isolation dates of 2006, 2006 and 2011 respectively.

A single SNP in the PaLoc region resulting in a synonymous change was found in five Korean strains in SL2 isolated between 2004 and 2008, however, visual inspection of the comparisons in the ACT application revealed no variations in the genes *tcdA* or *tcdB* compared with the reference control strain M68.

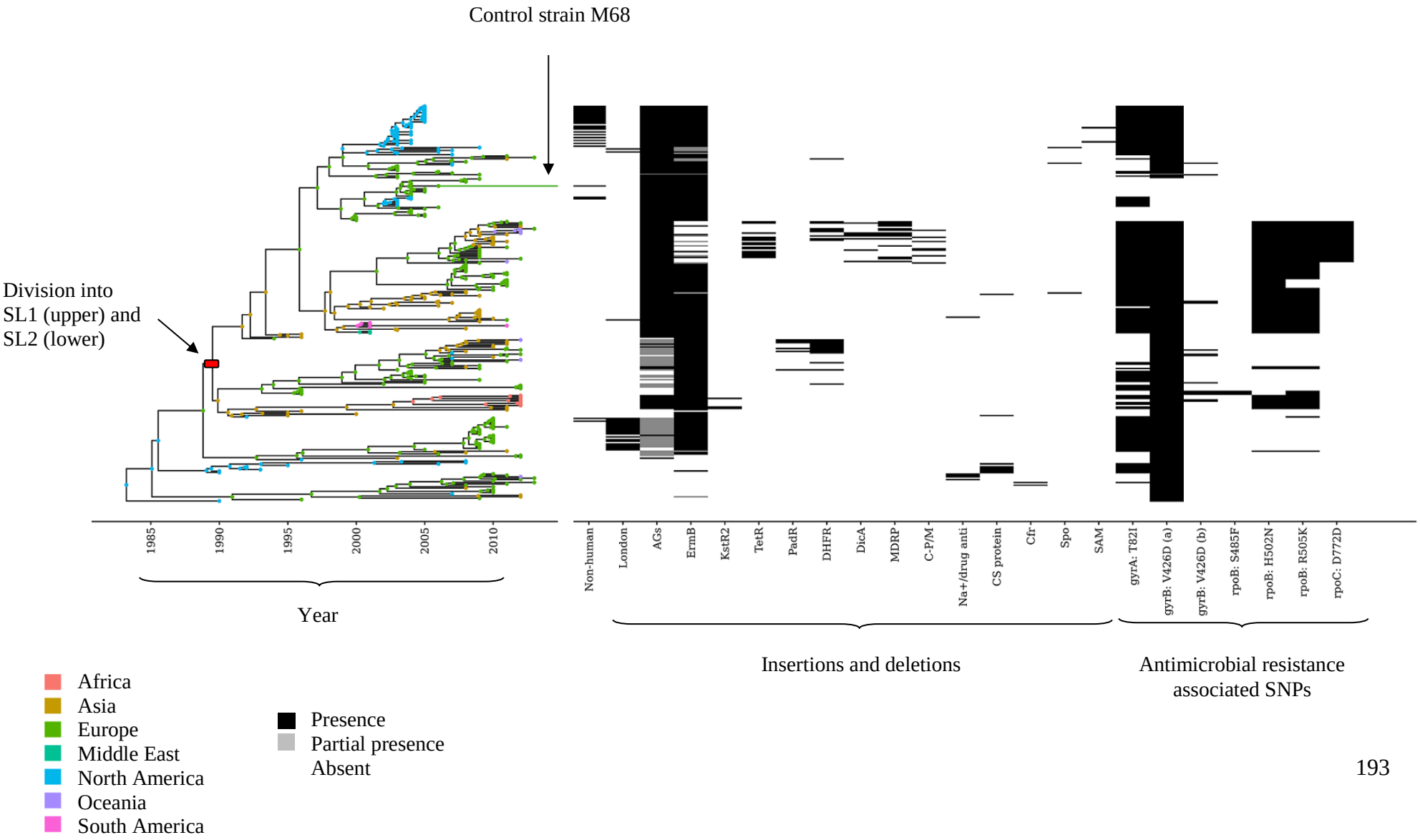

Control strain M68
Division into
SL1 (upper) and
SL2 (lower)
1985
1990
1995
2000
2005
2010
Year
Non-human
London
AGs
ErmB
KstR2
TetR
PadR
DHFR
DicA
MDRP
C-P/M
Na+/drug anti
CS protein
Cfr
Spo
SAM
gyrA: T82I
gyrB: V426D (a)
gyrB: V426D (b)
rpoB: S485F
rpoB: H502N
rpoB: R505K
rpoC: D772D
Insertions and deletions
Antimicrobial resistance
associated SNPs
Africa
Asia
Europe
Middle East
North America
Oceania
South America
Presence
Partial presence
Absent

Figure 5.2: Bayesian evolutionary analysis of the global *C. difficile* PCR ribotype 017 isolates

Figure adapted from a publication by Cairns *et al*., (Cairns et al., 2017). Bayesian evolutionary analysis of 277 global *C. difficile* PCR ribotype 017 isolates based on core-genome SNPs against the control strain M68. Using a geo-temporal model the evolution of the PCR ribotype 017 isolates can be orientated though time. The analysis indicates a split from SL1 (upper samples) into SL2 (lower samples) c1990, with the control strain M68 in SL1. The introduction of resistance associated SNPs (such as in *rpoC*) fall within closely related groups in the phylogeny. The heat map depicts the sub-lineage, presence/absence of insertions and antimicrobial resistance associated SNPs in relation to the isolates and continent. Details for insertions and deletions are shown in Table 5.3.

Table 5.3: Genes of interest found in either an insertion or deletion when compared to the M68 reference strain referred to in Figure 5.2

Abbreviation	Insertion/Deletion	Number of isolates	Sub -lineage	Gene/s of Interest/Putative Function
London insertion	Insertion	23	SL1 & SL2	Sporulation protein J and putative sortase B
AGs	Deletion	99	SL1 & SL2	Aminoglycoside 6-adenylyltransferase
ErmB	Deletion	53	SL1 & SL2	Dimethyladenosine transferase (*ermB*)
KstR2	Insertion	3	SL2	HTH-type transcriptional repressor KstR2
TetR	Insertion	15	SL1	TetR-family transcriptional regulator
PadR	Insertion	6	SL2	Transcriptional regulator PadR-like family protein
DHFR	Insertion	32	SL1 & SL2	Dihydrofolate reductase region
DicA	Insertion	7	SL1	Transcriptional repressor DicA
MDRP	Insertion	14	SL1	Multidrug resistance protein
C-P/M	Insertion	6	SL1	Putative corrin/porphyrin methyltransferase
Na^+ /drug anti	Insertion	5	SL1 & SL2	Putative drug/sodium antiporter
CS protein	Insertion	8	SL1 & SL2	Putative cell surface protein
Cfr	Insertion	2	SL2	Putative radical SAM enzyme and Cfr family
Spo	Insertion	3	SL1	Sporulation initiation inhibitor, Radical SAM protein, putative cell wall hydrolase, Stage 0 sporulation protein J
SAM	Insertion	2	SL1	SAM protein

5.4.2 Genotypic antimicrobial resistance determinants and phenotypic antimicrobial resistance

The data revealed SNPs commonly associated with antimicrobial resistance (Table 5.4 and Appendix 4, Page 265); 89.5% (248/277) and 5% (13/277) of isolates were found to have the amino acid substitutions found in the *gyrB* gene (V426D) and (V426I) respectively and 64.6% (179/277) to have the amino acid substitution found in the *gyrA* gene (T82I) which are known to confer resistance to the fluoroquinolone class of antibiotics. Additionally, substitutions in the 81 bp rifampicin resistance determining region of the *rpoB* gene; R505K, H502N and S485F were found in 32.5% (90/277), 33.2% (92/277) and 1.1% (3/277) respectively. These data are depicted and listed according to sub-lineage in Figure 5.2 and Table 5.4 respectively (also in Appendix 4, Page 265).

Genes potentially involved in antimicrobial resistance were also found in insertions and deletions; these are listed in Tables 5.5 and 5.6 respectively.

Table 5.4: SNPs associated with antimicrobial resistance found in the global isolates

Sub-lineage	Number of Isolates	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
		34,687	34,697	34,747	112,752	113,641	113,642	Position
		rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
		R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
SL1	163	73 (44.8%)	79 (48.5%)	0 (0%)	124 (76.1%)	134 (82.2%)	4 (2.5%)	
SL2	114	17 (15%)	13 (11.4%)	3 (2.6%)	55 (48.2%)	114 (100%)	9 (7.9%)	

* Reference residue/amino acid/ alternative residue.

Table 5.5: Genes found in insertions potentially involved in antimicrobial resistance

Insertion designation	Insertion site in M68 (base)	Size of insertion (bps)	Number of isolates	Sub-lineage	Gene/s of Interest/Putative Function	Potential antimicrobial resistance
A	697659	10,260	15	SL1	Drug/sodium antiporter & TetR-family transcriptional regulator	Tetracycline and others unknown
B	3873711	32,589	14	SL1	Streptogramin A acetyltransferase, antibiotic resistance ABC transporter & multidrug resistance protein	Streptogramin and others unknown
C	3879955	12,555	5	SL1 and SL2	Drug/sodium antiporter	Unknown
D	2633741	9,297	1	SL2	TetR-family transcriptional regulator	Tetracycline
E	3854778	6,571	1	SL2	Chloramphenicol o-acetyltransferase	Chloramphenicol

Table 5.6: Genes found in deletions potentially involved in antimicrobial resistance

Deletion designation	Deletion site in M68 (base)	Size of deletion (bps)	Number of isolates	Sub-lineage	Gene/s of Interest/Putative Function	Potential antimicrobial resistance
F	480308	27,982	67	SL1 and SL2	Dimethyladenosine transferase (*ermB*)	Erythromycin
G	2527343	9,890	99	SL1 and SL2	Aminoglycoside 6-adenylyltransferase	Aminoglycoside class
H	2969461	16,713	5	SL1 and SL2	Teicoplanin resistance gene & beta-lactamase	Teicoplanin & β-*lactam* antibiotics

To determine if any of the SNPs or genes found in insertions or deletions listed in Tables 5.4, 5.5 and 5.6 correlated with phenotypic antimicrobial resistance, seven isolates were selected for MIC testing based on maximising variation with regards to isolation date, geographical location, presence or absence of resistant SNPs and presence or absence of insertions and deletions containing genes potentially associated with antimicrobial resistance; S- 017.72, WA 1514, S- 017.92, S- 017.27, S- 017.74, I 6 and 01-116. MICs were determined (Section 2.2.7) against the antibiotics; chloramphenicol, rifampicin, tetracycline, erythromycin, naladixic acid, gentamicin, teicoplanin and ampicillin. MIC values are shown in Table 5.7.

All eight isolates were resistant to naladixic acid, ampicillin, gentamicin and all were sensitive to teicoplanin. Two isolates (2/8) were resistant and six (6/8) were sensitive to chloramphenicol. Four isolates (4/8) were resistant, two (2/8) were intermediate and two (2/8) were sensitive to rifampicin. Six isolates (6/8) were resistant and two isolates (2/8) were found with intermediate resistance to tetracycline. Seven isolates (7/8) were resistant and one isolates (1/8) was sensitive to erythromycin.

Table 5.7: Genotypic and phenotypic antimicrobial resistance data for the global isolates

	Strain	**M68**	**S- 017.72**	**WA 1514**	**S- 017.92**	**S- 017.27**	**S- 017.74**	**I 6**	**01-116**
	Location	**Ireland**	**Walsall**	**Australia**	**China**	**Wrexham**	**Walsall**	**Indonesia**	**Korea**
	Date Isolated	**2006**	**2011**	**2012**	**2009**	**1996**	**2011**	**2011**	**2001**
	Sub-lineage	SL1	SL1	SL1	SL1	SL2	SL2	SL2	SL1
	*Insertion			A, B	A		C	D, E	
	**Deletion	F, G, H		F	F, G	H	F, H	F, H	
Resistant SNPs	*rpoB* (R505K)			✓	✓	✓		✓	✓
	rpoB (H502N)		✓	✓	✓			✓	✓
	rpoB (S485F)					✓			
	gyrA (T82I)		✓	✓	✓			✓	
	gyrB (V426I)					✓			
	gyrB (V426D)		✓	✓	✓	✓	✓	✓	✓
Antimicrobial Agent	[a]Chloramphenicol	8 (S)	8 (S)	4 (S)	64 (R)	8 (S)	8 (S)	256 (R)	8 (S)
	[a]Rifampicin	0.008 (I)	2 (I)	0.004 (S)	>256 (R)	>256 (R)	0.004 (S)	>256 (R)	>256 (R)
	[b]Tetracycline	32 (R)	32 (R)	0.25 (I)	32 (R)	32 (R)	0.25 (I)	32 (R)	32 (R)
	[b]Erythromycin	>256 (R)	>256 (R)	>256 (R)	>256 (R)	>256 (R)	<2 (S)	>256 (R)	>256 (R)
	[b]Nalidixic acid	256 (R)	256 (R)	256 (R)	256 (R)	256 (R)	256 (R)	256 (R)	256 (R)
	[c]Gentamicin	>256 (R)	>256 (R)	256 (R)	>256 (R)	256 (R)	256 (R)	>256 (R)	>256 (R)
	[c]Teicoplanin	<1 (S)	<1 (S)	<1 (S)	<1 (S)	<1 (S)	<1 (S)	<1 (S)	<1 (S)
	[b]Ampicillin	8 (R)	8 (R)	8 (R)	8 (R)	8 (R)	4 (R)	4 (R)	8 (R)

Table adapted from a publication by Cairns *et al.*, (Cairns et al., 2017).

(S) = sensitive, (I) = intermediate resistance (R) = resistant

[a] Recommended by the European Committee on Antimicrobial Susceptibility Testing (EUCAST).
[b] Recommended by CLSI (M11-A8, 2012, and M-100-S23, 2013).
[c] No guidance from CLSI or EUCAST, cut-offs based on data according to the CLSI M-100-S23 (interpretative values for *Staphylococcus aureus*).

***Genes carried by insertions potentially involved in antimicrobial resistance:**
Insertion A: Putative drug/sodium antiporter and radical SAM protein TetR-family transcriptional regulator
Insertion B: Streptogramin A acetyltransferase and multidrug resistance protein
Insertion C: Putative drug/sodium antiporter
Insertion D: TetR-family transcriptional regulator
Insertion E: Chloramphenicol o-acetyltransferase (M68 has one copy of chloramphenicol)

**** Genes carried by deletions potentially involved in antimicrobial resistance:**
Deletion F: Dimethyladenosine transferase (*ermB*)
Deletion G: Putative teicoplanin resistance gene and putative beta-lactamase
Deletion H: Aminoglycoside 6-adenylyltransferase

To investigate the correlation between genotypic determinants of antimicrobial resistance with the actual phenotype data, both were compared. Based on the *gyrA* and *gyrB* SNP data, 76.2% (211/277) of isolates had a genotypic SNP associated with resistance to the fluoroquinolone class of antibiotics. All eight isolates (8/8) were phenotypically resistant to naladixic acid indicating resistance to the fluoroquinolone class of antimicrobials. Based on the *rpoB* SNP data, this study found 34.7% (96/277) of isolates with a genotypic SNP associated with resistance to the rifampicin class of antibiotics and 82% (152/185) of these mutations were found in SL1. Only two isolates (2/8) were sensitive to rifampicin with one isolate containing the R505K and H502N SNP indicating that these alone do not confer phenotypic resistance. The SNP S485F was found in three historical isolates from Wrexham, UK. All isolates (8/8) were sensitive to teicoplanin despite one isolate

with the deletion containing a putative teicoplanin resistance gene. Six isolates (6/8) were resistant and two isolates (2/8) were found with intermediate resistance to tetracycline, however, the two TetR-family transcriptional regulator genes found in insertions (insertions A and D) did not correlate with this variation in resistance indicating that neither of the two TetR-family transcriptional regulator genes alone conferred resistance or there is another mechanism involved. Seven isolates (7/8) were resistant and one isolates (1/8) was sensitive to erythromycin, however, the deletion that contained the dimethyladenosine transferase (*ermB*) gene (deletion F) did not correlate with this variation in phenotypic resistance. Two isolates (2/8) were resistant and six (6/8) were sensitive to chloramphenicol, however, the insertion that contained the chloramphenicol o-acetyltransferase gene (insertion E) did not correlate with this variation in phenotypic resistance. All isolates (8/8) were phenotypically resistant to gentamicin despite only three isolates with the deletion containing an aminoglycoside 6-adenylyltransferase (deletion H). All eight isolates (8/8) were resistant ampicillin despite the presence or absence of the genotypic determinants of antimicrobial resistance in Table 5.7.

By investigating the SNPs, insertions and deletions of a global collection of 277 PCR ribotype 017 isolates revealed two sub-lineages (SL1 and SL2), however, when comparing with phenotypic data, no SNPs, insertions, deletions or phenotypic antimicrobial resistance was associated with a clonal expansion of *C. difficile* PCR ribotype 017.

5.4.3 Genotypic comparison of isolates between geographical locations

To investigate if strains of *C. difficile* PCR ribotype 017 isolated from different geographical origins are phylogenetically distinct by WGS, Table 5.8 lists the isolates by geographical origin and a maximum-likelihood phylogenetic tree based on the 1288 SNPs was generated (Figure 5.3) with separation by continent (Figure 5.4). The majority of isolates are from Europe (137/277), Asia (59/277) and North America (59/277). The data in Table 5.8 and as illustrated in Figures 5.3 and 5.4 indicate that isolates from varying geographical locations are widespread amongst both sub-lineages.

Table 5.8: Number and percentage of isolates by origin and sub-lineage

Geographical Location	Number of isolates	Percentage of isolates	Sub-lineage/s
Europe	137	49.5%	1 and 2
Asia	59	21.3%	1 and 2
North America	59	21.3%	1 and 2
Africa	9	3.3%	1 and 2
Oceania	7	2.5%	1 and 2
South America	4	1.4%	1
Middle East	2	0.7%	1
TOTAL	**277**	**100%**	**1 and 2**

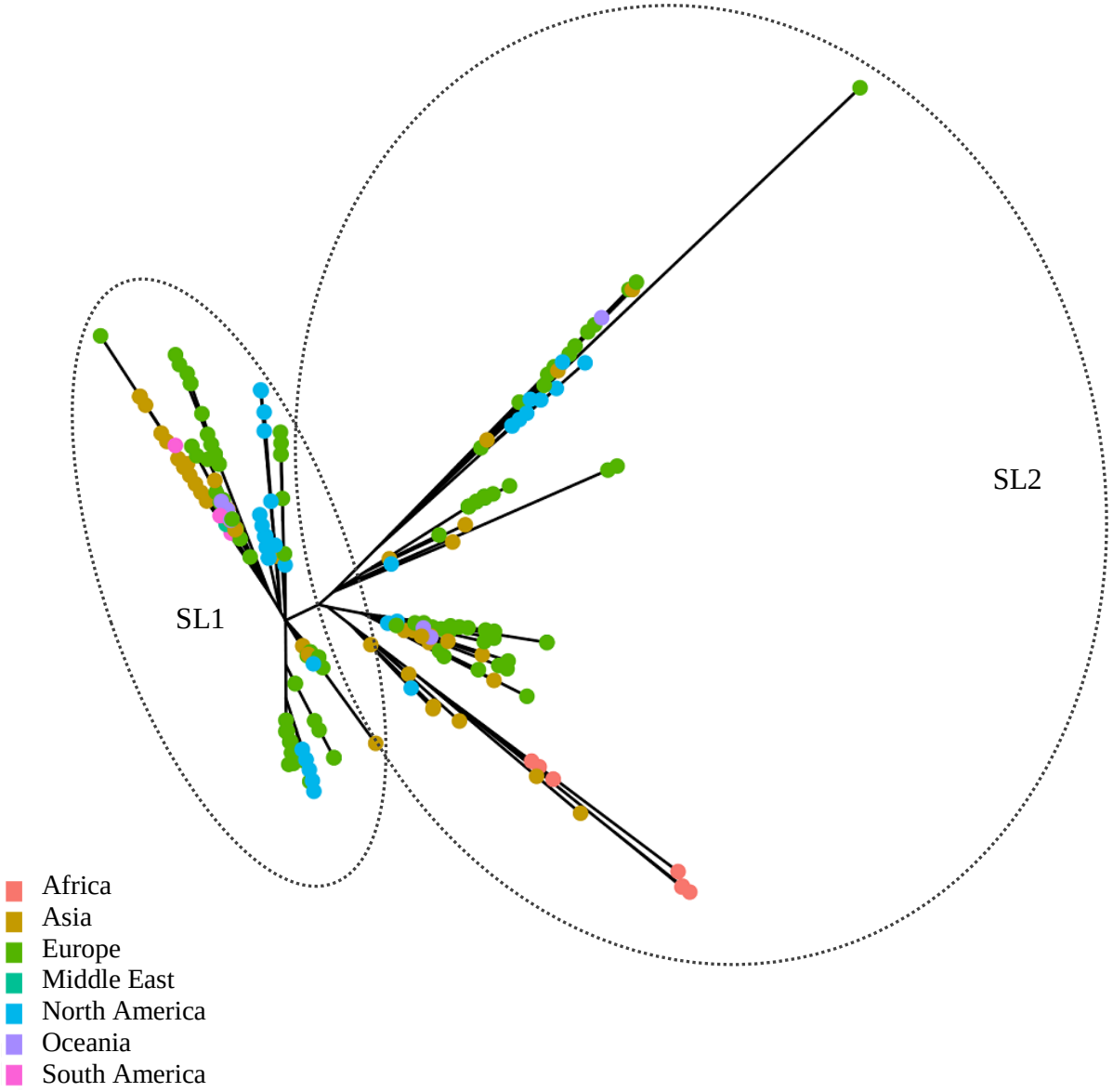


Figure 5.3: Maximum-likelihood phylogenetic analysis for the global *C. difficile* isolates

Figure adapted from a publication by Cairns *et al*., (Cairns et al., 2017). Maximum-likelihood phylogenetic analysis of 277 global *C. difficile* PCR ribotype 017 isolates based on core-genome SNPs against the M68 reference. Non-rare (>1% MAF) SNPs that were not in close proximity to insertions or deletions to determine the phylogenetic tree were used. The SL1 and SL2 sub-lineages were differentiated by four SNPs with the reference strain M68 falling into SL1. The coloured nodes indicate the geographical source of isolates.

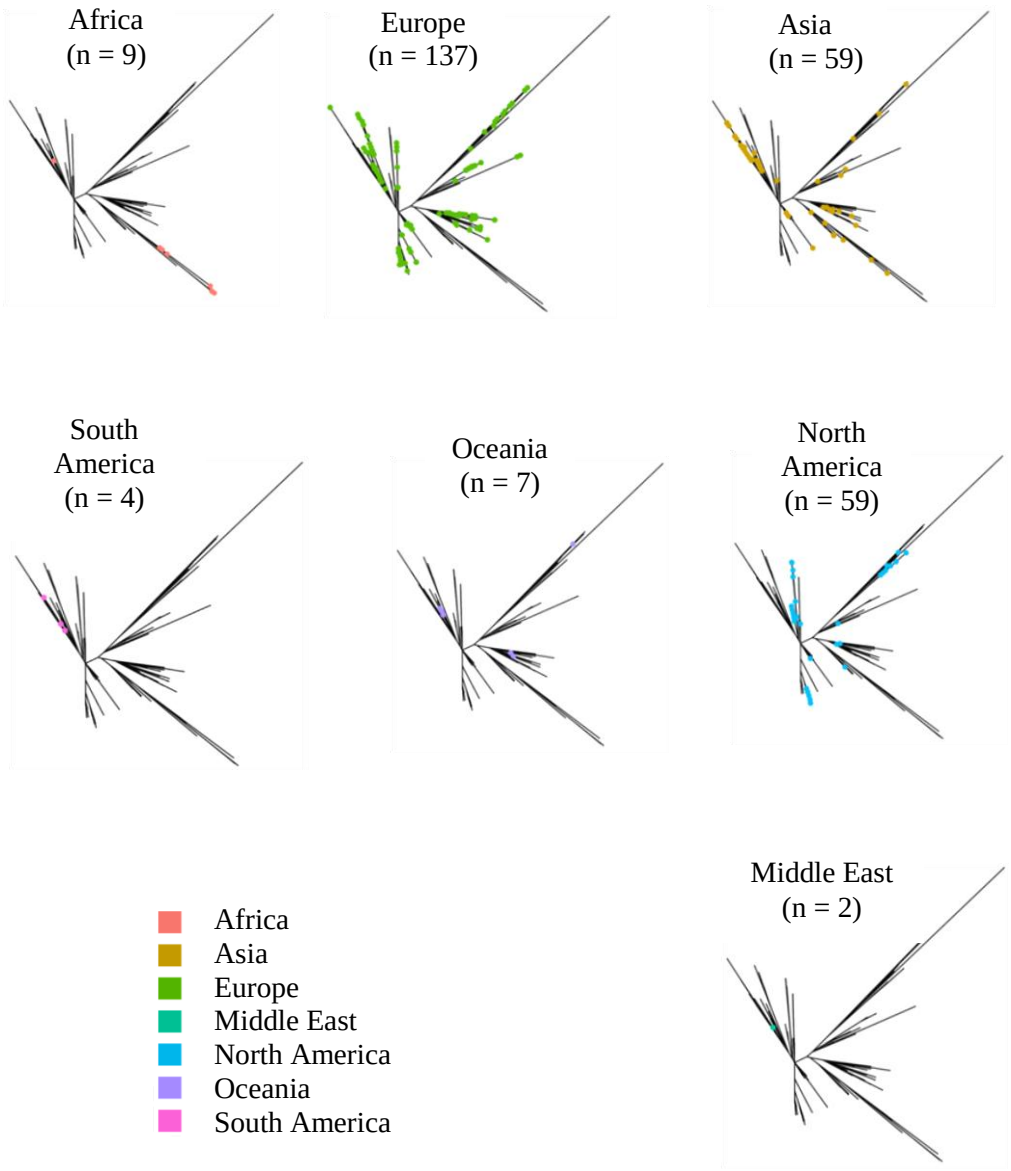


Figure 5.4: Maximum-likelihood phylogenetic analysis (separated by continent) for the global *C. difficile* PCR ribotype 017 isolates

Figure adapted from a publication by Cairns *et al*., (Cairns et al., 2017). Maximum-likelihood phylogenetic analysis of 277 global *C. difficile* PCR ribotype 017 isolates based on core-genome SNPs against the control strain M68 separated into individual panels corresponding to continent. Data from five out of seven continental designations (Africa, Europe, Asia, Oceania and USA) include SL1 and SL2 isolates indicating that both sub-lineages are global in nature.

5.4.4 Genotypic comparison of isolates between sources

To investigate if strains of *C. difficile* PCR ribotype 017 isolated from non-human sources are phylogenetically distinct by WGS, Table 5.9 lists the isolates by source and a maximum-likelihood phylogenetic tree based on the 1288 SNPs was generated with separation by source (Figure 5.5). There is no sub-lineage exclusively associated with non-human isolates, all non-human isolates are in sub-lineage 1 amongst human isolates.

Table 5.9: Number and percentage of isolates by source and sub-lineage

Geographical Location	Number of isolates	Percentage of isolates	Sub-Lineage/s
***Human**	253	91.3%	1 and 2
Canine	11	4%	1
Bovine	9	3.3%	1
Equine	4	1.4%	1
TOTAL	**277**	**100%**	**1 and 2**

*includes the two environmental strains isolated from a hospital ward

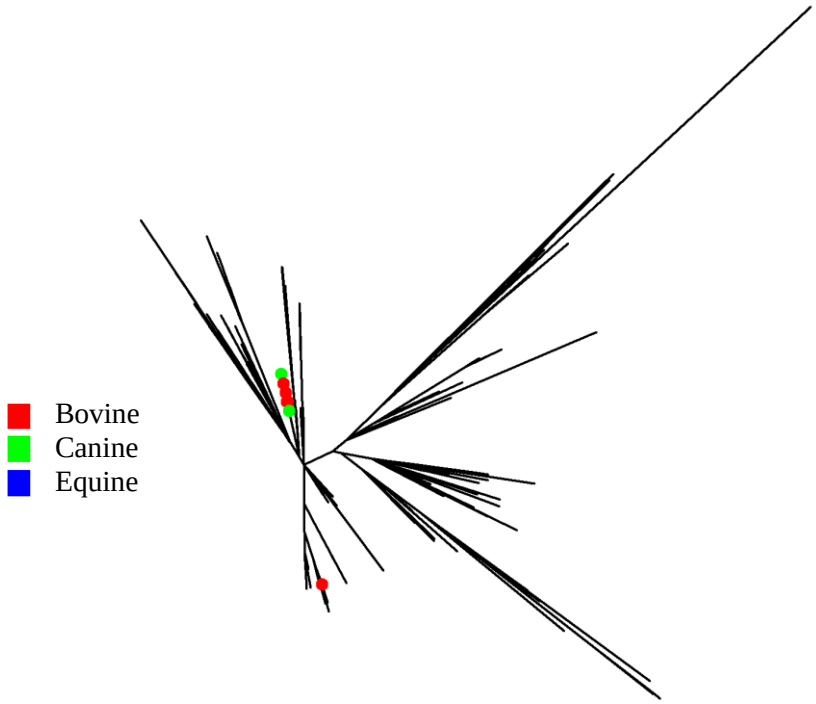


Figure 5.5: Maximum-likelihood phylogenetic analysis (separated by source) for the global *C. difficile* PCR ribotype 017 isolates

Figure adapted from a publication by Cairns *et al*., (Cairns et al., 2017). Maximum-likelihood phylogenetic analysis of the global *C. difficile* PCR ribotype 017 isolates based on core-genome SNPs against the control strain M68 depicting the 24 animal isolates by coloured nodes. Note the three equine isolates are positioned (and masked) by the bovine and canine cluster. The two bovine isolates on the bottom of the tree have a SNP distance of 17 from the bovine, canine, and equine cluster on the top. All animal isolates are from Ontario, Canada and isolated between 2002 and 2005.

5.4.5 The origins of *C. difficile* PCR ribotype 017

To investigate the belief that *C. difficile* PCR ribotype 017 originated in Asia, global transmission events inferred from Bayesian evolutionary analysis was performed and inferred global transmission as depicted in Figure 5.6. The data indicates a USA origin and suggests that it was Europe that introduced *C. difficile* PCR ribotype 017 to Asia and Australia, with subsequent spread from Asia to the Middle East, South America and South Africa. The analysis indicates over 40 movements back and forth over the span of 30 years.

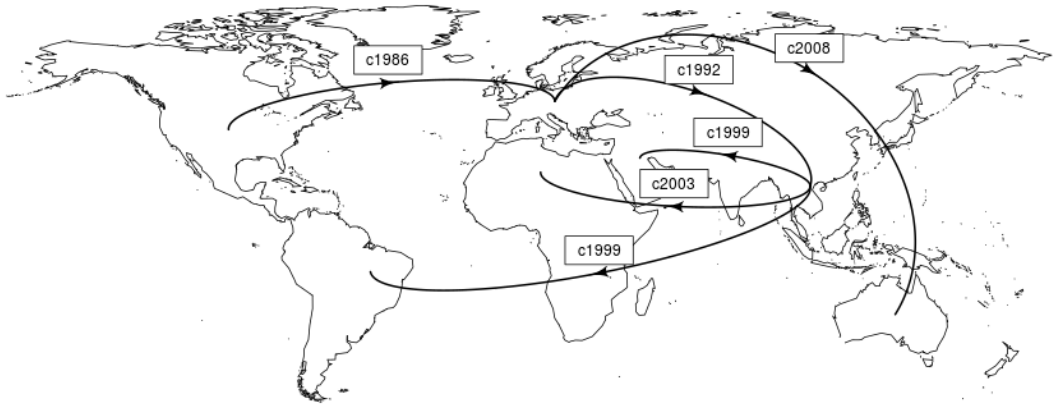


Figure 5.6: Inferred global transmission events of the global *C. difficile* PCR ribotype 017 isolates

Figure adapted from a publication by Cairns *et al*., (Cairns et al., 2017). Global transmission events inferred from bayesian evolutionary analysis of *C. difficile* PCR ribotype 017 isolates. From geo-temporal analyses the first movement into each continent of isolates in this study can be inferred, with the date and originating continent. The analysis indicates a USA origin with an expansion into Europe in the mid-1980s, followed by a move into Asia and on to Africa and South America through the 1990s and early 2000s. PCR ribotype 017 was not identified in Oceania (Australia) until the late 2000s, via a jump from Europe.

