



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

M. Mac Aogáin<sup>a,\*</sup>, G. Moloney<sup>a</sup>, S. Kilkenny<sup>a</sup>, M. Kelleher<sup>b</sup>, M. Kelleghan<sup>b</sup>,  
B. Boyle<sup>b</sup>, T.R. Rogers<sup>a,b</sup>

<sup>a</sup> Department of Clinical Microbiology, Sir Patrick Dun Translational Research Laboratory, School of Medicine, Trinity College Dublin, St James's Hospital, Dublin, Ireland

<sup>b</sup> Department of Microbiology, St James's Hospital, Dublin, Ireland

## ARTICLE INFO

### Article history:

Received 12 August 2014

Accepted 19 January 2015

Available online 26 February 2015

### Keywords:

*Clostridium difficile* infection

Recurrent infection

Transmission

Whole-genome sequencing



## SUMMARY

**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  $\geq 2$  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  $\geq 2$  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](mailto: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.

© 2015 The Healthcare Infection Society. Published by Elsevier Ltd. All rights reserved.

## 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.<sup>1–4</sup> Recurrence is defined as a CDI episode occurring within eight weeks of a previous infection.<sup>5</sup> 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.<sup>1,3,6,7</sup> 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.<sup>8,9</sup> 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.<sup>2,3,7,10–12</sup>

The use of whole-genome sequencing (WGS) has provided evidence for a higher degree of *C. difficile* strain diversity than previously acknowledged.<sup>13–15</sup> 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.<sup>15</sup> 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.<sup>15</sup> This underscores the existence of unidentified *C. difficile* transmission routes.<sup>15–17</sup> In a subsequent study, Eyre *et al.* applied WGS to recurrent CDI with the consideration that ribotyping may underestimate reinfections caused by endemic ribotypes.<sup>18</sup> 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.<sup>18</sup>

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  $\geq 2$  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<sub>2</sub>, 10% CO<sub>2</sub> and 80% N<sub>2</sub>) at 37°C for 48–72 h.

### Strain collection

Stool samples from patients suffering  $\geq 2$  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.<sup>19</sup> PCR-based ribotyping was performed on all isolates to establish strain relatedness.<sup>20</sup>

### 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.<sup>21,22</sup> Strains were sequenced to an average raw read depth of  $91.1 \pm 44.5$ -fold. Sequence types (ST) were determined using the *Clostridium difficile* Multi Locus Sequence Typing website (<http://pubmlst.org/cdifficile/>).<sup>23</sup> 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*.<sup>14</sup>

### 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).<sup>5</sup> 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.

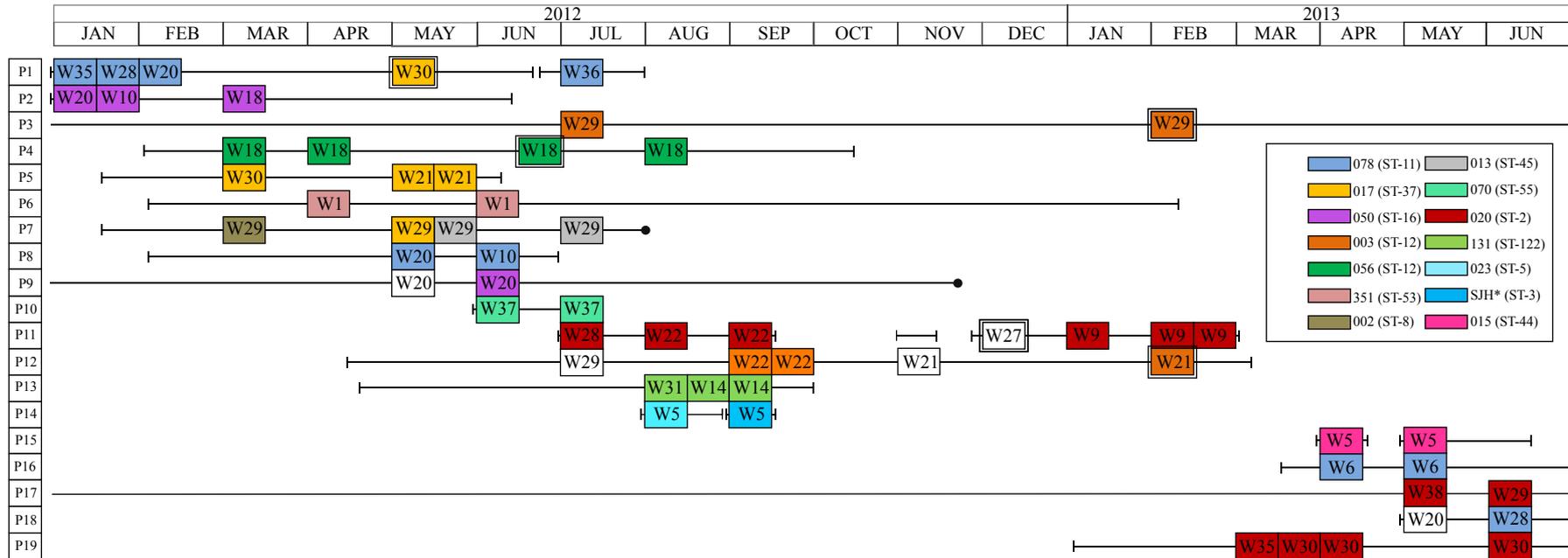
### Investigation of recurrent isolates by WGS

