Evaluation of Next-Generation Sequencing to Investigate the Epidemiology of *Clostridioides difficile* Infection

A thesis submitted to the University of Dublin, Trinity College,

for the Degree of Doctor of Philosophy

2022

Dr Geraldine Moloney

MB, BCh, BAO (Hons), MRCPI

Department of Clinical Microbiology,

School of Medicine, Trinity College Dublin

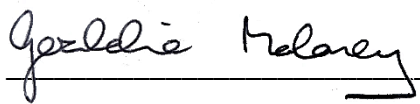Supervisor: Prof. Thomas Rogers

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I have acknowledged my contributions and those of co-authors to the academic publications included in this thesis.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signed: Dr Geraldine Moloney,

29th November 2020

Summary

Clostridioides difficile is an anaerobic bacteria, first identified as a commensal or harmless bacteria within the faecal samples of infants. It was recognised as the infectious aetiology of a severe gastrointestinal condition, pseudomembranous colitis, which occurred in epidemic proportions in the 1970s after the antibiotic clindamycin was introduced into clinical practice. A global epidemic of *C. difficile* infection occurred in the early 21st century, associated with the widespread use of fluoroquinolone antibiotics, especially ciprofloxacin. Since 2013, the Centers for Disease Control, USA, has classified *C. difficile* amongst the urgent threats to human health. They recognised 223,900 cases with 12,800 deaths in 2017. In Ireland, there was 2,053 cases in 2018, mostly diagnosed in hospital patients.

There have been major developments of techniques to identify sub-types of *C. difficile* since the global epidemic was recognised. The first method was PCR ribotype analysis, which separates subtypes, known as ribotypes, by the number of copies of the 16S-23S intergenic spacer region. As the number of copies varies, ribotypes have distinctive banding patterns, visible by pulsed-field gel electrophoresis. This typing mechanism indicated the overall diversity of *C. difficile* isolates, and associated the ribotype 027 with the global epidemic. However, PCR ribotype analysis does not provide sufficient discrimination to enable recognition of clusters of infection at local or regional level.

Sequencing by synthesis, otherwise known as next generation sequencing (NGS), has facilitated genome-based research on an unprecedented scale. When first applied to *C. difficile* isolates from ~1000 episodes of CDI, Eyre *et al.* found that only 33% of isolates had sufficient similarities, defined as a maximum of 2 single-nucleotide polymorphisms (SNPs) between genomes, for plausible CDI transmission within the hospital. These findings caused a paradigm shift in the understanding of *C. difficile* epidemiology, as CDI was previously considered as almost universally acquired in nosocomial environments. Applications of PCR ribotype and NGS analyses have also enabled better recognition of

CDI in people without recent healthcare exposure, known as community-acquired CDI, and that *C. difficile* can be a commensal or pathogenic bacteria for animals.

Use of antibiotics is a central risk factor for CDI. Antibiotics inhibit normal, healthy, intestinal bacteria in addition to bacteria at the site of suspected or confirmed primary infection. When antibiotics alter the microbiome, the normal bacterial community, *C. difficile* can proliferate as a secondary infection. Despite this recognised association, it is not yet clear which commensal bacterial species are most important to resist this proliferation. Sequencing by synthesis also allows for amplification of nucleic acid fragments present in all bacteria, and thereby assess the bacterial diversity and the relative abundance of key bacterial species of a microbiome.

The primary aim of this project was the application of NGS analysis to *C. difficile* isolates from a prospective cohort of CDI diagnosed within St. James's Hospital (SJH) in a 3-year period, to assess the nosocomial transmission rate. These isolates were compared to ones associated with community-acquired CDI, and with veterinary CDI. Clinical risk factors for CDI were considered with the genomic findings. Faecal microbiome samples from first episodes of CDI were compared to those of recurrent CDI, for potential microbiome indicators of increased risk for recurrence.

There were 335 *C. difficile* isolates analysed pertaining to clinical CDI. Most patients were inpatients of SJH at the time of diagnosis, but 17 isolates were provided from a national prevalence study of community-acquired CDI. The transmission rate identified was 21%, which is comparable to other studies with findings of 13-24%. Compared to those studies, the distribution of PCR ribotypes was strikingly different for the SJH isolates, with a markedly lower proportion of ribotype 027 isolates, and greater proportion of ribotype 078 isolates. Transmission was significantly associated with increased age of patients, but not their length of stay. More transmissions were associated with healthcare staff than with hospital wards. There were significant differences between the antibiotics preceding CDI diagnosis and those of all hospital inpatients during this time, which may account for the distribution of *C. difficile* ribotypes.

The *C. difficile* isolates of community-acquired CDI included the most prevalent ribotypes of episodes diagnosed in SJH. Furthermore, there were 20 pairs of isolates separated by 0–2 SNPs which could not be explained by available epidemiology.

NGS also enabled recognition of an outbreak of a unique subtype, *C. difficile* ST-295, which was of nosocomial origin and transmitted to another patient. This patient developed symptoms weeks later, at his residence in a local hostel. Four other hostel residents acquired this subtype of CDI and were admitted to SJH for management. This inadvertently led to two further transmissions within a hospital ward. Public Health staff addressed the infection prevention and control issues of this outbreak with the hostel. This outbreak highlights the shared epidemiology of hospital and community CDI.

As indicated, ribotype 078 was the most common *C. difficile* ribotype identified within the clinical isolates. During this study, a collaboration was developed with the Central Veterinary Research Laboratory, and PCR ribotype analysis was performed on isolates from porcine CDI. Ribotype 078 was most common, and NGS analysis was subsequently performed on these isolates. This allowed for the evaluation of Irish ribotype 078 *C. difficile* isolates associated with hospital, community, and veterinary CDI. This evaluation also included the available 078 genome sequences of a European CDI point prevalence study (EUCLID) and of a Dutch study of farm-associated *C. difficile*. We found genomic relationships (0–2 SNPs) between Irish clinical and porcine isolates, and between Irish clinical and EUCLID isolates. The rate of evolution between the Dutch farm isolates and Irish porcine isolates was lower than expected. The overall findings support greater inter-European transmission of *C. difficile* 078 isolates than previously recognised.

Microbiome analysis was completed for 70 samples, of 41 first episodes and of 29 recurrences. Alpha diversity measures of bacterial richness were similar between these groups. Beta diversity tests allowed separation between groups, despite significant differences within the groups. Two approaches to compare taxonomy each identified bacterial species with significantly different abundance between groups. *Bacteroides xylanisolvens* and *Bacteroides uniformis* were identified by both approaches as less

abundant in recurrent CDI samples, and have been considered elsewhere as potential probiotics.

In conclusion, the application of NGS analysis of *C. difficile* genomes and faecal microbiome samples provides much greater insight to the epidemiology of *C. difficile* infection and recurrence than what was previously understood.

Acknowledgements

To my mother, my sisters Aileen and Claire, my brothers-in-law Alan and Wilson, my nieces Shauna, Lauren, Ciara and Sophia, and my nephew Evan, I hope you will be proud to see me complete my thesis, and I thank you for your long-standing encouragement. To my friends who have provided consistent encouragement, it's been greatly appreciated.

Is mise le meas.

List of Abbreviations

Amplicon	Nucleic acid fragments that are the source and/or product of amplification or replication events
Anosim	Analysis of similarity
Anti-TNF	Anti-tumour necrosis factor
ANOVA	Analysis of variance
APC Microbiome Unit	Alimentary Pharmabiotic Centre
ATM	Amplification tagmentation mix, used in genomic tagmentation
B. dorei	*Bacteroides dorei*
B. thetaiotamicron	*Bacteroides thetaiotamicron*
B. uniformis	*Bacteroides uniformis*
B. xylanisolvens	*Bacteroides xylanisolvens*
BWA	Burrows-Wheeler Aligner
C. difficile	*Clostrioides difficile,* formerly *Clostridium difficile*
C. perfringens	*Clostridium perfringens*
CA CDI	Community acquired *Clostrioides difficile* infection
CC BY	Creative Commons license: credit must be given to the creator
CCEY	Cefoxitin cycloserine egg-yolk
CDAD	*Clostrioides difficile* associated disease
CDI	*Clostrioides difficile* infection
CO	Community onset (of symptoms)
CO$_2$	Carbon dioxide
CT	Computed tomography

CT	Cycle threshold
DNA	Deoxyribonucleic acid
DOB	Date of birth
EIAs	Enzyme immunoassays
ELISA	Enzyme-linked immunosorbent assay
ENA	European Nucleotide Archive
ESCMID	European Society of Clinical Microbiology and Infectious Diseases
FMT	Faecal microbiota transplantation
HA CDI	Hospital Acquired *Clostrioides difficile* infection
HO	Hospital onset (of symptoms)
HPSC	Health Protection and Surveillance Centre
H$_2$	Hydrogen
ICU	Intensive Care Unit
IDSA	Infectious Diseases Society of America
IV	Intravenous
JPEG	Joint Photographic Experts Group standard image format
LNS-1	Library Normalisation Storage buffer 1
LOS	Length of Stay
LTCF	Long-term care facilities
MLST	Multi-Locus Sequence Type
MLVA	Multiple Locus Variable number tandem repeat Analysis
MRN	Medical record number
mRNA	messenger ribonucleic acid
NA	Not applicable

NAAT	Nucleic acid amplification tests
NG	Nasogastric
NGS	Next Generation Sequence/Sequencing
NPM	Nextera PCR Mastermix, used in genomic tagmentation
NS	No (statistical) significance
NT	Neutralisation tagmentation buffer, used in genomic tagmentation
N₂	Nitrogen
OTUs	Operational Taxonomic Units
PCoA	Principal co-ordinate analysis
PCR	Polymerase Chain Reaction
PD	Phylogenetic diversity
PEG	Percutaneous endoscopic gastrostomy
PFGE	Pulsed-field gel electrophoresis
PMC	Pseudomembranous colitis
PPIs	Proton pump inhibitors
PPM	Parts per million
PubMLST	Public database of multi-locus sequence typing
REC	Research Ethics Committee
RCDI	Recurrent *Clostrioides difficile* infection
RPM	Revolutions per minute
SBA	Sheep's blood agar
SHEA	Society for Healthcare Epidemiology of America
SJH	St James's Hospital
SNP(s)	Single Nucleotide Polymorphism(s)

SNV(s)	Single Nucleotide Variation(s)
ST	Sequence Type
tcdA	Toxin A
tcdB	Toxin B
TD	Tagmentation DNA buffer, used in genomic tagmentation
TTROD	Time to resolution of diarrhoea
UCC	University College Cork
WCC	White cell count
WGS	Whole Genome Sequence/ Sequencing
16S rRNA	16s (subunit) ribosomal Ribonucleic acid
4C	Fluoroquinolones, clindamycin, co-amoxiclav, and cephalosporins
95% CI	95% confidence intervals

Table of Contents

List of Figures

List of Tables

1. General Introduction to *C. difficile* & Hypotheses

The first recognition of *Clostridium difficile*'s pathogenicity was made in the 1970s. This related to the epidemic of colitis which followed the introduction of the broad-spectrum antibiotic agent clindamycin into clinical practice [1]. The epidemiology of this infection has continued to evolve, with the emergence of a virulent strain that caused a global pandemic in the early 21st century [2]. Since that time, there has been considerable work undertaken regarding the biology of the Clostridial genus, including a proposal to re-name *Clostridium difficile* as *Clostridioides difficile* [3]. Either name can still be seen in academic publications [3]. For consistency, *C. difficile* will be used in this thesis.

    C. difficile was originally isolated as a normal bacterial constituent of infant gastrointestinal microbial flora [1]. It is an anaerobic Gram-positive bacillus, which is capable of producing toxins [1]. Spore formation is an effective strategy to improve survival and increase dissemination in the (aerobic) environment [4, 5], and toxin formation greatly augments the pathogenicity of the associated infection [6].

1.1. *C. difficile* Infection

C. difficile infection (CDI) in humans affects the gastrointestinal system, as indicated by the first reports of *C. difficile* and antibiotic-associated colitis [1]. Owens *et al.* reviewed the pivotal role of antibiotics in the development of CDI, with consideration of factors that would alleviate or augment this risk, and how further exposure to antibiotics was expected to influence the CDI recurrence [7]. This is shown schematically in Figure 1.1 [7].

    It is acknowledged that both the unknown infectious dose of bacteria and the uncertain incubation period of CDI limit the understanding of mechanisms of exposure and transmission of this infection [8]. Two reported estimates of the incubation period have identified a time not exceeding 4 weeks as compatible with the most probable transmission links observed in those studies [9]. *C. difficile* has been included in the Centers for Disease Control's category of 'Urgent Threats' in their publication series, Antibiotic Resistance Threats, since 2013 due to their concern regarding the extent of this

bacteria's effects on human health [10]. In 2017, there were 223,900 cases of CDI in the USA, of which at least 12,800 patients are known to have died [10].

Figure 1.1: Antibiotics and the pathogenesis of C. difficile infection [7]
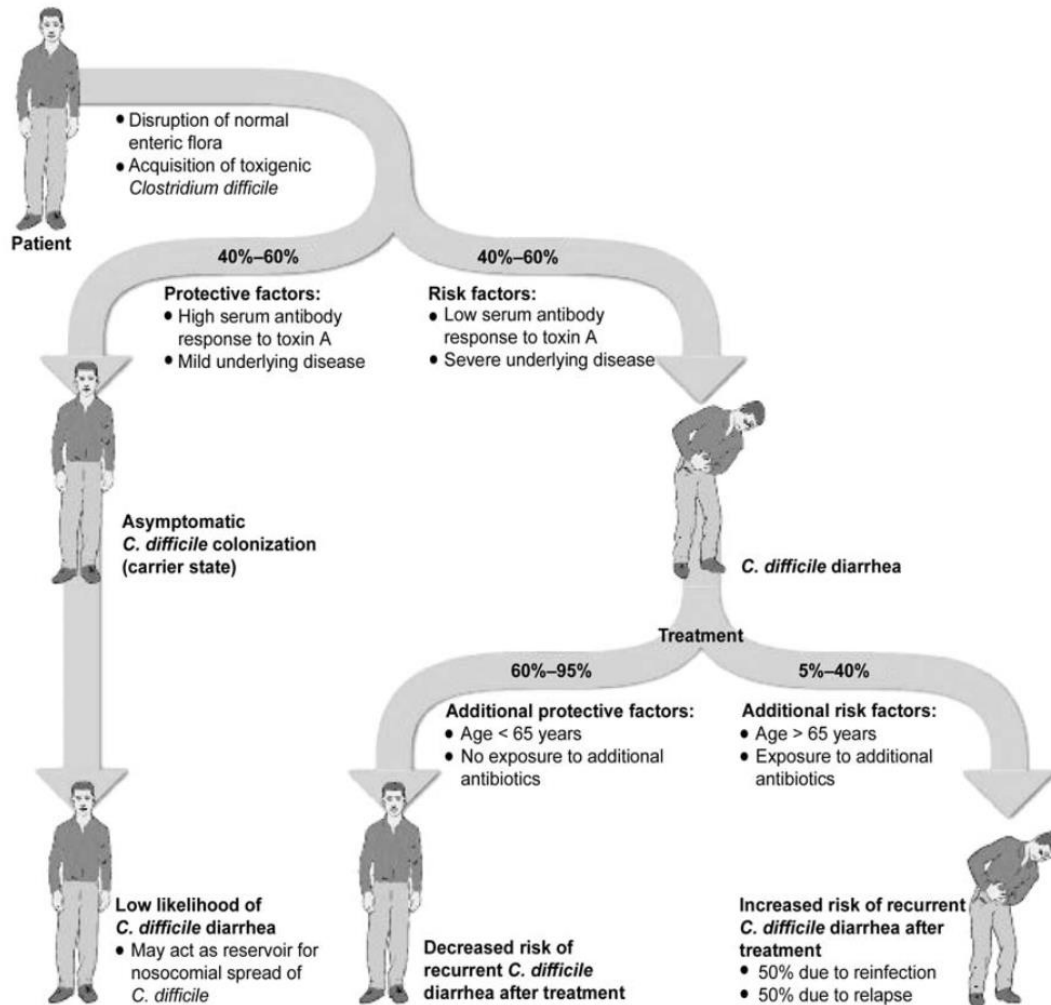
1.2. CDI in Ireland

CDI has been a statutorily notifiable infection in Ireland since 2008. All cases must be reported to the Health Protection and Surveillance Centre (HPSC) [11]. In a 2019 report, HPSC presented the incidence rate of hospital-acquired cases since 2010 [12]. This appears to have been relatively stable in recent years, withstanding the clinical advances such as the introduction of the antibiotic fidaxomicin for management of CDI [13]. It is

shown in Figure 1.2, as the quarterly national hospital acquired CDI rates between 2010 and 2019.

Figure 1.2: Quarterly national HA CDI rates, 2010 – 2019 [12]
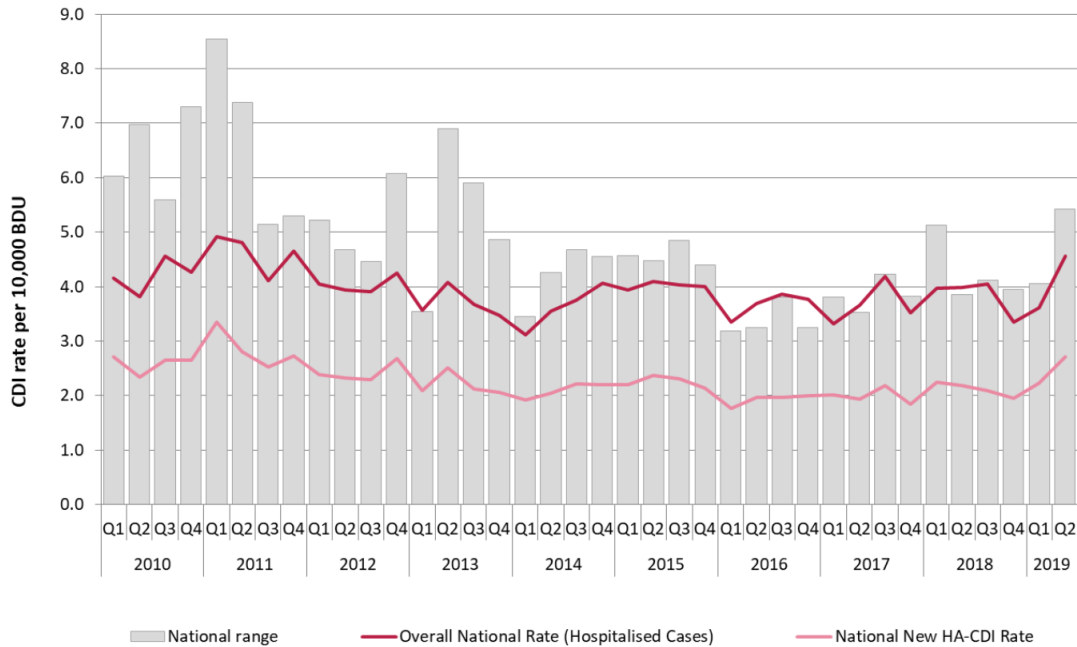
1.3. *C. difficile* Colonisation

There is no straightforward distinction between the concepts of *C. difficile* carriage and colonisation [14]. No significant differences have been identified to distinguish between the risk factors for asymptomatic colonisation and symptomatic infection, including the person's age, co-morbidities, recent antibiotics, or hospital admissions [15]. Patients who have *C. difficile* colonisation at the time of hospital admission are likely to remain colonised after their discharge [16]. Furuya-Kanamori *et al.* completed a narrative review of *C. difficile* colonisation [14]. They described the notable differences between studies regarding the observed duration of colonisation, and that no study had been designed to assess the role of repeated exposure, either from the environment or from other people with asymptomatic colonisation [14].

Crobach *et al.* have also published a comprehensive review of *C. difficile* colonisation [17]. They found considerable differences between the published rates of asymptomatic

colonisation, even allowing for the extensive variation between studies, regarding the working definitions of colonisation and diarrhoea used, and the laboratory methods [17].

Their review includes a summary of the role of antibody-mediated adaptive immune responses to *C difficile,* as mitigation against symptomatic infection or colonised states [17]. In one study, patients who remained asymptomatically colonised had greater titres of anti-TcdA IgG antibodies compared to those who progressed to CDI [17]. Antibodies against *C. difficile* surface proteins may protect against colonisation, whereas antibodies against *C. difficile* toxin can protect against disease [17].

Even after resolution of *C. difficile*-associated diarrhoea, contamination of patients' skin and immediate environment is still common. Persistent contamination has been associated with receipt of more antibiotics after a CDI episode [17].

Asymptomatic carriers are likely to contribute to transmission within the hospital environment. Longtin *et al.* conducted an interventional study with single room isolation for asymptomatic carriers of *C. difficile*, and found a subsequent reduction in the overall nosocomial rate of CDI, with presumed reduction of transmission from asymptomatic carriers as the reason [17].

Point prevalence studies have found that 4–15% of healthy adults had asymptomatic colonisation [17]. The prevalence of asymptomatic colonisation at admission to hospital ranged from 3–21% [17]. Rates of asymptomatic acquisition during hospital admission have also been shown to range from 3–21%, and are associated with increased length of stay (LOS) [17]. Exposure to antibiotics, and the presence of comorbid illnesses, both correlate with the progression to CDI [17].

Crobach *et al.* also emphasise the differences between colonisation in the community or at admission to hospital, and colonisation acquired during a hospital admission [17]. In the community, antibiotic exposure in the preceding 3 months has been found to increase risk of colonisation, and for healthy infants, colonisation was increased in those with pet dogs in the home [17]. Recognised risk factors for colonisation

at admission include recent hospital admission, chronic dialysis, corticosteroid/immunosuppressant use, proton pump inhibitors, and antibodies against toxin B [17]. Previous hospitalisation, chemotherapy, proton pump inhibitors and antibodies against toxin B are also recognised risk for acquiring colonisation as an inpatient [17].

1.4. Progression of *C. difficile* Colonisation to *C. difficile* Infection

Antibiotic exposure, especially to cephalosporins, is known to elevate the risk of new *C. difficile* acquisition and persistent colonisation during hospital admission [16]. Investigation of *C. difficile* colonisation in a long-term care facility (LTCF) identified a colonisation rate of 19% [18]. Of note, approximately 33% of colonised patients had a past history of CDI [18]. For all patients with previous CDI, those with colonisation detected had their episodes about six months before this point prevalence study, and those who had a negative screening test had CDI approximately 13 months before this test [18]. As there was no molecular typing of the *C. difficile* isolates in this study, it is difficult to ascertain if patients had stable colonisation with genetically identical *C. difficile*, or if patients had acquired a different molecular sub-type since their clinical recovery from CDI. Patients who have asymptomatic colonisation with a toxigenic strain of *C. difficile* may progress to symptomatic infection, but asymptomatic colonisation with a non-toxigenic strain does not appear to confer an elevated risk of CDI [17].

There are limited studies that have characterised the intestinal microbiota in patients with *C. difficile* colonisation. These studies support the hypotheses that decreased microbial diversity and richness appear to allow the development of *C. difficile* colonisation, and that the presence of certain bacterial taxa may reduce the progression to symptomatic infection [17].

1.5. Severity of *C. difficile* Infection

There is a spectrum of clinical manifestations of CDI, which can range from a mild infection to gastro-intestinal complications such as ileus or pseudomembranous colitis, and systemic complications, which can include hypotension or renal failure. Mild infection is typically experienced as diarrhoeal symptoms. The development of gastro-intestinal

complications and/or systemic complications may culminate in multi-organ failure and death.

Khanna *et al.* found that 20% of people in their cohort with CA CDI had severe infection, when they retrospectively applied the IDSA criteria [19]. However, in Ireland, only 1.5% of all CDI cases in 2012 were reported as severe, due to patients requiring either intensive care unit admission and/or surgery [11]. There have been substantial differences between international guidelines regarding how they grade the clinical and laboratory features considered as markers of severe infection [8, 11, 20, 21]. These are shown in Table 1.1. Authors of the current IDSA guidelines acknowledge that current criteria are based on expert opinion, as prospectively validated severity scores have not been available [22]. When Khanafer *et al.* applied different sets of criteria for CDI severity to their patients, they found rates of severe CDI varied between 11–59%, depending on the criteria used [23]. Khanafer *et al.* had omitted older age as a risk for severe infection as they felt there was inconsistent evidence for inclusion; they commented that if had they incorporated patients' age, the incidence rate of severe infection would have increased from 59% to 82% of their cohort [23].

Table 1.1: Criteria of Severe C. difficile Infection in Guidelines

Guideline, year:	Demographics	Physical examination	Laboratory investigations	Colonoscopy or sigmoidoscopy	Imaging	Complications
McDonald *et al*., 2007 [24]	N/A	N/A	N/A	N/A	N/A	Admission to intensive care unit, colectomy, and/or death within 30 days of CDI
Cohen *et al*., 2010, IDSA and SHEA [8]	Age	N/A	- Peak white blood cell count - Peak serum creatinine level	- Pseudomembranous colitis	- CT evidence of colitis or ascites	N/A
Health Protection and Surveillance Centre, Ireland, 2013 [11]	N/A	Fever, rigors, abdominal pain	- Leukocytosis - Rising serum creatinine	N/A	N/A	N/A
Public Health England, 2013 [20]	N/A	Temp > 38.5°C, Evidence of severe colitis	- Acutely rising blood creatinine, e.g. 50% increase above baseline	N/A	- Evidence of severe colitis	N/A
DeBast *et al*., 2014, ESCMID [21]	N/A	Fever, rigors, peritonitis, colonic ileus	- Marked leukocytosis - Marked left shift - Rise in serum creatinine (> 50% baseline) - Elevated serum lactate - Markedly reduced serum albumin	- Pseudomembranous colitis	- Distension of large intestine - Colonic wall thickening - Pericolonic fat stranding - Ascites not explained by other causes	- Haemodynamic instability, including shock - Respiratory failure requiring mechanical ventilation
McDonald *et al*., 2018, IDSA and SHEA [22]	N/A	N/A	- Leukocytosis - Rise in serum creatinine level > 1.5mg/dL	N/A	N/A	- Hypotension or shock - Ileus - Megacolon

1.6. Recurrent *C. difficile* Infection

Recurrent CDI (RCDI) is an important clinical issue, occurring in 20–40% of patients who experience a first episode of CDI [2]. RCDI has been defined as repeated episodes which occur within eight weeks of each other [24]. Lessa *et al.* estimated there were 83,000 first recurrences of CDI in the USA, 2011 [25]. In Ireland, there were approximately 180 episodes of RCDI reported in 2018, as 9% of all cases notified to the HPSC [26]. RCDI augments the risk of further recurrences, often in association with further exposure to CDI risk factors, including antibiotics [7, 22].

However, clinical symptoms can occur as either a relapse of the infection caused by the previous infecting *C. difficile* strain, or by the acquisition of a new strain. Chen *et al.*'s investigation of 62 episodes of RCDI with PCR ribotyping of the bacterial isolates found that 51.6% were due to the same ribotype, 42% were due to a different ribotype (suggesting re-infection), and 6.5% had evidence for genetic relapse and re-infection [27]. Barbut *et al.* found that 48% of clinical recurrences were caused by a different *C. difficile* ribotype, but Kamboj *et al.* reported that 88% of their clinical recurrences within 8 weeks were caused by identical ribotypes [28, 29]. When Eyre *et al.* used MLST to investigate isolates associated with clinical recurrences, they found that 77% had the same ST as the first CDI, with a peak incidence at 14 days after that episode. Reinfection, with a different ST, had a later peak incidence, at 30 days after first episode CDI [30]. When Martin *et al.* used whole genome sequence analysis on their recurrent CDI isolates, they found that > 80% met their criteria of probable relapse (0–2 Single Nucleotide Polymorphisms (SNPs) between isolates) and there was a median time of 26 days between these. Clinical recurrences with probable reinfection (> 10 SNPs apart) occurred slightly later, with median onset at 32 days after first episode [31]. There is considerable variation in the methods used in these studies, including the operational definition of recurrence, and the various molecular typing mechanisms used. PCR ribotyping and MLST analysis are not sensitive enough to identify two or more isolates within a ribotype or sequence type as clonally related or not [32], but the finding of a different ribotype or sequence type of *C. difficile* with a clinical recurrence of symptoms is sufficient to categorise as a re-infection.

Guidelines have advised that clinical assessment alone cannot distinguish between recurrence and reinfection [8, 22]. Recurrence may reflect ongoing carriage of that *C. difficile* strain in the patient's microbiome, or another exposure to that strain within the person's environment. It has been hypothesised that RCDI occurs when the patient's gastrointestinal microbiome cannot restore colonisation resistance after the initial CDI episode, or that the immune response to this infection was deficient [33]. Intestinal colonisation resistance is thought to be affected by the antibiotic used to treat CDI [34]. Differential effects of vancomycin and metronidazole on both *C. difficile* and the residual microbiome have been identified, when used as antibiotics to treat CDI [35]. In their exploratory study with three confirmed RCDI out of ten patients, Chang *et al.* identified stool samples from RCDI patients with RCDI had lower diversity and richness of bacterial species than either patients with CDI or controls [33].

Abou Chakra *et al.* described older age, use of antibiotics after CDI diagnosis, use of proton pump inhibitors and *C. difficile* strain type as the most frequent risk factors for RCDI in their systematic review [36]. Of interest, Fekety *et al.* found that recurrence of CDI was not associated with greater severity of infection [37] More recently, Appaneal *et al.* identified 32 risk factors for RCDI: these included medications used before, during, or after CDI. Medications that conferred risk were certain antibiotics, probiotics, laxatives, PPIs and immunosuppressants. Other predictors included comorbidities, especially with recognised immunosuppressive effects and/or medications [38]. Studies that focus on RCDI, from either a clinical and/or a molecular perspective, are presented in Table 1.2.

Clinical prediction rules for RCDI have been described, and have been included in Irish guidelines for management of CDI [11]. Caution is required, as these rules have been derived from and validated in small cohorts, without an impact analysis in clinical practice [39]. For example, Hu *et al.* had only 22 and 13 patients with RCDI in the derivation and validation cohorts respectively of their clinical prediction rule for RCDI [40]. When van Rossen *et al.* retrospectively applied two clinical prediction rules to a cohort across six hospitals in the Netherlands, they found that both rules performed poorly, and failed to predict most RCDI [41]. RCDI has been associated with a significantly increased risk for all-

cause mortality in the six months following the recurrence, compared to people who experience a single episode of CDI [42]

Table 1.2: Clinical and/or molecular investigations pertaining to RCDI

First author Year of publication	Type of study Clinical (C) and/or molecular details (M)	No. of patients with RCDI / no. of recurrent episodes	Major findings	Comment
Fekety, 1997 [37]	Retrospective analysis of prospective RCT (C, M)	37 patients with RCDI; self-reported total of 60 episodes RCDI before enrolment	Odds ratios: - Season of onset of RCDI 7.73 - History of 2+ CDI episodes 3.87 - No. of (CDI) unrelated antibiotics 2.97	- Exclusion of patients with significant immunocompromise - Follow-up specified to 2 months following the enrolment episode of CDI - Immunoblot findings as molecular typing available at that time - Significant findings of multivariate analysis outweighed demographic factors and clinical CDI of univariate analysis - More features of severe CDI in RCDI, not statistically significant in univariate/ multivariate analysis
Barbut, 2000 [28]	Retrospective (M)	93 patients with RCDI	45/93 clinical recurrences were molecular re-infections, caused by different ribotypes	No clinical factors assessed
Tang-Feldman, 2003 [43]	Retrospective (M)	18 patients with RCDI	6/18 clinical recurrences were molecular reinfections, caused by different ribotypes	No clinical factors assessed
Dial, 2004 [44]	Cohort and case-control study (C)	21 patients with RCDI	Univariate analysis: - PPIs OR 5.2 (95% CI 1.1–24.6) - Diabetes mellitus OR 2.7 (95% CI 0.8 – 9.2)	History of CDI was an exclusion criteria for the case-control study

			Multivariate analysis: - PPIs OR 5.1 (95% CI 1.1 –24.9)	
Hu, 2009 [40] (C)	Derivation and validation cohorts	Derivation cohort: 63 patients with CDI, of which 22 had RCDI Validation cohort: 89 patients with CDI, of which 13 had RCDI	Clinical prediction rule included age > 65 years, severe or fulminant illness, and additional antibiotics after CDI. Correct classification of 77. 3% (95% CI 62.2 – 88.5%) of derivation cohort and 71.9% (95% CI 59.2 – 82.4%) of validation cohort	Sensitivity at expense of specificity, small sample sizes are concerning regarding the validity of the rule in other settings
Kamboj, 2011 [29] (M)	Retrospective cohort	102 patients with 134 episodes of RCDI	PCR ribotype results were identical for 88% of 85 RCDI within 8 weeks of CDI, and 65% of 49 RCDI > 8 weeks from CDI.	Limitations of PCR ribotype analysis to determine molecular relapse vs. reinfection No clinical factors assessed
Bauer, 2012 [45] (C)	Database from two randomised controlled trials pertaining to CDI treatment	194 patients with RCDI	Renal failure, creatinine $\geq$ 133 μmol/L risk ratio 1.45 (1.05 –2.02)	Monitoring for recurrence was limited to 28 $\pm$ 2 days from end of treatment Interactions with the investigational medical products of the trial, as lower recurrence rate with fidaxomicin vs. vancomycin
Eyre, 2012 [30] (C, M)	Retrospective cohort	363 recurrences	Recurrence risks higher among patients with non-elective admissions, previous GI ward admissions, last discharged 4–12 weeks before CDI diagnosis, and CDI diagnosed at admission. Risks also included age, previous total hours admitted and C-reactive protein level at first CDI.	CDI diagnosed 2006 – 2010, with follow up until 2011. MLST findings suggest relapse occurs soon after completion of CDI antibiotics, and reinfection slightly later. Although they used the risk factors to create a score, it has not been validated.

			77% recurrences had same ST, peak incidence 14 days after 1st episode. 23% had different ST, peak incidence 30 days after 1st episode.	
Marsh, 2012 [46]	Retrospective (C, M)	82 patients	Multivariate analysis: - Relapse (51/82): - Infection with ribotype 027 OR 6.9 (1.7, 28.2) - Opiate use during previous episode OR 13.1 (3.2, 54) - Re-infection (31/82) - Non-CDI antibiotic (previous episode) 0.1 (0.01, 0.5) - Inflammatory bowel disease 0.04 (0, 0.5) - Antimicrobials in 12 weeks prior to 2nd episode 0.1 (0.01, 0.8) - 8-week cut-off misclassified 44/117 (38%) episodes as recurrence (MLVA)	
Doh, 2014 [47]	Retrospective cohort (C)	23 patients with early RCDI and 17 patients with 'delayed recurrence'	Risk factors for 'early' RCDI: - Nasogastric tube insertion OR 8.7 (1.34, 59.1) - PPI or H2-blocker medication OR 5.4 (1.0, 28.9) Risk factors for 'delayed' RCDI: - Age > 70 years OR 4.4 (1.0, 89.9) - Nasogastric tube insertion OR 40.1 (2.6, 608.1) - PPI or H2 blocker medication OR 1.1 (1.0, 1.2)	CDI diagnosed 2007–2008This study looked at 'delayed recurrence', specified as > 8 weeks after initial episode. No molecular typing of *C. difficile* isolates Small sample size, with high mortality (23%) during the index admission.

Ramanathan, 2014 [48]	Retrospective cohort of 1464 episodes of CDI (C)	315 episodes of RCDI	Risks for RCDI: - Concomitant fluoroquinolone antibiotics OR 1.39 (1.08, 1.8) - Concomitant tetracycline antibiotics OR 0.35 (0.14, 0.19) - Cerebrovascular accident OR 0.46 (0.25, 0.85)	Cohort determined by laboratory records
Olsen, 2015 [42]	Retrospective cohort (C)	421 patients with RCDI	RCDI associated with significantly increased risk of death within 180 days of initial CDI treatment compared to people with single CDI. Risk ratio 1.33 (1.12–1.58)	Cohort of 2003 – 2009 Significant risk of all-cause mortality within 180 days for people with RCDI, despite adjustments made for demographics and co-morbidities.
Martin, 2018 [31]	Retrospective cohort study (C + M)	114 episodes of RCDI	95 isolates within 0–2 SNPs of first isolate: probable relapse, median 26 days apart 16 isolates > 10 SNPs: probable reinfection, median 32 days apart	Cohort of August 2010 – April 2012 Predominance of Ribotype 027 C. *difficile*
Appaneal, 2019 [38]	Case-control study (C)	974 episodes of RCDI matched to controls, from 3020 identified	Significant predictors included medication use before, during, or after CDI. These medications included certain antibiotics, probiotics, laxatives, PPIs and immunosuppressants. Other predictors included comorbidities, including some with associated immunosuppressive effects and/or medications.	Cases from May 2010 – December 2014 In this study, RCDI defined as recurrence within 30 days of end of treatment of first episode of CDI

1.7. Epidemiology of *C. difficile* Infection

Current classification modes of acquisition of *C. difficile* depend upon the time of the patient's most recent healthcare exposure [8, 21]. Zarowitz *et al.* concluded that most CDI (84%) seen in LTCFs was more likely to have been acquired by recent acute hospital exposures (as a result of hospital admissions) than within the LTCFs [49].

1.8. Risk Factors

1.8.1. Antibiotics

It is acknowledged that prior antibiotic exposure of hospital inpatients contributes to their risk of developing CDI; an increase of ward-level antimicrobial use has been predicted to increase the rates of hospital-acquired CDI by 13–34% [50]. Durham *et al.* calculated the impact of selective pressure exerted by antimicrobial use in different settings; they found that for every unit increase in antimicrobial drug 'risk ratio', the CDI incidence increased by 160% in the hospital, 33% in LTCFs, and 6.4% in the community [51].

1.8.2. Proton-Pump Inhibitors

There has been much controversy regarding the potential contribution of proton-pump inhibitors (PPIs) to a person's susceptibility for acquiring *C. difficile* colonisation and/or developing CDI. One Dutch study could not find any association between PPIs and either diarrhoea or CDI, although the number of patients recruited in that study may have led it to lack sufficient statistical power to exclude this as a risk factor [52]. Similarly, an Australian study could not definitively exclude such an association, owing to the high prevalence of exposure to PPIs amongst the control group [53]. However, a Canadian study found use of PPIs as an independent risk factor for inpatient CDI, with an adjusted odds ratio of 2.7 (95% CI 1.4–5.2) [44]. A dose-response effect has been described elsewhere, with increasing intensity of acid suppression [54].

Evidence from mouse models demonstrates that PPIs can enhance the severity of intestinal inflammation and symptoms of CDI [55]. Greater levels of expression of *C. difficile* toxins was seen with exposure to PPI; more toxin A was produced by *C. difficile* ribotype 001, and more toxin B and binary toxin by *C. difficile* ribotype 027 [56]. Finally, there is emerging evidence of the impact of PPIs on the human microbiome, associated with diminishing Operational Taxonomic Unit (OTU) counts, which indicates a reduced number of bacterial species present in a particular sample [57]. A more recent Dutch cohort study found that people who took PPIs had differences consistently associated with a less healthy microbiome [58].

1.9. Molecular Typing of *C. difficile*

The development of laboratory methods to distinguish sub-types of *C. difficile* has led to significant gains in our understanding of *C. difficile* epidemiology [32]. PCR ribotype analysis has been one of the most widely used methods for the investigation of *C. difficile*. This technique allows for *C. difficile* categorisation by separating the varied number and sizes of copies of the intergenic spacer region of 16S–23S ribosomal DNA into different banding patterns [32]. These copies can be separated according to their molecular weight by gel electrophoresis, and the resulting patterns are compared to a library developed by this method [32]. As a molecular typing method, PCR ribotyping is considered to have good discriminatory power, with moderate typeability, reproducibility, ease of interpretation and low technical complexity [32]. PCR ribotype details can be used for establishing hospital or regional epidemiology profiles, but are not recommended to be used to determine outbreaks at a local level [32].

Multiple Locus Variable number tandem repeat Analysis (MLVA) has also been applied to the investigation of CDI [32]. In one study, which compared the application of PCR ribotype versus MLVA analysis of isolates, the latter technique was able to characterise 34% of isolates as highly related and 19% of isolates as unrelated [59]. This was more informative than the ribotype results, which could not provide any further distinction between the isolates [59].

The first complete *C. difficile* genome was published in 2006 [60]. He *et al.* were subsequently able to demonstrate the global spread of ribotype 027 isolates, with two distinct lineages, by obtaining and comparing whole genome sequence (WGS) information of these isolates [60]. Eyre *et al.* then applied both MLVA and WGS analysis to their collection of more than 500 *C. difficile* isolates, and found concordance of 95% between the two typing methods regarding the conclusions of potential outbreaks [61]. In a landmark publication, Eyre *et al.* investigated the genomic similarities of 957 *C. difficile* isolates associated with clinical infection in Oxford, and found that contrary to prior expectations, only 33% of these isolates had sufficient genomic similarities to support what was clinically considered to have been nosocomial transmissions [62]. Eyre *et al.* used a threshold of 0–2 SNPs for isolates identified within 4 months of each other to classify *C. difficile* isolates and corresponding CDI episodes as related to each other [62]. Didelot *et al.* had previously determined a mutation rate of 1.4 mutations per genome per annum based on their analysis of serially isolated genomes [63]. However, this mutation rate could potentially be influenced by environmental factors in vivo, and variations of the techniques or processes in the laboratory and the bioinformatic pipelines could affect subsequent interpretations of genomic relatedness [64]. The mutation threshold of 0–2 SNPs to infer transmission has been used in several studies, with thresholds and timing of isolates outlined in Table 1.3. Nonetheless, the European Centre for Disease Control has created a framework to enable the integration of whole genome sequence analysis into outbreak investigation and public health surveillance [65].

Table 1.3: Publications regarding thresholds for SNP analysis of C. difficile epidemiology

Author	Location of study, year of publication	No. of *C. difficile* isolates analysed	SNP threshold	Comment re timing of source isolates, and if restriction on interval between source isolates for analysis
Didelot *et al.* [63]	Oxford, 2012	Serial pairs of isolates from 91 patients with CDI	0 – 2	Isolates 2006 – 2010; separated by 1 – 561 days (unrestricted)
Eyre *et al.* [62]	Oxford, 2013	957	0 – 2	Isolates 2007 – 2011, where isolates detected less than 124 days apart
Eyre *et al.* [13]	Fidaxomicin Phase 3 trials, 2013	93	- 0 – 2: relapse - 3 – 10: indeterminate - ≥ 11: reinfection	Recurrence was assessed at 30 days following end of treatment
Knetsch *et al.* [66]	Netherlands, 2014	65	- 0 – 1: suspected transmission - 3 – 4: Potential transmission event a few years earlier - > 10: excludes direct transmission	Isolates from 2002 – 2011
Kumar *et al.* [67]	UK, 2015	108	0 – 2: identical genotype	Isolates 2008 – 2010
Knight *et al*. [68]	Australia, 2016	40	0 – 2: clonal	Isolates 2012 – 2013
Stoesser *et al.* [69]	Oxford, 2017	158	- 0 – 2: direct/ indirect transmission - 0 – 10: common origin within the last ~5 years	Isolates 2010 – 2012

Eyre et al. [70]	England, 2018	971	- 0 – 2: genetically linked isolates	Isolates 2013 – 2014; transmission analysis excluded a run-in period of 3 months
Kociolek et al. [71]	USA, 2018	131	- Putative transmission: 0 – 2 SNPs within 0–124 days - 0 – 3 SNPs within 124–364 days	Isolates 2012 – 2013
Martin et al. [31]	Leeds, 2018	640	- 0 – 2: recent acquisition/transmission - 3 – 10: Likely to share a common source in the last 5 years	2010 – 2012
Moradigaravand et al. [72]	East of England, 2018	186	0 – 2: highly related	October 2012 – April 2013; June 2014 – January 2015
Eyre et al. [73]	Wales, 2018	338	0 – 2: possible transmission	February – July 2015
Garcia-Fernandez et al. [74]	Spain, 2019	265	0 – 2: closely related	2014 –2016; transmission analysis excluded a run-in period of three months

1.10. *C. difficile* in the Hospital

It has long been considered that environmental contamination within the hospital can occur as a consequence of symptomatic CDI [4, 5]. This leads to more patients being exposed to *C. difficile*, who may also develop symptomatic CDI. Patient placement in a single room with private en-suite toilet facility is well recognised as a key component of *C. difficile* hospital infection prevention and control [75]. Surfaces including floors, commodes and bed frames are known to be easily contaminated by *C. difficile* in the presence of a symptomatic patient, and this contamination may persist for months or even years [75].

The contribution of patients with asymptomatic colonisation to nosocomial transmission frequency of *C. difficile* has been a challenging issue [14]. Durham *et al.* estimated that the rate of transmission from a patient with symptomatic CDI would be 15 times that of a patient with colonisation [51].

1.11. Infection Prevention and Control versus Antimicrobial Stewardship?

Despite the significant heterogeneity of studies, Khanafer *et al.* found evidence to support both infection control policies and antimicrobial stewardship reduce the prevalence of CDI in their review of best practices for hospital management [76]. Other studies have shown antimicrobial stewardship to have a much greater impact on reducing colonisation, rather than reduction of symptomatic CDI [77].

1.12. Community-acquired *C. difficile* Infection

Recognition of community-acquired *C. difficile* infection (CA CDI), where the person has developed clinical CDI without a personal history of hospital admission in recent weeks has been an important development in CDI epidemiology [8]. The prevalence of CA CDI appears to be increasing in Europe, North America, and Australia [3]. The reported risk factors for CA CDI have been inconsistent between studies performed to date. Early investigations, such as Hirschhorn *et al.*'s, examined age-adjusted antibiotic specific

attack rates, and found interactions between antibiotic use, increasing age and risk for CDI [78].

In a retrospective review of almost 1,000 cases of CA CDI across eight US states, 40% of patients had attended outpatient clinics as their only healthcare exposure and 18% had no identifiable healthcare exposure [79]. Almost 36% of patients had no recent antibiotic exposure preceding CDI, but 31% of this group had taken PPIs [79]. The majority of patients with either low or no personal hospital exposure were more likely to have had contact either with infants less than 1 year old or patients with CDI [79]. The finding that infants may have high rates of colonisation of *C. difficile* may be relevant [79].

Taori *et al.* undertook a prospective study in Scotland and found that a particular ribotype of *C. difficile* was significantly associated with CA CDI, whereas the risk factors of immunosuppression and receipt of antibiotics in the eight weeks preceding CDI diagnosis were significant for HA CDI [80]. They also found that four ribotypes of *C. difficile* were most common for both CA and HA CDI, but the relative proportions of each varied considerably [80].

1.13. *C. difficile* in the Community Environment

Chitnis *et al.* proposed that investigation of transmission of CDI within household settings is merited by the prevalence of CA CDI associated with exposure to young infants or patients with CDI [79]. There is evidence to support the presence of *C. difficile* in non-clinical environments. Two US studies that investigated household contamination found *C. difficile* in at least one domestic sample from 83% of households tested, and in all households with a resident known to have recurrent CDI [81, 82]. Shoe bottoms and vacuum cleaner samples were the most common sites of household contamination, at 40% and 41% respectively [81, 82]. Weese *et al.* found a lower prevalence of contamination, 31%, in a convenience sample of households with pets, albeit with a greater diversity of ribotypes of *C. difficile* [83]. Beyond the household, *C. difficile* has been identified at low frequency from retail baskets and trolleys at supermarkets in Saudi Arabia, and from river water in Slovenia [84, 85].

In addition to healthcare and community environments, *C. difficile* has been identified in retail samples of seafood in Texas and meat in Iran [86, 87]. It has been detected in pork sausages and raw beef [88, 89]. Janezic *et al.* demonstrated that *C. difficile* ribotypes associated with clinical illness in humans were also found in animals and river water [90]. Rodriguez *et al.* found further evidence of *C. difficile* isolates in beef and pork retail products with identical ribotypes to those isolated from clinical cases of CDI [91].

1.14. *C. difficile* as a One Health Issue

The 'One Health' concept links the relationships between the health of people, animals, and the environment [92]. Circumstantial evidence would suggest the possibility of zoonotic associations with cases of CDI, such as the emergence of the *C. difficile* ST-11/ribotype 078 in both human CDI and in association with piglets, calves and their immediate surroundings [93]. Janezic *et al.* also investigated the diversity of *C. difficile* ribotypes associated with animal sources across Europe and North America; the greatest diversity of ribotypes was identified in the countries which provided most samples for analysis [94]. A Canadian study investigated the point prevalence of *C. difficile* and found it present in 58% of the veterinary hospitals tested for specified pathogens [95]. *C. difficile* has been identified in a variety of zoonotic samples, taken within countries of almost every geographic region: Belgian cattle and pigs [96], Brazilian dogs, piglets, foals, and calves [97], American pigs, cattle, horses, and dogs [98], Canadian cattle [99], Japanese piglets [100] and veal calves in Canada and Slovenia [101, 102]. Many of these studies have found *C. difficile* ribotypes 078 and 014/020, which are also known to cause CDI in humans [94, 96–103].

By MLVA analysis, *C. difficile* ribotype 078 isolates from Japanese piglets were similar to both European porcine and European clinical cases [100]. Knetsch *et al.* investigated the similarity of ribotype 078 isolates from asymptomatic pigs (n = 19), asymptomatic farmers (n = 15) and clinical cases (n = 31) in the Netherlands [66]. They applied whole genome sequence analysis, and found that farmers and pigs were colonised with identical *C. difficile* isolates in 5/12 farms [66]. Their maximum likelihood

phylogeny showed genomic similarities among isolates of human and porcine origin; perhaps of most interest was the genomic similarity of a 2008 clinical isolate to a 2011 farm isolate [66]. Although there were four single nucleotide polymorphism (SNP) differences between these isolates-which can be considered as an indeterminate relationship allowing the rate of 1–2 mutations (SNP differences) per *C. difficile* per genome per annum may suggest a closer relationship between these two isolates than what was initially considered [66].

1.15. Interactions between *C. difficile* and the Human Microbiome

As indicated, antibiotics are understood to have a key role in the development of CDI [7]. Despite the first isolation of *C. difficile* as a harmless constituent of infant gastrointestinal flora, it was subsequently associated with colitis in adults who had been prescribed clindamycin [1]. There have been studies which have demonstrated a reduction in bacterial diversity and of bacterial species in the gastrointestinal microbiome that appears to be associated with CDI and other gastrointestinal pathogens [104]. This is illustrated in Figure 1.3 [104]. The mechanisms of this association are not fully understood, and may reflect the absence of certain bacterial species, altered metabolism, other interactions between bacteria present, or a combination of these and/or additional factors [105]. Nevertheless, the central importance of the microbiome to recovery from CDI is apparent from the efficacy of faecal microbiota transplantation (FMT) as a therapeutic intervention for patients with recalcitrant CDI [106, 107].

The application of techniques broadly similar to those used to obtain whole genome sequence data of individual bacterial species have also been successfully applied to microbiome analysis. For example, one study of high-throughput amplicon sequencing of microbiome samples found reduced numbers of bacterial species in patients with active CDI compared to patients with asymptomatic colonisation [108]. Gupta *et al.* compared the yield obtained from next-generation sequence (NGS) analysis of 16S rRNA amplicons to more traditional laboratory culture methods of processing pharyngeal and faecal microbiome samples. They found that NGS had a higher yield of bacterial species, compared to that of traditional culture methods [109].

There are still open questions regarding the relationship between *C. difficile* and the gastrointestinal microbiome [106]. These include lack of understanding of the microbiome constituents that play a central role in the establishment or maintenance of colonisation resistance to *C. difficile* [106]. However, there are valuable insights from studies pertaining to FMT. For example, the abundance of Streptococcus correlated with increased susceptibility to CDI in mouse models, and Bacteroidetes is considered to be protective against CDI in human studies [110–112]. Weingarden *et al.* found that after FMT, restoration of normal bacterial composition was quickly followed by normalisation of bile acid composition to secondary bile acids, less conducive for *C. difficile* growth [113].

Figure 1.3: Differences between diverse microbiota (a) and simplified microbiota (b) in the pathogenesis of gastrointestinal infection [104]
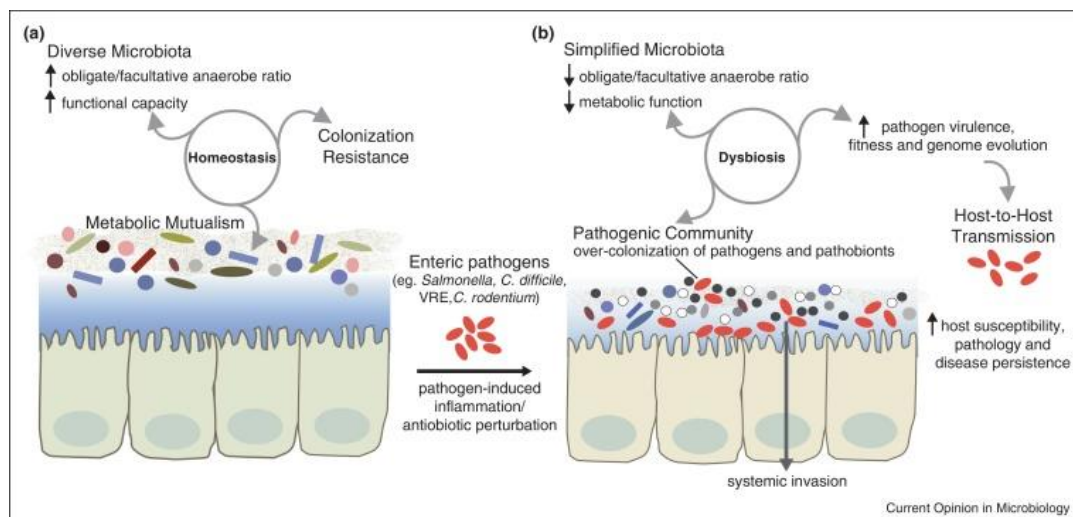
1.16. Aims & Objectives

The work presented in this thesis is intended to address key issues regarding the epidemiology of *C. difficile* infection:

1. Evaluation of the epidemiology of *C. difficile* infection in a major tertiary hospital using PCR ribotype and next-generation sequence analysis of *C. difficile* DNA extracted from selective culture of the faecal samples that have had test results indicative of *C. difficile* in the diagnostic laboratory.

2. Assessment of similarities and potential overlap between *C. difficile* isolates associated with local nosocomial, community and veterinary infections.

3. Exploration of the gastrointestinal microbiome associated with CDI and recurrence, for the possibility of signature bacterial species which may serve as markers for recurrence.

Note: There have been four articles published to date relating to this work. Three have been included in this thesis in their entirety in Chapters 3, 7 and 8. For each of these articles, the publishers state that authors retain the right to include the material in a thesis, as long as it is not for commercial publication. The figures, tables, and references of each article have been incorporated throughout.

2. Materials & Methods

2.1. Introduction

In this chapter, the methodology used to establish the cohort study at St James's Hospital (SJH) is presented. This includes the prospective identification of patients with test results indicative of *C. difficile* and selective stool culture to isolate *C. difficile* from the diagnostic faecal samples. The *C. difficile* isolates were used for PCR ribotype analysis, genomic DNA extraction, whole genome sequencing and analysis. Faecal samples were also used for microbiome analysis, by extracting DNA for use in 16S amplicon PCR sequence generation. As variations of the 16S gene sequence exist between different bacterial species, the 16S nucleic acid fragments, amplicons, can be used to identify the diversity and abundance of species present in a sample [114]. The processes for *C. difficile* genome and 16S amplicon sequence generation were performed on Illumina bench-top sequencers. An Illumina MiSeq machine, located at the TrinSeq laboratory, Trinity College Dublin, was most commonly used for this work. Data storage and analysis will also be outlined.

2.2. Methods

2.2.1. Clinical Identification and Data Collection

Institutional ethics approval was provided for this research by St James's Hospital/Tallaght Hospital Joint Research Ethics Committee (23/9 83/13), including the planned laboratory investigations and associated demographic and clinical details outlined in the specific case report form created. The feedback provided with this approval included their consideration of this work to be an infection control improvement project.

The Clinical Microbiology laboratory of St James's Hospital (SJH) performs *C. difficile* diagnostic tests on faecal samples submitted for patients of SJH, patients attending general practice within the geographic catchment region, and as a referral service for other healthcare facilities. Since September 2013, the diagnostic test in use is

the EntericBio toxin gene PCR assay of faecal samples. The transition from the prior test, a toxin enzyme immunoassay, had occurred in the months before this study began. Only faecal samples of a liquid consistency were tested for *C. difficile* toxin genes If a second faecal sample was submitted to the laboratory within 10 days of a positive PCR result, this second sample was not subject to repeat testing.

Samples with positive PCR result indicative of *C. difficile* were flagged for Clinical Microbiology medical staff to notify the medical and/or nursing staff providing care to the patient. They also liaised with the SJH Infection Prevention and Control team when the result related to an inpatient of the hospital. This was an established clinical practice, which was not affected by the prospective cohort study.

The SJH Laboratory Information System was queried each workday between September 2013 – August 2016 to ascertain positive samples. The sample request details were used to identify the patient's age, gender, requesting clinician (SJH admitting or out-patient speciality, GP, or other facility), the date when the sample was received by the laboratory, and the SJH patient record number, if available. No further details were available for patients who had samples submitted from primary care, other healthcare facilities or who had attended St James's Hospital as either an out-patient or for a day-case procedure. I reviewed the medical charts of hospital inpatients, with completion of a paper case report form (see Appendix A). This case report form was created using the HPSC's sample root cause analysis tool (Version 1, July 2012) of hospital-acquired *C. difficile* infection. Patient details were pseudonymised, and details were entered to a dedicated database using Microsoft Excel. Derived time variables were calculated. These included the number of days from the date of admission to *C. difficile* detection, the length of stay, duration of admission, and the time from *C. difficile* detection to hospital discharge. Details of prescribed medications were obtained from General Practitioners correspondence, such as referral letters for evaluation in the Emergency Department, the clinical and nursing notes of medical charts, and every drug Kardex in use for that patient during that admission. Medications of interest included antibiotics, proton pump inhibitors and immunosuppressants. Antibiotics were evaluated for their suitability for

the suspected or confirmed infection as described in the patient's clinical notes. An antibiotic was considered to be suitable if it was an agent recommended either by the empiric antimicrobial prescribing guidelines for SJH or by a consultant Clinical Microbiologist and/or Infectious Diseases consultant for that patient. Otherwise, the prescription was deemed unsuitable, and when this was still an active prescription, was brought to the attention of the SJH Antimicrobial Stewardship Pharmacist to review with the primary clinical team.

Regarding the institutional Infection Prevention and Control practice, patients were isolated with suspected *C. difficile* or diarrhoea of unknown cause, until 48 hours following return to normal bowel habit in the context of confirmed *C. difficile* infection. The isolation policy was for single room placement with en-suite bathroom or dedicated commode, hand decontamination with soap and water, and use of apron and gloves for contact with the patient. This study did not interrupt standard procedures of Infection Prevention and Control.

2.2.2. Laboratory Methods

2.2.2.1.   Culture of *C. difficile*

Faecal samples were obtained from the Enterics section of the SJH Clinical Microbiology. All samples received an individual project reference number for laboratory workflow and data analysis.