5.5 Discussion

Using PCR ribotyping, WGS and MIC assays, my study investigated a total of 277 *C. difficile* isolates (Appendix 2, Page 249). This collection of 277 strains were from; human (n=251), bovine (n=9), canine (n=11), equine (n=4) and hospital ward environment (n=2) with isolation dates between 1990 and 2013 and included isolates from six global continents (Europe, North America, South America, Asia, Oceania and Africa).

No SNP/s and/or insertions or deletions (including those associated with antimicrobial resistance) were found associated with the global spread of *C. difficile* PCR ribotype 017 in my study. It is possible that one or more were associated yet not identified using the methods described in this thesis. Reminiscent of the *C. difficile* PCR ribotype 027 lineage, two distinct sub-lineages of *C. difficile* PCR ribotype 017 that contain multiple independent clonal expansions were revealed in this study (Figures 5.1 and 5.2). These two sub-lineages differed by four SNPs; one present in a non-coding region and three non-synonymous SNPs (Table 5.2) with isolates in both sub-lineages being geographically and temporally diverse. My data shows that the multiple haplotypes revealed for *C. difficile* PCR ribotype 017 is similar to PCR ribotype 027 where >100 distinct genotypes were found in 151 isolates (He et al., 2013). The insertions and deletions were well distributed geographically and temporally suggesting either the rapid dissemination of strains or the multiple independent acquisition and loss of DNA regions. The insertion of different clusters of genes at the same site suggests ‘hot-spot’ sites for the uptake of DNA and a 49 kbp putative transposon-like putative mobile genetic element found exclusive to the

clonal cluster of 23/37 London isolates from chapter 4 were also found to insert at a different site in single isolates from Canada, the USA and the UK with isolation dates of 2006, 2006 and 2011 respectively (Figure 5.2) suggesting these isolates have independently acquired this insertion.

The data revealed SNPs, insertions and deletions associated with antimicrobial resistance and variation in phenotypic resistance (Tables 5.4, 5.5 and 5.6), however, no antimicrobial resistance genotype or phenotype was found associated with the geographical spread of *C. difficile* PCR ribotype 017 like that seen with PCR ribotype 027 and the SNP associated with fluoroquinolone resistance (He et al., 2013).

The majority of isolates (76%) were found to possess a SNP present in either the *gyrA* or *gyrB* genes which are associated with phenotypic resistance to the fluoroquinolone class of antibiotics (*gyrA* [T82I], *gyrB* [V426I] and *gyrB* [V426D]). All eight isolates tested phenotypically, were resistant to the fluoroquinolone naladixic acid. The SNP T82I found in the *gyrA* gene is the same mutation reported in the global outbreak of *C. difficile* PCR ribotype 027 (He et al., 2013) however, it was not shown to lead to the expansion of a sub-lineage like seen with PCR ribotype 027. Interestingly, the control strain M68 was resistant yet it did not possess any of the *gyrA* and *gyrB* associated SNPs or genes carried by insertions potentially involved in antimicrobial resistance suggesting that there is another mechanism for resistance to the fluoroquinolone class of antibiotics. This resistance amongst all eight strains tested is not unusual; other studies have found high levels of resistance to one or more of the fluoroquinolone class of antimicrobials. Barbut *et al.*, (Barbut

et al., 2007) found 33% (134/411) of isolates were intermediate susceptibility or resistant to moxifloxacin. Spigaglia *et al.,* (Spigaglia et al., 2008) further tested 83 of these isolates, eleven of which were PCR ribotype 017. All 11 isolates were resistant to the fluoroquinolones tested (moxifloxacin, ciprofloxacin, gatifloxacin, levofloxacin).

My study found 34.7% of isolates to possess a SNP present in the *rpoB* gene which is associated with phenotypic resistance to the rifampicin class of antibiotics. My phenotypic data found four, two and two isolates to be resistant, intermediate resistance and sensitive to rifampicin respectively. The SNPs R505K and H502N have previously been associated with rifampicin resistance in *C. difficile* (Curry et al., 2009) however, based on MIC data from my study, one of the two isolates that were sensitive to rifampicin contained the R505K and H502N SNPs indicating that these SNPs alone do not confer phenotypic resistance. My study also identified a third SNP in the *rpoB* gene (S485F) which was found in three historical isolates from Wrexham, UK. This resistance conferring SNP has not previously been reported in *C. difficile,* only in *Mycobacterium tuberculosis* (Bahrmand et al., 2009). All three isolates were phenotypically resistant to rifampicin, however, all three isolates also contained the SNP R505K and so confirming the SNPs (S485F) contribution to resistance was not possible. Gene expression profiling in tandem with phenotypic assays could be performed to confirm the contribution of this SNP to phenotypic resistance to rifampicin since not all genotypic determinants are expressed.

All eight isolates tested phenotypically were sensitive to teicoplanin despite one isolate with the deletion containing a putative teicoplanin resistance gene. This

indicates that the presence of the genetic region containing the putative teicoplanin resistance gene played no role alone or at all in the MIC of the seven isolates tested.

Phenotypic susceptibility to the antimicrobials tetracycline, ampicillin, erythromycin, gentamicin and chloramphenicol was found not to correlate with genotypic resistance determinants identified through WGS (Table 5.7) suggesting that there are other mechanisms required for phenotypic resistance to these antimicrobials. All isolates demonstrated phenotypic resistance to gentamicin. This is to be expected as *C. difficile* as an anaerobic organism is intrinsically resistant to the aminoglycoside class of antimicrobials. Anaerobes lack the oxygen-dependent transport mechanism required for cellular uptake (Bryan et al., 1979). This indicates that the determination of MICs to this class of antimicrobial in this thesis was futile.

C. difficile PCR ribotype 017 has previously been associated with high level antimicrobial resistance (Drudy et al., 2007c, Kuijper et al., 2001, Pituch et al., 2007, Freeman et al., 2015) and different countries appear to have different levels of antimicrobial resistance (for all PCR ribotypes) which is likely a result of varying antimicrobial prescribing and subsequent selective pressure (Freeman et al., 2015). With a large collection of *C. difficile* PCR ribotype 017 strains, it would be interesting to test the susceptibilities of more of the strain collection in this study. The geographical patterns of antimicrobial resistance for this toxin A-B+ PCR ribotype could then be determined.

C. difficile PCR ribotype 017 isolated from different geographical origins were not phylogenetically distinct by WGS. The transcontinental spread of both sub-lineages

containing strains from varied global locations showing no distinct pattern of spread (Figures 5.3 and 5.4) suggests that *C. difficile* PCR ribotype 017 isolated from different geographical origins are not phylogenetically distinct by WGS. This was not known previously and is a significant finding, as it suggests that neither sub-lineage of *C. difficile* PCR ribotype 017 has had a selective advantage in its ability to spread globally like that found with PCR ribotype 027 (He et al., 2013)

C. difficile PCR ribotype 017 isolated from human and non-human sources were not phylogenetically distinct by WGS (Figure 5.5). All non-human isolates were in SL1 only, however, the collection of non-human isolates was small (n=24) and were isolated from a similar location (Ontario, Canada) over a relatively short period of time (2002 and 2005) suggesting they were clonal by time and place. Importantly, all non-human isolates were placed phylogenetically amongst human isolates suggesting possible transmission between humans and animals. Although it is known that *C. difficile* can cause similar infection in both humans and animals (Limaye et al., 2000), transmission and phylogenetic associations have previously only been shown using MLST and not WGS. Therefore, this data shows that the same strain of PCR ribotype 017 can cause infection in both humans and animals using WGS.

C. difficile PCR ribotype 017 did not originate in Asia (Figure 5.6). Traditionally, it has been considered that *C. difficile* PCR ribotype 017 strains emerged from Asia due to its reported high incidence that could not relate to nor depend on toxin A-based assays for diagnosis (Collins et al., 2013). However, the global transmission events inferred from Bayesian evolutionary analyses in my study supports a USA origin for *C. difficile* PCR ribotype 017 with multiple, global transmission events

with its earliest movement into Europe in 1986 (Figure 5.6). This is similar to that found with PCR ribotype 027; the USA health system and practices appeared to have facilitated the evolution and epidemic spread of *C. difficile* PCR ribotype 027 (He et al., 2013) and now in this study with PCR ribotype 017. This is consistent with population movements of a globalised society (Soto, 2009).

The non-synonymous SNP found in the *dacF* gene was one of four SNPs that differentiated the two sub-lineages and was present in all 114/277 isolates in sub-lineage 2 (Table 5.2). The *dacF* gene codes for D-alanyl-D-alanine carboxypeptidase and has been shown to regulate peptidoglycan crosslinking in *Bacillus subtilis* (Popham et al., 1999). The β-lactam class of antibiotics target the peptidoglycan of the bacterial cell wall; the antibiotic inhibits the final step of peptidoglycan synthesis and inactivates penicillin-binding proteins that form the peptidoglycan layer of the cell wall. Due to the loss of the cell wall, this leads to cell death as a result of osmotic pressure (Tipper and Strominger, 1965). It is possible that a SNP in the *dacF* gene may result in a fitness advantage to strains possessing this genotype within the healthcare environment rendering the bacteria resistant to β-lactam antibiotics. Conversely, a study found that knockout mutants lacking the *dacF* gene in *C. perfringens,* reduced heat resistance, likely due to reduced cortex integrity when subjected to high heat (Orsburn et al., 2009). However, no difference in susceptibility to the β-lactam antibiotics teicoplanin or ampicillin between isolates in both sub-lineages was observed (Table 5.7) suggesting that this SNP did not affect susceptibility to β-lactam antibiotics nor did it lead to a clonal expansion of either sub-lineage like that seen with fluoroquinolone resistance and PCR ribotype 027 (He et al., 2013).

It has been well documented that fluoroquinolone use preceded resistance and subsequent global outbreaks of a clonal strain of *C. difficile* PCR ribotype 027 (He et al., 2013). PCR ribotype 017 is associated with higher rates of antimicrobial resistance when compared with other PCR ribotypes and in my study, a significant number of isolates had SNPs associated with phenotypic antimicrobial resistance (76% of isolates had a SNP present in either the *gyrA* or *gyrB* genes and 34.7% of isolates had a SNP present in the *rpoB* gene). It would therefore be plausible to consider that a clone of PCR ribotype 017 could be as problematic as PCR ribotype 027. However, my data shows no global clonal expansion of a strain of PCR ribotype 017. The strains used in my study also date back to 1990 which is prior to the concept of antimicrobial stewardship and its role in controlling *C. difficile* outbreaks in healthcare facilities. This would suggest that there are other phenotypic characteristics as well as antimicrobial resistance that are required for the global spread of a clonal strain like that seen with PCR ribotype 027. Future work to this study could include comparing the genotypic and phenotypic characteristics of PCR ribotype 017 and the clonal strain of PCR ribotype 027 to establish why the PCR ribotype 027 strain was clonal and spread globally and why PCR ribotype 017 has not. Previous studies have found PCR ribotype 027 to have phenotypic characteristics that could be associated with its global spread however; there is conflicting evidence for a phenotype specific to PCR ribotype 027 that would enable a strain to be clonal (Section 1.10, Page 50). No study has yet compared the phenotypic characteristics of the clonal strain of 027 with a large collection of isolates of other PCR ribotypes with known genotype. This could be performed with

select isolates from this study with the WGS data available for strains of PCR ribotype 017.

There were limitations in this study. Similar to chapter 4, there were limitations to the choice of control strain and methodologies used for detection of SNPs and *de novo* analysis of insertions and deletions (as described in Section 4.5, Page 175). Furthermore, the choice of control strain for the detection of phenotypic resistance was likely flawed. All isolates with MIC testing (both chapters 4 and 5) were resistant to naladixic acid including the reference control strain M68. Therefore, there was no negative control for this antibiotics effectiveness and achievement of a successful MIC assay. There may have been deterioration of the antibiotic, error in inoculum or the control strain was resistant to this antibiotic. Further phenotypic MIC studies should seek to include a fully sensitive control strain of *C. difficile*.

5.6 Conclusion

This study investigated the genetic diversity of 277 *C. difficile* PCR ribotype 017 isolates with isolation date, geographical, and source variation using WGS. Phylogenetic analysis of the SNPs suggests there are two main sub-lineages of PCR ribotype 017 that share an ancestry and are globally disseminated. Both sub-lineages contain isolates from diverse geographical locations and isolation dates, with animal isolates spread among human isolates. Together with the haplotype diversity and geographical and temporal diverse presence of insertions and deletions, these data suggest widespread transcontinental spread and recombination with independent acquisition and loss within different clusters. Antimicrobial resistant genotypic

determinants were well distributed geographically and none were found exclusive to phenotypic resistance suggesting there are other genotypic determinants associated with phenotypic antimicrobial resistance in this collection of strains.

Chapter 6

Final Discussion

6 Final Discussion

C. difficile continues to be a leading cause of healthcare-associated infections in the developed world. Its spores are able to survive, persist and spread in hostile environments and cause outbreaks of CDI amongst susceptible hosts in wards, hospitals and other healthcare facilities. This species was selected for study as there was a deficiency in our understanding of its phylogenetics and phylohistory. Genetic exploration using PCR ribotyping, MLST, WGS and phenotypic assays was performed and has provided an insight into how this species is evolving.

Clostridium difficile – a multifarious pathogen

Previous MLST and WGS studies identified five divergent phylogenetic lineages of *C. difficile* PCR ribotypes; 017, 027, 078, 023 and a group containing multiple PCR ribotypes (Griffiths et al., 2010, Dingle et al., 2011a, Lemee et al., 2004, Lemee et al., 2005). However, these studies were limited by only focusing on either; various hosts, geographic sources, various PCR ribotypes, or hospital and community and various PCR ribotypes. Chapter 3 of this thesis applied MLST to a larger collection of isolates (n=385) from diverse human, animal and food sources from three continents and of multiple PCR ribotypes and confirmed the five lineages of *C. difficile* previously identified. Chapter 3 also demonstrated that the typing techniques PCR ribotyping and REA do not correlate as well as previously thought and genotypes identified by one technique are incorrectly assumed a genotype based on a different typing technique. For example, strains previously assigned as PCR ribotype 027 but subsequently confirmed to be PCR ribotypes 176 and 198 but are sequence type 1 and cluster with lineage 2. Similarly, PCR ribotype 244 shares the

same genetic lineage as PCR ribotype 027 according to SNP analysis (Eyre et al., 2015) and like PCR ribotype 027, patients present with severe CDI and higher mortality (Lim et al., 2014). It is a concern that emerging PCR ribotypes phylogenetically similar to other PCR ribotypes by MLST and associated with hypervirulence and increased transmission are not initially associated as having a similar phenotype (Lim et al., 2014, Polivkova et al., 2016).

Since chapter 3 was performed, the number of PCR ribotypes and sequence types has dramatically increased; the PCR ribotype and MLST databases now comprises 877 PCR ribotypes (12/04/2018, Dr Warren Fawley, personal communication) and 488 sequence types (19/04/2018, MLST database https://pubmlst.org/cdifficile/).This suggests further evolution. This MLST analysis has broadened our understanding of the pathogen *C. difficile* by demonstrating how evolution can be observed using SNP variation in MLST alleles, however, investigating the individual lineages in more depth with finer resolution will enable us to monitor the emergence of evolving virulent and highly transmissible clones. There is need for immediate awareness of a hypervirulent clone which would prompt earlier intervention as has already been shown with other pathogens such as *Listeria monocytogenes* (Jackson et al., 2016, Kvistholm Jensen et al., 2016), *Streptococcus pyogenes* (Tagini et al., 2017), MRSA (Azarian et al., 2015), *Salmonella* (Inns et al., 2017), *S. aureus* and *C. difficile* (Eyre et al., 2012).

PCR ribotype 017 – Historical London clone

The MLST data in chapter 3 confirmed previous findings that *C. difficile* PCR ribotype 017 strains cluster as an individual lineage (Griffiths et al., 2010, Dingle et

al., 2011a, Stabler et al., 2012). This coupled with its unique toxin profile (Alfa et al., 2000), unusual global prevalence (Hawkey et al., 2013) and multiple clusters of CDI caused by PCR ribotype 017 in a London hospital elderly care ward between 2009 and 2011, led to hypothesising that these cluster isolates would be clonal and a phenotype that may explain a persistent trait could be characterised.

Phylogenetic analysis of WGS SNPs of 37 *C. difficile* PCR ribotype 017 isolates ascertained that a clone was present at UHL which included a historical isolate from 2005 and was thought to be the ancestral cluster strain. Additionally, a 49 kbp genetic region suggestive of being a transposon-like putative mobile genetic element was found exclusive to the clonal strains. This putative mobile genetic element contained genes potentially involved in virulence and/or transmissibility and with the clonal nature of this cluster, phenotypic assays were performed, however, no phenotype exclusive to the clonal cluster was characterised. It is unknown if the clonal strain identified in this chapter possessed genetic attributes that contributed to a persistence phenotypic trait and/or if any of the SNPs or genes carried by the 49 kbp genetic region contributed to this. It is equally possible that the clonal strain did not possess any attributable phenotype and was simply re-introduced into the new hospital build and was never eradicated from the environment.

Although the observations detailed within chapter 4 were derived from a small sub-set of isolates, this was the first report in the UK investigating the phylohistory of isolates from hospitalised patients with CDI due to PCR ribotype 017.

PCR ribotype 017 – a global perspective

In chapter 3 the evolution of individual lineages of *C. difficile* was confirmed and in chapter 4 a historical clone of PCR ribotype 017 unique to a hospital ward that exclusively possessed a genetic region suggestive of being a transposon-like putative mobile genetic element was identified. Studies have investigated the in-depth phylogenetics of *C. difficile* PCR ribotype 027 using WGS and significant to chapter 5, He *et al.*, revealed the global expansion of a fluoroquinolone resistant sub-lineage of PCR ribotype 027 (He et al., 2013). However, no studies have reported the global phylogeny of other lineages of *C. difficile*.

A hypothesis of chapter 5 was to investigate the global population structure of *C. difficile* PCR ribotype 017. It was postulated that PCR ribotype 017 evolved similar to PCR ribotype 027 but from a different geographical location; due to the unusual global prevalence of PCR ribotype 017 previously reported, it was hypothesised that PCR ribotype 017 evolved out of Asia unlike PCR ribotype 027 which evolved out of USA (He et al., 2013). An advantage of chapter 5 was the collection of isolates; it was large (n=277) and a global representation of PCR ribotype 017. Chapter 5 identified two distinct sub-lineages (SL1 and SL2) of PCR ribotype 017 containing multiple independent clonal expansions and multiple antimicrobial resistant SNP determinants. The *gyrA* T82I SNP reported in the global outbreak of PCR ribotype 027 (He et al., 2013) was found, however, this did not result in the global expansion of a strain like that seen with PCR ribotype 027, nor did any other antimicrobial genotype or phenotype. Similar to the PCR ribotype 027 study, the data did reveal multiple haplotypes suggesting both lineages have evolved in a similar fashion (He et al., 2013). The data alludes to possible transmission

between humans and animals with 24 animal strains placed amongst human isolates. The animal isolates were amongst human isolates which is not surprising as studies have already indicated this for PCR ribotype 078 using PCR ribotyping and MLST (Debast et al., 2009, Jhung et al., 2008, Goorhuis et al., 2007, Gould and Limbago, 2010). This data shows that the same strain of PCR ribotype 017 can cause CDI in both humans and animals using WGS.

Chapter 5 identified multiple insertions and deletions which were well distributed geographically and temporally. Isolates from Canada, USA and UK harboured the 49 kb transposon-like putative mobile genetic element found in chapter 4 although inserted at different sites suggesting integration 'hot-spot' sites for horizontal transfer of genetic material. These data suggest either the rapid dissemination of strains or the multiple independent acquisition and loss of DNA regions.

Like that found with PCR ribotype 027 (He et al., 2013), using geo-temporal analyses, the findings in chapter 5 support a USA origin for PCR ribotype 017 with multiple, global transmission events, movement to Europe in ~1986, and subsequent spread to Asia and Australia, and from Asia to the Middle East, South America and South Africa. The data indicates over 40 movements back and forth over the span of 30 years which is consistent with population movements of a globalised society (Soto, 2009). This is surprising considering the published data indicating an Asian origin for PCR ribotype 017, however, the USA health system and practices appeared to facilitate the ready evolution and epidemic spread of PCR ribotype 027 (He et al., 2013) and the same appears possible in this chapter with PCR ribotype 017.

Conclusions and future direction

The key findings of this thesis are that *C. difficile* as a species is continually evolving with the appearance of divergent sub-lineages and mini-clusters within lineages, with further description of SNPs associated with antimicrobial resistance, haplotypes and genetic deletions and insertions. This knowledge contributes to our understanding of the evolution of *C. difficile* and more specifically PCR ribotype 017. However, it remains to be identified why PCR ribotype 017 is so successful and has spread globally yet lacks toxin A. It could be postulated that toxins are not as important in strain dissemination as was previously thought. There may be other compensatory factors that PCR ribotype 017 has that have enabled it to spread globally and be so transmissible. For example, improved environmental survival by spore formation; may be this is more important for PCR ribotype 017. Interestingly, data subsequent to this thesis compared strains of PCR ribotype 023 from across Europe and China using WGS and showed great similarity between strains which was consistent with a recently emerged lineage that appears to be under little selective pressure to evolve (Shaw *et al.*, submitted, Page 23). This is similar to the findings in this thesis for PCR ribotype 017 and in contrast to PCR ribotype 027 where a sub-lineage was shown to spread globally as a result of antimicrobial selective pressure (He et al., 2013). It remains to be determined why evolutionary distinct lineages of *C. difficile* are simultaneously emerging to cause disease.

The application and utility of WGS for tracking transmission, outbreak investigation and studying the evolution of many pathogens has demonstrated WGS to be superior to routine typing tools. Examples include; the prompt investigation of a clonal strain of *Escherichia coli* (Rasko et al., 2011), improved epidemiological inferences and

linkages of *M. tuberculosis* (Gardy et al., 2011) and real-time investigation enabling actionable results and transmission monitoring of an outbreak of *Salmonella* (Quick et al., 2015). WGS has demonstrated to be superior to routine typing tools for outbreak investigation of *C. difficile* enabling the involvement of infection control to promptly reduce the incidence of CDI (Eyre et al., 2017). If WGS were available in 2003, detection of the *C. difficile* fluoroquinolone resistant genotype outbreak strain, PCR ribotype 027 would likely have been confirmed much earlier (He et al., 2013, Dingle et al., 2017) and would have enabled appropriate antimicrobial stewardship and subsequent reduction in the global outbreaks that this strain caused.

WGS allows for the identification and description of selective pressures acting on microbial pathogens; the global spread of *C. difficile* PCR ribotype 027 is an example of the selective pressure of antimicrobial prescribing where use of fluoroquinolones selected for this clone (He et al., 2013).

In addition to the detection of known genetic virulence determinants (i.e. the Xpert® *C. difficile* assay and detection of PCR ribotype 027), WGS performed alongside phenotypic studies enables the identification and characterisation of virulent genotypes, for example the identification of genetic markers associated with *C. difficile* PCR ribotype 027 (Stabler et al., 2009). This is valuable for the diagnostic laboratory and patient/s.

Despite this, WGS does have its limitations. To scientists and clinicians in the diagnostic field of microbiology, it can present unknown genotypes for which the phenotype and clinical significance maybe unknown. Relying on WGS will also not

detect new and emerging phenotypes as clearly as if phenotypic testing was performed. WGS requires sophisticated bioinformatics, quick data processing, large data storage capabilities and experienced bioinformaticians to analyse and interpret data and put it into clinical context (Tsai et al., 2016, Fricke and Rasko, 2014); this is currently too costly for most diagnostic microbiology laboratories. A number of global organisations, collaborative projects and laboratory consortia are actively working to overcome technical barriers, develop bioinformatic solutions and pilot studies of WGS-based typing for public health protection and to develop standards (European Centre for Disease Prevention and Control, 2016).

Before genotypic data can be relied upon clinically, sufficient data must be collated and compared with phenotype to ensure reliable correlation. Further studies comparing genotype to phenotype and expression profiles using WGS, knockout mutant studies (Lyras et al., 2009, Bradshaw et al., 2017, Ngernsombat et al., 2017) and proteomic studies (Pettit et al., 2014, Dresler et al., 2017) will lead the way to improving the inference and ultimately confidence in WGS.

Improvements to the limitations described will progress diagnostic microbiology. The integration of WGS into routine diagnostic laboratories will lead the way for real-time identification and monitoring of outbreaks (local and national), surveillance of global transmission, and identification of novel pathogenic genotypes which may have otherwise evaded detection by current diagnostic techniques. This will ultimately improve the management of patients and populations with infectious diseases.

Appendices

Appendix 1: Bacterial isolates used in chapter 3

NT = nontypeable (PCR ribotype not recognised by the London CDRN reference laboratory), * = Novel sequence type, ** = Isolates used in the PCR ribotype 027 depth study.

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
630	J. Brazier	012	54	1	4	7	1	1	3	3	Lineage 1	Human	Switzerland	1982
IS58	T. V. Riley	033	11	5	8	5	11	9	11	8	Lineage 5	Human	Australia	< 2004
2046	J. Brazier	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	unknown	unknown
2047	J. Brazier	NT	8	1	1	2	6	1	5	1	Lineage 1	unknown	unknown	unknown
2050	J. Brazier	015	44	2	5	2	1	1	3	1	Lineage 1	unknown	unknown	unknown
2052	J. Brazier	023	5	1	6	4	7	2	8	7	Lineage 3	unknown	unknown	unknown
****2053**	J. Brazier	027	1	1	1	1	10	1	3	5	Lineage 2	unknown	unknown	unknown
2054	J. Brazier	NT	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
2055	J. Brazier	106	42	1	1	2	1	1	7	1	Lineage 1	unknown	unknown	unknown
5342	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5350	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
****5359**	J. G. Songer	NT	1	1	1	1	10	1	3	5	Lineage 2	unknown	unknown	unknown
5361	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	26/12/2006
5363	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/03/2007
****5370**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Bovine	USA	26/06/2007
****5373**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Bovine	USA	20/03/2007

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
5379	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5384	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/03/2007
5397	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	26/06/2007
5407	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	26/09/2006
5408	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	26/06/2007
5416	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
5416	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
****5427**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Bovine	USA	31/07/2007
5428	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/03/2007
5429	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	15/10/2007
5432	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	23/01/2007
5444	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	15/10/2007
5468	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5898	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/03/2007
5904	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	15/10/2007
5911	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	17/04/2007
5912	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5917	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	15/10/2007
5920	J. G. Songer	NT	3	1	1	2	1	1	1	1	Lineage 1	Bovine	USA	31/07/2007
5921	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/11/2006

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
5927	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5927	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5933	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5938	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/11/2006
5946	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/03/2007
5954	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5963	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/11/2006
5968	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/03/2007
5982	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/11/2006
5983	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/03/2007
5984	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	15/10/2007
5986	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	31/07/2007
5987	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	15/10/2007
5992	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5996	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	15/10/2007
6004	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
6005	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	26/06/2007
6007	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2007
6007	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2007
****6014**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Bovine	USA	unknown

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
6015	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	15/10/2007
6021	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/03/2007
6065	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
6067	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
****8864**	J. Brazier	036	62	1	1	1	9	1	3	1	Lineage 2	Human	UK	< 2004
10/33	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
30256	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
31807	E. Kuijper	NT	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
13/03	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
19/07	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
19/09	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
17/10	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
3/11	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
80249	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	20/03/2007
****2004101**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2004
****2004102**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2004
****2004118**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2004
****2004163**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2004
****2005079**	J. G. Songer	NT	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2005
2005088	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Porcine	USA	2005

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
2005093	J. G. Songer	126	11	5	8	5	11	9	11	8	Lineage 5	Porcine	USA	2005
2005094	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Porcine	USA	2005
2005325	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2005
2005508	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Porcine	USA	2005
2005511	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Porcine	USA	2005
2005515	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Porcine	USA	2005
2005517	J. G. Songer	NT	48	1	1	2	1	1	5	1	Lineage 1	Porcine	USA	2005
2005519	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Porcine	USA	2005
****2006237**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Equine	USA	2006
2006238	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006239	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006240	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006241	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006243	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006244	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006245	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006246	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006253	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006254	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	2006
2006354	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2006

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
2006379	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2006
2006437	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Food	USA	2006
2006438	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Food	USA	2006
****2006439**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Food	USA	2006
2006460	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2006
2007007	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
2007011	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
2007019	J. G. Songer	126	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
2007024	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
****2007042**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Food	USA	2007
2007054	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
****2007098**	J. G. Songer	NT	32	1	1	11	1	1	3	2	Lineage 2	Human	USA	2007
2007134	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
****2007140**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2007
****2007195**	J. G. Songer	NT	41	1	1	9	9	1	3	2	Lineage 2	Human	USA	2007
2007206	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
****2007218**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Food	USA	2007
2007219	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Food	USA	2007
****2007222**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Food	USA	2007
****2007223**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Food	USA	2007

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
2007224	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Food	USA	2007
2007229	J. G. Songer	126	11	5	8	5	11	9	11	8	Lineage 5	Food	USA	2007
2007230	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Food	USA	2007
****2007235**	J. G. Songer	NT	1	1	1	1	10	1	3	5	Lineage 2	Food	USA	2007
2007334	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
2007361	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
2007380	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
2007600	J. G. Songer	126	11	5	8	5	11	9	11	8	Lineage 5	Human	Spain	2007
2007601	J. G. Songer	126	11	5	8	5	11	9	11	8	Lineage 5	Human	Spain	2007
2007606	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Food	USA	2007
2007607	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Food	USA	2007
2007786	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	Italy	2007
2007792	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	Italy	2007
****2007825**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2007
****2007826**	J. G. Songer	NT	61	1	5	11	1	1	3	2	Lineage 2	Human	USA	2007
****2007827**	J. G. Songer	262	67	1	1	9	9	1	3	5	Lineage 2	Human	USA	2007
****2007828**	J. G. Songer	NT	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2007
****2007829**	J. G. Songer	262	67	1	1	9	9	1	3	5	Lineage 2	Human	USA	2007
****2007830**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2007
2007831	J. G. Songer	001	3	1	1	2	1	1	1	1	Lineage 1	Human	USA	2007

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
**2007832	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2007
**2007833	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2007
2007834	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
2007835	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
2007838	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	2007
**2007839	J. G. Songer	NT	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	2007
2007841	J. G. Songer	NT	3	1	1	2	1	1	1	1	Lineage 1	Human	USA	2007
**2007843	J. G. Songer	NT	61	1	5	11	1	1	3	2	Lineage 2	Food	USA	2007
**2007844	J. G. Songer	NT	61	1	5	11	1	1	3	2	Lineage 2	Food	USA	2007
**2007850	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Household	USA	2007
**2007855	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Bovine	USA	2007
6600639	M. Holland	078	11	5	8	5	11	9	11	8	Lineage 5	Human	UK	unknown
6600726	M. Holland	050	18	1	1	2	5	1	3	1	Lineage 1	Human	UK	unknown
6601158	M. Holland	NT	17	1	1	2	1	1	5	3	Lineage 1	Human	UK	unknown
6601667	M. Holland	002	8	1	1	2	6	1	5	1	Lineage 1	Human	UK	unknown
6603036	M. Holland	015	10	2	1	2	1	1	3	1	Lineage 1	Human	UK	unknown
6603061	M. Holland	NT	2	1	1	2	1	5	3	1	Lineage 1	Human	UK	unknown
6604395	M. Holland	015	44	2	5	2	1	1	3	1	Lineage 1	Human	UK	unknown
6605117	M. Holland	005	*131	2	1	6	8	1	5	1	Lineage 1	Human	UK	unknown
6605475	M. Holland	015	44	2	5	2	1	1	3	1	Lineage 1	Human	UK	unknown