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.<sup>24</sup> 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  $\leq 2$  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*.<sup>18</sup>

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 <sup>a</sup>
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 <sup>a</sup>
Time to first recurrence of CDI <sup>b</sup>	
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.

<sup>a</sup> Data skewed by inclusion of two patients receiving long-stay care.

<sup>b</sup> 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  $\leq 2$  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  $\leq 2$  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.<sup>25</sup> However, our recurrent CDI cohort had an exceptional LOS which placed them in the minority (3.3%) of inpatient admissions nationally.<sup>25</sup> Even within this category, the national mean LOS is estimated to be 65.5 days, considerably shorter than our patients' experience.<sup>25</sup> 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.<sup>25</sup> 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.<sup>26</sup> Strains belonging to ribotypes 020 and 017 were also present. All three ribotypes have proven virulence potential and have been implicated in recurrent CDI.<sup>27–29</sup> 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 1; P1 and P7). Similar findings have previously been described and highlight the complex epidemiological scenarios that arise among patients with recurrent CDI.<sup>2,30</sup> 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 <sup>a</sup>	Variant <sup>b</sup>	Synonymous <sup>c</sup>	Protein alteration	Locus tag <sup>d</sup>	Gene function
P2	050	ST-16	1412874	T	C	No	V93A	CD630_12140	<i>spo0A</i> ; stage 0 sporulation protein A
			3243804	C	A	No	A176S	CD630_27870	
P5	017	ST-37	1826371	A	C	No	S215R	CD630_15770	<i>cwp84</i> ; cell surface protein <i>pgm</i> ; alpha-phosphoglucomutase <i>murC</i> ; UDP- <i>N</i> -acetylmuramate-L-alanine ligase
			4111335	G	A	No	G149S	CD630_35180	
P7	002	ST-45	1186090	G	T	No	P36Q	CD630_10160	Transcriptional regulator, MerR family Putative acetyltransferase
			1484356	G	A	Yes	na	CD630_12770	
P8	078	ST-11	3810191_3810192insT	AT	ATT	No	Frameshift	CD630_32550	<i>rgaR</i> ; two-component response regulator VirR-like
P12	003	ST-12	2708584_2708585insC	TC	TCC	No	Frameshift	CD630_23410	<i>abfD</i> ; 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 <i>atpA</i> <sup>e</sup> ; ATP synthase subunit alpha
			4061875	C	T	No	R283C	CD630_34700	
P1 (vs P5) <sup>f</sup>	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	<i>acoB</i> ; acetoin dehydrogenase E1 component

SNVs, single nucleotide variants; ST, sequence type; na, not applicable.

<sup>a</sup> Sequence identity at relative SNV locus in *C. difficile* 630 (AM180355) reference genome.