Approximately 10–15 µL of each faecal sample was mixed with 75µL 96% ethanol for twenty minutes, as a preliminary ethanol shock to select for *C. difficile*. 15µL of each mixture of faecal sample and ethanol was then transferred to an individual plate containing Brazier's cefoxitin cycloserine egg-yolk (CCEY) medium [115]. Plates were incubated anaerobically (10% H_2, 10% CO_2 and 80% N_2) at 34°C, and were inspected for growth at 48 hours. 'Broken-glass' colonies typical of *C. difficile* were transferred to blood agar plates, and placed in anaerobic incubation for another 48 hours before isolation of DNA for ribotyping and genome sequencing protocols. All manipulation of faecal samples

and *C. difficile* colonies was performed in a dedicated Category 2 level laboratory with a laminar flow cabinet.

2.2.2.2.   Other sources of *C. difficile* bacterial isolates

1. Bacterial stock samples of community-acquired CDI (CA CDI) *C. difficile* were provided from a national surveillance study of Ireland conducted April-June 2015. Ethanol shock was not required, as these were samples from pure *C. difficile* cultures in glycerol-based stocks and stored at -80°C. These isolates were permitted to thaw on ice, before 5ul was transferred to a Brazier's CCEY plate. PCR ribotype and whole genome sequence (WGS) analysis was performed on these isolates.

2. Isolates of *C. difficile* were recovered from samples of either porcine colonic contents or porcine faeces that had tested positive for *C. difficile* toxins A/B at the Central Veterinary Research Laboratory, Backweston, Ireland. These samples were treated with ethanol shock before selective culture for *C. difficile* was performed, as outlined earlier. PCR ribotype analysis was performed on these isolates, and ribotype 078 isolates were included in subsequent WGS analysis.

2.2.2.3.   PCR Ribotype analysis

Within the Department of Clinical Microbiology, Trinity College Dublin, Dr Micheál MacAogáin had developed an internal protocol for PCR ribotype analysis, derived from the procedure in use at the Anaerobic Reference Laboratory, Cardiff. A collection of *C. difficile* isolates from St James's Hospital had previously been subject to DNA extraction and ribotype analysis was performed in both the Anaerobic Reference Laboratory (Cardiff) and the Department of Clinical Microbiology, Trinity College Dublin. This served as a validation of the internal procedures of the latter.

<u>DNA extraction and Polymerase Chain Reaction</u>

A loopful of C. *difficile* colonies was mixed with 100μL Chelex-100 chelating resin to extract DNA. This mixture was subject to vortex, followed by heat treatment of 99°C in a

water bath for 12 minutes, and centrifugation at 13,000 rpm for 10 minutes to allow the separation of supernatant. 40µL of this supernatant was removed and stored as the DNA for ribotyping. A negative control was obtained with each batch of ribotyping DNA extractions, where nuclease-free water was added to the chelating resin. DNA was stored at -30°C, and was allowed to come to room temperature before PCR preparation.

The next step was to mix 25µL of master mix, 19µL nuclease-free water and 2.5µL each of the forward and reverse primers, for a total volume of 49µL per sample. 1µL DNA was added for starting volume of 50µL per PCR reaction tube.

The PCR was performed in the G-Storm Thermal Cycler, with the following programme:

1. Lid heated to 111°C
2. Temperature set to 95°C × 2 minutes
3. 30 cycles of
   a. Denaturation: 92°C × 1 minute
   b. Annealing: 55°C × 1 minute
   c. Elongation 72°C × 90 seconds
4. Final cooling to 5°C.

Following PCR, the tube contents were transferred to 1.5ml Eppendorf tubes, and placed on a prepared heat block, set to 75°C, for approx. 45 minutes, to allow concentration of products to approx. 20µl.

Agarose gel electrophoresis

For each PCR ribotype electrophoresis, a gel was prepared with 3% agarose, Tris-Borate-EDTA, and 2.5µL GelGreen Nucleic Acid Stain. This mixture was transferred to an electrophoresis casting tray, with placement of a 20-well comb, and allowed to set. The DNA ladders were formulated with 1µL GeneRuler 100bp DNA ladder, 1µL DNA loading dye and 4µL deionised water. Each test well was loaded with 6µL concentrated PCR product, with use of 6µL nuclease-free water as the negative control.

With a 20 well comb, 16 tests were loaded as shown in Figure 2.1, in wells 2–9 and 11–18, and the DNA ladders were placed in wells 1, 10 and 19. The negative control was placed in well 20. The PowerPack was set to 200mV, 200mA for 60 minutes. The gel was then removed from the tray.

Figure 2.1: Placement of test samples between DNA ladders & a negative control in the terminal lane

Well:																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
DNA Ladder	Test sample 1	2	3	4	5	6	7	8	DNA Ladder	Test sample 9	10	11	12	13	14	15	16	DNA Ladder	Neg. control

Image capture and analysis

The gel was placed under ultraviolet light in a closed chamber of Bio-Rad Universal Hood II, and visualised with use of Bio-Rad Quantity One (version 4-6-1) 1-dimensional analysis software. The image obtained was converted to jpeg format, and exported for comparison to the images of the samples which had been subject to external validation at the Anaerobic Reference Laboratory, Cardiff. PCR ribotypes were then allocated to the samples under investigation.

2.2.2.4. Whole Genome Sequence Preparation

Dr Micheál MacAogáin had adapted the guidance provided by Illumina for bacterial genomic DNA library preparation to prepare a local protocol for use within the Dept. of Clinical Microbiology, Trinity College Dublin. This included materials from other commercial proprietors, such as the Roche High-Pure PCR template preparation kit.

Colony preparation and genomic DNA extraction

For each isolate, a loopful of C. *difficile* colony was transferred from each blood agar plate to a sterile Eppendorf tube containing 600µL phosphate-buffered saline. 200µL of re-suspended cells was transferred to a tube with an equal volume of binding buffer, and

40μL of proteinase K. This mixture was incubated for 10 minutes at 70°C, then subject to centrifuge at 8000rpm for 10 minutes, to remove viable cells. For each sample, 320μL of supernatant was transferred to a fresh Eppendorf tube, with the addition of 80μL isopropanol. This was mixed, before transfer by pipette to the reservoir of a high filter collection tube. This tube was then subject to centrifugation of 10,000rpm for 1 minute.

Genomic DNA was obtained by ethanol precipitation, with use of the Roche High-pure PCR template kit in accordance with the manufacturer's instructions. Genomic DNA was stored at -30°C, and allowed to come to room temperature before completion of genomic DNA library preparation.

Pico Green DNA quantification

DNA quantification of genomic DNA was performed by a Qubit pico green fluorimeter, which has a quantitation range of 0.2–100ng. A working solution was made by mixing n × (199μL Qubit buffer + 1μL Qubit reagent), where n equals number of samples for DNA quantification plus two standards. For each standard, 190μL working solution was mixed with 10μL of standard, and for each sample, 198μL was mixed with 2μL genomic DNA sample. All were subject to vortex for 2–3 seconds, and incubated at room temperature for 5 minutes. Standards were used to calibrate the fluorimeter for each run, and the settings were selected for the input of 2μL DNA per Qubit tube. The final concentration provided by Qubit assay was X μg/ml. This reading was entered into an Excel spreadsheet, with unit conversion to μg/μL, ng/μL and 1:10 dilution to calculate the required volume (μL) to proceed. Samples were mixed with nuclease-free water to achieve the desired volume of 30μL per sample.

Tagmentation and PCR amplification

Nextera XT library preparation reagents (Illumina, Eindhoven, Netherlands) were used to generate multiplexed paired end sequencing libraries of the *C. difficile* genomic DNA. The Tagmentation reagents, primers, and DNA were allowed to thaw on ice. All were inverted and subject to brief vortex, to ensure no precipitate was present in any tube. For each of 10 DNA samples, 10μL Tagmentation DNA buffer was added to well of a 'TYC' plate before

addition of 5μL DNA (at concentration 0.2ng/ml). The plate was covered with Microseal B, and placed in a centrifuge for 1 minute at 280 × g, 20°C. The plate was transferred to the thermal cycler with the lid heated to 55°C for 5 minutes, before cooling to 10°C.

Microseal B was removed, and 5μL neutralisation Tagmentation buffer was added to each well by pipette mixture. The plate was covered with a fresh piece of Microseal B and subject to another minute's centrifugation, at the same conditions. It was then allowed to incubate at room temperature for 5 minutes.

At this stage, 15μL Nextera PCR Mastermix was added to each well, with 5μL of both forward and reverse primers. A unique combination of primers was added to each well for which a DNA sample would be added. The primer combination was selected in advance, and checked for conformity with Illumina MiSeq requirements by use of the Illumina Experiment Manager. An example of primer combinations is shown in Table 2.1. Each row of wells is labelled by consecutive letters and each column by consecutive numbers, to ensure a unique well identification number. The Illumina reverse base adapters are added to each column in turn, and the forward base adapters are added to a few wells of each row, to ensure a unique combination of primers.

Table 2.1: Example of Illumina primer combinations added for tagmentation of 10 samples

Well number:	A1	A2	A3	A4	A5
7 reverse base adapter	N701	N702	N703	N704	N705
5 forward base adapter	S502	S502	S502	S502	S502
Well number:	B1	B2	B3	B4	B5
7 reverse base adapter	N701	N702	N703	N704	N705
5 forward base adapter	S503	S503	S504	S504	S504

The contents of each well were gently mixed, the plate was covered with Microseal A and placed in a centrifuge, with similar settings. The plate was then transferred to the thermal cycler for the amplification programme consisting of:

1. Three minutes at 72°C

2. Thirty seconds at 95°C

3. Twelve cycles of:

    a. Ten seconds at 95°C

    b. Thirty seconds at 55°C

    c. Thirty seconds at 72°C

4. Five minutes at 72°C

5. Two minutes at 10°C

PCR clean-up and library normalisation

The plate containing PCR products was subject to centrifugation of 1 minute at 280 × g, at 20°C. Then 22.5µL of the 'clean amplified' product from each well was transferred to the corresponding well of a deep well 'Midi' plate. The AMPure XP beads are subject to vortex, and 12.5µL beads were added to each well. The plate was incubated at room temperature for 5 minutes, and then placed on a magnetic stand for 2 minutes. The supernatant was removed, and the beads were washed with 2 cycles of 80% ethanol. The plate was then incubated at room temperature for 15 minutes. After removing the plate from the magnetic stand, 26.15µl resuspension buffer was added to each well. The beads were re-suspended in this buffer, and incubated at room temperature for two minutes.

25µL supernatant of each well was transferred to the corresponding well of a fresh TCY plate. This could be stored at -20°C for several days, or used immediately for library normalisation.

The concentration of each well was determined using the Qubit pico green quantification method, as described earlier. The ratio of concentration of each sample to the highest concentration of the group (n = 10) was calculated using a Microsoft Excel worksheet, and the resulting ratios were multiplied by a convenient (arbitrary) dilution factor to determine the required volume of each sample. This volume was transferred from each well to a single Eppendorf tube, to obtain the pooled Tagmented library. An equal volume of fresh 0.2M NaOH was added to this library, with 5-minute incubation period at room temperature, to allow denaturation. An equal volume of LNS1 buffer is added (i.e. equal to volume of the pooled samples plus NaOH). The final library concentration was determined by use of the KAPA Library quantitative kit, in accordance with the manufacturers' instructions. A 1:1000 dilution library sample was prepared by adding 0.5µL pooled library to 499.5µL nuclease-free water.

Preparation of flow-cell and Illumina MiSeq benchtop sequencing by synthesis

An Illumina MiSeq V2 reagent kit (300 cycles) was used for each use of the Illumina MiSeq benchtop sequencer. The flow cell was prepared, with the addition of the final pooled library to the paired-end reagent plate, in accordance with the manufacturer's instructions. The initiation procedures recommended by Illumina for the MiSeq benchtop sequencer were completed. The sequence data produced was downloaded from Illumina BaseSpace for further use.

2.2.2.5. Microbiome Analysis by use of 16S Amplicons

These protocols were provided courtesy of Prof. Paul O'Toole and staff of the APC Microbiome Unit, University College Cork, Cork, Ireland.

<u>DNA Extraction by mechanical bead beating of faecal samples</u>

The first step was the preparation of a lysis buffer, consisting of 500mM NaCl, 50mM Tris-HCl (pH 8), 50mM EDTA and 4% sodium dodecyl sulphate, and a TE buffer (pH 8), comprised of 10mM Tris-HCl and 1mM EDTA. These buffers were kept in the cold room, temperature < 5°C.

Fresh 70% ethanol was prepared for DNA extraction. For example, 10.5ml pure ethanol was diluted with 4.5ml nuclease-free water, to process the recommended 12 samples.

Cell lysis

Screw-cap tubes with 0.2g of 0.1mm zirconia beads, 0.2g of 1mm zirconia beads and one 3mm glass bead were sterilised by autoclave. Frozen faecal samples were taken from storage at -80°C, and 0.25g of each sample was added to a sterile screw-cap tube with beads. An individual disposable spatula was used for each faecal sample. Lysis buffer (1ml) was added to each tube, before the tube was placed in the bead-beating machine. This machine was set to operate at maximum speed for one-minute cycles, for three minutes in total. Tubes were cooled on ice between the one-minute cycles. Tubes were transferred to a heat block set to 95°C, and heated for 15 minutes. Tubes were shaken manually every 5 minutes during this heat treatment, before centrifugation at 4°C for another 5 minutes to pellet the faecal particles. The supernatant (S1) was transferred to a 2ml Eppendorf tube and kept on ice.

Fresh lysis buffer was added to the screw-cap tubes containing the faecal pellets and the processes of mechanical bead-beating, heat treatment and centrifugation was repeated. The resulting supernatant (S2) was added to the Eppendorf tube with the earlier yield from that faecal pellet (S1).

Precipitation of nucleic acids

Ammonium acetate (347µl, 7.5M) was added to each lysate tube and mixed well. The tubes were incubated on ice for 5 minutes. Tubes were transferred to the centrifuge, set to 4°C for 10 min at full speed. Faecal pellets were discarded. This supernatant (S3) was

split into two 1.5ml microcentrifuge tubes, and mixed with an equal volume of isopropanol (approx. 635µl). These tubes were incubated on ice for 30 minutes. These tubes were transferred to the centrifuge, set to 4°C for 15 min at full speed. The resulting supernatant (S4) was discarded. The nucleic acid pellet was washed by the addition of 500µl freshly prepared 70% EtOH and centrifugation at 4°C for 2 min at full speed. The supernatant (S5) was discarded by pipette transfer. The nucleic acid pellets were left to dry at room temperature for approximately 15 minutes. Each pellet was then dissolved in 100µl of TE buffer per tube (S6). As each S3 was split into two tubes, there were two S6 aliquots obtained for each nucleic acid pellet. These aliquots were pooled, by the source faecal sample, in universal containers.

Removal of RNA, protein, and purification

(2µl of 10mg/ml) was added to each universal container and incubated at 37°C for 15min. The containers were subject to a brief cycle on the Vortex device for a 'spin down'. Proteinase K (15µl) was added to each container and mixed well. Then, the buffer 'AL' was added and mixed, and incubated at 70°C for 10 min. The tubes were subject to a second brief Vortex spin down.

Each tube had 200µl of EtOH added and mixed, before the tube contents were transferred to a QIAmp column and centrifuged for 1 minute at full speed. The flow through was discarded, before Buffer AW1 was added and the columns were placed in the centrifuge for another minute at full speed. The resulting flow through was discarded, before the addition of Buffer AW2 and the columns were subject to a third cycle of centrifugation. Each column was then transferred to a new collection tube, and placed in the centrifuge for two minutes.

Finally, each column was transferred to a new 1.5ml Eppendorf tube, and 100µl of Buffer AE was added to each column. These were left to incubate at room temperature for five minutes and then placed in the centrifuge, set to operate at full speed for one minute. The resulting eluate (containing DNA) was transferred by pipette back onto the column, with repetition of the incubation and centrifuge steps. Finally the eluate was

transferred to a new 0.5ml Eppendorf tube. This tube was labelled according to the identity of the faecal sample, and nomenclature was maintained between the steps of DNA extraction, nucleic acid precipitation and purification. These DNA extracts could be stored at -30°C if required.

Check DNA quality and prepare dilutions

A 1:10 dilution of the extracted DNA was prepared by adding 4µl DNA to 36µl of AE buffer in a new 0.5ml microcentrifuge tube. This dilution was mixed well, and 9µl were transferred to a new tube. This 9µl aliquot was used to measure the DNA concentration on the Nanodrop in accordance with manufacturers' instructions, and to prepare a 0.8% agarose gel electrophoresis to visualise the DNA quality. The DNA concentration obtained by the Nanodrop reading was used to calculate the required quantity of the other 31µl DNA solution with another 40µl of AE buffer to obtain a 7.5ng/µl solution.

16S Amplicon PCR

Preparation for this process commenced with switching on the safety cabinet, and placement of a waste bag within the cabinet. A PCR microplate was cut into two halves. All required consumable materials and pipettes were placed within the cabinet, and were subject to UV irradiation for twenty minutes. During this time, DNA samples (stored at -30°C), primers and Phusion Taq High-Fidelity reagents were thawed on ice.

The next step was to prepare the 16S Amplicon PCR mix for 20 samples in a 2ml microcentrifuge tube. The required volumes for this mixture are shown in Table 2.2.

Table 2.2: 16S Amplicon PCR Mastermix

Components	For single well (n= 1)	Total mix (n =20)
Phusion Polymerase (blue cap)	0.3	6
5x HF Buffer (red cap)	6	120
dNTPs (white cap)	0.6	12
Primer Forward (5µM)	1.2	24
Primer Reverse (5µM)	1.2	24
Nuclease-free water	18.7	374
Mix volume / well	28	
DNA Template	2	
Final volume (µl)	30	

Then, 28µl of the 16S Amplicon PCR Mastermix was added to each well of the half-cut PCR microplate. The number of wells equalled the number of DNA samples for 16S amplicon sequencing. An equal volume was transferred to a PCR tube, to serve as a negative control. This was followed by the addition of 2.0µl of each diluted DNA samples into a well. The contents of each well were mixed by pipette. The wells of the microplate were covered with PCR cap flat strips. Then the microplate was transferred to the centrifuge, set to 1000rpm for one minute. The PCR microplate and the negative control were both placed in the preheated 96-well thermal cycler, with these settings:

1. Initialisation: 30 minutes at 98°C
2. 25 cycles of:
   a. Denaturation: 10 minutes at 98°C
   b. Annealing: 15 minutes at 55°C
   c. Extension: 20 minutes at 72°C
3. Elongation: 5 minutes at 72°C

<u>Clean-up of 16S Amplicon PCR product</u>

The safety cabinet was switched on, and a waste container was placed in the cabinet. All tips, plates, pipettes, tube racks, markers, EB buffer and other required materials were put in the cabinet, which was set to UV irradiation for 20 minutes. Meanwhile, 40ml of fresh 85% ethanol was prepared by adding 34ml absolute ethanol to 6ml nuclease-free water.

The SPRIselect aliquot was set to Vortex spin for 1 min to re-suspend the SPRI beads and then poured into a pipetting reservoir. SPRIselect (20µl) was added to each well of a flat bottom 96-well plate, according to the number of wells used for the 16S Amplicon PCR microplate. The 16S Amplicon PCR products were transferred, in a volume of 20µl, to each well containing SPRIselect. The well contents were mixed by pipette, and incubated at room temperature for five minutes. The plate was then placed on the Magnetic Separation Rack for two minutes, to allow the SPRI beads to settle. The resulting supernatant was removed and discarded.

While the plate was still on the magnetic separation rack, 180µL of 85% ethanol was added to each well and the plate was permitted to incubate at room temperature for 30 seconds. The ethanol supernatant was removed and discarded. This ethanol wash was repeated, and residual ethanol evaporated during 15-minute incubation at room temperature.

The plate was removed from the magnetic separation rack at this point, and 50µL of EB Buffer was added to each well. Pipette mixing was used to re-suspend the beads, and the plate was left to incubate at room temperature for two minutes. The plate was returned to the magnetic separation rack for five minutes. The resulting eluate is the purified 16S Amplicon PCR product, and this was transferred to a new plate.

Index Nextera PCR

Preparation for PCR Reaction

The index primers were allowed to thaw for approximately 20 minutes and then spun briefly on the vortex machine. A unique combination of primers was selected for each well, i.e. each DNA sample, as described earlier in Tagmentation and PCR preparation.

A new 96-well non-skirt PCR plate was obtained, and 10µl of nuclease-free water into each required well. The purified Amplicon PCR product was transferred, in volume of 5µl into each well. The Nextera index forward and reverse primers were added in turn, according to the earlier allocation. Then, 25µl of Phusion Taq High-Fidelity Mix was added to each well of the plate, with further pipette mixing. The contents and volumes added to each well during this stage is shown in Table 2.3.

Table 2.3: Index Nextera PCR Components

Index Nextera PCR Components	Volume/well (µl)
DNA	5
Nuclease-Free Water	23.5
Nextera Index Forward Primer	5
Nextera Index Reverse Primer	5
2× Phusion Taq High-Fidelity Mix	
Pol (blue cap)	0.5
5× HF buffer (red cap)	10
dNTPs (white cap)	1
Total	50

The plate was covered with an adhesive PCR film seal. The plate was placed in the centrifuge, set to 1,000 x g at 20°C for 1 minute and then transferred to the thermal cycler. These were the thermal cycler settings for this PCR:

1. Initialisation: 30 minutes at 98°C
2. 8 cycles of:
   a. Denaturation: 10 minutes at 98°C

      b.   Annealing: 15 minutes at 55°C

      c.   Extension: 20 minutes at 72°C

  3. Elongation: 5 minutes at 72°C

(New caps were placed on the Nextera Index primers before they were returned to storage at -20°C.)

Clean-up of Index PCR product

The safety cabinet was switched on, and a waste container, tips, pipettes, tube racks, permanent markers, EB buffer and other required materials were placed inside. The cabinet was set to UV irradiation for 20 minutes.

Meanwhile 40 ml of fresh 85% ethanol was prepared by adding 34ml absolute ethanol to 6ml nuclease-free water. The Index Nextera PCR plate was transferred from the thermal cycler to the centrifuge, set to 1000rpm for one minute. The SPRIselect aliquot was set to Vortex spin for one minute to resuspend the SPRI beads and poured into a pipetting reservoir. SPRIselect (56 µL) was added to each well of the Index PCR plate by pipette mixture, and the plate was left to incubate at room temperature for five minutes.

The plate was transferred to the plate magnetic separation rack for two minutes, to allow the SPRIselect beads to settle. The resulting supernatant was removed and discarded.

Then, 180µl fresh 85% ethanol was added to each well, and the plate was left to incubate at room temperature for 30 seconds. These steps were repeated, and the remaining ethanol was permitted to evaporate by incubation at room temperature for 15 minutes.

Then the plate was removed from the magnetic separation rack, before 27.5µl of EB Buffer was added to each well. The total elution volume was mixed by pipette to re-suspend the beads, and the plate was incubated at room temperature for two minutes. The plate was transferred back to the magnetic separation rack for five minutes, to allow

the SPRIselect beads to settle. The purified Nextera Index PCR product was the resulting eluate, of which 25µl of each well was transferred to the corresponding well of a new plate, and covered with film.

Qubit quantification of Index Nextera PCR products

The concentration of each well was determined using the Qubit pico green quantification method, as described earlier.

Completion of Amplicon library preparation and sequencing

Subsequent steps of library normalisation, pooling, denaturation and loading of the amplicon library was performed in accordance with the instructions on pages 16–19 of 16S Metagenomic Sequencing Library Preparation guide 15044223 Re. B, Illumina, Netherlands. This document is included as Appendix B.

2.2.3. Data Management

Data was managed in Excel (Microsoft Office, 2010) with allocation of unique numbers for patients and for *C. difficile* isolates. Pseudonymised patient details from the case report forms and details pertaining to the bacterial isolates were held in separate Microsoft Excel repositories. Statistical analyses were performed in R Studio, in conjunction with R i386 3.1.0. These included chi-square test of proportions for categorical variables, and ANOVA for continuous variables.

2.2.3.1. *C. difficile* genome sequence mapping and variant calling

Paired end reads were mapped to the *C. difficile* 630 reference genome (AM180355) with the Burrows-Wheeler Aligner (BWA) and analysed with the SAMtools package. All sequences were trimmed to remove plasmids, mobile genetic elements and repetitive regions, preserving the core genome of each sequence. Each sequence was then analysed for percentage of all bases with definite calls: A, C, T or G, instead of 'N', which indicates ambiguity of the base call within the sequence generation. This threshold percentage was set to 70% definite base calls, which allowed 93% of sequences to pass to the next stage, cluster generation. The initial threshold for cluster definition was set to 500 SNPs as

maximum permitted distance between isolates. These clusters were visually inspected, using SeaView (version 4.6.1), and used to generate phylogenetic trees, with Clonal Frame scaled distances and PhyML scaled distances. Trees were optimised in FigTree software (version 1.4.3), and with use of the ggtree package in R Studio.

2.2.3.2. Multi-locus sequence type identification

Multi-locus sequence type (MLST) analysis was performed using the Clostridium *difficile* Multi Locus Sequence Typing website (http://pubmlst.org/cdifficile/) developed by Keith Jolley, with funding from the Wellcome Trust, and sited at the University of Oxford [116].

2.2.3.3. Epidemiological analysis and interpretation

There were three clinical classifications made in response to a positive NAAT:

1. *C. difficile* infection (CDI), where the patient had at least 3 unformed stools documented in one 24-hour period, and no other causes of diarrhoea were identified.

2. *C. difficile* recurrence (RCDI), where the patient had resolution of symptoms from an earlier episode, with subsequent development of symptoms meeting criteria of CDI.

3. *C. difficile* colonisation, where the patient did not experience 3 unformed stools within a 24-hour period, with identification of confounding causes of diarrhoea, e.g. concurrent laxative therapy.

CDI was categorised as severe with any one or more of these criteria [11]:

- **Clinical:** fevers, rigors, abdominal pain

- **Laboratory:** Leukocytosis > 15 × 109/L or rise of serum creatinine > 50% baseline, or serum creatinine > 133μmol/L

- **Endoscopy**: Detection of pseudomembranous colitis by visualisation at sigmoidoscopy

- **Radiology:** computed tomography (CT) evidence of colitis or ascites, when a CT had been requested by clinical personnel overseeing that patient's medical care

The case report forms included details of ward allocations during the patient's admission and the responsible clinical team(s). Where there were findings of two or more *C. difficile* isolates with only 0–2 SNP differences between them, the details of the corresponding clinical patients were reviewed for possible associations. As per the earlier methods used by Eyre *et al.*, the assumption was made that cases were infectious from 1 week before diagnosis until 8 weeks after, with a possible incubation period of 0–12 weeks [62]. More recently, this assumption has been refined in other studies to an interval of 90 days between the respective CDI episodes, and unrestricted in another [31, 69, 70]. In this study, possible associations were explored with and without a 90-day interval between the onset of the clinical symptoms of the CDI episodes. The genomic findings were categorised as either plausible or cryptic transmission events:

- Plausible transmission events: the patients with genomically related isolates were either admitted to the same ward, and/or were under the care of the same clinical teams for any part of their respective admissions.

- Cryptic transmission events: there were no associations of either ward or healthcare personnel identifiable between patients with genomically related isolates.

2.2.3.4. Bioinformatic pipeline analysis of 16S amplicon sequence data

Output from the Illumina MiSeq run was subject to the 16S QIIME pipeline, protocol version November 2016. This was performed on Bio-Linux (version 8), with Ubuntu 14.04.3 LTS, GNU bash version 4.3.11 and Qiime v1.9.1. This protocol was written by Hugh Harris and modified by Celine Ribiere, of the APC Microbiome Unit. Assistance for

the analysis of these samples was provided by David Mullins, also of the APC Microbiome Unit.

- Phase 1: filtering of reads and 'demultiplexing'
  Paired reads were joined, with removal of the forward primer. Samples were 'demultiplexed' with a quality finishing check. The reverse primer was removed.

- Phase 2: Chimera filtering and cluster generation
  Unique sequences were identified and sorted by decreasing length and abundance. Operational taxonomic units (OTUs) were clustered with chimera filtering, according to a database. Sequence headers were modified to read OTU_X, and each OTU was matched to its corresponding read in each sample. A filter was applied to select for the OTUs above a set threshold. The OTU table was modified to use in R statistical environment.

- Phase 3: Microbial diversity calculations
  The OTU table was converted to a matrix with read counts for each sample. This matrix was set to a specific format. The representative OTU sequences were aligned, and filtered to remove unimportant gaps. The OTU table was rarefied in R to calculate measures of alpha and beta diversity calculations.

- Phase 4: Assignment of taxonomy to OTUs
  Taxonomy was assigned to OTU sequences, with modification of the Mothur '.taxonomy' file for use within R, providing details of bacterial phyla to species-level information. The table generated was suitable as input for a Mann-Whitney-U evaluation and for use in the DeSeq package.

2.3.   Conclusions

The results of these methods and of the analyses performed will be presented in Chapters 3–8. Findings of the NGS analysis of *C. difficile* genomes pertaining to the prospective cohort study, with some additional CA CDI isolates, are presented in Chapter 4. Chapter 5 contains an overview of the prospective cohort, regarding the diagnostic stewardship of

CDI and observations made about CDI risk factors, clinical features, and antibiotic management. Chapter 6 compares the microbiome findings of patients with CDI and recurrent CDI. A published outbreak report of a unique sequence type of *C. difficile*, ST-295, is presented in Chapter 7. Chapter 8 is the published manuscript regarding *C. difficile* as an infection pertinent to the One Health concept, with a focused evaluation of ST-11/078 isolates.

3. A Preliminary Investigation of Recurrent *C. difficile* Infection

3.1. Introduction

This chapter presents the published findings of a preliminary study regarding RCDI. This investigation was the first application of NGS analysis to explore RCDI *C. difficile* isolates in St James's Hospital with the available epidemiological information [117]. As a joint first author of this publication, my contributions were to review the inpatients' medical notes, to conduct the epidemiological analysis for potential circumstances of transmission of related isolates, and to assist with the development of the text, figures, and tables of the manuscript [117].

3.2. Whole-genome sequencing improves discrimination of relapse from reinfection and identifies transmission events among patients with recurrent *Clostridium difficile* infections [117]

Introduction

Between 15% and 50% of patients who develop *Clostridium difficile* infection (CDI) will suffer subsequent CDI episodes, which adds to the clinical and economic burden of this disease [28, 40, 48, 118]. Recurrence is defined as a CDI episode occurring within eight weeks of a previous infection [24]. Accepted risk factors for recurrent CDI include older age (>65 years), prescribing of additional 'non-CDI' antibiotics, and cumulative time spent in the healthcare environment [30, 37, 40, 48]. Recurrent clinical episodes may be categorized as relapse, when due to the original strain, or reinfection, when caused by a newly acquired strain.

Molecular typing studies of *C. difficile* have provided insight into the proportions of cases with relapsed CDI as opposed to reinfection [32, 46]. Estimates for reinfections range from 12% to 35% of recurrent CDI episodes, within the limits of discrimination provided by conventional typing methods such as polymerase chain reaction (PCR)-based ribotyping, pulsed-field gel electrophoresis (PFGE) or multi-locus sequence typing (MLST) with the intervals to recurrence after a first episode of CDI ranging from 24 to 42 days [28–30, 43, 48, 119].

The use of whole-genome sequencing (WGS) has provided evidence for a higher degree of *C. difficile* strain diversity than previously acknowledged [62, 63, 120]. A recent study applied WGS to 1,223 *C. difficile* strains and found 45% of all isolates investigated to be genetically distinct, suggesting a considerable reservoir of endemic *C. difficile* strains [62]. Of the patients infected with genetically indistinguishable strains, Eyre *et al.* found that 38% had identifiable hospital contact with another symptomatic case and 36% had no recognizable shared epidemiology [62]. This underscores the existence of unidentified *C. difficile* transmission routes [62, 121, 122]. In a subsequent study, Eyre *et al.* applied WGS to recurrent CDI with the consideration that ribotyping may underestimate reinfections caused by endemic ribotypes [13]. This provided improved discrimination between relapse and reinfection through comparisons of paired isolates (index versus first recurrence) and revealed that 81% of recurrences were caused by the same strain, 15% by reinfections with 4% assigned to an indeterminate category [13]. We undertook prospective analysis of CDI episodes meeting clinical and microbiological criteria and identified all patients suffering recurrent CDI over an 18-month period. Strains causing index as well as first and subsequent CDI episodes were characterized using both conventional ribotyping and WGS to assess the level of concordance of these methods in view of the enhanced discriminatory power of WGS.

Methods

Setting

St James's Hospital (SJH) is a 1,015-bed acute tertiary care hospital with some 3,800 staff members and an immediate catchment population of about 350,000. Annual inpatient admissions exceed 25,000 with more than 220,000 outpatient and 46,000 emergency department visits per annum.

Study cohort

Between January 1st, 2012, and June 30th, 2013, all clinical cases of recurrent CDI were identified at St James's Hospital, Dublin, in accordance with national guidelines for recurrent CDI. In addition, any patient suffering 2 clinical CDI episodes was included in our

analysis, even if episodes occurred more than eight weeks apart. Laboratory confirmation of cases meeting clinical criteria was provided by the Premier toxin A and B enzymatic immune assay (Meridian Bioscience Inc., Cincinnati, OH, USA) performed either on direct stool samples ('toxin positive') or on cultured isolates ('culture positive') grown on Brazier's cefoxitin cycloserine egg-yolk (CCEY) agar under anaerobic conditions (10% H_2, 10% CO_2 and 80% N_2) at 37°C for 48–72 h.

Strain collection

Stool samples from patients suffering 2 identified CDI episodes, which were originally confirmed by the Diagnostic Laboratory, were recovered for further analysis. Of 58 identified CDI cases meeting this criterion, stool samples were available for 53 (91%). Stool samples were subjected to alcohol shock and plated on Brazier's CCEY agar to selectively isolate *C. difficile*. From these toxin-positive cultures, a single colony was taken and stored as a spore stock culture at –80°C as previously described [123]. PCR-based ribotyping was performed on all isolates to establish strain relatedness [124].

Whole-genome sequencing

DNA was extracted from *C. difficile* using the Roche High-pure PCR template preparation kit (Roche Diagnostics Ltd, Burgess Hill, UK). Nextera XT library preparation reagents (Illumina, Eindhoven, The Netherlands) were used to generate multiplexed paired end sequencing libraries of *C. difficile* genomic DNA. Resultant libraries were sequenced on an Illumina MiSeq instrument. All short-read data obtained in this study have been deposited in the European Nucleotide Archive (ENA), project accession number PRJEB6575.

Sequence mapping and variant calling

Paired end reads were mapped to the *C. difficile* 630 reference genome (AM180355) with the Burrows–Wheeler Aligner (BWA) and analysed with the SAMtools package [125, 126]. Strains were sequenced to an average raw read depth of 91.1 44.5-fold. Sequence types (ST) were determined using the *Clostridium difficile* Multi Locus Sequence Typing website

(http://pubmlst.org/cdifficile/) [116]. Single nucleotide variants (SNVs) were called using the SAMtools mpileup command consistent with the parameters described by Didelot *et al.* for SNV calling in *C. difficile* [63].

Ethics

This study proposal was reviewed by the hospital research ethics committee (REC ref: 23/9 83/13) and considered to be part of a service improvement for the infection control team.

Results

Recurrent CDI prevalence and associated patient demographics

Over the 18-month study period, a total of 230 CDI episodes were documented at SJH representing a CDI rate of 0.42/1000 occupied bed-days and a recurrence rate of 10% among hospitalized patients. Although recurrent CDI is generally defined as a positive CDI result dated within the preceding eight weeks of a prior CDI infection, for the purpose of this study, we extended our definition to include any CDI episode preceded by a prior episode in the same patient over the course of the entire study (18 months) [24].

Despite our liberal criteria for defining recurrence, 18/19 patients had at least one recurrent episode which conformed to the accepted criteria for recurrent CDI. Using our criteria, 19 index and 39 recurrent isolates were identified among the 58 episodes investigated. Five episodes fell outside the accepted eight-week boundary of the formal definition of recurrent CDI (Figure 3.1).

The demographic details of the 19 patients who suffered recurrence are summarized in Table 3.1. They had a mean age of 73.5 years (range: 35.5–94 years) and a median LOS of 144 days. Patients suffered between two and seven CDI episodes with a median time from admission to first CDI episode and first CDI recurrence of 71 and 32 days respectively. The majority of clinically defined cases were confirmed by detection of *C. difficile* toxins in faecal samples (68%) with the remainder confirmed by direct detection of toxin A/B production by *C. difficile* cultured from faeces. Patients

experienced an average of 2.5 ward transfers (range: 1–7) throughout their admission and were cared for by a range of clinical specialties. Two patients died (of complications unrelated to CDI) and two remained in the hospital receiving ongoing care over the course of the study within an onsite inpatient long-term care facility (LTCF). Of the 15 patients who survived to hospital discharge, nine were discharged to LTCFs and five were discharged home (one discharge location unknown).

Table 3.1: Demographics of patients with recurrent Clostridium difficile infection

Patient demographics	Value
Total patients	19
Age (mean, range; years)	73.5; 35—94
Gender (M; F)	10; 9
Length of stay (days)	
Per patient median, mean	141, 220
Range	9—1780[a]
Mean no. of admissions per patient	1.2
No. of ward placements per patient	2.5, 1—7
(mean, range)	
Clinical specialty ($N = 26$)	
Medical	19 (73%)
General medicine	5 (19%)
Medicine for elderly	5 (19%)
Endocrinology	2 (7%)
Gastrointestinal/hepatology	2 (7%)
Haematology	2 (7%)
Respiratory	2 (7%)
Nephrology	1 (3%)
Surgical	3 (11%)
General surgery	1 (3%)
Orthopaedics	1 (3%)
Plastics/reconstructive surgery	1 (3%)
Psychiatry	1 (3%)
No. of CDI episodes	
Total no.	58
Per patient mean, range	2, 2—7
Microbiological confirmation	
Direct toxin detection	39 (68%)
Identified by culture of toxigenic strain	19 (32%)
Time from admission to first CDI episode	
Per patient median, range (days)	71, 20—1444[a]
Time to first recurrence of CDI[b]	
Per patient median, range (days)	32, 5—191
Outcomes observed	
Survival to hospital discharge	15 (78.9%)
Ongoing inpatient LTCF	2 (10.5%)
Death terminating admission	2 (10.5%)
Of those discharged ($N = 15$)	
Discharge to LTCF	9 (60%)
Discharge home	5 (33%)
Unknown	1 (6%)

CDI, *Clostridium difficile* infection; LTCF, long-term care facility.
[a] Data skewed by inclusion of two patients receiving long-stay care.
[b] For the purpose of this study, new CDI episodes (separated by >8 weeks) in the same patient were considered to be recurrent ($N = 5$).

Figure 3.1: Timeline of Clostridium difficile infection (CDI) episodes illustrating ribotype prevalence and ward location of patients with recurrent CDI

Nineteen patients (P1-19) are each represented by horizontal lines spanning patient admissions observed over the 18-month study interval. Clinical CDI episodes are represented as rectangles. Black lines indicate the length of hospital admission times. Admission and discharge dates are bracketed by vertical lines; closed circles represent admissions terminated by death; and unbracketed lines indicate admissions that precede or success the study interval. The colours of each rectangle correspond to the identified ribotype as indicated in the key. Sequence types (ST) based on whole-genome sequencing analysis, are indicated in parenthesis. One isolate (SJH*) – did not match any reference strains in our ribotyping database and could not be assigned a ribotype. White rectangles represent CDI episodes for which a stool sample was not available. Rectangles with a double boarder indicate repeat CDI episodes which fall outside the eight-week definition of recurrent CDI, and which would be considered 'new infections' under existing guidelines. The ward location of the patient at the time of active CDI is indicated in each rectangle.

54

Investigation of recurrent isolates by PCR-based ribotyping

In 16/19 cases a single ribotype was identified per patient, consistent with relapse. PCR ribotyping results supported relapse in the majority (27/39) of recurrences, with only five reinfections identified (Figure 3.1). In the case of seven CDI episodes, the nature of CDI recurrence (relapse or reinfection) could not be confirmed due to missing samples (Figure 3.1; P9, P11, P12, P18). Two patients suffered both relapse and reinfection (Figure 3.1; P1 and P7). One of these patients (Figure 3.1, P1) had two recurrent episodes involving a ribotype 078 strain, and suffered a subsequent ribotype 017 reinfection, followed by a reinfection with the original 078 strain. Another patient suffered two reinfections, the second of which relapsed (Figure 3.1, P7). Of the 14 ribotype profiles identified, 078 and 020 predominated with a total of nine and 12 linked CDI episodes identified among four and three patients respectively. A ribotype 017 strain was isolated from five episodes among three patients (Figure 3.1). No patient harboured the 027/NAP1/BI strain.

Investigation of recurrent isolates by WGS

All isolates were subjected to WGS and comparative SNV analysis with reference to the *C. difficile* 630 genome (AM180355). MLST sequence types (ST), predicted from WGS data, were consistent with previously observed MLST–ribotype correlations [127]. This allowed assignment of one isolate, for which a ribotype designation could not be established, to ST-3 (Figure 3.1, P14). Strains of the same ribotypes, causing multiple infections in individual patients, were compared by WGS in an effort to confirm relapse with increased certainty. Overall, WGS analysis was consistent with ribotyping in defining reinfection and relapse; strains of the same ribotype from individual patients differed by 2 SNVs (Table 3.2, Appendix C Supplementary Table I). Thus the SNV differences observed among these strains were within the bounds of previously accepted criteria for inferred relapse in *C. difficile* [13].

One or two SNVs were identified on comparing first and last isolated strains in patients who relapsed. In almost half (7/16, 44%) of these patients, we observed the occurrence of within-strain SNVs emerging over the course of their recurrent CDI

infections. Where emergent SNVs were observed, the number of SNVs per strain ranged from one to two, or two to 15 SNVs per strain per year, when the observed time interval between isolation of first and last isolate in each individual patient was considered (Appendix C Supplementary Table I). The genomic locations of the SNVs which arose over the course of clinical CDI relapses are detailed in Table 3.2.

Patient-to-patient transmission of C. difficile inferred by PCR-ribotype analysis

Fourteen patients shared strains of the same ribotype; ribotype 078 was shared by four patients, ribotypes 020 and 017 each infected three patients, and ribotypes 050 and 003 were each found to be shared between two pairs of individual patients (Figure 3.1). The electronic records of patients infected by *C. difficile* of identical ribotype were investigated for epidemiological evidence supporting transmission including shared space and time on a ward, shared medical specialty team, and overlapping admission times. This identified 10 possible patient-to-patient transmission events (Figure 3.2A, A–J). Six such events were substantiated by clinical data including shared ward placement (Figure 3.2A, A–E) or shared medical specialty (Figure 3.2A, F). Four potential transmission events involved shared ward placement of symptomatic and non-symptomatic patients (Figure 3.2A, A–D). Ribotyping also highlighted four apparent transmission events without substantiating epidemiologic evidence other than overlapping hospital admission times (Figure 3.2A, G–J).

WGS analysis of ribotyping-inferred CDI transmission events

To further investigate transmission events that had been inferred by ribotyping, all isolates of the same ribotype were compared by WGS analysis. The numbers of SNVs identified among isolates implicated in transmission are detailed in Figure 3.2B. Among the 10 suspected transmission events, WGS analysis excluded five (Figure 3.2A; A, D, E, I, J) through the identification of strain divergences of between five and 86 SNVs (Figure 3.2B). Five strain transmissions were substantiated by WGS analysis (Figure 3.2A; B, C, F, G, H). Although strains implicated in events G and H differed by 2 SNVs, a difference of 1 vs 0 SNVs was observed for transmission events G (P1–P7) and H (P5–P7) respectively and

the analysis thus marginally favoured event H. Although three SNVs were found to separate strains implicated in transmission event 'B', two of these SNVs appeared to have arisen over the course of CDI relapse in patient 'P2' (Figure 3.2B). Thus, in spite of the three observed SNVs, the acknowledged cut-off of 2 SNVs for inferring strain relatedness was not breached and this transmission event was supported. The SNVs which emerged between transmission events 'B' and 'C' are detailed in Table 3.2. In total, four transmission events were inferred from WGS among the 19 patients investigated.

Table 3.2: Emergent within-strain SNVs and their predicted impact on gene function

Patient	Ribotype	ST	SNV locus	Reference[a]	Variant[b]	Synonymous[c]	Protein alteration	Locus tag[d]	Gene function
P2	050	ST-16	1412874	T	C	No	V93A	CD630_12140	*spo0A*; stage 0 sporulation protein A
			3243804	C	A	No	A176S	CD630_27870	*cwp84*; cell surface protein
P5	017	ST-37	1826371	A	C	No	S215R	CD630_15770	*pgm*; alpha-phosphoglucomutase
			4111335	G	A	No	G149S	CD630_35180	*murC*; UDP-*N*-acetylmuramate-L-alanine ligase
P7	002	ST-45	1186090	G	T	No	P36Q	CD630_10160	Transcriptional regulator, MerR family
			1484356	G	A	Yes	na	CD630_12770	Putative acetyltransferase
P8	078	ST-11	3810191_3810192insT	AT	ATT	No	Frameshift	CD630_32550	*rgaR*; two-component response regulator VirR-like
P12	003	ST-12	2708584_2708585insC	TC	TCC	No	Frameshift	CD630_23410	*abfD*; gamma-aminobutyrate metabolism dehydratase/isomerase
P13	131	ST-122	829898delC	ACC	AC	No	Frameshift	CD630_06840	Putative ATP-dependent peptidase, M41 family
P19	020	ST-2	4097019	C	T	No	G424E	CD630_35060	Conserved hypothetical protein
			4061875	C	T	No	R283C	CD630_34700	*atpA[e]*; ATP synthase subunit alpha
P1 (vs P5)[f]	017	ST-37	457298	G	A	na	na	intergenic; CD630_03540-CD630_03550	112 bp upstream of: CD630_03550 (*bglF*; PTS system, IIABC component)
P9 (vs P2)	050	ST-16	60312	G	A	No	A289V	CD630_00370	*acoB*; acetoin dehydrogenase E1 component

SNVs, single nucleotide variants; ST, sequence type; na, not applicable.
[a] Sequence identity at relative SNV locus in *C. difficile* 630 (AM180355) reference genome.
[b] Sequence identity at relative SNV locus in *C. difficile* clinical isolates.
[c] Functional status of SNV at protein level (synonymous or non-synonymous).
[d] Relative locus tag in *C. difficile* 630 reference genome in which within-host SNVs were observed.
[e] The atpA gene is used in the *C. difficile* MLST scheme. This mutation gives rise to a novel MLST profile which has been designated ST-295.
[f] Observed SNVs were identified on comparison of transmitted strains rather than over the course of relapse in individual patients.

Figure 3.2: Timeline of suspected Clostridium difficile infection (CDI) transmission events investigated by whole-genome sequence analysis (WGS) among patients with recurrent CDI

(A) Ten suspected transmission events were defined based on the identification of shared ribotypes among 13 patients with overlapping hospital stays (column one, A-J). These transmissions were further supported by epidemiologic data including either shared ward time (W##) or shared medical specialties (SMS*). For four suspected transmissions (G-J), no supportive clinical data other than overlapping hospital admissions (OA, †) was observed. The respective transmission events, supportive clinical data, and the patients involved are detailed in columns one, two, and three, respectively. Horizontal coloured bars represent overlapping patient ward placements, time under common medical specialties or overlapping admission times consistent with transmission. The colour of the bars corresponds to colour scheme used in Figure 5.1 and indicates strain ribotype. Closed circles represent CDI episodes that occurred on a ward where transmission was suspected. Open circles indicate an episode that occurred on a different ward to that of the suspected transmission. Only episodes caused by ribotypes implicated in transmission are shown.

(B) Analysis of suspected transmission events by WGS. All isolates implicated in transmission events were subjected to WGS analysis. The number of SNV differences between same-ribotype isolates is illustrated by pairwise comparison tables for all strains of a shard ribotype against each other. Each table is coloured according to ribotype, consistent with Figures 5.1 and 5.2A. The degree of coloration in each square corresponds to the degree of similarity (less observed SNVs) between strains.

58

Discussion

We investigated the molecular epidemiology of recurrent CDI cases at a tertiary referral hospital comparing conventional PCR-based ribotyping and WGS analysis. Overall, the age profile of patients with recurrence was reflective of national data for adult inpatients in Ireland [128]. However, our recurrent CDI cohort had an exceptional LOS which placed them in the minority (3.3%) of inpatient admissions nationally [128]. Even within this category, the national mean LOS is estimated to be 65.5 days, considerably shorter than our patients' experience [128]. This finding was likely attributable to underlying comorbidities as well as CDI. Although we did not undertake formal calculation of comorbidity, available clinical details suggested that this group had considerable medical issues and nursing requirements (data not shown). This is also reflected in the high percentage of the group discharged to long-term care (60%) compared to 4.7% of all adult inpatients nationally [128]. Patients thus compromised a vulnerable group who experienced multiple CDI episodes over prolonged hospital admissions.

Fourteen distinct ribotypes were identified including the 078 strain, which has been reported previously in recurrent CDI cases in Ireland [129]. Strains belonging to ribotypes 020 and 017 were also present. All three ribotypes have proven virulence potential and have been implicated in recurrent CDI [130–132]. Notably, the 027/ NAP1/B1 lineage, which has been associated with recurrent CDI, was not detected. Local ribotype prevalence data, for strains collected over the duration of this study, suggest that 078 and 020 strains are the most frequently occurring ribotypes at our hospital, each accounting for 19% of observed isolates, whereas the 027/NAP1/B1 lineage was less frequently observed (unpublished data). Thus, strain ribotype prevalence among our recurrent CDI cohort appeared to reflect local *C. difficile* epidemiology. Two patients suffered both relapse and reinfection (Figure 3.1; P1 and P7). Similar findings have previously been described and highlight the complex epidemiological scenarios that arise among patients with recurrent CDI [28, 133]. However, in our cohort, the majority of CDI episodes resulted from same-strain relapses with only one patient suffering reinfection as the sole cause of clinical recurrence.

To confirm persistent, same-strain relapse among recurrent CDI patients, we used WGS to distinguish relapse and reinfection with greater accuracy. All relapses (as identified by PCR ribotyping) were confirmed by WGS; strains causing relapse were found to be identical or differed by <2 SNVs at the whole-genome level which is considered an acceptable cut-off within the bounds of the predicted within-host evolutionary rate for *C. difficile* [13, 63]. Five patients experiencing intervals greater than eight weeks between CDI episodes, which thus fell outside accepted formal definitions for recurrent CDI, were nonetheless included in our analysis (Figure 3.1; P1, P3, P4, P11, P12). According to accepted guidelines, these should be considered as new rather than as recurrent CDI episodes in light of the exceptional interval between episodes [24]. Interestingly, WGS analysis demonstrated that all five patients suffered relapse by strains genetically indistinguishable from their index case, in spite of the long intervals between episodes. The longest interval observed between infections caused by identical strains was 191 days, which exceeds current definitions for recurrence by > 19 weeks. Previous WGS analysis of paired *C. difficile* isolates from cases separated by one to 561 days also identified apparent relapse (<2 SNVs between isolates) over exceptional timescales [63]. However, such lengthy intervals between index and relapse could also be interpreted as reinfections by genetically indistinguishable strains via common environmental contamination sources. A limitation of our study was the absence of WGS data on the broader population of *C. difficile* strains at this institution, including those causing non-recurrent CDI. This would have provided greater insight into transmission dynamics between recurrent CDI patients and the broader hospital population and whether environmental sources of genetically identical strains were present or, conversely, whether patients with relapse represent reservoirs for onward CDI transmission.

Longitudinal sequencing of *C. difficile* isolates from relapse episodes identified SNVs occurring over the course of recurrent CDI in individual patients (Table 3.2). Of the 11 within-strain SNVs identified, 10 led to predicted non-synonymous changes at the protein level. A mutation in the *spoOA* gene, encoding a key regulator of *C. difficile* sporulation, virulence, and metabolism, was observed in a ribotype 050 strain over the

course of relapsing CDI (Table 3.2) [134]. Mutational alteration of *spoOA* has been observed previously in a *C. difficile* strain from a fidaxomicin-treated patient with CDI relapse [13]. Other regulatory genes affected included *rgaR* – encoding a predicted two-component response regulator and a gene encoding a MerR-family transcriptional regulator. The emergence of mutations in central regulators of virulence *in vivo* can radically alter bacterial physiology, triggering adverse clinical outcomes [135, 136]. Whereas our study was not designed to investigate the correlation between the emergence of bacterial mutations and clinical outcome, such changes may have clinical relevance and, giving the growing adoption of WGS technology, they may become the focus of larger WGS studies addressing their clinical impact. Other genes in which mutations were observed included *cwp84*, encoding a protease involved in processing of the surface layer protein and biogenesis of the *C. difficile* cell wall, and *murC*, encoding an essential component of peptidoglycan biosynthesis [137, 138].

In many cases, sustained *C. difficile* infections recurred in our patients over prolonged intervals where multiple patient transfers between wards and medical specialties occurred. Given the potential for transmission of *C. difficile*, we focused our investigation on several apparent patient-to-patient transmission events among our recurrent CDI cohort. In total, 10 potential transmissions were suggested based on ribotyping analysis, including six that were supported by clinical data (Figure 3.2A). However, analysis of WGS substantiated only four transmissions identifying multiple SNVs separating purportedly transmitted strains. The four transmission events supported by WGS were linked to at least five subsequent CDI episodes including at least one which recurred (Figure 3.2, transmission event 'F'). WGS identified a ribotype 017 (ST-37) strain causing relapse in one patient which subsequently caused reinfection in two others (Figure 3.2A, transmission events C and G). Analysis of WGS data also highlighted at potential transmission event concerning ribotype 050 (ST-16) (Figure 3.2A, event B) which was contentious due to the identification of three SNVs (greater than the accepted cut-off of <2) between the strains (Figure 3.2B), in spite of supportive epidemiological evidence. More focused analysis revealed that, when within-strain SNVs arising in the transmitted

strain were considered, only a single SNV difference separated the two strains (Figure 3.2B and Table 3.2). This suggested that the transmission event occurred prior to the subsequent accumulation of SNVs in the index strain of patient 'P2', thus distorting interpreted strain divergence when only the temporally closest isolates were compared. This highlighted the advantage of considering multiple strains when trying to establish patient-to-patient transmission routes among patients with relapsing CDI. Furthermore, the importance of mixed infections in establishing transmission chains is increasingly acknowledged and the investigation of a single isolate per sample represents both a limitation of this study and an important consideration of WGS studies of transmission [139]. Nonetheless, WGS analysis provided insights into recurrent CDI epidemiology beyond that achievable by conventional PCR-based ribotyping.

The ability of WGS to rule out spurious epidemiological interpretations and resolve cryptic transmission events is a major advantage over other typing methods of lower discriminatory power. In contrast to previous WGS analysis of *C. difficile*, where < 40% of genetically identical strains had clinical evidence supporting transmission, the majority (three out of four) of our WGS-identified transmission events were substantiated by clinical data, albeit in a relatively small patient cohort [62]. This observation may highlight missed opportunities for infection control and that further intervention strategies (e.g. hand hygiene and environmental decontamination) are warranted in this vulnerable patient cohort. The confirmation of persistent infection by genetically indistinguishable strains over intervals greater than eight weeks was notable as current clinical definitions of recurrent CDI exclude such cases. Whether such protracted relapse intervals are indicative of chronic *C. difficile* colonization-infection cycles or are due to reinfection by common environmental sources is an intriguing question with implications for both CDI management and the definition of recurrent infection. The broader adoption of WGS technology in the clinical setting will undoubtedly help to address such questions.

3.3.   Conclusion

This study demonstrated the suitability of the methods used for selective isolation and culture of *C. difficile*, PCR ribotyping, whole genome sequence generation and analysis in

the context of available epidemiological data. The findings of this study illustrate the complex interconnections between clinical episodes of RCDI and nosocomial transmission in a tertiary hospital. This study was a key component of my successful application for the Health Research Board Research Training Fellowship, which supported the work presented in Chapters 4–8.

4. Investigation of a Prospective Cohort of CDI, and Genomic Comparisons with CA CDI isolates

4.1. Introduction

Clostridioides difficile infection (CDI) has until recently been considered as typically acquired in hospitalised patients. However Eyre *et al*. demonstrated a much lower rate of nosocomial transmission between symptomatic patients, 35%, by analysing *C. difficile* genomes [62]. Even then, only 38% of patients with genetically related isolates had shared time and space on the same hospital ward [62]. For patients who had no identifiable hospital-based contact, intriguingly, Eyre *et al.* identified a small proportion who were resident in the same postal code districts, or were noted to attend the same primary care practice [62].

Subsequent work revealed that 13% of Oxford CDI *C. difficile* strains were genetically related to strains associated with colonisation of local infants; these findings were consistent with recent direct/indirect transmission to/from Oxfordshire CDI cases, despite the lack of identifiable epidemiological risks [69]. However, transmission rates between symptomatic children appear to be far lower, as only 8% of children with genetically related isolates in one American study had mutual nosocomial exposures [71].

In Australia, one study reported an overlap between 79% of *C. difficile* ribotypes causing hospital and community associated infections, with exposure to antibiotics before hospital admission being a significant factor in the development of symptomatic CDI [140]. Brown *et al.* found that antibiotic use in acute care hospitals accounted for 72% of variation in CDI rates between acute care and long-term care facilities (LTCFs) [141]. McLure *et al.* have reported results of their modelling work that suggest:

1. Nosocomial transmissions alone cannot maintain endemic nosocomial CDI, without frequent admission of colonised patients [142].

2. Current definitions of CDI over-estimate hospital-acquired infection and under-estimate the role of community transmission, with consideration required regarding asymptomatic colonised adults, infants, and animal reservoirs [143, 144].

3. Reduction of antibiotic consumption within the community, especially regarding the seasonal variation and peaks, would be expected to decrease both hospital and community-acquired CDI [145].

It remains difficult to determine sources of the *C. difficile* strains that cause healthcare-associated CDI. The hospital environment is complex by nature, and it is challenging to assess the relative contributions of either potential lapses of infection control within the hospital or suboptimal antimicrobial stewardship to a patient's episode of CDI [70].

We conducted a prospective observational cohort study of *C. difficile* epidemiology within our hospital. Our primary aim was to assess the nosocomial transmission rate, based on next-generation sequence analysis of *C. difficile* isolates and integration with demographic and clinical variables. Secondary aims included the investigation of demographic and clinical variables between patients with probable nosocomial acquisition of *C. difficile* and patients who appeared to have another mode of *C. difficile* acquisition, and if the molecular epidemiology of community associated CDI resembled that of nosocomial infection.

4.2. Methods

4.2.1. Sources of C. difficile Isolates

1. Clinical isolates of *C. difficile* were collected between September 2013 and August 2016 as part of a prospective investigation in St James's Hospital, Dublin, Ireland. Stool samples submitted to the Diagnostic Microbiology laboratory were tested for *C. difficile* toxin B gene (tcdB) by EntericBio PCR kit (Serosep, Annacotty, Ireland as a diagnostic nucleic acid amplification test (NAAT). Stool samples had to

have a liquid consistency for approval of testing, and repeat samples within a 10-day interval following a positive NAAT were not processed. Positive NAAT results on faecal samples from hospital inpatients were notified by medical microbiologists to clinical and/or nursing staff as routine practice. Faecal samples were treated with ethanol shock before anaerobic incubation on cycloserine-cefoxitin egg-yolk (CCEY) medium.