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
6614282	M. Holland	023	5	1	6	4	7	2	8	7	Lineage 3	Human	UK	unknown
6614376	M. Holland	078	11	5	8	5	11	9	11	8	Lineage 5	Human	UK	unknown
6615591	M. Holland	002	8	1	1	2	6	1	5	1	Lineage 1	Human	UK	unknown
6615723	M. Holland	014	49	1	1	2	1	5	3	3	Lineage 1	Human	UK	unknown
6616023	M. Holland	078	11	5	8	5	11	9	11	8	Lineage 5	Human	UK	unknown
6616104	M. Holland	078	11	5	8	5	11	9	11	8	Lineage 5	Human	UK	unknown
001-01	M. Holland	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	UK	unknown
001-02	M. Holland	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	UK	unknown
001-03	M. Holland	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	UK	unknown
001-04	M. Holland	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	UK	unknown
001-05	M. Holland	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	UK	unknown
001-06	M. Holland	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	UK	unknown
001-07	M. Holland	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	UK	unknown
001-08	M. Holland	001	2	1	1	2	1	5	3	1	Lineage 1	unknown	UK	unknown
001-09	M. Holland	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	UK	unknown
001-10	M. Holland	001	3	1	1	2	1	1	1	1	Lineage 1	unknown	UK	unknown
078W	M. Holland	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
106- 01	M. Holland	106	42	1	1	2	1	1	7	1	Lineage 1	Human	UK	unknown
106- 02	M. Holland	106	42	1	1	2	1	1	7	1	Lineage 1	Human	UK	unknown
106- 04	M. Holland	106	42	1	1	2	1	1	7	1	Lineage 1	Human	UK	unknown

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
106- 05	M. Holland	106	42	1	1	2	1	1	7	1	Lineage 1	Human	UK	unknown
106- 06	M. Holland	106	42	1	1	2	1	1	7	1	Lineage 1	Human	UK	unknown
106-07	M. Holland	106	42	1	1	2	1	1	7	1	Lineage 1	Human	UK	unknown
106-10	M. Holland	106	42	1	1	2	1	1	7	1	Lineage 1	Human	UK	unknown
17/50	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
18/21	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
19/44	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
19/52	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
19/72	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
20/28	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
22/31	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
23/41	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
25/40	E. Kuijper	078	11	5	8	5	11	9	11	8	Lineage 5	unknown	unknown	unknown
3623 -03	J. G. Songer	NT	26	1	1	6	1	4	3	4	Lineage 1	Human	Germany	< 2004
5353 (3/20)	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/03/2007
****5354 (17/4)**	J. G. Songer	027	1	1	1	1	10	1	3	5	Lineage 2	Bovine	USA	unknown
5379 (26/12)	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
5404 (17/4)	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	17/04/2007
5404 (9/26)	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	26/09/2006
5424 (2/20)	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/02/2007

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
5945 (2/21)	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/11/2006
5964 (20/11)	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	20/11/2006
5964 (9/1)	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	09/01/2007
5994 (9/4)	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	04/09/2007
6612-065	M. Holland	002	8	1	1	2	6	1	5	1	Lineage 1	Human	UK	unknown
6612-590	M. Holland	015	44	2	5	2	1	1	3	1	Lineage 1	Human	UK	unknown
6612-647	M. Holland	NT	55	1	1	6	6	1	12	12	Lineage 1	Human	UK	unknown
6612-820	M. Holland	094	12	1	1	6	4	3	5	1	Lineage 1	Human	UK	unknown
6612-978	M. Holland	094	12	1	1	6	4	3	5	1	Lineage 1	Human	UK	unknown
6613-373	M. Holland	002	8	1	1	2	6	1	5	1	Lineage 1	Human	UK	unknown
6613-833	M. Holland	087	46	4	1	6	1	1	10	1	Lineage 1	Human	UK	unknown
AI149	T. V. Riley	126	11	5	8	5	11	9	11	8	Lineage 5	Kangaroo	Australia	unknown
AI15	T. V. Riley	237	11	5	8	5	11	9	11	8	Lineage 5	Porcine	Australia	unknown
AI152	T. V. Riley	NT	11	5	8	5	11	9	11	8	Lineage 5	Porcine	Australia	unknown
AI18	T. V. Riley	014	*132	5	1	6	1	5	3	1	Lineage 1	Porcine	Australia	unknown
AI24	T. V. Riley	237	11	5	8	5	11	9	11	8	Lineage 5	Porcine	Australia	unknown
AI35	T. V. Riley	237	11	5	8	5	11	9	11	8	Lineage 5	Porcine	Australia	unknown
****B1**	D. N. Gerding	NT	63	1	1	7	5	1	3	3	Lineage 1	Human	UK	1978
BI-1	D. N. Gerding	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	26/02/1988
****BI-10**	D. N. Gerding	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	10/08/2001

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
****BI-13**	D. N. Gerding	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	09/09/2004
****BI-15**	D. N. Gerding	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	09/09/2004
****BI-2**	D. N. Gerding	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	14/01/1991
****BI-5**	D. N. Gerding	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	25/08/1995
****BI-6**	D. N. Gerding	027 (176)	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	20/05/2003
****BI-6p**	D. N. Gerding	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	09/09/2004
****BI-7**	D. N. Gerding	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	20/05/2003
****BI-8**	D. N. Gerding	027	1	1	1	1	10	1	3	5	Lineage 2	Human	USA	22/01/2004
BI-9	D. N. Gerding	001	3	1	1	2	1	1	1	1	Lineage 1	Human	USA	unknown
CD DE2	E. Kuijper	NT	3	1	1	2	1	1	1	1	Lineage 1	unknown	unknown	unknown
CD#101	E. Kuijper	NT	63	1	1	7	5	1	3	3	Lineage 1	unknown	unknown	unknown
CD#17	E. Kuijper	NT	37	3	7	3	8	6	9	11	Lineage 4	unknown	unknown	unknown
CD#371	E. Kuijper	NT	3	1	1	2	1	1	1	1	Lineage 1	unknown	unknown	unknown
****CD1**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	12/07/2007
CD1010	CDRN	274	*133	1	1	7	1	1	3	3	Lineage 1	Human	UK	20/01/2010
CD1049	CDRN	013	45	4	1	6	1	1	5	1	Lineage 1	Human	UK	03/03/2010
CD1061	CDRN	283	*134	1	3	2	15	1	3	3	Lineage 1	Human	UK	19/03/2010
CD1075	CDRN	070	55	1	1	6	6	1	12	12	Lineage 1	Human	UK	04/03/2010
CD1077	CDRN	106	*135	1	1	2	8	1	7	1	Lineage 1	Human	UK	29/03/2010
CD1079	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Human	USA	27/03/2010

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
CD1080	CDRN	052	*136	2	1	6	16	1	5	13	Lineage 1	Human	UK	27/03/2010
CD1099	CDRN	059	53	1	2	2	1	1	5	1	Lineage 1	Human	UK	02/04/2010
****CD11**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	21/05/2007
CD1108	CDRN	097	21	2	2	2	1	1	1	3	Lineage 1	Human	UK	18/04/2010
CD1128	CDRN	118	42	1	1	2	1	1	7	1	Lineage 1	Human	UK	04/05/2010
CD1132	CDRN	216	33	1	1	2	1	6	5	3	Lineage 1	Human	UK	07/05/2010
CD1141	CDRN	020	2	1	1	2	1	5	3	1	Lineage 1	Human	UK	17/05/2010
CD1143	CDRN	140	26	1	1	6	1	4	3	4	Lineage 1	Human	UK	04/05/2010
CD1144	CDRN	003	57	1	1	6	4	3	5	13	Lineage 1	Human	UK	17/05/2010
CD1149	CDRN	186	51	1	1	2	6	1	7	6	Lineage 1	Human	UK	23/05/2010
CD1153	CDRN	139	52	1	1	2	16	1	12	1	Lineage 1	Human	UK	27/05/2010
CD1157	CDRN	046	35	2	5	8	1	1	3	6	Lineage 1	Human	UK	24/05/2010
CD1165	CDRN	053	63	1	1	7	5	1	3	3	Lineage 1	Human	UK	28/05/2010
CD1170	CDRN	018	17	1	1	2	1	1	5	3	Lineage 1	Human	UK	22/05/2010
CD1171	CDRN	029	*137	1	1	2	3	1	3	1	Lineage 1	Human	UK	26/05/2010
CD1199	CDRN	021	56	1	3	6	3	1	5	1	Lineage 1	Human	UK	09/06/2010
****CD12**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	06/07/2007
CD1201	CDRN	015	10	2	1	2	1	1	3	1	Lineage 1	Human	UK	11/06/2010
CD1202	CDRN	012	54	1	4	7	1	1	3	3	Lineage 1	Human	UK	09/05/2010
CD1210	CDRN	054	43	1	7	6	1	1	5	6	Lineage 1	Human	UK	18/06/2010

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
CD1214	CDRN	011	*138	2	1	2	16	1	5	1	Lineage 1	Human	UK	04/12/2009
CD1220	CDRN	005	6	2	1	6	1	1	5	1	Lineage 1	Human	UK	20/06/2010
CD1224	CDRN	107	*139	2	1	2	1	1	1	3	Lineage 1	Human	UK	14/06/2010
****CD20**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	07/08/2007
****CD25**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	31/07/2007
CD305	CDRN	023	22	1	1	4	7	2	8	7	Lineage 3	Human	UK	24/04/2008
CD453	CDRN	023	5	1	6	4	7	2	8	7	Lineage 3	Human	UK	30/09/2008
CD527	CDRN	030	48	1	1	2	1	1	5	1	Lineage 1	Human	UK	06/01/2009
CD586	CDRN	017	37	3	7	3	8	6	9	11	Lineage 4	Human	UK	11/03/2009
****CD59**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	11/08/2007
****CD60**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	04/10/2007
****CD630**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	04/04/2009
CD637	CDRN	017	37	3	7	3	8	6	9	11	Lineage 4	Human	UK	16/04/2009
****CD679**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	16/04/2009
****CD682**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	27/04/2009
****CD683**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	23/04/2009
CD689	CDRN	064	33	1	1	2	1	6	5	3	Lineage 1	Human	UK	11/05/2009
CD714	CDRN	050	16	1	1	2	6	1	3	1	Lineage 1	Human	UK	12/06/2009
CD718	CDRN	085	39	3	7	10	8	7	2	10	Lineage 4	Human	UK	08/06/2009
CD735	CDRN	081	*139	2	1	2	1	1	1	3	Lineage 1	Human	UK	03/07/2009

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
CD742	CDRN	243	*139	2	1	2	1	1	1	3	Lineage 1	Human	UK	01/01/2009
CD759	CDRN	050	6	2	1	6	1	1	5	1	Lineage 1	Human	UK	15/07/2009
****CD762**	CDRN	111	*140	1	1	9	9	1	1	2	Lineage 2	Human	UK	31/07/2009
CD767	CDRN	015	10	2	1	2	1	1	3	1	Lineage 1	Human	UK	27/07/2009
****CD790**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	02/08/2009
****CD806**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	02/09/2009
****CD81**	CDRN	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	28/08/2007
CD816	CDRN	017	37	3	7	3	8	6	9	11	Lineage 4	Human	UK	28/09/2009
CD825	CDRN	017	37	3	7	3	8	6	9	11	Lineage 4	Human	UK	01/10/2009
CD839	CDRN	017	37	3	7	3	8	6	9	11	Lineage 4	Human	UK	29/10/2009
CD853	CDRN	062	44	2	5	2	1	1	3	1	Lineage 1	Human	UK	21/10/2009
CD871	CDRN	259	*141	1	3	7	1	3	1	6	Lineage 1	Human	UK	06/11/2009
****CD877**	CDRN	135	41	1	1	9	9	1	3	2	Lineage 2	Human	UK	22/11/2009
CD886	CDRN	116	10	2	1	2	1	1	3	1	Lineage 1	Human	UK	06/12/2009
CD909	CDRN	264	*142	8	7	14	8	6	25	15	Lineage 4	Human	UK	17/12/2009
CD914	CDRN	010	15	1	1	6	1	8	5	1	Lineage 1	Human	UK	14/12/2009
CD915	CDRN	126	11	5	8	5	11	9	11	8	Lineage 5	Human	UK	02/12/2009
CD917	CDRN	022	66	1	1	2	6	1	5	3	Lineage 1	Human	UK	20/12/2009
CD955	CDRN	268	3	1	1	2	1	1	1	1	Lineage 1	Human	UK	15/01/2010
CD959	CDRN	262	*143	1	11	6	16	1	1	1	Lineage 1	Human	UK	13/01/2010

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
CD966	CDRN	196	*144	1	1	2	2	1	5	3	Lineage 1	Human	UK	22/01/2010
CD970	CDRN	271	6	2	1	6	1	1	5	1	Lineage 1	Human	UK	23/01/2010
CD973	CDRN	002	8	1	1	2	6	1	5	1	Lineage 1	Human	UK	24/11/2009
CD134	CDRN	002	8	1	1	2	6	1	5	1	Lineage 1	Human	UK	unknown
CF3	J. Brazier	017	37	3	7	3	8	6	9	11	Lineage 4	Human	Belgium	1995
CF5	J. Brazier	017	86	3	7	3	8	6	19	11	Lineage 4	Human	Belgium	1995
****DS209/06**	M. Holland	NT	1	1	1	1	10	1	3	5	Lineage 2	unknown	UK	unknown
E327 -98	T. V. Riley	126	11	5	8	5	11	9	11	8	Lineage 5	Equine	Switzerland	< 2004
ES173	T. V. Riley	017	37	3	7	3	8	6	9	11	Lineage 4	Human	Australia	18/12/2006
ES130	T. V. Riley	280	11	5	8	5	11	9	11	8	Lineage 5	Human	unknown	unknown
ES166	T. V. Riley	281	11	5	8	5	11	9	11	8	Lineage 5	Human	Australia	28/06/1905
ES67	T. V. Riley	014	13	1	1	6	1	5	3	1	Lineage 1	unknown	Australia	unknown
****ES84**	T. V. Riley	027	1	1	1	1	10	1	3	5	Lineage 2	unknown	Canada	unknown
J9	J. G. Songer	NT	3	1	1	2	1	1	1	1	Lineage 1	Human	USA	unknown
JGS 6047	J. G. Songer	NT	3	1	1	2	1	1	1	1	Lineage 1	Equine	USA	unknown
JGS 679	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
JGS344	J. G. Songer	NT	35	2	5	8	1	1	3	6	Lineage 1	Murine	USA	unknown
JGS355	J. G. Songer	002	35	2	5	8	1	1	3	6	Lineage 1	Murine	USA	unknown
JGS356	J. G. Songer	002	35	2	5	8	1	1	3	6	Lineage 1	Murine	USA	unknown
JGS357	J. G. Songer	002	35	2	5	8	1	1	3	6	Lineage 1	Murine	USA	unknown

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
JGS360	J. G. Songer	002	35	2	5	8	1	1	3	6	Lineage 1	Murine	USA	unknown
JGS6042	J. G. Songer	002	8	1	1	2	6	1	5	1	Lineage 1	Equine	USA	unknown
JGS6050	J. G. Songer	020	2	1	1	2	1	5	3	1	Lineage 1	Canine	USA	unknown
JGS655	J. G. Songer	NT	8	1	1	2	6	1	5	1	Lineage 1	Porcine	USA	unknown
JGS673	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
JGS674	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
JGS675	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
JGS676	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
JGS677	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
JGS688	J. G. Songer	126	11	5	8	5	11	9	11	8	Lineage 5	Porcine	USA	unknown
JGS691	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Porcine	USA	unknown
JGS692	J. G. Songer	002	48	1	1	2	1	1	5	1	Lineage 1	Porcine	USA	unknown
M1	D. Drudy	NT	64	1	1	6	1	1	13	1	Lineage 1	unknown	unknown	unknown
M120	D. Drudy	078	11	5	8	5	11	9	11	8	Lineage 5	Human	Ireland	unknown
M13	D. Drudy	NT	15	1	1	6	1	8	5	1	Lineage 1	Human	unknown	unknown
M68	D. Drudy	017	37	3	7	3	8	6	9	11	Lineage 4	Human	Ireland	2006
metal 1	J. G. Songer	NT	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
****O1-027**	M. Holland	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	unknown
O1-078	M. Holland	078	11	5	8	5	11	9	11	8	Lineage 5	Human	UK	unknown
PMH13	T. V. Riley	010	15	1	1	6	1	8	5	1	Lineage 1	Human	Australia	09/01/2007

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
PMH44	T. V. Riley	126	11	5	8	5	11	9	11	8	Lineage 5	Human	Australia	16/08/2008
****R10287**	J. Brazier	027	1	1	1	1	10	1	3	5	Lineage 2	Human	France	unknown
R1040	J. Brazier	212	5	1	6	4	7	2	8	7	Lineage 3	Human	unknown	unknown
R10459	J. Brazier	106	42	1	1	2	1	1	7	1	Lineage 1	Human	unknown	unknown
****R20291**	J. Brazier	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	2006
****R20352**	J. Brazier	027	1	1	1	1	10	1	3	5	Lineage 2	Human	Canada	2005
R279	J. Brazier	NT	5	1	6	4	7	2	8	7	Lineage 3	Human	unknown	unknown
R711	J. Brazier	031	29	1	1	2	16	1	3	1	Lineage 1	Human	unknown	unknown
R714	J. Brazier	001	3	1	1	2	1	1	1	1	Lineage 1	Human	unknown	unknown
R8366	J. Brazier	001	3	1	1	2	1	1	1	1	Lineage 1	Human	unknown	unknown
R839	J. Brazier	NT	3	1	1	2	1	1	1	1	Lineage 1	Human	unknown	unknown
R894	J. Brazier	002	8	1	1	2	6	1	5	1	Lineage 1	Human	unknown	unknown
RPH101	T. V. Riley	NT	11	5	8	5	11	9	11	8	Lineage 5	Human	Australia	27/01/2007
RPH13	T. V. Riley	NT	63	1	1	7	5	1	3	3	Lineage 1	Human	Australia	24/08/2006
RPH35	T. V. Riley	087	*145	4	1	6	1	1	10	12	Lineage 1	Human	Australia	30/08/2006
RPH56	T. V. Riley	014	2	1	1	2	1	5	3	1	Lineage 1	Human	Australia	02/06/2006
RPH61	T. V. Riley	005	6	2	1	6	1	1	5	1	Lineage 1	Human	Australia	04/11/2006
RT023	M. Holland	023	5	1	6	4	7	2	8	7	Lineage 1	unknown	unknown	unknown
RT026	M. Holland	026	7	1	1	7	1	1	5	1	Lineage 1	unknown	unknown	unknown
RT042	M. Holland	042	6	2	1	6	1	1	5	1	Lineage 1	unknown	unknown	unknown

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
RT050	M. Holland	050	18	1	1	2	5	1	3	1	Lineage 1	unknown	unknown	unknown
****RT176**	M. Holland	176	1	1	1	1	10	1	3	5	Lineage 2	unknown	unknown	unknown
****S10.1014**	M. Holland	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	2010
S10.1486	M. Holland	NT	13	1	1	6	1	5	3	1	Lineage 1	Human	UK	2010
S10.1920	M. Holland	002	*146	1	1	2	6	1	5	4	Lineage 1	Human	UK	2010
S10.358	M. Holland	078	11	5	8	5	11	9	11	8	Lineage 5	Human	UK	2010
****S10.564**	M. Holland	027	1	1	1	1	10	1	3	5	Lineage 2	Human	UK	2010
Slaughter 1	J. G. Songer	078	11	5	8	5	11	9	11	8	Lineage 5	Bovine	USA	unknown
T7	M. Holland	NT	3	1	1	2	1	1	1	1	Lineage 1	Human	unknown	unknown
VPI 10463	J. G. Songer	087	46	4	1	6	1	1	10	1	Lineage 1	Human	USA	1935
WA 107	T. V. Riley	127	11	5	8	5	11	9	11	8	Lineage 5	Human	Australia	01/03/2006
WA 12	T. V. Riley	239	*147	10	8	5	11	9	11	8	Lineage 5	Human	Australia	unknown
WA 122	T. V. Riley	002	8	1	1	2	6	1	5	1	Lineage 1	Human	Australia	14/04/2006
WA 13	T. V. Riley	291	*148	5	8	5	11	9	11	20	Lineage 5	Human	Australia	31/12/2005
WA 146	T. V. Riley	NT	55	1	1	6	6	1	12	12	Lineage 1	Human	Australia	27/12/2005
WA 15	T. V. Riley	002	8	1	1	2	6	1	5	1	Lineage 1	Human	Australia	05/01/2009
WA 151	T. V. Riley	237	11	5	8	5	11	9	11	8	Lineage 5	Human	Australia	11/07/2006
WA 156	T. V. Riley	012	54	1	4	7	1	1	3	3	Lineage 1	Human	Australia	14/06/2006
WA 158	T. V. Riley	NT	34	1	5	7	1	1	3	1	Lineage 1	Human	Australia	08/06/2006
WA 161	T. V. Riley	010	15	1	1	6	1	8	5	1	Lineage 1	Human	Australia	05/07/2006

Isolate	Provider	PCR ribotype	Sequence type	Alleles							Sequence type lineage	Origin	Location	Date isolated
				adk	*atpA*	*dxr*	*glyA*	*recA*	*sodA*	*tpi*				
WA 169	T. V. Riley	081	9	1	1	6	1	1	6	1	Lineage 1	Human	Australia	02/06/2006
WA 48	T. V. Riley	127	11	5	8	5	11	9	11	8	Lineage 5	Human	Australia	14/11/2005
WA 52	T. V. Riley	014	2	1	1	2	1	5	3	1	Lineage 1	Human	Australia	02/09/2005
WA 68	T. V. Riley	009	3	1	1	2	1	1	1	1	Lineage 1	Human	Australia	09/10/2005
WA 76	T. V. Riley	001	3	1	1	2	1	1	1	1	Lineage 1	Human	Australia	13/10/2005
WA 77	T. V. Riley	127	11	5	8	5	11	9	11	8	Lineage 5	Human	Australia	14/10/2005
WA 80	T. V. Riley	NT	55	1	1	6	6	1	12	12	Lineage 1	Human	Australia	19/10/2005
WA 93	T. V. Riley	054	43	1	7	6	1	1	5	6	Lineage 1	Human	Australia	28/01/2006
WA 94	T. V. Riley	078	11	5	8	5	11	9	11	8	Lineage 5	Human	Australia	19/01/2006

Appendix 2: Bacterial isolates used in chapter 5

Isolate	Provider	Location	Date Isolated	Origin
M04-34	K. Solomon	Dublin, Ireland	2004	Human
M04-21	K. Solomon	Dublin, Ireland	2004	Human
M04-30	K. Solomon	Dublin, Ireland	2004	Human
M04-40	K. Solomon	Dublin, Ireland	2004	Human
M04-50	K. Solomon	Dublin, Ireland	2004	Human
M05-213	K. Solomon	Dublin, Ireland	2005	Human
M05-195	K. Solomon	Dublin, Ireland	2005	Human
M05-214	K. Solomon	Dublin, Ireland	2005	Human
M04-9	K. Solomon	Dublin, Ireland	2004	Human
M08-127	K. Solomon	Dublin, Ireland	2008	Human
M08-150	K. Solomon	Dublin, Ireland	2008	Human
IE 1104	K. Solomon	Dublin, Ireland	2009	Human
SV04-54	K. Solomon	Dublin, Ireland	2005	Human
S3	K. Solomon	Dublin, Ireland	2006	Human
MOH849	J. S. Weese	Ontario, Canada	2003	Human
CD257	J. S. Weese	Ontario, Canada	2003	Human
CD334	J. S. Weese	Ontario, Canada	2003	Human
CD459	J. S. Weese	Ontario, Canada	2003	Human
S- 017.8	P.M Hawkey	Dublin, Ireland	2000	Human
S- 017.9	P.M Hawkey	Dublin, Ireland	2000	Human
S- 017.10	P.M Hawkey	Dublin, Ireland	2000	Human
MOH82	J. S. Weese	Ontario, Canada	2003	Human
CD656	J. S. Weese	Ontario, Canada	2004	Bovine
CD657	J. S. Weese	Ontario, Canada	2004	Bovine
S- 017.5	P.M Hawkey	Carlisle, UK	1994	Human
GAI 95601	H. Kato	Japan	1995	Human
S- 017.13	P.M Hawkey	Japan	1999	Human
B-33	CDRN	London, UK	2011	Human
M04-106	K. Solomon	Dublin, Ireland	2005	Human
M05-185	K. Solomon	Dublin, Ireland	2005	Human
7081082	E. J. Kuijper	The Netherlands	2007	Human
8083571	E. J. Kuijper	The Netherlands	2008	Human
MRL 1135	D. N. Gerding	North Dakota, USA	2007	Human
UHL-24	CDRN	London, UK	2013	Human
MRL 981	D. N. Gerding	California, USA	2007	Human
S- 017.12	P.M Hawkey	Japan	1999	Human
11-383	H. Kim	Korea	2011	Human
PT 1307	E. J. Kuijper	Portugal	2009	Human
S- 017.22	P.M Hawkey	Poole, UK	2003	Human
S- 017.23	P.M Hawkey	Poole, UK	2003	Human

Isolate	Provider	Location	Date Isolated	Origin
S- 017.24	P.M Hawkey	Poole, UK	2003	Human
S- 017.25	P.M Hawkey	Poole, UK	2003	Human
MRL 529	D. N. Gerding	Michigan, USA	2007	Human
SV04-9	K. Solomon	Dublin, Ireland	2004	Human
SV04-17	K. Solomon	Dublin, Ireland	2004	Human
SV04-29	K. Solomon	Dublin, Ireland	2004	Human
MRL 532	D. N. Gerding	Michigan, USA	2007	Human
CD552	J. S. Weese	Ontario, Canada	2002	Equine
CD577	J. S. Weese	Ontario, Canada	2003	Equine
CD409	J. S. Weese	Ontario, Canada	2003	Human
CD159	J. S. Weese	Ontario, Canada	2003	Human
CD210	J. S. Weese	Ontario, Canada	2003	Human
CD160	J. S. Weese	Ontario, Canada	2003	Human
CD425	J. S. Weese	Ontario, Canada	2003	Human
CD345	J. S. Weese	Ontario, Canada	2003	Human
CD538	J. S. Weese	Ontario, Canada	2003	Equine
CD562	J. S. Weese	Ontario, Canada	2004	Human
CD744	J. S. Weese	Ontario, Canada	2004	Bovine
CD472	J. S. Weese	Ontario, Canada	2004	Bovine
CD748	J. S. Weese	Ontario, Canada	2004	Bovine
CD754	J. S. Weese	Ontario, Canada	2004	Bovine
CD755	J. S. Weese	Ontario, Canada	2004	Bovine
CD746	J. S. Weese	Ontario, Canada	2004	Bovine
CD762	J. S. Weese	Ontario, Canada	2004	Bovine
CD479	J. S. Weese	Ontario, Canada	2005	Canine
CD493	J. S. Weese	Ontario, Canada	2005	Canine
CD504	J. S. Weese	Ontario, Canada	2005	Canine
CD515	J. S. Weese	Ontario, Canada	2005	Canine
CD523	J. S. Weese	Ontario, Canada	2005	Canine
CD527	J. S. Weese	Ontario, Canada	2005	Canine
CD528	J. S. Weese	Ontario, Canada	2005	Canine
CD529	J. S. Weese	Ontario, Canada	2005	Canine
CD531	J. S. Weese	Ontario, Canada	2005	Canine
CD578	J. S. Weese	Ontario, Canada	2005	Equine
CD524	J. S. Weese	Ontario, Canada	2005	Canine
CD514	J. S. Weese	Ontario, Canada	2005	Canine
S- 017.15	P.M Hawkey	Kuwait	2001	Human
S- 017.16	P.M Hawkey	Kuwait	2001	Human
DS 1383	T. D. Lawley	Singapore	2003	Human
SI 032	T. D. Lawley	Singapore	2008	Human
MOH207	J. S. Weese	Ontario, Canada	2003	Human
MRL 270	D. N. Gerding	Ontario, Canada	2006	Human

Isolate	Provider	Location	Date Isolated	Origin
MRL 102	D. N. Gerding	Ohio, USA	2006	Human
09-275	H. Kim	Korea	2009	Human
BG 181	E. J. Kuijper	Bulgaria	2009	Human
GR 0108	E. J. Kuijper	Greece	2009	Human
S- 017.43	P.M Hawkey	Hong Kong	2011	Human
S- 017.75	P.M Hawkey	Coventry, UK	2011	Human
S- 017.40	P.M Hawkey	Hong Kong	2011	Human
SCH5842529	S. Reid	South Africa	2012	Human
SI 064	T. V. Riley	Singapore	2012	Human
WX-36	CDRN	London, UK	2012	Human
S- 017.79	P.M Hawkey	Nottinghamshire, UK	2012	Human
02-210	H. Kim	Korea	2002	Human
SI 1101	E. J. Kuijper	Slovenia	2009	Human
S- 017.55	P.M Hawkey	Nottinghamshire, UK	2009	Human
S- 017.91	P.M Hawkey	China	2009	Human
S- 017.32	P.M Hawkey	Hong Kong	2010	Human
S- 017.33	P.M Hawkey	Hong Kong	2010	Human
WX-35	CDRN	London, UK	2011	Human
WA 1514	T. V. Riley	Australia	2012	Human
MRL 3836	D. N. Gerding	Illinois, USA	2009	Human
S- 017.77	P.M Hawkey	Coventry, UK	2012	Human
S- 017.92	P.M Hawkey	China	2009	Human
CZ 1201	E. J. Kuijper	Czech Republic	2009	Human
675	E. J. Kuijper	Romania	2009	Human
S- 017.20	P.M Hawkey	Poland	2003	Human
S- 017.1	P.M Hawkey	Argentina	2001	Human
S- 017.2	P.M Hawkey	Argentina	2001	Human
S- 017.3	P.M Hawkey	Argentina	2001	Human
NP-25	CDRN	London, UK	2008	Human
ES 531	T. V. Riley	Australia	2011	Human
S- 017.35	P.M Hawkey	Hong Kong	2010	Human
S- 017.37	P.M Hawkey	Hong Kong	2010	Human
S- 017.38	P.M Hawkey	Hong Kong	2010	Human
ES 580	T. V. Riley	Australia	2012	Human
ES 720	T. V. Riley	Australia	2012	Human
12H001405	CDRN	London, UK	2013	Human
6797	D. N. Gerding	Argentina	2011	Human
PL 1204	E. J. Kuijper	Poland	2009	Human
01-116	H. Kim	Korea	2001	Human
S- 017.19	P.M Hawkey	Poland	2003	Human
6050595	E. J. Kuijper	The Netherlands	2006	Human
7036732	E. J. Kuijper	The Netherlands	2007	Human