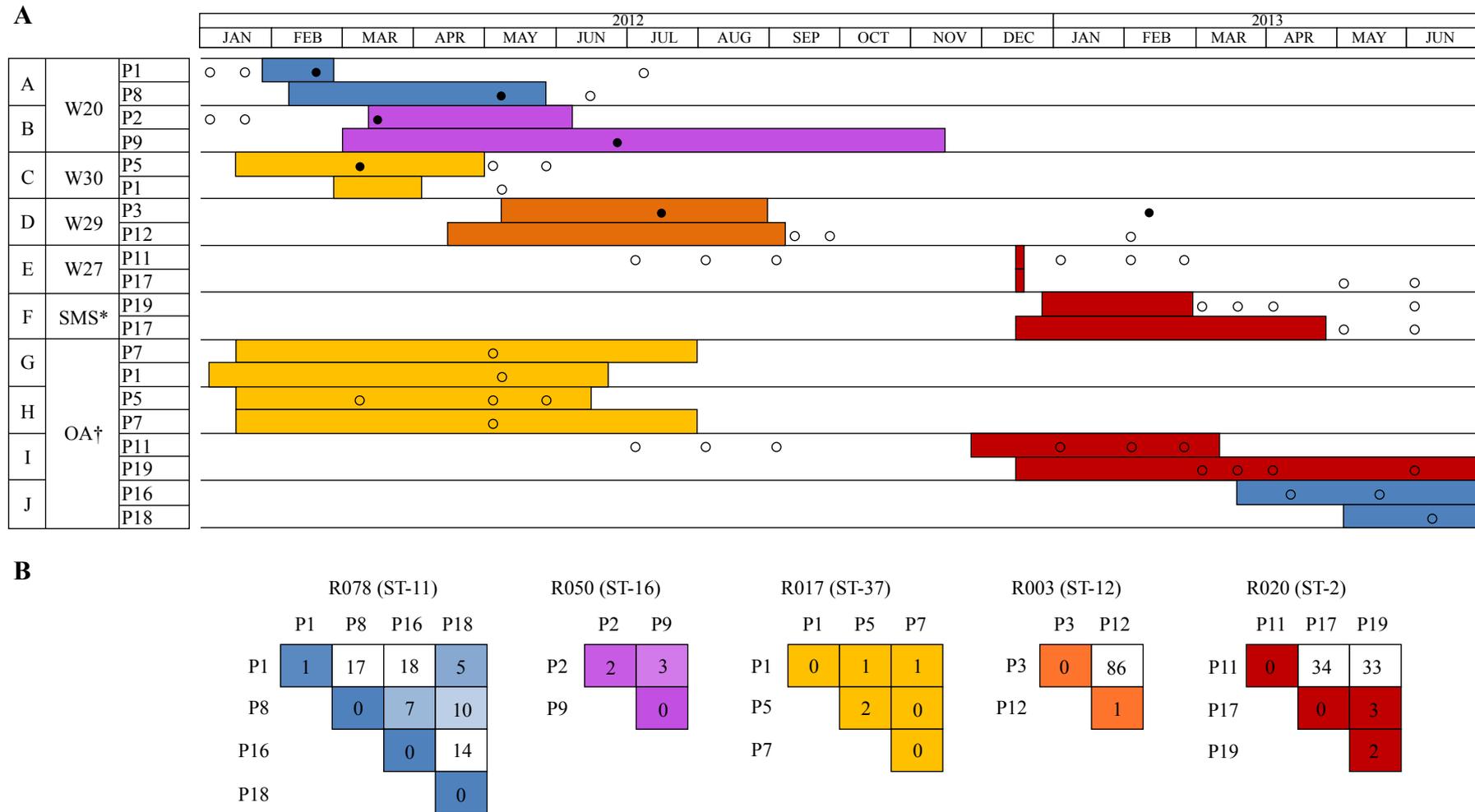
<sup>b</sup> Sequence identity at relative SNV locus in *C. difficile* clinical isolates.

<sup>c</sup> Functional status of SNV at protein level (synonymous or non-synonymous).

<sup>d</sup> Relative locus tag in *C. difficile* 630 reference genome in which within-host SNVs were observed.

<sup>e</sup> 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.

<sup>f</sup> 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  $\leq 2$  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*.<sup>14,18</sup> 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.<sup>5</sup> 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 ( $\leq 2$  SNVs between isolates) over exceptional timescales.<sup>14</sup> However, such lengthy intervals between index and relapse could also be interpreted as reinfections by genetically indistinguishable strains via common environmental contamination sources.<sup>2</sup> 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).<sup>31</sup> Mutational alteration of *spo0A* has been observed previously in a *C. difficile* strain from a fidaxomicin-treated patient with CDI relapse.<sup>18</sup> 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.<sup>32,33</sup> 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.<sup>34,35</sup>

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  $\leq 2$ ) 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.<sup>36</sup> 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.<sup>15</sup> 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

The authors would like to thank the staff at St James's Hospital Microbiology Department for their technical support and donation of isolates.

## Conflict of interest statement

T.R.R. is an advisory board member for Astellas Pharma Ireland.

### Funding sources

This work was supported by the Department of Clinical Microbiology, Trinity College Dublin, as well as a Small Research Grant from the Healthcare Infection Society and an unrestricted grant from Astellas Pharma Ireland.

### 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

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- He M, Sebahia 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.
- Didelot X, Eyre DW, Cule M, et al. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biol* 2012;13:R118.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25:1754–1760.
- Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010;11:595.
- Dingle KE, Griffiths D, Didelot X, et al. Clinical *Clostridium difficile*: clonality and pathogenicity locus diversity. *PLoS One* 2011;6:e19993.
- Economic and Social Research Institute. *Activity in acute public hospitals in Ireland, annual report 2011*. Report No. 45. Dublin: ESRI; 2011.
- 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.
- 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.
- Bauer MP, Notermans DW, van Benthem BH, et al. *Clostridium difficile* infection in Europe: a hospital-based survey. *Lancet* 2011;377(9759):63–73.
- 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.
- 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.
- 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.
- Sokurenko EV, Hasty DL, Dykhuizen DE. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol* 1999;7:191–195.
- 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.
- 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.
- Smith CA. Structure, function and dynamics in the mur family of bacterial cell wall ligases. *J Mol Biol* 2006;362:640–655.
- 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.