2. Stocks of *C. difficile* isolates were provided to the investigators from a national surveillance study of community-acquired CDI undertaken in April–June 2015. Approximately 5–10μL of stock samples were transferred to individual CCEY plates for anaerobic incubation.

4.2.2.   C. difficile DNA Extraction and Whole Genome Sequencing

DNA was extracted from resulting 'broken glass' colonies for PCR ribotype analysis and Illumina genomic library preparation as previously described [117]. Whole genome sequencing was performed on the following benchtop sequence platforms:

- Illumina MiSeq platform, TrinSeq Genome Sequencing Laboratory, Trinity College, Dublin, Ireland

- Illumina MiniSeq platform, Dept. of Clinical Microbiology, Trinity College, Dublin, Ireland

- Illumina HiSeq platform at Nuffield Department of Clinical Medicine, John Radcliffe Hospital and University of Oxford, Oxford, United Kingdom.

4.2.3.   C. difficile Genome Sequence Analysis

High-quality variants of resulting sequence reads were identified as described previously [62, 146]. Sequences were compared using single nucleotide polymorphisms (SNPs), obtaining differences between sequences from maximum likelihood phylogenies corrected for recombination [147]. Clusters were generated, with a maximum permitted distance of 500 SNPs within sequences. Based on rates of *C. difficile* evolution and within-

host diversity, isolates which only differed by 0–2 SNPs were considered for plausible transmission events.

4.2.4. Clinical Details

Medical notes of hospital inpatients were reviewed by the investigators within 72 hours of a positive NAAT result, and at intervals for the remainder of their hospital admission. Clinical data and administrative details, including dates of admission and discharge, ward placements and admitting medical speciality were entered into a case report form, which was pseudonymised and stored in a dedicated Microsoft Excel repository.

4.2.5. Infection Prevention and Control

As the majority of hospital beds are located in 4–6 bedded bays within the hospital, patients are typically admitted to a multi-bedded bay with a shared bathroom. When a significant Infection Prevention and Control risk is identified upon admission, a single room with en suite would be prioritised. This would include infections with significant airborne and/or droplet transmission, and includes patients with symptomatic diarrhoeal illnesses pending further investigations. Otherwise, patients with suspected or confirmed CDI during their admission were moved from a multi-bedded bay to single room until at least 48 hours following cessation of diarrhoea.

4.2.6. Antimicrobial Stewardship Activity

An active antimicrobial stewardship programme was in place at the hospital, preceding and during this study, under the guidance of the Antimicrobial Stewardship Committee. Activities included at least weekly antimicrobial stewardship rounds performed by an antimicrobial stewardship pharmacist and a consultant physician in either Infectious Diseases or Clinical Microbiology. Antimicrobial prescriptions could be referred for review by any clinical pharmacist, any member of nursing staff on a ward, and/or any member of a medical team. Otherwise the antimicrobial stewardship rounds would be conducted to review antimicrobial prescriptions either in an individual ward, or of a nominated antibiotic such as meropenem or daptomycin. The activities of the Antimicrobial

Stewardship Committee were in addition to other clinical activities of both Clinical Microbiology and Infectious Diseases departments. Senior Clinical Microbiology doctors reviewed the antibiotics prescribed for all patient with positive blood cultures and other samples of note e.g. cerebrospinal fluid. Other departmental activities of relevance include daily multidisciplinary ICU rounds, and weekly multidisciplinary team discussions with Haematology, regarding inpatients with haematological malignancies at various stages of either bone marrow transplantation or stem cell transplantation. Both departments provide consultation services for any clinical query of either diagnosis or management of infections, available for any patient.

4.2.7. *Epidemiological Analysis and Interpretation*

Epidemiological links were evaluated for demonstrable associations of overlapping ward admissions and/or care provided by the same hospital personnel, for patients with *C. difficile* isolates with 0–2 SNP differences between them. Details of ward locations and relevant personnel were available for hospital inpatients, and personnel details were available for patients who had attended either an out-patient clinic or for a day-case procedure at the hospital. Where a likely cause of transmission could be ascertained, e.g. both patients were admitted to the same ward and/or under the care of the same clinical personnel, this mutual exposure was categorised as a plausible transmission event. Where no likely cause could be ascertained, yet two or more patients have genetically identical isolates, this was categorised as a cryptic transmission event. Epidemiological links are presented with a 90-day interval between identification of *C. difficile* isolates [31, 73] and without time restrictions.

Demographic and clinical details were compared for inpatients with *C. difficile* isolates of 0–2 SNPs difference, inclusive of both plausible and cryptic transmission events, to inpatients with bacterial isolates which did not have a genomic relationship to other isolates within this collection. Categorical variables were compared by chi-square test of proportions, and continuous variables by ANOVA. Statistical analysis was conducted in R studio. No epidemiological details were made available to the investigators for the Irish CA CDI isolates. *C. difficile* PCR ribotypes, sequence types (STs),

and the presence or absence of genetic similarities of 0–2 SNPs between these and any other *C. difficile* isolate of our collection was the extent of analysis performed.

This investigation was conducted with the prior approval from the St James's Hospital/Tallaght Research Ethics Committee (23/9 83/13).

4.3.   Results

There were 335 *C. difficile* isolates included in this analysis, obtained from 302 patients. These isolates belonged to 41 sequence types (STs). Most patients were admitted at the time of *C. difficile* detection in their stool samples (198/302, 66%). Requests for *C. difficile* testing of faecal samples were made on behalf of 28% of patients from their attendance at the hospital out-patient clinics, day ward facilities, or General Practice within the hospital catchment area. There were 17 (5%) samples provided as bacterial stocks from the national CA CDI study. This is shown in Figure 4.1.

Figure 4.1 Outline of patients included, according to the source of referral of positive C. difficile test
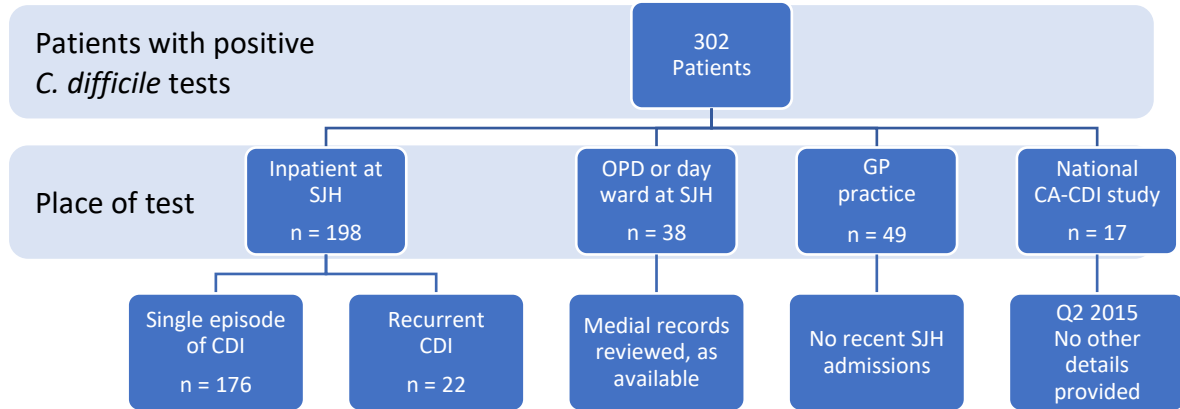
Overall, 149/335 (44%) isolates had 0–2 SNP differences to an isolate associated with another episode of CDI. This included 98 findings of 0 SNP differences, 36 findings of 1 SNP between 2 isolates and 15 findings of 2 SNPs between 2 isolates. If the criterion for related isolates was relaxed to 0–3 SNPs between isolates, there would have been an additional 6 findings of related isolates.

Of these 149 genomically related isolates, 21% were hospital isolates identified within 90 days of each other, 16% from episodes of RCD, 4% from hospital patients with clinical symptoms and positive NAAT results separated by > 90 days, and 3% from samples where the patient was not known to have any admissions at SJH. This is displayed in Table 4.1.

Table 4.1: Sources of C. difficile isolates with 0–2 SNP differences

Source of isolates	Number of isolates (%)
HO CDI episodes, with clinical onset of symptoms within 90 days	72 (21%)
RCDI episodes	53 (16%)
HO CDI episodes, with > 90 days between clinical onset of symptoms	14 (4%)
CDI episodes, without history of inpatient admissions at SJH	10 (3%)

4.3.1. Patient Demographics and Clinical Details

There were 149 isolates with genomic relationships of 0–2 SNP differences, pertaining to the clinical episodes of CDI of 109/302 (36%) individual patients. Of these, 10 patients had RCDI without any associated transmission events, and 16 patients had no known admissions at this hospital. This results in 75 inpatients with *C. difficile* isolates genomically related to the isolate of another inpatient. There were 123 inpatients, whose *C. difficile* isolates were not known to be genomically related to any other inpatient. Demographics and clinical details were compared for these two groups, i.e. 75 patients with related *C. difficile* isolates and 123 patients with unrelated *C. difficile* isolates, and presented in Table 4.2. Patients with related *C. difficile* isolates were significantly older, more likely to be admitted by Medicine for the Elderly and to be discharged to a long-term care facility after their admission. Patients with unrelated *C. difficile* isolates were significantly more likely to be under the care of other medical specialties, and to have received immunosuppressant medication. One third of this group were under the care of Haematology, Oncology or Radiation Oncology services, with a broad range of immunosuppressants. Corticosteroids were prescribed to 15% of this group, by medical specialties including Respiratory and Endocrinology teams. Corticosteroids and anti-TNF

monoclonal antibody treatments were observed in 6%, under the care of Gastroenterology and Rheumatology teams. No significant differences were found regarding exposure to proton-pump inhibitors, the mean number of inpatient antibiotic prescriptions preceding *C. difficile* detection, or the type of antibiotic prescribed. Length-of-stay (LOS) was recorded as the duration in days between the date of the index admission to hospital and the date of discharge or death. The detection of *C. difficile* from the patient's stool sample could have occurred at any time during this admission.

Table 4.2: Demographics and clinical variables for inpatients with related versus unrelated isolates

Demographics and clinical variables	Patients with related isolates	Patients with unrelated isolates	Relative Risk ratio	Significance
Number of patients	75	123		NS
Mean age (years)	74	67		0.005657
Mean LOS (days)	125	152		NS
PPI prior to admission	75%	78%	0.96	
Immunosuppressants during admission	16%	45%	0.35	
Enteral feeding during admission	23%	21%	1.1	
Admitting speciality:				
- Medicine for Elderly	35%	15%	2.3	
- Other medical specialities	42%	64%	0.66	
- Surgical specialities	20%	19%	1.05	
Mean no. of antibiotic prescriptions per inpatient, prior to *C. difficile* detection	2.9	2.9		NS
Outcome of admission:				
- Discharge home	30%	54%	.55	0.00056
- LTCF	51%	29%	1.75	0.001
- Death	19%	15%	1.26	NS

4.3.2.   Genomic Analysis of RCDI Isolates

Most RCDI (58%) occurred within the accepted 8-week interval of the initial episode [24]. There were three events where patients experienced clinical episodes compatible with RCDI but genomic analysis revealed these episodes to be re-infections. One patient had 2

ST-2 isolates, separated by 28 SNPs, another had 2 ST-77 isolates separated by 30 SNPs, and the third had 2 ST-12 isolates separated by 2,683 SNPs. A fourth patient had 2 ST-98 isolates which were distinguishable by 5 SNPs, so the genomic inference of recurrence versus re-infection could not be ascertained.

4.3.3. Potential for Transmission between Patients with Genomically Related C. difficile Isolates

Within 90 days, there were 45 occurrences of two or more isolates with only 0–2 SNP differences), and clinical details identified plausible transmission events for 25/45 (55%). These isolates belonged to fifteen *C. difficile* sequence types, and 10/15 sequence types each had only a single pair of related isolates, i.e. a single plausible transmission event. Clinical details could not account for 20/45 (45%) occurrences; these isolates belonged to 13 sequence types. Here, 7/13 sequence types each had only a single cryptic transmission event. There is significant overlap of the sequence types associated with plausible and cryptic transmission events, as detailed in Table 4.3.

Of the plausible transmission events, 14/25 (56%) related to admissions under the care of the same speciality, 7/25 (28%) related to admissions on the same hospital ward, and 4/25 (16%) to admissions under both the care of the same speciality and presence on the same ward. There were five different specialities associated with plausible transmissions. As patients may spend time in multiple hospital wards during the course of a single admission, there were eight hospital wards and a rehabilitation/convalescence facility elsewhere in Dublin associated with plausible transmissions. There were 14 transmission events within the 90-day cut-off in which at least one of the patients experienced RCDI.

Without restricting analysis to isolates of CDI episodes identified within 90 days of each other, there were an additional eight transmission events inferred by the genomic similarities between isolates. This included two plausible transmissions. One plausible transmission event related to a pair of ST-12 isolates separated by 103 days, associated with two patients who were admitted to the same ward and under the care of the same

specialty. The other event related to a pair of ST-3 isolates separated by 317 days, associated with two patients who were admitted under the care of the same speciality.

There were six additional cryptic transmission events identified by removing the 90-day restriction of analysis. These include:

1) Two ST-36 isolates, separated by 0 SNPs and 302 days. One isolate related to a nosocomial episode, and the other related to an episode within a LTCF, who had not previously been admitted to St James's Hospital.

2) Two ST-10 isolates, separated by 0 SNPs and 319 days. Both isolates related to patients attending out-patient clinics. There were no identifiable demographic or clinical factors common to these patients.

3) Two ST-11 isolates, separated by 1 SNP and 444 days. Both isolates related to nosocomial episodes, with no shared epidemiology between these patients.

4) Three ST-11 isolates, where the third isolate is separated by 2 SNPs and 577 days from the other two isolates. All three isolates were from nosocomial episodes; however, the two earlier isolates were separated by 1 SNP and 1 day, with those two patients under the care of the same speciality.

5) Two ST-48 isolates, separated by 0 SNPs and 618 days. Both isolates related to nosocomial episodes.

6) Two ST-37 isolates, separated by 0 SNPs and 824 days. Both isolates related to nosocomial episodes.

4.3.4. *Diversity of Sequence Types and Associated Epidemiology*

Table 4.3 shows the distribution of isolates by sequence type. As ST-11 was the single most common sequence type identified during this study, it was also the most common sequence type associated with RCDI, as well as plausible and cryptic transmission events. Four other sequence types, ST-2, ST-3, ST-6, and ST-8, also had associated RCDI, plausible and cryptic transmission events. There were 17/41 sequence types with no associated RCDI and/or transmission events.

4.3.5. CA CDI Isolates

The 17 CA CDI isolates belonged to 11 *C. difficile* sequence types. Two of these isolates had genomic relationships with isolates of hospital inpatients; both of these findings reflect the most prevalent *C. difficile* sequence types identified during this cohort study. There was one pair of ST-11 isolates, where the CA CDI isolate differed by 2 SNPs from a nosocomial isolate dated July 2015. There was also a pair of ST-2 isolates, of which the CA CDI isolate differed by 1 SNP from a nosocomial isolate from February 2014. Neither of these findings can be explained by the available epidemiological information. The most common *C. difficile* sequence type for the CA CDI isolates was ST-44, seen in 4/17 isolates (24%). The CA CDI isolates were associated with five sequence types without any recognised RCDI or transmission events: ST-9, ST-33, ST-41, ST-43, and ST-54. All other sequence types associated with CA CDI isolates were associated with RCDI and/or transmission events, albeit not with the isolates of CA CDI origin. The CA CDI isolates were the sole ST-41 and ST-43 isolates identified within this study.

Table 4.3: Distribution of C. difficile isolates by sequence type and association of isolates with recurrent CDI, plausible transmission events and cryptic transmission events.

ST	No. of isolates	Proportion of isolates associated with RCDI (% of no. of total isolates of ST)	Proportion of isolates with plausible transmission events (% of no. of total isolates of ST)	Proportion of isolates with cryptic transmission events (% of no. of total isolates of ST)
1	3	0	0	2 (67%)
2	39	3 (8%)	5 (13%)	5 (13%)
3	13	2 (15%)	2 (15%)	3 (23%)
4	2	2 (100%)	0	0
6	24	3 (13%)	2 (8%)	5 (21%)
8	29	2 (7%)	5 (17%)	4 (14%)
10	11	0	2 (18%)	2 (18%)
11	56	13 (23%)	9 (16%)	7 (13%)
12	11	6 (55%)	5 (45%)	0
14	12	2 (17%)	0	0
18	2	0	2 (100%)	0
35	9	0	2 (22%)	2 (22%)
36	9	0	2 (22%)	4 (44%)
37	4	0	0	2 (50%)
45	3	0	2 (67)	0
46	6	2 (33%)	2 (33%)	0
48	3	0	0	2 (67%)
49	8	2 (25%)	2 (25%)	0
55	5	0	0	2 (40%)
56	6	4 (67%)	0	2 (33%)
77	3	2 (67%)	0	0
98	3	2 (67%)	0	0
295	8	6 (75%)	7 (88%)	0
358	3	2 (67%)	2 (67%)	0
Others	63	0	0	0

4.4.   Discussion

We identified 72/335 isolates with genomic relationships based on 0–2 SNP differences, which suggests an overall transmission rate of 21% for the interval of September 2013–August 2016. This is comparable to findings published elsewhere, which includes transmission rates of 11–27% per hospital in North Wales, 2015 [73], and 19% in a Madrid hospital, 2014–2016 [74]. The rates reported in six English hospitals appear marginally lower, with reported transmission rates of 7–24% per hospital, from 70–153 isolates sequenced per hospital [70]. Eyre *et al.* calculated that once 50 isolates per hospital had been successfully sequenced, the estimated proportion of linked isolates per hospital would be relatively stable [70]. These studies have also found significantly higher rates of transmission amongst ST-1/Ribotype 027 *C. difficile* isolates [70, 73, 74]. It is uncertain how a significantly lower prevalence of ST-1/Ribotype 027 isolates would alter Eyre *et al.*'s model of stability regarding the proportion of linked isolates identified per hospital [70].

Garcia-Fernandez *et al.* identified potential transmission rates of 3% for ribotype 014/020 to 60% for ribotype 027 isolates [74]. Martin *et al.* reported an overall rate of 35% of isolates from CDI cases being within 0–2 SNPs from a prior case in Leeds, but this varied between 11% of cases associated with ST-11/ribotype 078 to 64% of ST-1/ribotype 027 [31]. When considered by sequence type, our findings of isolates with genomic relationships (0–2 SNP differences) reflect the prevalence of sequence types within our cohort. As *C. difficile* ST-11 was the single most prevalent genotype, it was associated with more RCDI, plausible and cryptic transmission events than any other ST within our analysis. Genome sequence analysis of the isolates from the European multi-centre prospective bi-annual point prevalence study of CDI (EUCLID) in hospitalised patients with diarrhoea found that ribotypes 001/072 and 027 demonstrated within-country clustering, but ribotypes 078, 015, 002, and 014/020 did not show evidence of clusters occurring within the individual European countries [147].

Regarding the potential sources of transmission that we identified in this analysis, there were more plausible transmission events relating to admissions under the care of a

shared clinical team (14/25; 56%), than findings of patients with related *C. difficile* isolates having been admitted to a single ward (7/25; 28%). This is in contrast to other studies, such as findings in North Wales, where 21/43 (48%) of their linked cases had shared ward space, and only 2/43 (5%) had overlapping admissions without any shared ward time [73]. Garcia-Fernandez *et al.* could only identify direct ward contacts for a minority of transmission recipients, where only 7 pairs of 41 linked cases had been on the same ward at the same time [74]. Although they found that 18 pairs (44%) of linked cases had no shared time within the hospital, they inferred that the re-appearance of specific *C. difficile* genotypes after intervals of months, or even a year, was indicative of (unknown) hospital reservoirs [74]. Although persuasive, a finding of 0–2 SNP differences between two *C. difficile* isolates is not irrefutable evidence of a direct transmission event between two patients [71]. McLure *et al.*'s model could not account for sustained nosocomial CDI unless admission of colonised patients was a frequent event [142]. We have previously reported an outbreak of a unique *C. difficile* sequence type, ST-295, where we identified the (nosocomial) origin of the signature mutation and the transmission events associated with a hospital ward and subsequently within a local hostel [148]. This outbreak report is presented in Chapter 7. There are several studies which have used the criterion of 0–2 SNPs between *C. difficile* isolates to consider them as genomically related: [13, 31, 62, 63, 66–74]. Here, if this criterion was relaxed to allow 3 SNPs between isolates, there would be another 5 unique findings of related isolates. Of these, 4 had no known shared epidemiology and 1 was related to patients under the care of the same clinical team albeit 247 days apart from each other.

Despite the greater mean length of stay for patients whose isolates were not genetically related to those of any other inpatient, the significant findings of older age and greater likelihood of discharge of patients with genetically related isolates to LTCFs suggests that frailty was the underlying risk for nosocomial acquisition of *C. difficile*. We did not find any evidence of differences between either the number or type of inpatient antibiotic prescriptions when comparing patients with genetically related or unrelated *C. difficile* isolates. Martin *et al.* had also found that antibiotics within either 7 days or 90

days before CDI diagnosis did not appear to influence the transmission of *C. difficile* [31]. Dingle *et al.* have demonstrated the impact of reducing community and hospital prescriptions of fluoroquinolone and cephalosporin antibiotics in England on the incidence of fluoroquinolone resistant CDI [149]. Their findings suggest that smaller studies may not be sufficient to detect the effects of antibiotic consumption trends on CDI epidemiology at a local level.

The 90-day interval used by others [31, 73] allowed for the identification of almost all plausible transmission events within our cohort. Of the two events which exceeded this interval, the ST-12 pair of isolates shared convincing epidemiological factors of shared time on a single ward, and being under the care of the same clinical specialty. We consider the link between ST-3 isolates of patients under the care of a shared speciality, separated by 317 days, to be more tenuous. Although we could not identify any other common demographic details, an exposure unrelated to the hospital could provide a more satisfactory explanation [62, 69]. We found a greater proportion of cryptic transmission events for which detection of the isolates was separated by 302–824 days. This supports McLure *et al.*'s hypothesis of frequent importation of *C. difficile* from patients exposed to community sources [142]. We have previously demonstrated a likely incubation/carriage of *C. difficile* of 16 weeks between a nosocomial transmission, and the recipient's subsequent re-admission with CDI [148].

In the prospective cohort study, most genomic recurrences (58%) occurred within the recognised interval of 8 weeks following CDI [24]. In addition, four clinical episodes which were within this 8-week interval did not meet genomic criteria for RCDI. Similarly, Garcia-Fernandez *et al.* reported that 54% of the genetically related recurrences had occurred within 8 weeks of CDI, and Martin *et al.* reported a range of 15–103 days between genetically related recurrences, in their respective studies [31, 74].

Limitations of our study include the absence of investigation of asymptomatic colonisation of inpatients and/or staff, which may have reduced our overall findings of transmission. We did not have an objective measure of frailty as a clinical variable

observed within this cohort, so we cannot conclude that this is the most important factor underlying nosocomial transmission. However, in our prospective cohort with a uniquely low prevalence of *C. difficile* ST-1/ribotype 027 isolates, we believe these findings of a nosocomial transmission rate of 21% and the molecular epidemiology of CA CDI isolates enhance our current understanding of *C. difficile* epidemiology.

5. Diagnostic and Antimicrobial Stewardship

5.1. Introduction

Clostridioides difficile is a toxin-producing bacterium, associated with colonisation and infection of the human gastrointestinal system, with significant morbidity and mortality. Although *C. difficile* infection (CDI) is typically associated with nosocomial infections, community-associated CDI (CA CDI) is increasingly recognised [21, 106].

Diagnosis of CDI is typically made by a combination of compatible clinical symptoms and laboratory detection of this bacterium and/or its toxin [21, 106]. In recent years, nucleic acid amplification tests (NAATs) have been introduced to aid laboratory identification with a sensitivity of 73–100% for diagnosis of CDI. NAATs require less 'hands-on' time for laboratory staff, and results are available more quickly than enzyme immunoassays (EIAs) [150, 151], but NAATs cannot distinguish colonisation from infection. In an early review, Brecher *et al.* highlighted the importance of testing samples only from patients with clinically significant diarrhoea and who have risk factors for CDI [151]. They predicted that false positive diagnosis of CDI based on NAATs used alone would lead to unnecessary antibiotic prescriptions [151]. EIAs had been associated with poor sensitivity causing missed infections, but it soon became apparent that the use of NAATs could lead to significantly increased rates of *C. difficile* detection without meeting criteria for CDI [151, 152]. In one study, rates of reported CDI increased by 43–67% in three American states, despite accepting fewer stool samples for testing [152]. Stable rates of CDI identification by their control laboratories, which did not change to NAATs, suggested that study findings were reflective of the altered methodology, rather than a widespread change of regional epidemiology [152]. The authors commented that further research would be required to understand the clinical impact of NAATs on CDI epidemiology and management [152]. Davies *et al.* also found that significant differences in the rates of CDI diagnosis between European hospitals were attributable to the laboratory methods used, even with comparable rates of testing [153]. There was a 3.5-fold increase in diagnosis for hospitals using stand-alone NAATs [153]. Truong *et al.* found similar rates of positive NAATs for patients with either asymptomatic colonisation (11.8%)

or clinically compatible symptoms of CDI (15.4%) [154]. They found no significant differences between either the median test cycle threshold (CT) values or demographic and clinical variables to separate colonisation vs. infection [154]. In a further study, Truong *et al.* showed how their interventions, which targeted 'over-diagnosis', led to reduced rates of hospital onset CDI (HO CDI) diagnosis, and consequent reduction of therapeutic oral vancomycin [155].

In this study we describe the demographics and clinical variables of hospitalised patients with *C. difficile* colonisation and infection, from a prospective cohort after NAATs had been introduced in our laboratory as the sole diagnostic test for CDI. Our primary aim was to reconcile the patient's clinical status with the positive NAAT result as one of *C. difficile* colonisation, infection, or recurrence. The *C. difficile*-related medication prescribed for each patient was then reviewed to assess if it was appropriate or not for their clinical status.

5.2.   Methods

Stool samples submitted to the Diagnostic Microbiology laboratory of an Irish tertiary referral hospital between September 2013 and August 2016 were tested for *C. difficile* toxin B gene (*tcdB*) by EntericBio PCR kit (Serosep, Annacotty, Ireland). Stool samples had to have a liquid consistency for approval of testing, and repeat samples within a 10-day interval of a positive NAAT were not processed. Medical microbiologists notified positive NAAT results of hospital inpatients to clinical and/or nursing staff as routine practice. Medical notes were reviewed by the investigators of this cohort study. Relevant clinical data were entered into a case report form, which was pseudonymised and stored in a dedicated Microsoft Excel repository. Case notes were reviewed within 72 hours of a positive NAAT, and at intervals for the remainder of the hospital admission. This investigation was conducted with prior approval from the hospital's research ethics committee. Clinical cases and data were recorded between 1st September 2013 and 31st August 2016. The analysis of *C. difficile* isolates pertaining to this cohort is described in Chapter 3.

There were three clinical classifications in response to a positive NAAT:

1) *C. difficile* infection (CDI), where the patient had at least 3 unformed stools documented in one 24-hour period, and no other causes of diarrhoea were identified.

2) *C. difficile* recurrence (RCDI), where the patient had resolution of symptoms from an earlier episode, with subsequent development of symptoms meeting criteria of CDI.

3) *C. difficile* colonisation, where the patient did not experience 3 unformed stools within a 24-hour period, with identification of confounding causes of diarrhoea, e.g. concurrent laxative therapy.

CDI was categorised as severe with any one or more of these criteria [11]:

- **Clinical:** fevers, rigors, abdominal pain
- **Laboratory:** Leukocytosis > 15 × 109/L or rise of serum creatinine > 50% baseline, or serum creatinine > 133μmol/L
- **Endoscopy**: Detection of pseudomembranous colitis by visualisation at sigmoidoscopy
- **Radiology:** computed tomography (CT) evidence of colitis or ascites, when a CT had been requested by clinical personnel overseeing that patient's medical care

Inpatient antibiotic prescriptions that preceded the positive NAAT report were recorded, and assessed if the antibiotic therapy prescribed was in accordance with either the contemporary hospital empiric guidelines and/or had been recommended by an infection specialist. This could be either a consultant Clinical Microbiologist or Infectious Diseases physician. Antibiotic prescriptions that did not fit with the hospital empiric guidelines and were not recommended by an infection specialist were Inpatient antibiotic prescriptions that preceded the positive NAAT report were recorded, and assessed if the antibiotic therapy prescribed was in accordance with either the contemporary hospital empiric guidelines and/or had been recommended by an infection specialist. This could

be either a consultant Clinical Microbiologist or Infectious Diseases physician. Antibiotic prescriptions that did not fit with the hospital empiric guidelines and were not recommended by an infection specialist were classified as 'gap' antibiotics. Antibiotics were also compared to the overall antibiotic consumption data for the hospital for the same period as the study, as reported to the Health Protection & Surveillance Centre, Ireland.

Antibiotics used for management of CDI were also recorded and assessed as to whether they were in keeping with the hospital empiric guidelines and/or recommended by an infection specialist. These prescriptions were also assessed if they were or were not appropriate for the patient's clinical status. For those considered inappropriate, prescriptions recommended for mild-moderate CDI but provided to patients with features of severe CDI were recorded as antibiotics for a lesser clinical status. Prescriptions recommended for severe CDI but provided to patients with features of either mild-moderate CDI or who did not meet clinical criteria for CDI were recorded as antibiotics in excess of clinical status.

Categorical variables were analysed by chi-square test of proportions, with p-value < .01 required for significance. Statistical tests were conducted in R studio.

5.3. Results

There were 381 positive NAATs for 283 hospital inpatients, with clinical notes available for review. Of these results, 299 were the first positive NAAT for an individual patient, and 82 were at least the second positive NAAT. Clinical classification as CDI, RCDI or colonisation, as outlined in Table 5.1, resulted in 181 cases of CDI, 53 recurrences and 147 episodes of colonisation. These classifications have a significant impact on the derived CDI recurrence rates. If recurrence is only considered based on the number of patients who have at least 2 positive NAATs, then the recurrence is rate is 21% (82/381). However, with re-classification of each positive test result with clinical criteria, the recurrence rate is significantly lower, 14% (53/381; p value .005).

Table 5.1: Classification of positive NAATs

Classification	Sub-classification	Number
Infection	Initial infection	171
	Infection after colonisation	10
Recurrence	First recurrence	33
	Second recurrence	13
	Third recurrence	3
	Fourth recurrence	2
	Fifth recurrence	2
Colonisation	Initial colonisation	128
	Colonisation after infection	15
	Colonisation after recurrence	4

Patient demographics are presented in Table 5.2, There were significantly more females with colonisation than with CDI, 66 vs. 51% (p = .008), and although not statistically significant, more RCDI than CDI, 60 vs. 51%. Patients with *C. difficile* colonisation were older than patients with either CDI or RCDI, with mean age of 73.3 years vs 68.7 and 67.6 years respectively. Exposure to antibiotics and proton-pump inhibitors preceding hospital admission was common; this prevalence was significantly increased in CDI and RCDI cases compared to those with colonisation. Hospital admissions were common: 65 and 66% of patients with CDI or *C. difficile* colonisation respectively, and 97% of patients with RCDI had prior admissions within three months preceding the positive NAAT. Immunosuppressive medications were more common, and enteral feeding was more prevalent for CDI vs. colonisation. The overwhelming majority of patients, 86–89%, had additional inpatient antibiotic exposure prior to the detection of *C. difficile*. There was no significant difference in the inpatient antibiotic exposure between groups, for the overall prevalence, or the proportions of antibiotic prescriptions either compliant with empiric guidelines, recommended by an infection specialist, or the gap antibiotics.

Table 5.2: Demographics and risk factors for inpatients with C. difficile colonisation, CDI and RCDI

Demographics & Risk Factors	Colonisation (n = 147)	Infection (n = 181)	Recurrence (n = 53)	p value
Female (%)	66%	51%	60%	0.008
Mean age (years)	73.3	68.7	67.6	0.02 (NS)
Preceding antibiotics (%)	26%	40%	66%	< 0.001
PPI (%)	54%	77%	85%	< 0.001
Immunosuppression (%)	29%	41%	40%	NS
Enteral feeding (%)	9.1%	21%	21%	0.01
Prior admissions (%)	66%	65%	97%	< 0.001
Proportion who received antibiotics as inpatient, preceding *C. difficile* (%)	86%	89%	89%	NS
Total antibiotic prescriptions (n)	355	539	172	
Guideline-concordant antibiotics (n; %)	135; 41%	177; 35%	49; 39%	NS
Recommended antibiotics (n; %)	103; 32%	152; 30%	33; 26%	NS
'Gap' antibiotics (n; %)	111; 34%	202; 39%	56; 44%	NS

However, significant differences were identified between the preceding antibiotic exposure of patients with a positive *C. difficile* NAAT and the overall inpatient population. Table 5.3 shows the proportion of prescriptions accounted for by different antibiotics within the different patient groups. Patients with associated *C. difficile* had significantly greater exposure to piperacillin/tazobactam and IV vancomycin, but significantly less exposure to co-amoxiclav and other agents. There was no significant differences in their exposure to carbapenems or fluoroquinolones compared to the overall hospital consumption.

Table 5.3: Proportion (%) of antibiotic agents of all prescriptions for patients with colonisation, CDI, RCDI and total inpatient hospital consumption

Antibiotic	% Prescriptions in colonisation cases (n = 326)	% Prescriptions in CDI (n = 530)	% Prescriptions in RCDI (n = 143)	% of all hospital prescriptions	Significance
Co-amoxiclav	16.6	15.5	18.9	23.7	< 0.001
Piperacillin/tazobactam	25	27.3	22.4	10	< 0.001
Carbapenems	8.2	9.2	9.1	8.7	NS
Fluoroquinolones	6.2	6.3	5.6	8	NS
IV vancomycin	7.4	8.7	10.5	4.6	0.001
Other agents	36.5	33	33.5	45	< 0.001

5.3.1. Clinical Features

Clinical features considered as markers of CDI are shown in Table 5.4. Fever, abdominal pain, altered white cell count (WCC), and elevated creatinine were all significantly more common in patients with CDI or RCDI, versus those with colonisation. Complications of CDI, including pseudomembranous colitis (PMC), hypotension and ileus were rare within this cohort. No patient required surgical intervention, and intensive care unit (ICU) admissions were infrequent.

Complications mostly occurred in patients with their first episode of CDI. Criteria for severe CDI were met in 45.6% of CDI and 60% of RCDI episodes. Although not meeting criteria for CDI, significantly fewer patients with colonisation had clinical and/or laboratory criteria which were interpreted by clinical staff as 'severe CDI'. There were fewer complications of hypotension and ileus seen in episodes of colonisation vs. CDI or RCDI.

Table 5.4: Clinical features and complications observed in patients with colonisation, CDI and RCDI

Clinical Features	Colonisation (n = 147)	Infection (n = 181)	Recurrence (n = 53)	p value
Fever (%)	3%	12%	6%	0.008
Abdominal pain (%)	12.5%	26%	21%	0.008
Altered WCC (%)	11%	26%	30%	< 0.001
Creatinine > 133 (%)	6.3%	15.2%	9%	0.025 (NS)
Features compatible with severe CDI (%)	11%	45.6%	60%	< 0.001
Abnormal radiology of large bowel (% of available results)	4.6%	14.6%	9%	0.01
PMC (n)	0	3	0	NS
Hypotension (%)	0.7%	7%	7%	0.01
Ileus (%)	0.7%	6%	7%	0.02 (NS)
ICU admission (%)	0%	3%	0.3%	NS

5.3.2. CDI Treatment

Table 5.5 shows *C. difficile* related treatment rates observed within this cohort. Overall, 61.8% of patients who did not have symptoms meeting the threshold for clinical CDI received *C. difficile*-related antibiotics. Oral metronidazole accounted for 70% of these prescriptions. Other prescriptions for single antibiotics accounted for 14.4%, and combination prescriptions for 15.6%.

Table 5.5: C. difficile related antibiotic prescriptions for patients with colonisation, CDI or RCDI

CDI Treatment	Colonisation (n = 147)	Infection (n = 181)	Recurrence (n = 53)	p value
Proportion who received any CDI-related treatment (%)	61.8%	86.4%	94%	< 0.001
Proportion who received appropriate treatment (%)	37.5%	53.8%	72%	< 0.001
Proportion who received antibiotic agents in excess of clinical status (%)	62.5%	22.5%	22%	< 0.001
Proportion who received antibiotic agents of a lesser clinical status (%)	N/A	24%	5%	0.001
PO Metronidazole only, as proportion of CDI-related prescriptions (%)	46.8%	49.1%	21%	< 0.001
PO Vancomycin only, as proportion of CDI-related prescriptions (%)	4%	5%	25%	< 0.001
PO Fidaxomicin, as proportion of CDI-related prescriptions (%)	4%	6%	15%	0.02 (NS)
PO Vancomycin & IV metronidazole, as proportion of CDI-related prescriptions (%)	4.7%	15.8%	27%	< 0.001
PO Vancomycin & PO metronidazole, as proportion of CDI-related prescriptions (%)	2.3%	10.5%	6%	< 0.01
Proportion of patients meeting criteria for severe CDI (%)	11%	45.6%	60%	< 0.001
Proportion of prescriptions for severe, complicated illness (%)	7%	26.3%	33%	< 0.001
Proportion of appropriate treatment for severe, completed illness	NA	32%	72%	< 0.001

Patients with a first episode of CDI received oral metronidazole as the most common therapy (49.1%). Significantly more prescriptions were compliant with hospital and/or national guidelines for patients with RCDI vs. initial CDI (72% vs. 53.8%, respectively). For the latter group, non-compliant prescriptions were almost equally likely to provide either 'over-treatment' or 'under-treatment'. Antibiotic management for severe CDI was also significantly better for patients with RCDI vs. initial infection (72% vs. 32%). The most common deviation was the prescription of oral metronidazole for severe

CDI. Overall, 35% of prescriptions deemed incongruent with clinical status comprised of combination therapy with 2–3 antibiotics.

5.4. Discussion

Overall, 38% of patients in this study with a positive *C. difficile* NAAT did not have symptoms meeting the criteria of clinical CDI [21]. The most common reason for not meeting the criteria was that patients only had 1–2 episodes of loose stool, 46%. Otherwise, 39% received laxatives and 15 had another cause of gastrointestinal symptoms. Similar findings have been reported elsewhere. For example, Kelly *et al.* considered only 19.6% of tests to have been ordered appropriately [156]. They also demonstrated the contribution of these inappropriate tests to the standardised infection rates for their hospital [156]. More recently, Ilies *et al.* have reported the impact of laboratory switch to NAATs from EIAs accounting for an average increase of 75% in reported CDI rates, across 47 hospitals over 9 years [157].

Truong *et al.* reported similar demographic and clinical risk factors between their symptomatic and colonised patients, with younger age of the former (median 57 years vs. 66 years) [154]. As the NAAT CT values were comparable between symptomatic and colonised patients in their study, this suggests the absolute CT value is unlikely to be a reliable means of discrimination [154]. Dubberke *et al.* assessed the performance and utility of 8 different diagnostic assays, and their interpretation, together with consideration of patients' clinical status. They found that the positive predictive value of NAATs was < 50% in a population where 55% of their patients did not have clinically significant diarrhoea [158].

Truong *et al.* identified greater antibiotic exposure as the only significantly different variable between symptomatic and colonised patients (87% vs. 53.5%) [154]. CDI is a recognised complication of antibiotic exposure, and reduction of CDI rate is often cited as a benefit of better antimicrobial stewardship [159]. We found significantly greater prior exposure to antibiotics preceding hospital admission for patients with *C. difficile* colonisation, infection, and recurrence (26%, 40% and 66%, respectively). Certain

antibiotics have been associated much greater risks of CA CDI: one meta-analysis identified an overall odds ratio (OR) of 6.91 for all antibiotic exposures, with range of 1.84–20.83 as the ORs of antibiotic classes with increased risk [160]. Dingle *et al.* have reported the impact of reducing community and hospital prescriptions on the national incidence of CDI in England between 2007–2013, especially for infections caused by *C. difficile* ribotype 027 [149]. Although the greater prevalence of recent hospital admissions seen for the patients with RCDI may be a correlate of their history of CDI, it is still of concern that they were most likely to receive antibiotic prescriptions before their admissions with RCDI. It is possible that these prescriptions may reflect a greater concern for hospital-acquired infections by their general practitioners.

By contrast, we found that antibiotic exposure during hospital admission was comparable between groups (86%, 89% and 89%). Of these prescriptions, 35–44% did not follow hospital guidelines and were not specifically recommended for by an infection specialist. This is comparable to other studies where 20–40% of antibiotic prescriptions have been deemed inappropriate or unnecessary [159]. Although the prevalence of guideline concordant prescriptions was similar between groups in our study, the trends of fewer recommendations by infection specialists and more gap prescriptions for RCDI are of concern. Watson *et al.* also found no improvement of antibiotic prescriptions for patients after their first episode of CDI compared to their earlier prescriptions [161]. It is difficult to predict the effects of antimicrobial stewardship on overall *C. difficile* epidemiology [162]. In our study, patients with positive *C. difficile* NAATs had lower exposure to fluoroquinolones, and significantly greater exposure to piperacillin/tazobactam and IV vancomycin, compared to overall hospital consumption.

There is credible evidence for antibiotics and PPIs as contributing factors to the risk of RCDI within this prospective cohort study, by the gradients or 'dose-effects' observed. A plausible mechanism would be the effects of these medications on the microbiome of each patient. There is literature to support PPIs both as a risk factor for CDI and as causative of detrimental changes to the microbiome, and the deleterious effects of antibiotics on the microbiome are generally accepted [7, 44, 54, 58, 163]. An

exploration of the microbiome associated with RCDI in this cohort will be presented in Chapter 6.

The clinical features suggestive of severe CDI are not unique to CDI. There is still neither a single specific test or a prospectively validated severity score to identify severe CDI cases [106]. However, we found significantly more of these inpatients with criteria meeting CDI, namely 45–60% of patients with CDI or RCDI, vs. 11% of patients with colonisation. No CDI complications occurred in patients with colonisation.

There is scope for targeted antimicrobial stewardship for inpatients with positive *C. difficile* NAATs within our hospital. We found antibiotic prescriptions were far more likely to be appropriate for patients with first or recurrent CDI, dur to the high prevalence of prescriptions for patients with *C. difficile* colonisation only. Other common issues at odds with hospital, national and international guidelines included the prescription of oral metronidazole for patients with criteria for severe CDI [11]. There was a significant discrepancy between the overall proportion of patients who received the antibiotic combination of vancomycin and IV metronidazole and the proportion who met the criteria of severe CDI. We identified significantly more use of vancomycin and fidaxomicin for recurrent CDI, as deemed appropriate by national guidelines [11]. Peterson *et al.* had found that 39% of their patients did not meet minimum symptom criteria and excluded those samples from their subsequent analysis of NAAT accuracy [164]. Fabre *et al.* described the impact of a *C. difficile* action team, with review of individual cases and provision of specific recommendations during intervention periods [165]. Only 43% of their recommendations were accepted, and there was no reduction in treatment of patients with colonisation [165].

We have described issues relating to diagnostic and antimicrobial stewardship for *C. difficile* in a tertiary referral hospital. Inappropriate NAAT requesting and/or interpretation is not unique to our hospital, and reflects the challenges of contemporary medical practice. As this was not a prospective or systematic investigation of *C. difficile* colonisation, the analysis and interpretation has been restricted to the available

information. The overall prevalence of colonisation and progression to CDI cannot be inferred. However, our findings were consistent across the 3-year period, and it is likely to present a reasonable reflection of practices within our hospital. We believe our findings have international relevance by identifying interventions that can improve future practice, and identify how reporting of CDI rates can be made more accurate.

6. Microbiome Analysis in Patients with CDI

6.1. Introduction

The human gut microbiota is estimated to comprise 15,000 to 36,000 different microbes, which have mutualistic functions with both each other and their human host, and of which only a minority can be detected by culture in the laboratory [108]. Identification of bacterial composition, without requirement for bacterial culture, can be achieved by sequencing 16S ribosomal rDNA amplicons. High-throughput sequencing technologies, including the Illumina platforms, have allowed for greater yields from individual samples and parallel sequencing of multiple samples, thereby facilitating more complex projects than Sanger sequencing [166]. This is exemplified by the work of the Human Microbiome consortium, which proposed the concept of the 'core microbiome', anticipated as present at a given site in all humans, but subject to influence of human host factors and exposures [167].

The 16S rRNA gene is highly conserved among bacterial species, and has been used successfully as a surrogate marker to identify microbial phyla within a community [104]. The 16S gene sequences are typically classified into Operational Taxonomic Units (OTUs) where sequences with ≥97% nucleotide identity can be assigned to a single species [104]. This technique allows for the identification of bacterial species present in a sample, although it cannot offer the functional output attained by sequencing mRNA (metatranscriptomics), quantification of proteome (proteomics) or active metabolites (metabolomics) [104].

A 'typical' intestinal microbiota is considered to be dominated by obligate anaerobes, of phyla Bacteroidetes, Firmicutes and Actinobacteria, and facultative anaerobes of the phylum Proteobacteria [104]. The microbiota composition and structure show significant variance, both between individuals, and at different sites within the gastro-intestinal tract of a person [104]. Such diversity does not easily fit an intuitive categorisation, and early concepts of 'enterotypes' from different studies have been inconsistent [168]. However, the concept of linking patterns of microbial composition

with function is both biologically plausible and clinically relevant [168]. For example, murine models of intestinal infection with *Salmonella Typhimurium* and *Citrobacter rodentium* either precipitated or stabilised the dysbiosis associated with these infections [104].

First identified from faecal samples of infants, *C. difficile* is known to be a coloniser in early childhood; associated taxonomic findings include *Ruminococcus gnavus* and *Klebsiella pneumoniae* with colonisation, and *Bifidobacterium longum* with resistance to colonisation [169]. Disruption of the healthy, functioning gut microbiota, or 'dysbiosis', is a pivotal step leading to the development of symptomatic CDI, as illustrated in Figure 6.1. An inpatient case-control study found that hospital-acquired diarrhoeal samples had 2-fold lower diversity than those of the non-diarrhoeal controls, but were not significantly different from each other [170]. This was true of CDI-associated diarrhoea and diarrhoeal controls, and when tested by either the Shannon or inverse Simpson indices as measures of bacterial diversity [170]. Other studies have also demonstrated some reduction of taxonomic diversity and species membership of the microbiota in association with CDI [104]. This could be the effect of the absence of particular commensal species, alteration of mutualistic bacterial interactions, or a combination of these factors [105]. For example, antibiotic use increases the availability of mucosal carbohydrates normally consumed by commensal *Bacteroides* species[104]. However, findings pertaining to particular species cannot always be replicated between studies, due to context-dependent interactions and minor environmental changes [171].

Figure 6.1: Differences between diverse microbiota (a) and simplified microbiota (b) in the pathogenesis of gastrointestinal infection [104]
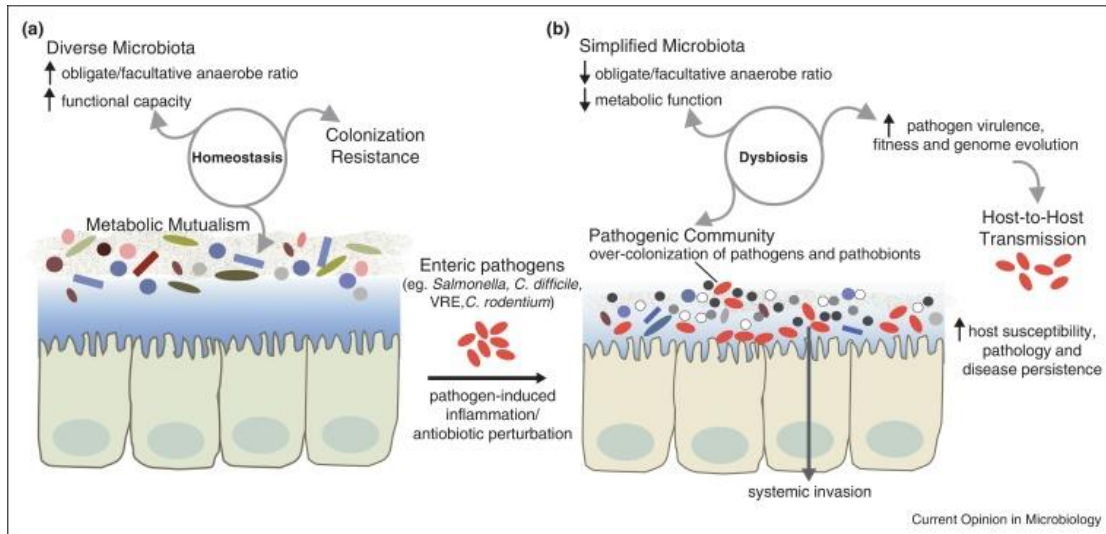
In one study, Rea *et al.* performed high-throughput amplicon sequencing of subjects' microbiota, and found fewer identifiable bacterial taxa in patients with active CDI compared to those with asymptomatic colonisation [108]. This finding withstood the broad-spectrum antibiotic use in 50% of the latter group [108]. Their additional observations included significant differences at bacterial family level between *C. difficile* culture-positive and culture-negative participants; those with *C. difficile* had a relative decrease in *Enterococcaceae* and an increase in *Lactobacillaceae* and *Enterobacteriaceae* [108].

Recurrence of CDI, where patients experience a relapse of clinical symptoms within weeks of their initial episode despite the resolution of earlier symptoms, is reported to affect approximately 20–40% of patients with CDI [107]. Recurrence is considered to reflect ongoing dysbiosis after the initial episode of CDI; the intervention of providing faecal microbiota transplantation is intended to correct this dysbiosis, with clinical success rates of approximately 90% according to one meta-analysis [107]. Chang *et al.* found different microbiome profiles in a group of patients with recurrent CDI, compared to those with incident CDI or healthy controls [33]. In their exploratory study of seven patients, they found that the controls and the patients with incident CDI had microbiome profiles with predominant Bacteroidetes and Firmicutes, but those with

recurrent CDI had consistently lower species richness, with Proteobacteria or Verrucomicrobia as dominant phyla [33]. They postulated that reduced phylotype richness during a person's first episode of CDI may act as a marker of their risk of recurrence [33]. Seekatz *et al.* compared the index faecal samples of patients who did or did not develop recurrence of CDI, and found differences between the bacterial diversity and community structure of these samples [172]. The index samples had a non-significant trend towards lower diversity for patients who subsequently developed recurrent CDI, compared to samples from patients who had no such recurrence [172].

The aim of this work is to explore the intestinal microbiome profiles pertaining to CDI in a prospective clinical cohort, to investigate the differences of bacterial richness, diversity, or the composition of individual bacterial species between patients with a single episode of CDI and patients with recurrent CDI.

6.2. Methods

Initial plans were made for patient recruitment, and longitudinal sampling, and Ethics approval was granted for this. However, the initial exclusion criteria proved prohibitive to patient recruitment, and there would have been insufficient samples for any meaningful analysis or interpretation. Table 6.1 shows the breakdown of exclusion criteria, from 1st December 2015 – 31st March 2016:

Table 6.1: Exclusion criteria of the first proposal for patient recruitment for microbiome analysis

Exclusion criteria	Number of patients
Malignancy/ immunosuppression	19
Difficulty accessing patient (intubation, hospital discharge/ transfer etc.)	17
Cognitive impairment	13
Clinical condition not meeting clinical CDI criteria	10
GI conditions/surgery	6
Blood-borne viral infection as co-morbidity	1

Due to these difficulties, and the consideration of a more recent publication by Seekatz *et al.* [172], whereby the remains of diagnostic faecal samples were retrieved and used for microbiome analysis, as well as patient recruitment for longitudinal sampling by research staff, an updated proposal was submitted to the Research Ethics Committee. This proposal was approved, and enabled the remaining portion of diagnostic samples which had a positive result for *C. difficile* toxin B gene to be used for microbiome analysis, with all patients considered eligible for inclusion. Targets for groups, e.g. recurrent versus incident CDI, were set at minimum of 16 patients per group, following discussions at APC Microbiome Unit, UCC, regarding sample sizes needed for statistical power. Demographics and clinical details were obtained from patients' medical notes, as outlined in Chapters 2 – 4, with classification of the clinical episode of CDI as either first episode or recurrence, and severe CDI or not. Only samples pertaining to clinical episodes consistent with CDI were considered for 16S microbiome analysis.

Following retrieval from the Enterics section of the Clinical Microbiology laboratory, St James's Hospital, faecal samples were stored at -80°C, and DNA extraction performed on batches of 5–12 samples. DNA extraction of 8 samples was repeated at least once, with maximum of three replicates (see Table 6.2). The replicate extractions were performed on different months, to assess the stability of DNA in storage at -80°C.

Protocols from Prof. Paul O'Toole's group in the APC Microbiome Unit, University College Cork were followed for DNA extraction, sequencing and analysis of sequences generated:

- 16S DNA extraction by mechanical bead beating and Qiagen DNeasy Blood & Tissue Kit,

- 16S amplicon PCR generation,

- Illumina library preparation with MiSeq reagent kit v3 (600 cycle).

These protocols are described in greater detail in Chapter 2. Sequence generation of 16S amplicons was performed on the TrinSeq MiSeq benchtop sequencing instrument,

with 5 samples sequenced as an internal proof of concept, 8/7/2016, and the others were completed by a second sequencing run, 4/4/2017.

DNA pipeline analytic steps were completed in the RStudio statistical environment:

1. Phase 1: Read filters, and de-multiplexing, to allow identification of results by the (faecal) sample identification.

2. Phase 2: Chimera filtering and cluster generation
Sequences are initially sorted by the length of the sequence, and then by decreasing abundance. This allows operational taxonomic unit (OTU) clustering with chimera filtering by database and matching of OTUs to corresponding read of the sample.

3. Phase 3: Microbial diversity calculations are conducted with rarefraction of the OTU table, and establishment of the selected depth of rarefraction. These include measures of alpha and beta diversity, and differential abundance tests of individual bacterial species. Adjusted p-values were calculated by use of the Benjamini and Hochberg method.

6.2.1. Alpha Diversity

Alpha diversity reflects the number (richness) and distribution (evenness) of taxa expected within a single population [173]. Before calculating measures of alpha diversity, samples were tested for normal distribution. Normality was tested by histogram, boxplot, normal Q-Q plot, and standard deviation from a normal distribution for the Shannon index, chao-1 index, Simpson index and the observed species counts. These measures of alpha diversity were subsequently calculated in the R statistical computing environment.

6.2.2. Beta Diversity

Beta diversity measures assess the difference of microbial abundances or communities from different samples. These include the Bray Curtis index, which compares abundance or OTU read count data, and UniFrac Principal Coordinate Analysis (PCoA) which

compares sequence distances between the OTUs, as either unweighted UniFrac PCoA with sole use of the sequence distances, or weighted UniFrac PCoA, which incorporates the abundance of OTUs in addition to the sequence distances.

6.3. Results

There were 72 individual samples accessed for 16S DNA extraction, and an additional DNA sample was provided by Prof. Paul O'Toole's group (APC Microbiome Institute, University College Cork) for comparison. Four extractions failed to yield sufficient DNA for inclusion in further sequencing. Four of the individual samples were used as internal controls, with replication of DNA extraction on 1 – 3 occasions. Details of samples are provided in Table 6.2.

Table 6.2: Outline of sample types and numbers for 16S amplicon sequence generation

Samples	Number of samples attempted	Comment	Number of samples used for amplicon sequencing and analysis
Molecular grade water	2	Negative control	2
DNA extract (APC Microbiome unit, UCC)	1	Positive control: sample provided without any associated metadata	1
Single DNA extraction	72	3 samples had inadequate DNA concentrations	69
DNA extracted on a second occasion	4	1 replicate had an inadequate DNA concentration	3
DNA extracted on 3 occasions	1		3
DNA extracted on 4 occasions	2		8
Total number of samples	90		86

The negative controls were successful, i.e. there was no amplicon generation or sequence information generated. Thirteen samples failed to yield sufficient depth of coverage for FASTQ files from 4/4/2017; these samples could not be used in downstream analysis. The positive control was compared to previous results available within the APC

unit, and confirmed to have similar findings. This sample was removed from subsequent analysis of CDI-associated microbiome. Overall 81% of the 16S samples with extracted 16S rRNA were used in the final analysis.

Regarding the samples which could not be included, these do not appear to have skewed the final analysis which was conducted. Table 6.3 shows key descriptors of samples included and excluded (at any stage) in the analysis. There were no significant differences identified regarding age or proportion of females, severe or recurrent CDI, by unpaired t-test and Chi-square tests respectively.

Table 6.3: Comparison of samples included in the bioinformatic analysis with those excluded, by details of the source patient

Descriptor	Samples included	Samples excluded	Statistical significance
Number of samples	70	21	
Mean age of patient (years)	64.9	64.8	None
% Female	54	58	None
% Severe CDI	28	26	None
% Recurrent CDI	40	37	None

The samples for which analysis was completed included 41 samples from patients with a single episode of CDI, and 29 from patients with recurrent CDI. Clinical notes indicated that 20 patients had features of severe CDI, and 50 did not. There were 38 samples from female patients and 32 samples from male patients.

Table 6.4 shows the distribution of the number of observed species counts for samples classified according to the patient's clinical status of either having a single episode of CDI or an episode of recurrent CDI. There are no significant differences between these distributions, despite the median numbers suggesting a larger number of observed species in samples pertaining to a single episode of CDI.

Table 6.4: Distribution of the observed bacterial species counts, according to the clinical status of CDI versus RCDI

Classification of sample by patient's clinical status	Single episode of CDI (n = 41)	Recurrent CDI (n = 29)
Minimum count of observed bacterial species per sample (no)	36	5
1st quartile (no)	61	63.5
Median count of observed bacterial species per sample (no)	91	73
Mean count of observed bacterial species per sample (no)	92.08	90.07
3rd quartile (no)	111	112.5
Maximum count of observed bacterial species per sample (no)	187	277

6.3.1. Alpha Diversity

Normality of samples was tested by histogram, boxplot, normal Q-Q plot, and standard deviation from a normal distribution for the Shannon index, chao-1 index, Simpson index and the observed species counts. These are shown in Figure 6.2.