Isolate	Provider	Location	Date Isolated	Origin
S- 017.58	P.M Hawkey	Nottinghamshire, UK	2010	Human
03-182	H. Kim	Korea	2003	Human
06-378	H. Kim	Korea	2006	Human
07-613	H. Kim	Korea	2007	Human
0 9-02	H. Kim	Korea	2009	Human
08-191	H. Kim	Korea	2008	Human
06-472	H. Kim	Korea	2006	Human
07-527	H. Kim	Korea	2007	Human
04-247	H. Kim	Korea	2004	Human
05-269	H. Kim	Korea	2005	Human
0 8-26	H. Kim	Korea	2008	Human
S- 017.48	P.M Hawkey	Coventry, UK	2008	Human
S- 017.47	P.M Hawkey	Coventry, UK	2008	Human
S- 017.60	P.M Hawkey	Nottinghamshire, UK	2010	Human
S- 017.62	P.M Hawkey	Nottinghamshire, UK	2010	Human
S- 017.68	P.M Hawkey	Coventry, UK	2011	Human
S- 017.69	P.M Hawkey	Coventry, UK	2011	Human
S- 017.46	P.M Hawkey	Coventry, UK	2008	Human
S- 017.50	P.M Hawkey	Coventry, UK	2008	Human
S- 017.49	P.M Hawkey	Coventry, UK	2008	Human
S- 017.51	P.M Hawkey	Coventry, UK	2008	Human
S- 017.52	P.M Hawkey	Coventry, UK	2008	Human
05-302	H. Kim	Korea	2005	Human
S- 017.72	P.M Hawkey	Walsall, UK	2011	Human
S- 017.73	P.M Hawkey	Walsall, UK	2011	Human
S- 017.64	P.M Hawkey	Walsall, UK	2011	Human
S- 017.65	P.M Hawkey	Walsall, UK	2011	Human
S- 017.66	P.M Hawkey	Walsall, UK	2011	Human
S- 017.90	P.M Hawkey	China	2009	Human
S- 017.57	P.M Hawkey	Nottinghamshire, UK	2009	Human
S- 017.59	P.M Hawkey	Nottinghamshire, UK	2010	Human
S- 017.70	P.M Hawkey	Lincolnshire, UK	2011	Human
S- 017.86	P.M Hawkey	China	2009	Human
S- 017.87	P.M Hawkey	China	2009	Human
S- 017.95	P.M Hawkey	China	2010	Human
S- 017.88	P.M Hawkey	China	2009	Human
S- 017.89	P.M Hawkey	China	2009	Human
S- 017.93	P.M Hawkey	China	2009	Human
S- 017.94	P.M Hawkey	China	2009	Human
5992	D. N. Gerding	Illinois, USA	1996	Human
S- 017.80	P.M Hawkey	Wolverhampton, UK	2012	Human
S- 017.78	P.M Hawkey	Wolverhampton, UK	2012	Human

Isolate	Provider	Location	Date Isolated	Origin
S- 017.36	P.M Hawkey	Hong Kong	2010	Human
95-24	H. Kim	Korea	1995	Human
95-25	H. Kim	Korea	1995	Human
S- 017.14	P.M Hawkey	Japan	1999	Human
T-14	P. J. Tsai	Taiwan	2011	Human
MRL 991	D. N. Gerding	California, USA	2007	Human
QM-34	CDRN	London, UK	2011	Human
S- 017.11	P.M Hawkey	Fife, UK	2002	Human
S- 017.21	P.M Hawkey	Poland	2003	Human
MRL 738	D. N. Gerding	Washington, USA	2007	Human
S- 017.30	P.M Hawkey	Hong Kong	2010	Human
S- 017.31	P.M Hawkey	Hong Kong	2010	Human
S- 017.41	P.M Hawkey	Hong Kong	2011	Human
S- 017.45	P.M Hawkey	Hong Kong	2011	Human
S- 017.44	P.M Hawkey	Hong Kong	2011	Human
WA 1196	T. V. Riley	Australia	2012	Human
S- 017.4	P.M Hawkey	Aylesbury, UK	2005	Human
00-108	H. Kim	Korea	2000	Human
5572	D. N. Gerding	Minnesota, USA	1992	Human
6084529	E. J. Kuijper	The Netherlands	2006	Human
SI 010	T. V. Riley	Singapore	2008	Human
S- 017.56	P.M Hawkey	Birmingham, UK	2009	Human
7038446	E. J. Kuijper	The Netherlands	2007	Human
7047343	E. J. Kuijper	The Netherlands	2007	Human
WA 1428	T. V. Riley	Australia	2012	Human
SI 004	T. V. Riley	Singapore	2008	Human
S- 017.67	P.M Hawkey	Coventry, UK	2011	Human
S- 017.71	P.M Hawkey	Staffordshire, UK	2011	Human
S- 017.81	P.M Hawkey	Walsall, UK	2012	Human
SI 007	T. V. Riley	Singapore	2008	Human
S- 017.76	P.M Hawkey	Coventry, UK	2012	Human
DS 21	T. D. Lawley	Singapore	2003	Human
I 6	T. V. Riley	Indonesia	2011	Human
9029054	E. J. Kuijper	The Netherlands	2009	Human
S- 017.7	P.M Hawkey	Dorchester, UK	1996	Human
NM-27	CDRN	Middlesex, UK	2005	Human
S- 017.6	P.M Hawkey	Dorchester, UK	1996	Human
4241	D. N. Gerding	Minnesota, USA	1990	Human
S- 017.53	P.M Hawkey	Worcestershire, UK	2009	Human
SI 047	T. V. Riley	Singapore	2012	Human
SI 099	T. V. Riley	Singapore	2012	Human
B-26	CDRN	London, UK	2009	Human

Isolate	Provider	Location	Date Isolated	Origin
S- 017.27	P.M Hawkey	Wrexham, UK	1996	Human
S- 017.28	P.M Hawkey	Wrexham, UK	1996	Human
S- 017.29	P.M Hawkey	Wrexham, UK	1996	Human
T-79	P. J. Tsai	Taiwan	2011	Human
T-16	P. J. Tsai	Taiwan	2011	Human
SV04-6	K. Solomon	Dublin, Ireland	2004	Human
SV05-3	K. Solomon	Dublin, Ireland	2005	Human
SV05-25	K. Solomon	Dublin, Ireland	2005	Human
SV04-10	K. Solomon	Dublin, Ireland	2004	Human
SV04-18	K. Solomon	Dublin, Ireland	2004	Human
SV04-19	K. Solomon	Dublin, Ireland	2004	Human
SCH6148880	S. Reid	South Africa	2012	Human
UHL-4	CDRN	London, UK	2009	Human
SCH5806999	S. Reid	South Africa	2012	Human
SCH6098043	S. Reid	South Africa	2012	Human
SCH5824693	S. Reid	South Africa	2012	Human
SCH5845556	S. Reid	South Africa	2012	Human
MRL 985	D. N. Gerding	California, USA	2007	Human
UHL-1	CDRN	London, UK	2005	Human
UHL-6	CDRN	London, UK	2009	Human
UHL-8	CDRN	London, UK	2009	Human
UHL-5	CDRN	London, UK	2009	Human
UHL-7	CDRN	London, UK	2009	Human
UHL-2	CDRN	London, UK	2009	Human
UHL-3	CDRN	London, UK	2009	Human
UHL-9	CDRN	London, UK	2009	Human
UHL-12	CDRN	London, UK	2010	Human
UHL-10	CDRN	London, UK	2010	Human
UHL-11	CDRN	London, UK	2010	Human
UHL-13	CDRN	London, UK	2010	Human
UHL-14	CDRN	London, UK	2010	Human
UHL-22	CDRN	London, UK	2010	Human
SI 006	T. V. Riley	Singapore	2008	Human
CDP08WTH7	T. D. Lawley	Liverpool, UK	2010	Human
UHL-15	CDRN	London, UK	2010	Human
UHL-16	CDRN	London, UK	2010	Human
UHL-17	CDRN	London, UK	2010	Human
UHL-21	CDRN	London, UK	2010	Human
UHL-23	CDRN	London, UK	2011	Human
UHL-19	CDRN	London, UK	2010	Hospital ward
UHL-21	CDRN	London, UK	2010	Hospital ward
UHL-18	CDRN	London, UK	2010	Human

Isolate	Provider	Location	Date Isolated	Origin
MRL 2923	D. N. Gerding	Colorado, USA	2008	Human
5733	D. N. Gerding	Illinois, USA	1995	Human
8070899	E. J. Kuijper	The Netherlands	2008	Human
CX-32	CDRN	London, UK	2011	Human
MRL 2259	D. N. Gerding	California, USA	2008	Human
4092	D. N. Gerding	Minnesota, USA	1990	Human
4139	D. N. Gerding	Minnesota, USA	1990	Human
H-219	D. N. Gerding	Illinois, USA	2006	Human
12H400159	CDRN	London, UK	2013	Human
5264	D. N. Gerding	Minnesota, USA	1992	Human
5265	D. N. Gerding	Minnesota, USA	1992	Human
5340	D. N. Gerding	Minnesota, USA	1993	Human
S- 017.84	P.M Hawkey	Northamptonshire, UK	2012	Human
S- 017.82	P.M Hawkey	Northamptonshire, UK	2012	Human
S- 017.83	P.M Hawkey	Northamptonshire, UK	2012	Human
GOSH-28	CDRN	London, UK	2010	Human
GOSH-29	CDRN	London, UK	2010	Human
GOSH-30	CDRN	London, UK	2010	Human
RF-31	CDRN	London, UK	2010	Human
S- 017.61	P.M Hawkey	Birmingham, UK	2010	Human
S- 017.74	P.M Hawkey	Walsall, UK	2011	Human
S- 017.85	P.M Hawkey	Northamptonshire, UK	2012	Human
WA 0908	T. V. Riley	Australia	2012	Human
GOSH-37	CDRN	London, UK	2013	Human
SCH5865760	S. Reid	South Africa	2012	Human
SCH5864722	S. Reid	South Africa	2012	Human
SCH6163235	S. Reid	South Africa	2012	Human

Appendix 3: Chapter 5 isolates: non-synonymous SNPs and predicted function (includes isolates from chapter 4)

Position in genome	Reference Base	Alternative Base	Amino Acid	Gene/Predicted Function and/or Potential Impact	Number of isolates with SNP
23868	G	T	6	*rpsJ* (tigecyline resistance)	265
1907433	T	G	282	*msrAB* (altered response to environmental stress)	256
113641	A	T	426	*gyrB* (fluoroquinolone resistance)	248
288057	C	A	56	*phnM* (degradation of phosphonate compounds)	248
1869520	A	G	144	*hisC* (histidinol-phosphate aminotransferase)	248
112752	G	A	82	*gyrA* (fluoroquinolone resistance)	179
650374	A	G	19	*MerR* (altered response to environmental stimuli)	114
2914248	A	G	257	*dacF* Reduced resistance to heat or β-lactam antibiotics	114
3604289	C	A	329	Hypothetical protein	114
34697	G	T	502	*rpoB* (rifampicin resistance)	92
34687	C	T	505	*rpoB* (rifampicin resistance)	90
1918864	T	C	292	*bioB* (altered biotin production, limiting its availability and in-turn increasing toxin synthesis)	59
3419885	A	C	314	Signalling protein	59
800885	C	T	20	PTS system, IIa component (altered sensitivity to bacteriocins)	55
72976	C	T	218	Beta-glucosidase (hydrolysis of cellobiose to two molecules of glucose)	39
3355379	C	A	346	Signalling protein	39
1204039	G	T	141	HAD superfamily hydrolase (hydrolytic enzyme reactions)	36
213561	C	T	193	Hypothetical protein	29
345335	A	C	31	Protein-tyrosine-phosphatase (control of the biosynthesis of capsular and extracellular polysaccharides)	29
465423	T	C	97	Hypothetical protein	29
1666328	T	C	87	Signalling protein	29

Position in genome	Reference Base	Alternative Base	Amino Acid	Gene/Predicted Function and/or Potential Impact	Number of isolates with SNP
1669112	G	T	392	Signalling protein	29
2155937	G	A	134	Glyoxalase (altered response to environmental stress)	29
2761655	T	C	104	MerR (altered response to environmental stimuli)	29
3304067	T	G	559	Sigma-54 (controls expression of nitrogen related genes)	29
3366100	G	T	164	Permease (membrane transporter)	29
400308	C	A	321	*cbiK* (iron transport system)	26
2764775	C	T	676	Signalling protein	26
3067079	C	G	497	Penicillin-binding protein (B-lactam resistance)	26
4025381	G	A	11	*FlgG* (altered formation of flagella)	26
3066391	G	A	726	Penicillin-binding protein (B-lactam resistance)	25
3202066	T	A	332	ABC transporter, permease/ATP-binding protein	25
578215	C	T	252	Iron-only hydrogenase	23
707105	C	A	98	ABC transporter, permease protein	23
1123155	G	A	125	Membrane protein	23
1541265	G	A	394	Sensor histidine kinase	23
2236997	C	T	429	aroA	23
3072208	C	G	66	maf	23
3403871	G	T	115	plfA	23
2881638	T	G	223	tepA	20
457651	C	A	37	Multidrug efflux protein	19
1178734	G	T	223	Cell surface protein	19
1942550	T	A	158	Signalling protein	19
1997615	A	T	133	Xanthine/uracile permease	19
3053095	C	T	162	obg	19
3364434	G	A	227	CoA-transferase	19

Position in genome	Reference Base	Alternative Base	Amino Acid	Gene/Predicted Function and/or Potential Impact	Number of isolates with SNP
3916418	G	A	239	Hypothetical protein	19
3511614	T	C	9	Transporter	18
132573	G	A	330	Amino acid aminotransferase	16
850525	C	T	163	bglG1	16
2764219	G	T	491	Signalling protein	16
1032720	C	T	275	ABC transporter, substrate-binding protein	15
2060217	G	A	67	argC	15
2075525	G	T	854	clpB	15
2800650	C	A	356	Oxidoreductase	15
3382194	G	A	555	Penicillin-binding protein (B-lactam resistance)	15
4024527	G	A	31	flgG	15
113642	C	T	426	gyrB	13
2284466	G	T	174	Nitrite and sulfite reductase subunit	13
2733799	G	A	55	folP	13
3399853	G	A	310	Ca2+/Na+ antiporter	13
4184694	T	A	197	buk	13
50281	C	A	317	Dual-specificity prolyl/cysteinyl-tRNA synthetase	12
1115288	G	A	48	Aromatic amino acid aminotransferase	12
1303022	G	A	9	Lipoprotein	12
1394996	C	A	145	TetR-family transcriptional regulator	12
3066199	A	G	790	Penicillin-binding protein (B-lactam resistance)	12
3066407	A	G	721	Penicillin-binding protein (B-lactam resistance)	12
2764414	G	A	556	Diguanylate kinase signalling protein	11
3066235	G	A	778	Penicillin-binding protein (B-lactam resistance)	11
1382195	A	G	641	Exported protein	10

Position in genome	Reference Base	Alternative Base	Amino Acid	Gene/Predicted Function and/or Potential Impact	Number of isolates with SNP
1977841	G	T	212	Xanthine/uracil permease	10
2961596	C	T	498	smc	10
3174341	A	C	90	Cell surface protein (Putative N-acetylmuramoyl-L-alanine amidase)	10
3174343	T	C	89	Cell surface protein (Putative N-acetylmuramoyl-L-alanine amidase)	10
388057	G	T	323	cbiP	9
470760	G	C	264	bclA3	9
585426	C	T	50	Peptidase	9
657360	G	T	515	Sensor histidine kinase	9
1048935	G	A	67	Drug/sodium antiporter	9
1276180	C	A	78	Hypothetical protein	9
1381396	A	G	375	Exported protein	9
1458207	C	T	358	Bifunctional protein	9
1474488	C	T	364	Sodium: dicarboxylate symporter family protein	9
1538843	T	C	249	selB	9
2495508	G	A	81	Exported protein	9
2646577	A	G	339	Iron-sulfur protein	9
2656639	C	T	691	feoB2	9
2902717	T	C	473	acd	9
3066331	G	A	746	Penicillin-binding protein (B-lactam resistance)	9
3166570	C	T	449	Helicase	9
3229965	C	A	675	Two-component sensor histidine kinase	9
3300923	C	A	281	PTS system lactose/cellobiose-family transporter subunit IIC	9
3514262	G	A	302	Transposase-like protein b	9
3514307	A	G	287	Transposase-like protein b	9
3514802	C	T	122	Transposase-like protein b	9

Position in genome	Reference Base	Alternative Base	Amino Acid	Gene/Predicted Function and/or Potential Impact	Number of isolates with SNP
3514805	G	A	121	Transposase-like protein b	9
567449	G	T	632	Two-component sensor histidine kinase	8
1252965	A	C	86	Two-component response regulator	8
1491685	T	C	27	kstR2	8
1949814	T	G	265	Membrane protein	8
3239437	G	A	337	leuC	8
3382096	A	G	522	Penicillin-binding protein (B-lactam resistance)	8
3704987	C	T	19	sleB	8
3751894	T	G	1315	Cell surface protein (Putative hemagglutinin/adhesin) precursor	8
644128	C	T	259	Two-component sensor histidine kinase	7
2629377	C	A	475	Signalling protein	7
2636947	A	G	631	ABC transporter, permease protein	7
382360	T	G	171	Aldose epimerase00416	6
682915	G	A	305	xdhA3	6
763280	C	T	152	Hypothetical protein	6
839287	C	T	148	Transcription antiterminator	6
1165584	G	T	537	secA2	6
1828018	C	A	289	Signalling protein	6
4157880	G	A	395	PTS system, IIc component	6
15593	C	T	20	rpsE	5
51607	C	A	487	proS	5
221342	C	T	274	pyrAB1_2	5
232498	C	T	200	Hypothetical protein	5
383181	C	A	26	Hypothetical protein	5
395438	A	G	3	cbiT	5

Position in genome	Reference Base	Alternative Base	Amino Acid	Gene/Predicted Function and/or Potential Impact	Number of isolates with SNP
850279	C	T	81	bglG1	5
906912	G	A	39	glvC	5
939561	A	G	61	Hypothetical protein	5
1208476	A	G	478	Signalling protein	5
1242643	A	T	4	Aminotransferase	5
1303484	G	A	163	Lipoprotein	5
1508773	C	A	18	Two-component sensor histidine kinase	5
1694380	G	A	221	Oligoendopeptidase	5
1698282	G	A	77	grdE	5
1927579	G	T	59	Serine/threonine protein kinase	5
3066400	G	A	723	Penicillin-binding protein (B-lactam resistance)	5
3147022	G	A	117	Phosphoesterase	5
3267110	A	G	24	speE	5
51218	A	T	4	Dual-specificity prolyl/cysteinyl-tRNA synthetase	4
127314	A	G	47	soj_1	4
1873506	G	T	199	Acetyl-CoA synthetase	4
2438942	G	T	158	Lipoprotein	4
2444939	C	T	143	Iron compound ABC transporter, permease protein	4
2830244	A	T	256	Phage protein	4
2933682	C	A	225	topA	4
3581263	G	C	496	hrsA_2	4
3784489	C	T	373	Penicillin-binding protein (B-lactam resistance)	4
3850253	A	G	86	Hypothetical protein	4
34747	G	A	485	*rpoB* (rifampicin resistance)	3
464168	G	A	11	Transcriptional regulator	3

Position in genome	Reference Base	Alternative Base	Amino Acid	Gene/Predicted Function and/or Potential Impact	Number of isolates with SNP
511931	C	T	43	hydN1	3
557896	C	A	342	feoB3	3
569182	C	T	181	Hypothetical protein	3
612296	C	A	21	bclA2_1	3
660220	T	C	285	ABC transporter (salivaricin lantibiotic)	3
714135	G	A	112	Hypothetical protein	3
918006	T	C	299	Hypothetical protein	3
941032	A	G	96	Thioredoxin	3
941686	C	T	287	Aaminotransferase	3
996373	G	A	212	ABC transporter, substrate-binding protein	3
1050326	C	T	709	Signalling protein	3
1064970	A	C	394	Sigma-54-dependent transcriptional regulator	3
1161805	T	C	74	slpA	3
1162228	A	G	215	slpA	3
1185215	C	T	451	pgm2	3
1206681	G	A	259	N-acetylmuramoyl-L-alanine amidase"	3
1220630	C	T	417	Radical SAM family protein	3
1250085	C	T	138	Transferase	3
1359584	G	T	272	Extracellular solute-binding protein	3
1405479	T	C	247	Carbon starvation	3
1463838	G	A	207	Sugar transporter, permease protein	3
1553544	G	C	104	Lipoprotein	3
1795359	G	A	343	PTS system, IIbc component	3
1907960	G	T	89	Hypothetical protein	3
1920960	G	T	8	ATP/GTP-binding protein	3

Position in genome	Reference Base	Alternative Base	Amino Acid	Gene/Predicted Function and/or Potential Impact	Number of isolates with SNP
1932695	G	A	43	Membrane protein	3
2036515	T	C	212	Hypothetical protein	3
2060944	C	T	309	argC	3
2084143	T	G	96	TetR-family transcriptional regulator	3
2373940	A	G	159	Molybdopterin-guanine biosynthesis protein	3
2401429	G	T	101	Membrane protein	3
2457775	C	T	580	Calcium-transporting ATPase	3
2523615	G	A	142	Arylesterase	3
2704311	C	T	180	Ruberythrin	3
2719354	G	A	143	ABC transporter, ATP-binding protein	3
2764540	G	A	598	Signalling protein	3
2801003	G	A	239	Oxidoreductase	3
2943677	G	A	66	Hypothetical protein	3
2987909	C	A	103	Penicillin-binding protein (B-lactam resistance)	3
3036429	G	A	47	fabH	3
3067082	T	C	496	Penicillin-binding protein (B-lactam resistance)	3
3253744	G	T	343	opuCC	3
3306641	T	A	136	oppF	3
3340579	A	C	78	Oxidative stress regulatory protein	3
3382086	T	G	519	Penicillin-binding protein (B-lactam resistance)	3
3402470	C	T	445	plfB	3
3522538	G	A	135	RNA polymerase sigma factor	3
3626504	T	G	222	Two-component sensor histidine kinase	3
3627605	T	A	160	Hypothetical protein	3
3660789	G	A	144	Chemosensory protein	3

Position in genome	Reference Base	Alternative Base	Amino Acid	Gene/Predicted Function and/or Potential Impact	Number of isolates with SNP
3784055	C	A	518	Penicillin-binding protein (B-lactam resistance)	3
3831154	T	C	51	DNA-binding transcriptional activator YeiL	3
3832857	G	A	63	Acetyltransferase	3
3924724	G	T	628	ABC transporter, permease protein	3
4000580	T	C	5	ABC transporter, permease protein	3
4130077	G	A	76	Amidohydrolase/peptidase	3
4169330	G	A	37	fabZ	3
4180249	C	T	148	Phosphoglucomutase/phosphomannomutase mutase	3

Appendix 4: Chapter 5 isolates: total genotypic breakdown (includes isolates from chapter 4)

Non-human isolates are indicated by strain names followed by "C" (canine), "B" (bovine), "E" (equine) and "HW" (hospital ward).

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
1	M04-34	Dublin, Ireland	2004	1	0									
1	M04-21	Dublin, Ireland	2004	1	0									
1	M04-30	Dublin, Ireland	2004	1	0									
1	M04-40	Dublin, Ireland	2004	1	0									
1	M04-50	Dublin, Ireland	2004	1	0									
1	M05-213	Dublin, Ireland	2005	1	0									
1	M05-195	Dublin, Ireland	2005	1	0									
1	M05-214	Dublin, Ireland	2005	1	0									
1	M04-9	Dublin, Ireland	2004	2	2									
1	M08-127	Dublin, Ireland	2008	3	3									
1	M08-150	Dublin, Ireland	2008	3	3									
1	IE 1104	Dublin, Ireland	2009	3	3									
1	SV04-54	Dublin, Ireland	2005	4	5									
1	S3	Dublin, Ireland	2006	4	5									
1	MOH849	Ontario, Canada	2003	5	6						✓			
1	CD257	Ontario, Canada	2003	5	6						✓			
1	CD334	Ontario, Canada	2003	5	6						✓			
1	CD459	Ontario, Canada	2003	5	6						✓			

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
1	S- 017.8	Dublin, Ireland	2000	6	7									
1	S- 017.9	Dublin, Ireland	2000	6	7									
1	S- 017.10	Dublin, Ireland	2000	6	7									
1	MOH82	Ontario, Canada	2003	7	8						✓			
1	CD656 [B]	Ontario, Canada	2004	7	8						✓			
1	CD657 [B]	Ontario, Canada	2004	7	8						✓			
1	S- 017.5	Carlisle, UK	1994	8	9	✓						✓		
1	GAI 95601	Japan	1995	8	9							✓		
1	S- 017.13	Japan	1999	8	9							✓		
1	B-33	London, UK	2011	8	9							✓		
1	M04-106	Dublin, Ireland	2005	9	10	✓								
1	M05-185	Dublin, Ireland	2005	9	10									
1	7081082	The Netherlands	2007	10	10	✓						✓		
1	8083571	The Netherlands	2008	10	10							✓		
1	MRL 1135	North Dakota, USA	2007	11	10		✓					✓		
1	UHL-24	London, UK	2013	12	11							✓		
1	MRL 981	California, USA	2007	13	11		✓					✓	✓	
1	S- 017.12	Japan	1999	14	11							✓		
1	11-383	Korea	2011	15	11	✓	✓				✓	✓		
1	PT 1307	Portugal	2009	16	13	✓	✓	✓	✓		✓	✓		
1	S- 017.22	Poole, UK	2003	17	13							✓		
1	S- 017.23	Poole, UK	2003	17	13							✓		
1	S- 017.24	Poole, UK	2003	17	13							✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
1	S- 017.25	Poole, UK	2003	17	13							✓		
1	MRL 529	Michigan, USA	2007	18	13						✓	✓		
1	SV04-9	Dublin, Ireland	2004	19	14	✓								
1	SV04-17	Dublin, Ireland	2004	19	14	✓								
1	SV04-29	Dublin, Ireland	2004	19	14	✓								
1	MRL 532	Michigan, USA	2007	20	15	✓	✓				✓	✓		
1	CD552 [E]	Ontario, Canada	2002	21	15						✓	✓		
1	CD577 [E]	Ontario, Canada	2003	21	15						✓	✓		
1	CD409	Ontario, Canada	2003	21	15						✓	✓		
1	CD159	Ontario, Canada	2003	21	15						✓	✓		
1	CD210	Ontario, Canada	2003	21	15						✓	✓		
1	CD160	Ontario, Canada	2003	21	15						✓	✓		
1	CD425	Ontario, Canada	2003	21	15						✓	✓		
1	CD345	Ontario, Canada	2003	21	15		✓				✓	✓		
1	CD538 [E]	Ontario, Canada	2003	21	15						✓	✓		
1	CD562	Ontario, Canada	2004	21	15						✓	✓		
1	CD744 [B]	Ontario, Canada	2004	21	15						✓	✓		
1	CD472 [B]	Ontario, Canada	2004	21	15		✓				✓	✓		
1	CD748 [B]	Ontario, Canada	2004	21	15						✓	✓		
1	CD754 [B]	Ontario, Canada	2004	21	15						✓	✓		
1	CD755 [B]	Ontario, Canada	2004	21	15						✓	✓		
1	CD746 [B]	Ontario, Canada	2004	21	15						✓	✓		
1	CD762 [B]	Ontario, Canada	2004	21	15						✓	✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
1	CD479 [C]	Ontario, Canada	2005	21	15	✓					✓	✓		
1	CD493 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	CD504 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	CD515 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	CD523 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	CD527 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	CD528 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	CD529 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	CD531 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	CD578 [E]	Ontario, Canada	2005	21	15						✓	✓		
1	CD524 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	CD514 [C]	Ontario, Canada	2005	21	15						✓	✓		
1	S- 017.15	Kuwait	2001	22	15			✓	✓		✓	✓		
1	S- 017.16	Kuwait	2001	22	15			✓	✓		✓	✓		
1	DS 1383	Singapore	2003	23	16			✓	✓		✓	✓		
1	SI 032	Singapore	2008	23	16			✓	✓		✓	✓		
1	MOH207	Ontario, Canada	2003	24	16	✓					✓	✓		
1	MRL 270	Ontario, Canada	2006	24	16		✓				✓	✓		
1	MRL 102	Ohio, USA	2006	24	16	✓	✓				✓	✓		
1	09-275	Korea	2009	25	16	✓	✓	✓	✓		✓	✓		
1	BG 181	Bulgaria	2009	25	16	✓	✓	✓	✓		✓	✓		
1	GR 0108	Greece	2009	25	16		✓	✓	✓		✓	✓		
1	S- 017.43	Hong Kong	2011	25	16	✓	✓	✓	✓		✓	✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
1	S- 017.75	Coventry, UK	2011	25	16	✓	✓	✓	✓		✓	✓		
1	S- 017.40	Hong Kong	2011	25	16	✓	✓	✓	✓		✓	✓		
1	SCH5842529	South Africa	2012	25	16	✓	✓	✓	✓		✓	✓		
1	SI 064	Singapore	2012	25	16		✓	✓	✓		✓	✓		
1	WX-36	London, UK	2012	25	16	✓	✓	✓	✓		✓	✓		
1	S- 017.79	Nottinghamshire, UK	2012	25	16	✓	✓	✓	✓		✓	✓		
1	02-210	Korea	2002	27	17		✓	✓	✓		✓	✓		
1	SI 1101	Slovenia	2009	28	17	✓	✓	✓	✓		✓	✓		
1	S- 017.55	Nottinghamshire, UK	2009	28	17	✓	✓	✓	✓		✓	✓		
1	S- 017.91	China	2009	28	17		✓	✓	✓		✓	✓		
1	S- 017.32	Hong Kong	2010	28	17	✓	✓	✓	✓		✓	✓		
1	S- 017.33	Hong Kong	2010	28	17	✓	✓	✓	✓		✓	✓		
1	WX-35	London, UK	2011	28	17	✓	✓	✓	✓		✓	✓		
1	WA 1514	Australia	2012	29	17	✓	✓	✓	✓		✓	✓		
1	MRL 3836	Illinois, USA	2009	30	17	✓	✓				✓	✓		
1	S- 017.77	Coventry, UK	2012	31	18	✓	✓	✓	✓		✓	✓		
1	S- 017.92	China	2009	32	18	✓	✓	✓	✓		✓	✓		
1	CZ 1201	Czech Republic	2009	32	18		✓	✓	✓		✓	✓		
1	675	Romania	2009	32	18		✓	✓	✓		✓	✓		
1	S- 017.20	Poland	2003	35	18							✓		
1	S- 017.1	Argentina	2001	37	18			✓	✓		✓	✓		
1	S- 017.2	Argentina	2001	37	18		✓	✓	✓		✓	✓		
1	S- 017.3	Argentina	2001	37	18			✓	✓		✓	✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
1	NP-25	London, UK	2008	39	19	✓		✓	✓		✓	✓		
1	ES 531	Australia	2011	39	19	✓	✓	✓	✓		✓	✓		
1	S- 017.35	Hong Kong	2010	40	19	✓	✓	✓	✓		✓	✓		
1	S- 017.37	Hong Kong	2010	40	19	✓	✓	✓	✓		✓	✓		
1	S- 017.38	Hong Kong	2010	40	19	✓	✓	✓	✓		✓	✓		
1	ES 580	Australia	2012	41	19	✓	✓	✓	✓		✓	✓		
1	ES 720	Australia	2012	41	19	✓	✓	✓	✓		✓	✓		
1	12H001405	London, UK	2013	41	19	✓	✓	✓	✓		✓	✓		
1	6797	Argentina	2011	43	19			✓	✓		✓	✓		
1	PL 1204	Poland	2009	49	20							✓	✓	
1	01-116	Korea	2001	56	20			✓	✓			✓		
1	S- 017.19	Poland	2003	57	21						✓	✓		
1	6050595	The Netherlands	2006	57	21						✓	✓		
1	7036732	The Netherlands	2007	57	21						✓	✓		
1	S- 017.58	Nottinghamshire, UK	2010	62	22				✓		✓	✓		
1	03-182	Korea	2003	63	22	✓	✓	✓	✓		✓	✓		
1	06-378	Korea	2006	64	22	✓		✓	✓		✓	✓		
1	07-613	Korea	2007	64	22		✓	✓	✓		✓	✓		
1	0 9-02	Korea	2009	64	22	✓		✓	✓		✓	✓		
1	08-191	Korea	2008	65	22	✓	✓	✓	✓		✓	✓		
1	06-472	Korea	2006	66	23	✓		✓	✓		✓	✓	✓	
1	07-527	Korea	2007	66	23	✓		✓	✓		✓	✓	✓	
1	04-247	Korea	2004	67	23		✓	✓	✓		✓	✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
1	05-269	Korea	2005	67	23			✓	✓		✓	✓		
1	0 8-26	Korea	2008	67	23			✓	✓		✓	✓		
1	S- 017.48	Coventry, UK	2008	68	23	✓		✓	✓		✓	✓		
1	S- 017.47	Coventry, UK	2008	68	23			✓	✓		✓	✓		
1	S- 017.60	Nottinghamshire, UK	2010	68	23			✓	✓		✓	✓		
1	S- 017.62	Nottinghamshire, UK	2010	68	23			✓	✓		✓	✓		
1	S- 017.68	Coventry, UK	2011	68	23			✓	✓		✓	✓		
1	S- 017.69	Coventry, UK	2011	68	23			✓	✓		✓	✓		
1	S- 017.46	Coventry, UK	2008	72	24			✓	✓		✓	✓		
1	S- 017.50	Coventry, UK	2008	72	24			✓	✓		✓	✓		
1	S- 017.49	Coventry, UK	2008	72	24			✓	✓		✓	✓		
1	S- 017.51	Coventry, UK	2008	72	24			✓	✓		✓	✓		
1	S- 017.52	Coventry, UK	2008	72	24			✓	✓		✓	✓		
1	05-302	Korea	2005	75	25		✓	✓	✓		✓	✓		
1	S- 017.72	Walsall, UK	2011	79	27				✓		✓	✓		
1	S- 017.73	Walsall, UK	2011	79	27				✓		✓	✓		
1	S- 017.64	Walsall, UK	2011	79	27				✓		✓	✓		
1	S- 017.65	Walsall, UK	2011	79	27				✓		✓	✓		
1	S- 017.66	Walsall, UK	2011	79	27				✓		✓	✓		
1	S- 017.90	China	2009	82	30		✓	✓	✓		✓	✓		
1	S- 017.57	Nottinghamshire, UK	2009	83	31			✓	✓		✓	✓		
1	S- 017.59	Nottinghamshire, UK	2010	83	31			✓	✓		✓	✓		
1	S- 017.70	Lincolnshire, UK	2011	84	31	✓	✓	✓	✓		✓	✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
1	S- 017.86	China	2009	94	34		✓	✓	✓		✓	✓		
1	S- 017.87	China	2009	94	34			✓	✓		✓	✓		
1	S- 017.95	China	2010	94	34			✓	✓		✓	✓		
1	S- 017.88	China	2009	95	35			✓	✓		✓	✓		
1	S- 017.89	China	2009	95	35			✓	✓		✓	✓		
1	S- 017.93	China	2009	95	35		✓	✓	✓		✓	✓		
1	S- 017.94	China	2009	95	35			✓	✓		✓	✓		
2	5992	Illinois, USA	1996	26	17	✓						✓		
2	S- 017.80	Wolverhampton, UK	2012	33	18	✓	✓					✓		
2	S- 017.78	Wolverhampton, UK	2012	33	18	✓	✓					✓		
2	S- 017.36	Hong Kong	2010	34	18	✓	✓					✓		
2	95-24	Korea	1995	36	18	✓	✓					✓		
2	95-25	Korea	1995	36	18	✓	✓					✓		
2	S- 017.14	Japan	1999	38	19	✓						✓		
2	T-14	Taiwan	2011	38	19	✓						✓		
2	MRL 991	California, USA	2007	42	19	✓	✓					✓		
2	QM-34	London, UK	2011	42	19	✓						✓		
2	S- 017.11	Fife, UK	2002	44	19	✓	✓					✓		
2	S- 017.21	Poland	2003	44	19	✓	✓					✓		
2	MRL 738	Washington, USA	2007	44	19	✓	✓					✓		
2	S- 017.30	Hong Kong	2010	44	19		✓					✓		
2	S- 017.31	Hong Kong	2010	44	19	✓	✓					✓		
2	S- 017.41	Hong Kong	2011	44	19		✓					✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
2	S- 017.45	Hong Kong	2011	44	19	✓	✓					✓		
2	S- 017.44	Hong Kong	2011	44	19	✓	✓					✓		
2	WA 1196	Australia	2012	44	19	✓	✓					✓		
2	S- 017.4	Aylesbury, UK	2005	45	19	✓						✓	✓	
2	00-108	Korea	2000	46	19	✓						✓		
2	5572	Minnesota, USA	1992	47	20	✓		✓			✓	✓		
2	6084529	The Netherlands	2006	48	20							✓		
2	SI 010	Singapore	2008	48	20	✓						✓		
2	S- 017.56	Birmingham, UK	2009	50	20	✓						✓		
2	7038446	The Netherlands	2007	51	20	✓					✓	✓		
2	7047343	The Netherlands	2007	51	20	✓	✓				✓	✓		
2	WA 1428	Australia	2012	52	20	✓						✓		
2	SI 004	Singapore	2008	53	20	✓	✓					✓		
2	S- 017.67	Coventry, UK	2011	54	20	✓	✓					✓		
2	S- 017.71	Staffordshire, UK	2011	55	20	✓	✓					✓	✓	
2	S- 017.81	Walsall, UK	2012	55	20	✓						✓	✓	
2	SI 007	Singapore	2008	58	21			✓	✓			✓		
2	S- 017.76	Coventry, UK	2012	59	21	✓	✓					✓	✓	
2	DS 21	Singapore	2003	60	21	✓	✓					✓		
2	I 6	Indonesia	2011	61	22	✓	✓	✓	✓		✓	✓		
2	9029054	The Netherlands	2009	69	23		✓	✓	✓		✓	✓		
2	S- 017.7	Dorchester, UK	1996	70	23	✓						✓		
2	NM-27	Middlesex, UK	2005	71	24	✓					✓	✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
2	S- 017.6	Dorchester, UK	1996	73	25	✓					✓	✓		
2	4241	Minnesota, USA	1990	74	25	✓						✓		
2	S- 017.53	Worcestershire, UK	2009	74	25	✓						✓		
2	SI 047	Singapore	2012	74	25	✓	✓					✓		
2	SI 099	Singapore	2012	74	25	✓	✓					✓		
2	B-26	London, UK	2009	76	26	✓					✓	✓		
2	S- 017.27	Wrexham, UK	1996	77	26	✓		✓		✓		✓	✓	
2	S- 017.28	Wrexham, UK	1996	77	26	✓		✓		✓		✓	✓	
2	S- 017.29	Wrexham, UK	1996	77	26	✓		✓		✓		✓	✓	
2	T-79	Taiwan	2011	78	27		✓	✓	✓		✓	✓		
2	T-16	Taiwan	2011	78	27		✓	✓	✓		✓	✓		
2	SV04-6	Dublin, Ireland	2004	80	28	✓					✓	✓		
2	SV05-3	Dublin, Ireland	2005	80	28	✓					✓	✓		
2	SV05-25	Dublin, Ireland	2005	80	28	✓					✓	✓		
2	SV04-10	Dublin, Ireland	2004	81	30	✓					✓	✓		
2	SV04-18	Dublin, Ireland	2004	81	30	✓					✓	✓		
2	SV04-19	Dublin, Ireland	2004	81	30	✓					✓	✓		
2	SCH6148880	South Africa	2012	85	31		✓	✓	✓		✓	✓		
2	UHL-4	London, UK	2009	86	31	✓	✓				✓	✓		
2	SCH5806999	South Africa	2012	87	32			✓	✓		✓	✓		
2	SCH6098043	South Africa	2012	87	32			✓	✓		✓	✓		
2	SCH5824693	South Africa	2012	88	32			✓	✓			✓	✓	
2	SCH5845556	South Africa	2012	88	32			✓	✓			✓	✓	