Figure 6.2: Normality plots for measures of alpha diversity
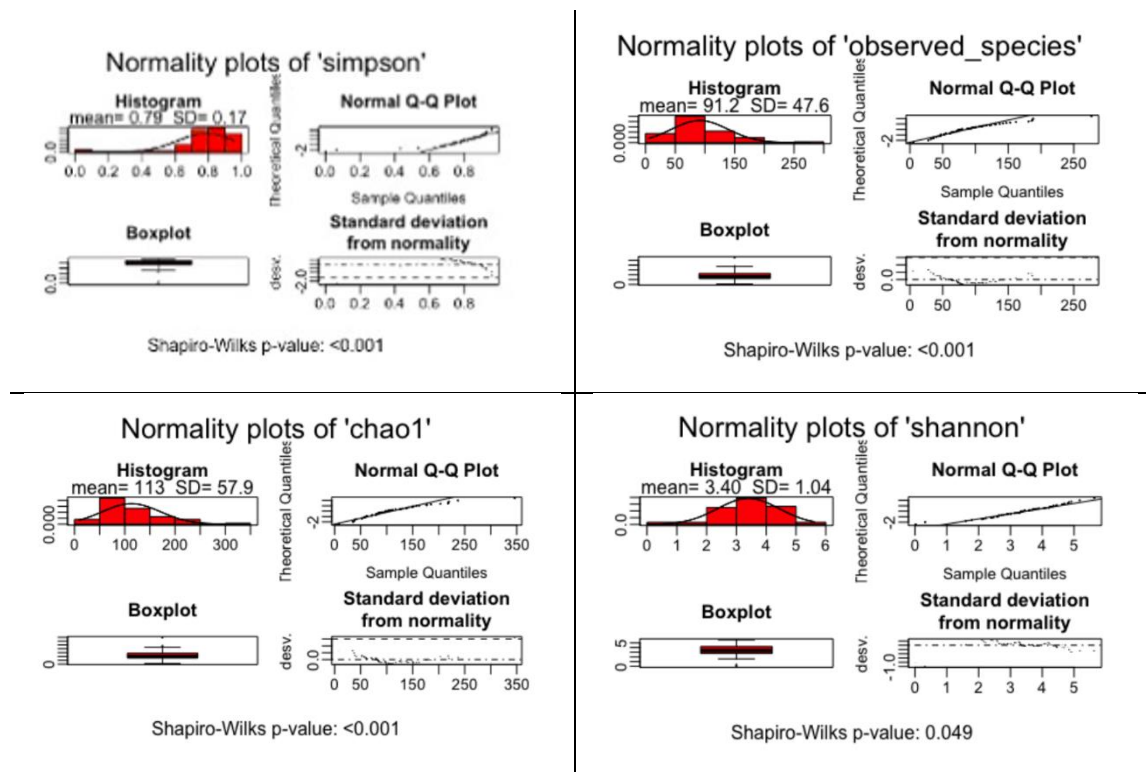
Alpha diversity was assessed for the population, according to patient status as having recurrent CDI ("Recurrence – Yes") or incident CDI ("Recurrence – No"). Results are shown in Figure 6.3. No significant findings were identified by this analysis. However, non-significant trends include lower observed species count in recurrent CDI, and lower Chao-1 and phylogenetic diversity (PD_whole_tree) indices. Groups appear to have comparable Shannon index values, and the recurrent group appeared to have a greater Simpson index than those with incident CDI.

Figure 6.3: Plots of alpha diversity of samples representing CDI ('N') or RCDI ('Y')
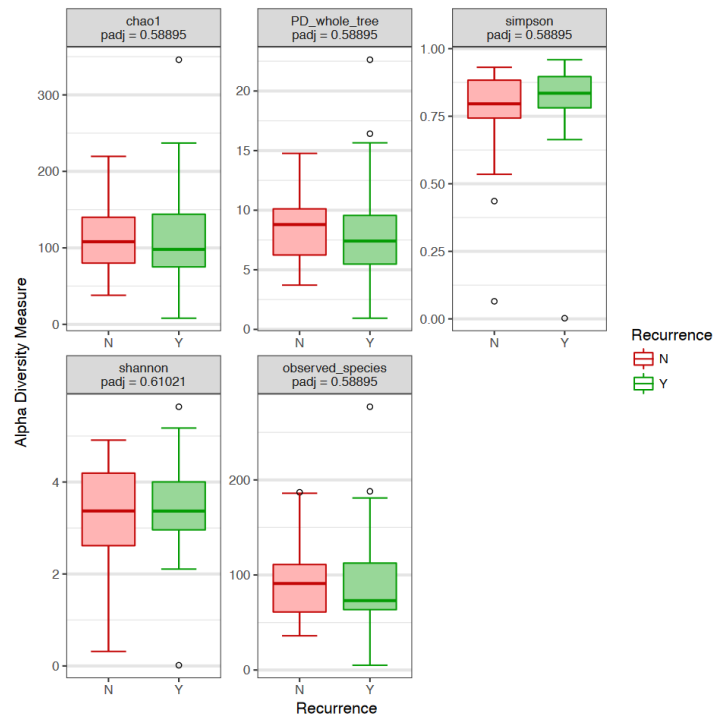
6.3.2. Beta Diversity

The beta diversity (dissimilarity) indices, with both two-dimensional and three-dimensional calculations, are shown in Table 6.5 and Figures 6.4 – 6.7. There is consistency between the scales of the axes in two-dimensional and three-dimensional projections. There may be some differential OTU representation between the recurrent and incident CDI groups, which is easier to visualise in three-dimensional representation.

Table 6.5: Results of beta diversity indices

Beta diversity index	2-Dimensional analysis		3-Dimensional analysis		
	PC1 axis	PC2 axis	PC1 axis	PC2 axis	PC3 axis
Bray Curtis	24.9%	10%	24.9%	10%	9.1%
Unweighted UniFrac PCoA	23.9%	21.1%	23.9%	21.1%	6.8%
Weighted UniFrac PCoA	40.4%	20.6%	40.4%	20.6%	19.6%

Figure 6.4: Bray Curtis dissimilarity between samples with CDI ('N') or RCDI ('Y')
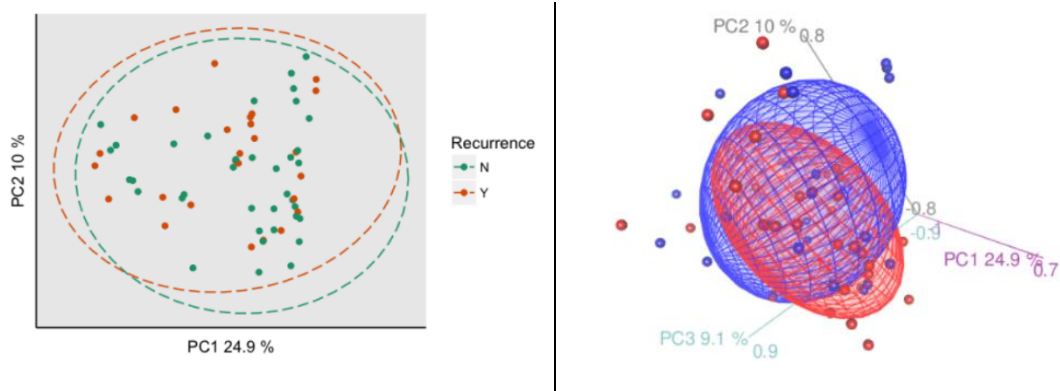
Figure 6.5: Unweighted UniFrac PCoA of CDI ('N') versus RCDI ('Y')
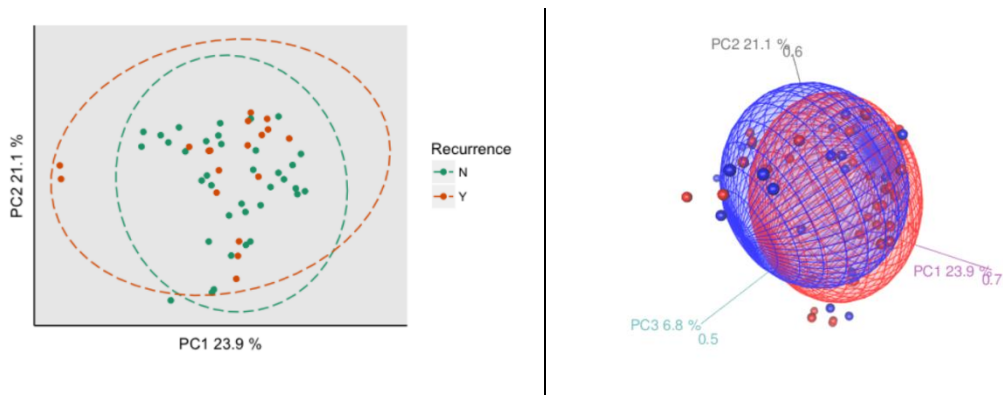
The anosim (analysis of similarity) statistical test was performed to investigate if there was a statistical difference between the microbial communities of the two sample groups, i.e. faecal samples from either patients with recurrent CDI or patients with single episode CDI. The result obtained was an R value of -0.021, with a non-significant p-value of 0.705. This R value indicates there are greater differences within these two groups than there are between them [174].

Figure 6.6: Anosim test of differences between CDI ('N') and RCDI ('Y')
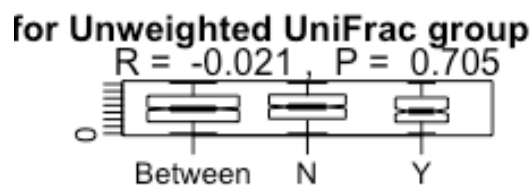
Figure 6.7: Weighted UniFrac PCoA of CDI ('N') versus RCDI ('Y')
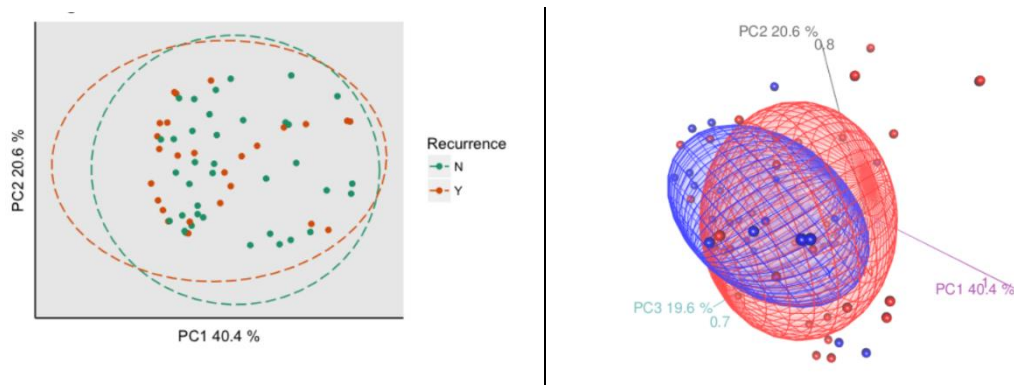
6.3.3. Taxonomy Differences

The OTUs were compared for abundance between the two categories of sample, i.e. 'no recurrence' or single episode CDI, and 'yes (regarding) recurrence' or recurrent CDI. These comparisons are shown by phyla, class, and group in Figures 6.8 – 6.10, to illustrate the varied proportions of the OTUs, ultimately representing the bacterial species present. Even at the level of bacterial phyla, there appears to be a greater mean proportion of Firmicutes and fewer Bacteroidetes in the samples of patients with recurrent CDI. This is illustrated in Figure 6.8.

Figure 6.8: Proportions of bacterial classes, by status of CDI ('N') or RCDI ('Y')
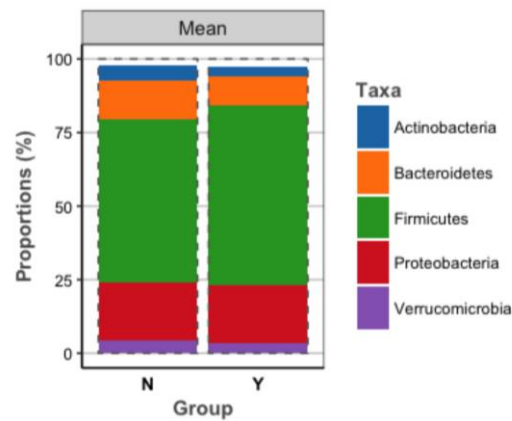
When OTUs are compared by bacterial classes of the groups, there appears to be greater variation between the two groups. This includes a greater proportion of Firmicutes/Bacilla bacterial species and smaller proportions of both Actinobacteria/Actinobacteria and

Bacteroidetes/Bacteroidia in the samples from patients with RCDI. This comparison is shown in Figure 6.9.

Figure 6.9: Proportions of bacterial classes, by status of CDI ('N') or RCDI ('Y')
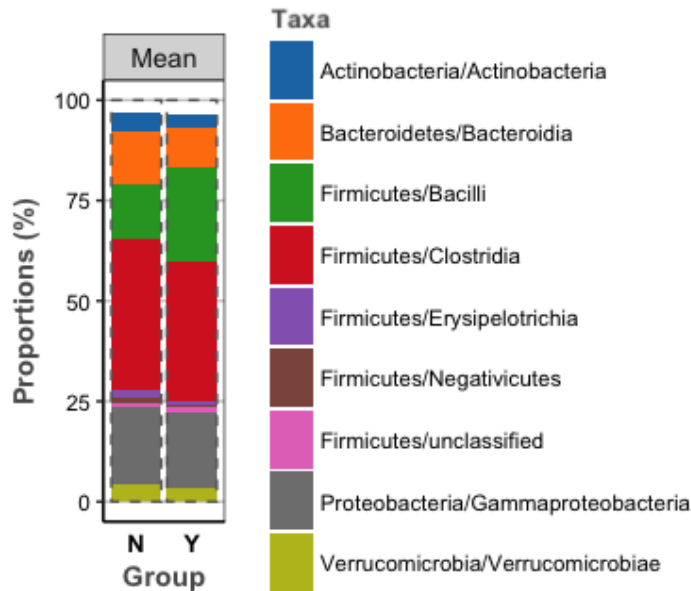
Comparison of bacterial families between the two groups shows results which appear to be consistent with the findings of the comparison of bacterial classes between groups. The comparison of bacterial families is shown in Figure 6.10. For patients with RCDI, the faecal samples showed a greater proportion of Enterococcaceae and Lactobacillaceae families of Firmicutes/Bacilli class, and smaller proportions of the Bifidobacteriaceae and Coriobacteriaeceae families of the Actinobacteria class, and the Bacteroidaceae and Porphyromonadaceae families of the Bacteroidia class. In addition, there appears to be a greater proportion of Peptostreptococcaceae family of the Clostridia class.

Figure 6.10: Proportions of bacterial families, by status of CDI ('N') or RCDI ('Y')
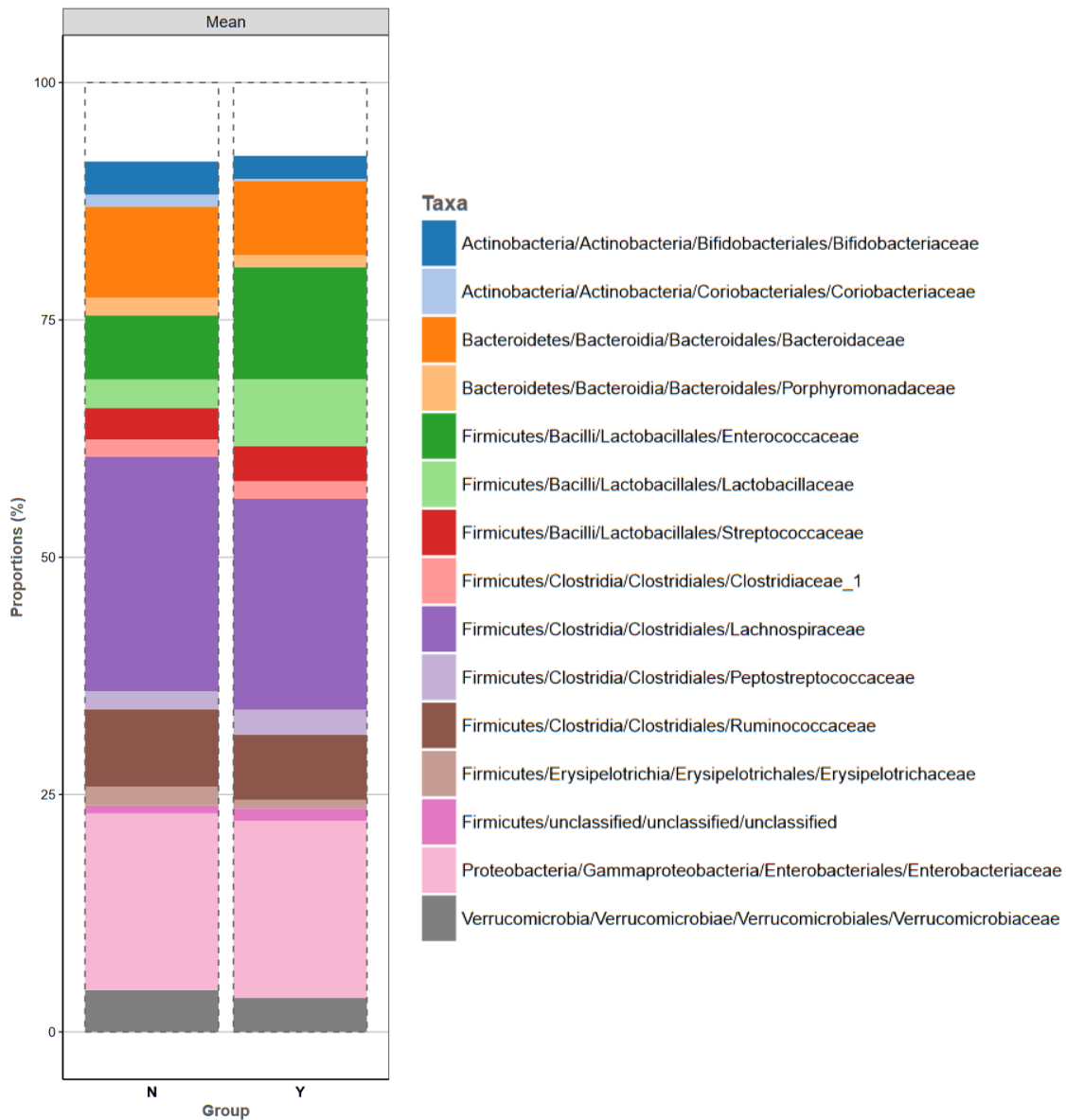
6.3.4.   Bacterial Species

There were 24 bacterial species with sufficient differences between the two CDI categories to test for differential abundance by Mann-Whitney U test. These species are shown in Table 6.6. Regarding bacterial phyla of these species, there are fifteen Firmicutes species, four Bacteroidetes species, three Proteobacteria, and one each of Actinobacteria and Verrucomicrobia phyla. The Mann-Whitney U test demonstrated significant p-values for the four Bacteroidetes species: *B. xylanisolvens* (p = 0.029741841), *B. dorei* (p = 0.021177864), *B. uniformis* (p = 0.014828916) and *B. thetaiotamicron* (p =

0.007319304). There was also a significantly different abundance of an unclassified Citrobacter species (p = 0.027388684).

Bacterial species were also tested for differential abundance by the DeSeq package in R. This analysis found two bacterial species were significantly less abundant in the faecal samples of patients with RCDI: *Bacteroides uniformis* and *Bacteroides xylanisolvens*.

The graphical output of this analysis is shown in Figure 6.11, with the left panel displaying the relative abundance of *Bacteroides uniformis* and the right panel displaying the relative abundance of *Bacteroides xylanisolvens*.

Figure 6.11: Bacterial species found to be significantly different in abundance between CDI ('N') and RCDI ('Y')

Differential abundance of:

B. uniformis   **B. xylanisolvens**
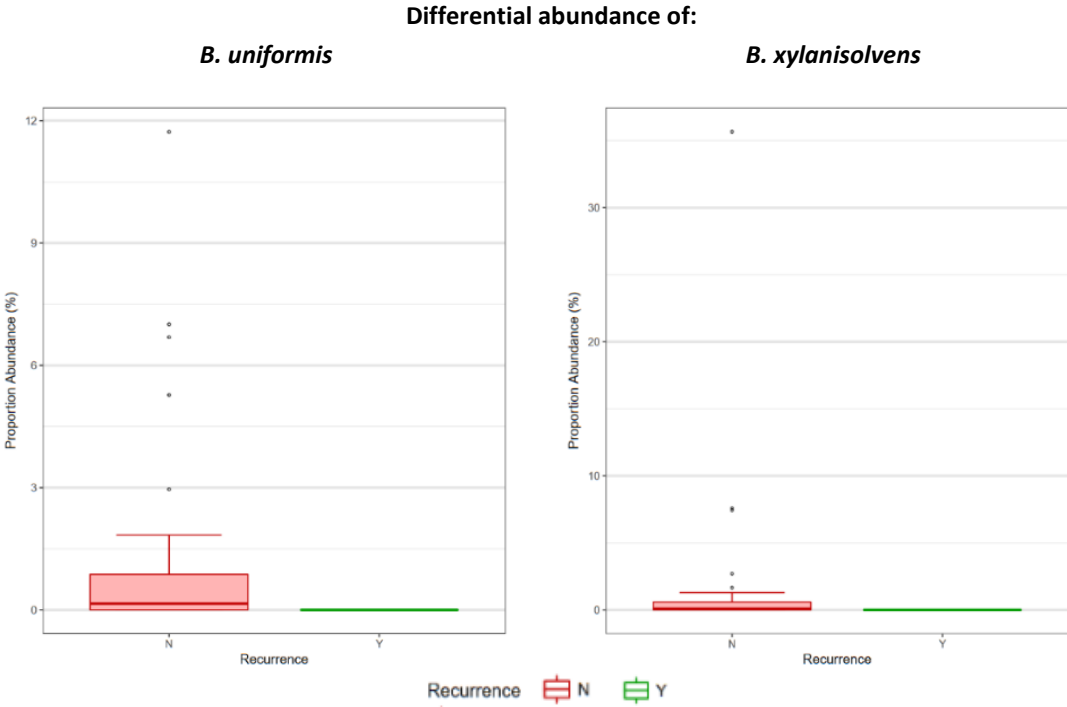
Table 6.6: Bacterial species included in Mann-Whitney U test of differential abundance

Phylum	Family/genus	Species	p value	Significant p value
Actinobacteria	*Bifidobacterium*	*longum*	0.338887066	No
Bacteroidetes	*Bacteroides*	*dorei*	0.021177864	Yes
Bacteroidetes	*Bacteroides*	*thetaiotamicron*	0.007319304	Yes
Bacteroidetes	*Bacteroides*	*uniformis*	0.014828916	Yes
Bacteroidetes	*Bacteroides*	*xylanisolvens*	0.029741841	Yes
Firmicutes	*Enterococcus*	unclassified	0.102295499	No
Firmicutes	*Lactobacillus*	unclassified	0.096673373	No
Firmicutes	*Streptococcus*	unclassified	0.226396509	No
Firmicutes	*Clostridium_sensu_stricto*	unclassified	0.323453597	No
Firmicutes	*Blautia*	*faecis*	0.206127179	No
Firmicutes	*Blautia*	*luti*	0.113587892	No
Firmicutes	*Blautia*	unclassified	0.371566060	No
Firmicutes	*Clostridium_XlVa*	unclassified	0.951516492	No
Firmicutes	*Ruminococcus*	*torques*	0.883714603	No
Firmicutes	*Lachnospiraceae*	unclassified	0.255670038	No
Firmicutes	*Peptoclostridium*	*difficile*	0.133196281	No
Firmicutes	*Faecalibacterium*	*prausnitzii*	0.493222716	No
Firmicutes	*Ruminococcaceae*	unclassified	0.812556107	No
Firmicutes	*Clostridium*	*ramosum*	0.460179721	No
Firmicutes	*unclassified*	unclassified	0.301303254	No
Proteobacteria	*Citrobacter*	unclassified	0.027388684	Yes
Proteobacteria	*Shigella*	unclassified	0.648478308	No
Proteobacteria	*Klebsiella*	unclassified	0.966044368	No
Verrucomicrobia	*Akkermansia*	*muciniphila*	0.985046490	No

6.4.   Discussion

The initial plan for recruitment of suitable patients and identification of appropriate

samples for investigation proved to have been of low yield within the clinical context of

this tertiary hospital. While provision of faecal samples specifically for the purpose of 16S rRNA analysis was considered to have been the ideal, Seekatz *et al.* published a proof-of-concept study, regarding the use of the portion of faecal samples remaining after completion of diagnostic tests [172]. The impact of prolonged storage must also be considered for the potential to alter microbiome profiles. One study investigated the impact of storage at -80°C found that results were not significantly different within two years of freezing [175]. The authors report their findings as consistent with those of a previous publication [175]. When the approval for the amended protocol, regarding the use of remainder of diagnostic samples, was provided for this study, the minimum number of 16 samples per group was easily attained; the final analysis had 41 samples representing single episodes of CDI and 29 samples representing episodes of recurrent CDI.

There were five measures of alpha diversity calculated for these samples, of which results did not indicate any statistically significant differences. Although non-significant, the samples of recurrent CDI appeared to have lower phylogenetic diversity and observed species scores, comparable Shannon and Chao-1 scores, and marginally higher Simpson index scores. This reflects the complexity of application and interpretation of alpha diversity measures, and the extent of variation between measures [176]. Seekatz *et al.* had reported the inverse Simpson index of index microbiota samples, as the sole indicator of alpha diversity [172]. However, transforming the Simpson index $(1 - \lambda)$ to the inverse Simpson index $(1/\lambda)$ did not provide a significant result for this analysis.

In contrast to the alpha diversity, which assess the number of OTUs present in single populations (e.g. individual samples), beta diversity considers how OTUs are shared between populations, i.e. examination of the relative similarities and differences between populations [173]. There are multiple methods for calculating beta diversity of populations; like the alpha diversity indices, performance may differ between them, but Bashan *et al.* found consistency of the beta diversity measures when applied to the gastrointestinal microbiome profiles of the Human Microbiome Project and Student Microbiome Project [173, 177].

Within this analysis, there is greater consistency between measures of beta diversity than what was found for measures of alpha diversity. This includes two and three-dimensional axial separation of PC1 in Bray-Curtis and unweighted UniFrac PCoA, and PC2 of unweighted and weighted UniFrac PCoA analyses. The differentiation of PC1 (axis 1) is greater than that identified by Seekatz *et al*., whereas the differentiation of PC2 (axis 2) is comparable.

Unweighted UniFrac analysis is the beta diversity measure most sensitive to sequencing depth, and thus, increased sampling and sequencing may allow for greater resolution [178]. The anosim result indicates the extent of variation of findings within the two groups, as at least equal to the variation between them. The weighted UniFrac analysis, which considers relative OTU abundance, allows for the greatest axial separation of PC1 in this study, but it is also affected by number of samples included in the analysis [178].

There was consistency in the overall findings of bacterial phyla and classes pertaining to bacterial taxonomy. Greater resolution is provided by analysis of bacterial families, genera, and species. Samples of RCDI had greater proportions of the Enterococcaceae and Lactobacillaceae families of Firmicutes/Bacilli class, and the Peptostreptococcaceae family of the Clostridia class. They also had smaller proportions of the Bifidobacteriaceae and Coriobacteriaeceae families of the Actinobacteria class and the Bacteroidaceae and Porphyromonadaceae families of the Bacteroidia class. In a review of available literature on microbiota and faecal microbiota transplantation (FMT), Streptococcus was one of the genera which had a correlation with increased susceptibility to CDI in mouse models, and Bacteroidetes was considered to be protective against CDI in human studies [110–112]. Weingarden *et al.* found that restoration of normal bacterial composition of faecal microbiota was accompanied by rapid normalisation of faecal bile acid composition (to secondary bile acids) after FMT [113]. Normalisation of bacterial composition include increasing the abundance of *Bacteroidetes* [111]. Our findings have some consistency with previous studies, such as Schubert *et al.*'s report that their subjects with CDI were more likely to have *Enterococcaceae*, *Lachnospiraceae*, and

Erysipelotrichaceae, and less likely to have *Bacteroides* within their microbiome samples [170]. A nested case-control study also found an association with increased *Lactobacillaceae* and *Enterococcaceae* among patients with *C. difficile*, but not with *Enterobacteriaceae* [179].

Differential abundance of bacterial species by two methods identified significantly less *Bacteroides uniformis* and *Bacteroides xylanisolvens* in patients with recurrent CDI. *B. uniformis* strains are capable of degrading mucin, which enables their colonisation of the gastro-intestinal mucosa [180]. *B. uniformis* has also been shown to attenuate lipopolysaccharide-induced interleukin-8 release [181] In mouse models, administration of *Bacteroides uniformis* strain ameliorated the associated metabolic and immune dysfunction associated with diet-induced obesity [182]. Mullish *et al.* investigated bile metabolism pre- and post-FMT: pre-FMT status was characterised by a negative correlation between (greater) taurocholic acid levels and the abundance of bacterial species that produce bile salt hydrolase enzymes [183]. *Bacteroides ovatus* can reduce the germination of *C. difficile* mediated by taurocholic acid [183]. As *B. ovatus* is the species most closely related to *B. xylanisolvens* [184], they may share this property. Both B. uniformis and B. xylanisolvens have potential for use as probiotics [185, 186]

This study adds to the knowledge of dysbiosis in CDI, especially regarding the potential distinct microbiome composition of RCDI in a real-world clinical setting. However, further work would be required to explore the underlying mechanisms of dysbiosis, and the contributions of both the metabolic activity of this altered state and interactions with the host immune system [187].

7. An Outbreak of a Novel *Clostridium Difficile* Sequence Type 295

7.1. Introduction

This paper describes the successful resolution of an outbreak of *C. difficile* infection, due to the integration of whole-genome sequence analysis with the available epidemiological information pertaining to individual patients [148].

I was solely responsible for the *C. difficile* isolation, culture, ribotype analysis and the epidemiological analysis relating to inpatient admissions. I performed the genomic DNA extraction and genomic DNA library preparation for the Illumina MiSeq benchtop sequencing machine. I wrote the first draft of the manuscript and contributed to revisions and editing required for publication, and I created Figure 7.1 as visualisation of this outbreak. This paper is included as Appendix D.

7.2. Possible Interplay Between Hospital and Community Transmission of a Novel *Clostridium Difficile* Sequence Type 295 Recognized by Next-Generation Sequencing [148]

Introduction

Clostridium difficile intestinal infection (CDI) is a leading cause of morbidity and mortality, both within healthcare environments and in the community [188]. Although the epidemic strain B1/NAP1/027 was first identified using pulsed-field gel electrophoresis (PFGE) and polymerase chain reaction (PCR) ribotyping techniques [188], the application of next-generation sequencing (NGS) to investigate the epidemiology of *C. difficile* has demonstrated greater diversity within genotypes than was previously recognised [62]. In a seminal study by Eyre *et al*., 36% of patients had no identifiable hospital or community contacts as a source for their infection despite *C. difficile* genomic differences of only 0–2 single-nucleotide variants (SNVs), which suggests complex modes of *C. difficile* acquisition [62].

Without a clearly defined incubation period for CDI, current Society for Healthcare Epidemiology of America (SHEA)-Infectious Diseases Society of America (IDSA) definitions

classify nosocomial versus community acquisition according to the time interval since the patient's most recent healthcare exposure [8]. Although patients who develop community-acquired CDI have not, themselves, had healthcare exposure, it has been acknowledged that their household or residential contacts with recent hospital admissions could provide a source for community acquisition of *C. difficile* [79]. The potential for this mode of transmission is likely to reflect the complexity of the residence in terms of numbers and health status of its residents. Long-term housing provision within a 'hostel' facility for chronically homeless adults, most of whom have chronic illness, exemplifies a high order of residence complexity. Case management strategies that include provision of hostel-type accommodation have been shown to reduce emergence and inpatient admissions of such residents [189], and tuberculosis and blood-borne virus transmission events have been identified in this context [190, 191]. Here, we report a cluster of cases of CDI involving 4 hostel residents among 7 patients with CDI due to *C. difficile* ST-295. Identification of this unique strain enabled resolution of its transmission from its recognised nosocomial origin (patient 1) [117], possible carriage to the hostel (patient 2), and subsequent nosocomial transmission following hospital admission of a hostel patient (patient 4) with recurrent CDI.

Methods

Patients' clinical details and their *C. difficile* isolates were collected prospectively with the approval of our institutional review board. CDI was diagnosed in accordance with Irish national guidelines [117]. Following identification of *C. difficile* DNA using an EntericBio PCR kit (Serosep, Annacotty, Ireland) at the diagnostic enteric laboratory, stool specimens were treated with ethanol shock before anaerobic incubation, using cycloserine-cefoxitin egg-yolk medium. Next-generation whole-genome sequencing of *C. difficile* colonies was performed at the TrinSeq Genome Sequencing Laboratory (Trinity College, Dublin, Ireland) with downstream read mapping and SNV calling performed as previously described [117]. Strain phylogenetic comparisons based on genome-wide SNV calls were performed by neighbour joining (BIONJ) using PhyML (ATGC, Montpellier, France) [192]. Genomes were assigned sequence types according to the PubMLST database [193], and

an epidemiologic analysis was performed following the approach of Eyre *et al.* [62]. A PCR-based ribotype analysis was also performed on each isolate as previously described [117]. CDI episodes were classified according to current IDSA/SHEA definitions [8]. The hospital is a 1,015-bed acute tertiary care facility with annual inpatient admissions exceeding 25,000 and an immediate catchment population of ~350,000. The hostel is within this catchment population; ~66% of its residents are allocated to single en-suite bedrooms.

Results

The details of CDI episodes confirmed to have been caused by *C. difficile* ST-295 are presented in Figure 7.1. Patient 1 had a prolonged hospital admission on ward A, during which he developed nosocomial CDI in March 2013, with 4 recurrences by September 2013. We identified the defining single-nucleotide variant (SNV) for *C. difficile* ST-295 as one that had evolved in this patient in June 2013 from an ST-2 isolate [117]. Patient 2 was admitted to ward A in July 2013, while patient 1 was symptomatic. Patient 2 had subsequent hospital admissions (from the hostel), with no further known contact with patient 1 or ward A, and this patient presented in December 2013 with community-onset symptoms of hospital-acquired CDI.

Patient 3 had been diagnosed with hospital onset of hospital-acquired CDI. He had no identifiable shared hospital exposures to either patients 1 or 2, but he was a resident of the hostel at the same time as patient 2. His only symptoms of CDI occurred as a hospital inpatient.

Patient 4 presented with community-onset symptoms in September 2014. Because he had an inpatient admission during the previous month, this episode was classified as community onset, hospital-acquired CDI. However, he had known exposure to patient 2's community onset of CDI symptoms in August 2014 (Figure 7.1). This factor raised concern over the possibility of patient 4 having community-acquired infection. Patient 4 had a recurrence in October 2014 that warranted hospital readmission.

Patient 5 was diagnosed with CDI in December 2014 as a hospital inpatient. He was also a resident of the hostel at times when both patients 2 and 4 experienced community-onset of symptomatic CDI.

Figure 7.1: Cluster of ST-295 C. difficile infection. Horizontal axis illustrates the patient's location at the time of inferred C. difficile transmission, with separation of locations by line style. Vertical axis illustrates time of cluster, from September 2013 to April 2015. Patient details are presented as pseudonym, age, gender, and CDI episodes with date and panels to represent IDSA/SHEA classification code. Hospital-onset, hospital-acquired CDI episodes are displayed as rectangular panel with light and dark gray fill ▭. Community-onset, hospital-acquired CDI episodes are displayed as rectangular panel with light and dark gray fill ▬. Solid arrows (⟶) illustrate the transmission event from exposure to one symptomatic patient. Dashed arrows (⇢) indicate exposure to two symptomatic patients. Dashed extension to patient details represents identified ward contact for transmissions, meeting criteria of Eyre et al.
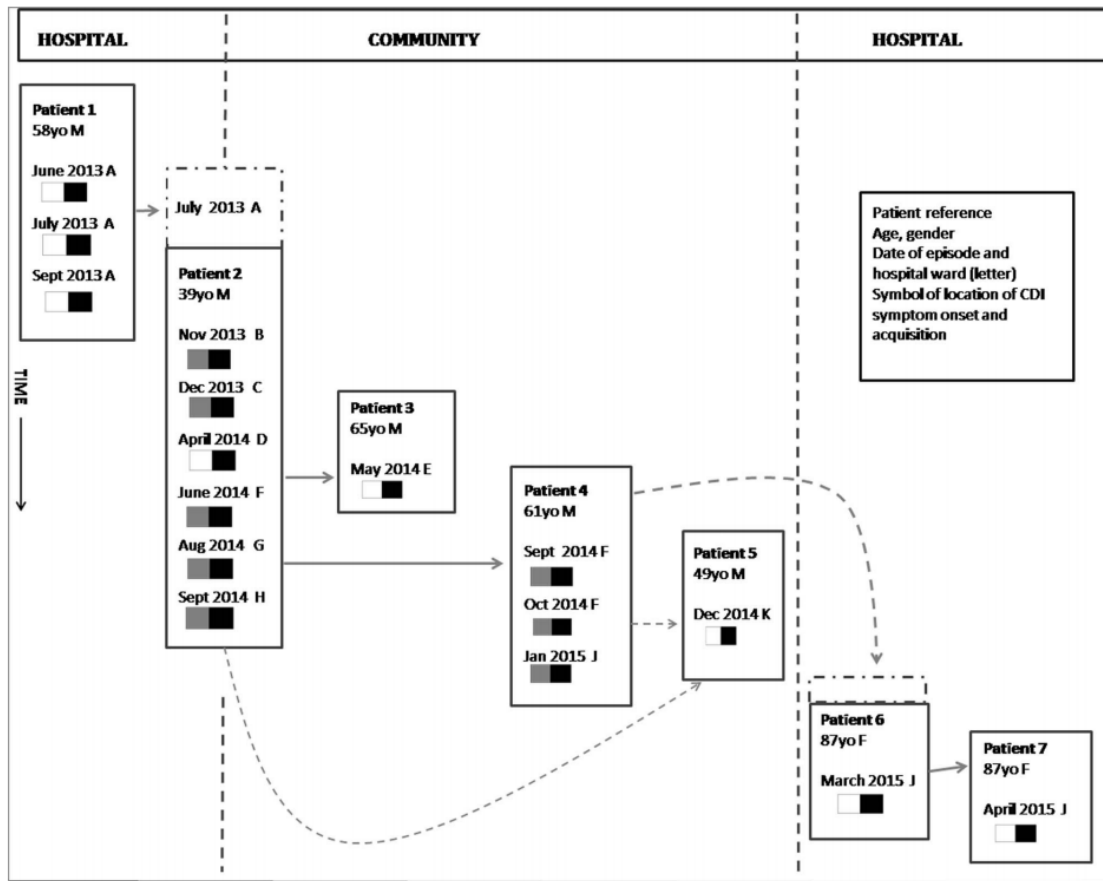
With new symptoms of CDI, patient 4 had another hospital admission in January–February 2015. CDI diagnoses were made for patient 6 in March 2015, and patient 7 in April 2015. *C. difficile* ST-295 was subsequently identified in the faecal samples of patient 6 in March 2015 and in patient 7 in April 2015. Some areas within the hospital ward (J) were common to both patients 4 and 6, and later, to patients 6 and 7.

In PCR-based ribotype analyses, all *C. difficile* ST-295 isolates were classified as ribotype 020. Phylogenetic comparison of genomic data showed that all ST-295 isolates clustered together as a distinct branch from ribotype 020 isolates belonging to ST2 (Figure 7.2). Within our collection of 200 additional *C. difficile* clinical isolates with NGS information over the course of the outbreak investigation, only 5 ST-2 isolates were within 10 SNVs of ST-295. There is no plausible clinical explanation to support alternate hypotheses for the evolution and transmission of ST-295 isolates.

Figure 7.2: The phylogenetic relationship between ST-2 and closely related ST-295 isolates. A neighbor-joining tree illustrates the phylogenetic relationship between ST-295 (black circles) and closely related ST-2 (white circles) isolates observed during the outbreak investigation. Larger circles represent clusters of genetically indistinguishable isolates. In addition, 16 additional ST-2 isolates identified during our investigation are not shown because they differed from the index case (Patient 1) by >10 SNVs. All identified isolates that differed from the index case by <10 SNVs are included in the tree.

In October 2014, the CDI outbreak cases from the hostel (patients 2,3, and 4) were reported to the responsible public health department; the first recognition of a possible CDI transmission within the hostel was made by a primary care physician. We subsequently learned that another patient from the same hostel was admitted to another hospital with CDI in September 2014, but, unfortunately, that isolate could not be recovered for investigation. The community hostel is under the governance of a voluntary organisation. It is not classified as a healthcare facility, and cleaning standards are not equivalent to those of a healthcare institution [194]. When these cases were recognised, public health action was undertaken. Residents were encouraged to practice good hand hygiene; environmental surfaces were treated with either 1,000 ppm of chlorine agent or bleach preparations and dedicated equipment was used to clean bedrooms of symptomatic residents [11].

A root-cause analysis was undertaken by the hospital's infection prevention and control team for nosocomial onset cases in response to the link identified between patients 6 and 7.

Table 7.1 presents an evaluation of the epidemiological classification of each patient's CDI episode(s) according IDSA/SHEA guidelines [8] and in light of the NGS results. Findings for patients residing in the community hostel (patients 2,3, 4 and 5) were discordant between the SHEA/IDSA classification of their initial episodes and the acquisition source inferred by NGS analysis.

Table 7.1: Comparison of SHEA – IDSA classification of CDI episodes with findings of NGS analysis

Patient No.	Date of CDI Episode	IDSA/SHEA Classification[3]	Source of *C difficile* by NGS Analysis	Time Since Exposure to Symptomatic ST-295 CDI	Time to Recurrence of ST-295 CDI	NGS vs IDSA/SHEA Classifications
Patient 1	June 2013	HO-HA	HA	Index ST-295		Concordance
	July 2013		Recurrence		25 d	
	September 2013		Recurrence		52 d	
Patient 2	November 2013	CO-HA	HA	>16 wk		Variance: time since ST-295 exposure
					25 d	Discordance: community ST-295
	December 2013		Recurrence		19 wk[a]	
	April 2014	HO-HA	Recurrence		10 wk[a]	
	July 2014	CO-HA	Recurrence		36 d	
	August 2014		Recurrence		35 d	
	September 2014		Recurrence			
Patient 3	May 2014	HO- HA	CA	>24 wk		Discordance: community ST-295 exposure despite CDI symptom onset after 48 h of hospital admission
Patient 4	September 2014	CO-HA	CA	6–48 d		Discordance: community ST-295 exposure despite recent hospital admission
	October 2014		Recurrence		31 d	
	January 2015	CO-HA	Recurrence		14 wk[a]	
Patient 5	December 2014	HO-HA	CA	10–15 wk		Discordance: community ST-295 exposure despite CDI symptom onset after 48 h of hospital admission
Patient 6	March 2015	HO-HA	HA	29 d		Concordance
Patient 7	April 2015	HO-HA	HA	37 d		Concordance

NOTE. HO, hospital onset of symptoms; HA, hospital-acquired infection; CO, community onset of symptoms; CA, community-acquired infection.
[a]Cannot exclude reinfection from environmental spores versus intestinal carriage.

Discussion

The resolution of a link between these cases by virtue of their sharing a *C. difficile* strain with sequence type 295 provides new insight into current surveillance definitions. Although prolonged carriage of *C. difficile* prior to a first episode of CDI has not been reported previously, our findings suggest that patient 2 may have had nosocomial acquisition of *C. difficile* ST-295 16 weeks before his symptoms began. This time lapse exceeds the maximum interval of 12 weeks between hospital discharge and symptom onset recognised by earlier surveillance recommendations [24]. These surveillance definitions acknowledge the complexity of attributing the source of *C. difficile* exposure when patients have been admitted to multiple facilities [24]. We cannot exclude the possibility of other mutual contacts across multiple hospital wards; however Eyre *et al.*'s analysis of such intermediate contacts, either asymptomatic or with negative enzyme immunoassay results, suggests that this is more likely a chance finding than a source of

transmission [62]. Our NGS results favour community acquisition of *C. difficile* by patients 3, 4, and 5, all of whom had episodes categorised as hospital-acquired CDI by current IDSA/SHEA definitions [8]. We believe that, with the increasing complexity of the epidemiology of CDI, NGS enhances the capacity to distinguish between community and hospital acquisition as well as the 'trafficking' between them.

The community hostel setting was an interesting aspect of this case cluster. Patients 2, 3, and 5 had single bedrooms in close proximity. As a consequence of their medical issues, patients 2, 3, 4, and 5 all had antibiotic and proton-pump inhibitor exposure prior to their first CDI episode. Appropriate infection prevention and control measures were taken in the hostel under the guidance of public health personnel. No new ST-295 infections have been detected in the 14 months since this intervention.

Hospitalised patients with suspected and/or confirmed CDI are normally placed in single rooms with en-suite facilities, accompanied by enhanced cleaning and disinfection of the room and equipment during the symptomatic period. Additional investigations have been conducted as a result of the likely link between patients 6 and 7.

We believe patient comorbidities, with further antibiotic exposure, to be the predominant cause for the recurrent disease experienced by patients 1, 2, and 4, but we cannot definitely exclude further environmental acquisition.

Although only 1 SNV distinguishes ST-295 from ST-2, its occurrence within the housekeeping gene *atpA* generated a new allele and new sequence type that served as a molecular marker for case recognition in this cluster [117, 193]. Ribotype analysis could not provide this degree of resolution, which supports its replacement by NGS [62]. To our knowledge, this is the first description of an ST-295-associated cluster and the first evidence of possible *C. difficile* transmission among residents of hostel facilities for homeless adults with chronic illness.

7.3. Conclusion

- The ST-295 isolates formed a unique subgroup of the ST-2 isolates identified as part of this prospective cohort study. As presented in Ch. 4, there were 335 isolates with whole genome sequence data, of which ST-2 and ST-295 formed a cluster of 39 (12%) isolates. The ST-2 isolates were identified throughout the timeframe of the cohort study: 8 during 2013, 15 during 2014 and 8 during 2015. This outbreak investigations provides further evidence for the contribution of recurrent CDI to *C. difficile* transmission in this hospital [117].
  - Three of the seven patients with *C. difficile* ST-295 had recurrent CDI
  - The signature mutation occurred during Patient 1's experiences of multiple episodes, and a ward transmission is a plausible explanation for Patient 2's acquisition of this strain, despite the prolonged period of incubation and/or carriage before his first episode of CDI.
  - Patient 2 had six episodes of CDI, each associated with *C. difficile* ST-295, and his experience of symptoms while resident in the hostel can account for the introduction of this strain to the hostel.
  - Patient 4 had three episodes of CDI, and his admissions for the management of recurrent CDI inadvertently led to two further episodes of nosocomial transmission.
- It also supports the hypothesis of McLure *et al.*, who proposed from their modelling work that nosocomial CDI could not be sustained without multiple introductions of *C. difficile* from the community [142].
- The hostel appeared to have been a community reservoir for this outbreak. In the EUCLID study, Eyre and colleagues found no evidence of clusters of *C. difficile* ribotypes 014 or 020 (corresponding to ST-2/ ST-295) within hospitals or countries [147]. As the ST-295 outbreak was captured within a prospective cohort study, samples obtained over a longer period may allow for better identification of local transmission, and thus, our contrasting findings within this outbreak.

8. Evidence for Inter-European Country Transmission of *Clostridioides difficile* Ribotype 078

8.1. Introduction

Since the earliest applications of PCR ribotype analysis to CDI epidemiology, different ribotypes have been more prevalent in either HA CDI, such as ribotype 027, or for CA CDI, ribotype 078 [80, 140, 195]. There is overlap between the ribotypes associated with HA and CA CDI, as demonstrated in the Netherlands, where the prevalence of ribotype 078 was found to increase from 3% in 2005 to 13% in 2008, and was associated with both severe CDI and CA CDI [196]. Ribotype 078 has also emerged as the most prevalent ribotype associated with *C. difficile* colonisation and infection of animals, including farm animals. A zoonotic origin for this ribotype has been proposed in accordance with such findings [3].

In Chapter 4, *C. difficile* ribotype 078 was shown to be the most common ribotype identified in this prospective study of 335 clinical isolates analysed. This accounts for approximately 17% of all isolates. As the most common ribotype overall, it was also the most prevalent ribotype of RCDI, plausible transmission events and cryptic transmission events. It was the second most common ribotype of the 17 CA CDI isolates. Although ribotype information was available for only 19% of all the CDI cases in Ireland notified to the HPSC for 2018, 078 was the second most frequently identified *C. difficile* ribotype, 11% [26].

During the prospective cohort study, *C. difficile* isolates from porcine necropsies at the Central Veterinary Research Hospital (CVRL) were obtained for the purpose of ribotype analysis. These isolates had been detected by EIA in samples from pigs with typhlocolitis identified by histopathological examination, albeit with additional findings to suggest bacterial co-infection. This was published as a case series, as the first report of *C. difficile* ribotypes 078 and 110 contributing to typhlocolitis of Irish pigs [197], and is included as Appendix E. More *C. difficile* isolates from porcine samples were made

available, and as ribotype 078 was most prevalent, they were included in the plans for WGS analysis of the prospective cohort and CA CDI isolates.

8.2. Methods

The following manuscript has been published in Emerging Infectious Diseases [198]. This paper is included as Appendix F.

8.3. Evidence for Inter-European Country Transmission of *Clostridioides difficile* Ribotype 078

Introduction

Clostridioides (formerly Clostridium) difficile was considered to be a predominantly nosocomial pathogen until findings of several whole genome sequencing studies suggested a more complex epidemiology. For example, Eyre *et al.* reported that only 35% of nosocomial *C. difficile* infections (CDIs) were potentially attributable to other cases based on genomic data, and only 19% were additionally linked through sharing possible hospital-based contact [62]. This finding suggests that a major proportion of *C. difficile* from CDI cases occurring in healthcare institutions originates from other sources, including the community [148].

Community-associated CDI (CA CDI) is now well recognized, accounting for ≈25% of cases in Australia, ≤25% of cases in Europe, and 33% of cases in the United States [3, 199]. There is increasing recognition that *C. difficile* is a near ubiquitous environmental organism, and that humans have widespread environmental exposure to it*. C. difficile* has been detected in samples from parks (24.6%); water sources, including rivers, lakes, and sea water; homes (17.1%); commercial stores; and other premises (6.5 – 8.1%), in addition to hospitals (16.5%) [200, 201].

Isolates of *C. difficile* from these studies underwent ribotype analysis: overall, ribotype 027 isolates were most commonly identified in hospital samples, and ribotype 014-020 isolates predominated in other environmental samples. Isolates of the most common ribotypes were not restricted to any particular location [200]. These findings

support the possibility that there are different sources for exposure to each *C. difficile* ribotype.

Occurrence of CDI due to *C. difficile* ribotype 027 has been greatly reduced in the United Kingdom, most likely the result of the combination of antimicrobial stewardship and hospital infection prevention and control measures. However, these interventions have not reduced the incidence of infections caused by other ribotypes, including ribotype 078 [149].

Findings of genomic analysis of isolates from the European, Multi-center, Prospective, Biannual, Point-Prevalence Study of *Clostridium difficile* Infection in Hospitalised Patients with Diarrhoea (EUCLID) showed that specific *C. difficile* ribotypes were associated with healthcare clusters, and other ribotypes had an international distribution across Europe [147]. For example, ribotype 078 isolates did not cluster by their country of origin, indicating a complex distribution unrelated to nosocomial transmission. The mechanisms of transmission have not been identified, but might be related to the movement of food, other animal-derived products, or people across Europe [147].

C. difficile carriage and infection has been well described in livestock and other animals [3]; certain ribotypes of *C. difficile* are considered to be major ribotypes from a One Health perspective. These ribotypes includes ribotype 078, carriage of which has been reported in 9–100% of piglets from North America, Europe, Asia, and Australia [3]. Carriage rates in calves (56%) and cows (13%) have been lower. Although many studies did not identify any major carriage in adult pigs, one study in the Netherlands reported a rate ranging from 6.6–100% [3].

We have reported *C. difficile* ribotype 078 in cases of typhlocolitis in neonatal piglets in Ireland [197], and Knetsch *et al.* found that ribotype 078 isolates carried by farmers in the Netherlands and their pigs were identical by whole-genome sequence analysis [66]. These findings suggest that *C. difficile* isolates may be shared between

humans and pigs when in close proximity. However, the mechanisms, and directions of transmission are not known.

In this study we investigated the genomic relationships between *C. difficile* ribotype 078 isolates of both human and porcine origin collected from Ireland and compared these with international ribotype 078 isolates. We investigated the extent to which geographical proximity could explain clusters of clonal isolates.

Methods

Samples and Settings

Clinical isolates of *C. difficile* ribotype 078 were collected prospectively as part of an investigation of consecutive episodes of CDI conducted at St James's Hospital (Dublin, Ireland), a 900-bed tertiary referral center, during 2013–2016. Stool samples, sent from patients with diarrhoea, had the *C. difficile* toxin B gene identified by using the EntericBio PCR kit (Serosep, https://www.serosep.com). We reviewed medical notes of inpatients to obtain relevant clinical data, including antimicrobial drugs and proton pump inhibitors prescribed before the onset of diarrhea, features indicative of severe CDI with or without complications, and the antimicrobial drugs used for management of CDI. These data were pseudonymized and stored in a dedicated database.

We retrieved an additional 9 *C. difficile* 078 isolates from a study of recurrent CDI at St James' Hospital during 2012–2013 [117]. Five additional *C. difficile* ribotype 078 isolates were provided from those submitted to a national surveillance study of CA CDI in Ireland conducted during 2015. Isolates of *C. difficile* were recovered from pigs that had been referred for autopsy at the Central Veterinary Research Laboratory (CVRL; Backweston, Ireland) during 2014–2015, irrespective of the suspected cause of death, by sampling colonic contents or feces that had tested positive for *C. difficile* toxins A/B using Premier Elisa Kit (Meridian BioScience Inc., https://www.meridianbioscience.com). We treated human fecal and porcine colonic/fecal samples with ethanol shock before anaerobic incubation on cycloserine-cefoxitin egg-yolk medium. DNA was extracted from

resulting colonies for PCR ribotype analysis and Illumina genomic library preparation as previously described [117].

Whole-Genome Sequencing

Whole-genome sequencing was performed either on an Illumina MiSeq or MiniSeq platform at Trinity College, Dublin, Ireland, or on the Illumina HiSeq platform at Wellcome Centre for Human Genetics, University of Oxford (Oxford, UK). Sequence data generated have been deposited in the National Center for Biotechnology Information Short Read Archive (https://www.ncbi.nlm.nih.gov/sra) under BioProject PRJNA692997.

We mapped sequence reads to the ribotype 078 reference genome M120 (GenBank accession no. FN665653.1) and identified high-quality variants using an approach developed and calibrated for *C. difficile* (Eyre NEJM) with later refinements [146] (Appendix of publication). We obtained published comparison sequences from the EUCLID pan-European cross-sectional survey conducted in 2012–13 [147] and from farm animal and human isolates from the Netherlands (2002–11), described by Knetsch *et al.* [66].

Sequence Comparisons

We compared sequences by using single nucleotide polymorphisms (SNPs) and obtained differences between sequences from maximum-likelihood phylogenies corrected for recombination (Appendix of publication). We reviewed phylogenetic analysis of closely related genomes in conjunction with available epidemiological data. Within the clinical database, CDI recurrence was defined as the identification of two isolates within 10 SNPs from a single patient [62], for which that patient had clearly documented clinical resolution of symptoms after their first episode. On the basis of rates of *C. difficile* evolution and within-host diversity [62], we defined plausible short-term transmission/mutual exposure as isolates differing by 0–2 SNPs.

We made epidemiological matches between patients who had inpatient admissions and demonstrable links with respect to time, location, or healthcare staff,

where their *C. difficile* isolates were within 0–2 SNPs. Because epidemiological details were not available for either the CA CDI investigation in Ireland or the EUCLID isolates, we analyzed linkage between cases on the basis of genetic similarity alone. These genomic pairs were named by the isolate sources in chronological order of identification.

Ethics

Investigation of hospital-associated CDI (HA-CDI) cases at St James's Hospital was conducted after obtaining approval from the St James's Hospital/Tallaght Research Ethics Committee. Porcine isolates were exempt from requiring ethics approval.

Results

A total of 171 *C. difficile* ribotype 078 isolates were included in the analysis. There were 53 isolates from CDI episodes in 44 inpatients diagnosed at SJH (including 5 community-associated infections), 20 Irish porcine isolates, 67 Dutch clinical, farmer and porcine isolates and 31 clinical EUCLID isolates. Details of their country of origin, source, and date of isolation are shown in Table 8.1. The EUCLID isolates were A total of 171 *C. difficile* ribotype 078 isolates were included in the analysis: 53 isolates from CDI episodes in 44 inpatients at St James's Hospital, including 5 community-associated isolates; 20 porcine isolates from Ireland; 67 clinical, farmer and porcine isolates from the Netherlands and 31 clinical EUCLID isolates. We provide details of their country of origin, source, and date of isolation (Table 8.1). The EUCLID isolates were obtained from 9 European countries. Six countries, including Ireland, submitted $\geq$ 2 isolates.

Of the 53 CDI isolates causing CDI in Ireland, 9 were from recurrent CDI episodes in 7 patients (7 subsequent isolates were 0 SNPs different from the baseline isolate, 1 was 1 SNP different and 1 was 8 SNPs different). Only the first isolate from each patient was considered in subsequent analyses. We provide genomic relationships between the remaining 162 ribotype 078 isolates (Figure 8.1). Despite the diverse sampling frame, only limited diversity was seen; the greatest root to tip distance in the phylogenetic tree was 48 SNPs.

Isolates from Ireland were found throughout the tree, but specific clusters of these isolates were seen, including, as shown at the $\approx 120^0$ (≈ 8 o'clock) position (Figure 8.1), a cluster of cases that included isolates from HA-CDI and CA CDI cases as well as cases from pigs. Within this cluster, several porcine isolates were directly ancestral to 1 HA-CDI case. Another 5 CDI cases, including 1 CA CDI, had another porcine isolate directly ancestral. This finding suggests a porcine origin for these cases, either directly or by ≥ 1 or more intermediate (unsampled) transmission routes. This same cluster also contained an isolate from a pig and a farmer in the Netherlands. Several other clinical isolates from the Netherlands were closely related to porcine isolates (Figure 8.1).

We provide epidemiologic links between genetically related isolates within 0–2 SNPs (Table 8.2). Although nearly all genomic pairs occurred among isolates with the same country of origin, the epidemiologic information available can explain only a small proportion of transmissions/mutual exposures.

Figure 8.1: Recombination-adjusted maximum-likelihood phylogenetic tree of sequences from human and porcine Clostridioides difficile isolates from Ireland and 9 other countries in Europe. Isolates are shown as triangles for healthcare-associated C. difficile cases and circles for community-associated C. difficile cases. Isolates from pigs are shown as crosses and those from farmers as squares. The color at each tip indicates the country of origin of the isolate. The tree was based on 4,861 variable sites before correction for recombination, based on a median (interquartile ranges) of 93.4% (93.0%–93.8%) and (83.1%–96.2%) of the reference genome being called. Scale bar indicates single-nucleotide polymorphisms.

Table 8.1: Countries from which Clostridioides. difficile 078 isolates originated, their identified sources, and date of collection

Country of origin of isolates	Source of isolates	Timeframe of collection	No. of isolates
Ireland	HA CDI	2012 – 2016	48
	Porcine	2014 – 2015	20
	CA CDI	April – June 2015	5
Netherlands	CDI	2002 – 2011	31
	Porcine	2009, 2011	20
	Healthy farmers	2011	16
EUCLID, including:	HA CDI	Dec 2012 – Aug 2013	31
- Germany			9
- Italy			7
- United Kingdom			4
- France			3
- Portugal			3
- Ireland			2
- Spain			1
- Greece			1
- Austria			1

Table 8.2 shows the epidemiological links between genetically related isolates within 0–2 SNPs. While almost all pairs occurred among isolates with the same country of origin, the epidemiological information available can explain only a small proportion of transmissions/mutual exposures.