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
2	MRL 985	California, USA	2007	89	32	✓						✓		
2	UHL-1	London, UK	2005	90	32	✓	✓				✓	✓		
2	UHL-6	London, UK	2009	90	32	✓	✓				✓	✓		
2	UHL-8	London, UK	2009	90	32	✓	✓				✓	✓		
2	UHL-5	London, UK	2009	90	32	✓	✓				✓	✓		
2	UHL-7	London, UK	2009	90	32	✓	✓				✓	✓		
2	UHL-2	London, UK	2009	90	32	✓	✓				✓	✓		
2	UHL-3	London, UK	2009	90	32	✓	✓				✓	✓		
2	UHL-9	London, UK	2009	90	32	✓	✓				✓	✓		
2	UHL-12	London, UK	2010	90	32						✓	✓		
2	UHL-10	London, UK	2010	90	32	✓	✓				✓	✓		
2	UHL-11	London, UK	2010	90	32	✓	✓				✓	✓		
2	UHL-13	London, UK	2010	90	32	✓					✓	✓		
2	UHL-14	London, UK	2010	90	32	✓					✓	✓		
2	UHL-22	London, UK	2010	90	32	✓	✓				✓	✓		
2	SI 006	Singapore	2008	91	33	✓	✓					✓		
2	CDP08WTH7	Liverpool, UK	2010	91	33	✓						✓		
2	UHL-15	London, UK	2010	92	33	✓	✓				✓	✓		
2	UHL-16	London, UK	2010	92	33	✓	✓				✓	✓		
2	UHL-17	London, UK	2010	92	33	✓	✓				✓	✓		
2	UHL-21	London, UK	2010	92	33	✓	✓				✓	✓		
2	UHL-23	London, UK	2011	92	33	✓	✓				✓	✓		
2	UHL-19 [HW]	London, UK	2010	93	34	✓	✓				✓	✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
2	UHL-21 [HW]	London, UK	2010	93	34	✓	✓				✓	✓		
2	UHL-18	London, UK	2010	93	34	✓	✓				✓	✓		
2	MRL 2923	Colorado, USA	2008	96	36	✓	✓					✓		
2	5733	Illinois, USA	1995	97	37	✓	✓				✓	✓		
2	8070899	The Netherlands	2008	98	37	✓	✓					✓		
2	CX-32	London, UK	2011	98	37	✓						✓		
2	MRL 2259	California, USA	2008	99	38	✓						✓		
2	4092	Minnesota, USA	1990	100	38	✓	✓				✓	✓		
2	4139	Minnesota, USA	1990	100	38	✓	✓				✓	✓		
2	H-219	Illinois, USA	2006	100	38	✓	✓				✓	✓		
2	12H400159	London, UK	2013	101	39	✓	✓				✓	✓		
2	5264	Minnesota, USA	1992	102	39	✓	✓				✓	✓		
2	5265	Minnesota, USA	1992	102	39	✓	✓				✓	✓		
2	5340	Minnesota, USA	1993	102	39	✓	✓				✓	✓		
2	S- 017.84	Northamptonshire, UK	2012	103	39	✓					✓	✓		
2	S- 017.82	Northamptonshire, UK	2012	103	39	✓					✓	✓		
2	S- 017.83	Northamptonshire, UK	2012	103	39	✓					✓	✓		
2	GOSH-28	London, UK	2010	104	41	✓						✓		
2	GOSH-29	London, UK	2010	105	47	✓						✓		
2	GOSH-30	London, UK	2010	105	47	✓						✓		
2	RF-31	London, UK	2010	105	47	✓	✓					✓		
2	S- 017.61	Birmingham, UK	2010	106	48	✓	✓					✓		
2	S- 017.74	Walsall, UK	2011	106	48	✓	✓					✓		

Sub-lineage	Strain Name	Location	Isolation Year	Haplotype	No. of SNPs	Insertion	Deletion	Rifampicin resistance			Fluoroquinolone resistance			Resistance inferred
								34,687	34,697	34,747	112,752	113,641	113,642	Position
								rpoB	*rpoB*	*rpoB*	*gyrA*	*gyrB*	*gyrB*	Gene
								R505K	H502N	S485F	T82I	V426D	V426I	*Amino acid change
2	S- 017.85	Northamptonshire, UK	2012	106	48	✓	✓					✓		
2	WA 0908	Australia	2012	106	48	✓	✓					✓		
2	GOSH-37	London, UK	2013	107	49	✓						✓		
2	SCH5865760	South Africa	2012	108	51			✓	✓			✓		
2	SCH5864722	South Africa	2012	109	52		✓	✓	✓		✓	✓		
2	SCH6163235	South Africa	2012	109	52			✓	✓		✓	✓		

Communications arising from this thesis

The global emergence, evolution and transmission of the *Clostridium difficile* toxin A-B+ PCR Ribotype 017.

Michelle D. Cairns [1,2,3] Richard A. Stabler [2] and Brendan W. Wren [2].

[1] Public Health Laboratory London, 3rd Floor Pathology & Pharmacy Building, 80 Newark Street, London E1 2ES. UK.
[2] Department of Pathogen Molecular Biology, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK.
[3] UCL Centre for Clinical Microbiology, University College London, Royal Free Campus, Rowland Hill Street, London, NW3 2PF, UK.

INTRODUCTION

Clostridium difficile is an important nosocomial pathogen, and the rapid spread of the infamous PCR ribotype (RT) 027 strain has detracted attention from other potentially virulent RTs. The reported incidence of CDI has changed dramatically over the last decade and phylogenetic analysis suggests this is related to the emergence of distinct clonal lineages that are highly transmissible and cause severe *C. difficile* infection (CDI). These include RTs; 027, 023, 017 and 078 and a general grouping of most other RTs[1].

As well as forming an individual lineage, RT 017 strains are interestingly toxin A-B+ yet still able to cause severe CDI and outbreaks like toxin A+B+ strains such as RT 027[2]. In the last decade, RT 017 have emerged in parts of Asia and Europe with documented reports from; The Netherlands, Israel, Japan, Dublin, Poland, China, Taiwan, Korea, South Africa and Buenos Aires[2]. The epidemiology of *C. difficile* in the Asia/Pacific regions and eastern parts of Europe also differ from elsewhere where the prevalence of RT 017 strains is higher.

In this study, we aim to define a population structure of RT 017 and identify new sub lineages and clusters.

METHODS

Global isolate collection
A total of 359 *C. difficile* isolates representing human, bovine, canine, equine, environmental and both outbreak associated and sporadic CDI were collected. Global coverage is illustrated in figure 1.

PCR ribotyping
Performed as previously described[1].

Genomic DNA extraction
Genomic DNA was extracted using phenol chloroform methodology.

Whole genome sequencing (WGS)
Performed at the Wellcome Trust Sanger Institute (WTSI) using the Illumina Genome Analyzer Iix.

Mapping and SNP detection
Raw sequence data reads were aligned and mapped with Burrows-Wheeler Aligner (BWA) against the genome of the control strain M68 to produce a binary alignment map (BAM) file. SAMtools mpileup was used to identify SNPs and to produce the variant call format (VCF) file. A visual comparison in Artemis of the BAM and VCF files was used to validate the SNP calling.

Phylogenetic analysis
SeaView was used to visualise the concatenated SNPs and generate a phylogenetic tree using neighbour-joining with 100 bootstrap iterations (figure 2).

RESULTS

To date, 198 isolates have been WGS from diverse sources, geographical locations and with isolation dates between 1995 and 2012. Test sequences were compared against the M68 reference sequenced strain.

A phylogenetic tree was produced from the WGS SNPs of those isolates sequenced (figure 2), and reveals five main clades representing; Korea, London, South Africa, China, Europe and Ireland (figure 2).

DISCUSSION

The control strain M68 clusters together with the Irish isolates; the M68 strain was isolated in Ireland in 2006 from an outbreak that affected the same two hospitals whose isolates were included in this study suggesting that they are highly similar, if not the same strain. Asian isolates have limited geographical clustering and are widely spread suggesting global transmission. Whereas isolates from South Africa demonstrate a geographically isolated cluster.

This is the first study of its kind to investigate such a diverse collection of RT 017 isolates. Completion of WGS for all isolates in this study and further analysis will provide a clearer understanding of the population genetics of this interesting lineage of *C. difficile*.

Figure 1: Global Collection *of C. difficile* isolates.

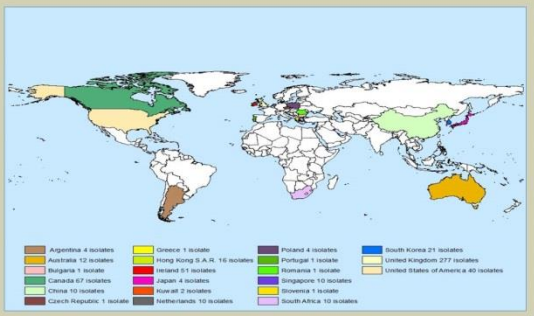
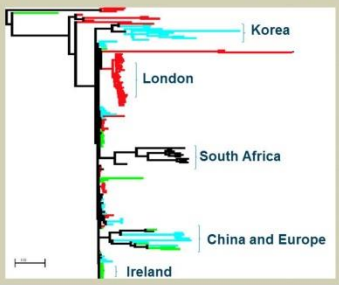


Figure 2: Maximum-likelihood phylogeny of 198 *C. difficile* RT 017 isolates based on core genome SNPs. Branch colours indicate the geographical source of the isolates and infer geographic ancestry.



CONCLUSIONS

- Completion of WGS for all isolates will enable us to produce global, national and local level phylogenetic trees which in turn will allow us to investigate the population genetics of RT 017.
- The rate of RT 017 in parts of Asia are likely an underestimate due to lack of testing and using detection methods that detect toxin B only.
- Circulating strains in Asia, as in other regions, have the potential to spread internationally. We should therefore have a heightened awareness of its prevalence and molecular epidemiology of CDI in the region.

ACKNOWLEDGEMENTS

We wish to acknowledge the National Institute for Health Research (NIHR), the Wellcome Trust (grant number 086418) and MRC (grant number PF451) for funding the research and the London *C. difficile* Ribotyping Network (CDRN) for support with PCR ribotyping.

MC is funded by a Doctoral Research Fellowship award from the NIHR. This report is independent research arising from a CSO Healthcare Scientist Award supported by the National Institute for Health Research and the CSO.

We thank; Dr Haru Kato, Dr Hee-Jung Kim, Dr Katie Solomon, Professor Ed Kuijper, Dr Koh Tse Hsien, Professor Peter Hawkey, Professor Val Hall, Professor Scott Weese, Professor Dale Gerding, Professor Tom Riley, Professor Sharon J Reid and the London CDRN for contributing strains. We also thank Dr Trevor Lawley and David Harris at the WTSI for sequencing.

REFERENCES

1. Stabler R.A. *et al;* Macro and Micro Diversity of *Clostridium difficile* Isolates from Diverse Sources and Geographical Locations. *PloS one* 7(3), e31559 (2012).
2. Cairns M.D. *et al;* The continually evolving *Clostridium difficile* species. Future Medicine. Vol. 7, No. 8, Pages 945-957 (2012).

Public Health England

Protecting and improving the nation's health

Phenotypic characterisation of toxin A negative, *Clostridium difficile* PCR Ribotype 017 strains of different sub-lineages in London, England.

Michelle D. Cairns [1,2,3] Esmeralda Valiente[2] and Brendan W. Wren [2].

[1] Public Health Laboratory London, 3rd Floor Pathology & Pharmacy Building, 80 Newark Street, London E1 2ES. UK.
[2] Department of Pathogen Molecular Biology, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK.
[3] UCL Centre for Clinical Microbiology, University College London, Royal Free Campus, Rowland Hill Street, London, NW3 2PF, UK.

INTRODUCTION

Clostridium difficile is an important nosocomial pathogen and its reported incidence has changed dramatically over the last decade. Phylogenetic analysis suggests this is related to the emergence of distinct clonal lineages that are highly transmissible and cause severe *C. difficile* infection (CDI). These include RTs; 027, 023, 017 and 078 and a general grouping of most other RTs[1]. As well as forming an individual lineage, RT017 strains are interestingly toxin A-B+ yet still able to cause severe CDI and outbreaks like toxin A+B+ strains such as RT027[2] and in the last decade, have emerged in parts of Asia, Europe, Australia, Middle East and North and South America. It is thought that the emergence of *C. difficile* associated with an increased incidence and propensity to cause outbreaks is in part due to phenotypic traits that some strains/PCR ribotypes/lineages possess which may contribute to virulence and transmissibility.

We recently performed whole genome sequencing (WGS) on human isolates of RT017 circulating in London, United Kingdom and identified three variants with one being clonal and had persisted in a London hospital ward for at least five years[3]. Here, we try to identify a phenotype for the clonal London variant isolates that may play a role in environmental persistence and/or transmission.

METHODS

Isolates Tested

Five isolates were chosen from the London WGS study based on their temporal and geographical variation and origin.

Phenotypic Assays

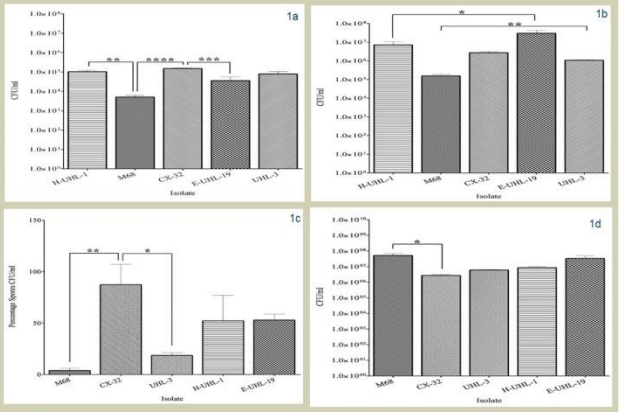
The isolates were investigated to identify characteristics that would contribute to the ability of the London variant clone to persistence and spread in a population and environment using the following phenotypic assays:

- Growth kinetics
- All cell recovery
- Sporulation characteristics
- Minimum-inhibitory concentrations
- Susceptibility to the disinfectant 'Actichlor Plus'

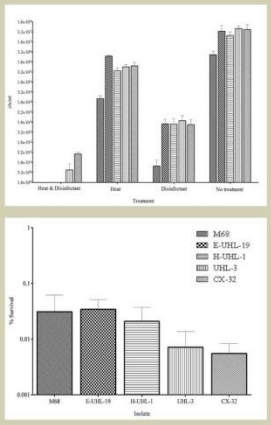
Statistical analysis

Statistical significance was assessed using Two-Way ANOVA followed with Tukey's multiple comparison tests in GraphPad Prism.

RESULTS



Figures 1a to 1d: All spores recovered at **1a)** 24 hours and **1b)** 144 hours incubation, **1c:** Percentage spores recovered at 72 hours incubation and **1d:** All cells recovered at 72 hours incubation



Figures 2a) All Achtichlor Plus assay data and **2b)** Percentage survival with Actichlor Plus treatment

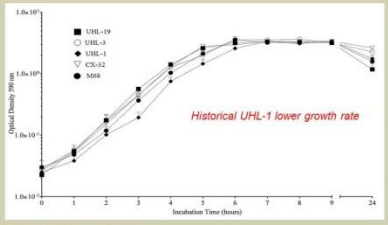


Figure 3: Growth Kinetics

Table 1: Minimum inhibitory concentrations of the five isolates tested with an antibiotic range 0 to 256µg/ml.

Isolate	MIC (µg/ml)							
	Metronidazole	Vancomycin	Erythromycin	Lincomycin	Fusidic Acid	Naladixic Acid	Rifampicin	Rifaximin
M68	1.5	1	≥ 256	≥ 256	4	128	≤ 0.5	≤ 0.5
UHL-1	1.5	1	≥ 256	≥ 256	4	128	≤ 0.5	2
UHL-19	1.5	1	≥ 256	≥ 256	4	128	≤ 0.5	≤ 0.5
UHL-3	1.5	1	≥ 256	≥ 256	4	128	≤ 0.5	≤ 0.5
CX-32	1.5	1.5	≤ 0.5	32	4	128	≤ 0.5	≤ 0.5

RESULTS

- The historical UHL-1 isolate had a statistically significant lower growth rate and a higher MIC to the antibiotic rifaximin.
- The M68 strain has a slower sporulation rate at 24 and 72 hours, but all isolates peak at a similar CFU by 144 hours.
- There was no significant differences observed in the percentage spore production at 24 or 144 h or the cell counts at 24 and 144 hours.
- There were no significant variations in susceptibility to the disinfectant Actichlor Plus.
- The non-UHL isolate CX-32 had significantly lower MICs to the macrolides erythromycin and lincomycin.

CONCLUSIONS

Although there was phenotypic variation between isolates, this was not found to be specific to the London variant clone. How the in vitro phenotypic characteristics relate to those in vivo is still unknown.

ACKNOWLEDGEMENTS

The National Institute for Health Research (NIHR). MC is funded by a Doctoral Research Fellowship award from the NIHR. This report is independent research arising from a CSO Healthcare Scientist Award supported by the National Institute for Health Research and the CSO.

REFERENCES

1. Stabler R.A. *et al;* Macro and Micro Diversity of *Clostridium difficile* Isolates from Diverse Sources and Geographical Locations. *PloS one* 7(3), e31559 (2012).
2. Cairns M.D. *et al;* The continually evolving *Clostridium difficile* species. Future Medicine. Vol. 7, No. 8, Pages 945-957 (2012).
3. Cairns M.D. *et al; Genomic Epidemiology of a Protracted Hospital Outbreak Caused by a Toxin A-Negative Clostridium difficile Sublineage PCR Ribotype 017 Strain in London, England.* JCM. Vol: 53, Issue 10.

Bibliography

AANENSEN, D. M., FEIL, E. J., HOLDEN, M. T., DORDEL, J., YEATS, C. A., FEDOSEJEV, A., GOATER, R., CASTILLO-RAMIREZ, S., CORANDER, J., COLIJN, C., CHLEBOWICZ, M. A., SCHOULS, L., HECK, M., PLUISTER, G., RUIMY, R., KAHLMETER, G., AHMAN, J., MATUSCHEK, E., FRIEDRICH, A. W., PARKHILL, J., BENTLEY, S. D., SPRATT, B. G., GRUNDMANN, H. & EUROPEAN, S. R. L. W. G. 2016. Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive Staphylococcus aureus in Europe. *MBio,* 7.

ACKERMANN, G., TANG, Y. J., KUEPER, R., HEISIG, P., RODLOFF, A. C., SILVA, J., JR. & COHEN, S. H. 2001. Resistance to moxifloxacin in toxigenic Clostridium difficile isolates is associated with mutations in gyrA. *Antimicrob Agents Chemother,* 45**,** 2348-53.

ADLER, A., MILLER-ROLL, T., BRADENSTEIN, R., BLOCK, C., MENDELSON, B., PARIZADE, M., PAITAN, Y., SCHWARTZ, D., PELED, N., CARMELI, Y. & SCHWABER, M. J. 2015. A national survey of the molecular epidemiology of Clostridium difficile in Israel: the dissemination of the ribotype 027 strain with reduced susceptibility to vancomycin and metronidazole. *Diagn Microbiol Infect Dis,* 83**,** 21-4.

AL SAIF, N. & BRAZIER, J. S. 1996. The distribution of Clostridium difficile in the environment of South Wales. *J Med Microbiol,* 45**,** 133-7.

ALFA, M. J., KABANI, A., LYERLY, D., MONCRIEF, S., NEVILLE, L. M., AL-BARRAK, A., HARDING, G. K., DYCK, B., OLEKSON, K. & EMBIL, J. M. 2000. Characterization of a toxin A-negative, toxin B-positive strain of Clostridium difficile responsible for a nosocomial outbreak of Clostridium difficile-associated diarrhea. *J Clin Microbiol,* 38**,** 2706-14.

ANDERSON, T. F. 1949. The mechanism of adsorption of bacteriophages on host cells. *Nature of Bacterial Surfaces***,** 76-95.

ANDREWS, J. M. 2001. Determination of minimum inhibitory concentrations. *J Antimicrob Chemother,* 48 Suppl 1**,** 5-16.

ANOSOVA, N. G., COLE, L. E., LI, L., ZHANG, J., BROWN, A. M., MUNDLE, S., ZHANG, J., RAY, S., MA, F., GARRONE, P., BERTRAMINELLI, N., KLEANTHOUS, H. & ANDERSON, S. F. 2015. A Combination of Three Fully Human Toxin A- and Toxin B-Specific Monoclonal Antibodies Protects against Challenge with Highly Virulent Epidemic Strains of Clostridium difficile in the Hamster Model. *Clin Vaccine Immunol,* 22**,** 711-25.

ARONIADIS, O. C., BRANDT, L. J., GREENBERG, A., BORODY, T., KELLY, C. R., MELLOW, M., SURAWICZ, C., CAGLE, L., NESHATIAN, L., STOLLMAN, N., GIOVANELLI, A., RAY, A. & SMITH, R. 2016. Long-term Follow-up Study of Fecal Microbiota Transplantation for Severe and/or Complicated Clostridium difficile Infection: A Multicenter Experience. *J Clin Gastroenterol,* 50**,** 398-402.

ARROYO, L. G., ROUSSEAU, J. D., STAEMPFLI, H. R. & WEESE, J. S. 2005. Suspected Clostridium difficile-associated hemorrhagic diarrhea in a 1-week-old elk calf. *Can Vet J,* 46**,** 1130-1.

ASHTON, P. M., NAIR, S., PETERS, T. M., BALE, J. A., POWELL, D. G., PAINSET, A., TEWOLDE, R., SCHAEFER, U., JENKINS, C., DALLMAN, T. J., DE PINNA, E. M., GRANT, K. A. & SALMONELLA WHOLE GENOME SEQUENCING IMPLEMENTATION, G. 2016. Identification of Salmonella for public health surveillance using whole genome sequencing. *PeerJ,* 4**,** e1752.

ASSEFA, S., KEANE, T. M., OTTO, T. D., NEWBOLD, C. & BERRIMAN, M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics,* 25**,** 1968-9.

AUBRY, A., HUSSACK, G., CHEN, W., KUOLEE, R., TWINE, S. M., FULTON, K. M., FOOTE, S., CARRILLO, C. D., TANHA, J. & LOGAN, S. M. 2012. Modulation of toxin production by the flagellar regulon in Clostridium difficile. *Infect Immun,* 80**,** 3521-32.

AZARIAN, T., COOK, R. L., JOHNSON, J. A., GUZMAN, N., MCCARTER, Y. S., GOMEZ, N., RATHORE, M. H., MORRIS, J. G. & SALEMI, M. 2015. Whole-genome sequencing for outbreak investigations of methicillin-resistant Staphylococcus aureus in the neonatal intensive care unit: time for routine practice? *Infect Control Hosp Epidemiol,* 36**,** 777-85.

BABAKHANI, F., GOMEZ, A., ROBERT, N. & SEARS, P. 2011. Postantibiotic effect of fidaxomicin and its major metabolite, OP-1118, against Clostridium difficile. *Antimicrob Agents Chemother,* 55**,** 4427-9.

BABAN, S. T., KUEHNE, S. A., BARKETI-KLAI, A., CARTMAN, S. T., KELLY, M. L., HARDIE, K. R., KANSAU, I., COLLIGNON, A. & MINTON, N. P. 2013. The role of flagella in Clostridium difficile pathogenesis: comparison between a non-epidemic and an epidemic strain. *PLoS One,* 8**,** e73026.

BABCOCK, G. J., BROERING, T. J., HERNANDEZ, H. J., MANDELL, R. B., DONAHUE, K., BOATRIGHT, N., STACK, A. M., LOWY, I., GRAZIANO, R., MOLRINE, D., AMBROSINO, D. M. & THOMAS, W. D., JR. 2006. Human monoclonal antibodies directed against toxins A and B prevent Clostridium difficile-induced mortality in hamsters. *Infect Immun,* 74**,** 6339-47.

BAHRMAND, A. R., TITOV, L. P., TASBITI, A. H., YARI, S. & GRAVISS, E. A. 2009. High-level rifampin resistance correlates with multiple mutations in the rpoB gene of pulmonary tuberculosis isolates from the Afghanistan border of Iran. *J Clin Microbiol,* 47**,** 2744-50.

BALE, J., MEUNIER, D., WEILL, F. X., DEPINNA, E., PETERS, T. & NAIR, S. 2016. Characterization of new Salmonella serovars by whole-genome sequencing and traditional typing techniques. *J Med Microbiol,* 65**,** 1074-1078.

BARBUT, F., DECRE, D., LALANDE, V., BURGHOFFER, B., NOUSSAIR, L., GIGANDON, A., ESPINASSE, F., RASKINE, L., ROBERT, J., MANGEOL, A., BRANGER, C. & PETIT, J. C. 2005. Clinical features of Clostridium difficile-associated diarrhoea due to binary toxin (actin-specific ADP-ribosyltransferase)-producing strains. *J Med Microbiol,* 54**,** 181-5.

BARBUT, F., MASTRANTONIO, P., DELMEE, M., BRAZIER, J., KUIJPER, E., POXTON, I. & EUROPEAN STUDY GROUP ON CLOSTRIDIUM, D. 2007. Prospective study of Clostridium difficile infections in Europe with phenotypic and genotypic characterisation of the isolates. *Clin Microbiol Infect,* 13**,** 1048-57.