Table 8.2: Pairs of C. difficile ribotype 078 isolates matched by country of origin and source case, with associated epidemiology

Country 1	Source of *C. difficile* isolates(s)	Country 2	Source of *C. difficile* isolates(s)	No. of pairs of isolates	Associated epidemiology
Ireland	CA CDI	Ireland	CA CDI	2	No known links
Ireland	CA CDI	Ireland	HA CDI	2	No known links
Ireland	HA CDI	Ireland	HA CDI	10	Possible transmission 6 pairs [1] Unknown for 4 pairs
Ireland	Porcine	Ireland	HA CDI	3	No known links
Ireland	Porcine	Ireland	Porcine	12	Eight pairs at one farm Three pairs at one farm One pair at one farm No pairs between farms
Ireland	CA CDI	Italy	HA CDI	1	Unknown
Ireland	HA CDI	UK	HA CDI	1	Unknown
Germany	HA CDI	Germany	HA CDI	1	Unknown
Netherlands	HA CDI	Netherlands	HA CDI	1	Unknown
Netherlands	CDI	Netherlands	Farmer	1	No known links
Netherlands	CDI	Netherlands	Porcine	1	No known links
Netherlands	Farmer	Netherlands	Farmer	3	Unknown
Netherlands	Farmer	Netherlands	Porcine	10	Farm exposures
Netherlands	Porcine	Netherlands	Porcine	1	No known links
Portugal	HA CDI	Portugal	HA CDI	1	Unknown

CA CDI, community-associated *C. difficile* infection; HA-CDI, healthcare-associated *C. difficile* infection.

[1] The 6 possible healthcare-associated transmission pairs shared time and space on the same hospital ward (n=3) or time on different hospital wards while under the care of the same medical team (n=3).

Discussion

Our findings support a complex regional and international distribution of *C. difficile* ribotype 078 isolates. In contrast to the EUCLID study, which obtained samples on single days in winter and summer, more dense sampling was undertaken in our study. In the EUCLID study, no evidence of clustering of ribotype 078 within countries was seen, which is consistent with a complex pattern of dissemination in Europe over timescales spanning years (Figure 8.1). However, our study shows evidence of sublineages of ribotype 078 that are predominantly found in isolates from the Netherlands and others predominantly found in isolates from Ireland (Figure 8.1). It is likely that this denser sampling has enabled recent, local, onward transmission to be better captured. We also identify a EUCLID isolate from Italy (2013) and a CA CDI isolate from Dublin, Ireland (from 2014) that are within 2 SNPs, consistent with a temporally related transmission. However, we do not know of any epidemiologic link between these 2 cases.

For 10 pairs of isolates within 2 SNPs from inpatients who had HA-CDI, possible healthcare-based epidemiologic links could be made for 6 of these pairs, but not the other 4. Plausible ward-based transmission only accounted for 3 pairs. For other genetically related isolates pertaining to inpatients in our study, there was a median of 559 days between their associated CDI episodes (range of 147–651 days) without overlapping hospital admissions or appointments. Overall, nosocomial transmission accounted for 15% of closely genetically related (≤2 SNPs) *C. difficile* ribotype 078 cases in this study, with equal proportions were attributable to farms and unknown transmission routes. In a study in Leeds, UK, which had comparable phylogenetic analysis, hospital ward-based epidemiological linkage was reported as 11% for ribotype 078 cases, versus 64% for ribotype 027 case [31].

A EUCLID isolate from Ireland (2013) forms a genomic cluster with 1 CA CDI isolate (2015) and 2 HA-CDI isolates (July 2015 and Dec 2015). These 4 isolates were from patients in 3 Dublin healthcare facilities and from 1 case of CA CDI that had been collected within a three-year period. This finding suggests shared exposure across the greater Dublin area, and that nosocomial transmission is not the dominant route of

acquisition of *C difficile* ribotype 078. This observation is consistent with the EUCLID study findings [147].

It is not clearly understood how people with CA CDI have acquired their infection because they do not have the risk factors for HA-CDI [19]. Anderson *et al.* described proximity to livestock farms, agricultural industry, and nursing home facilities as risk factors for CA CDI in North Carolina, but they did not include analysis of *C. difficile* molecular data in their models [202]. In contrast, Van Dorp *et al.* found no evidence of either localised point sources or livestock exposure as risk factors for *C. difficile* acquisition in the Netherlands [203]. They included ribotype detail in their analysis but found no evidence of geographical clustering of ribotype 078 CDI cases [203]. This finding is consistent with that of Knetsch *et al.*, who reported clonal isolates of farm and clinical origin, without a geographic basis for those clusters [66].

Knetsch *et al.* identified another genomic cluster of *C. difficile* ribotype 078 isolates which included an isolate of animal origin from Canada (2004) and 8 isolates of clinical origin from the United Kingdom (2008–2012) [204]. We also identified a cluster of clinical and porcine 078 isolates from Ireland, where there was no known occupational exposure of the affected patients who lived in urban locations far from relevant pig farms. Knight *et al.* reported clonal ribotype 014 isolates from Australia that were considerable geographic distances from each other, which is suggestive of long-range transmission and major community reservoirs. They concluded that this transmission was unlikely to have been caused by direct contact between the humans and animals involved, and suggested that by-products, such as manure or compost, could enable indirect transmission from animals and humans [68]. In a study from the United States, biosolid-based compost had the highest rate of *C. difficile* recovery that included ribotype 078 isolates [205], which was also the most common ribotype in an investigation of manure from Japan [206].

Findings based on ribotype analysis alone are insufficient for clear identification of transmission events pertaining to community reservoirs [32]. Moradigaravand *et al.*

identified ≈90% of their collection of clinical and wastewater isolates as clade 1 (231/256), and only 10 (3.9%) as clade 5/ribotype 078 [72]. When their ribotype 078 isolates were compared to the same isolates from the Netherlands included in our analysis, they found divergence of ≈20 years between the isolates from the United Kingdom and the Netherlands. This finding suggests that water is not the primary reservoir or route for dissemination of *C. difficile* ribotype 078 isolates. It is still considered possible that dissemination of ribotype 078 isolates occurs by the food chain, the environment, or both [93, 207]. This view is supported by the presence and distribution of tetracycline-resistant determinants in *C. difficile* genomes, reflecting the antimicrobial drug selection pressure from tetracycline use in agriculture or veterinary practice, and thereby facilitating the emergence and spread of ribotype 078 bacteria [207].

It is not completely understood how some livestock might have asymptomatic *C. difficile* colonisation while others show development of infection [208]. The porcine isolates from Ireland in this analysis were from available samples processed at the CVRL. Three isolates included samples from neonatal piglets that had typhlocolitis [197]. We have identified genomic similarities among isolates causing human and veterinary infections. This finding augments the need for a One Health approach for *C. difficile* ribotype 078.

The strengths of this analysis include the large number of *C. difficile* ribotype 078 isolates included, from different sources including humans and animal species, and geographic origin. The limitations of this study include the lack of epidemiologic data available to the investigators for CA CDI and the limited number of porcine strains from samples available at CVRL. In conclusion, our analysis of *C. difficile* ribotypes 078 isolates from Ireland and 9 other countries in Europe showed close overlap between isolates from humans and pigs, including the occurrence of plausible transmission, either directly or by an unknown intermediate source.

8.4. Conclusion

The findings of this paper illustrate the complexity of *C. difficile* transmission. Current epidemiologic definitions of CDI cannot account for the findings of three pairs of identical genomes between patients in a Dublin hospital and pigs reared in other parts of Ireland. The definitions also cannot readily explain the two findings of identical genomes which defy international borders (Ireland CA CDI and Italy HA-CDI isolates; Ireland HA-CDI and UK HA-CDI isolates). Although not discussed within the paper, there was a small number of porcine isolates from Ireland and the Netherlands with 3–10 SNP differences between them. Allowing for the three-six years between the collection of these isolates, and the accepted rate of 1–2 mutations per *C. difficile* genome per annum, this could provide further support for long-range transmission [13, 63].

9. Conclusions

9.1. Introduction

As *C. difficile* infection remains a challenge in contemporary Irish clinical practice, this evaluation of next-generation sequence analysis adds considerable depth to the understanding of nosocomial *C. difficile* epidemiology in a major Dublin tertiary hospital, and provides some insights of broader *C. difficile* epidemiology in a regional and international context.

9.2. Methods

As described in Chapter 2, this prospective cohort study was conducted during a three-year interval, September 2013 – August 2016. Demographic and clinical details were ascertained for 381 clinical episodes relating to the detection of *C. difficile* toxin B gene NAAT in stool samples submitted for inpatients of St James's Hospital. This allowed for the establishment of a dedicated database, to assess the prevalence of established risk factors for CDI. The timing of this cohort study was set so that the diagnostic laboratory had already switched to using the NAAT as the sole test for *C. difficile* of liquid faecal samples. There was an existing practice of direct communication of positive *C. difficile* NAAT results by Medical Microbiology clinical staff to the clinical staff providing care for that patient. Nevertheless, as presented in Chapter 5, when reviewing medical records specifically to gather clinical information pertinent to CDI, I discovered that 147/381 (38%) of episodes did not meet the clinical criteria of CDI. This was most commonly due to patients experiencing 1–2 episodes of diarrhoea within a 24-hour period (46%), the administration of laxatives in the days prior to faecal sample submission to the diagnostic laboratory (39%), and due to the identification of other causes of diarrhoea (15%).

Before the investigation of this prospective cohort was initiated, suitable methods had been introduced within the department for the selective isolation and culture of *C. difficile*, PCR ribotyping, whole genome sequence generation and analysis in the context of available epidemiological data [117].

9.3. Findings

9.3.1. Risk Factors for C. difficile Infection

I consider the classification made of episodes as CDI, recurrent CDI, or *C. difficile* colonisation as an accurate representation of the medical notes available. As discussed in Chapter 5, prescription details were more readily available for inpatient prescriptions than those preceding hospital admission. Notwithstanding, analysis of the available data identified statistically significant gradients of medication exposure for both antibiotics and proton-pump inhibitors preceding hospital admission. For patients with colonisation, patients with CDI, and patients with RCDI, the prevalence of antibiotics preceding hospital admission was 26%, 40%, and 66% respectively. The prevalence of PPIs was 54%, 77%, and 85% respectively.

There were no significant differences of the prevalence of antibiotic exposure during hospital admissions, with 86%, 89% and 89% of patients respectively prescribed antibiotics. These prescriptions also followed similar proportions in their issuances: 35–41% were concordant with the empiric prescribing guidelines of the hospital, 26–32% were recommended by infection specialist staff, and 34–44% were selected by the clinical teams independently of the hospital guidelines and/or specialist recommendations. This is comparable to other studies, where 20–40% of antibiotic prescriptions are deemed inappropriate or unnecessary [159].

With the availability of the hospital's antimicrobial consumption data, the proportions of antibiotic agents prescribed to this cohort of patients were compared to the overall hospital use for this time. This revealed significant differences in the antibiotic exposures of patients within the *C. difficile* cohort, including significantly more piperacillin/tazobactam (22.4–27% vs. 10%, $p = .00001$), more IV vancomycin (7.4–10.5% vs. 4.6%, $p = .001$) and less co-amoxiclav (15.5–18.9% vs. 23.7%, $p = .0005$). It is difficult to predict the effects of (better) antimicrobial stewardship on the incidence of CDI [162]. Nevertheless, it is noteworthy that the '4C' antibiotics conventionally held as high risk for

CDI [209] were not the more commonly prescribed agents for patients who had positive NAAT results for *C. difficile* in this cohort.

Antimicrobial stewardship opportunities for the management of *C. difficile* were also identified within this cohort. Prescription of 'anti-*C. difficile*' antibiotics was frequent, with significant differences observed between the choices made for patients with colonisation, CDI, and RCDI. Patients with colonisation were most likely to receive over-treatment, as they often received antibiotics, despite recommendations to the contrary. Other findings included the observation that 22% of patients with CDI and RCDI received antibiotics for severe CDI, when their clinical features did not meet that criteria. In contrast, patients with CDI were significantly more likely to receive antibiotic undertreatment of severe infection vs. patients with RCDI (24% vs. 5%, p = .001). This was most often due to the use of concomitant oral metronidazole instead of the intravenous form, with oral vancomycin.

It was also identified that the lack of recognition of *C. difficile* colonisation in routine clinical practice would have a significant impact on the hospital's rate of RCDI. Without consideration of colonisation, there would have been a recurrence rate of 21%. With this consideration, there is a significantly lower rate of 14% (p = .005) for this cohort.

Inappropriate NAAT requesting and/or interpretation is not unique to this hospital, and reflects the challenges of contemporary medical practice. As this was not a systematic investigation of *C. difficile* colonisation, the analysis and interpretation has been restricted to the available information. The overall prevalence of colonisation and progression to CDI cannot be inferred. However, our findings were consistent across the 3-year period, and it is likely to present a reasonable reflection of practices within our hospital. We believe these findings have international relevance by identifying interventions that can improve future practice, and identify how reporting of CDI rates can be made more accurate.

In Chapter 4, risk factors were also evaluated for patients who were most likely to have nosocomial acquisition of *C. difficile*, by virtue of the genomic similarity of their *C.*

difficile isolates to other nosocomial isolates, compared to patients whose *C. difficile* isolates had no such genomic similarities. In this analysis, patients with inferred nosocomial acquisition were statistically older (mean age 74 vs. 67 years, p = .005), more likely to be discharged to a long-term care facility (51% vs. 29%, p = .001), more likely to be admitted by the clinical speciality, Care of the Elderly (35% vs. 15%, p = .0001), and less likely to have received immunosuppressant medication (16% vs. 45%, p = .00001). Although not reaching statistical significance, the mean length-of-stay was greater for patients without inferred nosocomial admission, at 152 vs. 125 days. This indicates that nosocomial acquisition is not simply a consequence of the duration of admission. Nosocomial acquisition could reflect greater frailty of the patients based on genomic similarities of their *C. difficile* isolates. Measures of frailty were not performed or recorded as part of this cohort study, but could be considered within future epidemiological projects.

9.3.2. *Genomic Epidemiology of Recurrent CDI*

Although not part of the prospective cohort study, the preliminary study included in Chapter 3 provided valuable context of recent genomic epidemiology related to RCDI within SJH. PCR ribotype information provided some discriminatory information, such as the demonstration of a molecular re-infection despite the clinical features of an episode of RCDI; analysis and comparison of SNPs between genomes allowed for much greater resolution of potential transmission events [117]. This was not unexpected, as there was emerging evidence of the utility of whole genome sequence analysis [32, 62]. The exceptional duration of those patients' admission was highlighted in this paper, which may have been associated with underlying frailty [117]. In retrospect, there appears to have been a relatively high yield of transmission events recognised for a small number of patients and clinical episodes [117]. This may have been influenced by the patients exceptional length of stay.

The role of RCDI in transmission of *C. difficile* was also evident in the findings of the outbreak associated with a unique sequence type of *C. difficile*, ST-295 [148]. This outbreak report is presented in Chapter 7. Three of the seven affected patients

experienced RCDI. Didelot *et al.* have established the anticipated rate of mutations within *C. difficile* as 1–2 mutations per genome per annum [63]. As Patient 1 of this outbreak experienced RCDI, a single mutation occurred within the housekeeping gene of his *C. difficile* isolates, *atpA*, which created a new sequence type, ST-295. A ward-based transmission is a plausible explanation for Patient 2's acquisition of this isolate, despite the prolonged period (16 weeks) of incubation and/or carriage before his first episode of CDI. Patient 2 had six episodes of CDI in total, all associated with *C. difficile* ST-295. His experience of symptoms while resident in the hostel can account for the introduction of this *C. difficile* sequence type to the hostel. Patient 4 had three episodes of CDI; his admissions for the management of recurrent CDI inadvertently led to two further episodes of nosocomial transmission [148].

The *C. difficile* ST-295 isolates associated with this outbreak formed a distinctive subgroup related to ST-2 isolates of the prospective cohort study. In total, there were 335 isolates with whole genome sequence data, with 39 isolates (12%) identified as ST-2 or ST-295. The ST-2 isolates were identified throughout the cohort study: 8 during 2013, 15 during 2014, and 8 during 2015. No ST-2 or ST-295 isolates were identified during January – August 2016, when this prospective cohort reached its planned cessation point.

The details of this outbreak also support the hypothesis of McLure *et al.*, who proposed from their modelling work that nosocomial CDI could not be sustained without multiple introductions of *C. difficile* from the community [142]. While the hostel appeared to have been a community reservoir for this outbreak, this investigation describes the findings pertaining to a relatively small outbreak and may not reflect the overall patterns of transmission within the hospital [148].

Within the prospective cohort study, most patients with RCDI experienced a single episode of recurrence. Many of these episodes occurred within eight weeks of the earlier episode. There were also three patients with clinical episodes suggestive of RCDI, but re-infection could only be recognised by the SNP differences between their respective *C. difficile* isolates. One patient had two ST-2 isolates, with 28 SNP differences. The second

patient had two ST-77 isolates, with 30 SNP differences, and the third had two ST-12 isolates with 2683 SNP differences. In addition, a fourth patient had two ST-98 isolates with 8 SNP differences, so a conclusion of genomic recurrence versus re-infection could not be reached. The epidemiological criterion of a maximum interval of eight weeks between an initial episode of CDI and a recurrence was established prior to the availability of molecular sub-typing of *C. difficile* [24]. Other contemporary publications have also described findings of genomic recurrence which have occurred later than eight weeks after CDI [31, 74]. It would be difficult to update this criterion in clinical practice without the comprehensive availability of next generation sequence analysis for clinically significant *C. difficile* genomes. Although there is more widespread availability of analytical tools, there could be differences between the findings of analysis of SNP versus core genome MLST analysis [210].

9.3.3. Genomic Epidemiology of CDI within a Tertiary Hospital

As described in Chapter 4, there were 335 successfully sequenced isolates of *C. difficile*, pertaining to faecal samples of 302 patients. There was considerable diversity amongst these isolates, with identification of 41 different sequence types. There were more faecal samples relating to inpatient admissions (66%) than samples relating to attendance at the hospital out-patient clinics, day ward facilities, or General Practice (28%).

When isolates with only 0–2 SNP differences that were identified within 90 days of each other were analysed with corresponding epidemiological data, there were 45 occurrences of two or more related isolates. These identified plausible transmission events for 25/45 (55%). Clinical details could not account for 20/45 (45%) occurrences, which are subsequently described as cryptic transmission events. There is significant overlap of the *C. difficile* sequence types associated with plausible and cryptic transmission events.

Of the plausible transmission events, many (56%) related to admissions under the care of the same speciality, some (28%) to admissions on the same hospital ward, and a lesser number (16%) to admissions under both the care of the same speciality and

presence on the same ward. This is in some contrast to other publications, where likely transmissions have most often been associated with inpatient wards, rather than associations with clinical staff [31, 62, 73, 74].

When all isolates with only 0–2 SNP differences were analysed with the epidemiological data, an additional eight occurrences of related isolates were identified. Seven of these could represent re-introduction of the specific *C. difficile* isolates to the hospital following community-based acquisition. Although one of these occurrences was associated with two patients under the care of one clinical specialty, this could be a false positive identification of a plausible transmission event. The clinical episodes of CDI in question were separated by 317 days, and the associated sequence type (ST-3) accounted for 3% of isolates.

In general, representation of sequence types associated with RCDI, plausible and cryptic transmission events reflected its prevalence in this cohort study. The most striking findings regarding the prevalence of *C. difficile* sequence types within this cohort are the exceptionally low prevalence of ST-1/ribotype 027 (<1%), and the dominance of ST-11/ribotype 078 (17%). This may reflect the nature of antibiotics received by patients with a positive result of *C. difficile* NAAT of faecal sample; they received significantly more piperacillin/tazobactam and IV vancomycin compared to the overall hospital antibiotic consumption records. Dingle *et al.* have demonstrated the effects of reducing fluoroquinolone prescriptions on the English incidence of CDI associated with ST-1/ribotype 027 [149]. As such, it is likely that the molecular epidemiology of CDI within this hospital is influenced by local patterns of antibiotic use.

The observed transmission rate of 21% for this hospital for the period of this cohort study is comparable to the rates within the range of 11–27% reported elsewhere [31, 70, 73, 74]. Eyre *et al.*'s landmark investigation of Oxford CDI occurred during a time of greater prevalence of ST-1/ribotype 027 associated infection in England [62, 149]. It is therefore possible that Eyre's initial finding of 35% of linked isolates reflects that higher prevalence of ST-1/ribotype 027 *C. difficile* [62, 149]. In contrast, the observed

transmission rate of 21% in this prospective cohort may be a more stable finding, given the uniquely low prevalence of ST-1/ribotype 027.

Limitations of this prospective cohort study include the absence of investigation of asymptomatic colonisation of inpatients and/or staff, which may have reduced the overall findings of transmission.

9.3.4. Consideration of Community-Acquired CDI Isolates

Although 28% of the positive NAAT results from the diagnostic laboratory of SJH did not represent faecal samples of inpatients, this was insufficient to classify those episodes as CA CDI within the prospective cohort study. (Medical records were not available for this purpose).These patients had attended either outpatients or day case procedures at SJH, or a general practice within the hospital's catchment area. If their medical records been available for review, some may have met criteria of CA CDI. However, twenty isolates were made available from a national point prevalence study of CDI conducted in April– June 2014, regarding cases reported to the Health Protection and Surveillance Centre as CA CDI. Seventeen of these isolates were successfully recovered from bacterial stocks, and had whole genome sequence data for comparison with the isolates sourced from this hospital.

Surprisingly, two of the CA CDI isolates had only 0–2 SNP differences with nosocomial isolates relating to patients within the prospective cohort. Both findings reflected common sequence types within the prospective cohort, namely ST-11/ribotype 078 and ST-2/ribotype 014–020. It is possible that these findings may represent additional cryptic transmission events, which could ultimately be associated with community-based reservoirs. However, as no epidemiological data was made available for this doctoral study, I am reluctant to assign inferences in this regard. Otherwise, 5/11 sequence types of the CA CDI isolates were not associated with any RCDI, plausible, or cryptic transmission events identified within the prospective cohort.

9.3.5. *CDI as a One Health Issue*

In addition to the recognition of CA CDI, CDI has emerged as a 'One Health' issue, as *C. difficile* can also cause infection in animals [197]. During the course of this research, *C. difficile* isolates obtained at porcine necropsies at the Central Veterinary Research Laboratory were made available to this study. These isolates were initially subject to PCR ribotype analysis, with identification of ribotypes comparable to those associated with the nosocomial isolates of the prospective cohort study. The PCR ribotypes included 078, 110 and 014–020 amongst others. This work regarding *C. difficile* ribotype association with typhlocolitis in pigs was the first such case report regarding Irish pigs [197].

9.3.6. *Complexity of Transmission of C. difficile ST-11/ribotype 078*

As indicated, *C. difficile* ST-11/Ribotype 078 was the single most common sequence type identified in the prospective cohort study, accounting for 56/335 (17%) of isolates. It was a well-represented sequence type within the relatively small number of CA CDI isolates made available for inclusion within this project (3/17). In addition, whole genome sequence analysis was completed for twenty isolates of veterinary porcine origin, after the initial PCR ribotype analysis.

Our aim was to investigate these ST-11/Ribotype 078 isolates, of Irish nosocomial, community and porcine origin, by next-generation sequence analysis for a more precise assessment of genomic evolution and relationships than PCR ribotype detail can provide [32]. Access to more sequences was required to place this analysis within a broader context. Genome sequence data of 31 nosocomial ST-11/ribotype 078 isolates, originating from nine European countries, was available from the EUCLID study [147]. Other ST-11/ribotype 078 genome sequences were available from Knetsch *et al.*'s investigation of porcine (n = 20), farmers (n = 16) and clinical (n = 31) isolates from the Netherlands [66]. The EUCLID analysis did not show clustering of those ST-11 isolates by country of origin, while Knetsch *et al.* found clusters relating to particular farms, and other findings suggestive of transmission across a wider region of the Netherlands [66, 147].

As presented in Chapter 8, we corroborated the SNP-based analysis of the earlier works of the EUCLID and Dutch isolates, and revealed additional linked isolates, i.e. pairs with 0–2 SNP differences. Regarding the nosocomial isolates of the prospective cohort, 6/10 genomically related pairs were considered to represent plausible transmission events, with the remaining 4/10 pairs as cryptic transmission events. There were 12 pairs of links within the Irish porcine isolates, associated within the three source farms. There were no findings of transmission across these farms, only within their respective herds.

More intriguingly, there were three pairs of related isolates that each comprised an Irish porcine and an Irish nosocomial isolate from the prospective cohort study. None of these pig farms were in Dublin, and these three patients were all resident in Dublin with no likely occupational and/or residential connections, so there is no clear explanation how they have acquired *C. difficile* isolates identical to those pigs.

While the EUCLID analysis had identified a small number of linked ST-11 isolates of similar national origin, this analysis revealed two pairs of linked isolates of international origin. This included one Irish CA CDI isolate, genomically identical to an Italian EUCLID isolate, and one Irish nosocomial isolate, genomically identical to a EUCLID isolate of UK origin.

Although not included in the published manuscript paper, there was a small number of Irish and Dutch porcine isolates with 3–10 SNP differences between them. Allowing for the three-six years between the collection of these isolates, and the accepted rate of 1–2 mutations per *C. difficile* genome per annum, this could provide further support of international transmission of ST-11/ ribotype 078 isolates [13, 63, 66, 147]. The findings of this analysis of *C. difficile* ST-11/ribotype 078 isolates highlights the genetic similarities associated with nosocomial, community and veterinary infections, and supports the interplay between both nosocomial and community CDI, and CDI as a 'One Health' issue.

9.3.7. GI Microbiome Signature of Risk for C. difficile Recurrence

Re-evaluation of patient and sample recruitment was required to ensure that the minimum recruitment target of sixteen samples per patient group would be met. This was performed with emerging literature regarding the stability of faecal samples stored at -80°C, and the feasibility of using the remainder of the samples submitted for diagnostic evaluation as part of routine clinical practice [172, 175]. This change of recruitment plan was approved by the institutional Research Ethics Committee, and recruitment targets were met.

Microbiome profiles were successfully obtained from 70 faecal samples and analysed, representing 41 patients with CDI and 29 patients with RCDI. While statistical conditions for completing alpha diversity measures were met, there were no significant differences between these two patient groups. As all patients had clinically significant CDI at the time of faecal sample submission for diagnostic testing, it is perhaps not surprising that in this clinical condition associated with dysbiosis, the alpha diversity measures of the richness and evenness of bacterial taxa could not separate a first episode from recurrence [104, 173].

Separation of these groups was assessed by the principal co-ordinate analysis methodology of assessing beta diversity, or the differences between microbial communities of different samples, as indicated by operational taxonomic units [172, 173]. Here, the magnitude of axial separation was comparable to that demonstrated by Seekatz *et al.* [172]. However, the anosim statistical test of these samples suggested that there were greater differences within the two groups of samples, rather than between the two samples, which could explain why a greater magnitude of separation could not be obtained.

Comparison of OTU content, and ultimately the bacterial species content of these samples, revealed statistically significant differences of five species by Mann-Whitney U comparison of the two groups, and two species by the DeSeq package analysis for differential abundance. As the two bacterial species identified by the latter method were

also amongst the significant differences identified by the former test, these two species show the most promise as potential microbiome signature of risk of recurrent CDI. These two species are *Bacteroides uniformis* and *Bacteroides xylanisolvens*, both of which have been considered elsewhere for their potential as probiotics to correct dysbiosis [185, 186]. Both species were significantly less abundant in the samples pertaining to RCDI.

This work, as described in Chapter 6, adds to the knowledge of dysbiosis in CDI, especially regarding the potential distinct microbiome composition of RCDI in a real-world clinical setting. However, further work would be required to explore the underlying mechanisms of dysbiosis, and the contributions of both the metabolic activity of this altered state and interactions with the host immune system.

9.4. Conclusion

The establishment of this prospective study of *C. difficile* at a tertiary hospital has allowed for the realisation of both a well-defined cohort of patients and their associated *C. difficile* isolates. This work was enhanced by the provision of the hospital antibiotic consumption data for the duration of the study, the provision of *C. difficile* isolates of community origin, isolates of veterinary origin and the availability of *C. difficile* sequence data regarding European nosocomial infection and Dutch porcine/farm origin, none of which were anticipated at the outset. The research presented in this thesis extends the available knowledge of the key questions posed throughout, and will serve as a basis for further studies regarding *C. difficile* epidemiology.

Bibliography

1.   Carroll, K.C. and J.G. Bartlett, *Biology of Clostridium difficile: implications for epidemiology and diagnosis.* Annu Rev Microbiol, 2011. **65**: p. 501-21.

2.   Ananthakrishnan, A., *Clostridium difficile infection: epidemiology, risk factors and management.* Nat Rev Gastroenterol Hepatol, 2011. **8**: p. 17 - 26.

3.   Knight, D.R. and T.V. Riley, *Genomic Delineation of Zoonotic Origins of Clostridium difficile.* Front Public Health, 2019. **7**: p. 164.

4.   Malamou-Ladas, H., *et al.*, *Isolation of Clostridium difficile from patients and the environment of hospital wards.* J Clin Pathol, 1983. **36**(1): p. 88-92.

5.   Testore, G.P., *et al.*, *Evidence for cross-infection in an outbreak of Clostridium difficile-associated diarrhoea in a surgical unit.* J Med Microbiol, 1988. **26**(2): p. 125-8.

6.   Permpoonpattana, P., *et al.*, *Surface layers of Clostridium difficile endospores.* J Bacteriol, 2011. **193**(23): p. 6461-70.

7.   Owens, R.C., Jr., *et al.*, *Antimicrobial-associated risk factors for Clostridium difficile infection.* Clin Infect Dis, 2008. **46 Suppl 1**: p. S19-31.

8.   Cohen, S.H., *et al.*, *Clinical practice guidelines for Clostridium difficile infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA).* Infect Control Hosp Epidemiol, 2010. **31**(5): p. 431-55.

9.   Otete, E.H., *et al.*, *Parameters for the mathematical modelling of Clostridium difficile acquisition and transmission: a systematic review.* PLoS One, 2013. **8**(12): p. e84224.

10.   CDC, *Antibiotic Resistance Threats in the United States, 2019.* 2019: Atlanta, GA: USA

11.   Centre, H.P.a.S., *<Surveillance, Diagnosis and Management of C. difficile Infection in Ireland Update of 2008 Guidance>.* 2013.

12.   Centre, H.P.a.S., *Enhanced Surveillance of Clostridioides (Clostridium) difficile Infection: Ireland - Q2 2019 National Report* H.P.a.S. Centre, Editor. 2019. p. 1-7.

13.   Eyre, D.W., *et al.*, *Whole Genome Sequencing Demonstrates Fidaxomicin is Superior to Vancomycin for Prevention of Clostridium difficile Relapse and Reinfection.* J Infect Dis, 2013.

14.   Furuya-Kanamori, L., *et al.*, *Asymptomatic Clostridium difficile colonization: epidemiology and clinical implications.* BMC Infect Dis, 2015. **15**: p. 516.

15.   Alasmari, F., *et al.*, *Prevalence and Risk Factors for Asymptomatic Clostridium difficile Carriage.* Clinical Infectious Diseases, 2014. **59**(2): p. 216-222.

16.   Dubberke, E.R., *et al.*, *Risk Factors for Acquisition and Loss of Clostridium difficile Colonization in Hospitalized Patients.* Antimicrob Agents Chemother, 2015. **59**(8): p. 4533-43.

17.   Crobach, M.J.T., *et al.*, *Understanding Clostridium difficile Colonization.* Clinical microbiology reviews, 2018. **31**(2): p. e00021-17.

18.   Prasad, N., *et al.*, *Asymptomatic rectal colonization with carbapenem-resistant Enterobacteriaceae and Clostridium difficile among residents of a long-term care facility in New York City.* Am J Infect Control, 2016. **44**(5): p. 525-32.

19.   Khanna, S., *et al.*, *Outcomes in community-acquired Clostridium difficile infection.* Alimentary pharmacology & therapeutics, 2012. **35**(5): p. 613-618.

20.   England, P.H., *Updated guidance on the management and treatment of Clostridium difficile infection.* London, 2013. **Department of Health**.

21.   Debast, S.B., *et al.*, *European Society of Clinical Microbiology and Infectious Diseases: update of the treatment guidance document for Clostridium difficile infection.* Clin Microbiol Infect, 2014. **20 Suppl 2**: p. 1-26.

22.   McDonald, L.C., *et al.*, *Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA).* Clin Infect Dis, 2018. **66**(7): p. 987-994.

23.   Khanafer, N., *et al.*, *Factors predictive of severe Clostridium difficile infection depend on the definition used.* Anaerobe, 2016. **37**: p. 43-8.

24.   McDonald, L.C., *et al.*, *Recommendations for surveillance of Clostridium difficile-associated disease.* Infect Control Hosp Epidemiol, 2007. **28**(2): p. 140-5.

25.   Lessa, F.C., *et al.*, *Burden of Clostridium difficile Infection in the United States.* New England Journal of Medicine, 2015. **372**(9): p. 825-834.

26.   Centre, H.P.a.S., *Clostridioides difficile infection in Ireland 2018*, H. HPSC, Editor. 2019 Dublin

27.   Chen, Y., *et al.*, *Molecular characteristics of Clostridium difficile strains from patients with a first recurrence more than 8 weeks after the primary infection.* J Microbiol Immunol Infect, 2017. **50**(4): p. 532-536.

28.   Barbut, F., *et al.*, *Epidemiology of recurrences or reinfections of Clostridium difficile-associated diarrhea.* J Clin Microbiol, 2000. **38**(6): p. 2386-8.

29.   Kamboj, M., *et al.*, *Relapse versus reinfection: surveillance of Clostridium difficile infection.* Clin Infect Dis, 2011. **53**(10): p. 1003-6.

30.   Eyre, D.W., *et al.*, *Predictors of first recurrence of Clostridium difficile infection: implications for initial management.* Clin Infect Dis, 2012. **55 Suppl 2**: p. S77-87.

31.   Martin, J.S.H., *et al.*, *Patient and Strain Characteristics Associated With Clostridium difficile Transmission and Adverse Outcomes.* Clin Infect Dis, 2018. **67**(9): p. 1379-1387.

32.   Knetsch, C.W., *et al.*, *Current application and future perspectives of molecular typing methods to study Clostridium difficile infections.* Euro Surveill, 2013. **18**(4): p. 20381.

33.   Chang, J.Y., *et al.*, *Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea.* J Infect Dis, 2008. **197**(3): p. 435-8.

34.   Abujamel, T., *et al.*, *Defining the vulnerable period for re-establishment of Clostridium difficile colonization after treatment of C. difficile infection with oral vancomycin or metronidazole.* PLoS One, 2013. **8**(10): p. e76269.

35.   Louie, T.J., *et al.*, *Differences of the Fecal Microflora With Clostridium difficile Therapies.* Clin Infect Dis, 2015. **60 Suppl 2**: p. S91-7.

36.   Abou Chakra, C.N., J. Pepin, and L. Valiquette, *Prediction tools for unfavourable outcomes in Clostridium difficile infection: a systematic review.* PLoS One, 2012. **7**(1): p. e30258.

37.   Fekety, R., *et al.*, *Recurrent Clostridium difficile diarrhea: characteristics of and risk factors for patients enrolled in a prospective, randomized, double-blinded trial.* Clin Infect Dis, 1997. **24**(3): p. 324-33.

38.   Appaneal, H.J., *et al.*, *Predictors of Clostridioides difficile recurrence across a national cohort of veterans in outpatient, acute, and long-term care settings.* Am J Health Syst Pharm, 2019. **76**(9): p. 581-590.

39.   Reilly, B.M. and A.T. Evans, *Translating clinical research into clinical practice: impact of using prediction rules to make decisions.* Ann Intern Med, 2006. **144**(3): p. 201-9.

40.   Hu, M.Y., *et al.*, *Prospective derivation and validation of a clinical prediction rule for recurrent Clostridium difficile infection.* Gastroenterology, 2009. **136**(4): p. 1206-14.

41.   van Rossen, T.M., *et al.*, *External validation of two prediction tools for patients at risk for recurrent Clostridioides difficile infection.* Therap Adv Gastroenterol, 2021. **14**: p. 1756284820977385.

42.   Olsen, M.A., *et al.*, *Recurrent Clostridium difficile infection is associated with increased mortality.* Clin Microbiol Infect, 2015. **21**(2): p. 164-70.

43.   Tang-Feldman, Y., *et al.*, *Molecular analysis of Clostridium difficile strains isolated from 18 cases of recurrent clostridium difficile-associated diarrhea.* J Clin Microbiol, 2003. **41**(7): p. 3413-4.

44.   Dial, S., *et al.*, *Risk of Clostridium difficile diarrhea among hospital inpatients prescribed proton pump inhibitors: cohort and case-control studies.* CMAJ, 2004. **171**(1): p. 33-8.

45.   Bauer, M.P., *et al.*, *Renal failure and leukocytosis are predictors of a complicated course of Clostridium difficile infection if measured on day of diagnosis.* Clin Infect Dis, 2012. **55 Suppl 2**: p. S149-53.

46.   Marsh, J.W., *et al.*, *Association of relapse of Clostridium difficile disease with BI/NAP1/027.* J Clin Microbiol, 2012. **50**(12): p. 4078-82.

47.   Doh, Y.S., *et al.*, *Long-Term Clinical Outcome of Clostridium difficile Infection in Hospitalized Patients: A Single Center Study.* Intest Res, 2014. **12**(4): p. 299-305.

48.   Ramanathan, S., *et al.*, *Recurrence of Clostridium difficile infection among veterans with spinal cord injury and disorder.* Am J Infect Control, 2014. **42**(2): p. 168-73.

49.   Zarowitz, B.J., *et al.*, *Risk Factors, Clinical Characteristics, and Treatment Differences Between Residents With and Without Nursing Home- and Non-Nursing*

Home-Acquired Clostridium difficile Infection. J Manag Care Spec Pharm, 2015. **21**(7): p. 585-95.

50. DiDiodato, G. and L. McArthur, *Evaluating the Effectiveness of an Antimicrobial Stewardship Program on Reducing the Incidence Rate of Healthcare-Associated Clostridium difficile Infection: A Non-Randomized, Stepped Wedge, Single-Site, Observational Study.* PLoS One, 2016. **11**(6): p. e0157671.

51. Durham, D.P., *et al.*, *Quantifying Transmission of Clostridium difficile within and outside Healthcare Settings.* Emerg Infect Dis, 2016. **22**(4): p. 608-16.

52. Hensgens, M.P., *et al.*, *Clostridium difficile infection in an endemic setting in the Netherlands.* Eur J Clin Microbiol Infect Dis, 2011. **30**(4): p. 587-93.

53. Leonard, A.D., K.M. Ho, and J. Flexman, *Proton pump inhibitors and diarrhoea related to Clostridium difficile infection in hospitalised patients: a case-control study.* Intern Med J, 2012. **42**(5): p. 591-4.

54. Howell, M.D., *et al.*, *Iatrogenic gastric acid suppression and the risk of nosocomial Clostridium difficile infection.* Arch Intern Med, 2010. **170**(9): p. 784-90.

55. Hung, Y.P., *et al.*, *Proton-Pump Inhibitor Exposure Aggravates Clostridium difficile-Associated Colitis: Evidence From a Mouse Model.* J Infect Dis, 2015. **212**(4): p. 654-63.

56. Stewart, D.B., Sr. and J.P. Hegarty, *Correlation between virulence gene expression and proton pump inhibitors and ambient pH in C. difficile: Results of an in vitro study.* J Med Microbiol, 2013.

57. Seto, C.T., *et al.*, *Prolonged use of a proton pump inhibitor reduces microbial diversity: implications for susceptibility.* Microbiome, 2014. **2**: p. 42.

58. Imhann, F., *et al.*, *Proton pump inhibitors affect the gut microbiome.* Gut, 2016. **65**(5): p. 740-8.

59.   Fawley, W.N., *et al.*, *An enhanced DNA fingerprinting service to investigate potential Clostridium difficile infection case clusters sharing the same PCR ribotype.* Journal of clinical microbiology, 2011. **49**(12): p. 4333-4337.

60.   He, M., *et al.*, *Emergence and global spread of epidemic healthcare-associated Clostridium difficile.* Nat Genet, 2013. **45**(1): p. 109-13.

61.   Eyre, D.W., *et al.*, *Comparison of Multilocus Variable-Number Tandem-Repeat Analysis and Whole-Genome Sequencing for Investigation of Clostridium difficile Transmission.* Journal of Clinical Microbiology, 2013. **51**(12): p. 4141-4149.

62.   Eyre, D.W., *et al.*, *Diverse Sources of C. difficile Infection Identified on Whole-Genome Sequencing.* New England Journal of Medicine, 2013. **369**(13): p. 1195-1205.

63.   Didelot, X., *et al.*, *Microevolutionary analysis of Clostridium difficile genomes to investigate transmission.* Genome Biol, 2012. **13**(12): p. R118.

64.   Stimson, J., *et al.*, *Beyond the SNP Threshold: Identifying Outbreak Clusters Using Inferred Transmissions.* Molecular Biology and Evolution, 2019. **36**(3): p. 587-603.

65.   Control, E.C.f.D.P.a., *ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations 2019-2021*. 2019, ECDC Stockholm.

66.   Knetsch, C., *et al.*, *Whole genome sequencing reveals potential spread of Clostridium difficile between humans and farm animals in the Netherlands, 2002 to 2011.* Euro Surveill, 2014. **19**(45).

67.   Kumar, N., *et al.*, *Genome-Based Infection Tracking Reveals Dynamics of Clostridium difficile Transmission and Disease Recurrence.* Clinical Infectious Diseases, 2015. **62**(6): p. 746-752.

68.   Knight, D.R., *et al.*, *Genome Analysis of Clostridium difficile PCR Ribotype 014 Lineage in Australian Pigs and Humans Reveals a Diverse Genetic Repertoire and Signatures of Long-Range Interspecies Transmission.* Front Microbiol, 2016. **7**: p. 2138.

69.   Stoesser, N., *et al.*, *Epidemiology of Clostridium difficile in infants in Oxfordshire, UK: Risk factors for colonization and carriage, and genetic overlap with regional C. difficile infection strains.* PLoS One, 2017. **12**(8): p. e0182307.

70.   Eyre, D.W., *et al.*, *Comparison of Control of Clostridium difficile Infection in Six English Hospitals Using Whole-Genome Sequencing.* Clin Infect Dis, 2017. **65**(3): p. 433-441.

71.   Kociolek, L.K., *et al.*, *Clostridium difficile Whole Genome Sequencing Reveals Limited Transmission Among Symptomatic Children: A Single-Center Analysis.* Clin Infect Dis, 2018. **67**(2): p. 229-234.

72.   Moradigaravand, D., *et al.*, *Genomic survey of Clostridium difficile reservoirs in the East of England implicates environmental contamination of wastewater treatment plants by clinical lineages.* Microb Genom, 2018. **4**(3).

73.   Eyre, D.W., *et al.*, *WGS to determine the extent of Clostridioides difficile transmission in a high incidence setting in North Wales in 2015.* J Antimicrob Chemother, 2019. **74**(4): p. 1092-1100.

74.   Garcia-Fernandez, S., *et al.*, *Whole-genome sequencing reveals nosocomial Clostridioides difficile transmission and a previously unsuspected epidemic scenario.* Sci Rep, 2019. **9**(1): p. 6959.

75.   Vonberg, R.P., *et al.*, *Infection control measures to limit the spread of Clostridium difficile.* Clin Microbiol Infect, 2008. **14 Suppl 5**: p. 2-20.

76.   Khanafer, N., *et al.*, *Hospital management of Clostridium difficile infection: a review of the literature.* Journal of Hospital Infection.

77.   Yakob, L., *et al.*, *Clostridium difficile exposure as an insidious source of infection in healthcare settings: an epidemiological model.* BMC Infect Dis, 2013. **13**(1): p. 376.

78.   Hirschhorn, L.R., *et al.*, *Epidemiology of community-acquired Clostridium difficile-associated diarrhea.* J Infect Dis, 1994. **169**(1): p. 127-33.

79.   Chitnis, A.S., *et al.*, *Epidemiology of Community-Associated Clostridium difficile Infection, 2009 Through 2011.* JAMA Intern Med, 2013. **173**(14): p. 1359-67.

80.   Taori, S.K., *et al.*, *A prospective study of community-associated Clostridium difficile infections: The role of antibiotics and co-infections.* J Infect, 2014.

81.   Shaughnessy, M.K., *et al.*, *Environmental Contamination in Households of Patients with Recurrent Clostridium difficile Infection.* Appl Environ Microbiol, 2016. **82**(9): p. 2686-2692.

82.   Alam, M.J., *et al.*, *Investigation of potentially pathogenic Clostridium difficile contamination in household environs.* Anaerobe, 2014. **27**: p. 31-3.

83.   Weese, J.S., *et al.*, *Evaluation of Clostridium difficile in dogs and the household environment.* Epidemiol Infect, 2010. **138**(8): p. 1100-4.

84.   Alqumber, M.A., *Clostridium difficile in retail baskets, trolleys, conveyor belts, and plastic bags in Saudi Arabia.* Saudi Med J, 2014. **35**(10): p. 1274-7.

85.   Zidaric, V., *et al.*, *The occurrence and high diversity of Clostridium difficile genotypes in rivers.* Anaerobe, 2010. **16**(4): p. 371-5.

86.   Esfandiari, Z., *et al.*, *Isolation and characterization of Clostridium difficile in farm animals from slaughterhouse to retail stage in Isfahan, Iran.* Foodborne Pathog Dis, 2015. **12**(10): p. 864-6.

87.   Norman, K.N., *et al.*, *Survey of Clostridium difficile in retail seafood in College Station, Texas.* Food Addit Contam Part A Chem Anal Control Expo Risk Assess, 2014. **31**(6): p. 1127-9.

88.   Curry, S.R., *et al.*, *Prevalence of Clostridium difficile in uncooked ground meat products from Pittsburgh, Pennsylvania.* Appl Environ Microbiol, 2012. **78**(12): p. 4183-6.

89.   Esfandiari, Z., *et al.*, *Occurrence of Clostridium difficile in seasoned hamburgers and seven processing plants in Iran.* BMC Microbiol, 2014. **14**(1): p. 283.

90.   Janezic, S., *et al.*, *Clostridium difficile genotypes other than ribotype 078 that are prevalent among human, animal and environmental isolates.* BMC Microbiol, 2012. **12**: p. 48.

91.   Rodriguez, C., *et al.*, *Multilocus sequence typing analysis and antibiotic resistance of Clostridium difficile strains isolated from retail meat and humans in Belgium.* Food Microbiol, 2014. **42**: p. 166-71.

92.   Destoumieux-Garzón, D., *et al.*, *The One Health Concept: 10 Years Old and a Long Road Ahead.* Frontiers in Veterinary Science, 2018. **5**(14).

93.   Hensgens, M.P., *et al.*, *Clostridium difficile infection in the community: a zoonotic disease?* Clin Microbiol Infect, 2012. **18**(7): p. 635-45.

94.   Janezic, S., *et al.*, *International Clostridium difficile animal strain collection and large diversity of animal associated strains.* BMC Microbiol, 2014. **14**: p. 173.

95.   Murphy, C.P., *et al.*, *Escherichia coli and selected veterinary and zoonotic pathogens isolated from environmental sites in companion animal veterinary hospitals in southern Ontario.* Can Vet J, 2010. **51**(9): p. 963-72.

96.   Rodriguez, C., *et al.*, *Investigation of Clostridium difficile interspecies relatedness using multilocus sequence typing, multilocus variable-number tandem-repeat analysis and antimicrobial susceptibility testing.* Vet J, 2015. **206**(3): p. 349-55.

97.   Silva, R.O., *et al.*, *Clostridium difficile ribotypes in humans and animals in Brazil.* Mem Inst Oswaldo Cruz, 2015. **110**(8): p. 1062-5.

98.   Keel, K., *et al.*, *Prevalence of PCR ribotypes among Clostridium difficile isolates from pigs, calves, and other species.* J Clin Microbiol, 2007. **45**(6): p. 1963-4.

99.   Costa, M.C., *et al.*, *Prevalence and molecular characterization of Clostridium difficile isolated from feedlot beef cattle upon arrival and mid-feeding period.* BMC Vet Res, 2012. **8**: p. 38.

100.   Usui, M., *et al.*, *Genetic relatedness between Japanese and European isolates of Clostridium difficile originating from piglets and their risk associated with human health.* Front Microbiol, 2014. **5**: p. 513.

101.   Costa, M.C., *et al.*, *Epidemiology of Clostridium difficile on a veal farm: prevalence, molecular characterization and tetracycline resistance.* Vet Microbiol, 2011. **152**(3-4): p. 379-84.

102.   Zidaric, V., *et al.*, *Different antibiotic resistance and sporulation properties within multiclonal Clostridium difficile PCR ribotypes 078, 126, and 033 in a single calf farm.* Appl Environ Microbiol, 2012. **78**(24): p. 8515-22.

103.   Spigaglia, P., *et al.*, *Antibiotic resistance patterns and PCR-ribotyping of Clostridium difficile strains isolated from swine and dogs in Italy.* Anaerobe, 2015. **31**: p. 42-6.

104.   Pham, T.A. and T.D. Lawley, *Emerging insights on intestinal dysbiosis during bacterial infections.* Curr Opin Microbiol, 2014. **17**: p. 67-74.

105. Knecht, H., *et al.*, *Effects of beta-lactam antibiotics and fluoroquinolones on human gut microbiota in relation to Clostridium difficile associated diarrhea.* PLoS One, 2014. **9**(2): p. e89417.

106. McDonald, L.C., *et al.*, *Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA).* Clinical Infectious Diseases, 2018. **66**(7): p. e1-e48.

107. Ooijevaar, R.E., *et al.*, *Clinical Application and Potential of Fecal Microbiota Transplantation.* Annu Rev Med, 2019. **70**: p. 335-351.

108. Rea, M., *et al.*, *Clostridium difficile carriage in elderly subjects and associated changes in the intestinal microbiota.* J Clin Microbiol, 2012. **50**: p. 867 - 875.

109. Gupta, S., *et al.*, *Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing.* Communications Biology, 2019. **2**(1): p. 291.

110. Hudson, L.E., *et al.*, *Gleaning Insights from Fecal Microbiota Transplantation and Probiotic Studies for the Rational Design of Combination Microbial Therapies.* Clinical Microbiology Reviews, 2017. **30**(1): p. 191-231.

111. Theriot, C.M. and V.B. Young, *Interactions Between the Gastrointestinal Microbiome and Clostridium difficile.* Annu Rev Microbiol, 2015. **69**: p. 445-61.

112. Hernández, M., *et al.*, *Fecal Microbiota of Toxigenic Clostridioides difficile-Associated Diarrhea.* Front Microbiol, 2018. **9**: p. 3331.

113. Weingarden, A.R., *et al.*, *Microbiota transplantation restores normal fecal bile acid composition in recurrent Clostridium difficile infection.* Am J Physiol Gastrointest Liver Physiol, 2014. **306**(4): p. G310-9.

114.   Claesson, M.J., *et al.*, *Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions.* Nucleic Acids Res, 2010. **38**(22): p. e200.

115.   Lister, M., *et al.*, *Comparison of culture based methods for the isolation of Clostridium difficile from stool samples in a research setting.* Anaerobe, 2014. **28**: p. 226-9.

116.   Jolley, K. and M. Maiden, *BIGSdb: Scalable analysis of bacterial genome variation at the population level.* BMC Bioinformatics, 2010. **11**(1): p. 595.

117.   Mac Aogáin, M., *et al.*, *Whole-genome sequencing improves discrimination of relapse from reinfection and identifies transmission events among patients with recurrent Clostridium difficile infections.* Journal of Hospital Infection, 2015. **90**(2): p. 108-16.

118.   Wiegand, P.N., *et al.*, *Clinical and economic burden of Clostridium difficile infection in Europe: a systematic review of healthcare-facility-acquired infection.* J Hosp Infect, 2012. **81**(1): p. 1-14.

119.   Oka, K., *et al.*, *Molecular and microbiological characterization of Clostridium difficile isolates from single, relapse, and reinfection cases.* J Clin Microbiol, 2012. **50**(3): p. 915-21.

120.   He, M., *et al.*, *Evolutionary dynamics of Clostridium difficile over short and long time scales.* Proc Natl Acad Sci U S A, 2010. **107**(16): p. 7527-32.

121.   Riggs, M.M., *et al.*, *Asymptomatic carriers are a potential source for transmission of epidemic and nonepidemic Clostridium difficile strains among long-term care facility residents.* Clin Infect Dis, 2007. **45**(8): p. 992-8.

122.   Dubberke, E.R., *et al.*, *Prevalence of Clostridium difficile environmental contamination and strain variability in multiple health care facilities.* Am J Infect Control, 2007. **35**(5): p. 315-8.

123. Merrigan, M.M., *et al.*, *Prevention of fatal Clostridium difficile-associated disease during continuous administration of clindamycin in hamsters.* J Infect Dis, 2003. **188**(12): p. 1922-7.

124. Stubbs, S.L., *et al.*, *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, 1999. **37**(2): p. 461-3.

125. Li, H., *et al.*, *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

126. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform.* Bioinformatics, 2010. **26**(5): p. 589-95.

127. Dingle, K.E., *et al.*, *Clinical Clostridium difficile: clonality and pathogenicity locus diversity.* PLoS One, 2011. **6**(5): p. e19993.

128. Institute, E.a.S.R., *Activity in acute public hospitals in Ireland, annual report 2011 Report No. 45*. 2011.

129. Burns, K., *et al.*, *Infection due to C. difficile ribotype 078: first report of cases in the Republic of Ireland.* J Hosp Infect, 2010. **75**(4): p. 287-91.

130. Magnusson, C., *et al.*, *Ribotyping of Clostridium difficile strains associated with nosocomial transmission and relapses in a Swedish County.* Apmis, 2013. **121**(2): p. 153-7.

131. Bauer, M.P., *et al.*, *Clostridium difficile infection in Europe: a hospital-based survey.* Lancet, 2011. **377**(9759): p. 63-73.

132. Drudy, D., *et al.*, *Isolation and characterisation of toxin A-negative, toxin B-positive Clostridium difficile in Dublin, Ireland.* Clin Microbiol Infect, 2007. **13**(3): p. 298-304.

133. Alonso, R., *et al.*, *Molecular analysis of relapse vs re-infection in HIV-positive patients suffering from recurrent Clostridium difficile associated diarrhoea.* Journal of Hospital Infection, 2001. **48**(2): p. 86-92.

134. Pettit, L.J., *et al.*, *Functional genomics reveals that Clostridium difficile Spo0A coordinates sporulation, virulence and metabolism.* BMC Genomics, 2014. **15**: p. 160.

135. Sokurenko, E.V., D.L. Hasty, and D.E. Dykhuizen, *Pathoadaptive mutations: gene loss and variation in bacterial pathogens.* Trends Microbiol, 1999. **7**(5): p. 191-5.

136. Chotirmall, S.H., *et al.*, *Effect of estrogen on pseudomonas mucoidy and exacerbations in cystic fibrosis.* N Engl J Med, 2012. **366**(21): p. 1978-86.

137. de la Riva, L., *et al.*, *Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of Clostridium difficile.* J Bacteriol, 2011. **193**(13): p. 3276-85.

138. Smith, C.A., *Structure, function and dynamics in the mur family of bacterial cell wall ligases.* J Mol Biol, 2006. **362**(4): p. 640-55.

139. Eyre, D.W., *et al.*, *Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in Clostridium difficile transmission.* PLoS Comput Biol, 2013. **9**(5): p. e1003059.

140. Furuya-Kanamori, L., *et al.*, *Comparison of Clostridium difficile Ribotypes Circulating in Australian Hospitals and Communities.* J Clin Microbiol, 2017. **55**(1): p. 216-225.

141. Brown, K.A., *et al.*, *The Drivers of Acute and Long-term Care Clostridium difficile Infection Rates: A Retrospective Multilevel Cohort Study of 251 Facilities.* Clin Infect Dis, 2017. **65**(8): p. 1282-1288.

142.   McLure, A., *et al.*, *Healthcare-Associated Clostridium difficile Infections are Sustained by Disease from the Community.* Bull Math Biol, 2017. **79**(10): p. 2242-2257.

143.   McLure, A., *et al.*, *Clostridium difficile classification overestimates hospital-acquired infections.* J Hosp Infect, 2018. **99**(4): p. 453-460.

144.   McLure, A., *et al.*, *Modelling diverse sources of Clostridium difficile in the community: importance of animals, infants and asymptomatic carriers.* Epidemiol Infect, 2019. **147**: p. e152.

145.   McLure, A., *et al.*, *Seasonality and community interventions in a mathematical model of Clostridium difficile transmission.* J Hosp Infect, 2019. **102**(2): p. 157-164.

146.   De Silva, D., *et al.*, *Whole-genome sequencing to determine transmission of Neisseria gonorrhoeae: an observational study.* Lancet Infect Dis, 2016. **16**(11): p. 1295-1303.

147.   Eyre, D.W., *et al.*, *Two Distinct Patterns of Clostridium difficile Diversity Across Europe Indicating Contrasting Routes of Spread.* Clin Infect Dis, 2018. **67**(7): p. 1035-1044.

148.   Moloney, G., *et al.*, *Possible Interplay Between Hospital and Community Transmission of a Novel Clostridium Difficile Sequence Type 295 Recognized by Next-Generation Sequencing.* Infect Control Hosp Epidemiol, 2016. **37**(6): p. 680-4.

149.   Dingle, K.E., *et al.*, *Effects of control interventions on Clostridium difficile infection in England: an observational study.* Lancet Infect Dis, 2017. **17**(4): p. 411-421.

150.   Catanzaro, M. and J. Cirone, *Real-time polymerase chain reaction testing for Clostridium difficile reduces isolation time and improves patient management in a small community hospital.* Am J Infect Control, 2012. **40**(7): p. 663-6.

151. Brecher, S.M., S.M. Novak-Weekley, and E. Nagy, *Laboratory Diagnosis of Clostridium difficile Infections: There Is Light at the End of the Colon.* Clinical Infectious Diseases, 2013. **57**(8): p. 1175-1181.

152. Gould, C.V., *et al.*, *Effect of Nucleic Acid Amplification Testing on Population-Based Incidence Rates of Clostridium difficile Infection.* Clin Infect Dis, 2013.

153. Davies, K., *et al.*, *Variability in testing policies and impact on reported Clostridium difficile infection rates: results from the pilot Longitudinal European Clostridium difficile Infection Diagnosis surveillance study (LuCID).* Eur J Clin Microbiol Infect Dis, 2016. **35**(12): p. 1949-1956.

154. Truong, C., *et al.*, *Clostridium difficile rates in asymptomatic and symptomatic hospitalized patients using nucleic acid testing.* Diagn Microbiol Infect Dis, 2017. **87**(4): p. 365-370.

155. Truong, C.Y., *et al.*, *Real-Time Electronic Tracking of Diarrheal Episodes and Laxative Therapy Enables Verification of Clostridium difficile Clinical Testing Criteria and Reduction of Clostridium difficile Infection Rates.* J Clin Microbiol, 2017. **55**(5): p. 1276-1284.