BARKETI-KLAI, A., MONOT, M., HOYS, S., LAMBERT-BORDES, S., KUEHNE, S. A., MINTON, N., COLLIGNON, A., DUPUY, B. & KANSAU, I. 2014. The flagellin FliC of Clostridium difficile is responsible for pleiotropic gene regulation during in vivo infection. *PLoS One,* 9**,** e96876.

BARTLETT, J. G. 2002. Clinical practice. Antibiotic-associated diarrhea. *N Engl J Med,* 346**,** 334-9.

BAUER, M. P., NOTERMANS, D. W., VAN BENTHEM, B. H., BRAZIER, J. S., WILCOX, M. H., RUPNIK, M., MONNET, D. L., VAN DISSEL, J. T., KUIJPER, E. J. & GROUP, E. S. 2011. Clostridium difficile infection in Europe: a hospital-based survey. *Lancet,* 377**,** 63-73.

BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., KEIRA CHEETHAM, R., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A., BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CHIARA, E. C. M., CHANG, S., NEIL COOLEY, R., CRAKE, N. R., DADA, O. O., DIAKOUMAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J., EGBUJOR, U. C., ELMORE, D. W., ETCHIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FUENTES FAJARDO, K. V., SCOTT FUREY, W., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D., HEYER, N. I., HIMS, M. M., HO, J. T., HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-9.

BERG, H. C. 2003. The rotary motor of bacterial flagella. *Annu Rev Biochem,* 72**,** 19-54.

BIANCO, M., FEDELE, G., QUATTRINI, A., SPIGAGLIA, P., BARBANTI, F., MASTRANTONIO, P. & AUSIELLO, C. M. 2011. Immunomodulatory activities of surface-layer proteins obtained from epidemic and hypervirulent Clostridium difficile strains. *J Med Microbiol,* 60**,** 1162-7.

BIGNARDI, G. E. 1998. Risk factors for Clostridium difficile infection. *J Hosp Infect,* 40**,** 1-15.

BORRIELLO, S. P., WREN, B. W., HYDE, S., SEDDON, S. V., SIBBONS, P., KRISHNA, M. M., TABAQCHALI, S., MANEK, S. & PRICE, A. B. 1992. Molecular, immunological, and biological characterization of a toxin A-negative, toxin B-positive strain of Clostridium difficile. *Infect Immun,* 60**,** 4192-9.

BOUDEAU, J., BARNICH, N. & DARFEUILLE-MICHAUD, A. 2001. Type 1 pili-mediated adherence of Escherichia coli strain LF82 isolated from Crohn's disease is involved in bacterial invasion of intestinal epithelial cells. *Mol Microbiol,* 39**,** 1272-84.

BOWMAN, R. A., O'NEILL, G. L. & RILEY, T. V. 1991. Non-radioactive restriction fragment length polymorphism (RFLP) typing of Clostridium difficile. *FEMS Microbiol Lett,* 63**,** 269-72.

BRADSHAW, W. J., KIRBY, J. M., ROBERTS, A. K., SHONE, C. C. & ACHARYA, K. R. 2017. Cwp2 from Clostridium difficile exhibits an extended three domain fold and cell adhesion in vitro. *FEBS J,* 284**,** 2886-2898.

BRAUN, V., HUNDSBERGER, T., LEUKEL, P., SAUERBORN, M. & VON EICHEL-STREIBER, C. 1996. Definition of the single integration site of the pathogenicity locus in Clostridium difficile. *Gene,* 181**,** 29-38.

BRAZIER, J. S., FAWLEY, W., FREEMAN, J. & WILCOX, M. H. 2001. Reduced susceptibility of Clostridium difficile to metronidazole. *J Antimicrob Chemother,* 48**,** 741-2.

BRODA, D. M., DELACY, K. M., BELL, R. G. & PENNEY, N. 1996. Association of psychrotrophic Clostridium spp. with deep tissue spoilage of chilled vacuum-packed lamb. *Int J Food Microbiol,* 29**,** 371-8.

BRODRICK, H. J., RAVEN, K. E., HARRISON, E. M., BLANE, B., REUTER, S., TOROK, M. E., PARKHILL, J. & PEACOCK, S. J. 2016. Whole-genome sequencing reveals transmission of vancomycin-resistant Enterococcus faecium in a healthcare network. *Genome Med,* 8**,** 4.

BROUWER, M. S., ROBERTS, A. P., HUSSAIN, H., WILLIAMS, R. J., ALLAN, E. & MULLANY, P. 2013. Horizontal gene transfer converts non-toxigenic Clostridium difficile strains into toxin producers. *Nat Commun,* 4**,** 2601.

BROUWER, M. S., WARBURTON, P. J., ROBERTS, A. P., MULLANY, P. & ALLAN, E. 2011. Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of Clostridium difficile. *PLoS One,* 6**,** e23014.

BROWN, A. C., BRYANT, J. M., EINER-JENSEN, K., HOLDSTOCK, J., HOUNIET, D. T., CHAN, J. Z., DEPLEDGE, D. P., NIKOLAYEVSKYY, V., BRODA, A., STONE, M. J., CHRISTIANSEN, M. T., WILLIAMS, R., MCANDREW, M. B., TUTILL, H., BROWN, J., MELZER, M., ROSMARIN, C., MCHUGH, T. D., SHORTEN, R. J., DROBNIEWSKI, F., SPEIGHT, G. & BREUER, J. 2015. Rapid Whole-Genome Sequencing of Mycobacterium tuberculosis Isolates Directly from Clinical Samples. *J Clin Microbiol,* 53**,** 2230-7.

BRYAN, L. E., KOWAND, S. K. & VAN DEN ELZEN, H. M. 1979. Mechanism of aminoglycoside antibiotic resistance in anaerobic bacteria: Clostridium perfringens and Bacteroides fragilis. *Antimicrob Agents Chemother,* 15**,** 7-13.

BUGGY, B. P., WILSON, K. H. & FEKETY, R. 1983. Comparison of methods for recovery of Clostridium difficile from an environmental surface. *J Clin Microbiol,* 18**,** 348-52.

BURGOS, M. J., AGUAYO, M. C., PULIDO, R. P., GALVEZ, A. & LOPEZ, R. L. 2014. Multilocus sequence typing and antimicrobial resistance in Enterococcus faecium isolates from fresh produce. *Antonie Van Leeuwenhoek,* 105**,** 413-21.

BURNS, D. A., HEAP, J. T. & MINTON, N. P. 2010a. Clostridium difficile spore germination: an update. *Res Microbiol,* 161**,** 730-4.

BURNS, D. A., HEAP, J. T. & MINTON, N. P. 2010b. The diverse sporulation characteristics of Clostridium difficile clinical isolates are not associated with type. *Anaerobe,* 16**,** 618-22.

BURNS, D. A., HEEG, D., CARTMAN, S. T. & MINTON, N. P. 2011. Reconsidering the sporulation characteristics of hypervirulent Clostridium difficile BI/NAP1/027. *PLoS One,* 6**,** e24894.

BYRD, D. R. & MATSON, S. W. 1997. Nicking by transesterification: the reaction catalysed by a relaxase. *Mol Microbiol,* 25**,** 1011-22.

CAIRNS, M. D., PRESTON, M. D., HALL, C. L., GERDING, D. N., HAWKEY, P. M., KATO, H., KIM, H., KUIJPER, E. J., LAWLEY, T. D., PITUCH, H., REID, S., KULLIN, B., RILEY, T. V., SOLOMON, K., TSAI, P. J., WEESE, J. S., STABLER, R. A. & WREN, B. W. 2017. Comparative Genome Analysis and Global Phylogeny of the Toxin Variant Clostridium difficile PCR Ribotype 017 Reveals the Evolution of Two Independent Sublineages. *J Clin Microbiol,* 55**,** 865-876.

CAIRNS, M. D., PRESTON, M. D., LAWLEY, T. D., CLARK, T. G., STABLER, R. A. & WREN, B. W. 2015. Genomic epidemiology of a protracted hospital outbreak caused by a toxin A negative, Clostridium difficile sublineage PCR Ribotype 017 strain in London, England. *J Clin Microbiol*.

CAMPBELL, A., BERG, D. E., BOTSTEIN, D., LEDERBERG, E. M., NOVICK, R. P., STARLINGER, P. & SZYBALSKI, W. 1979. Nomenclature of transposable elements in prokaryotes. *Gene,* 5**,** 197-206.

CARTER, G. P., LYRAS, D., ALLEN, D. L., MACKIN, K. E., HOWARTH, P. M., O'CONNOR, J. R. & ROOD, J. I. 2007. Binary toxin production in Clostridium difficile is regulated by CdtR, a LytTR family response regulator. *J Bacteriol,* 189**,** 7290-301.

CHESNEAU, O., MORVAN, A., AUBERT, S. & EL SOLH, N. 2000. The value of rRNA gene restriction site polymorphism analysis for delineating taxa in the genus Staphylococcus. *Int J Syst Evol Microbiol,* 50 Pt 2**,** 689-97.

CHIN, C. S., SORENSON, J., HARRIS, J. B., ROBINS, W. P., CHARLES, R. C., JEAN-CHARLES, R. R., BULLARD, J., WEBSTER, D. R., KASARSKIS, A., PELUSO, P., PAXINOS, E. E., YAMAICHI, Y., CALDERWOOD, S. B., MEKALANOS, J. J., SCHADT, E. E. & WALDOR, M. K. 2011. The origin of the Haitian cholera outbreak strain. *N Engl J Med,* 364**,** 33-42.

CHONG, P. M., LYNCH, T., MCCORRISTER, S., KIBSEY, P., MILLER, M., GRAVEL, D., WESTMACOTT, G. R., MULVEY, M. R. & CANADIAN NOSOCOMIAL INFECTION SURVEILLANCE, P. 2014. Proteomic analysis of a NAP1 Clostridium difficile clinical isolate resistant to metronidazole. *PLoS One,* 9**,** e82622.

CHUMBLER, N. M., FARROW, M. A., LAPIERRE, L. A., FRANKLIN, J. L. & LACY, D. B. 2016. Clostridium difficile Toxins TcdA and TcdB Cause Colonic Tissue Damage by Distinct Mechanisms. *Infect Immun,* 84**,** 2871-7.

CLABOTS, C. R., JOHNSON, S., BETTIN, K. M., MATHIE, P. A., MULLIGAN, M. E., SCHABERG, D. R., PETERSON, L. R. & GERDING, D. N. 1993. Development of a rapid and efficient restriction endonuclease analysis typing system for Clostridium difficile and correlation with other typing systems. *J Clin Microbiol,* 31**,** 1870-5.

CLINICAL AND LABORATORY STANDARDS INSTITUTE. 2012. Methods for antimicrobial susceptibility testing of anaerobic bacteria. 8th edition ed.

CLINICAL AND LABORATORY STANDARDS INSTITUTE. 2013. RE: Performance standards for antimicrobial susceptibility testing.

COLLINS, D. A., HAWKEY, P. M. & RILEY, T. V. 2013. Epidemiology of Clostridium difficile infection in Asia. *Antimicrob Resist Infect Control,* 2**,** 21.

CURRY, S. R., MARSH, J. W., MUTO, C. A., O'LEARY, M. M., PASCULLE, A. W. & HARRISON, L. H. 2007. tcdC genotypes associated with severe TcdC truncation in an epidemic clone and other strains of Clostridium difficile. *J Clin Microbiol,* 45**,** 215-21.

CURRY, S. R., MARSH, J. W., SHUTT, K. A., MUTO, C. A., O'LEARY, M. M., SAUL, M. I., PASCULLE, A. W. & HARRISON, L. H. 2009. High frequency of rifampin resistance identified in an epidemic Clostridium difficile clone from a large teaching hospital. *Clin Infect Dis,* 48**,** 425-9.

DAVIES, K. A., ASHWIN, H., LONGSHAW, C. M., BURNS, D. A., DAVIS, G. L., WILCOX, M. H. & GROUP, E. S. 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. *Euro Surveill,* 21.

DAWSON, L. F., VALIENTE, E., DONAHUE, E. H., BIRCHENOUGH, G. & WREN, B. W. 2011. Hypervirulent Clostridium difficile PCR-ribotypes exhibit resistance to widely used disinfectants. *PLoS One,* 6**,** e25754.

DE ALMEIDA, M. N., HEFFERNAN, H., DERVAN, A., BAKKER, S., FREEMAN, J. T., BHALLY, H., TAYLOR, S. L., RILEY, T. V. & ROBERTS, S. A. 2013. Severe Clostridium difficile infection in New Zealand associated with an emerging strain, PCR-ribotype 244. *N Z Med J,* 126**,** 9-14.

DE BEEN, M., PINHOLT, M., TOP, J., BLETZ, S., MELLMANN, A., VAN SCHAIK, W., BROUWER, E., ROGERS, M., KRAAT, Y., BONTEN, M., CORANDER, J., WESTH, H., HARMSEN, D. & WILLEMS, R. J. 2015. A core genome MLST scheme for high-resolution typing of Enterococcus faecium. *J Clin Microbiol*.

DEAKIN, L. J., CLARE, S., FAGAN, R. P., DAWSON, L. F., PICKARD, D. J., WEST, M. R., WREN, B. W., FAIRWEATHER, N. F., DOUGAN, G. & LAWLEY, T. D. 2012. The Clostridium difficile spo0A gene is a persistence and transmission factor. *Infect Immun,* 80**,** 2704-11.

DEBAST, S. B., VAN LEENGOED, L. A., GOORHUIS, A., HARMANUS, C., KUIJPER, E. J. & BERGWERFF, A. A. 2009. Clostridium difficile PCR ribotype 078 toxinotype V found in diarrhoeal pigs identical to isolates from affected humans. *Environ Microbiol,* 11**,** 505-11.

DEPITRE, C., DELMEE, M., AVESANI, V., L'HARIDON, R., ROELS, A., POPOFF, M. & CORTHIER, G. 1993. Serogroup F strains of Clostridium difficile produce toxin B but not toxin A. *J Med Microbiol,* 38**,** 434-41.

DIDELOT, X. & MAIDEN, M. C. 2010. Impact of recombination on bacterial evolution. *Trends Microbiol,* 18**,** 315-22.

DINGLE, K. E., DIDELOT, X., QUAN, T. P., EYRE, D. W., STOESSER, N., GOLUBCHIK, T., HARDING, R. M., WILSON, D. J., GRIFFITHS, D.,

VAUGHAN, A., FINNEY, J. M., WYLLIE, D. H., OAKLEY, S. J., FAWLEY, W. N., FREEMAN, J., MORRIS, K., MARTIN, J., HOWARD, P., GORBACH, S., GOLDSTEIN, E. J. C., CITRON, D. M., HOPKINS, S., HOPE, R., JOHNSON, A. P., WILCOX, M. H., PETO, T. E. A., WALKER, A. S., CROOK, D. W. & MODERNISING MEDICAL MICROBIOLOGY INFORMATICS, G. 2017. Effects of control interventions on Clostridium difficile infection in England: an observational study. *Lancet Infect Dis,* 17**,** 411-421.

DINGLE, K. E., GRIFFITHS, D., DIDELOT, X., EVANS, J., VAUGHAN, A., KACHRIMANIDOU, M., STOESSER, N., JOLLEY, K. A., GOLUBCHIK, T., HARDING, R. M., PETO, T. E., FAWLEY, W., WALKER, A. S., WILCOX, M. & CROOK, D. W. 2011a. Clinical Clostridium difficile: clonality and pathogenicity locus diversity. *PLoS One,* 6**,** e19993.

DINGLE, T. C., MULVEY, G. L. & ARMSTRONG, G. D. 2011b. Mutagenic analysis of the Clostridium difficile flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. *Infect Immun,* 79**,** 4061-7.

DRESLER, J., KRUTOVA, M., FUCIKOVA, A., KLIMENTOVA, J., HRUZOVA, V., DURACOVA, M., HOUDKOVA, K., SALOVSKA, B., MATEJKOVA, J., HUBALEK, M., PAJER, P., PISA, L. & NYC, O. 2017. Analysis of proteomes released from in vitro cultured eight Clostridium difficile PCR ribotypes revealed specific expression in PCR ribotypes 027 and 176 confirming their genetic relatedness and clinical importance at the proteomic level. *Gut Pathog,* 9**,** 45.

DRUDY, D., FANNING, S. & KYNE, L. 2007a. Toxin A-negative, toxin B-positive Clostridium difficile. *Int J Infect Dis,* 11**,** 5-10.

DRUDY, D., HARNEDY, N., FANNING, S., HANNAN, M. & KYNE, L. 2007b. Emergence and control of fluoroquinolone-resistant, toxin A-negative, toxin B-positive Clostridium difficile. *Infect Control Hosp Epidemiol,* 28**,** 932-40.

DRUDY, D., HARNEDY, N., FANNING, S., O'MAHONY, R. & KYNE, L. 2007c. Isolation and characterisation of toxin A-negative, toxin B-positive Clostridium difficile in Dublin, Ireland. *Clin Microbiol Infect,* 13**,** 298-304.

DRUDY, D., KYNE, L., O'MAHONY, R. & FANNING, S. 2007d. gyrA mutations in fluoroquinolone-resistant Clostridium difficile PCR-027. *Emerg Infect Dis,* 13**,** 504-5.

DRUDY, D., QUINN, T., O'MAHONY, R., KYNE, L., O'GAORA, P. & FANNING, S. 2006. High-level resistance to moxifloxacin and gatifloxacin associated with a novel mutation in gyrB in toxin-A-negative, toxin-B-positive Clostridium difficile. *J Antimicrob Chemother,* 58**,** 1264-7.

DRUMMOND, A. J., SUCHARD, M. A., XIE, D. & RAMBAUT, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol,* 29**,** 1969-73.

DUPUY, B., GOVIND, R., ANTUNES, A. & MATAMOUROS, S. 2008. Clostridium difficile toxin synthesis is negatively regulated by TcdC. *J Med Microbiol,* 57**,** 685-9.

ELDAR, A., LAWHON, S., FRELIER, P. F., ASSENTA, L., SIMPSON, B. R., VARNER, P. W. & BERCOVIER, H. 1997. Restriction fragment length polymorphisms of 16S rDNA and of whole rRNA genes (ribotyping) of Streptococcus iniae strains from the United States and Israel. *FEMS Microbiol Lett,* 151**,** 155-62.

ELLIOTT, B., REED, R., CHANG, B. J. & RILEY, T. V. 2009. Bacteremia with a large clostridial toxin-negative, binary toxin-positive strain of Clostridium difficile. *Anaerobe,* 15**,** 249-51.

ELLIOTT, B., SQUIRE, M. M., THEAN, S., CHANG, B. J., BRAZIER, J. S., RUPNIK, M. & RILEY, T. V. 2011. New types of toxin A-negative, toxin B-positive strains among clinical isolates of Clostridium difficile in Australia. *J Med Microbiol,* 60**,** 1108-11.

ENRIGHT, M. C., SPRATT, B. G., KALIA, A., CROSS, J. H. & BESSEN, D. E. 2001. Multilocus sequence typing of Streptococcus pyogenes and the relationships between emm type and clone. *Infect Immun,* 69**,** 2416-27.

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL 2013. Point prevalence survey of healthcare associated infections and antimicrobial use in European acute care hospital.

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL. 2016. Expert opinion on whole genome sequencing for public health surveillance. Available: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/whole-genome-sequencing-for-public-health-surveillance.pdf.

EYRE, D. W., CULE, M. L., WILSON, D. J., GRIFFITHS, D., VAUGHAN, A., O'CONNOR, L., IP, C. L., GOLUBCHIK, T., BATTY, E. M., FINNEY, J. M., WYLLIE, D. H., DIDELOT, X., PIAZZA, P., BOWDEN, R., DINGLE, K. E., HARDING, R. M., CROOK, D. W., WILCOX, M. H., PETO, T. E. & WALKER, A. S. 2013. Diverse sources of C. difficile infection identified on whole-genome sequencing. *N Engl J Med,* 369**,** 1195-205.

EYRE, D. W., FAWLEY, W. N., RAJGOPAL, A., SETTLE, C., MORTIMER, K., GOLDENBERG, S. D., DAWSON, S., CROOK, D. W., PETO, T. E. A., WALKER, A. S. & WILCOX, M. H. 2017. Comparison of Control of Clostridium difficile Infection in Six English Hospitals Using Whole-Genome Sequencing. *Clin Infect Dis,* 65**,** 433-441.

EYRE, D. W., GOLUBCHIK, T., GORDON, N. C., BOWDEN, R., PIAZZA, P., BATTY, E. M., IP, C. L., WILSON, D. J., DIDELOT, X., O'CONNOR, L., LAY, R., BUCK, D., KEARNS, A. M., SHAW, A., PAUL, J., WILCOX, M. H., DONNELLY, P. J., PETO, T. E., WALKER, A. S. & CROOK, D. W. 2012. A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. *BMJ Open,* 2.

EYRE, D. W., TRACEY, L., ELLIOTT, B., SLIMINGS, C., HUNTINGTON, P. G., STUART, R. L., KORMAN, T. M., KOTSIOU, G., MCCANN, R., GRIFFITHS, D., FAWLEY, W. N., ARMSTRONG, P., DINGLE, K. E., WALKER, A. S., PETO, T. E., CROOK, D. W., WILCOX, M. H. & RILEY, T. V. 2015. Emergence and spread of predominantly community-onset Clostridium difficile PCR ribotype 244 infection in Australia, 2010 to 2012. *Euro Surveill,* 20**,** 21059.

FAGAN, R. P. & FAIRWEATHER, N. F. 2014. Biogenesis and functions of bacterial S-layers. *Nat Rev Microbiol,* 12**,** 211-22.

FARROW, K. A., LYRAS, D. & ROOD, J. I. 2000. The macrolide-lincosamide-streptogramin B resistance determinant from Clostridium difficile 630 contains two erm(B) genes. *Antimicrob Agents Chemother,* 44**,** 411-3.

FAWLEY, W. N., UNDERWOOD, S., FREEMAN, J., BAINES, S. D., SAXTON, K., STEPHENSON, K., OWENS, R. C., JR. & WILCOX, M. H. 2007. Efficacy of hospital cleaning agents and germicides against epidemic Clostridium difficile strains. *Infect Control Hosp Epidemiol,* 28**,** 920-5.

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 Genet,* 9**,** e1003660.

FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J. F., DOUGHERTY, B. A., MERRICK, J. M. & ET AL. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science,* 269**,** 496-512.

FOGLIA, G., SHAH, S., LUXEMBURGER, C. & PIETROBON, P. J. 2012. Clostridium difficile: development of a novel candidate vaccine. *Vaccine,* 30**,** 4307-9.

FOWLER, S., WEBBER, A., COOPER, B. S., PHIMISTER, A., PRICE, K., CARTER, Y., KIBBLER, C. C., SIMPSON, A. J. & STONE, S. P. 2007. Successful use of feedback to improve antibiotic prescribing and reduce Clostridium difficile infection: a controlled interrupted time series. *J Antimicrob Chemother,* 59**,** 990-5.

FRALEY, C., RAFERTY, A. E., MURPHY, B. & SCRUCCA, L. 2012. *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation* [Online]. Available: https://www.researchgate.net/publication/257428214_MCLUST_Version_4_for_R_Normal_Mixture_Modeling_for_Model-Based_Clustering_Classification_and_Density_Estimation.

FREEMAN, J., BAINES, S. D., SAXTON, K. & WILCOX, M. H. 2007. Effect of metronidazole on growth and toxin production by epidemic Clostridium difficile PCR ribotypes 001 and 027 in a human gut model. *J Antimicrob Chemother,* 60**,** 83-91.

FREEMAN, J., BAUER, M. P., BAINES, S. D., CORVER, J., FAWLEY, W. N., GOORHUIS, B., KUIJPER, E. J. & WILCOX, M. H. 2010. The changing epidemiology of Clostridium difficile infections. *Clin Microbiol Rev,* 23**,** 529-49.

FREEMAN, J., VERNON, J., MORRIS, K., NICHOLSON, S., TODHUNTER, S., LONGSHAW, C., WILCOX, M. H. & PAN-EUROPEAN LONGITUDINAL SURVEILLANCE OF ANTIBIOTIC RESISTANCE AMONG PREVALENT CLOSTRIDIUM DIFFICILE RIBOTYPES' STUDY, G. 2015. Pan-European longitudinal surveillance of antibiotic resistance among prevalent Clostridium difficile ribotypes. *Clin Microbiol Infect,* 21**,** 248 e9-248 e16.

FRICKE, W. F. & RASKO, D. A. 2014. Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nat Rev Genet,* 15**,** 49-55.

FROST, L. S., LEPLAE, R., SUMMERS, A. O. & TOUSSAINT, A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol,* 3**,** 722-32.

GARDY, J. L., JOHNSTON, J. C., HO SUI, S. J., COOK, V. J., SHAH, L., BRODKIN, E., REMPEL, S., MOORE, R., ZHAO, Y., HOLT, R., VARHOL, R., BIROL, I., LEM, M., SHARMA, M. K., ELWOOD, K.,

JONES, S. J., BRINKMAN, F. S., BRUNHAM, R. C. & TANG, P. 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med,* 364**,** 730-9.

GARDY, J. L., NAUS, M., AMLANI, A., CHUNG, W., KIM, H., TAN, M., SEVERINI, A., KRAJDEN, M., PUDDICOMBE, D., SAHNI, V., HAYDEN, A. S., GUSTAFSON, R., HENRY, B. & TANG, P. 2015. Whole-Genome Sequencing of Measles Virus Genotypes H1 and D8 During Outbreaks of Infection Following the 2010 Olympic Winter Games Reveals Viral Transmission Routes. *J Infect Dis,* 212**,** 1574-8.

GAREY, K. W., SETHI, S., YADAV, Y. & DUPONT, H. L. 2008. Meta-analysis to assess risk factors for recurrent Clostridium difficile infection. *J Hosp Infect,* 70**,** 298-304.

GARNIER, F., TAOURIT, S., GLASER, P., COURVALIN, P. & GALIMAND, M. 2000. Characterization of transposon Tn1549, conferring VanB-type resistance in Enterococcus spp. *Microbiology,* 146 (Pt 6)**,** 1481-9.

GERDING, D. N., JOHNSON, S., RUPNIK, M. & AKTORIES, K. 2014. Clostridium difficile binary toxin CDT: mechanism, epidemiology, and potential clinical importance. *Gut Microbes,* 5**,** 15-27.

GERDING, D. N., MUTO, C. A. & OWENS, R. C., JR. 2008a. Measures to control and prevent Clostridium difficile infection. *Clin Infect Dis,* 46 Suppl 1**,** S43-9.

GERDING, D. N., MUTO, C. A. & OWENS, R. C., JR. 2008b. Treatment of Clostridium difficile infection. *Clin Infect Dis,* 46 Suppl 1**,** S32-42.

GIEL, J. L., SORG, J. A., SONENSHEIN, A. L. & ZHU, J. 2010. Metabolism of bile salts in mice influences spore germination in Clostridium difficile. *PLoS One,* 5**,** e8740.

GINALDI, L., LORETO, M. F., CORSI, M. P., MODESTI, M. & DE MARTINIS, M. 2001. Immunosenescence and infectious diseases. *Microbes Infect,* 3**,** 851-7.

GOLAN, Y. & EPSTEIN, L. 2012. Safety and efficacy of fidaxomicin in the treatment of Clostridium difficile-associated diarrhea. *Therap Adv Gastroenterol,* 5**,** 395-402.

GOORHUIS, A., BAKKER, D., CORVER, J., DEBAST, S. B., HARMANUS, C., NOTERMANS, D. W., BERGWERFF, A. A., DEKKER, F. W. & KUIJPER, E. J. 2008. Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. *Clin Infect Dis,* 47**,** 1162-70.

GOORHUIS, A., LEGARIA, M. C., VAN DEN BERG, R. J., HARMANUS, C., KLAASSEN, C. H., BRAZIER, J. S., LUMELSKY, G. & KUIJPER, E. J. 2009. Application of multiple-locus variable-number tandem-repeat analysis to determine clonal spread of toxin A-negative Clostridium difficile in a general hospital in Buenos Aires, Argentina. *Clin Microbiol Infect,* 15**,** 1080-6.

GOORHUIS, A., VAN DER KOOI, T., VAESSEN, N., DEKKER, F. W., VAN DEN BERG, R., HARMANUS, C., VAN DEN HOF, S., NOTERMANS, D. W. & KUIJPER, E. J. 2007. Spread and epidemiology of Clostridium difficile polymerase chain reaction ribotype 027/toxinotype III in The Netherlands. *Clin Infect Dis,* 45**,** 695-703.

GOUDARZI, M., GOUDARZI, H., ALEBOUYEH, M., AZIMI RAD, M., SHAYEGAN MEHR, F. S., ZALI, M. R. & ASLANI, M. M. 2013. Antimicrobial susceptibility of clostridium difficile clinical isolates in iran. *Iran Red Crescent Med J,* 15**,** 704-11.

GOULD, L. H. & LIMBAGO, B. 2010. Clostridium difficile in food and domestic animals: a new foodborne pathogen? *Clin Infect Dis,* 51**,** 577-82.

GREENBERG, R. N., MARBURY, T. C., FOGLIA, G. & WARNY, M. 2012. Phase I dose finding studies of an adjuvanted Clostridium difficile toxoid vaccine. *Vaccine,* 30**,** 2245-9.

GRIFFITHS, D., FAWLEY, W., KACHRIMANIDOU, M., BOWDEN, R., CROOK, D. W., FUNG, R., GOLUBCHIK, T., HARDING, R. M., JEFFERY, K. J., JOLLEY, K. A., KIRTON, R., PETO, T. E., REES, G., STOESSER, N., VAUGHAN, A., WALKER, A. S., YOUNG, B. C., WILCOX, M. & DINGLE, K. E. 2010. Multilocus sequence typing of Clostridium difficile. *J Clin Microbiol,* 48**,** 770-8.

HACHLER, H., BERGER-BACHI, B. & KAYSER, F. H. 1987. Genetic characterization of a Clostridium difficile erythromycin-clindamycin resistance determinant that is transferable to Staphylococcus aureus. *Antimicrob Agents Chemother,* 31**,** 1039-45.

HALL, I. C. & O'TOOLE, E. 1935. Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, Bacillus difficilis. *Am J Dis Child***,** 390–402.

HARRIS, S. R., FEIL, E. J., HOLDEN, M. T., QUAIL, M. A., NICKERSON, E. K., CHANTRATITA, N., GARDETE, S., TAVARES, A., DAY, N., LINDSAY, J. A., EDGEWORTH, J. D., DE LENCASTRE, H., PARKHILL, J., PEACOCK, S. J. & BENTLEY, S. D. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science,* 327**,** 469-74.

HARRISON, E. M., LUDDEN, C., BRODRICK, H. J., BLANE, B., BRENNAN, G., MORRIS, D., COLL, F., REUTER, S., BROWN, N. M., HOLMES, M. A., O'CONNELL, B., PARKHILL, J., TOROK, M. E., CORMICAN, M. & PEACOCK, S. J. 2016. Transmission of methicillin-resistant Staphylococcus aureus in long-term care facilities and their related healthcare networks. *Genome Med,* 8**,** 102.

HASLAM, S. C., KETLEY, J. M., MITCHELL, T. J., STEPHEN, J., BURDON, D. W. & CANDY, D. C. 1986. Growth of Clostridium difficile and production of toxins A and B in complex and defined media. *J Med Microbiol,* 21**,** 293-7.

HAWKEY, P. M., MARRIOTT, C., LIU, W. E., JIAN, Z. J., GAO, Q., LING, T. K., CHOW, V., SO, E., CHAN, R., HARDY, K., XU, L. & MANZOOR, S. 2013. Molecular epidemiology of Clostridium difficile infection in a major chinese hospital: an underrecognized problem in Asia? *J Clin Microbiol,* 51**,** 3308-13.

HAWRELAK, J. A. & MYERS, S. P. 2004. The causes of intestinal dysbiosis: a review. *Altern Med Rev,* 9**,** 180-97.