156. Kelly, S.G., *et al.*, *Inappropriate Clostridium difficile Testing and Consequent Overtreatment and Inaccurate Publicly Reported Metrics.* Infect Control Hosp Epidemiol, 2016. **37**(12): p. 1395-1400.

157. Ilies, I., *et al.*, *Impact of molecular testing on reported Clostridoides difficile infection rates.* Infect Control Hosp Epidemiol, 2020. **41**(3): p. 306-312.

158. Dubberke, E.R., *et al.*, *Impact of clinical symptoms on interpretation of diagnostic assays for Clostridium difficile infections.* J Clin Microbiol, 2011. **49**(8): p. 2887-93.

159. Srinivasan, A., *Antibiotic stewardship: Why we must, how we can.* Cleve Clin J Med, 2017. **84**(9): p. 673-679.

160. Deshpande, A., *et al.*, *Community-associated Clostridium difficile infection and antibiotics: a meta-analysis.* Journal of Antimicrobial Chemotherapy, 2013. **68**(9): p. 1951-1961.

161. Watson, R.L. and C.J. Graber, *Lack of improvement in antimicrobial prescribing after a diagnosis of Clostridium difficile and impact on recurrence.* Am J Infect Control, 2018. **46**(12): p. 1370-1374.

162. Hecker, M.T., *et al.*, *Efficacy of a stewardship intervention focused on reducing unnecessary use of non-Clostridioides difficile antibiotics in patients with Clostridioides difficile infection.* Infect Control Hosp Epidemiol, 2019: p. 1-3.

163. McDonald, E.G., *et al.*, *Continuous Proton Pump Inhibitor Therapy and the Associated Risk of Recurrent Clostridium difficile Infection.* JAMA Intern Med, 2015. **175**(5): p. 784-91.

164. Peterson, L.R., *et al.*, *Detection of toxigenic Clostridium difficile in stool samples by real-time polymerase chain reaction for the diagnosis of C. difficile-associated diarrhea.* Clin Infect Dis, 2007. **45**(9): p. 1152-60.

165. Fabre, V., *et al.*, *Impact of Case-Specific Education and Face-to-Face Feedback to Prescribers and Nurses in the Management of Hospitalized Patients With a Positive Clostridium difficile Test.* Open Forum Infect Dis, 2018. **5**(10): p. ofy226.

166. Klindworth, A., *et al.*, *Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.* Nucleic Acids Res, 2013. **41**(1): p. e1.

167. Turnbaugh, P.J., *et al.*, *The human microbiome project.* Nature, 2007. **449**(7164): p. 804-10.

168. Jeffery, I.B., *et al.*, *Categorization of the gut microbiota: enterotypes or gradients?* Nat Rev Microbiol, 2012. **10**(9): p. 591-2.

169.   Rousseau, C., *et al.*, *Clostridium difficile colonization in early infancy is accompanied by changes in intestinal microbiota composition.* J Clin Microbiol, 2011. **49**(3): p. 858-65.

170.   Schubert, A.M., *et al.*, *Microbiome data distinguish patients with Clostridium difficile infection and non-C. difficile-associated diarrhea from healthy controls.* MBio, 2014. **5**(3): p. e01021-14.

171.   Steinway, S.N., *et al.*, *Inference of Network Dynamics and Metabolic Interactions in the Gut Microbiome.* PLoS Comput Biol, 2015. **11**(5): p. e1004338.

172.   Seekatz, A.M., *et al.*, *Dynamics of the fecal microbiome in patients with recurrent and nonrecurrent Clostridium difficile infection.* Genome Med, 2016. **8**(1): p. 47.

173.   Morgan, X.C. and C. Huttenhower, *Chapter 12: Human Microbiome Analysis.* PLOS Computational Biology, 2012. **8**(12): p. e1002808.

174.   CLARKE, K.R., *Non-parametric multivariate analyses of changes in community structure.* Australian Journal of Ecology, 1993. **18**(1): p. 117-143.

175.   Shaw, A.G., *et al.*, *Latitude in sample handling and storage for infant faecal microbiota studies: the elephant in the room?* Microbiome, 2016. **4**(1): p. 40.

176.   Morris, E.K., *et al.*, *Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories.* Ecology and Evolution, 2014. **4**(18): p. 3514-3524.

177.   Bashan, A., *et al.*, *Universality of human microbial dynamics.* Nature, 2016. **534**(7606): p. 259-62.

178.   Lozupone, C., *et al.*, *UniFrac: an effective distance metric for microbial community comparison.* The ISME journal, 2011. **5**(2): p. 169-172.

179. Manges, A., *et al.*, *Comparative metagenomic study of alterations to the intestinal microbiota and risk of nosocomial Clostridum difficile-associated disease.* J Infect Dis, 2010. **202**: p. 1877 - 1884.

180. Benítez-Páez, A., E.M. Gómez Del Pulgar, and Y. Sanz, *The Glycolytic Versatility of Bacteroides uniformis CECT 7771 and Its Genome Response to Oligo and Polysaccharides.* Front Cell Infect Microbiol, 2017. **7**: p. 383.

181. Hiippala, K., *et al.*, *Isolation of Anti-Inflammatory and Epithelium Reinforcing Bacteroides and Parabacteroides Spp. from A Healthy Fecal Donor.* Nutrients, 2020. **12**(4).

182. Gauffin Cano, P., *et al.*, *Bacteroides uniformis CECT 7771 ameliorates metabolic and immunological dysfunction in mice with high-fat-diet induced obesity.* PLoS One, 2012. **7**(7): p. e41079.

183. Mullish, B.H., *et al.*, *Microbial bile salt hydrolases mediate the efficacy of faecal microbiota transplant in the treatment of recurrent Clostridioides difficile infection.* Gut, 2019. **68**(10): p. 1791-1800.

184. Chassard, C., *et al.*, *Bacteroides xylanisolvens sp. nov., a xylan-degrading bacterium isolated from human faeces.* Int J Syst Evol Microbiol, 2008. **58**(Pt 4): p. 1008-13.

185. El Hage, R., E. Hernandez-Sanabria, and T. Van de Wiele, *Emerging Trends in "Smart Probiotics": Functional Consideration for the Development of Novel Health and Industrial Applications.* Frontiers in microbiology, 2017. **8**: p. 1889-1889.

186. Ulsemer, P., *et al.*, *Preliminary safety evaluation of a new Bacteroides xylanisolvens isolate.* Applied and environmental microbiology, 2012. **78**(2): p. 528-535.

187.   De La Cochetiere, M., *et al.*, *Effect of antibiotic therapy on human fecal microbiota and the relation to the development of Clostridium difficile.* Microb Ecol, 2008. **56**: p. 395 - 402.

188.   Gerding, D.N. and F.C. Lessa, *The Epidemiology of Clostridium difficile Infection Inside and Outside Health Care Institutions.* Infect Dis Clin North Am, 2015. **29**(1): p. 37-50.

189.   Sadowski, L.S., *et al.*, *Effect of a housing and case management program on emergency department visits and hospitalizations among chronically ill homeless adults: a randomized trial.* Jama, 2009. **301**(17): p. 1771-8.

190.   Kmietowicz, Z., *NICE advises screening for TB in hostels and prisons to reduce UK cases.* Bmj, 2012. **344**: p. e2309.

191.   Neale, J. and C. Stevenson, *Routine exposure to blood within hostel environments might help to explain elevated levels of hepatitis C amongst homeless drug users: insights from a qualitative study.* Int J Drug Policy, 2012. **23**(3): p. 248-50.

192.   Guindon, S., *et al.*, *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.* Syst Biol, 2010. **59**(3): p. 307-21.

193.   Griffiths, D., *et al.*, *Multilocus sequence typing of Clostridium difficile.* J Clin Microbiol, 2010. **48**(3): p. 770-8.

194.   *Housekeeping Manual for Municipally Operated Shelters & Hostel Services*, S.a.H.A. Toronto Shelter, Editor. 2013.

195.   Brazier, J.S., *Typing of Clostridium difficile.* Clin Microbiol Infect, 2001. **7**(8): p. 428-31.

196.   Goorhuis, A., *et al.*, *Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078.* Clin Infect Dis, 2008. **47**(9): p. 1162-70.

197.   McElroy, M.C., *et al.*, *Typhlocolitis associated with Clostridium difficile ribotypes 078 and 110 in neonatal piglets from a commercial Irish pig herd.* Irish Veterinary Journal, 2016. **69**(1): p. 10.

198.   Moloney, G., *et al.*, *Human and Porcine Transmission of Clostridioides difficile Ribotype 078, Europe.* Emerging Infectious Disease journal, 2021. **27**(9): p. 2294.

199.   Control., E.C.f.D.P.a., *Healthcare-associated infections: Clostridium difficile infections.*, in *ECDC. Annual epidemiological report for 2016. Stockholm: ECDC; 2018.* 2018.

200.   Alam, M.J., *et al.*, *Community Environmental Contamination of Toxigenic Clostridium difficile.* Open Forum Infect Dis, 2017. **4**(1): p. ofx018.

201.   al Saif, N. and J.S. Brazier, *The distribution of Clostridium difficile in the environment of South Wales.* J Med Microbiol, 1996. **45**(2): p. 133-7.

202.   Anderson, D.J., *et al.*, *Identification of novel risk factors for community-acquired Clostridium difficile infection using spatial statistics and geographic information system analyses.* PLoS One, 2017. **12**(5): p. e0176285.

203.   van Dorp, S.M., *et al.*, *Spatial clustering and livestock exposure as risk factor for community-acquired Clostridium difficile infection.* Clin Microbiol Infect, 2019. **25**(5): p. 607-612.

204.   Knetsch, C.W., *et al.*, *Zoonotic Transfer of Clostridium difficile Harboring Antimicrobial Resistance between Farm Animals and Humans.* J Clin Microbiol, 2018. **56**(3).

205.   Dharmasena, M. and X. Jiang, *Isolation of Toxigenic Clostridium difficile from Animal Manure and Composts Being Used as Biological Soil Amendments.* Appl Environ Microbiol, 2018. **84**(16).

206.   Usui, M., *et al.*, *Survival and prevalence of Clostridium difficile in manure compost derived from pigs.* Anaerobe, 2017. **43**: p. 15-20.

207.   Dingle, K.E., *et al.*, *A Role for Tetracycline Selection in Recent Evolution of Agriculture-Associated Clostridium difficile PCR Ribotype 078.* MBio, 2019. **10**(2).

208.   Weese, J.S., *Clostridium (Clostridioides) difficile in animals.* J Vet Diagn Invest, 2020. **32**(2): p. 213-221.

209.   Lawes, T., *et al.*, *Effect of a national 4C antibiotic stewardship intervention on the clinical and molecular epidemiology of Clostridium difficile infections in a region of Scotland: a non-linear time-series analysis.* Lancet Infect Dis, 2017. **17**(2): p. 194-206.

210.   Janezic, S. and M. Rupnik, *Development and Implementation of Whole Genome Sequencing-Based Typing Schemes for Clostridioides difficile.* Frontiers in Public Health, 2019. **7**(309).

Appendices

A. Case Report Form

B. 16S Amplicon metagenomic sequencing library preparation guide

C. Whole-genome sequencing improves discrimination of relapse from reinfection and identifies transmission events among patients with recurrent *C. difficile* infections

D. Possible Interplay Between Hospital and Community Transmission of a Novel *C. difficile* Sequence Type 295 Recognized by Next-Generation Sequencing

E. Typhlocolitis associated with *C. difficile* ribotypes 078 and 110 in neonatal piglets from a commercial Irish pig herd

F. Human and Porcine Transmission of *Clostridioides difficile* Ribotype 078, Europe

Case Report Form, submitted as Appendix 1 for REC approval

CASE REPORT FORM V1.1

1. Demographics	MRN_____ Surname_____ Forename_____ DOB_____ Consultant_____ Ward _____

2. C diff results
 Date of specimen
 	
 Lab no:
	Toxin PCR: Pos □ Culture: Pos □ Neg □
	 Symptomatic: Y □ N □
	        If yes, no of diarrhoeal episodes: __
	      CDI episode – new □ same episode □
	      Recurrence □
	
	 If asymptomatic, reason for test (if known) _____
	 Reason unknown □

3. CDI: Symptoms	Date of onset of symptoms: _____
	 Date of admission: _____ Acq:____
	 Time to acquisition: _____
	 Date of isolation:_____
	 Isolation to single room: Y □ N □
	 If no, reason (if known): _____
	
	 Hospital placements:
	 Ward Duration (days) Dates:
	    1. _____
	    2. _____
	    3. _____
	    4. _____
	    5. _____
	
	 TTROD (days): _____
	 Step-down of isolation precautions: Y □ N □
	 Subsequent placements:
	 Ward Duration (days) Dates:
	    1. _____
	    2. _____
	    3. _____
	    4. _____
	    5. _____

	Onset: Healthcare □ Community □ Origin: Healthcare □ Community □ Unknown □

4. CDI Episode:	Risks:
	
	Previous hospital admissions: Y □ N □
	
	▪ Time since prior admission: ____
	▪ Total number of in-patient admissions (>48 hours) in past 12 months
	LTCF resident: Y □ N □
	
	Recent antimicrobials: Y □ N □ Unk □
	
	Agent Duration (days) Dates
	
	1._____
	
	2._____
	
	3._____
	
	4._____
	
	5._____
	
	Enteral feeding: Y □ N □
	      NG □ PEG/RIG □ PEJ □
	PPI: Y □ N □
	      If yes, did use precede admission: Y □ N □
	Immunosuppressants: Y □ N □
	      If yes: agent
	      Agent/course initiated during admission: Y □ N □
	IBD: Crohns □ UC□ unspecified/other□ not IBD □
	GI surgery: Y □ N □
	      If yes, procedure:_____
	      Duration between procedure and CDI:____

5. CDI episode:	Episode: new □ same □ relapse □ Date of specimen: Fever: Y □ N □ Rigors: Y □ N □ Abdominal pain: Y □ N □ WCC >15: Y □ N □ Creat ▪ 133: Y □ N □

	▪ > 50% above baseline: Y □ N □ CT abd/pelvis: Findings of colitis Y □ N □ Ascites Y □ N □ Endoscopy: PMC Y □ N □ Complications: ▪ Hypotension: Y □ N □ ▪ Shock: Y □ N □ ▪ Rising serum lactate levels: Y □ N □ ▪ Ileus: Y □ N □ ▪ Megacolon: Y □ N □

6. CDI Mx:
(symptomatic)
	On treatment: Y □ N □ Start date: End date:
	
	Agent: Metronidazole PO □
	          Vancomycin PO □
	          Metronidazole IV □
	          Other: Agent: Route:
	
	ICU admission due to CDI:
	Duration of ICU admission: ___ Outcome: _____
	Surgical intervention attributable to CDI: Y □ N □
	
	Outcome:
	Duration of SJH admission:
	TTROD:
	Discharge:
	          Home: □
	          LTCF: □ If yes, is placement new Y □ N □
	          Other:□ Specify _____
	          Death:□

7. Antimicrobials preceding episode	(As per antibiotics outlined in 4.CDI Episode) 1. Agent ▪ Documented indication ▪ Choice in keeping with SJH guidelines ▪ Recommendation by Clinical Micro or ID ▪ AMS review ▪ Other 2. Agent ▪ Documented indication

Text code:

- Black font – data accessed through established Infection Control/ Clinical Microbiology database
- Blue font – enhanced clinical data obtained from Electronic Patient Record and medical notes

Case Definitions (1)

1. C *difficile*-associated disease(CDAD) case: ≥1 of following
   - Diarrhoeal stools or toxic megacolon, with either positive EIA for toxin in stool, or toxin-producing C. *difficile* organism detected in stool via culture
   - Pseudomembranous colitis revealed by lower gastrointestinal endoscopy
   - Colonic histopathology characteristic of C. *difficile* infection

2. Severe CDAD case: any 1 of following
   - Admission to ICU for treatment of CDAD or complications
   - Surgery for toxic megacolon, perforation or refractory colitis
   - Death within 30 days after diagnosis if CDAD primary or a contributive cause
   - Admission to healthcare facility for treatment of community associated CDAD

3. Recurrent CDAD case
- Patient with episode of CDAD that occurs within 8 weeks following onset of previous episode provided CDAD symptoms from earlier episode resolved with or without therapy

Onset of CDAD

- Healthcare: symptoms start during stay in healthcare facility

- Community: symptoms start during stay in community setting, outside healthcare facilities

Origin of CDAD

- **Healthcare-associated:**
  Healthcare-onset: symptom onset at least 48 hours following admission to healthcare facility
  Community onset: symptom onset in the community within 4 weeks following discharge from a healthcare facility

- **Community-associated**:
  Healthcare-onset: symptom onset within 48 hours following admission to a healthcare facility without residence in a healthcare facility within previous 12 weeks
  Community-onset: symptom onset while outside healthcare facility, and without discharge from a healthcare facility within previous 12 weeks

- Unknown
  CDAD patient who was discharged from healthcare facility 4-12 weeks before onset of symptoms.

(Version 1.1, July 31st 2013)

(1) Health Protection and Surveillance Centre, 'Surveillance, Diagnosis and Management of C. difficile Infection in Ireland Update of 2008 Guidance' , published 2013.

16S Metagenomic Sequencing Library Preparation

Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System

Introduction

Metagenomic studies are commonly performed by analyzing the prokaryotic 16S ribosomal RNA gene (16S rRNA), which is approximately 1,500 bp long and contains nine variable regions interspersed between conserved regions. Variable regions of 16S rRNA are frequently used in phylogenetic classifications such as genus or species in diverse microbial populations.

Which 16S rRNA region to sequence is an area of debate, and your region of interest might vary depending on things such as experimental objectives, design, and sample type. This protocol describes a method for preparing samples for sequencing the variable V3 and V4 regions of the 16S rRNA gene. This protocol can also be used for sequencing other regions with different region-specific primers. This protocol combined with a benchtop sequencing system, on-board primary analysis, and secondary analysis using MiSeq Reporter or BaseSpace, provides a comprehensive workflow for 16S rRNA amplicon sequencing.

Workflow Summary:

1   Order amplicon primers–The protocol includes the primer pair sequences for the V3 and V4 region that create a single amplicon of approximately ~460 bp. The protocol also includes overhang adapter sequences that must be appended to the primer pair sequences for compatibility with Illumina index and sequencing adapters. Illumina does not sell these primers. They must be ordered from a third party. See Amplicon Primers, on page 3 for more information on amplicon primers.

2   Prepare library–The protocol describes the steps to amplify the V3 and V4 region and using a limited cycle PCR, add Illumina sequencing adapters and dual-index barcodes to the amplicon target. Using the full complement of Nextera XT indices, up to 96 libraries can be pooled together for sequencing.

3   Sequence on MiSeq–Using paired 300-bp reads, and MiSeq v3 reagents, the ends of each read are overlapped to generate high-quality, full-length reads of the V3 and V4 region in a single 65-hour run. The MiSeq run output is approximately > 20 million reads and, assuming 96 indexed samples, can generate > 100,000 reads per sample, commonly recognized as sufficient for metagenomic surveys.

4   Analyze on MSR or BaseSpace–The Metagenomics workflow is a secondary analysis option built into the MiSeq Reporter (on-system software) or available on BaseSpace (cloud-based software). The Metagenomics Workflow performs a taxonomic classification using the Greengenes database showing genus or species level classification in a graphical format.

This protocol can be used to sequence alternative regions of the 16S rRNA gene and for other targeted amplicon sequences of interest. When using this protocol for amplicon sequencing other than 16S rRNA, use the Generate FASTQ Workflow (secondary analysis option). For more information, see MiSeq Reporter Metagenomics Workflow, on page 20.

DISCLAIMER
The information in this Illumina Demonstrated Protocol is being provided as a courtesy; in some cases reagents are required to be purchased from non-authorized third-party suppliers. Illumina does not guarantee nor promises technical support for the performance of our products used with reagents purchased from a non-authorized third-party supplier.

Figure 1   16S V3 and V4 Amplicon Workflow
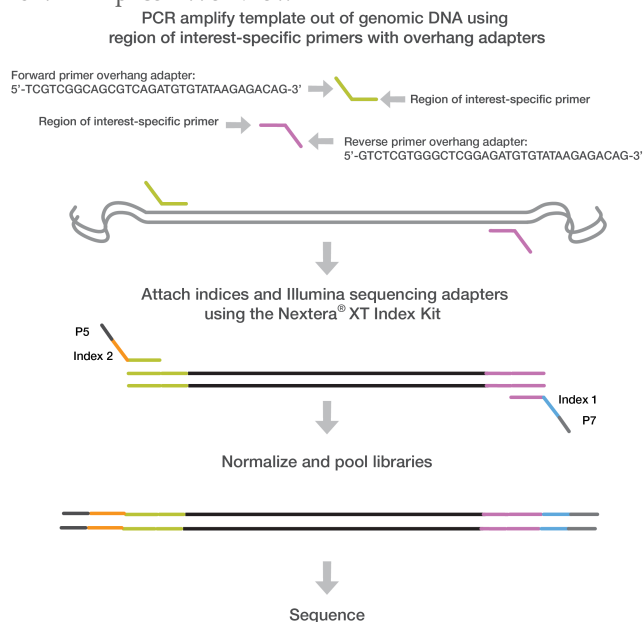
User-defined forward and reverse primers that are complementary upstream and downstream of the
  region of interest are designed with overhang adapters, and used to amplify templates from genomic
  DNA. A subsequent limited-cycle amplification step is performed to add multiplexing indices and
  Illumina sequencing adapters. Libraries are normalized and pooled, and sequenced on the MiSeq
  system using v3 reagents.

Amplicon Primers

- The gene-specific sequences used in this protocol target the 16S V3 and V4 region. They
  are selected from the Klindworth et al. publication (Klindworth A, Pruesse E, Schweer T,
  Peplles J, Quast C, et al. (2013) Evaluation of general 16S ribosomal RNA gene PCR
  primers for classical and next-generation sequencing-based diversity studies. Nucleic
  Acids Res 41(1).) as the most promising bacterial primer pair. Illumina adapter
  overhang nucleotide sequences are added to the gene-specific sequences. The full length
  primer sequences, using standard IUPAC nucleotide nomenclature, to follow the protocol
  targeting this region are:
  16S Amplicon PCR Forward Primer = 5'
  TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG
  16S Amplicon PCR Reverse Primer = 5'
  GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC
- This method can also be utilized to target other regions on the genome (either for 16S
  with other sets of primer pairs, or non-16S regions throughout the genome; ie any
  amplicon). The overhang adapter sequence must be added to the locus-specific primer
  for the region to be targeted (Figure 1). The Illumina overhang adapter sequences to be
  added to locus-specific sequences are:
  Forward overhang: 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[locus-
  specific sequence]
  Reverse overhang: 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[locus-
  specific sequence]

- The following considerations are recommended for designing other locus-specific primers:

  a Illumina recommends targeting regions that result in an amplicon that when sequenced with paired-end reads has at least ~50 bp of overlapping sequence in the middle. For example, if running 2x300 bp paired-end reads Illumina recommends having an insert size of 550 bp or smaller so that the bases sequenced at the end of each read overlap.

  b The locus-specific portion of primer (not including overhang sequence) must have a melting temperature (Tm) of 60°–65°C. You can use online PCR primer sequence analysis tools (e.g. http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/) to check the properties of primer designs. For the Tm calculation only, the gene-specific portion must be used in calculation. For hairpin and dimer calculations, the fully-assembled primer sequence (including the overhang) should be used.

  c Illumina recommends using standard desalting purification when ordering oligo primer sets.

  > **NOTE**
  > For more information on reagents used in the protocol, see Consumables and Equipment, on page 21.

16S Library Preparation Workflow

The following diagram illustrates the workflow using the 16S Library Preparation Protocol. Safe stopping points are marked between steps.

Figure 2   16S Library Preparation Workflow
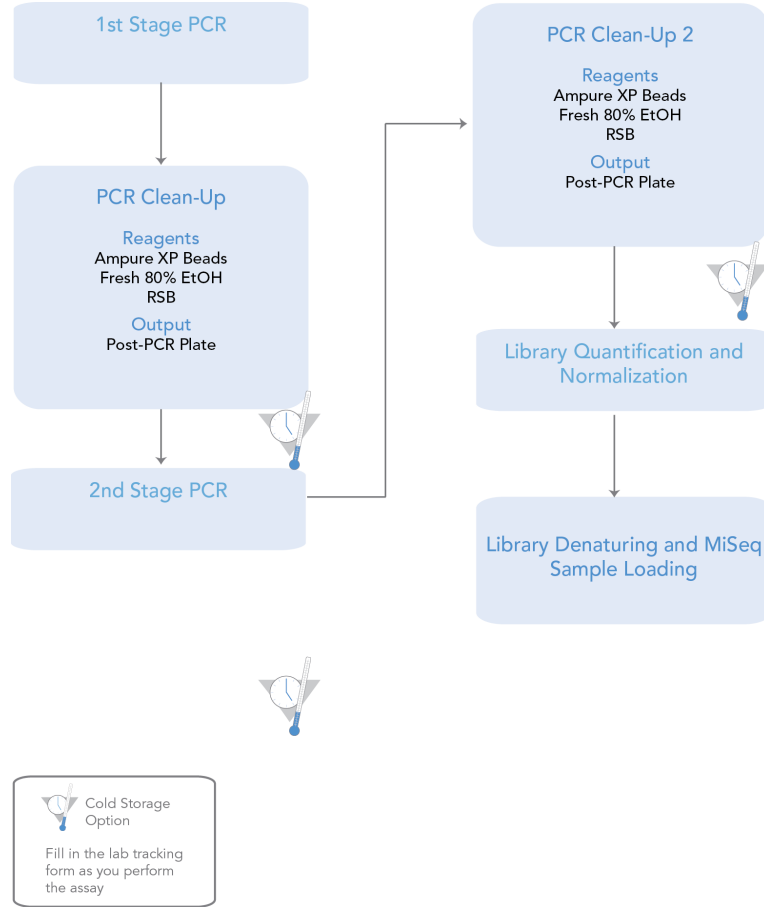
Amplicon PCR

This step uses PCR to amplify template out of a DNA sample using region of interest-specific primers with overhang adapters attached. For more information on primer sequences, see Amplicon Primers, on page 3.

Consumables

📌 **NOTE**
For more information on consumables and equipment for this protocol see Consumables and Equipment, on page 21.

Item	Quantity	Storage
Microbial Genomic DNA (5 ng/µl in 10 mM Tris pH 8.5)	2.5 µl per sample	-15° to -25°C
Amplicon PCR Reverse Primer (1 µM)	5 µl per sample	-15° to -25°C
Amplicon PCR Forward Primer (1 µM)	5 µl per sample	-15° to -25°C
2x KAPA HiFi HotStart ReadyMix	12.5 µl per sample	-15° to -25°C
Microseal 'A' film		
96-well 0.2 ml PCR plate	1 plate	
[Optional] Bioanalyzer chip (Agilent DNA 1000 kit catalog # 5067-1504)		

Procedure

1   Set up the following reaction of DNA, 2x KAPA HiFi HotStart ReadyMix, and primers:

	Volume
Microbial DNA (5 ng/µl)	2.5 µl
Amplicon PCR Forward Primer 1 µM	5 µl
Amplicon PCR Reverse Primer 1 µM	5 µl
2x KAPA HiFi HotStart ReadyMix	12.5 µl
Total	**25 µl**

2   Seal plate and perform PCR in a thermal cycler using the following program:
- 95°C for 3 minutes
- 25 cycles of:
  - 95°C for 30 seconds
  - 55°C for 30 seconds
  - 72°C for 30 seconds
- 72°C for 5 minutes
- Hold at 4°C

3   **[Optional]** Run 1 μl of the PCR product on a Bioanalyzer DNA 1000 chip to verify the size. Using the V3 and V4 primer pairs in the protocol, the expected size on a Bioanalyzer trace after the Amplicon PCR step is ~550 bp.

Figure 3   Example Bioanalyzer Trace after Amplicon PCR Step
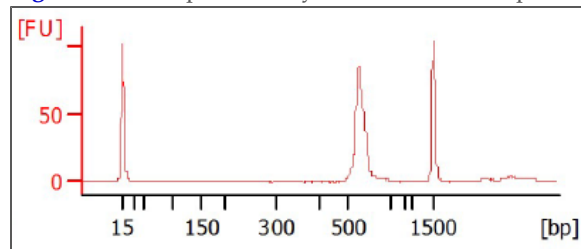
PCR Clean-Up

This step uses AMPure XP beads to purify the 16S V3 and V4 amplicon away from free primers and primer dimer species.

Consumables

Item	Quantity	Storage
10 mM Tris pH 8.5	52.5 μl per sample	-15° to -25°C
AMPure XP beads	20 μl per sample	2° to 8°C
Freshly Prepared 80% Ethanol (EtOH)	400 μl per sample	
96-well 0.2 ml PCR plate	1 plate	
[Optional] Microseal 'B' film		
[Optional] 96-well MIDI plate	1 plate	

Preparation

- Bring the AMPure XP beads to room temperature.

Procedure

1   Centrifuge the Amplicon PCR plate at 1,000 × g at 20°C for 1 minute to collect condensation, carefully remove seal.

2   **[Optional - for use with shaker for mixing]** Using a multichannel pipette set to 25 μl, transfer the entire Amplicon PCR product from the PCR plate to the MIDI plate. Change tips between samples.

> **NOTE**
> Transfer the sample to a 96-well MIDI plate if planning to use a shaker for mixing. If mixing by pipette, the sample can remain in the 96-well PCR plate.

3   Vortex the AMPure XP beads for 30 seconds to make sure that the beads are evenly dispersed. Add an appropriate volume of beads to a trough depending on the number of samples processing.

4   Using a multichannel pipette, add 20 μl of AMPure XP beads to each well of the Amplicon PCR plate. Change tips between columns.

5   Gently pipette entire volume up and down 10 times if using a 96-well PCR plate or seal plate and shake at 1800 rpm for 2 minutes if using a MIDI plate.

6   Incubate at room temperature without shaking for 5 minutes.

7   Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared.

8   With the Amplicon PCR plate on the magnetic stand, use a multichannel pipette to remove and discard the supernatant. Change tips between samples.

9   With the Amplicon PCR plate on the magnetic stand, wash the beads with freshly prepared 80% ethanol as follows:

   a   Using a multichannel pipette, add 200 μl of freshly prepared 80% ethanol to each sample well.

   b   Incubate the plate on the magnetic stand for 30 seconds.

   c   Carefully remove and discard the supernatant.

10 With the Amplicon PCR plate on the magnetic stand, perform a second ethanol wash as follows:

   a   Using a multichannel pipette, add 200 μl of freshly prepared 80% ethanol to each sample well.

   b   Incubate the plate on the magnetic stand for 30 seconds.

   c   Carefully remove and discard the supernatant.

   d   Use a P20 multichannel pipette with fine pipette tips to remove excess ethanol.

11 With the Amplicon PCR plate still on the magnetic stand, allow the beads to air-dry for 10 minutes.

12 Remove the Amplicon PCR plate from the magnetic stand. Using a multichannel pipette, add 52.5 μl of 10 mM Tris pH 8.5 to each well of the Amplicon PCR plate.

13 Gently pipette mix up and down 10 times, changing tips after each column (or seal plate and shake at 1800 rpm for 2 minutes). Make sure that beads are fully resuspended.

14 Incubate at room temperature for 2 minutes.

15 Place the plate on the magnetic stand for 2 minutes or until the supernatant has cleared.

16 Using a multichannel pipette, carefully transfer 50 μl of the supernatant from the Amplicon PCR plate to a new 96-well PCR plate. Change tips between samples to avoid cross-contamination.

> **SAFE STOPPING POINT**
> If you do not immediately proceed to *Index PCR*, seal plate with Microseal "B" adhesive seal and store it at -15° to -25°C for up to a week.

Index PCR

This step attaches dual indices and Illumina sequencing adapters using the Nextera XT Index Kit.

Consumables

Item	Quantity	Storage
2x KAPA HiFi HotStart ReadyMix	25 µl per sample	-15° to -25°C
Nextera XT Index 1 Primers (N7XX) from the Nextera XT Index kit (FC-131-1001 or FC-131-1002)	5 µl per sample	-15° to -25°C
Nextera XT Index 2 Primers (S5XX) from the Nextera XT Index kit (FC-131-1001 or FC-131-1002)	5 µl per sample	-15° to -25°C
PCR Grade Water	10 µl per sample	
TruSeq Index Plate Fixture (FC-130-1005)	1	
96-well 0.2 ml PCR plate	1 plate	
Microseal 'A' film	1	

Procedure

1   Using a multichannel pipette, transfer 5 µl from each well to a new 96-well plate. The remaining 45 µl is not used in the protocol and can be stored for other uses.

2   Arrange the Index 1 and 2 primers in a rack (i.e. the TruSeq Index Plate Fixture) using the following arrangements as needed:

   a   Arrange Index 2 primer tubes (white caps, clear solution) vertically, aligned with rows A through H.

   b   Arrange Index 1 primer tubes (orange caps, yellow solution) horizontally, aligned with columns 1 through 12.

   For more information on index selection, see Dual Indexing Principle, on page 23.

Figure 4   TruSeq Index Plate Fixture
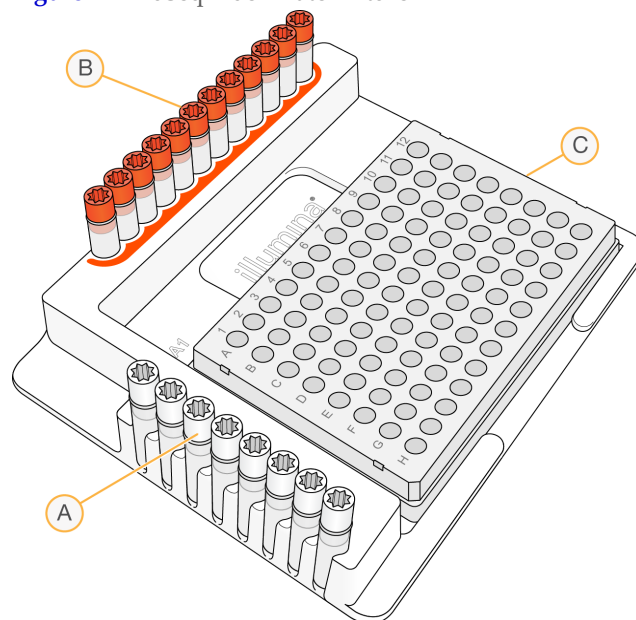
  A   Index 2 primers (white caps)
  B   Index 1 primers (orange caps)
  C   96-well plate

3   Place the 96-well PCR plate with the 5 µl of resuspended PCR product DNA in the TruSeq Index Plate Fixture.

4   Set up the following reaction of DNA, Index 1 and 2 primers, 2x KAPA HiFi HotStart ReadyMix, and PCR Grade water:

	Volume
DNA	5 µl
Nextera XT Index Primer 1 (N7xx)	5 µl
Nextera XT Index Primer 2 (S5xx)	5 µl
2x KAPA HiFi HotStart ReadyMix	25 µl
PCR Grade water	10 µl
Total	50 µl

5   Gently pipette up and down 10 times to mix.

6   Cover the plate with Microseal 'A'.

7   Centrifuge the plate at 1,000 × g at 20°C for 1 minute.

8 Perform PCR on a thermal cycler using the following program:
- 95°C for 3 minutes
- 8 cycles of:
    — 95°C for 30 seconds
    — 55°C for 30 seconds
    — 72°C for 30 seconds
- 72°C for 5 minutes
- Hold at 4°C

PCR Clean-Up 2

This step uses AMPure XP beads to clean up the final library before quantification.

Consumables

Item	Quantity	Storage
10 mM Tris pH 8.5	27.5 µl per sample	-15° to -25°C
AMPure XP beads	56 µl per sample	2° to 8°C
Freshly Prepared 80% Ethanol (EtOH)	400 µl per sample	
96-well 0.2 ml PCR plate	1 plate	
[Optional] Microseal 'B' film		
[Optional] 96-well MIDI plate	1 plate	

Procedure

1   Centrifuge the Index PCR plate at 280 × g at 20°C for 1 minute to collect condensation.

2   **[Optional - for use with shaker for mixing]** Using a multichannel pipette set to 50 µl, transfer the entire Index PCR product from the PCR plate to the MIDI plate. Change tips between samples.

> **NOTE**
> Transfer the sample to a 96-well MIDI plate if planning to use a shaker for mixing. If mixing by pipette, the sample can remain in the 96-well PCR plate.

3   Vortex the AMPure XP beads for 30 seconds to make sure that the beads are evenly dispersed. Add an appropriate volume of beads to a trough.

4   Using a multichannel pipette, add 56 µl of AMPure XP beads to each well of the Index PCR plate.

5   Gently pipette mix up and down 10 times if using a 96-well PCR plate or seal plate and shake at 1800 rpm for 2 minutes if using a MIDI plate.

6   Incubate at room temperature without shaking for 5 minutes.

7   Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared.

8   With the Index PCR plate on the magnetic stand, use a multichannel pipette to remove and discard the supernatant. Change tips between samples.

9   With the Index PCR plate on the magnetic stand, wash the beads with freshly prepared 80% ethanol as follows:
    a   Using a multichannel pipette, add 200 µl of freshly prepared 80% ethanol to each sample well.
    b   Incubate the plate on the magnetic stand for 30 seconds.
    c   Carefully remove and discard the supernatant.

10 With the Index PCR plate on the magnetic stand, perform a second ethanol wash as follows:

    a Using a multichannel pipette, add 200 µl of freshly prepared 80% ethanol to each sample well.

    b Incubate the plate on the magnetic stand for 30 seconds.

    c Carefully remove and discard the supernatant.

    d Use a P20 multichannel pipette with fine pipette tips to remove excess ethanol.

11 With the Index PCR plate still on the magnetic stand, allow the beads to air-dry for 10 minutes.

12 Remove the Index PCR plate from the magnetic stand. Using a multichannel pipette, add 27.5 µl of 10 mM Tris pH 8.5 to each well of the Index PCR plate.

13 If using a 96-well PCR plate, gently pipette mix up and down 10 times until beads are fully resuspended, changing tips after each column. If using a MIDI plate, seal plate and shake at 1800 rpm for 2 minutes.

14 Incubate at room temperature for 2 minutes.

15 Place the plate on the magnetic stand for 2 minutes or until the supernatant has cleared.

16 Using a multichannel pipette, carefully transfer 25 µl of the supernatant from the Index PCR plate to a new 96-well PCR plate. Change tips between samples to avoid cross-contamination.

> **SAFE STOPPING POINT**
> If you do not plan to proceed to *Library Quantification, Normalization, and Pooling, on page 16*, seal the plate with Microseal "B" adhesive seal. Store the plate at -15° to -25°C for up to a week.

[Optional] Validate Library

Run 1 µl of a 1:50 dilution of the final library on a Bioanalyzer DNA 1000 chip to verify the size. Using the V3 and V4 primer pairs in the protocol, the expected size on a Bioanalyzer trace of the final library is ~630 bp.

Figure 5   Example Bioanalyzer Trace of Final Library
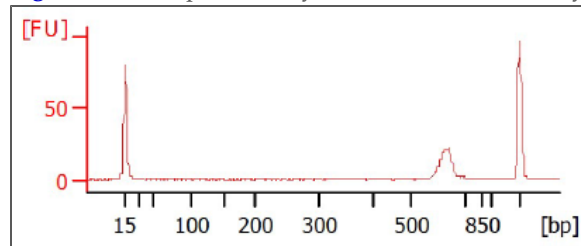
Library Quantification, Normalization, and Pooling

Illumina recommends quantifying your libraries using a fluorometric quantification method that uses dsDNA binding dyes.

Calculate DNA concentration in nM, based on the size of DNA amplicons as determined by an Agilent Technologies 2100 Bioanalyzer trace:

$$\frac{(\text{concentration in ng/µl})}{(660 \text{ g/mol} \times \text{average library size})} \quad \times \quad 10^6 \quad = \quad \text{concentration in nM}$$

For example:

$$\frac{15 \text{ ng/µl}}{(660 \text{ g/mol} \times 500)} \quad \times \quad 10^6 \quad = \quad 45 \text{ nM}$$

Dilute concentrated final library using Resuspension Buffer (RSB) or 10 mM Tris pH 8.5 to 4 nM. Aliquot 5 µl of diluted DNA from each library and mix aliquots for pooling libraries with unique indices. Depending on coverage needs, up to 96 libraries can be pooled for one MiSeq run.

For metagenomics samples, >100,000 reads per sample is sufficient to fully survey the bacterial composition. This number of reads allows for sample pooling to the maximum level of 96 libraries, given the MiSeq output of > 20 million reads.

Library Denaturing and MiSeq Sample Loading

In preparation for cluster generation and sequencing, pooled libraries are denatured with NaOH, diluted with hybridization buffer, and then heat denatured before MiSeq sequencing. Each run must include a minimum of 5% PhiX to serve as an internal control for these low-diversity libraries. Illumina recommends using MiSeq v3 reagent kits for improved run metrics.

Consumables

Item	Quantity	Storage
10 mM Tris pH 8.5 or RSB (Resuspension Buffer)	6 µl	-15° to -25°C
HT1 (Hybridization Buffer)	1540 µl	-15° to -25°C
0.2 N NaOH (less than a week old)	10 µl	-15° to -25°C
PhiX Control Kit v3 (FC-110-3001)	4 µl	-15° to -25°C
MiSeq reagent cartridge	1 cartridge	-15° to -25°C
1.7 ml microcentrifuge tubes (screw cap recommended)	3 tubes	
2.5 L ice bucket		

Preparation

1 Set a heat block suitable for 1.7 ml microcentrifuge tubes to 96°C

2 Remove a MiSeq reagent cartridge from -15°C to -25°C storage and thaw at room temperature.

3 In an ice bucket, prepare an ice-water bath by combining 3 parts ice and 1 part water.

Denature DNA

1 Combine the following volumes of pooled final DNA library and freshly diluted 0.2 N NaOH in a microcentrifuge tube:
   - 4 nM pooled library (5 µl)
   - 0.2 N NaOH (5 µl)

2 Set aside the remaining dilution of 0.2 N NaOH to prepare a PhiX control within the next 12 hours.

3 Vortex briefly to mix the sample solution, and then centrifuge the sample solution at 280 × g at 20°C for 1 minute.

4 Incubate for 5 minutes at room temperature to denature the DNA into single strands.

5 Add the following volume of pre-chilled HT1 to the tube containing denatured DNA:
   - Denatured DNA (10 µl)

- Pre-chilled HT1 (990 μl)

Adding the HT1 results in a 20 pM denatured library in 1 mM NaOH.

6 Place the denatured DNA on ice until you are ready to proceed to final dilution.

Dilute Denatured DNA

1 Dilute the denatured DNA to the desired concentration using the following example:

> 📌 **NOTE**
> Illumina recommends targeting 800–1000 K/mm² raw cluster densities using MiSeq v3 reagents. It is suggested to start your first run using a 4 pM loading concentration and adjust subsequent runs appropriately.

Final Concentration	2 pM	4 pM	6 pM	8 pM	10 pM
20 pM denatured library	60 μl	120 μl	180 μl	240 μl	300 μl
Pre-chilled HT1	540 μl	480 μl	420 μl	360 μl	300 μl

2 Invert several times to mix and then pulse centrifuge the DNA solution.

3 Place the denatured and diluted DNA on ice.

Denature and Dilution of PhiX Control

Use the following instructions to denature and dilute the 10 nM PhiX library to the same loading concentration as the Amplicon library. The final library mixture must contain at least 5% PhiX.

1 Combine the following volumes to dilute the PhiX library to 4 nM:
- 10 nM PhiX library (2 μl)
- 10 mM Tris pH 8.5 (3 μl)

2 Combine the following volumes of 4 nM PhiX and 0.2 N NaOH in a microcentrifuge tube:
- 4 nM PhiX library (5 μl)
- 0.2 N NaOH (5 μl)

3 Vortex briefly to mix the 2 nM PhiX library solution.

4 Incubate for 5 minutes at room temperature to denature the PhiX library into single strands.

5 Add the following volumes of pre-chilled HT1 to the tube containing denatured PhiX library to result in a 20 pM PhiX library:
- Denatured PhiX library (10 μl)
- Pre-chilled HT1 (990 μl)

6 Dilute the denatured 20 pM PhiX library to the same loading concentration as the Amplicon library as follows:

Final Concentration	2 pM	4 pM	6 pM	8 pM	10 pM
20 pM denatured library	60 µl	120 µl	180 µl	240 µl	300 µl
Pre-chilled HT1	540 µl	480 µl	420 µl	360 µl	300 µl

7 Invert several times to mix and then pulse centrifuge the DNA solution.

8 Place the denatured and diluted PhiX on ice.

Combine Amplicon Library and PhiX Control

> **NOTE**
> The recommended PhiX control spike-in of ≥ 5% for low diversity libraries is possible with RTA v1.17.28 or later, which is bundled with MCS v2.2. For optimal performance, update to v3 software (MCS 2.3). If you are using an older version of the MiSeq software or sequencing these libraries on the GA or HiSeq, Illumina recommends using ≥ 25% PhiX control spike-in.

1 Combine the following volumes of denatured PhiX control library and your denatured amplicon library in a microcentrifuge tube:
   • Denatured and diluted PhiX control (30 µl)
   • Denatured and diluted amplicon library (570 µl)

2 Set the combined sample library and PhiX control aside on ice until you are ready to heat denature the mixture immediately before loading it onto the MiSeq v3 reagent cartridge.

3 Using a heat block, incubate the combined library and PhiX control tube at 96°C for 2 minutes.

4 After the incubation, invert the tube 1–2 times to mix and immediately place in the ice-water bath.

5 Keep the tube in the ice-water bath for 5 minutes.

> **NOTE**
> Perform the heat denaturation step immediately before loading the library into the MiSeq reagent cartridge to ensure efficient template loading on the MiSeq flow cell.

MiSeq Reporter Metagenomics Workflow

After samples are loaded, the MiSeq system provides on-instrument secondary analysis using the MiSeq Reporter software (MSR). MSR provides several options for analyzing MiSeq sequencing data. For this demonstrated 16S protocol, select the Metagenomics workflow.

By following this 16S Metagenomics protocol, the Metagenomics workflow classifies organisms from your V3 and V4 amplicon using a database of 16S rRNA data. The classification is based on the Greengenes database (http://greengenes.lbl.gov/). The output of this workflow is a classification of reads at several taxonomic levels: kingdom, phylum, class, order, family, genus, and species. The analysis output includes:

- Clusters Graph – shows numbers of raw cluster, clusters passing filter, clusters that did not align, clusters not associated with an index, and duplicates.
- Sample Table – summarizes the sequencing results for each sample.
- Cluster Pie Chart – a graphical representation of the classification breakdown for each sample.

See the *MiSeq Reporter Metagenomics Workflow – Reference Guide* (Part # 15042317) for detailed instructions and guidance.

The method described in this 16S Metagenomics protocol can be used for any targeted amplicon sequencing, relevant to virus research, mutation detection, or other microbiology-related studies. If you use the protocol for other targeted amplicon sequencing studies, select the MiSeq Reporter Generate FASTQ Workflow for on-instrument generation of FASTQ files for downstream analysis. For specific guidance on the Generate FASTQ Workflow, see the *MiSeq Reporter Generate FASTQ Workflow – Reference Guide* (Part # 15042322).

Supporting Information

The protocols described in this guide assume that you are familiar with the contents of this section and have obtained all of the requisite equipment and consumables.

Acronyms

Table 1 Acronyms

Acronym	Definition
HT1	Hybridization Buffer
IEM	Illumina Experiment Manager
MSR	MiSeq Reporter
PCR	Polymerase Chain Reaction
rRNA	Ribosomal RNA
RSB	Resuspension Buffer

Consumables and Equipment

Check to make sure that you have all of the necessary user-supplied consumables and equipment before proceeding to sample preparation.

Table 2 User-Supplied Consumables

Consumable	Supplier
1.7 ml microcentrifuge tubes	General lab supplier
10 μl barrier pipette tips	General lab supplier
10 μl multichannel pipettes	General lab supplier
10 μl single channel pipettes	General lab supplier
20 μl barrier pipette tips	General lab supplier
20 μl multichannel pipettes	General lab supplier
20 μl single channel pipettes	General lab supplier
200 μl barrier pipette tips	General lab supplier
200 μl multichannel pipettes	General lab supplier
200 μl single channel pipettes	General lab supplier
1000 μl barrier pipette tips	General lab supplier

Consumable	Supplier
1000 µl multichannel pipettes	General lab supplier
1000 µl single channel pipettes	General lab supplier
96-well 0.2 ml skirtless PCR plates or Twin.Tec 96-well PCR plates	Bio-Rad, part # MSP-9601
Agencourt AMPure XP 60 ml kit	Beckman Coulter Genomics, part # A63881
Ethanol 200 proof (absolute) for molecular biology (500 ml)	Sigma-Aldrich, part # E7023
Amplicon PCR Forward Primer (Standard desalting)	
Amplicon PCR Reverse Primer (Standard desalting)	
KAPA HiFi HotStart ReadyMix (2X)	KAPA Biosystems, part # KK2601
Microseal 'A' adhesive seals	Bio-Rad, part # MSA-5001
Microseal 'B' adhesive seals	Bio-Rad, part # MSB-1001
MiSeq Reagent Kit v3 (600 cycle)	Illumina, catalog # MS-102-3003
Nextera XT Index Kit	Illumina, catalog # FC-131-1001 or Illumina, catalog # FC-131-1002
PhiX Control Kit v3	Illumina, catalog # FC-110-3001
PCR grade water	General lab supplier
Fluorometric quantitation with dsDNA binding dye reagents	General lab supplier
RNase/DNase-free 8-well PCR strip tubes and caps	General lab supplier
RNase/DNase-free multichannel reagent reservoirs, disposable	VWR, part # 89094-658
Tris-HCl 10 mM, pH 8.5	General lab supplier
[Optional] 96-well storage plates, round well, 0.8 ml ("MIDI" plate)	Fisher Scientific, part # AB-0859

Table 3 User-Supplied Equipment

Equipment	Supplier
2.5 L ice bucket	General lab supplier
96-well thermal cycler (with heated lid)	General lab supplier

Equipment	Supplier
Fluorometer for quantitation with dsDNA binding dyes	General lab supplier
Magnetic stand-96	Life Technologies, catalog # AM10027
Microplate centrifuge	General lab supplier
TruSeq Index Plate Fixture Kit (reusable)	Illumina, catalog # FC-130-1005
[Optional] 2100 Bioanalyzer Desktop System	Agilent, part # G2940CA
[Optional] Agilent DNA 1000 Kit	Agilent, part # 5067-1504
[Optional] High Speed Micro Plate Shaker	VWR, catalog # 13500-890 (110V/120V) or VWR, catalog # 14216-214 (230V)

Dual Indexing Principle

The dual indexing strategy uses two 8 base indices, Index 1 (i7) adjacent to the P7 sequence, and Index 2 (i5) adjacent to the P5 sequence. Dual indexing is enabled by adding a unique Index 1 (i7) and Index 2 (i5) to each sample. The 96 sample Nextera XT Index Kit (FC-131–1002) use 12 different Index 1 (i7) adapters (N701–N712) and 8 different Index 2 (i5) adapters (S501–S508). The 24 sample Nextera XT Index Kit (FC-131–1001) uses 6 different Index 1 (i7) adapters (N701–N706) and 4 different Index 2 (i5) adapters (S501–S504). In the Index adapter name, the N or S refers to Nextera XT sample preparation, 7 or 5 refers to Index 1 (i7) or Index 2 (i5), respectively. The 01–12 refers to the Index number. A list of index sequences is provided for generating sample sheets to demultiplex the samples:

Index 1 (i7)	Sequence	Index 2 (i5)	Sequence
N701	TAAGGCGA	S501	TAGATCGC
N702	CGTACTAG	S502	CTCTCTAT
N703	AGGCAGAA	S503	TATCCTCT
N704	TCCTGAGC	S504	AGAGTAGA
N705	GGACTCCT	S505	GTAAGGAG
N706	TAGGCATG	S506	ACTGCATA
N707	CTCTCTAC	S507	AAGGAGTA
N708	CAGAGAGG	S508	CTAAGCCT
N709	GCTACGCT		
N710	CGAGGCTG		
N711	AAGAGGCA		
N712	GTAGAGGA		

Low Plexity Pooling Guidelines

Illumina uses a green laser or LED to sequence G/T and a red laser or LED to sequence A/C. At each cycle, at least one of two nucleotides for each color channel are read to ensure proper registration. It is important to maintain color balance for each base of the index read being sequenced, otherwise index read sequencing could fail due to registration failure. If you choose the dual-indexed sequencing workflow, always use at least two unique and

compatible barcodes for each index (index 1 and index 2). The following tables illustrate possible pooling strategies:

Table 4 Libraries Pooled: 6 or fewer; Sequencing Workflow: Single Index

Plex	Index 1 (i7) Selection	Index 2 (i5) Selection
1-plex (no pooling)	Any Index 1 adapter	Any Index 2 adapter
2-plex	• [option 1] N702 and N701 • [option 2] N702 and N704	
3-plex	• [option 1] N701, N702, and N704 • [option 2] N703, N705, and N706	
4- or 5-plex	• [option 1] N701, N702, N704, and any other Index 1 adapter • [option 2] N703, N705, N706, and any other Index 1 adapter	
6-plex	N701, N702, N703, N704, N705, and N706	

Table 5 Sequencing Workflow: Single or Dual Index

Plex	Index 1 (i7) Selection	Index 2 (i5) Selection
7–12 plex, Dual Index	• [option 1] N701, N702, N704, and any other Index 1 adapter (as needed) • [option 2] N703, N705, N706, and any other Index 1 adapter (as needed)	• [option 1] S501 and S502 • [option 2] S503 and S504 • [option 3] S505 and S506
7–12 plex, Single Index (96 sample Nextera Index adapter kit)	• N701–N706 and any other Index 1 adapter (as needed)	• Any Index 2 (i5) adapter
Greater than 12-plex	N701, N702, N703, N704, N705, N706, and any other Index 1 adapter	• [option 1] S501, S502, and any other Index 2 adapter (as needed) • [option 2] S503, S504, and any other Index 2 adapter (as needed) • [option 3] S505, S506, and any other Index 2 adapter (as needed)

These strategies represent only some of the acceptable combinations. Alternatively, check the real sequences of each index in the tables to make sure that each base position has a signal in both color channels for the index read:

Good				Bad			
Index 1		Index 2		Index 1		Index 2	
705	GGACTCCT	503	TATCCTCT	705	GGACTCCT	502	CTCTCTAT
706	TAGGCATG	503	TATCCTCT	706	TAGGCATG	502	CTCTCTAT
701	TAAGGCGA	504	AGAGTAGA	701	TAAGGCGA	503	TATCCTCT
702	CGTACTAG	504	AGAGTAGA	702	CGTACTAG	503	TATCCTCT
	√√√√√√√		√√√√√√√		√√√√√√√		√√√√xxxx

√=signal in both color
x=signal missing in one color channel

Prevent PCR Product Contamination

The PCR process is commonly used in the laboratory to amplify specific DNA sequences. Unless proper laboratory hygiene is used, PCR products can contaminate reagents, instrumentation, and genomic DNA samples, causing inaccurate and unreliable results. PCR product contamination can shut down lab processes and significantly delay normal operations.

Make sure that the lab is set up appropriately to reduce the risk of PCR product contamination:

- **Physically Separate Pre-PCR and Post-PCR Areas**
  - Physically separate laboratory space where pre-PCR processes are performed (DNA extraction, quantification, and normalization) from the laboratory space where PCR products are made and processed (post-PCR processes).
  - Never use the same sink to wash pre-PCR and post-PCR troughs.
  - Never share water purification systems for pre-PCR and post-PCR processes.
  - Store all supplies used in the protocols in the pre-PCR area, and transfer to the post-PCR area as needed.
- **Use Dedicated Equipment and Supplies**
  - Dedicate separate full sets of equipment and supplies (pipettes, centrifuges, oven, heat block, etc.) to pre-PCR and post-PCR lab processes, and never share between processes.
  - Dedicate separate storage areas (freezers and refrigerators) to pre-PCR and post-PCR consumables.

Because the pre- and post-amplification reagents are shipped together, it is important to unpack the reagents in the pre-PCR lab area. After unpacking the reagents, move the post-amplification reagents to the proper post-PCR storage area.

Pre-PCR and Post-PCR Lab Procedures

To prevent PCR product contamination, it is important to establish lab procedures and follow best practices. Illumina recommends daily and weekly cleaning of lab areas using

0.5% Sodium Hypochlorite (10% Bleach).

> **CAUTION**
> To prevent sample or reagent degradation, make sure that all vapors from the cleaning solution have fully dissipated before beginning any processes.

Daily Cleaning of Pre-PCR Area

A daily cleaning of the pre-PCR area using a 0.5% Sodium Hypochlorite (10% Bleach) solution helps to eliminate PCR product that has entered the pre-PCR area.

Identify pre-PCR areas that pose the highest risk of contamination, and clean these areas with a 0.5% Sodium Hypochlorite (10% Bleach) solution before beginning any pre-PCR processes. High-risk areas might include, but are not limited to, the following items:

- Benchtops
- Door handles
- Refrigerator/freezer door handles
- Computer mouse
- Keyboards

Daily Cleaning of Post-PCR Area

Reducing the amount of PCR product in the post-PCR area helps reduce the risk of contamination in the pre-PCR area. Daily cleaning of the post-PCR area using a 0.5% Sodium Hypochlorite (10% Bleach) solution helps reduce the risk of contamination.

Identify post-PCR areas that pose the highest risk of contamination, and clean these areas with a 0.5% Sodium Hypochlorite (10% Bleach) solution daily. High-risk areas might include, but are not limited to, the following items:

- Thermal cyclers
- Bench space used to process amplified DNA
- Door handles
- Refrigerator/freezer door handles
- Computer mouse
- Keyboards

Weekly Cleaning of All Lab Areas

One time a week, perform a thorough cleaning of the pre-PCR and post-PCR areas using 0.5% Sodium Hypochlorite (10% Bleach).

- Clean all benchtops and laboratory surfaces.
- Clean all instruments that are not cleaned daily.
- Thoroughly mop lab floors.
- Make sure that personnel responsible for weekly cleaning are properly trained on prevention of PCR product contamination.

Items Fallen to the Floor

The floor is contaminated with PCR product transferred on the shoes of individuals coming from the post-PCR area; therefore, anything falling to the floor must be treated as contaminated.

- Disposable items that have fallen to the floor, such as empty tubes, pipette tips, gloves, lab coat hangers, must be discarded.

- Non-disposable items that have fallen to the floor, such as a pipette or an important sample container, must be immediately and thoroughly cleaned. Use a 0.5% Sodium Hypochlorite (10% Bleach) solution to remove PCR product contamination.
- Clean any lab surface that has come in contact with the contaminated item. Individuals handling anything that has fallen to the floor, disposable or non-disposable, must discard their lab gloves and put on a new pair.

Best Practices

When preparing libraries for sequencing, always adhere to good molecular biology practices. Read through the entire protocol before starting to make sure that all of the required materials are available and your equipment is programmed and ready to use.

Handling Liquids

Good liquid handling measures are essential, particularly when quantifying libraries or diluting concentrated libraries for making clusters.

- Small differences in volumes (±0.5 µl) can sometimes cause large differences in cluster numbers (~100,000).
- Small volume pipetting can be a source of potential error in protocols requiring the generation of standard curves, such as qPCR, or small but precise volumes, such as the Agilent Bioanalyzer.
- If small volumes are unavoidable, use due diligence to make sure that pipettes are correctly calibrated.
- Make sure that pipettes are not used at the volume extremes of their performance specifications.
- Prepare the reagents for multiple samples simultaneously, to minimize pipetting errors, especially with small volume enzyme additions. As a result, pipette one time from the reagent tubes with a larger volume, rather than many times with small volumes. Aliquot to individual samples in a single pipetting movement to allow for standardization across multiple samples.

Handling Magnetic Beads

NOTE
Cleanup procedures have only been validated using the 96-well plates and the magnetic stand specified in the *Consumables and Equipment* list. Comparable performance is not guaranteed when using a microcentrifuge tube or other formats, or other magnets.

- Before use, allow the beads to come to room temperature.
- Do not reuse beads. Always add fresh beads when performing these procedures.
- Immediately before use, vortex the beads until they are well dispersed and the color of the liquid is homogeneous.
- When pipetting the beads, pipette slowly and dispense slowly due to the viscosity of the solution.
- Take care to minimize bead loss, which can affect final yields.
- Change the tips for each sample, unless specified otherwise.
- Let the mixed samples incubate at room temperature for the time indicated in the protocol for maximum recovery.

- When removing and discarding supernatant from the wells, use a single channel or multichannel pipette and take care not to disturb the beads.
- When aspirating the cleared solution from the reaction plate and wash step, it is important to keep the plate on the magnetic stand and not disturb the separated magnetic beads. Aspirate slowly to prevent the beads from sliding down the sides of the wells and into the pipette tips.
- To prevent the carryover of beads after elution, approximately 2.5 µl of supernatant is left when the eluates are removed from the bead pellet.
- Be sure to remove all of the ethanol from the bottom of the wells, as it can contain residual contaminants.
- Keep the reaction plate on the magnetic stand and let it air-dry at room temperature to prevent potential bead loss due to electrostatic forces. Allow for the complete evaporation of residual ethanol, because the presence of ethanol affects the performance of the subsequent reactions. Illumina recommends at least minutes drying time, but a longer drying time can be required. Remaining ethanol can be removed with a 10 µl pipette.
- Avoid over drying the beads, which can impact final yields.
- Do not scrape the beads from the edge of the well using the pipette tip.
- To maximize sample recovery during elution, incubate the sample/bead mix for 2 minutes at room temperature before placing the samples onto the magnet.

Avoiding Cross-Contamination

Practice the following to avoid cross-contamination:
- Open only one adapter tube at a time.
- Change the tips for each sample, unless specified otherwise.
- Pipette carefully to avoid spillage.
- Clean pipettes and change gloves between handling different adapter stocks.
- Clean work surfaces thoroughly before and after the procedure.

Potential DNA Contaminants

When handling and processing samples using this protocol, use best practices to avoid PCR contamination, as you would when preparing PCR amplicons.

Temperature Considerations

Temperature is an important consideration for making libraries:
- Keep libraries at temperatures ≤37°C, except where specifically noted.
- Place reagents on ice after thawing at room temperature.

Equipment

- Review the programming instructions for your thermal cycler user guide to make sure that it is programmed appropriately using the heated lid function.
- It is acceptable to use the thermal cycler tracked heating lid function.

Whole-genome sequencing improves discrimination of relapse from reinfection and identifies transmission events among patients with recurrent *Clostridium difficile* infections

M. Mac Aogáin [a,*], G. Moloney [a], S. Kilkenny [a], M. Kelleher [b], M. Kelleghan [b], B. Boyle [b], T.R. Rogers [a,b]

[a] *Department of Clinical Microbiology, Sir Patrick Dun Translational Research Laboratory, School of Medicine, Trinity College Dublin, St James's Hospital, Dublin, Ireland*
[b] *Department of Microbiology, St James's Hospital, Dublin, Ireland*

A R T I C L E   I N F O

CrossMark

S U M M A R Y

Background: Recurrent *Clostridium difficile* infection (CDI) represents a significant healthcare challenge. Patients may suffer multiple episodes of CDI with the index strain (relapse) or become infected by another strain acquired nosocomially (reinfection).
Aim: We aimed to characterize *C. difficile* isolates causing recurrent CDI at a tertiary referral hospital by whole-genome sequencing (WGS) to assess strain similarities at the highest level of genetic resolution and accurately detect relapse, reinfection, and putative strain transmission events.
Methods: An 18-month prospective study of recurrent CDI was undertaken. *Clostridium difficile* was cultured from stool samples collected longitudinally from any patients suffering ≥2 clinically defined CDI episodes. Patient demographics and clinical data were recorded, and strain relatedness investigated by both polymerase chain reaction (PCR)-based ribotyping and WGS.
Findings: Nineteen patients were identified with ≥2 clinically defined CDI episodes who cumulatively suffered 39 recurring CDI episodes (58 total episodes). Patients had a median length of stay (LOS) of 144 days and experienced between two and seven CDI episodes. Ribotyping indicated 27 apparent same-strain relapses, five reinfections and the predominance of ribotypes 078 (ST-11) and 020 (ST-2). WGS allowed characterization of relapse with increased certainty and identified emergent within-strain single nucleotide variants (SNVs) with potential functional impact on diverse genes. Shared ribotypes among 14 patients with recurrent CDI suggested 10 possible patient-to-patient transmission events. However, WGS revealed greater diversity at the sub-ribotype level, excluding all but four transmission events.
Conclusion: WGS exhibits several advantages over PCR-based ribotyping in terms of its ability to distinguish relapse from reinfection, to identify patient-to-patient transmission

---

* Corresponding author. Address: Department of Clinical Microbiology, Trinity College Dublin, Room 1.26, Sir Patrick Dun Translational Research Labs, St James's Hospital, Dublin 8, Ireland. Tel.: +353 1 896 3079.
  E-mail address: m.macaogain@tcd.ie (M. Mac Aogáin).

events, and to exact fine structure characterization of recurrent CDI epidemiology. This offers the potential for more focused infection prevention strategies to eliminate strain transmission among patients with recurrent CDI.

Introduction

Between 15% and 50% of patients who develop *Clostridium difficile* infection (CDI) will suffer subsequent CDI episodes, which adds to the clinical and economic burden of this disease.[1–4] Recurrence is defined as a CDI episode occurring within eight weeks of a previous infection.[5] Accepted risk factors for recurrent CDI include older age (>65 years), prescribing of additional 'non-CDI' antibiotics, and cumulative time spent in the healthcare environment.[1,3,6,7] Recurrent clinical episodes may be categorized as relapse, when due to the original strain, or reinfection, when caused by a newly acquired strain.

Molecular typing studies of *C. difficile* have provided insight into the proportions of cases with relapsed CDI as opposed to reinfection.[8,9] Estimates for reinfections range from 12% to 35% of recurrent CDI episodes, within the limits of discrimination provided by conventional typing methods such as polymerase chain reaction (PCR)-based ribotyping, pulsed-field gel electrophoresis (PFGE) or multi-locus sequence typing (MLST) with the intervals to recurrence after a first episode of CDI ranging from 24 to 42 days.[2,3,7,10–12]

The use of whole-genome sequencing (WGS) has provided evidence for a higher degree of *C. difficile* strain diversity than previously acknowledged.[13–15] A recent study applied WGS to 1223 *C. difficile* strains and found 45% of all isolates investigated to be genetically distinct, suggesting a considerable reservoir of endemic *C. difficile* strains.[15] Of the patients infected with genetically indistinguishable strains, Eyre *et al.* found that 38% had identifiable hospital contact with another symptomatic case and 36% had no recognizable shared epidemiology.[15] This underscores the existence of unidentified *C. difficile* transmission routes.[15–17] In a subsequent study, Eyre *et al.* applied WGS to recurrent CDI with the consideration that ribotyping may underestimate reinfections caused by endemic ribotypes.[18] This provided improved discrimination between relapse and reinfection through comparisons of paired isolates (index versus first recurrence) and revealed that 81% of recurrences were caused by the same strain, 15% by reinfections with 4% assigned to an indeterminate category.[18]

We undertook prospective analysis of CDI episodes meeting clinical and microbiological criteria and identified all patients suffering recurrent CDI over an 18-month period. Strains causing index as well as first and subsequent CDI episodes were characterized using both conventional ribotyping and WGS to assess the level of concordance of these methods in view of the enhanced discriminatory power of WGS.

Methods

Setting

St James's Hospital (SJH) is a 1015-bed acute tertiary care hospital with some 3800 staff members and an immediate catchment population of about 350,000. Annual inpatient admissions exceed 25,000 with more than 220,000 outpatient and 46,000 emergency department visits per annum.

Study cohort

Between January 1st, 2012 and June 30th, 2013, all clinical cases of recurrent CDI were identified at St James's Hospital, Dublin, in accordance with national guidelines for recurrent CDI. In addition, any patient suffering ≥2 clinical CDI episodes was included in our analysis, even if episodes occurred more than eight weeks apart. Laboratory confirmation of cases meeting clinical criteria was provided by the Premier toxin A and B enzymatic immune assay (Meridian Bioscience Inc., Cincinnati, OH, USA) performed either on direct stool samples ('toxin positive') or on cultured isolates ('culture positive') grown on Brazier's cefoxitin cycloserine egg-yolk (CCEY) agar under anaerobic conditions (10% H_2, 10% CO_2 and 80% N_2) at 37°C for 48–72 h.

Strain collection

Stool samples from patients suffering ≥2 identified CDI episodes, which were originally confirmed by the Diagnostic Laboratory, were recovered for further analysis. Of 58 identified CDI cases meeting this criterion, stool samples were available for 53 (91%). Stool samples were subjected to alcohol shock and plated on Brazier's CCEY agar to selectively isolate *C. difficile*. From these toxin-positive cultures, a single colony was taken and stored as a spore stock culture at –80°C as previously described.[19] PCR-based ribotyping was performed on all isolates to establish strain relatedness.[20]

Whole-genome sequencing

DNA was extracted from *C. difficile* using the Roche High-pure PCR template preparation kit (Roche Diagnostics Ltd, Burgess Hill, UK). Nextera XT library preparation reagents (Illumina, Eindhoven, The Netherlands) were used to generate multiplexed paired end sequencing libraries of *C. difficile* genomic DNA. Resultant libraries were sequenced on an Illumina MiSeq instrument. All short-read data obtained in this study have been deposited in the European Nucleotide Archive (ENA), project accession number PRJEB6575.

Sequence mapping and variant calling

Paired end reads were mapped to the *C. difficile* 630 reference genome (AM180355) with the Burrows–Wheeler Aligner (BWA) and analysed with the SAMtools package.[21,22] Strains were sequenced to an average raw read depth of 91.1 ± 44.5-fold. Sequence types (ST) were determined using the *Clostridium difficile* Multi Locus Sequence Typing website (http://pubmlst.org/cdifficile/).[23] Single nucleotide variants

(SNVs) were called using the SAMtools mpileup command consistent with the parameters described by Didelot *et al.* for SNV calling in *C. difficile*.[14]

Ethics

This study proposal was reviewed by the hospital research ethics committee (REC ref: 23/9 83/13) and considered to be part of a service improvement for the infection control team.

Results

Recurrent CDI prevalence and associated patient demographics

Over the 18-month study period, a total of 230 CDI episodes were documented at SJH representing a CDI rate of 0.42/1000 occupied bed-days and a recurrence rate of 10% among hospitalized patients. Although recurrent CDI is generally defined as a positive CDI result dated within the preceding eight weeks of a prior CDI infection, for the purpose of this study, we extended our definition to include any CDI episode preceded by a prior episode in the same patient over the course of the entire study (18 months).[5] Despite our liberal criteria for defining recurrence, 18/19 patients had at least one recurrent episode which conformed to the accepted criteria for recurrent CDI. Using our criteria, 19 index and 39 recurrent isolates were identified among the 58 episodes investigated. Five episodes fell outside the accepted eight-week boundary of the formal definition of recurrent CDI (Figure 1).

The demographic details of the 19 patients who suffered recurrence are summarized in Table I. They had a mean age of 73.5 years (range: 35.5–94 years) and a median LOS of 144 days. Patients suffered between two and seven CDI episodes with a median time from admission to first CDI episode and first CDI recurrence of 71 and 32 days respectively. The majority of clinically defined cases were confirmed by detection of *C. difficile* toxins in faecal samples (68%) with the remainder confirmed by direct detection of toxin A/B production by *C. difficile* cultured from faeces. Patients experienced an average of 2.5 ward transfers (range: 1–7) throughout their admission and were cared for by a range of clinical specialties. Two patients died (of complications unrelated to CDI) and two remained in the hospital receiving ongoing care over the course of the study within an onsite inpatient long-term care facility (LTCF). Of the 15 patients who survived to hospital discharge, nine were discharged to LTCFs and five were discharged home (one discharge location unknown).

Investigation of recurrent isolates by PCR-based ribotyping

In 16/19 cases a single ribotype was identified per patient, consistent with relapse. PCR ribotyping results supported relapse in the majority (27/39) of recurrences, with only five reinfections identified (Figure 1). In the case of seven CDI episodes, the nature of CDI recurrence (relapse or reinfection) could not be confirmed due to missing samples (Figure 1; P9, P11, P12, P18). Two patients suffered both relapse and reinfection (Figure 1; P1 and P7). One of these patients (Figure 1, P1) had two recurrent episodes involving a ribotype 078 strain,

and suffered a subsequent ribotype 017 reinfection, followed by a reinfection with the original 078 strain. Another patient suffered two reinfections, the second of which relapsed (Figure 1, P7). Of the 14 ribotype profiles identified, 078 and 020 predominated with a total of nine and 12 linked CDI episodes identified among four and three patients respectively. A ribotype 017 strain was isolated from five episodes among three patients (Figure 1). No patient harboured the 027/NAP1/BI strain.

Investigation of recurrent isolates by WGS

All isolates were subjected to WGS and comparative SNV analysis with reference to the *C. difficile* 630 genome (AM180355). MLST sequence types (ST), predicted from WGS data, were consistent with previously observed MLST–ribotype correlations.[24] This allowed assignment of one isolate, for which a ribotype designation could not be established, to ST-3 (Figure 1, P14). Strains of the same ribotypes, causing multiple infections in individual patients, were compared by WGS in an effort to confirm relapse with increased certainty. Overall, WGS analysis was consistent with ribotyping in defining reinfection and relapse; strains of the same ribotype from individual patients differed by ≤2 SNVs (Table II, Supplementary Table I). Thus the SNV differences observed among these strains were within the bounds of previously accepted criteria for inferred relapse in *C. difficile*.[18]

One or two SNVs were identified on comparing first and last isolated strains in patients who relapsed. In almost half (7/16, 44%) of these patients, we observed the occurrence of within-strain SNVs emerging over the course of their recurrent CDI infections. Where emergent SNVs were observed, the number of SNVs per strain ranged from one to two, or two to 15 SNVs per strain per year, when the observed time interval between isolation of first and last isolate in each individual patient was considered (Table II, Supplementary Table I). The genomic locations of the SNVs which arose over the course of clinical CDI relapses are detailed in Table II.

Patient-to-patient transmission of C. difficile inferred by PCR-ribotype analysis

Fourteen patients shared strains of the same ribotype; ribotype 078 was shared by four patients, ribotypes 020 and 017 each infected three patients, and ribotypes 050 and 003 were each found to be shared between two pairs of individual patients (Figure 1). The electronic records of patients infected by *C. difficile* of identical ribotype were investigated for epidemiological evidence supporting transmission including shared space and time on a ward, shared medical specialty team and overlapping admission times. This identified 10 possible patient-to-patient transmission events (Figure 2A, A–J). Six such events were substantiated by clinical data including shared ward placement (Figure 2A, A–E) or shared medical specialty (Figure 2A, F). Four potential transmission events involved shared ward placement of symptomatic and non-symptomatic patients (Figure 2A, A–D). Ribotyping also highlighted four apparent transmission events without substantiating epidemiologic evidence other than overlapping hospital admission times (Figure 2A, G–J).
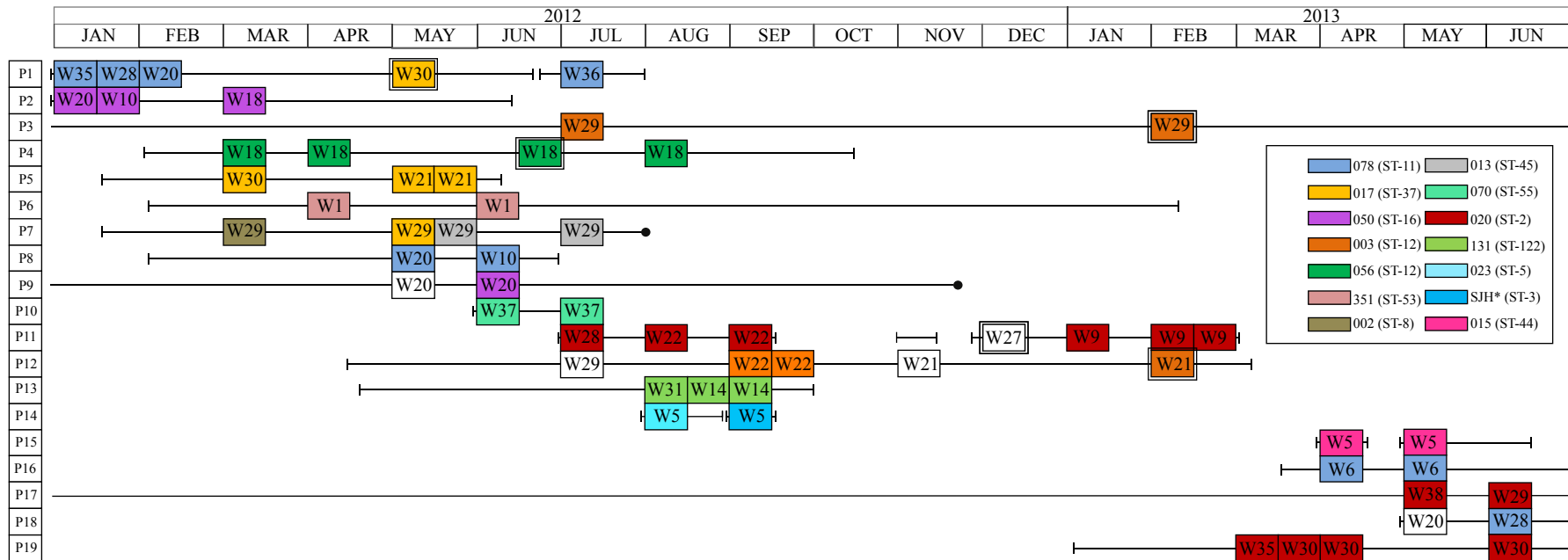
Figure 1. Timeline of *Clostridium difficile* infection (CDI) episodes illustrating ribotype prevalence and ward location of patients with recurrent CDI. Nineteen patients (P1–19) are each represented by horizontal lines spanning patient admissions observed over the 18-month study interval. Clinical CDI episodes are represented as rectangles. Black lines indicate the length of hospital admission times. Admission and discharge dates are bracketed by vertical lines; closed circles represent admissions terminated by death; and unbracketed lines indicate admissions that precede or succeed the study interval. The colour of each rectangle corresponds to the identified ribotype as indicated in the key. Sequence types (ST), based on whole-genome sequencing analysis, are indicated in parenthesis. One isolate (SJH*) — did not match any reference strains in our ribotyping database and could not be assigned a ribotype. White rectangles represent CDI episodes for which a stool sample was not available. Rectangles with a double border indicate repeat CDI episodes which fall outside the eight-week definition of recurrent CDI and which would be considered 'new infections' under existing guidelines. The ward location of the patient at the time of active CDI is indicated in each rectangle.

Table I
Demographics of patients with recurrent *Clostridium difficile* infection

Patient demographics	Value
Total patients	19
Age (mean, range; years)	73.5; 35−94
Gender (M; F)	10; 9
Length of stay (days)	
Per patient median, mean	141, 220
Range	9−1780[a]
Mean no. of admissions per patient	1.2
No. of ward placements per patient (mean, range)	2.5, 1−7
Clinical specialty (N = 26)	
Medical	19 (73%)
General medicine	5 (19%)
Medicine for elderly	5 (19%)
Endocrinology	2 (7%)
Gastrointestinal/hepatology	2 (7%)
Haematology	2 (7%)
Respiratory	2 (7%)
Nephrology	1 (3%)
Surgical	3 (11%)
General surgery	1 (3%)
Orthopaedics	1 (3%)
Plastics/reconstructive surgery	1 (3%)
Psychiatry	1 (3%)
No. of CDI episodes	
Total no.	58
Per patient mean, range	2, 2−7
Microbiological confirmation	
Direct toxin detection	39 (68%)
Identified by culture of toxigenic strain	19 (32%)
Time from admission to first CDI episode	
Per patient median, range (days)	71, 20−1444[a]
Time to first recurrence of CDI[b]	
Per patient median, range (days)	32, 5−191
Outcomes observed	
Survival to hospital discharge	15 (78.9%)
Ongoing inpatient LTCF	2 (10.5%)
Death terminating admission	2 (10.5%)
Of those discharged (N = 15)	
Discharge to LTCF	9 (60%)
Discharge home	5 (33%)
Unknown	1 (6%)

CDI, *Clostridium difficile* infection; LTCF, long-term care facility.
[a] Data skewed by inclusion of two patients receiving long-stay care.
[b] For the purpose of this study, new CDI episodes (separated by >8 weeks) in the same patient were considered to be recurrent (N = 5).

WGS analysis of ribotyping-inferred CDI transmission events

To further investigate transmission events that had been inferred by ribotyping, all isolates of the same ribotype were compared by WGS analysis. The numbers of SNVs identified among isolates implicated in transmission are detailed in Figure 2B. Among the 10 suspected transmission events, WGS analysis excluded five (Figure 2A; A, D, E, I, J) through the identification of strain divergences of between five and 86 SNVs

(Figure 2B). Five strain transmissions were substantiated by WGS analysis (Figure 2A; B, C, F, G, H). Although strains implicated in events G and H differed by ≤2 SNVs, a difference of 1 vs 0 SNVs was observed for transmission events G (P1−P7) and H (P5−P7) respectively and the analysis thus marginally favoured event H. Although three SNVs were found to separate strains implicated in transmission event 'B', two of these SNVs appeared to have arisen over the course of CDI relapse in patient 'P2' (Figure 2B). Thus, in spite of the three observed SNVs, the acknowledged cut-off of ≤2 SNVs for inferring strain relatedness was not breached and this transmission event was supported. The SNVs which emerged between transmission events 'B' and 'C' are detailed in Table II. In total, four transmission events were inferred from WGS among the 19 patients investigated.

Discussion

We investigated the molecular epidemiology of recurrent CDI cases at a tertiary referral hospital comparing conventional PCR-based ribotyping and WGS analysis. Overall, the age profile of patients with recurrence was reflective of national data for adult inpatients in Ireland.[25] However, our recurrent CDI cohort had an exceptional LOS which placed them in the minority (3.3%) of inpatient admissions nationally.[25] Even within this category, the national mean LOS is estimated to be 65.5 days, considerably shorter than our patients' experience.[25] This finding was likely attributable to underlying comorbidities as well as CDI. Although we did not undertake formal calculation of comorbidity, available clinical details suggested that this group had considerable medical issues and nursing requirements (data not shown). This is also reflected in the high percentage of the group discharged to long-term care (60%) compared to 4.7% of all adult inpatients nationally.[25] Patients thus comprised a vulnerable group who experienced multiple CDI episodes over prolonged hospital admissions.

Fourteen distinct ribotypes were identified including the 078 strain, which has been reported previously in recurrent CDI cases in Ireland.[26] Strains belonging to ribotypes 020 and 017 were also present. All three ribotypes have proven virulence potential and have been implicated in recurrent CDI.[27−29] Notably, the 027/NAP1/B1 lineage, which has been associated with recurrent CDI, was not detected. Local ribotype prevalence data, for strains collected over the duration of this study, suggest that 078 and 020 strains are the most frequently occurring ribotypes at our hospital, each accounting for 19% of observed isolates, whereas the 027/NAP1/BI lineage was less frequently observed (unpublished data). Thus, strain ribotype prevalence among our recurrent CDI cohort appeared to reflect local *C. difficile* epidemiology. Two patients suffered both relapse and reinfection (Figure 1; P1 and P7). Similar findings have previously been described and highlight the complex epidemiological scenarios that arise among patients with recurrent CDI.[2,30] However, in our cohort, the majority of CDI episodes resulted from same-strain relapses with only one patient suffering reinfection as the sole cause of clinical recurrence.

To confirm persistent, same-strain relapse among recurrent CDI patients, we used WGS to distinguish relapse and reinfection with greater accuracy. All relapses (as identified by PCR ribotyping) were confirmed by WGS; strains causing relapse

Table II
Emergent within-strain SNVs and their predicted impact on gene function

Patient	Ribotype	ST	SNV locus	Reference[a]	Variant[b]	Synonymous[c]	Protein alteration	Locus tag[d]	Gene function
P2	050	ST-16	1412874	T	C	No	V93A	CD630_12140	*spo0A*; stage 0 sporulation protein A
			3243804	C	A	No	A176S	CD630_27870	*cwp84*; cell surface protein
P5	017	ST-37	1826371	A	C	No	S215R	CD630_15770	*pgm*; alpha-phosphoglucomutase
			4111335	G	A	No	G149S	CD630_35180	*murC*; UDP-N-acetylmuramate-L-alanine ligase
P7	002	ST-45	1186090	G	T	No	P36Q	CD630_10160	Transcriptional regulator, MerR family
			1484356	G	A	Yes	na	CD630_12770	Putative acetyltransferase
P8	078	ST-11	3810191_3810192insT	AT	ATT	No	Frameshift	CD630_32550	*rgaR*; two-component response regulator VirR-like
P12	003	ST-12	2708584_2708585insC	TC	TCC	No	Frameshift	CD630_23410	*abfD*; gamma-aminobutyrate metabolism dehydratase/isomerase
P13	131	ST-122	829898delC	ACC	AC	No	Frameshift	CD630_06840	Putative ATP-dependent peptidase, M41 family
P19	020	ST-2	4097019	C	T	No	G424E	CD630_35060	Conserved hypothetical protein
			4061875	C	T	No	R283C	CD630_34700	*atpA*[e]; ATP synthase subunit alpha
P1 (vs P5)[f]	017	ST-37	457298	G	A	na	na	intergenic; CD630_03540-CD630_03550	112 bp upstream of: CD630_03550 (bglF; PTS system, IIABC component)
P9 (vs P2)	050	ST-16	60312	G	A	No	A289V	CD630_00370	*acoB*; acetoin dehydrogenase E1 component

SNVs, single nucleotide variants; ST, sequence type; na, not applicable.

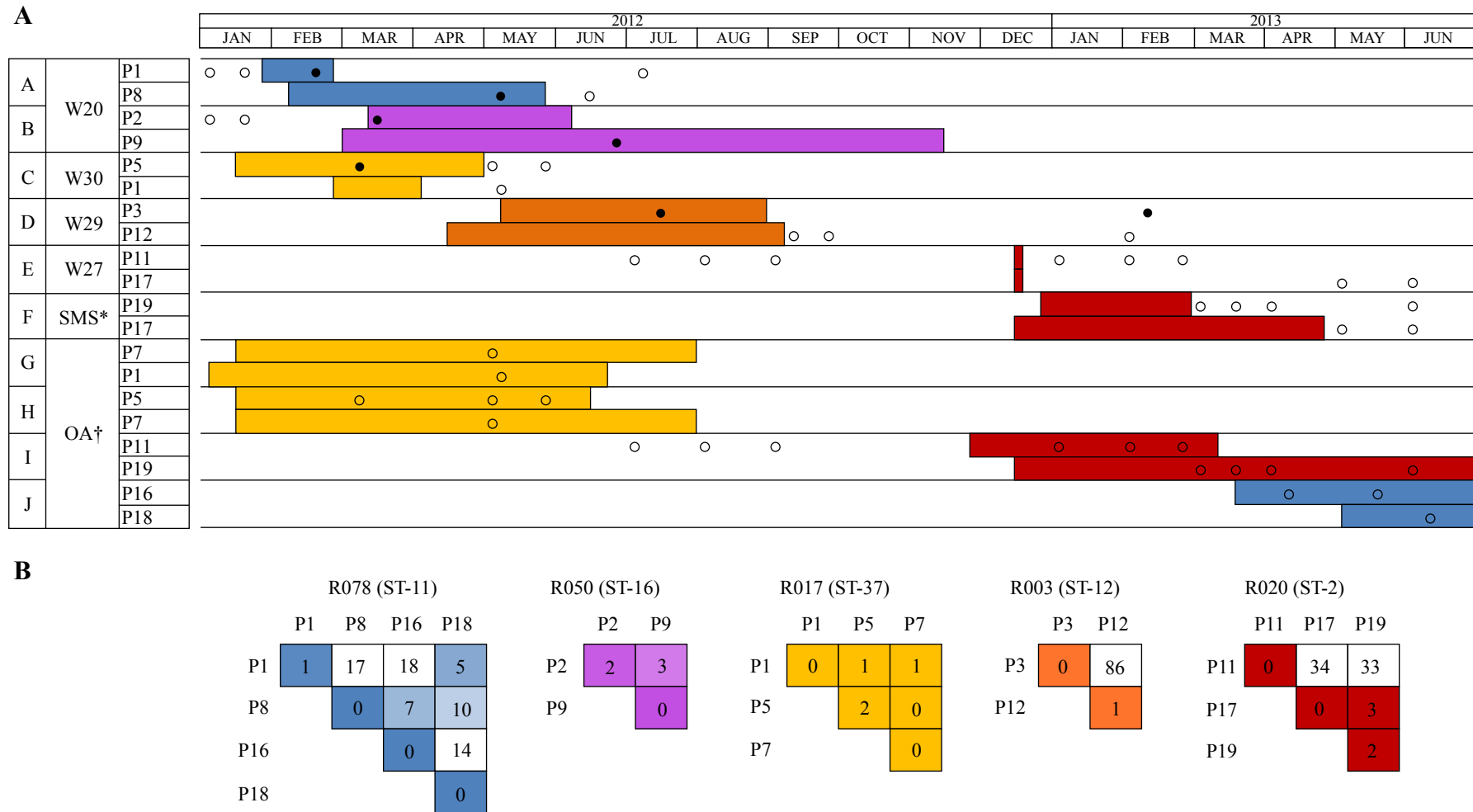[a] Sequence identity at relative SNV locus in *C. difficile* 630 (AM180355) reference genome.
[b] Sequence identity at relative SNV locus in *C. difficile* clinical isolates.
[c] Functional status of SNV at protein level (synonymous or non-synonymous).
[d] Relative locus tag in *C. difficile* 630 reference genome in which within-host SNVs were observed.
[e] The atpA gene is used in the *C. difficile* MLST scheme. This mutation gives rise to a novel MLST profile which has been designated ST-295.
[f] Observed SNVs were identified on comparison of transmitted strains rather than over the course of relapse in individual patients.

Figure 2. Timeline of suspected *Clostridium difficile* infection (CDI) transmission events investigated by whole-genome sequencing analysis (WGS) among patients with recurrent CDI. (A) Ten suspected transmission events were defined based on the identification of shared ribotypes among 13 patients with overlapping hospital stays (column one, A–J). These transmissions were further supported by epidemiologic data including either shared ward time (W##) or shared medical specialties (SMS*). For four suspected transmissions (G–J), no supportive clinical data other than overlapping hospital admissions (OA, †) was observed. The respective transmission events, supportive clinical data, and the patients involved are detailed in columns one, two, and three, respectively. Horizontal coloured bars represent overlapping patient ward placements, time under common medical specialties or overlapping admission times consistent with transmission. The colour of the bars corresponds to colour scheme used in Figure 1 and indicates strain ribotype. Closed circles represent CDI episodes that occurred on a ward where transmission was suspected. Open circles indicate an episode that occurred on a different ward to that of the suspected transmission. Only episodes caused by ribotypes implicated in transmission are shown. (B) Analysis of suspected transmission events by WGS. All isolates implicated in transmission events were subjected to WGS analysis. The number of SNV differences between same-ribotype isolates is illustrated by pairwise comparison tables for all strains of a shared ribotype against each other. Each table is coloured according to ribotype, consistent with Figures 1 and 2A. The degree of coloration in each square corresponds to the degree of similarity (less observed SNVs) between strains.

were found to be identical or differed by ≤2 SNVs at the whole-genome level which is considered an acceptable cut-off within the bounds of the predicted within-host evolutionary rate for *C. difficile*.[14,18] Five patients experiencing intervals greater than eight weeks between CDI episodes, which thus fell outside accepted formal definitions for recurrent CDI, were nonetheless included in our analysis (Figure 1; P1, P3, P4, P11, P12). According to accepted guidelines, these should be considered as new rather than as recurrent CDI episodes in light of the exceptional interval between episodes.[5] Interestingly, WGS analysis demonstrated that all five patients suffered relapse by strains genetically indistinguishable from their index case, in spite of the long intervals between episodes. The longest interval observed between infections caused by identical strains was 191 days, which exceeds current definitions for recurrence by >19 weeks. Previous WGS analysis of paired *C. difficile* isolates from cases separated by one to 561 days also identified apparent relapse (≤2 SNVs between isolates) over exceptional timescales.[14] However, such lengthy intervals between index and relapse could also be interpreted as reinfections by genetically indistinguishable strains via common environmental contamination sources.[2] A limitation of our study was the absence of WGS data on the broader population of *C. difficile* strains at this institution, including those causing non-recurrent CDI. This would have provided greater insight into transmission dynamics between recurrent CDI patients and the broader hospital population and whether environmental sources of genetically identical strains were present or, conversely, whether patients with relapse represent reservoirs for onward CDI transmission.

Longitudinal sequencing of *C. difficile* isolates from relapse episodes identified SNVs occurring over the course of recurrent CDI in individual patients (Table II). Of the 11 within-strain SNVs identified, 10 led to predicted non-synonymous changes at the protein level. A mutation in the *spo0A* gene, encoding a key regulator of *C. difficile* sporulation, virulence and metabolism, was observed in a ribotype 050 strain over the course of relapsing CDI (Table II).[31] Mutational alteration of *spo0A* has been observed previously in a *C. difficile* strain from a fidaxomicin-treated patient with CDI relapse.[18] Other regulatory genes affected included *rgaR* — encoding a predicted two-component response regulator — and a gene encoding a MerR-family transcriptional regulator. The emergence of mutations in central regulators of virulence *in vivo* can radically alter bacterial physiology, triggering adverse clinical outcomes.[32,33] Whereas our study was not designed to investigate the correlation between the emergence of bacterial mutations and clinical outcome, such changes may have clinical relevance and, given the growing adoption of WGS technology, they may become the focus of larger WGS studies addressing their clinical impact. Other genes in which mutations were observed included *cwp84*, encoding a protease involved in processing of the surface layer protein and biogenesis of the *C. difficile* cell wall, and *murC*, encoding an essential component of peptidoglycan biosynthesis.[34,35]

In many cases, sustained *C. difficile* infections recurred in our patients over prolonged intervals where multiple patient transfers between wards and medical specialties occurred. Given the potential for transmission of *C. difficile*, we focused our investigation on several apparent patient-to-patient transmission events among our recurrent CDI cohort. In total, 10 potential transmissions were suggested based on ribotyping

analysis, including six that were supported by clinical data (Figure 2A). However, analysis of WGS substantiated only four transmissions identifying multiple SNVs separating purportedly transmitted strains. The four transmission events supported by WGS were linked to at least five subsequent CDI episodes including at least one which recurred (Figure 2, transmission event 'F'). WGS identified a ribotype 017 (ST-37) strain causing relapse in one patient which subsequently caused reinfection in two others (Figure 2A, transmission events C and G). Analysis of WGS data also highlighted a potential transmission event concerning ribotype 050 (ST-16) (Figure 2A, event B) which was contentious due to the identification of three SNVs (greater than the accepted cut-off of ≤2) between the strains (Figure 2B), in spite of supportive epidemiological evidence. More focused analysis revealed that, when within-strain SNVs arising in the transmitted strain were considered, only a single SNV difference separated the two strains (Figure 2B and Table II). This suggested that the transmission event occurred prior to the subsequent accumulation of SNVs in the index strain of patient 'P2', thus distorting interpreted strain divergence when only the temporally closest isolates were compared. This highlighted the advantage of considering multiple strains when trying to establish patient-to-patient transmission routes among patients with relapsing CDI. Furthermore, the importance of mixed infections in establishing transmission chains is increasingly acknowledged and the investigation of a single isolate per sample represents both a limitation of this study and an important consideration for WGS studies of transmission.[36] Nonetheless, WGS analysis provided insights into recurrent CDI epidemiology beyond that achievable by conventional PCR-based ribotyping.

The ability of WGS to rule out spurious epidemiological interpretations and resolve cryptic transmission events is a major advantage over other typing methods of lower discriminatory power. In contrast to previous WGS analysis of *C. difficile*, where <40% of genetically identical strains had clinical evidence supporting transmission, the majority (three out of four) of our WGS-identified transmission events were substantiated by clinical data, albeit in a relatively small patient cohort.[15] This observation may highlight missed opportunities for infection control and that further intervention strategies (e.g. hand hygiene and environmental decontamination) are warranted in this vulnerable patient cohort. The confirmation of persistent infection by genetically indistinguishable strains over intervals greater than eight weeks was notable as current clinical definitions of recurrent CDI exclude such cases. Whether such protracted relapse intervals are indicative of chronic *C. difficile* colonization–infection cycles or are due to reinfection by common environmental sources is an intriguing question with implications for both CDI management and the definition of recurrent infection. The broader adoption of WGS technology in the clinical setting will undoubtedly help to address such questions.

Acknowledgements

Conflict of interest statement

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jhin.2015.01.021.

References

1. Hu MY, Katchar K, Kyne L, *et al*. Prospective derivation and validation of a clinical prediction rule for recurrent *Clostridium difficile* infection. *Gastroenterology* 2009;**136**:1206–1214.
2. Barbut F, Richard A, Hamadi K, Chomette V, Burghoffer B, Petit JC. Epidemiology of recurrences or reinfections of *Clostridium difficile*-associated diarrhea. *J Clin Microbiol* 2000;**38**:2386–2388.
3. Ramanathan S, Johnson S, Burns SP, *et al*. Recurrence of *Clostridium difficile* infection among veterans with spinal cord injury and disorder. *Am J Infect Control* 2014;**42**:168–173.
4. Wiegand PN, Nathwani D, Wilcox MH, Stephens J, Shelbaya A, Haider S. Clinical and economic burden of *Clostridium difficile* infection in Europe: a systematic review of healthcare-facility-acquired infection. *J Hosp Infect* 2012;**81**:1–14.
5. McDonald LC, Coignard B, Dubberke E, Song X, Horan T, Kutty PK. Recommendations for surveillance of *Clostridium difficile*-associated disease. *Infect Control Hosp Epidemiol* 2007;**28**:140–145.
6. Fekety R, McFarland LV, Surawicz CM, Greenberg RN, Elmer GW, Mulligan ME. Recurrent *Clostridium difficile* diarrhea: characteristics of and risk factors for patients enrolled in a prospective, randomized, double-blinded trial. *Clin Infect Dis* 1997;**24**:324–333.
7. Eyre DW, Walker AS, Wyllie D, *et al*. Predictors of first recurrence of *Clostridium difficile* infection: implications for initial management. *Clin Infect Dis* 2012;**55**(Suppl. 2):S77–S87.
8. Knetsch CW, Lawley TD, Hensgens MP, Corver J, Wilcox MW, Kuijper EJ. Current application and future perspectives of molecular typing methods to study *Clostridium difficile* infections. *Euro Surveill* 2013;**18**:20381.
9. Marsh JW, Arora R, Schlackman JL, Shutt KA, Curry SR, Harrison LH. Association of relapse of *Clostridium difficile* disease with BI/NAP1/027. *J Clin Microbiol* 2012;**50**:4078–4082.
10. Kamboj M, Khosa P, Kaltsas A, Babady NE, Son C, Sepkowitz KA. Relapse versus reinfection: surveillance of *Clostridium difficile* infection. *Clin Infect Dis* 2011;**53**:1003–1006.
11. Oka K, Osaki T, Hanawa T, *et al*. Molecular and microbiological characterization of *Clostridium difficile* isolates from single, relapse, and reinfection cases. *J Clin Microbiol* 2012;**50**:915–921.
12. Tang-Feldman Y, Mayo S, Silva Jr J, Cohen SH. Molecular analysis of *Clostridium difficile* strains isolated from 18 cases of recurrent *Clostridium difficile*-associated diarrhea. *J Clin Microbiol* 2003;**41**:3413–3414.
13. He M, Sebaihia M, Lawley TD, *et al*. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci USA* 2010;**107**:7527–7532.
14. Didelot X, Eyre DW, Cule M, *et al*. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biol* 2012;**13**:R118.
15. Eyre DW, Cule ML, Wilson DJ, *et al*. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med* 2013;**369**:1195–1205.
16. Riggs MM, Sethi AK, Zabarsky TF, Eckstein EC, Jump RL, Donskey CJ. Asymptomatic carriers are a potential source for transmission of epidemic and nonepidemic *Clostridium difficile* strains among long-term care facility residents. *Clin Infect Dis* 2007;**45**:992–998.
17. Dubberke ER, Reske KA, Noble-Wang J, *et al*. Prevalence of *Clostridium difficile* environmental contamination and strain variability in multiple health care facilities. *Am J Infect Control* 2007;**35**:315–318.
18. Eyre DW, Babakhani F, Griffiths D, *et al*. Whole-genome sequencing demonstrates that fidaxomicin is superior to vancomycin for preventing reinfection and relapse of infection with *Clostridium difficile*. *J Infect Dis* 2014;**209**:1446–1451.
19. Merrigan MM, Sambol SP, Johnson S, Gerding DN. Prevention of fatal *Clostridium difficile*-associated disease during continuous administration of clindamycin in hamsters. *J Infect Dis* 2003;**188**:1922–1927.
20. Stubbs SL, Brazier JS, O'Neill GL, Duerden BI. 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* 1999;**37**:461–463.
21. Li H, Handsaker B, Wysoker A, *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078–2079.
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;**25**:1754–1760.
23. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010;**11**:595.
24. Dingle KE, Griffiths D, Didelot X, *et al*. Clinical *Clostridium difficile*: clonality and pathogenicity locus diversity. *PloS One* 2011;**6**:e19993.
25. Economic and Social Research Institute. *Activity in acute public hospitals in Ireland, annual report 2011*. Report No. 45. Dublin: ESRI; 2011.
26. Burns K, Morris-Downes M, Fawley WN, Smyth E, Wilcox MH, Fitzpatrick F. Infection due to *C. difficile* ribotype 078: first report of cases in the Republic of Ireland. *J Hosp Infect* 2010;**75**:287–291.
27. Magnusson C, Wullt M, Lofgren S, Iveroth P, Akerlund T, Matussek A. Ribotyping of *Clostridium difficile* strains associated with nosocomial transmission and relapses in a Swedish County. *APMIS* 2013;**121**:153–157.
28. Bauer MP, Notermans DW, van Benthem BH, *et al*. *Clostridium difficile* infection in Europe: a hospital-based survey. *Lancet* 2011;**377**(9759):63–73.
29. Drudy D, Harnedy N, Fanning S, O'Mahony R, Kyne L. Isolation and characterisation of toxin A-negative, toxin B-positive *Clostridium difficile* in Dublin, Ireland. *Clin Microbiol Infect* 2007;**13**:298–304.
30. Alonso R, Gros S, Pelaez T, Garcia-de-Viedma D, Rodriguez-Creixems M, Bouza E. Molecular analysis of relapse vs re-infection in HIV-positive patients suffering from recurrent *Clostridium difficile* associated diarrhoea. *J Hosp Infect* 2001;**48**:86–92.
31. Pettit LJ, Browne HP, Yu L, *et al*. Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. *BMC Genomics* 2014;**15**:160.
32. Sokurenko EV, Hasty DL, Dykhuizen DE. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol* 1999;**7**:191–195.
33. Chotirmall SH, Smith SG, Gunaratnam C, *et al*. Effect of estrogen on pseudomonas mucoidy and exacerbations in cystic fibrosis. *N Engl J Med* 2012;**366**:1978–1986.
34. de la Riva L, Willing SE, Tate EW, Fairweather NF. Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of *Clostridium difficile*. *J Bacteriol* 2011;**193**:3276–3285.
35. Smith CA. Structure, function and dynamics in the mur family of bacterial cell wall ligases. *J Mol Biol* 2006;**362**:640–655.
36. Eyre DW, Cule ML, Griffiths D, *et al*. Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in *Clostridium difficile* transmission. *PLoS Comput Biol* 2013;**9**:e1003059.

Supplementary Table I

Estimated rate of SNV accumulation in strains from relapsing patients

Patient	Ribotype	ST	SNVs[a]	Weeks[b]	Minimum SNVs per strain per year
P7	002	ST45	2	7	14.9
P2	050	ST16	2	10	10.4
P5	017	ST37	2	11	9.5
P19	020	ST2	2	14	7.4
P15	015	ST44	1	5	10.4
P8	078	ST11	1	6	8.7
P13	131	ST122	1	6	8.7
P12	003	ST12	1	22	2.4
P1	078	ST11	1	24	2.2
P3	003	ST12	0	28	0.0
P4	056	ST58	0	20	0.0
P6	053	ST55	0	9	0.0
P10	070	ST55	0	5	0.0
P11	020	ST2	0	32	0.0
P16	078	ST11	0	5	0.0
P17	020	ST2	0	5	0.0
P9	na[c]	na	na	na	na
P14	na	na	na	na	na
P18	na	na	na	na	na

SNVs, single nucleotide variants; ST, sequence type; na, not applicable.

[a]Number of SNVs identified on comparing first and last strains from patients with CDI relapse.

[b]Number of weeks between isolation of first and last isolate from relapsing patients.

[c]Estimation of within-strain SNV rate not possible due to lack of paired isolates.

ORIGINAL ARTICLE

Possible Interplay Between Hospital and Community Transmission of a Novel *Clostridium Difficile* Sequence Type 295 Recognized by Next-Generation Sequencing

Geraldine Moloney, MRCPI;[1] Micheál Mac Aogáin, PhD;[1] Maureen Kelleghan, PGDipIPC;[2]
Brian O'Connell, MD, FRCPath, FRCPI;[2] Caroline Hurley, MScPH;[3] Elizabeth Montague, GNM;[3]
Mary Conlon, MPH;[3] Helena Murray, FFPHMI;[3] Thomas R. Rogers, FRCPath[1,2]

OBJECTIVE.   To use next-generation sequencing (NGS) analysis to enhance epidemiological information to identify and resolve a *Clostridium difficile* outbreak and to evaluate its effectiveness beyond the capacity of current standard PCR ribotyping.

METHODS.   NGS analysis was performed as part of prospective surveillance of all detected *C. difficile* isolates at a university hospital. An outbreak of a novel *C. difficile* sequence type (ST)-295 was identified in a hospital and a community hostel for homeless adults. Phylogenetic analysis was performed of all ST-295 and closest ST-2 isolates. Epidemiological details were obtained from hospital records and the public health review of the community hostel.

RESULTS.   We identified 7 patients with *C. difficile* ST-295 infections between June 2013 and April 2015. Of these patients, 3 had nosocomial exposure to this infection and 3 had possible hostel exposure. Current Society for Healthcare Epidemiology of America (SHEA)— Infectious Diseases Society of America (IDSA) surveillance definitions (2010) were considered in light of our NGS findings. The initial transmission was not detectable using current criteria, because of 16 weeks between ST-295 exposure and symptoms. We included 3 patients with hostel exposure who met surveillance criteria of hospital-acquired infection due to their hospital admissions.

CONCLUSION.   NGS analysis enhanced epidemiological information and helped identify and resolve an outbreak beyond the capacity of standard PCR ribotyping. In this cluster of cases, NGS was used to identify a hostel as the likely source of community-based *C. difficile* transmission.

Infect Control Hosp Epidemiol 2016;37:680–684

Clostridium difficile intestinal infection (CDI) is a leading cause of morbidity and mortality, both within healthcare environments and in the community.[1] Although the epidemic strain BI/NAP1/027 was first identified using pulsed-field gel electrophoresis (PFGE) and polymerase chain reaction (PCR) ribotyping techniques,[1] the application of next-generation sequencing (NGS) to investigate the epidemiology of *C. difficile* has demonstrated greater diversity within genotypes than was previously recognized.[2] In a seminal study by Eyre et al,[2] 36% of patients had no identifiable hospital or community contacts as a source for their infection despite *C. difficile* genomic differences of only 0–2 single-nucleotide variants (SNVs), which suggests complex modes of *C. difficile* acquisition.

Without a clearly defined incubation period for CDI, current Society for Healthcare Epidemiology of America (SHEA) — Infectious Diseases Society of America (IDSA) definitions classify nosocomial versus community acquisition according to the time interval since the patient's most recent healthcare exposure.[3] Although patients who develop community-acquired CDI have not, themselves, had healthcare exposure, it has been acknowledged that their household or residential contacts with recent hospital admissions could provide a source for community acquisition of *C. difficile*.[4] The potential for this mode of transmission is likely to reflect the complexity of the residence in terms of numbers and health status of its residents. Long-term housing provision within a 'hostel' facility for chronically homeless adults, most of whom have chronic illness, exemplifies a high order of residence complexity. Case management strategies that include provision of hostel-type accommodation have been shown to reduce emergency and inpatient admissions of such residents,[5] and tuberculosis and blood-borne virus transmission events have been identified in this context.[6,7] Here, we report a cluster of cases of CDI involving 4 hostel residents among 7 patients with
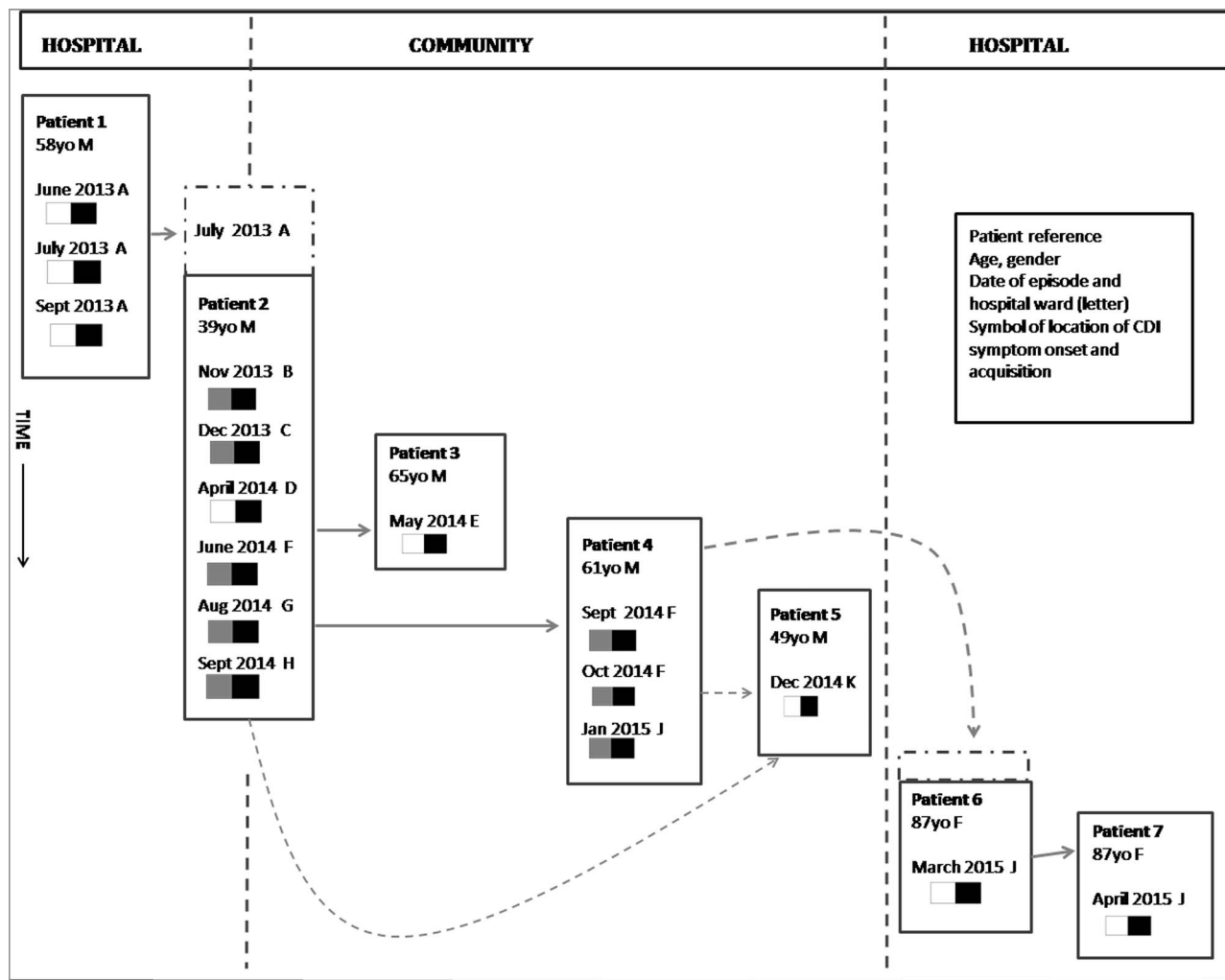
FIGURE 1.   Cluster of ST-295 *C. difficile* infection. Horizontal axis illustrates the patient's location at the time of inferred *C. difficile* transmission, with separation of locations by line style. Vertical axis illustrates time of cluster, from September 2013 to April 2015. Patient details are presented as pseudonym, age, gender, and CDI episodes with date and panels to represent IDSA/SHEA classification code. Hospital-onset, hospital-acquired CDI episodes are displayed as rectangular panel with white and dark gray fill ⬜■. Community-onset, hospital-acquired CDI episodes are displayed as rectangular panel with light and dark gray fill ▣■. Solid arrows (⟶) illustrate the transmission event from exposure to one symptomatic patient. Dashed arrows (⤍) indicate exposure to two symptomatic patients. Dashed extension to patient details represents identified ward contact for transmission, meeting criteria of Eyre et al.[2]

CDI due to *C. difficile* ST-295. Identification of this unique strain enabled resolution of its transmission from its recognized nosocomial origin (patient 1),[8] possible carriage to the hostel (patient 2), and subsequent nosocomial transmission following hospital admission of a hostel patient (patient 4) with recurrent CDI.

METHODS

Patients' clinical details and their *C. difficile* isolates were collected prospectively with the approval of our institutional review board. CDI was diagnosed in accordance with Irish national guidelines.[8] Following identification of *C. difficile*

DNA using a EntericBio PCR kit (Serosep, Annacotty, Ireland) at the diagnostic enteric laboratory, stool specimens were treated with ethanol shock before anaerobic incubation, using cycloserine-cefoxitin egg-yolk medium. Next-generation whole-genome sequencing of *C. difficile* colonies was performed at the TrinSeq Genome Sequencing Laboratory (Trinity College, Dublin, Ireland) with downstream read mapping and SNV calling performed as previously described.[8] Strain phylogenetic comparisons based on genome-wide SNV calls were performed by neighbor joining (BIONJ) using PhyML (ATGC, Montpellier, France).[9] Genomes were assigned sequence types according to the PubMLST database,[10] and an epidemiologic analysis was performed following
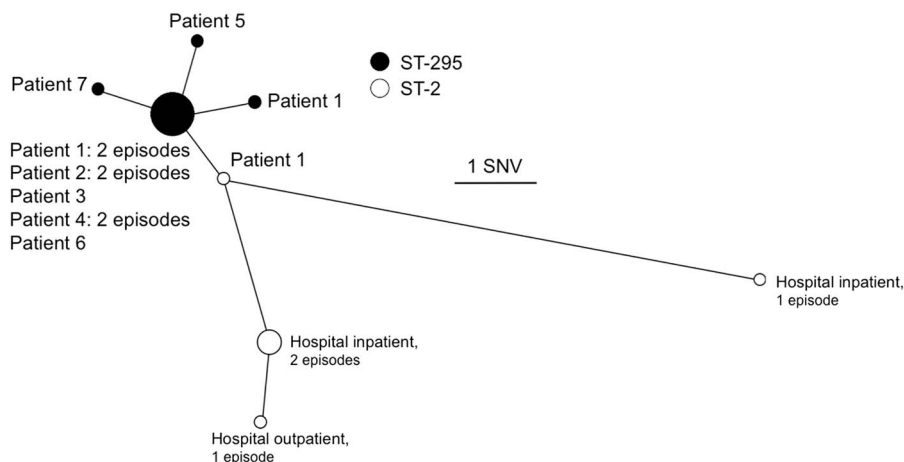
FIGURE 2. The phylogenetic relationship between ST-295 and closely related ST-2 isolates. A neighbor-joining tree illustrates the phylogenetic relationship between ST-295 (black circles) and closely related ST-2 (white circles) isolates observed during the outbreak investigation. Larger circles represent clusters of genetically indistinguishable isolates. In addition, 16 additional ST-2 isolates identified during our investigation are not shown because they differed from the index case (Patient 1) by >10 SNVs. All identified isolates that differed from the index case by <10 SNVs are included in the tree.

the approach of Eyre et al.[2] A PCR-based ribotype analysis was also performed on each isolate as previously described.[8] CDI episodes were classified according to current IDSA/SHEA definitions.[3] The hospital is a 1,015-bed acute tertiary care facility with annual inpatient admissions exceeding 25,000 and an immediate catchment population of ~350,000. The hostel is within this catchment population; ~66% of its residents are allocated to single en-suite bedrooms.

RESULTS

The details of CDI episodes confirmed to have been caused by *C. difficile* ST-295 are presented in Figure 1. Patient 1 had a prolonged hospital admission on ward A, during which he developed nosocomial CDI in March 2013, with 4 recurrences by September 2013. We identified the defining single-nucleotide variant (SNV) for *C. difficile* ST-295 as one that had evolved in this patient in June 2013 from an ST-2 isolate.[8] Patient 2 was admitted to ward A in July 2013, while patient 1 was symptomatic. Patient 2 had subsequent hospital admissions (from the hostel), with no further known contact with patient 1 or ward A, and this patient presented in December 2013 with community-onset symptoms of hospital-acquired CDI.

Patient 3 had been diagnosed with hospital onset of hospital-acquired CDI. He had no identifiable shared hospital exposures to either patients 1 or 2, but he was a resident of the hostel at the same time as patient 2. His only symptoms of CDI occurred as a hospital inpatient.

Patient 4 presented with community-onset symptoms in September 2014. Because he had an inpatient admission during the previous month, this episode was classified as community onset, hospital-acquired CDI. However, he had

known exposure to patient 2's community onset of CDI symptoms in August 2014 (Figure 1). This factor raised concern over the possibility of patient 4 having community-acquired infection. Patient 4 had a recurrence in October 2014 that warranted hospital readmission.

Patient 5 was diagnosed with CDI in December 2014 as a hospital inpatient. He was also a resident of the hostel at times when both patients 2 and 4 experienced community-onset of symptomatic CDI.

With new symptoms of CDI, patient 4 had another hospital admission in January–February 2015. CDI diagnoses were made for patient 6 in March 2015, and patient 7 in April 2015. *C. difficile* ST-295 was subsequently identified in the fecal samples of patient 6 in March 2015 and in patient 7 in April 2015. Some areas within the hospital ward (J) were common to both patients 4 and 6, and later, to patients 6 and 7.