HE, M., MIYAJIMA, F., ROBERTS, P., ELLISON, L., PICKARD, D. J., MARTIN, M. J., CONNOR, T. R., HARRIS, S. R., FAIRLEY, D., BAMFORD, K. B., D'ARC, S., BRAZIER, J., BROWN, D., COIA, J. E., DOUCE, G., GERDING, D., KIM, H. J., KOH, T. H., KATO, H., SENOH, M., LOUIE, T., MICHELL, S., BUTT, E., PEACOCK, S. J., BROWN, N. M., RILEY, T.,

SONGER, G., WILCOX, M., PIRMOHAMED, M., KUIJPER, E., HAWKEY, P., WREN, B. W., DOUGAN, G., PARKHILL, J. & LAWLEY, T. D. 2013. Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nat Genet,* 45**,** 109-13.

HE, M., SEBAIHIA, M., LAWLEY, T. D., STABLER, R. A., DAWSON, L. F., MARTIN, M. J., HOLT, K. E., SETH-SMITH, H. M., QUAIL, M. A., RANCE, R., BROOKS, K., CHURCHER, C., HARRIS, D., BENTLEY, S. D., BURROWS, C., CLARK, L., CORTON, C., MURRAY, V., ROSE, G., THURSTON, S., VAN TONDER, A., WALKER, D., WREN, B. W., DOUGAN, G. & PARKHILL, J. 2010. Evolutionary dynamics of Clostridium difficile over short and long time scales. *Proc Natl Acad Sci U S A,* 107**,** 7527-32.

HEALTHCARE COMMISSION. 2007. Investigation into outbreaks of Clostridium difficile at Maidstone and Tunbridge Wells.

HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107**,** 1-8.

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**,** e32381.

HOMAN, W. L., TRIBE, D., POZNANSKI, S., LI, M., HOGG, G., SPALBURG, E., VAN EMBDEN, J. D. A. & WILLEMS, R. J. L. 2002. Multilocus Sequence Typing Scheme for Enterococcus faecium. *J Clin Microbiol,* 40**,** 1963-1971.

HOWERTON, A., PATRA, M. & ABEL-SANTOS, E. 2013. A new strategy for the prevention of Clostridium difficile infection. *J Infect Dis,* 207**,** 1498-504.

HPA, D. A. December 2008. Clostridium difficile infection: How to deal with the problem. Crown copyright 2009.

HUANG, H., WEINTRAUB, A., FANG, H., WU, S., ZHANG, Y. & NORD, C. E. 2010. Antimicrobial susceptibility and heteroresistance in Chinese Clostridium difficile strains. *Anaerobe,* 16**,** 633-5.

HUELSENBECK, J. P. & RONQUIST, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics,* 17**,** 754-5.

HUSSAIN, H. A., ROBERTS, A. P. & MULLANY, P. 2005. Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. *J Med Microbiol,* 54**,** 137-41.

IIIUMINA ®. 2016. *Nextera® XT DNA Library Prep Reference Guide* [Online]. Available: http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-guide-15031942-01.pdf.

INNS, T., ASHTON, P. M., HERRERA-LEON, S., LIGHTHILL, J., FOULKES, S., JOMBART, T., REHMAN, Y., FOX, A., DALLMAN, T., E, D. E. P., BROWNING, L., COIA, J. E., EDEGHERE, O. & VIVANCOS, R. 2017. Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of Salmonella Enteritidis. *Epidemiol Infect,* 145**,** 289-298.

JABBAR, U., LEISCHNER, J., KASPER, D., GERBER, R., SAMBOL, S. P., PARADA, J. P., JOHNSON, S. & GERDING, D. N. 2010. Effectiveness of alcohol-based hand rubs for removal of Clostridium difficile spores from hands. *Infect Control Hosp Epidemiol,* 31**,** 565-70.

JACKSON, B. R., TARR, C., STRAIN, E., JACKSON, K. A., CONRAD, A., CARLETON, H., KATZ, L. S., STROIKA, S., GOULD, L. H., MODY, R. K., SILK, B. J., BEAL, J., CHEN, Y., TIMME, R., DOYLE, M., FIELDS, A., WISE, M., TILLMAN, G., DEFIBAUGH-CHAVEZ, S., KUCEROVA, Z., SABOL, A., ROACHE, K., TREES, E., SIMMONS, M., WASILENKO, J., KUBOTA, K., POUSEELE, H., KLIMKE, W., BESSER, J., BROWN, E., ALLARD, M. & GERNER-SMIDT, P. 2016. Implementation of Nationwide Real-time Whole-genome Sequencing to Enhance Listeriosis Outbreak Detection and Investigation. *Clin Infect Dis,* 63**,** 380-6.

JANEZIC, S., POTOCNIK, M., ZIDARIC, V. & RUPNIK, M. 2016. Highly Divergent Clostridium difficile Strains Isolated from the Environment. *PLoS One,* 11**,** e0167101.

JANSEN VAN RENSBURG, M. J., SWIFT, C., CODY, A. J., JENKINS, C. & MAIDEN, M. C. 2016. Exploiting Bacterial Whole-Genome Sequencing Data for Evaluation of Diagnostic Assays: Campylobacter Species Identification as a Case Study. *J Clin Microbiol,* 54**,** 2882-2890.

JEYARATNAM, D., WHITTY, C. J., PHILLIPS, K., LIU, D., OREZZI, C., AJOKU, U. & FRENCH, G. L. 2008. Impact of rapid screening tests on acquisition of meticillin resistant Staphylococcus aureus: cluster randomised crossover trial. *BMJ,* 336**,** 927-30.

JHUNG, M. A., THOMPSON, A. D., KILLGORE, G. E., ZUKOWSKI, W. E., SONGER, G., WARNY, M., JOHNSON, S., GERDING, D. N., MCDONALD, L. C. & LIMBAGO, B. M. 2008. Toxinotype V Clostridium difficile in humans and food animals. *Emerg Infect Dis,* 14**,** 1039-45.

JOHNSON, S. 2009. Recurrent Clostridium difficile infection: a review of risk factors, treatments, and outcomes. *J Infect,* 58**,** 403-10.

JOHNSON, S., KENT, S. A., O'LEARY, K. J., MERRIGAN, M. M., SAMBOL, S. P., PETERSON, L. R. & GERDING, D. N. 2001. Fatal pseudomembranous colitis associated with a variant clostridium difficile strain not detected by toxin A immunoassay. *Ann Intern Med,* 135**,** 434-8.

JOHNSON, S., SAMBOL, S. P., BRAZIER, J. S., DELMEE, M., AVESANI, V., MERRIGAN, M. M. & GERDING, D. N. 2003. International typing study of toxin A-negative, toxin B-positive Clostridium difficile variants. *J Clin Microbiol,* 41**,** 1543-7.

JOHNSON, S., SAMORE, M. H., FARROW, K. A., KILLGORE, G. E., TENOVER, F. C., LYRAS, D., ROOD, J. I., DEGIROLAMI, P., BALTCH, A. L., RAFFERTY, M. E., PEAR, S. M. & GERDING, D. N. 1999. Epidemics of diarrhea caused by a clindamycin-resistant strain of Clostridium difficile in four hospitals. *N Engl J Med,* 341**,** 1645-51.

JOUNG, K. B. & COTE, J. C. 2002. Evaluation of ribosomal RNA gene restriction patterns for the classification of Bacillus species and related genera. *J Appl Microbiol,* 92**,** 97-108.

KARCZEWSKI, J., ZORMAN, J., WANG, S., MIEZEIEWSKI, M., XIE, J., SORING, K., PETRESCU, I., ROGERS, I., THIRIOT, D. S., COOK, J. C., CHAMBERLIN, M., XOCONOSTLE, R. F., NAHAS, D. D., JOYCE, J. G., BODMER, J. L., HEINRICHS, J. H. & SECORE, S. 2014. Development of a recombinant toxin fragment vaccine for Clostridium difficile infection. *Vaccine,* 32**,** 2812-8.

KASSAM, Z., HUNDAL, R., MARSHALL, J. K. & LEE, C. H. 2012. Fecal transplant via retention enema for refractory or recurrent Clostridium difficile infection. *Arch Intern Med,* 172**,** 191-3.

KASSAM, Z., LEE, C. H., YUAN, Y. & HUNT, R. H. 2013. Fecal microbiota transplantation for Clostridium difficile infection: systematic review and meta-analysis. *Am J Gastroenterol,* 108**,** 500-8.

KATO, H., KATO, N., KATOW, S., MAEGAWA, T., NAKAMURA, S. & LYERLY, D. M. 1999. Deletions in the repeating sequences of the toxin A gene of toxin A-negative, toxin B-positive Clostridium difficile strains. *FEMS Microbiol Lett,* 175**,** 197-203.

KATO, H., KITA, H., KARASAWA, T., MAEGAWA, T., KOINO, Y., TAKAKUWA, H., SAIKAI, T., KOBAYASHI, K., YAMAGISHI, T. & NAKAMURA, S. 2001. Colonisation and transmission of Clostridium difficile in healthy individuals examined by PCR ribotyping and pulsed-field gel electrophoresis. *J Med Microbiol,* 50**,** 720-7.

KEEL, K., BRAZIER, J. S., POST, K. W., WEESE, S. & SONGER, J. G. 2007. Prevalence of PCR ribotypes among Clostridium difficile isolates from pigs, calves, and other species. *J Clin Microbiol,* 45**,** 1963-4.

KELLY, C. P. & LAMONT, J. T. 2008. Clostridium difficile--more difficult than ever. *N Engl J Med,* 359**,** 1932-40.

KELLY, C. R., IHUNNAH, C., FISCHER, M., KHORUTS, A., SURAWICZ, C., AFZALI, A., ARONIADIS, O., BARTO, A., BORODY, T., GIOVANELLI, A., GORDON, S., GLUCK, M., HOHMANN, E. L., KAO, D., KAO, J. Y., MCQUILLEN, D. P., MELLOW, M., RANK, K. M., RAO, K., RAY, A., SCHWARTZ, M. A., SINGH, N., STOLLMAN, N., SUSKIND, D. L., VINDIGNI, S. M., YOUNGSTER, I. & BRANDT, L. 2014. Fecal microbiota transplant for treatment of Clostridium difficile infection in immunocompromised patients. *Am J Gastroenterol,* 109**,** 1065-71.

KILLGORE, G., THOMPSON, A., JOHNSON, S., BRAZIER, J., KUIJPER, E., PEPIN, J., FROST, E. H., SAVELKOUL, P., NICHOLSON, B., VAN DEN BERG, R. J., KATO, H., SAMBOL, S. P., ZUKOWSKI, W., WOODS, C., LIMBAGO, B., GERDING, D. N. & MCDONALD, L. C. 2008. Comparison of seven techniques for typing international epidemic strains of Clostridium difficile: restriction endonuclease analysis, pulsed-field gel electrophoresis, PCR-ribotyping, multilocus sequence typing, multilocus variable-number tandem-repeat analysis, amplified fragment length polymorphism, and surface layer protein A gene sequence typing. *J Clin Microbiol,* 46**,** 431-7.

KIM, H., JEONG, S. H., ROH, K. H., HONG, S. G., KIM, J. W., SHIN, M. G., KIM, M. N., SHIN, H. B., UH, Y., LEE, H. & LEE, K. 2010a. Investigation of toxin gene diversity, molecular epidemiology, and antimicrobial resistance of Clostridium difficile isolated from 12 hospitals in South Korea. *Korean J Lab Med,* 30**,** 491-7.

KIM, K. H., FEKETY, R., BATTS, D. H., BROWN, D., CUDMORE, M., SILVA, J., JR. & WATERS, D. 1981. Isolation of Clostridium difficile from the environment and contacts of patients with antibiotic-associated colitis. *J Infect Dis,* 143**,** 42-50.

KIM, S. J., KIM, H., SEO, Y., YONG, D., JEONG, S. H., CHONG, Y. & LEE, K. 2010b. Molecular characterization of toxin A-negative, toxin B-positive

variant strains of Clostridium difficile isolated in Korea. *Diagn Microbiol Infect Dis,* 67**,** 198-201.

KLINE, K. A., FALKER, S., DAHLBERG, S., NORMARK, S. & HENRIQUES-NORMARK, B. 2009. Bacterial adhesins in host-microbe interactions. *Cell Host Microbe,* 5**,** 580-92.

KNETSCH, C. W., TERVEER, E. M., LAUBER, C., GORBALENYA, A. E., HARMANUS, C., KUIJPER, E. J., CORVER, J. & VAN LEEUWEN, H. C. 2012. Comparative analysis of an expanded Clostridium difficile reference strain collection reveals genetic diversity and evolution through six lineages. *Infect Genet Evol,* 12**,** 1577-85.

KOMATSU, M., KATO, H., AIHARA, M., SHIMAKAWA, K., IWASAKI, M., NAGASAKA, Y., FUKUDA, S., MATSUO, S., ARAKAWA, Y., WATANABE, M. & IWATANI, Y. 2003. High frequency of antibiotic-associated diarrhea due to toxin A-negative, toxin B-positive Clostridium difficile in a hospital in Japan and risk factors for infection. *Eur J Clin Microbiol Infect Dis,* 22**,** 525-9.

KRUTOVA, M., MATEJKOVA, J. & NYC, O. 2014. C. difficile ribotype 027 or 176? *Folia Microbiol (Praha),* 59**,** 523-6.

KRUTOVA, M., NYC, O., MATEJKOVA, J., KUIJPER, E. J., JALAVA, J. & MENTULA, S. 2017. The recognition and characterisation of Finnish Clostridium difficile isolates resembling PCR-ribotype 027. *J Microbiol Immunol Infect*.

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**,** 711-3.

KUIJPER, E. J., DE WEERDT, J., KATO, H., KATO, N., VAN DAM, A. P., VAN DER VORM, E. R., WEEL, J., VAN RHEENEN, C. & DANKERT, J. 2001. Nosocomial outbreak of Clostridium difficile-associated diarrhoea due to a clindamycin-resistant enterotoxin A-negative strain. *Eur J Clin Microbiol Infect Dis,* 20**,** 528-34.

KUIJPER, E. J., VAN DEN BERG, R. J., DEBAST, S., VISSER, C. E., VEENENDAAL, D., TROELSTRA, A., VAN DER KOOI, T., VAN DEN HOF, S. & NOTERMANS, D. W. 2006. Clostridium difficile ribotype 027, toxinotype III, the Netherlands. *Emerg Infect Dis,* 12**,** 827-30.

KULIKOVSKY, A., PANKRATZ, H. S. & SADOFF, H. L. 1975. Ultrastructural and chemical changes in spores of Bacillus cereus after action of disinfectants. *J Appl Bacteriol,* 38**,** 39-46.

KVISTHOLM JENSEN, A., NIELSEN, E. M., BJORKMAN, J. T., JENSEN, T., MULLER, L., PERSSON, S., BJERAGER, G., PERGE, A., KRAUSE, T. G., KIIL, K., SORENSEN, G., ANDERSEN, J. K., MOLBAK, K. & ETHELBERG, S. 2016. Whole-genome Sequencing Used to Investigate a Nationwide Outbreak of Listeriosis Caused by Ready-to-eat Delicatessen Meat, Denmark, 2014. *Clin Infect Dis,* 63**,** 64-70.

LABBE, A. C., POIRIER, L., MACCANNELL, D., LOUIE, T., SAVOIE, M., BELIVEAU, C., LAVERDIERE, M. & PEPIN, J. 2008. Clostridium difficile infections in a Canadian tertiary care hospital before and during a regional epidemic associated with the BI/NAP1/027 strain. *Antimicrob Agents Chemother,* 52**,** 3180-7.

LARSON, H. E., PRICE, A. B., HONOUR, P. & BORRIELLO, S. P. 1978. Clostridium difficile and the aetiology of pseudomembranous colitis. *Lancet,* 1**,** 1063-6.

LAWLEY, T. D., CLARE, S., WALKER, A. W., STARES, M. D., CONNOR, T. R., RAISEN, C., GOULDING, D., RAD, R., SCHREIBER, F., BRANDT, C., DEAKIN, L. J., PICKARD, D. J., DUNCAN, S. H., FLINT, H. J., CLARK, T. G., PARKHILL, J. & DOUGAN, G. 2012. Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice. *PLoS Pathog,* 8**,** e1002995.

LAWLEY, T. D., CROUCHER, N. J., YU, L., CLARE, S., SEBAIHIA, M., GOULDING, D., PICKARD, D. J., PARKHILL, J., CHOUDHARY, J. & DOUGAN, G. 2009. Proteomic and genomic characterization of highly infectious Clostridium difficile 630 spores. *J Bacteriol,* 191**,** 5377-86.

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) Prevot 1938. *Anaerobe,* 40**,** 95-9.

LAWSON, P. A. & RAINEY, F. A. 2015. Proposal to restrict the genus Clostridium (Prazmowski) to Clostridium butyricum and related species. *Int J Syst Evol Microbiol*.

LEEDS, J. A., SACHDEVA, M., MULLIN, S., BARNES, S. W. & RUZIN, A. 2014. In vitro selection, via serial passage, of Clostridium difficile mutants with reduced susceptibility to fidaxomicin or vancomycin. *J Antimicrob Chemother,* 69**,** 41-4.

LEMEE, L., BOURGEOIS, I., RUFFIN, E., COLLIGNON, A., LEMELAND, J. F. & PONS, J. L. 2005. Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of Clostridium difficile. *Microbiology,* 151**,** 3171-80.

LEMEE, L., DHALLUIN, A., PESTEL-CARON, M., LEMELAND, J. F. & PONS, J. L. 2004. Multilocus sequence typing analysis of human and animal Clostridium difficile isolates of various toxigenic types. *J Clin Microbiol,* 42**,** 2609-17.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LIM, S. K., STUART, R. L., MACKIN, K. E., CARTER, G. P., KOTSANAS, D., FRANCIS, M. J., EASTON, M., DIMOVSKI, K., ELLIOTT, B., RILEY, T. V., HOGG, G., PAUL, E., KORMAN, T. M., SEEMANN, T., STINEAR, T. P., LYRAS, D. & JENKIN, G. A. 2014. Emergence of a ribotype 244 strain of Clostridium difficile associated with severe disease and related to the epidemic ribotype 027 strain. *Clin Infect Dis,* 58**,** 1723-30.

LIMAYE, A. P., TURGEON, D. K., COOKSON, B. T. & FRITSCHE, T. R. 2000. Pseudomembranous colitis caused by a toxin A(-) B(+) strain of Clostridium difficile. *J Clin Microbiol,* 38**,** 1696-7.

LIN, J., HUANG, S. & ZHANG, Q. 2002. Outer membrane proteins: key players for bacterial adaptation in host niches. *Microbes Infect,* 4**,** 325-31.

LINDER, J. A., HUANG, E. S., STEINMAN, M. A., GONZALES, R. & STAFFORD, R. S. 2005. Fluoroquinolone prescribing in the United States: 1995 to 2002. *Am J Med,* 118**,** 259-68.

LOCKHART, S. R., ETIENNE, K. A., VALLABHANENI, S., FAROOQI, J., CHOWDHARY, A., GOVENDER, N. P., COLOMBO, A. L., CALVO, B., CUOMO, C. A., DESJARDINS, C. A., BERKOW, E. L., CASTANHEIRA, M., MAGOBO, R. E., JABEEN, K., ASGHAR, R. J., MEIS, J. F., JACKSON, B., CHILLER, T. & LITVINTSEVA, A. P. 2017. Simultaneous Emergence of Multidrug-Resistant Candida auris on 3 Continents Confirmed by Whole-Genome Sequencing and Epidemiological Analyses. *Clin Infect Dis,* 64**,** 134-140.

LOO, V. G., POIRIER, L., MILLER, M. A., OUGHTON, M., LIBMAN, M. D., MICHAUD, S., BOURGAULT, A. M., NGUYEN, T., FRENETTE, C., KELLY, M., VIBIEN, A., BRASSARD, P., FENN, S., DEWAR, K., HUDSON, T. J., HORN, R., RENE, P., MONCZAK, Y. & DASCAL, A. 2005. A predominantly clonal multi-institutional outbreak of Clostridium difficile-associated diarrhea with high morbidity and mortality. *N Engl J Med,* 353**,** 2442-9.

LOUIE, T. J., MILLER, M. A., MULLANE, K. M., WEISS, K., LENTNEK, A., GOLAN, Y., GORBACH, S., SEARS, P., SHUE, Y. K. & GROUP, O. P. T. C. S. 2011. Fidaxomicin versus vancomycin for Clostridium difficile infection. *N Engl J Med,* 364**,** 422-31.

LOWY, I., MOLRINE, D. C., LEAV, B. A., BLAIR, B. M., BAXTER, R., GERDING, D. N., NICHOL, G., THOMAS, W. D., JR., LENEY, M., SLOAN, S., HAY, C. A. & AMBROSINO, D. M. 2010. Treatment with monoclonal antibodies against Clostridium difficile toxins. *N Engl J Med,* 362**,** 197-205.

LUBELSKI, J., KONINGS, W. N. & DRIESSEN, A. J. 2007. Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiol Mol Biol Rev,* 71**,** 463-76.

LYERLY, D. M., BARROSO, L. A., WILKINS, T. D., DEPITRE, C. & CORTHIER, G. 1992. Characterization of a toxin A-negative, toxin B-positive strain of Clostridium difficile. *Infect Immun,* 60**,** 4633-9.

LYRAS, D., O'CONNOR, J. R., HOWARTH, P. M., SAMBOL, S. P., CARTER, G. P., PHUMOONNA, T., POON, R., ADAMS, V., VEDANTAM, G., JOHNSON, S., GERDING, D. N. & ROOD, J. I. 2009. Toxin B is essential for virulence of Clostridium difficile. *Nature,* 458**,** 1176-9.

MACCANNELL, D. R., LOUIE, T. J., GREGSON, D. B., LAVERDIERE, M., LABBE, A. C., LAING, F. & HENWICK, S. 2006. Molecular analysis of Clostridium difficile PCR ribotype 027 isolates from Eastern and Western Canada. *J Clin Microbiol,* 44**,** 2147-52.

MADOUI, M. A., ENGELEN, S., CRUAUD, C., BELSER, C., BERTRAND, L., ALBERTI, A., LEMAINQUE, A., WINCKER, P. & AURY, J. M. 2015. Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics,* 16**,** 327.

MALDARELLI, G. A., DE MASI, L., VON ROSENVINGE, E. C., CARTER, M. & DONNENBERG, M. S. 2014. Identification, immunogenicity, and cross-reactivity of type IV pilin and pilin-like proteins from Clostridium difficile. *Pathog Dis,* 71**,** 302-14.

MALDARELLI, G. A., PIEPENBRINK, K. H., SCOTT, A. J., FREIBERG, J. A., SONG, Y., ACHERMANN, Y., ERNST, R. K., SHIRTLIFF, M. E., SUNDBERG, E. J., DONNENBERG, M. S. & VON ROSENVINGE, E. C.

2016. Type IV pili promote early biofilm formation by Clostridium difficile. *Pathog Dis,* 74.

MANI, N. & DUPUY, B. 2001. Regulation of toxin synthesis in Clostridium difficile by an alternative RNA polymerase sigma factor. *Proc Natl Acad Sci U S A,* 98**,** 5844-9.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

MARSH, J. W., O'LEARY, M. M., SHUTT, K. A., SAMBOL, S. P., JOHNSON, S., GERDING, D. N. & HARRISON, L. H. 2010. Multilocus variable-number tandem-repeat analysis and multilocus sequence typing reveal genetic relationships among Clostridium difficile isolates genotyped by restriction endonuclease analysis. *J Clin Microbiol,* 48**,** 412-8.

MATAMOUROS, S., ENGLAND, P. & DUPUY, B. 2007. Clostridium difficile toxin expression is inhibited by the novel regulator TcdC. *Mol Microbiol,* 64**,** 1274-88.

MCBEE, R. H. 1960. Intestinal Flora of Some Antarctic Birds and Mammals. *J Bacteriol,* 79**,** 311-2.

MCCARTHY, A. 2010. Third generation DNA sequencing: pacific biosciences' single molecule real time technology. *Chem Biol,* 17**,** 675-6.

MCDONALD, L. C., KILLGORE, G. E., THOMPSON, A., OWENS, R. C., JR., KAZAKOVA, S. V., SAMBOL, S. P., JOHNSON, S. & GERDING, D. N. 2005. An epidemic, toxin gene-variant strain of Clostridium difficile. *N Engl J Med,* 353**,** 2433-41.

MCELLISTREM, M. C., CARMAN, R. J., GERDING, D. N., GENHEIMER, C. W. & ZHENG, L. 2005. A hospital outbreak of Clostridium difficile disease associated with isolates carrying binary toxin genes. *Clin Infect Dis,* 40**,** 265-72.

MCNULTY, C., LOGAN, M., DONALD, I. P., ENNIS, D., TAYLOR, D., BALDWIN, R. N., BANNERJEE, M. & CARTWRIGHT, K. A. 1997. Successful control of Clostridium difficile infection in an elderly care unit through use of a restrictive antibiotic policy. *J Antimicrob Chemother,* 40**,** 707-11.

MEATS, E., FEIL, E. J., STRINGER, S., CODY, A. J., GOLDSTEIN, R., KROLL, J. S., POPOVIC, T. & SPRATT, B. G. 2003. Characterization of Encapsulated and Noncapsulated Haemophilus influenzae and Determination of Phylogenetic Relationships by Multilocus Sequence Typing. *J Clin Microbiol,* 41**,** 1623-1636.

MERRIGAN, M., VENUGOPAL, A., MALLOZZI, M., ROXAS, B., VISWANATHAN, V. K., JOHNSON, S., GERDING, D. N. & VEDANTAM, G. 2010. Human hypervirulent Clostridium difficile strains exhibit increased sporulation as well as robust toxin production. *J Bacteriol,* 192**,** 4904-11.

METCALF, D., AVERY, B. P., JANECKO, N., MATIC, N., REID-SMITH, R. & WEESE, J. S. 2011. Clostridium difficile in seafood and fish. *Anaerobe,* 17**,** 85-6.

MITEVA, V., BOUDAKOV, I., IVANOVA-STOYANCHEVA, G., MARINOVA, B., MITEV, V. & MENGAUD, J. 2001. Differentiation of Lactobacillus delbrueckii subspecies by ribotyping and amplified ribosomal DNA restriction analysis (ARDRA). *J Appl Microbiol,* 90**,** 909-18.

MIYAJIMA, F., ROBERTS, P., SWALE, A., PRICE, V., JONES, M., HORAN, M., BEECHING, N., BRAZIER, J., PARRY, C., PENDLETON, N. & PIRMOHAMED, M. 2011. Characterisation and carriage ratio of Clostridium difficile strains isolated from a community-dwelling elderly population in the United Kingdom. *PLoS One,* 6**,** e22804.

MOLLE, V., FUJITA, M., JENSEN, S. T., EICHENBERGER, P., GONZALEZ-PASTOR, J. E., LIU, J. S. & LOSICK, R. 2003. The Spo0A regulon of Bacillus subtilis. *Mol Microbiol,* 50**,** 1683-701.

MOORE, P., KYNE, L., MARTIN, A. & SOLOMON, K. 2013. Germination efficiency of clinical Clostridium difficile spores and correlation with ribotype, disease severity and therapy failure. *J Med Microbiol,* 62**,** 1405-13.

MULLANY, P., PALLEN, M., WILKS, M., STEPHEN, J. R. & TABAQCHALI, S. 1996. A group II intron in a conjugative transposon from the gram-positive bacterium, Clostridium difficile. *Gene,* 174**,** 145-50.

MULVEY, M. A., LOPEZ-BOADO, Y. S., WILSON, C. L., ROTH, R., PARKS, W. C., HEUSER, J. & HULTGREN, S. J. 1998. Induction and evasion of host defenses by type 1-piliated uropathogenic Escherichia coli. *Science,* 282**,** 1494-7.

MUTO, C. A., BLANK, M. K., MARSH, J. W., VERGIS, E. N., O'LEARY, M. M., SHUTT, K. A., PASCULLE, A. W., POKRYWKA, M., GARCIA, J. G., POSEY, K., ROBERTS, T. L., POTOSKI, B. A., BLANK, G. E., SIMMONS, R. L., VELDKAMP, P., HARRISON, L. H. & PATERSON, D. L. 2007. Control of an outbreak of infection with the hypervirulent Clostridium difficile BI strain in a university hospital using a comprehensive "bundle" approach. *Clin Infect Dis,* 45**,** 1266-73.

MYLONAKIS, E., RYAN, E. T. & CALDERWOOD, S. B. 2001. Clostridium difficile--Associated diarrhea: A review. *Arch Intern Med,* 161**,** 525-33.

NACHAMKIN, I., YANG, X. H. & STERN, N. J. 1993. Role of Campylobacter jejuni flagella as colonization factors for three-day-old chicks: analysis with flagellar mutants. *Appl Environ Microbiol,* 59**,** 1269-73.

NGAMSKULRUNGROJ, P., SANMEE, S., PUTSATHIT, P., PIEWNGAM, P., ELLIOTT, B., RILEY, T. V. & KIRATISIN, P. 2015. Correction: Molecular Epidemiology of Clostridium difficile Infection in a Large Teaching Hospital in Thailand. *PLoS One,* 10**,** e0134771.

NGERNSOMBAT, C., SREESAI, S., HARNVORAVONGCHAI, P., CHANKHAMHAENGDECHA, S. & JANVILISRI, T. 2017. CD2068

potentially mediates multidrug efflux in Clostridium difficile. *Sci Rep,* 7**,** 9982.

NI EIDHIN, D. B., O'BRIEN, J. B., MCCABE, M. S., ATHIE-MORALES, V. & KELLEHER, D. P. 2008. Active immunization of hamsters against Clostridium difficile infection using surface-layer protein. *FEMS Immunol Med Microbiol,* 52**,** 207-18.

NIEMANN, H. H., SCHUBERT, W. D. & HEINZ, D. W. 2004. Adhesins and invasins of pathogenic bacteria: a structural view. *Microbes Infect,* 6**,** 101-12.

NIES, A. T., DAMME, K., SCHAEFFELER, E. & SCHWAB, M. 2012. Multidrug and toxin extrusion proteins as transporters of antimicrobial drugs. *Expert Opin Drug Metab Toxicol,* 8**,** 1565-77.

NYC, O., PITUCH, H., MATEJKOVA, J., OBUCH-WOSZCZATYNSKI, P. & KUIJPER, E. J. 2011. Clostridium difficile PCR ribotype 176 in the Czech Republic and Poland. *Lancet,* 377**,** 1407.

O'CONNOR, J. R., GALANG, M. A., SAMBOL, S. P., HECHT, D. W., VEDANTAM, G., GERDING, D. N. & JOHNSON, S. 2008. Rifampin and rifaximin resistance in clinical isolates of Clostridium difficile. *Antimicrob Agents Chemother,* 52**,** 2813-7.

ORSBURN, B., SUCRE, K., POPHAM, D. L. & MELVILLE, S. B. 2009. The SpmA/B and DacF proteins of Clostridium perfringens play important roles in spore heat resistance. *FEMS Microbiol Lett,* 291**,** 188-94.

OZAKI, E., KATO, H., KITA, H., KARASAWA, T., MAEGAWA, T., KOINO, Y., MATSUMOTO, K., TAKADA, T., NOMOTO, K., TANAKA, R. & NAKAMURA, S. 2004. Clostridium difficile colonization in healthy adults: transient colonization and correlation with enterococcal colonization. *J Med Microbiol,* 53**,** 167-72.

PALLEN, M. J. & WREN, B. W. 2007. Bacterial pathogenomics. *Nature,* 449**,** 835-42.

PALMER, A. C. & KISHONY, R. 2013. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat Rev Genet,* 14**,** 243-8.

PAREDES, C. J., ALSAKER, K. V. & PAPOUTSAKIS, E. T. 2005. A comparative genomic view of clostridial sporulation and physiology. *Nat Rev Microbiol,* 3**,** 969-78.

PARKHILL, J. & WREN, B. W. 2011. Bacterial epidemiology and biology - lessons from genome sequencing. *Genome Biol,* 12**,** 230.

PASQUALE, V., ROMANO, V. J., RUPNIK, M., DUMONTET, S., CIZNAR, I., ALIBERTI, F., MAURI, F., SAGGIOMO, V. & KROVACEK, K. 2011. Isolation and characterization of Clostridium difficile from shellfish and marine environments. *Folia Microbiol (Praha),* 56**,** 431-7.

PEAR, S. M., WILLIAMSON, T. H., BETTIN, K. M., GERDING, D. N. & GALGIANI, J. N. 1994. Decrease in nosocomial Clostridium difficile-associated diarrhea by restricting clindamycin use. *Ann Intern Med,* 120**,** 272-7.

PELAEZ, T., CERCENADO, E., ALCALA, L., MARIN, M., MARTIN-LOPEZ, A., MARTINEZ-ALARCON, J., CATALAN, P., SANCHEZ-SOMOLINOS, M. & BOUZA, E. 2008. Metronidazole resistance in Clostridium difficile is heterogeneous. *J Clin Microbiol,* 46**,** 3028-32.

PETROF, E. O., GLOOR, G. B., VANNER, S. J., WEESE, S. J., CARTER, D., DAIGNEAULT, M. C., BROWN, E. M., SCHROETER, K. & ALLEN-VERCOE, E. 2013. Stool substitute transplant therapy for the eradication of Clostridium difficile infection: 'RePOOPulating' the gut. *Microbiome,* 1**,** 3.

PETTIT, L. J., BROWNE, H. P., YU, L., SMITS, W. K., FAGAN, R. P., BARQUIST, L., MARTIN, M. J., GOULDING, D., DUNCAN, S. H., FLINT, H. J., DOUGAN, G., CHOUDHARY, J. S. & LAWLEY, T. D. 2014. Functional genomics reveals that Clostridium difficile Spo0A coordinates sporulation, virulence and metabolism. *BMC Genomics,* 15**,** 160.

PIEPENBRINK, K. H., MALDARELLI, G. A., MARTINEZ DE LA PENA, C. F., DINGLE, T. C., MULVEY, G. L., LEE, A., VON ROSENVINGE, E., ARMSTRONG, G. D., DONNENBERG, M. S. & SUNDBERG, E. J. 2015. Structural and evolutionary analyses show unique stabilization strategies in the type IV pili of Clostridium difficile. *Structure,* 23**,** 385-96.

PITUCH, H., BRAZIER, J. S., OBUCH-WOSZCZATYNSKI, P., WULTANSKA, D., MEISEL-MIKOLAJCZYK, F. & LUCZAK, M. 2006. Prevalence and association of PCR ribotypes of Clostridium difficile isolated from symptomatic patients from Warsaw with macrolide-lincosamide-streptogramin B (MLSB) type resistance. *J Med Microbiol,* 55**,** 207-13.

PITUCH, H., OBUCH-WOSZCZATYNSKI, P., ROUYAN, G. S., MARTIROSIAN, G. & MEISEL-MIKOLAJCZYK, F. 1998. [Detection of toxin producing strains of Clostridium difficile using rapid diagnostic methods]. *Med Dosw Mikrobiol,* 50**,** 55-61.

PITUCH, H., VAN DEN BRAAK, N., VAN LEEUWEN, W., VAN BELKUM, A., MARTIROSIAN, G., OBUCH-WOSZCZATYNSKI, P., LUCZAK, M. & MEISEL-MIKOLAJCZYK, F. 2001. Clonal dissemination of a toxin-A-negative/toxin-B-positive Clostridium difficile strain from patients with antibiotic-associated diarrhea in Poland. *Clin Microbiol Infect,* 7**,** 442-6.

PITUCH, H., VAN LEEUWEN, W., MAQUELIN, K., WULTANSKA, D., OBUCH-WOSZCZATYNSKI, P., NURZYNSKA, G., KATO, H., REIJANS, M., MEISEL-MIKOLAJCZYK, F., LUCZAK, M. & VAN BELKUM, A. 2007. Toxin profiles and resistances to macrolides and newer fluoroquinolones as epidemicity determinants of clinical isolates of Clostridium difficile from Warsaw, Poland. *J Clin Microbiol,* 45**,** 1607-10.

POLIVKOVA, S., KRUTOVA, M., PETRLOVA, K., BENES, J. & NYC, O. 2016. Clostridium difficile ribotype 176 - A predictor for high mortality and risk of nosocomial spread? *Anaerobe,* 40**,** 35-40.

POPHAM, D. L., GILMORE, M. E. & SETLOW, P. 1999. Roles of low-molecular-weight penicillin-binding proteins in Bacillus subtilis spore peptidoglycan synthesis and spore properties. *J Bacteriol,* 181**,** 126-32.

POPOFF, M. R., RUBIN, E. J., GILL, D. M. & BOQUET, P. 1988. Actin-specific ADP-ribosyltransferase produced by a Clostridium difficile strain. *Infect Immun,* 56**,** 2299-306.

POSTNOVA, T., GOMEZ-DUARTE, O. G. & RICHARDSON, K. 1996. Motility mutants of Vibrio cholerae O1 have reduced adherence in vitro to human small intestinal epithelial cells as demonstrated by ELISA. *Microbiology,* 142 (Pt 10)**,** 2767-76.

PRESTON, M. D., ASSEFA, S. A., OCHOLLA, H., SUTHERLAND, C. J., BORRMANN, S., NZILA, A., MICHON, P., HIEN, T. T., BOUSEMA, T.,

DRAKELEY, C. J., ZONGO, I., OUEDRAOGO, J. B., DJIMDE, A. A., DOUMBO, O. K., NOSTEN, F., FAIRHURST, R. M., CONWAY, D. J., ROPER, C. & CLARK, T. G. 2014. PlasmoView: a web-based resource to visualise global Plasmodium falciparum genomic variation. *J Infect Dis,* 209**,** 1808-15.

PUBLIC HEALTH ENGLAND 2016. Clostridium difficile Ribotyping Network (CDRN) for England and Northern Ireland, 2013 to 2015.

PUBLIC HEALTH ENGLAND 2017. C. difficile infections: quarterly counts by acute trust and CCG, and financial year counts and rates by acute trust and CCG, up to financial year 2016 to 2017. *In:* DATA, N. S. C. D. I. A. (ed.).

PUBLIC HEALTH ENGLAND. 2016. Clostridium difficile Ribotyping Network (CDRN) for England and Northern Ireland, 2013 to 2015.

PURCELL, E. B., MCKEE, R. W., BORDELEAU, E., BURRUS, V. & TAMAYO, R. 2015. Regulation of Type IV Pili Contributes to Surface Behaviors of Historical and Epidemic Strains of Clostridium difficile. *J Bacteriol,* 198**,** 565-77.

QUICK, J., ASHTON, P., CALUS, S., CHATT, C., GOSSAIN, S., HAWKER, J., NAIR, S., NEAL, K., NYE, K., PETERS, T., DE PINNA, E., ROBINSON, E., STRUTHERS, K., WEBBER, M., CATTO, A., DALLMAN, T. J., HAWKEY, P. & LOMAN, N. J. 2015. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. *Genome Biol,* 16**,** 114.

RAMOS, J. L., MARTINEZ-BUENO, M., MOLINA-HENARES, A. J., TERAN, W., WATANABE, K., ZHANG, X., GALLEGOS, M. T., BRENNAN, R. & TOBES, R. 2005. The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev,* 69**,** 326-56.

RASKO, D. A., WEBSTER, D. R., SAHL, J. W., BASHIR, A., BOISEN, N., SCHEUTZ, F., PAXINOS, E. E., SEBRA, R., CHIN, C. S., ILIOPOULOS, D., KLAMMER, A., PELUSO, P., LEE, L., KISLYUK, A. O., BULLARD, J., KASARSKIS, A., WANG, S., EID, J., RANK, D., REDMAN, J. C., STEYERT, S. R., FRIMODT-MOLLER, J., STRUVE, C., PETERSEN, A. M., KROGFELT, K. A., NATARO, J. P., SCHADT, E. E. & WALDOR, M. K. 2011. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med,* 365**,** 709-17.

RAYNER, R. E., SAVILL, J., HAFNER, L. M. & HUYGENS, F. 2015. Modified MLVA for Genotyping Queensland Invasive Streptococcus pneumoniae. *PLoS One,* 10**,** e0121870.

RAZAQ, N., SAMBOL, S., NAGARO, K., ZUKOWSKI, W., CHEKNIS, A., JOHNSON, S. & GERDING, D. N. 2007. Infection of hamsters with historical and epidemic BI types of Clostridium difficile. *J Infect Dis,* 196**,** 1813-9.

READ, T. D. & MASSEY, R. C. 2014. Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. *Genome Med,* 6**,** 109.

REDELINGS, M. D., SORVILLO, F. & MASCOLA, L. 2007. Increase in Clostridium difficile-related mortality rates, United States, 1999-2004. *Emerg Infect Dis,* 13**,** 1417-9.

REEVES, A. E., KOENIGSKNECHT, M. J., BERGIN, I. L. & YOUNG, V. B. 2012. Suppression of Clostridium difficile in the gastrointestinal tracts of

germfree mice inoculated with a murine isolate from the family Lachnospiraceae. *Infect Immun,* 80**,** 3786-94.

REEVES, A. E., THERIOT, C. M., BERGIN, I. L., HUFFNAGLE, G. B., SCHLOSS, P. D. & YOUNG, V. B. 2011. The interplay between microbiome dynamics and pathogen dynamics in a murine model of Clostridium difficile Infection. *Gut Microbes,* 2**,** 145-58.

REUTER, S., TOROK, M. E., HOLDEN, M. T., REYNOLDS, R., RAVEN, K. E., BLANE, B., DONKER, T., BENTLEY, S. D., AANENSEN, D. M., GRUNDMANN, H., FEIL, E. J., SPRATT, B. G., PARKHILL, J. & PEACOCK, S. J. 2016. Building a genomic framework for prospective MRSA surveillance in the United Kingdom and the Republic of Ireland. *Genome Res,* 26**,** 263-70.

RICHARDSON, K. 1991. Roles of motility and flagellar structure in pathogenicity of Vibrio cholerae: analysis of motility mutants in three animal models. *Infect Immun,* 59**,** 2727-36.

RIGGS, M. M., SETHI, A. K., ZABARSKY, T. F., ECKSTEIN, E. C., JUMP, R. L. & DONSKEY, C. J. 2007. Asymptomatic carriers are a potential source for transmission of epidemic and nonepidemic Clostridium difficile strains among long-term care facility residents. *Clin Infect Dis,* 45**,** 992-8.

ROBERTS, A. P., CHANDLER, M., COURVALIN, P., GUEDON, G., MULLANY, P., PEMBROKE, T., ROOD, J. I., SMITH, C. J., SUMMERS, A. O., TSUDA, M. & BERG, D. E. 2008. Revised nomenclature for transposable genetic elements. *Plasmid,* 60**,** 167-73.

ROBERTS, A. P. & MULLANY, P. 2009. A modular master on the move: the Tn916 family of mobile genetic elements. *Trends Microbiol,* 17**,** 251-8.

ROBERTS, A. P., MULLANY, P. & WILSON, M. 2001. Gene transfer in bacterial biofilms. *Methods Enzymol,* 336**,** 60-5.

RODRIGUEZ-PALACIOS, A., REID-SMITH, R. J., STAEMPFLI, H. R., DAIGNAULT, D., JANECKO, N., AVERY, B. P., MARTIN, H., THOMSPON, A. D., MCDONALD, L. C., LIMBAGO, B. & WEESE, J. S. 2009. Possible seasonality of Clostridium difficile in retail meat, Canada. *Emerg Infect Dis,* 15**,** 802-5.

ROETZER, A., DIEL, R., KOHL, T. A., RUCKERT, C., NUBEL, U., BLOM, J., WIRTH, T., JAENICKE, S., SCHUBACK, S., RUSCH-GERDES, S., SUPPLY, P., KALINOWSKI, J. & NIEMANN, S. 2013. Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. *PLoS Med,* 10**,** e1001387.

ROMANO, V., PASQUALE, V., KROVACEK, K., MAURI, F., DEMARTA, A. & DUMONTET, S. 2012. Toxigenic Clostridium difficile PCR ribotypes from wastewater treatment plants in southern Switzerland. *Appl Environ Microbiol,* 78**,** 6643-6.

RONQUIST, F. & HUELSENBECK, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-4.

RUPNIK, M., AVESANI, V., JANC, M., VON EICHEL-STREIBER, C. & DELMEE, M. 1998. A novel toxinotyping scheme and correlation of toxinotypes with serogroups of Clostridium difficile isolates. *J Clin Microbiol,* 36**,** 2240-7.

RUPNIK, M., BRAZIER, J. S., DUERDEN, B. I., GRABNAR, M. & STUBBS, S. L. 2001. Comparison of toxinotyping and PCR ribotyping of Clostridium difficile strains and description of novel toxinotypes. *Microbiology,* 147**,** 439-47.

RUPNIK, M. & JANEZIC, S. 2016. An Update on Clostridium difficile Toxinotyping. *J Clin Microbiol,* 54**,** 13-8.

RUPNIK, M., WIDMER, A., ZIMMERMANN, O., ECKERT, C. & BARBUT, F. 2008. Clostridium difficile toxinotype V, ribotype 078, in animals and humans. *J Clin Microbiol,* 46**,** 2146.

RUPNIK, M., WILCOX, M. H. & GERDING, D. N. 2009. Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol,* 7**,** 526-36.

RUSK, N. 2010. Torrents of sequence. *Nat Methods,* 8**,** 44.

RUSK, N. 2011. One genome, two haplotypes. *Nat Methods,* 8**,** 107.

RYAN, A., LYNCH, M., SMITH, S. M., AMU, S., NEL, H. J., MCCOY, C. E., DOWLING, J. K., DRAPER, E., O'REILLY, V., MCCARTHY, C., O'BRIEN, J., NI EIDHIN, D., O'CONNELL, M. J., KEOGH, B., MORTON, C. O., ROGERS, T. R., FALLON, P. G., O'NEILL, L. A., KELLEHER, D. & LOSCHER, C. E. 2011. A role for TLR4 in Clostridium difficile infection and the recognition of surface layer proteins. *PLoS Pathog,* 7**,** e1002076.

SABAT, A. J., BUDIMIR, A., NASHEV, D., SA-LEAO, R., VAN DIJL, J., LAURENT, F., GRUNDMANN, H., FRIEDRICH, A. W. & MARKERS, E. S. G. O. E. 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill,* 18**,** 20380.

SAITOU, N. & NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol,* 4**,** 406-25.

SALMENLINNA, S. & VUOPIO-VARKILA, J. 2001. Recognition of two groups of methicillin-resistant Staphylococcus aureus strains based on epidemiology, antimicrobial susceptibility, hypervariable-region type, and ribotype in Finland. *J Clin Microbiol,* 39**,** 2243-7.

SAMBOL, S. P., MERRIGAN, M. M., LYERLY, D., GERDING, D. N. & JOHNSON, S. 2000. Toxin gene analysis of a variant strain of Clostridium difficile that causes human clinical disease. *Infect Immun,* 68**,** 5480-7.

SAMRA, Z., TALMOR, S. & BAHAR, J. 2002. High prevalence of toxin A-negative toxin B-positive Clostridium difficile in hospitalized patients with gastrointestinal disease. *Diagn Microbiol Infect Dis,* 43**,** 189-92.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SATIAPUTRA, J., SHEARWIN, K. E., BOOKER, G. W. & POLYAK, S. W. 2016. Mechanisms of biotin-regulated gene expression in microbes. *Synth Syst Biotechnol,* 1**,** 17-24.

SATTA, G., WITNEY, A. A., SHORTEN, R. J., KARLIKOWSKA, M., LIPMAN, M. & MCHUGH, T. D. 2016. Genetic variation in Mycobacterium tuberculosis isolates from a London outbreak associated with isoniazid resistance. *BMC Med,* 14**,** 117.

SCHWAN, C., STECHER, B., TZIVELEKIDIS, T., VAN HAM, M., ROHDE, M., HARDT, W. D., WEHLAND, J. & AKTORIES, K. 2009. Clostridium difficile toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. *PLoS Pathog,* 5**,** e1000626.

SEBAIHIA, M., WREN, B. W., MULLANY, P., FAIRWEATHER, N. F., MINTON, N., STABLER, R., THOMSON, N. R., ROBERTS, A. P., CERDENO-TARRAGA, A. M., WANG, H., HOLDEN, M. T., WRIGHT, A., CHURCHER, C., QUAIL, M. A., BAKER, S., BASON, N., BROOKS, K., CHILLINGWORTH, T., CRONIN, A., DAVIS, P., DOWD, L., FRASER, A., FELTWELL, T., HANCE, Z., HOLROYD, S., JAGELS, K., MOULE, S., MUNGALL, K., PRICE, C., RABBINOWITSCH, E., SHARP, S., SIMMONDS, M., STEVENS, K., UNWIN, L., WHITHEAD, S., DUPUY, B., DOUGAN, G., BARRELL, B. & PARKHILL, J. 2006. The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nat Genet,* 38**,** 779-86.

SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 30**,** 2068-9.

SETLOW, P. 2007. I will survive: DNA protection in bacterial spores. *Trends Microbiol,* 15**,** 172-80.

SHAHINAS, D., SILVERMAN, M., SITTLER, T., CHIU, C., KIM, P., ALLEN-VERCOE, E., WEESE, S., WONG, A., LOW, D. E. & PILLAI, D. R. 2012. Toward an understanding of changes in diversity associated with fecal microbiome transplantation based on 16S rRNA gene deep sequencing. *MBio,* 3.

SHENDURE, J., PORRECA, G. J., REPPAS, N. B., LIN, X., MCCUTCHEON, J. P., ROSENBAUM, A. M., WANG, M. D., ZHANG, K., MITRA, R. D. & CHURCH, G. M. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science,* 309**,** 1728-32.

SHIN, B. M., KUAK, E. Y., YOO, S. J., SHIN, W. C. & YOO, H. M. 2008. Emerging toxin A-B+ variant strain of Clostridium difficile responsible for pseudomembranous colitis at a tertiary care hospital in Korea. *Diagn Microbiol Infect Dis,* 60**,** 333-7.

SILBY, M. W., FERGUSON, G. C., BILLINGTON, C. & HEINEMANN, J. A. 2007. Localization of the plasmid-encoded proteins TraI and MobA in eukaryotic cells. *Plasmid,* 57**,** 118-30.

SIMANGO, C. & MWAKURUDZA, S. 2008. Clostridium difficile in broiler chickens sold at market places in Zimbabwe and their antimicrobial susceptibility. *Int J Food Microbiol,* 124**,** 268-70.

SLOAN, L. M., DURESKO, B. J., GUSTAFSON, D. R. & ROSENBLATT, J. E. 2008. Comparison of real-time PCR for detection of the tcdC gene with four toxin immunoassays and culture in diagnosis of Clostridium difficile infection. *J Clin Microbiol,* 46**,** 1996-2001.

SMITH, L. D. & KING, E. O. 1962. Occurrence of Clostridium difficile in infections of man. *J Bacteriol,* 84**,** 65-7.

SNYDER, M. L. 1937. Further Studies on Bacillus Difficilis (Hall and O'Toole) *Journal of Infectious Diseases,* 60**,** 223-231.

SONGER, J. G. 2004. The emergence of Clostridium difficile as a pathogen of food animals. *Anim Health Res Rev,* 5**,** 321-6.

SONGER, J. G., TRINH, H. T., KILLGORE, G. E., THOMPSON, A. D., MCDONALD, L. C. & LIMBAGO, B. M. 2009. Clostridium difficile in retail meat products, USA, 2007. *Emerg Infect Dis,* 15**,** 819-21.

SORG, J. A. & SONENSHEIN, A. L. 2008. Bile salts and glycine as cogerminants for Clostridium difficile spores. *J Bacteriol,* 190**,** 2505-12.

SOTO, S. M. 2009. Human migration and infectious diseases. *Clin Microbiol Infect,* 15 Suppl 1**,** 26-8.

SPIGAGLIA, P., BARBANTI, F., DIONISI, A. M. & MASTRANTONIO, P. 2010. Clostridium difficile isolates resistant to fluoroquinolones in Italy: emergence of PCR ribotype 018. *J Clin Microbiol,* 48**,** 2892-6.

SPIGAGLIA, P., BARBANTI, F., MASTRANTONIO, P., BRAZIER, J. S., BARBUT, F., DELMEE, M., KUIJPER, E., POXTON, I. R. & EUROPEAN STUDY GROUP ON CLOSTRIDIUM, D. 2008. Fluoroquinolone resistance in Clostridium difficile isolates from a prospective study of C. difficile infections in Europe. *J Med Microbiol,* 57**,** 784-9.

SPIGAGLIA, P., BARKETI-KLAI, A., COLLIGNON, A., MASTRANTONIO, P., BARBANTI, F., RUPNIK, M., JANEZIC, S. & KANSAU, I. 2013. Surface-layer (S-layer) of human and animal Clostridium difficile strains and their behaviour in adherence to epithelial cells and intestinal colonization. *J Med Microbiol,* 62**,** 1386-93.

SPIRIG, T., WEINER, E. M. & CLUBB, R. T. 2011. Sortase enzymes in Gram-positive bacteria. *Mol Microbiol,* 82**,** 1044-59.

SPRATT, B. G. 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Curr Opin Microbiol,* 2**,** 312-6.

STABLER, R. A., DAWSON, L. F., VALIENTE, E., CAIRNS, M. D., MARTIN, M. J., DONAHUE, E. H., RILEY, T. V., SONGER, J. G., KUIJPER, E. J., DINGLE, K. E. & WREN, B. W. 2012. Macro and Micro Diversity of Clostridium difficile Isolates from Diverse Sources and Geographical Locations. *PLoS One,* 7**,** e31559.

STABLER, R. A., GERDING, D. N., SONGER, J. G., DRUDY, D., BRAZIER, J. S., TRINH, H. T., WITNEY, A. A., HINDS, J. & WREN, B. W. 2006. Comparative phylogenomics of Clostridium difficile reveals clade specificity and microevolution of hypervirulent strains. *J Bacteriol,* 188**,** 7297-305.

STABLER, R. A., HE, M., DAWSON, L., MARTIN, M., VALIENTE, E., CORTON, C., LAWLEY, T. D., SEBAIHIA, M., QUAIL, M. A., ROSE, G., GERDING, D. N., GIBERT, M., POPOFF, M. R., PARKHILL, J., DOUGAN, G. & WREN, B. W. 2009. Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol,* 10**,** R102.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-3.

STUBBS, S., RUPNIK, M., GIBERT, M., BRAZIER, J., DUERDEN, B. & POPOFF, M. 2000. Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of Clostridium difficile. *FEMS Microbiol Lett,* 186**,** 307-12.

STUBBS, S. L., BRAZIER, J. S., O'NEILL, G. L. & DUERDEN, B. I. 1999. PCR targeted to the 16S-23S rRNA gene intergenic spacer region of Clostridium difficile and construction of a library consisting of 116 different PCR ribotypes. *J Clin Microbiol,* 37**,** 461-3.

SUNDRIYAL, A., ROBERTS, A. K., LING, R., MCGLASHAN, J., SHONE, C. C. & ACHARYA, K. R. 2010. Expression, purification and cell cytotoxicity of actin-modifying binary toxin from Clostridium difficile. *Protein Expr Purif,* 74**,** 42-8.

SWANSON, J. 1973. Studies on gonococcus infection. IV. Pili: their role in attachment of gonococci to tissue culture cells. *J Exp Med,* 137**,** 571-89.

TAGINI, F., AUBERT, B., TROILLET, N., PILLONEL, T., PRAZ, G., CRISINEL, P. A., PROD'HOM, G., ASNER, S. & GREUB, G. 2017. Importance of whole genome sequencing for the assessment of outbreaks in diagnostic laboratories: analysis of a case series of invasive Streptococcus pyogenes infections. *Eur J Clin Microbiol Infect Dis,* 36**,** 1173-1180.

TASTEYRE, A., BARC, M. C., COLLIGNON, A., BOUREAU, H. & KARJALAINEN, T. 2001. Role of FliC and FliD flagellar proteins of Clostridium difficile in adherence and gut colonization. *Infect Immun,* 69**,** 7937-40.

TENOVER, F. C., AKERLUND, T., GERDING, D. N., GOERING, R. V., BOSTROM, T., JONSSON, A. M., WONG, E., WORTMAN, A. T. & PERSING, D. H. 2011. Comparison of strain typing results for Clostridium difficile isolates from North America. *J Clin Microbiol,* 49**,** 1831-7.

THE EUROPEAN COMMITTEE ON ANTIMICROBIAL SUSCEPTIBILITY TESTING. 2017. Breakpoint tables for interpretation of MICs and zone diameters.

TIPPER, D. J. & STROMINGER, J. L. 1965. Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc Natl Acad Sci U S A,* 54**,** 1133-41.

TORRES, J. F. 1991. Purification and characterisation of toxin B from a strain of Clostridium difficile that does not produce toxin A. *J Med Microbiol,* 35**,** 40-4.

TSAI, E. A., SHAKBATYAN, R., EVANS, J., ROSSETTI, P., GRAHAM, C., SHARMA, H., LIN, C. F. & LEBO, M. S. 2016. Bioinformatics Workflow for Clinical Whole Genome Sequencing at Partners HealthCare Personalized Medicine. *J Pers Med,* 6.

VALIENTE, E., BOUCHE, L., HITCHEN, P., FAULDS-PAIN, A., SONGANE, M., DAWSON, L. F., DONAHUE, E., STABLER, R. A., PANICO, M., MORRIS, H. R., BAJAJ-ELLIOTT, M., LOGAN, S. M., DELL, A. & WREN, B. W. 2016. Role of glycosyltransferases modifying type B flagellin of emerging hypervirulent Clostridium difficile lineages and their impact on motility and biofilm formation. *J Biol Chem*.

VALIENTE, E., CAIRNS, M. D. & WREN, B. W. 2014. The Clostridium difficile PCR ribotype 027 lineage: a pathogen on the move. *Clin Microbiol Infect,* 20**,** 396-404.

VALIENTE, E., DAWSON, L. F., CAIRNS, M. D., STABLER, R. A. & WREN, B. W. 2012. Emergence of new PCR ribotypes from the hypervirulent Clostridium difficile 027 lineage. *J Med Microbiol,* 61**,** 49-56.

VAN NOOD, E., VRIEZE, A., NIEUWDORP, M., FUENTES, S., ZOETENDAL, E. G., DE VOS, W. M., VISSER, C. E., KUIJPER, E. J., BARTELSMAN, J. F., TIJSSEN, J. G., SPEELMAN, P., DIJKGRAAF, M. G. & KELLER, J. J. 2013. Duodenal infusion of donor feces for recurrent Clostridium difficile. *N Engl J Med,* 368**,** 407-15.

VISCIDI, R., WILLEY, S. & BARTLETT, J. G. 1981. Isolation rates and toxigenic potential of Clostridium difficile isolates from various patient populations. *Gastroenterology,* 81**,** 5-9.

VOHRA, P. & POXTON, I. R. 2011. Comparison of toxin and spore production in clinically relevant strains of Clostridium difficile. *Microbiology,* 157**,** 1343-53.

VON EICHEL-STREIBER, C., LAUFENBERG-FELDMANN, R., SARTINGEN, S., SCHULZE, J. & SAUERBORN, M. 1992. Comparative sequence analysis of the Clostridium difficile toxins A and B. *Mol Gen Genet,* 233**,** 260-8.

VON EICHEL-STREIBER, C., ZEC-PIRNAT, I., GRABNAR, M. & RUPNIK, M. 1999. A nonsense mutation abrogates production of a functional enterotoxin A in Clostridium difficile toxinotype VIII strains of serogroups F and X. *FEMS Microbiol Lett,* 178**,** 163-8.

VOTH, D. E. & BALLARD, J. D. 2005. Clostridium difficile toxins: mechanism of action and role in disease. *Clin Microbiol Rev,* 18**,** 247-63.

WALKER, T. M., KOHL, T. A., OMAR, S. V., HEDGE, J., DEL OJO ELIAS, C., BRADLEY, P., IQBAL, Z., FEUERRIEGEL, S., NIEHAUS, K. E., WILSON, D. J., CLIFTON, D. A., KAPATAI, G., IP, C. L., BOWDEN, R., DROBNIEWSKI, F. A., ALLIX-BEGUEC, C., GAUDIN, C., PARKHILL, J., DIEL, R., SUPPLY, P., CROOK, D. W., SMITH, E. G., WALKER, A. S., ISMAIL, N., NIEMANN, S., PETO, T. E. & MODERNIZING MEDICAL MICROBIOLOGY INFORMATICS, G. 2015. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis,* 15**,** 1193-202.

WARNY, M., PEPIN, J., FANG, A., KILLGORE, G., THOMPSON, A., BRAZIER, J., FROST, E. & 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**,** 1079-84.

WEISMAN, A., CHOU, B., O'BRIEN, J. & SHEA, K. J. 2015. Polymer antidotes for toxin sequestration. *Adv Drug Deliv Rev,* 90**,** 81-100.

WIEDENBECK, J. & COHAN, F. M. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev,* 35**,** 957-76.

WREN, B. W., MULLANY, P., CLAYTON, C. & TABAQCHALI, S. 1988. Molecular cloning and genetic analysis of a chloramphenicol acetyltransferase determinant from Clostridium difficile. *Antimicrob Agents Chemother,* 32**,** 1213-7.

WREN, M. W., KINSON, R., SIVAPALAN, M., SHEMKO, M. & SHETTY, N. R. 2009. Detection of Clostridium difficile infection: a suggested laboratory diagnostic algorithm. *Br J Biomed Sci,* 66**,** 175-9.

WU, L. J. & ERRINGTON, J. 2003. RacA and the Soj-Spo0J system combine to effect polar chromosome segregation in sporulating Bacillus subtilis. *Mol Microbiol,* 49**,** 1463-75.

WUST, J. & HARDEGGER, U. 1983. Transferable resistance to clindamycin, erythromycin, and tetracycline in Clostridium difficile. *Antimicrob Agents Chemother,* 23**,** 784-6.

WUST, J., SULLIVAN, N. M., HARDEGGER, U. & WILKINS, T. D. 1982. Investigation of an outbreak of antibiotic-associated colitis by various typing methods. *J Clin Microbiol,* 16**,** 1096-101.

XU, Y., ZHU, Y., WANG, Y., CHANG, Y. F., ZHANG, Y., JIANG, X., ZHUANG, X., ZHU, Y., ZHANG, J., ZENG, L., YANG, M., LI, S., WANG, S., YE, Q., XIN, X., ZHAO, G., ZHENG, H., GUO, X. & WANG, J. 2016. Whole

genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic Leptospira. *Sci Rep,* 6**,** 20020.

YAMAKAWA, K., KARASAWA, T., IKOMA, S. & NAKAMURA, S. 1996. Enhancement of Clostridium difficile toxin production in biotin-limited conditions. *J Med Microbiol,* 44**,** 111-4.

YANAGAWA, R., OTSUKI, K. & TOKUI, T. 1968. Electron microscopy of fine structure of Corynebacterium renale with special reference to pili. *Jpn J Vet Res,* 16**,** 31-7.

YURLOVA, N. A., UIJTHOF, J. M. & DE HOOG, G. S. 1996. Distinction of species in Aureobasidium and related genera by PCR-ribotyping. *Antonie Van Leeuwenhoek,* 69**,** 323-9.

YUTIN, N. & GALPERIN, M. Y. 2013. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ Microbiol,* 15**,** 2631-41.

ZANKARI, E., HASMAN, H., KAAS, R. S., SEYFARTH, A. M., AGERSO, Y., LUND, O., LARSEN, M. V. & AARESTRUP, F. M. 2013. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J Antimicrob Chemother,* 68**,** 771-7.

ZERBINO, D. R. 2010. Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics,* Chapter 11**,** Unit 11 5.

ZHANG, F., LUO, W., SHI, Y., FAN, Z. & JI, G. 2012. Should we standardize the 1,700-year-old fecal microbiota transplantation? *Am J Gastroenterol,* 107**,** 1755; author reply p 1755-6.

ZIDARIC, V., BEIGOT, S., LAPAJNE, S. & RUPNIK, M. 2010. The occurrence and high diversity of Clostridium difficile genotypes in rivers. *Anaerobe,* 16**,** 371-5.