In PCR-based ribotype analyses, all *C. difficile* ST-295 isolates were classified as ribotype 020. Phylogenetic comparison of genomic data showed that all ST-295 isolates clustered together as a distinct branch from ribotype 020 isolates belonging to ST-2 (Figure 2). Within our collection of 200 additional *C. difficile* clinical isolates with NGS information over the course of the outbreak investigation, only 5 ST-2 isolates were within 10 SNVs of ST-295. There is no plausible clinical explanation to support alternate hypotheses for the evolution and transmission of ST-295 isolates.

In October 2014, the CDI outbreak cases from the hostel (patients 2, 3, and 4) were reported to the responsible public health department; the first recognition of a possible CDI transmission within the hostel was made by a primary care physician. We subsequently learned that another patient from the same hostel was admitted to another hospital with CDI in September 2014, but, unfortunately, that isolate could not be

TABLE 1. Comparison of Society for Healthcare Epidemiology of America (SHEA)—Infectious Diseases Society of America (IDSA) Classification of *Clostridium difficile* Infection Episodes with Next-Generation Sequencing Analysis

Patient No.	Date of CDI Episode	IDSA/SHEA Classification[3]	Source of *C difficile* by NGS Analysis	Time Since Exposure to Symptomatic ST-295 CDI	Time to Recurrence of ST-295 CDI	NGS vs IDSA/SHEA Classifications
Patient 1	June 2013	HO-HA	HA	Index ST-295		Concordance
	July 2013		Recurrence		25 d	
	September 2013		Recurrence		52 d	
Patient 2	November 2013	CO-HA	HA	>16 wk		Variance: time since ST-295 exposure
					25 d	Discordance: community ST-295
	December 2013		Recurrence		19 wk[a]	
	April 2014	HO-HA	Recurrence		10 wk[a]	
	July 2014	CO-HA	Recurrence		36 d	
	August 2014		Recurrence		35 d	
	September 2014		Recurrence			
Patient 3	May 2014	HO- HA	CA	>24 wk		Discordance: community ST-295 exposure despite CDI symptom onset after 48 h of hospital admission
Patient 4	September 2014	CO-HA	CA	6–48 d		Discordance: community ST-295 exposure despite recent hospital admission
	October 2014		Recurrence		31 d	
	January 2015	CO-HA	Recurrence		14 wk[a]	
Patient 5	December 2014	HO-HA	CA	10–15 wk		Discordance: community ST-295 exposure despite CDI symptom onset after 48 h of hospital admission
Patient 6	March 2015	HO-HA	HA	29 d		Concordance
Patient 7	April 2015	HO-HA	HA	37 d		Concordance

NOTE. HO, hospital onset of symptoms; HA, hospital-acquired infection; CO, community onset of symptoms; CA, community-acquired infection.
[a]Cannot exclude reinfection from environmental spores versus intestinal carriage.

recovered for investigation. The community hostel is under the governance of a voluntary organization. It is not classified as a healthcare facility, and cleaning standards are not equivalent to those of a healthcare institution.[11] When these cases were recognized, public health action was undertaken. Residents were encouraged to practice good hand hygiene; environmental surfaces were treated with either 1,000 ppm of chlorine agent or bleach preparations;[12] and dedicated equipment was used to clean bedrooms of symptomatic residents.

A root-cause analysis was undertaken by the hospital's infection prevention and control team for nosocomial onset cases in response to the link identified between patients 6 and 7.

Table 1 presents an evaluation of the epidemiological classification of each patient's CDI episode(s) according to IDSA/SHEA guidelines[3] and in light of the NGS results. Findings for patients residing in the community hostel (patients 2, 3, 4 and 5) were discordant between the SHEA/IDSA classifications of their initial episodes and the acquisition source inferred by NGS analysis.

DISCUSSION

The resolution of a link between these cases by virtue of their sharing a *C. difficile* strain with sequence type 295 provides new insight into current surveillance definitions. Although prolonged carriage of *C. difficile* prior to a first episode of CDI has not been reported previously, our findings suggest that patient 2 may have had nosocomial acquisition of *C. difficile* ST-295 16 weeks before his symptoms began. This time lapse exceeds the maximum interval of 12 weeks between hospital discharge and symptom onset recognized by earlier surveillance recommendations.[13] These surveillance definitions acknowledge the complexity of attributing the source of *C. difficile* exposure when patients have been admitted to multiple facilities.[13] We cannot exclude the possibility of other mutual contacts across multiple hospital wards; however, Eyre et al's analysis of such intermediate contacts, either asymptomatic or with negative enzyme immunoassay results, suggests that this is more likely a chance finding than a source of

transmission.[2] Our NGS results favor community acquisition of *C. difficile* by patients 3, 4, and 5, all of whom had episodes categorized as hospital-acquired CDI by current IDSA/SHEA definitions.[3] We believe that, with the increasing complexity of the epidemiology of CDI, NGS enhances the capacity to distinguish between community and hospital acquisition as well as the 'trafficking' between them.

The community hostel setting was an interesting aspect of this case cluster. Patients 2, 3, and 5 had single bedrooms in close proximity. As a consequence of their medical issues, patients 2, 3, 4, and 5 all had antibiotic and proton-pump-inhibitor exposure prior to their first CDI episode. Appropriate infection prevention and control measures were taken in the hostel under the guidance of public health personnel. No new ST-295 infections have been detected in the 14 months since this intervention.

Hospitalized patients with suspected and/or confirmed CDI are normally placed in single rooms with en-suite facilities, accompanied by enhanced cleaning and disinfection of the room and equipment during the symptomatic period. Additional investigations have been conducted as a result of the likely link between patients 6 and 7.

We believe patient comorbidities, with further antibiotic exposure, to be the predominant cause for the recurrent disease experienced by patients 1, 2, and 4, but we cannot definitively exclude further environmental acquisition.

Although only 1 SNV distinguishes ST-295 from ST-2, its occurrence within the housekeeping gene *atpA* generated a new allele[10] and new sequence type[8] that served as a molecular marker for case recognition in this cluster. Ribotype analysis could not provide this degree of resolution, which supports its replacement by NGS.[2] To our knowledge, this is the first description of an ST-295–associated cluster and the first evidence of possible *C. difficile* transmission among residents of hostel facilities for homeless adults with chronic illness.

Address correspondence to Geraldine Moloney, Department of Clinical Microbiology, Sir Patrick Dun Translational Research Laboratory, Trinity College Dublin, St James's Hospital, Dublin, Ireland.

REFERENCES

1. Gerding DN, Lessa FC. The epidemiology of *Clostridium difficile* infection inside and outside health care institutions. *Infect Dis Clin North Am* 2015;29:37–50.
2. Eyre DW, Cule ML, Wilson DJ, et al. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med* 2013;369:1195–1205.
3. Cohen SH, Gerding DN, Johnson S, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). *Infect Control Hosp Epidemiol* 2010;31:431–455.
4. Chitnis AS, Holzbauer SM, Belflower RM, et al. Epidemiology of community-associated *Clostridium difficile* infection, 2009 through 2011. *JAMA Intern Med* 2013;173:1359–1367.
5. Sadowski LS, Kee RA, VanderWeele TJ, Buchanan D. Effect of a housing and case management program on emergency department visits and hospitalizations among chronically ill homeless adults: a randomized trial. *JAMA* 2009;301:1771–1778.
6. Kmietowicz Z. NICE advises screening for TB in hostels and prisons to reduce UK cases. *BMJ* 2012;344:e2309.
7. Neale J, Stevenson C. Routine exposure to blood within hostel environments might help to explain elevated levels of hepatitis C amongst homeless drug users: insights from a qualitative study. *Int J Drug Policy* 2012;23:248–250.
8. Mac Aogáin M, Moloney G, Kilkenny S, Kelleher M, Kelleghan M, Boyle B. Whole-genome sequencing improves discrimination of relapse from reinfection and identifies transmission events among patients with recurrent *Clostridium difficile* infections. *J Hosp Infect* 2015;90:108–116.
9. Guindon S, Dufayard JR, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
10. Griffiths D, Fawley W, Kachrimanidou M, et al. Multilocus sequence typing of *Clostridium difficile*. *J Clin Microbiol* 2010;48:770–778.
11. Housekeeping Manual for Municipally Operated Shelters. Hostel Services, Toronto Shelter, Support and Housing Administration website. http://www1.toronto.ca/City%20Of%20Toronto/Shelter%20Support%20&%20Housing%20Administration/Files/pdf/H/Housekeeping%20Manual%20-%20March%202013%20Version_1.pdf. Published 2013. Accessed July 23, 2015.
12. National Clinical Effectiveness Committee. Surveillance, Diagnosis and Management of *Clostridium difficile* Infection in Ireland. National Guideline No. 3. Health Protection Surveillance Centre website. http://www.hpsc.ie/A-Z/Gastroenteric/Clostridiumdifficile/Guidelines/File,13950,en.pdf. Published 2014. Accessed July 23, 2015.
13. McDonald LC, Coignard B, Dubberke E, et al. Recommendations for surveillance of *Clostridium difficile*-associated disease. *Infect Control Hosp Epidemiol* 2007;28:140–145.

CrossMark

Typhlocolitis associated with *Clostridium difficile* ribotypes 078 and 110 in neonatal piglets from a commercial Irish pig herd

Máire C. McElroy[1*], Martin Hill[1], Geraldine Moloney[2], Micheál Mac Aogáin[2], Shane McGettrick[1], Áine O'Doherty[1] and Thomas R. Rogers[2]

Abstract

Background: *Clostridium difficile* is a recognised cause of typhlocolitis and diarrhoea in neonatal pigs but has never been confirmed in association with pathology and disease in Irish pigs.

Case Presentation: Four neonatal piglets, with a history of diarrhoea were referred to the Central Veterinary Research Laboratory, Backweston for necropsy. They were from a fully integrated, commercial pig farm with approximately 1000 sows. Three piglets had acute, superficial, erosive and suppurative typhlocolitis and the other had mild suppurative mesocolitis. *Clostridium difficile* (*C. difficile*) toxins A/B were detected using ELISA in the colonic contents from each piglet. *C. difficile* isolates from two of the piglets were PCR-ribotyped as 078 and an isolate from a third pig was ribotyped as 110.

Conclusions: This is the first report confirming *C. difficile* in association with typhlocolitis in Irish pigs.

Keywords: *Clostridium difficile*, Typhlocolitis, Pigs, PCR ribotyping

Abbreviations: CDAD, *Clostridium difficile*-associated disease; PCR, Polymerase chain reaction

Background

Clostridium difficile is a toxin producing, Gram-positive, spore-forming, anaerobic enteropathogen of humans and animals. It has recently emerged as a major cause of porcine neonatal diarrhoea in America [1]. It has also been reported in association with neonatal diarrhoea in Europe although one study found no clear association between *C. difficile* isolation and diarrhoea [2, 3]. Porcine *Clostridium difficile* -associated disease (CDAD) typically manifests itself as early-onset diarrhoea and sudden death in piglets 1–7 days of age. Gross lesions may include mesocolonic oedema and large intestines may be filled with pasty to watery yellowish contents. Histopathological mucosal lesions are limited to the caecum and colon. They are typically mild, but vary from grossly inapparent, multifocal necrosis of surface epithelial cells to transmural necrosis. The classic lesions are segmental erosions of the epithelium with effusion of fibrin and neutrophils into the lumen, so-called "volcano ulcers" [4].

CDAD occurs when *C. difficile* proliferates after endogenous intestinal flora is disrupted, either by a change in diet or antimicrobial treatment [5]. *C. difficile* produces two major toxins, Toxin A (TcdA) and Toxin B (TcdB) that act synergistically to cause apoptosis of mucosal epithelial cells and disruption of intracellular actin filaments responsible for cell to cell adhesion. Consequently, there is increased permeability of mucosal surfaces. Toxins A and B also initiate an inflammatory cascade that can result in increased damage to host tissues and fluid exudation [6]. The requirements for development of CDAD are disruption of normal intestinal or colonic flora, presence of the organism in the environment, and the production of toxins [5].

The standard for diagnosis of porcine CDAD is detection of toxins A and B in faeces or colonic contents, generally using commercially available enzyme immunoassays. Cultivation of *C. difficile* is difficult to interpret

* Correspondence: maire.mcelroy@agriculture.gov.ie
[1]Pathology Division, Central Veterinary Research Laboratory, Backweston Campus, Celbridge, Co. Kildare, Ireland
Full list of author information is available at the end of the article

McElroy *et al. Irish Veterinary Journal* (2016) 69:10

Page 2 of 4

because it can be found in healthy pigs, therefore its isolation may have little diagnostic relevance [4].

C. difficile is also one of the most important nosocomial pathogens of humans, primarily associated with intestinal dysbiosis due to antibiotic administration. In recent years the epidemiology of human disease is changing with more community-acquired infections and emergence of strains in humans that are common in domestic animals [5]. Therefore, considerable interest is developing in potential zoonotic capabilities of *C. difficile*.

The aims of this paper are to document typhlocolitis associated with *C. difficile* in an outbreak of diarrhoea in neonatal pigs from a commercial pig farm in Ireland and to report the strain typing results of the *C. difficile* isolates.

Case presentation

Four piglets, 3-4-days old, that had died during an outbreak of high morbidity, low mortality, neonatal diarrhoea in a 1000 sow commercial pig herd were submitted for necropsy. Details of treatment prior to death were not available.

At gross necropsy all four carcasses were well preserved with adequate body fat reserves. Stomachs were filled with milk. There was mild mesocolonic oedema, and small intestinal and colonic contents were soft and yellow in all four. No gross changes were noted in intestinal, caecal or colonic mucosa. Other body systems were unremarkable.

Histopathology

At least 6 sections from representative areas along the length of the small intestine, a section of caecum and at least six sections of spiral colon were sampled for histopathology. They were fixed in buffered formalin, processed routinely and stained with haematoxylin and eosin.

On histopathological examination three of the piglets had mild ($n = 1$) to severe ($n = 2$), multifocal, superficial fibrinosuppurative and erosive colitis with neutrophils and fibrin spilling from lamina propria through the eroded epithelium and into the lumen ('volcano lesions') (Fig. 1). In the other piglet there was a mild, multifocal, suppurative mesocolitis. In addition, in two of the piglets there was acute, mild, superficial, suppurative enteritis, with superficial necrosis and microthrombosis in one of the piglets. All four piglets also had mild atrophic enteritis.

C. difficile toxin testing, isolation and PCR Ribotyping

Colonic contents were positive for *C. difficile* toxins A/B using using Premier Elisa Kit Toxins A&B (Meridian Bioscience Inc.)

Fifteen µL colonic contents were treated with 50 µL (96 %) ethanol. Fifteen µL of each mixture was transferred individually to plates containing Brazier's cefoxitin cyclo-serine egg-yolk (CCEY) medium (Lister, 2014). Plates were incubated anaerobically ($10\%H_2$, $10\%CO_2$ and $80\%N_2$) at 34 °C for 48 h. 'Broken glass' colonies, typical of *C. difficile*, were transferred to blood agar plates and incubated for a further 48 h. Chelex-100 chelating resin was used for whole-cell DNA extraction of the resulting colonies. PCR-amplification of the DNA was performed with BioMix Red mastermix (Bioline) and CD-16s primers, 5′-CTG GGG TGA AGT CGT AAC AAG G-3′, 6′-GCG CCC TTT GTA GCT TGA CC-3′ (Eurofins MWG). PCR products were transferred to a heating block, set to 75 °C, for 45 min to concentrate products to c.20 µL, before electrophoresis on a 3 % agarose gel with GelGreen nucleic
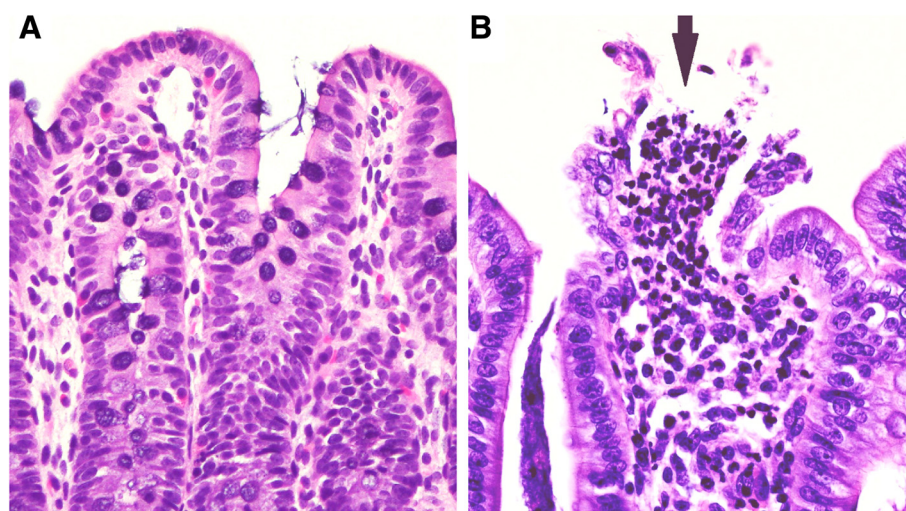
Fig. 1 Pig. Colon. **a** Normal colon lined by columnar epithelial cells. **b** Superficial, erosive colitis with infiltration of neutrophils into the lamina propria and effusion into the lumen – 'volcano lesion' (*arrow*) (Haematoxylin and eosin 20×)

McElroy *et al. Irish Veterinary Journal* (2016) 69:10

Page 3 of 4

acid stain (Biotium). The gel was placed under ultraviolet light in a closed chamber of Bio-Rad Universal Hood II. Results were visualised with Bio-Rad Quantity One (4-6-1) 1-dimensional analysis software. PCR ribotypes were successfully obtained for three piglets; two were ribotype 078 and one was ribotype 110. The fourth piglet's sample failed to yield a pure culture after treatment with 96 % ethanol and anaerobic incubation with CCEY medium.

Other laboratory testing

Clostridium perfringens (*C. perfringens*) was isolated by direct anaerobic culture of intestinal contents using pre-reduced 5 % Columbia sheep blood agar (SBA) and fastidious anaerobe blood agar with nheomycin (E&O Laboratories, Scotland). *C. perfringens* alpha toxin was detected in a pooled sample of small intestinal contents using a sandwich ELISA, testing for alpha (α), beta (β), epsilon (ε) toxins and *C. perfringens Antigen* (BioX Diagnostics, Belgium).

Rotavirus group B was detected in small intestinal contents in two of the piglets using modified versions of previously described PCR methods [7, 8]. PCRs for porcine coronaviruses and porcine reproductive and respiratory virus were negative using modified versions of previously described methods [9–12]. No antigen was detected using anti-*Cryptosporidium parvum* monoclonal antibody labelled with fluorescein isothiocyanate (Bio-X Diagnostics, Belgium, Catalogue Number BIO 073).

Conclusions

This is the first confirmed report of typhlocolitis associated with *C. difficile* in Irish pigs. *C. difficile* infection in the four piglets here appeared to be acting only as one component in a multifactorial diarrhoea that also involved Group B rotaviruses and possibly *Clostridium perfringens* Type A. Ongoing surveillance is required to determine the relative significance of *C. difficile* in porcine neonatal diarrhoea in Ireland.

This is also the first report of PCR-ribotyping of *C. difficile* isolates from clinically affected Irish pigs. There are a number of different molecular methods available for strain typing, most commonly polymerase chain reaction (PCR)-based ribotyping, multilocus variable number tandem repeat analysis (MLVA), pulsed field gel electrophoresis (PFGE) and whole genome sequencing (WGS) [13, 14]. PCR ribotype 078 is the most commonly reported isolate from pigs in most studies, including a recent Irish study in pigs of different ages [2, 14–16]. In one study in humans this strain was shown to have increased by at least 6-fold from 2000 to 2008 [17]. Moreover, genetically indistinguishable *C. difficile* 078 stains have been found in pigs and farmers, indicating interspecies transmission but the route of transmission

has not been determined [18]. Therefore there is increasing interest in *C. difficile* as a 'One Health' issue. Our findings of two ribotypes in this one herd are consistent with previous reports of diversity of *C. difficile* ribotypes amongst pigs in an international study of animal associated strains, and a previous report from Germany [19, 20].

There are approximately 1600 new cases of human *C. difficile* infections per annum in Ireland, 21.5 % of which are recognised to be community-acquired infections i.e. patients had not had admission to a healthcare facility for 12 weeks preceding symptom onset [21]. A UK study has shown that only 38 % of CDAD in hospital in-patients can be attributed to transmissions within the hospital [22]. This has led to increased interest in the 'One Health' epidemiology of *C. difficile*. PCR ribotype results were available for only 16 % of human cases in Ireland in 2014: ribotypes 078 and 014 were most frequently reported, both 11 % of known results [21].

This report has confirmed that *C. difficile* is present in Irish pigs and is associated with typhlocolitis as part of multifactorial diarrhoea. Further studies are warranted to determine the prevalence of *C. difficile* in Irish pigs and other animal species. In addition, molecular typing studies using conventional methods such as PCR-based ribotyping or whole genome sequencing may provide information on any relationship that may exist between strains from animals and humans in Ireland.

Availability of data and materials
Paraffin-embedded tissue sections and *Clostridium difficile* isolates are available.

Authors' contributions
MMcE and SMcG carried out the necropsies. MMcE interpreted the histopathology and drafted the paper. GM did the PCR ribotyping and drafted part of the paper. MH and AOD carried out bacteriology and toxin ELISAs. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
This paper relates to animals submitted for necropsy. All information relating to the farm of origin is anonymous. The need for written consent is not applicable.

Ethics approval and consent to participate
This case report relates to dead animals referred for necropsy so is exempt from ethics approval.

Author details
[1]Pathology Division, Central Veterinary Research Laboratory, Backweston Campus, Celbridge, Co. Kildare, Ireland. [2]Department of Clinical Microbiology, Trinity College Dublin, Dublin, Ireland.

McElroy *et al. Irish Veterinary Journal* (2016) 69:10

Page 4 of 4

References

1. Songer JG. The emergence of *Clostridium difficile* as a pathogen of food animals. Anim Health Res Rev. 2004;5:321–6.
2. Debast SB, van Leengoed LA, Goorhuis A, Harmanus C, Kuijper EJ, Bergwerff AA. *Clostridium difficile* PCR ribotype 078 toxinotype V found in diarrhoeal pigs identical to isolates from affected humans. Environ Microbiol. 2009;2:505–11.
3. Álvarez-Pérez S, Blanco JL, Bouza E, Alba P, Gibert X, Maldonado J, García ME. Prevalence of *Clostridium difficile* in diarrhoeic and non-diarrhoeic piglets. Vet Microbiol. 2009;137:302–5.
4. Songer JG, Uzal FA. Clostridial enteric infections in pigs. J Vet Diagn Invest. 2005;17:528–36.
5. Squire MM, Riley TV. Clostridium difficile infection in humans and piglets: a 'One Health' opportunity. Curr Top Microbiol Immunol. 2013;365:299–314.
6. Keel MK, Songer JG. The comparative pathology of Clostridium difficile-associated disease. Vet Pathol. 2006;43:225–40.
7. Marthaler D, Homwong N, Rossow K, Culhane M, Goyal S, Collina J, et al. Rapid detection and high occurrence of porcine rotavirus A, B, and C by RT-qPCR in diagnostic samples. J Virol Methods. 2014;209:30–4.
8. Jothikumar N, Kang G, Hill VR. Broadly Reactive TaqMan® assay for real-time RT-PCR detection of rotavirus in clinical and environmental samples. J Virol Methods. 2009;155:126–31.
9. Seong-Hee K, In-Joong K, Hyun-Mi P, Dong-Seob T, Jae-Young S, Bang-Hun H. Multiplex real-time RT-PCR for the simultaneous detection and quantification of transmissible gastroenteritis virus and porcine epidemic diarrhea virus. J Virol Methods. 2007;146:172–7.
10. Mardassi H, Mounir S, Dea S. Identification of major differences in the nucleocapsid protein genes of a Quebec strain and European strains of porcine reproductive and respiratory syndrome virus. J Gen Virol. 1994;75:681–5.
11. Egli C, Thur B, Liu L, Hofmann M. Quantative TaqMan RT-PCR for the detection and differentiation of European and North American strains of porcine reproductive and respiratory syndrome virus. J Virol Methods. 2001;98:63–75.
12. Frossard JP, Fearnley C, Naidu B, Errington J, Westcott DG, Drew TD. Porcine reproductive and respiratory syndrome virus: Antigenic and molecular diversity of British isolates and implication of diagnosis. Vet Microbiol. 2012;158:308–15.
13. Brazier JS. Typing of Clostridium difficile. Clin Microbiol Infect. 2001;7:428–31.
14. Marsh JW, Harrison LH. Clostridium difficile: a food safety concern. In: Chen CY, Yan X and Jackson CR, editors. Antimicrobial Resistance and Food safety: Methods and Techniques. London: Elsevier; 2015. p. 181–206.
15. Keel K, Brazier JS, Post KW, Weese S, Songer JG. Prevalence of PCR ribotypes among Clostridium difficile isolates from pigs, calves, and other species. J Clin Microbiol. 2007;45:1963–4.
16. Stein K, Egan S, Lynch H, Harmanus C, Kyne L, Herra C, et al. Prevalence and PCR ribotype distribution of *Clostridium difficile* in pigs in Ireland. In: Proceedings 26th European Congress of Clinical Microbiology and Infectious Diseases. 2016.
17. Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, et al. Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clin Infect Dis. 2008;47:1162–70.
18. Knetsch CW, Connor TR, Mutreja A, van Dorp SM, Sanders IM, Browne HP, et al. Whole genome sequencing reveals potential spread of *Clostridium difficile* between humans and farm animals in the Netherlands, 2002 to 2011. Euro Surveill. 2014;19(45). Article DOI: http://dx.doi.org/10.2807/1560-7917.ES2014.19.45.20954. Accessed 17 Aug 2016.
19. Janezic S, Zidaric V, Pardon B, Indra A, Kokotovic B, Blanco JL, et al. International Clostridium difficile animal strain collection and large diversity of animal-associated strains. BMC Microbiol. 2014;14:173–83.
20. Schneeberg A, Neubauer H, Schmoock G, Baier S, Harlizius J, Nienhoff H, Brase K, Zimmermann S, Seyboldt C. *Clostridium difficile* genotypes in piglet populations in Germany. J Clin Microbiol. 2013;51:3796–803.
21. *Clostridium difficile* infection. In: Health Protection Surveillance Centre Annual Epidemiological Report 2014 http://www.hpsc.ie/AboutHPSC/AnnualReports/File,15505,en.pdf. Accessed 17 Aug 2016.
22. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med. 2013;369:1195–205.

Human and Porcine Transmission of *Clostridioides difficile* Ribotype 078, Europe

Geraldine Moloney,[1] David W. Eyre,[1] Micheál Mac Aogáin, Máire C. McElroy, Alison Vaughan, Tim E.A. Peto, Derrick W. Crook, Thomas R. Rogers

Genomic analysis of a diverse collection of *Clostridioides difficile* ribotype 078 isolates from Ireland and 9 countries in Europe provided evidence for complex regional and international patterns of dissemination that is not restricted to humans. These isolates are associated with *C. difficile* colonization and clinical illness in humans and pigs.

Clostridioides (formerly *Clostridium*) *difficile* was considered to be a predominantly nosocomial pathogen until findings of several whole-genome sequencing studies suggested a more complex epidemiology. For example, Eyre et al. reported that only 35% of nosocomial *C. difficile* infections (CDIs) were potentially attributable to other cases on the basis of genomic data, and only 19% were additionally linked through sharing possible hospital-based contact (*1*). This finding suggests that a major proportion of *C. difficile* from CDI cases occurring in healthcare institutions originates from other sources, including the community (*2*).

Community-associated CDI (CA-CDI) is now well recognized, accounting for ≈25% of cases in Australia, <25% of cases in Europe, and 33% of cases in the United States (*3,4*). There is increasing recognition that *C. difficile* is a near ubiquitous environmental organism and that humans have widespread environmental exposure to it. *C. difficile* has been detected in samples from parks (24.6%); water sources, including rivers, lakes, and sea water; homes (17.1%); commercial stores; and other premises (6.5%–8.1%), in addition to hospitals (16.5%) (*5,6*). Isolates of *C. difficile*

Author affiliations: Trinity College Dublin, Dublin, Ireland
(G. Moloney, M. Mac Aogáin, T.R. Rogers); University of Oxford, Oxford, UK (D.W. Eyre, A. Vaughan, T.E.A. Peto, D.W. Crook); Central Veterinary Research Laboratory, Celbridge, Ireland (M.C. McElroy); St. James's Hospital, Dublin (T.R. Rogers)

from these studies underwent ribotype analysis. Overall, ribotype 027 isolates were most commonly identified in hospital samples, and ribotype 014–020 isolates predominated in other environmental samples. Isolates of the most common ribotypes were not restricted to any particular location (*5*). These findings support the possibility that there are different sources for exposure to each *C. difficile* ribotype.

Occurrence of CDI caused by *C. difficile* ribotype 027 has been greatly reduced in the United Kingdom, most likely the result of the combination of antimicrobial stewardship and hospital infection prevention and control measures. However, these interventions have not reduced the incidence of infections caused by other ribotypes, including ribotype 078 (*7*).

Findings of genomic analysis of isolates from the European, Multi-Center, Prospective, Biannual, Point-Prevalence Study of *Clostridium difficile* Infection in Hospitalized Patients with Diarrhea (EUCLID) showed that specific *C. difficile* ribotypes were associated with healthcare clusters, and other ribotypes had an international distribution across Europe (*8*). For example, ribotype 078 isolates did not cluster by their country of origin, indicating a complex distribution unrelated to nosocomial transmission. The mechanisms of transmission have not been identified, but might be related to the movement of food, other animal-derived products, or persons across Europe (*8*).

C. difficile carriage and infection has been well described in livestock and other animals (*3*); certain ribotypes of *C. difficile* are considered to be major ribotypes from a One Health perspective. These ribotypes include ribotype 078, carriage of which has been reported in 9%–100% of piglets from North America, Europe, Asia, and Australia (*3*). Carriage rates in calves (56%) and cows (13%) have been lower.

Although many studies did not identify any major carriage in adult pigs, 1 study in the Netherlands reported a rate ranging from 6.6% to 100% (*3*).

We have reported *C. difficile* ribotype 078 in cases of typhlocolitis in neonatal piglets in Ireland (*9*), and Knetsch et al. found that ribotype 078 isolates carried by farmers in the Netherlands and their pigs were identical by whole-genome sequence analysis (*10*). These findings suggest that *C. difficile* isolates might be shared between humans and pigs when in close proximity. However, the mechanisms and directions of transmission are not known.

In this study, we investigated the genomic relationships between *C. difficile* ribotype 078 isolates of human and porcine origin collected from Ireland and compared these with international ribotype 078 isolates. We also investigated the extent to which geographic proximity could explain clusters of clonal isolates.

Methods

Samples and Settings

Clinical isolates of *C. difficile* ribotype 078 were collected prospectively as part of an investigation of consecutive episodes of CDI conducted at St. James's Hospital (Dublin, Ireland), a 900-bed tertiary referral center, during 2013–2016. Stool samples, sent from patients with diarrhea, had the *C. difficile* toxin B gene identified by using the EntericBio PCR Kit (Serosep, https://www.serosep.com). We reviewed medical notes of inpatients to obtain relevant clinical data, including antimicrobial drugs and proton pump inhibitors prescribed before the onset of diarrhea, features indicative of severe CDI with or without complications, and the antimicrobial drugs used for management of CDI. These data were pseudonymized and stored in a dedicated database.

We retrieved an additional 9 *C. difficile* 078 isolates from a study of recurrent CDI at St. James's Hospital during 2012–2013 (*11*). Five additional *C. difficile* ribotype 078 isolates were provided from those submitted to a national surveillance study of CA-CDI in Ireland conducted during 2015. Isolates of *C. difficile* were recovered from pigs that had been referred for autopsy at the Central Veterinary Research Laboratory (CVRL; Backweston, Ireland) during 2014–2015, irrespective of the suspected cause of death, by sampling colonic contents or feces that had positive results for *C. difficile* toxins A/B by using the Premier Elisa Kit (Meridian BioScience Inc., https://www.meridianbioscience.com). We treated human fecal and porcine colonic/fecal samples with ethanol shock before anaerobic incubation on cycloserine cefoxitin egg yolk medium. DNA was extracted from resulting colonies for PCR ribotype analysis and Illumina (https://www.illumina.com) genomic library preparation as described (*11*).

Whole-Genome Sequencing

Whole-genome sequencing was performed either on an Illumina MiSeq or MiniSeq platform at Trinity College (Dublin, Ireland) or on the Illumina HiSeq platform at the Wellcome Centre for Human Genetics, University of Oxford (Oxford, UK). Sequence data generated have been deposited in the National Center for Biotechnology Information Short Read Archive (https://www.ncbi.nlm.nih.gov/sra) under BioProject PRJNA692997.

We mapped sequence reads to the ribotype 078 reference genome M120 (GenBank accession no. FN665653.1), and identified high-quality variants by using an approach developed and calibrated for *C. difficile* (*1*) with later refinements (*12*) (Appendix, https://wwwnc.cdc.gov/EID/article/27/9/20-3468-App1.pdf). We obtained published comparison sequences from the EUCLID pan-European cross-sectional survey conducted during in 2012–2013 (*8*) and from farm animal and human isolates from the Netherlands (2002–2011) described by Knetsch et al. (*10*).

Sequence Comparisons

We compared sequences by using single-nucleotide polymorphisms (SNPs) and obtained differences between sequences from maximum-likelihood phylogenies corrected for recombination (Appendix). We reviewed phylogenetic analysis of closely related genomes in conjunction with available epidemiologic data. Within the clinical database, CDI recurrence was defined as identification of 2 isolates within 10 SNPs from 1 patient (*1*) for which that patient had clearly documented clinical resolution of symptoms after their first episode. On the basis of rates of *C. difficile* evolution and within-host diversity (*1*), we defined plausible, short-term, transmission/mutual exposure as isolates differing by 0–2 SNPs.

We made epidemiologic matches between patients who had in-patient admissions and demonstrable links with respect to time, location, or healthcare staff, where their *C. difficile* isolates were within 0–2 SNPs. Because epidemiologic details were not available for either the CA-CDI investigation in Ireland or the EUCLID isolates, we analyzed linkage between cases on the basis of genetic similarity alone. These genomic pairs were named by the isolate sources in chronologic order of identification.

Ethics

Investigation of hospital-associated CDI (HA-CDI) cases at St James's Hospital was conducted after obtaining approval from the St. James's Hospital/Tallaght Research Ethics Committee. Porcine isolates were exempt from requiring ethics approval.

Results

A total of 171 *C. difficile* ribotype 078 isolates were included in the analysis: 53 isolates from CDI episodes in 44 inpatients at St. James's Hospital, including 5 community-associated isolates; 20 porcine isolates from Ireland; 67 clinical, farmer, and porcine isolates from the Netherlands; and 31 clinical EUCLID isolates. We provide details of their country of origin, source, and date of isolation (Table 1). The EUCLID isolates were obtained from 9 countries in Europe. Six countries, including Ireland, submitted >2 isolates.

Of the 53 isolates causing CDI in Ireland, 9 were from recurrent CDI episodes in 7 patients (7 subsequent isolates were 0 SNPs different from, the baseline isolate, 1 was 1 SNP different, and 1 was 8 SNPs different). Only the first isolate from each patient was considered in subsequent analyses. We provide genomic relationships between the remaining 162 ribotype 078 isolates (Figure). Despite the diverse sampling frame, only limited diversity was seen; the greatest root-to-tip distance in the phylogenetic tree was 48 SNPs.

Isolates from Ireland were found throughout the tree, but specific clusters of these isolates were seen,

including, as shown at the ≈240° (≈8 o'clock) position (Figure), a cluster of cases that included isolates from HA-CDI and CA-CDI cases as well as cases from pigs. Within this cluster, several porcine isolates were directly ancestral to 1 HA-CDI case. Another 5 CDI cases, including 1 CA-CDI, had another porcine isolate directly ancestral. This finding suggests a porcine origin for these cases, either directly or by >1 or more intermediate (unsampled) transmission routes. This same cluster also contained an isolate from a pig and a farmer from the Netherlands. Several other clinical isolates from the Netherlands were closely related to porcine isolates (Figure).

We provide epidemiologic links between genetically related isolates within 0–2 SNPs (Table 2). Although nearly all genomic pairs occurred among isolates with the same country of origin, the epidemiologic information available can explain only a small proportion of transmissions/mutual exposures.

Discussion

Our findings support a complex regional and international distribution of *C. difficile* ribotype 078 isolates. In contrast to the EUCLID study, which obtained samples on single days in winter and summer, more dense sampling was undertaken in our study. In the EUCLID study, no evidence of clustering of ribotype 078 within countries was seen, which is consistent with a complex pattern of dissemination in Europe over timescales spanning years (Figure). However, our study showed evidence of sublineages of ribotype 078 that are predominantly found in isolates from the Netherlands and others predominantly found in isolates from Ireland (Figure). It is likely that this denser sampling has enabled recent, local, onward transmission to be better captured. We also identify a EUCLID isolate from Italy (2013) and a CA-CDI isolate from Dublin, Ireland (2014), that are within 2 SNPs, which is consistent with a temporally related transmission. However, we do not know of any epidemiologic link between these 2 cases.

For 10 pairs of isolates within 2 SNPs from inpatients who had HA-CDI, possible healthcare-based epidemiologic links could be made for 6 of these pairs but not the other 4. Plausible ward-based transmission only accounted for 3 pairs. For other genetically related isolates pertaining to inpatients in our study, there was a median of 559 days between their associated CDI episodes (range 147–651 days) without overlapping hospital admissions or appointments. Overall, nosocomial transmission accounted for 15% of closely genetically related (<2 SNPs) *C. difficile* ribotype 078 cases in this study, and

Table 1. Countries from which *Clostridioides difficile* 078 isolates originated, their identified sources, and timeframe of collection*

Origin and source of isolates	Timeframe of collection	No. isolates
Ireland (*11*)		
HA-CDI	2012–2016	48†
Porcine	2014–2015	20
CA-CDI	2015 Apr–Jun	5
Netherlands (*10*)		
CDI	2002–2011	31
Porcine	2009, 2011	20
Healthy farmers	2011	16
EUCLID (*8*), HA-CDI	2012 Dec–2013 Aug	
Germany		9
Italy		7
United Kingdom		4
France		3
Portugal		3
Ireland		2
Spain		1
Greece		1
Austria		1

*CDI, *C. difficile* infection; EUCLID, European, Multi-Center, Prospective, Biannual, Point-Prevalence Study of *Clostridium difficile* Infection in Hospitalized Patients with Diarrhea; HA-CDI, hospital-associated CDI.
†Includes 9 isolates from HA-CDI cases (*11*).
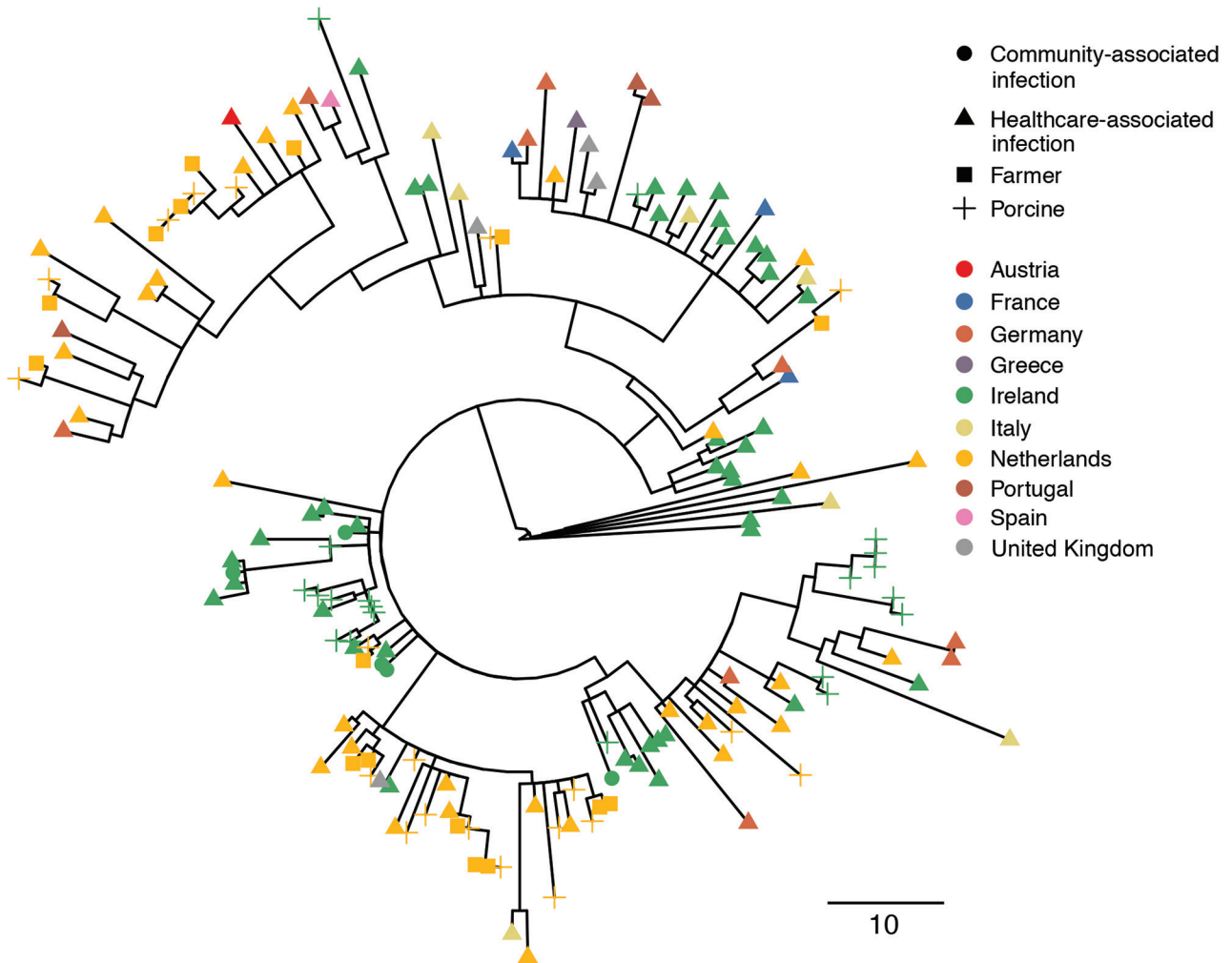
Figure. Recombination-adjusted maximum-likelihood phylogenetic tree of sequences from human and porcine *Clostridioides difficile* isolates from Ireland and 9 other countries in Europe. Isolates are shown as triangles for healthcare-associated *C. difficile* cases and circles for community-associated *C. difficile* cases. Isolates from pigs are shown as crosses and those from farmers as squares. The color at each tip indicates the country of origin of the isolate. The tree was based on 4,861 variable sites before correction for recombination, based on a median (interquartile ranges) of 93.4% (93.0%–93.8%) and (83.1%–96.2%) of the reference genome being called. Scale bar indicates single-nucleotide polymorphisms.

equal proportions were attributable to farms and unknown transmission routes. In a study in Leeds, UK, which had comparable phylogenetic analysis, hospital ward-based epidemiologic linkage was reported as 11% for ribotype 078 cases versus 64% for ribotype 027 cases (*13*).

A EUCLID isolate from Ireland (2013) forms a genomic cluster with 1 CA-CDI isolate (2015) and 2 HA-CDI isolates (July 2015 and December 2015). These 4 isolates were from patients in 3 Dublin healthcare facilities and from 1 case of CA-CDI that had been collected within a 3-year period. This finding suggests shared exposure across the greater Dublin area, and that nosocomial transmission is not the dominant route of acquisition of *C. difficile* ribotype 078.

This observation is consistent with the EUCLID study findings (*8*).

It is not clearly understood how persons who have CA-CDI acquired their infection because they do not have the risk factors for HA-CDI (*14*). Anderson et al. described proximity to livestock farms, agricultural industry, and nursing home facilities as risk factors for CA-CDI in North Carolina, USA, but they did not include analysis of *C. difficile* molecular data in their models (*15*). In contrast, Van Dorp et al. found no evidence of either localized point sources or livestock exposure as risk factors for *C. difficile* acquisition in the Netherlands (*16*). They included ribotype detail in their analysis, but found no evidence of geographic clustering of ribotype 078 CDI cases (*16*). This finding

Table 2. Pairs of *Clostridioides difficile* ribotype 078 isolates matched by country of origin and source case, with associated epidemiology*

Country	Source of isolate(s)	Country 2	Source of isolate(s)	No. pairs of isolates	Associated epidemiology
Ireland	CA-CDI	Ireland	CA-CDI	2	No known links
Ireland	CA-CDI	Ireland	HA-CDI	2	No known links
Ireland	HA-CDI	Ireland	HA-CDI	10	Possible transmission 6 pairs,† unknown for 4 pairs
Ireland	Porcine	Ireland	HA-CDI	3	No known links
Ireland	Porcine	Ireland	Porcine	12	8 pairs at 1 farm, 3 pairs at 1 farm, 1 pair at 1 farm, no pairs between farms
Ireland	CA-CDI	Italy	HA-CDI	1	Unknown
Ireland	HA-CDI	United Kingdom	HA-CDI	1	Unknown
Germany	HA-CDI	Germany	HA-CDI	1	Unknown
Netherlands	HA-CDI	Netherlands	HA-CDI	1	Unknown
Netherlands	CDI	Netherlands	Farmer	1	No known links
Netherlands	CDI	Netherlands	Porcine	1	No known links
Netherlands	Farmer	Netherlands	Farmer	3	Unknown
Netherlands	Farmer	Netherlands	Porcine	10	Farm exposures
Netherlands	Porcine	Netherlands	Porcine	1	No known links
Portugal	HA-CDI	Portugal	HA-CDI	1	Unknown

*CA-CDI, community-assocated *C. difficile* infection; HA-CDI, healthcare-associated *C. difficile* infection.
†The 6 possible healthcare-associated transmission pairs shared time and space on the same hospital ward (n = 3) or time on different hospital wards while under the care of the same medical team (n = 3).

is consistent with that of Knetsch et al., who reported clonal isolates of farm and clinical origin without a geographic basis for those clusters (*10*).

Knetsch et al. identified another genomic cluster of *C. difficile* ribotype 078 isolates, which included an isolate of animal origin from Canada (2004) and 8 isolates of clinical origin from the United Kingdom (2008–2012) (*17*). We also identified a cluster of clinical and porcine 078 isolates from Ireland, where there was no known occupational exposure of the affected patients who lived in urban locations far from relevant pig farms. Knight et al. reported clonal ribotype 014 isolates from Australia that were considerable geographic distances from each other, which is suggestive of long-range transmission and major community reservoirs (*18*). They concluded that this transmission was unlikely to have been caused by direct contact between the humans and animals involved, and suggested that by-products, such as manure or compost, could enable indirect transmission from animals and humans (*18*). In a study from the United States, biosolid-based compost had the highest rate of *C. difficile* recovery that included ribotype 078 isolates (*19*), which was also the most common ribotype in an investigation of manure from Japan (*20*).

Findings based on ribotype analysis alone are insufficient for clear identification of transmission events pertaining to community reservoirs (*21*). Moradigaravand et al. identified ≈90% of their collection of clinical and wastewater isolates as clade 1 (231/256), and only 10 (3.9%) as clade 5/ribotype 078 (*22*). When their ribotype 078 isolates were com-

pared with the same isolates from the Netherlands included in our analysis, they found divergence of ≈20 years between the isolates from the United Kingdom and the Netherlands. This finding suggests that water is not the primary reservoir or route for dissemination of *C. difficile* ribotype 078 isolates. It is still considered possible that dissemination of ribotype 078 isolates occurs by the food chain, the environment, or both (*23,24*). This view is supported by the presence and distribution of tetracycline-resistant determinants in *C. difficile* genomes, reflecting the antimicrobial drug selection pressure from tetracycline use in agriculture or veterinary practice, and thereby facilitating emergence and spread of ribotype 078 bacteria (*24*).

It is not completely understood how some livestock might have asymptomatic *C. difficile* colonization, whereas others show development of infection (*25*). The porcine isolates from Ireland in this analysis were from available samples processed at the CVRL. These isolates included samples from neonatal piglets that had typhlocolitis (*9*). We have identified genomic similarities among isolates causing human and veterinary infections. This finding augments the need for a One Health approach for *C. difficile* ribotype 078.

The strengths of this analysis include the large number of *C. difficile* ribotype 078 isolates included, from different sources including humans and animal species, and geographic origin. The limitations of this study include the lack of epidemiologic data available to the investigators for CA-CDI and the limited number of porcine strains from samples available at

the CVRL. In conclusion, our analysis of *C. difficile* ribotypes 078 isolates from Ireland and 9 other countries in Europe showed close overlap between isolates from humans and pigs, including the occurrence of plausible transmission, either directly or by an unknown intermediate source.

About the Author

Dr. Moloney is an infectious diseases physician at Cork University Hospital, Cork, Ireland. Her primary research interest is infections with *Clostridioides difficile*.

References

1. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med. 2013;369:1195–205. https://doi.org/10.1056/NEJMoa1216064
2. Moloney G, Mac Aogáin M, Kelleghan M, O'Connell B, Hurley C, Montague E, et al. Possible interplay between hospital and community transmission of a novel *Clostridium difficile* sequence type 295 recognized by next-generation sequencing. Infect Control Hosp Epidemiol. 2016;37:680–4. https://doi.org/10.1017/ice.2016.52
3. Knight DR, Riley TV. Genomic delineation of zoonotic origins of *Clostridium difficile*. Front Public Health. 2019;7:164. https://doi.org/10.3389/fpubh.2019.00164
4. European Centre for Disease Prevention and Control. Healthcare-associated infections: *Clostridium difficile* infections. In: Annual epidemiological report for 2016. Stockholm: The Centre, 2018 [cited 2021 Jun 5]. https://www.ecdc.europa.eu/en/publications-data/healthcare-associated-infections-clostridium-difficile-infections-annual
5. Alam MJ, Walk ST, Endres BT, Basseres E, Khaleduzzaman M, Amadio J, et al. Community environmental contamination of toxigenic *Clostridium difficile*. Open Forum Infect Dis. 2017;4:ofx018. https://doi.org/10.1093/ofid/ofx018
6. al Saif N, Brazier JS. The distribution of *Clostridium difficile* in the environment of South Wales. J Med Microbiol. 1996;45:133–7. https://doi.org/10.1099/00222615-45-2-133
7. Dingle KE, Didelot X, Quan TP, Eyre DW, Stoesser N, Golubchik T, et al.; Modernising Medical Microbiology Informatics Group. Effects of control interventions on *Clostridium difficile* infection in England: an observational study. Lancet Infect Dis. 2017;17:411–21. https://doi.org/10.1016/S1473-3099(16)30514-X
8. Eyre DW, Davies KA, Davis G, Fawley WN, Dingle KE, De Maio N, et al.; EUCLID Study Group. Two distinct patterns of *Clostridium difficile* diversity across Europe indicating contrasting routes of spread. Clin Infect Dis. 2018;67:1035–44. https://doi.org/10.1093/cid/ciy252
9. McElroy MC, Hill M, Moloney G, Mac Aogáin M, McGettrick S, O'Doherty Á, et al. Typhlocolitis associated with *Clostridium* difficile ribotypes 078 and 110 in neonatal piglets from a commercial Irish pig herd. Ir Vet J. 2016;69:10. https://doi.org/10.1186/s13620-016-0070-9
10. Knetsch CW, Connor TR, Mutreja A, van Dorp SM, Sanders IM, Browne HP, et al. Whole genome sequencing reveals potential spread of *Clostridium difficile* between humans and farm animals in the Netherlands, 2002 to 2011. Euro Surveill. 2014;19:20954. https://doi.org/10.2807/1560-7917.ES2014.19.45.20954
11. Mac Aogáin M, Moloney G, Kilkenny S, Kelleher M, Kelleghan M, Boyle B, et al. Whole-genome sequencing improves discrimination of relapse from reinfection and identifies transmission events among patients with recurrent *Clostridium difficile* infections. J Hosp Infect. 2015;90:108–16. https://doi.org/10.1016/j.jhin.2015.01.021
12. De Silva D, Peters J, Cole K, Cole MJ, Cresswell F, Dean G, et al. Whole-genome sequencing to determine transmission of *Neisseria gonorrhoeae*: an observational study. Lancet Infect Dis. 2016;16:1295–303. https://doi.org/10.1016/S1473-3099(16)30157-8
13. Martin JS, Eyre DW, Fawley WN, Griffiths D, Davies K, Mawer DP, et al. Patient and strain characteristics associated with *Clostridium difficile* transmission and adverse outcomes. Clin Infect Dis. 2018;67:1379–87. https://doi.org/10.1093/cid/ciy302
14. Khanna S, Pardi DS, Aronson SL, Kammer PP, Baddour LM. Outcomes in community-acquired *Clostridium difficile* infection. Aliment Pharmacol Ther. 2012;35:613–8. https://doi.org/10.1111/j.1365-2036.2011.04984.x
15. Anderson DJ, Rojas LF, Watson S, Knelson LP, Pruitt S, Lewis SS, et al.; CDC Prevention Epicenters Program. Identification of novel risk factors for community-acquired *Clostridium difficile* infection using spatial statistics and geographic information system analyses. PLoS One. 2017;12:e0176285. https://doi.org/10.1371/journal.pone.0176285
16. van Dorp SM, Hensgens MP, Dekkers OM, Demeulemeester A, Buiting A, Bloembergen P, et al. Spatial clustering and livestock exposure as risk factor for community-acquired *Clostridium difficile* infection. Clin Microbiol Infect. 2019;25:607–12. https://doi.org/10.1016/j.cmi.2018.07.018
17. Knetsch CW, Kumar N, Forster SC, Connor TR, Browne HP, Harmanus C, et al. Zoonotic transfer of *Clostridium difficile* harboring antimicrobial resistance between farm animals and humans. J Clin Microbiol. 2018;56:e01384-17. https://doi.org/10.1128/JCM.01384-17
18. Knight DR, Squire MM, Collins DA, Riley TV. Genome analysis of *Clostridium difficile* PCR ribotype 014 lineage in Australian pigs and humans reveals a diverse genetic repertoire and signatures of long-range interspecies transmission. Front Microbiol. 2017;7:2138. https://doi.org/10.3389/fmicb.2016.02138
19. Dharmasena M, Jiang X. Isolation of toxigenic *Clostridium difficile* from animal manure and composts being used as biological soil amendments. Appl Environ Microbiol. 2018;84:e00738-18. https://doi.org/10.1128/AEM.00738-18
20. Usui M, Kawakura M, Yoshizawa N, San LL, Nakajima C, Suzuki Y, et al. Survival and prevalence of *Clostridium difficile* in manure compost derived from pigs. Anaerobe. 2017;43:15–20. https://doi.org/10.1016/j.anaerobe.2016.11.004

21. Knetsch CW, Lawley TD, Hensgens MP, Corver J, Wilcox MW, Kuijper EJ. Current application and future perspectives of molecular typing methods to study *Clostridium difficile* infections. Euro Surveill. 2013;18:20381. https://doi.org/10.2807/ese.18.04.20381-en
22. Moradigaravand D, Gouliouris T, Ludden C, Reuter S, Jamrozy D, Blane B, et al. Genomic survey of *Clostridium difficile* reservoirs in the East of England implicates environmental contamination of wastewater treatment plants by clinical lineages. Microb Genom. 2018;4. https://doi.org/10.1099/mgen.0.000162
23. Hensgens MP, Keessen EC, Squire MM, Riley TV, Koene MG, de Boer E, et al.; European Society of Clinical Microbiology and Infectious Diseases Study Group for Clostridium difficile (ESGCD). *Clostridium difficile*
24. Dingle KE, Didelot X, Quan TP, Eyre DW, Stoesser N, Marwick CA, et al. A Role for tetracycline selection in recent Evolution of agriculture-associated *Clostridium difficile* PCR ribotype 078. MBio. 2019;10:e02790-18. https://doi.org/10.1128/mBio.02790-18
25. Weese JS. *Clostridium* (*Clostridioides*) *difficile* in animals. J Vet Diagn Invest. 2020;32:213–21. https://doi.org/10.1177/1040638719899081

infection in the community: a zoonotic disease? Clin Microbiol Infect. 2012;18:635–45. https://doi.org/10.1111/j.1469-0691.2012.03853.x

Address for correspondence: Geraldine Moloney, Department of Infectious Diseases, Cork University Hospital, Wilton, Cork, Ireland; email: geraldinemoloney@physicians.ie

Human and Porcine Transmission of *Clostridioides difficile* Ribotype 078, Europe

Appendix

Supplementary Methods

Mapping and Variant Calling

Reads were mapped by using Stampy version 1.0.23 (https://www.well.ox.ac.uk/research/research-groups/lunter-group/lunter-group/stampy) without Burrows-Wheeler Aligner premapping by using an expected substitution rate of 0.01. Samples were compared by using single-nucleotide polymorphisms (SNPs) identified with Samtools mpileup version 1.4.1 (www.htslib.org/doc/1.4.1/samtools.html) with the extended base-alignment quality flag. Python scripts with inputs from Samtools, Genome Analysis Toolkit version 3.7.0 (https://gatk.broadinstitute.org), Picard tools version 1.123 (https://broadinstitute.github.io/picard/), and vcftools version 0.1.9 (https://vcftools.github.io/index.html) were used to generate annotated variant call format files and for subsequent quality filtering. Filters included requiring an SNP quality score $\geq$25, a per base mapping score $\geq$30, a consensus $\geq$90% to support a SNP, and calls were required to be homozygous under a diploid model. Only SNPs supported by $\geq$5 reads, including 1 in each direction were accepted. SNPs were not called in repetitive regions of the genome identified by BLAST (https://blast.ncbi.nlm.nih.gov) to search for repeat regions >100 bp in length. Filtering rules were based on previous sequencing of technical replicates of bacterial genomes by using the same DNA pool (e.g., in Eyre et al. [*1*]), including visual inspection of alignments and chosen to keep the false-positive SNP rate to $\approx$1/100 Mb of genome sequenced. A containerized implementation of the pipeline used is available (https://github.com/oxfordmmm/CompassCompact).

Sequence Comparisons

Sequences in which <70% of the reference sequence was mapped were excluded from the analysis. To improve computational efficiency in identifying closely related sequences,

sequences within ≤500 SNPs of any other sequence were initially pooled into groups. For each group of sequences within ≤500 SNPs, initial maximum-likelihood phylogenetic trees were constructed by using PhyML version 3.0 (http://www.atgc-montpellier.fr), a generalized time-reversible substitution model, and the BEST tree topology search operation option. These trees were then adjusted to remove unrecombining regions by using ClonalFrameML version 1.25 (https://github.com/xavierdidelot/ClonalFrameML) and default parameters. Each recombination adjusted phylogenetic tree was used to determine the number of SNPs between all pairs of sequences (i.e., the patristic distance between them). An example implementation of this approach is available (https://github.com/davideyre/runListCompare).

Reference

1. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med. 2013;369:1195–205. PubMed https://doi.org/10.1056/NEJMoa1216064