Chapter 4

Hospital transmission and persistence of *C. difficile* from a whole genome sequencing perspective

4.1 Introduction

The majority of *C. difficile*-associated disease cases are diagnosed within hospitals and healthcare facilities, although community-acquired *C. difficile* disease exist. Indeed, almost all *C. difficile*-associated outbreaks have occurred in hospitals and healthcare facilities. *C. difficile* spores are resistant to heat and commonly used disinfectants, including 70% ethanol (Lawley *et al.*, 2009) and the resilient nature of *C. difficile* spores contributes to their high transmissibility. Spores can potentially persist in the environment for months under traditional routine cleaning regimes, while maintaining their transmissible nature (Gerding *et al.*, 2008). Additionally, *C. difficile* can persist in the spore form in gnotobiotic mice (Onderdonk *et al.*, 1980). It is unclear, however, what fraction of CDI cases arise due to infections mediated by spores directly originating from the environment.

One major difficulty in treating *C. difficile*-associated diseases is the recurrence of infection. Recurrent CDI is fairly typical, it can be found in 5%-35% of patients (Bakken, 2009; Johnson, 2009) and may occur months or years after the initial infection was resolved (Johnson, 2009). The symptoms

of recurrent cases are frequently indistinguishable from the previous infection scenarios (Bartlett, 2010). One of the factors behind recurrent CDI has been postulated to be the persistence of spores, either endogenously within the host or in the environment. Theoretically, re-infection can be caused by the same or a different strain independent of the nature of the source (Johnson, 2009; Johnson *et al.*, 1989). Early studies have used REA to type isolates from the same patient at multiple time points in order to discriminate between relapse and re-infection (Johnson *et al.*, 1989; ONeill *et al.*, 1991). These results generally indicate that approximately 50% of re-occurring infections are caused by a new *C. difficile* strain, supporting re-infection rather than relapse (Johnson *et al.*, 1989; ONeill *et al.*, 1991). A recent review summarized five published studies and concluded that 33%-75% of the recurrent cases are due to re-infection with a new strain. However, in these cases the question of whether re-infection was caused by strains from endogenous sources or the environment remained unresolved (ONeill *et al.*, 1991). Additionally, it is unclear whether patients can be colonized simultaneously with multiple *C. difficile* strains. O'Neill *et al.* investigated the possibility of multiple carriage using REA and claimed that such incidents are rare (ONeill *et al.*, 1991).

The accuracy of these studies depends on suitable typing methods. An important question is, how should we define strain types? REA may have sufficient discriminatory power to identify isolates belonging to different ribotypes, but to differentiate isolates within the same ribotype, particularly ribotype 027, which shares a highly similar genomic backbone, a more discriminatory method is needed. The power of whole genome sequencing in discriminating between ribotype 027 isolates was demonstrated by the data outlined in the previous chapter. Here these techniques were used to analyze 027 isolates from the same hospital and its associated areas, including isolates from the same patient. In addition to differentiating relapse from re-infection, a goal of this study was to gain some level of understanding of local transmission of *C. difficile* through an analysis that combined phylogenetic information and spatial/temporal data of isolates. Comparative analysis of multiple samples from the same patient can also potentially provide insights

into genetic changes occurring in *C. difficile* over time while being carried by individual patients.

As a comparison, a murine infection model (Lawley *et al.*, 2009) was utilized to monitor the genome change of ribotype 027 *C. difficile* during long-term colonization in mice. In an experiment conducted by Lawley *et al.*, five mice exhibit different outcomes in colonization levels and disease following infection with the same *C. difficile* ribotype 027 strain and subjected to the same antibiotic treatment. The hypothesis is that the differences detected in the genome over time could be attributed to factors such as the host immune system or competition within the intestinal microbiota of individual mice. Selective pressures could be reflected in changes in the *C. difficile* genome before and after colonization.

The aims of the analysis in this chapter were: -

- to differentiate between relapse and re-infection cases in diseased patients;
- to investigate the possibility of carriage of different strains by a single patient;
- to explore the use of whole genome sequencing in understanding local transmission of ribotype 027;
- to assess the level of genome changes in ribotype 027 in mice and humans, and the possible genetic and phenotypic consequences of these changes

4.2 Materials and methods

4.2.1 Bacterial isolates

4.2.1.1 Ribotype 027 isolates from patients

The *C. difficile* hospital collection included 127 *C. difficile* isolates sampled between July 2008 and May 2010 and two retrospective isolates from May

and July of 2007. These isolates were sampled from patients at the Royal Liverpool University Hospital (117 isolates) and neighbouring hospitals, including some visiting general practitioners (12 isolates). These isolates were selected and cultured by Paul Roberts and Fabio Miyajima and colleagues at Liverpool University Hospital, in the following steps: Faecal specimens were collected from patients who are suspected of having CDI based on clinical symptoms. The stool specimens were then tested for *C. difficile* toxins A and B using ELISA. *C. difficile* was then cultured from toxin-positive stool samples on Brazier's plates. Five to ten colonies were then inoculated onto fastidious anaerobe agar plates to check for purity. At least one purified colony from a single patient was collected and used to determine PCR ribotype. Only confirmed ribotype 027 isolates were included in this study.

For fourteen patients (named patient A to N), two or three isolates were collected from different infection episodes at an interval of one to eight months. Additionally, five more colonies were sampled from the primary culture at the first infection episodes of patients C, F, I, and the second infection episode of patient H. The full details of these isolates, including sampling dates and hospital locations where the patients are residing, are given in Appendix B.

4.2.1.2 *C. difficile* BI-7 (ribotype 027) isolates obtained over time during a mouse colonization experiment

Colonization with ribotype 027 *C. difficile* strain BI-7 (clindamycin resistant human isolate) was established by infecting C3H/HeN mice via oral gavage containing 10^7 CFU of culture grown organisms. The mice were subsequently treated with clindamycin for seven days. The resulting colonization condition was monitored by culturing *C. difficile* directly from the faeces (Figure 4.1). In this model, mice 1 and 3 became low-level carriers of *C. difficile* ($<10^2$ CFU/gram faeces); mice 2 and 5 exhibit moderate-level carriage (10^4-10^6 CFU/gram faeces), while mouse 4 remained a high-level excretor of *C. difficile* ($>10^8$ CFU/gram faeces) for extended periods. The experiment was carried

out over a 90 day period post-infection. Mouse 4 displayed chronic intestinal inflammation, while the other four mice did not exhibit any significant intestinal pathologies.

The samples of *C. difficile* selected for whole genome sequencing were: twelve colonies from the day 0 input inoculum; twelve colonies from each of mouse 3, 4, and 5 immediately after clindamycin treatment, and similar samples collected at 90 days post-infection. This added up to a total of 84 samples. All the *in vivo* work described in 4.2.1.2 was carried out by Drs. Trevor Lawley and Simon Clare at WTSI.

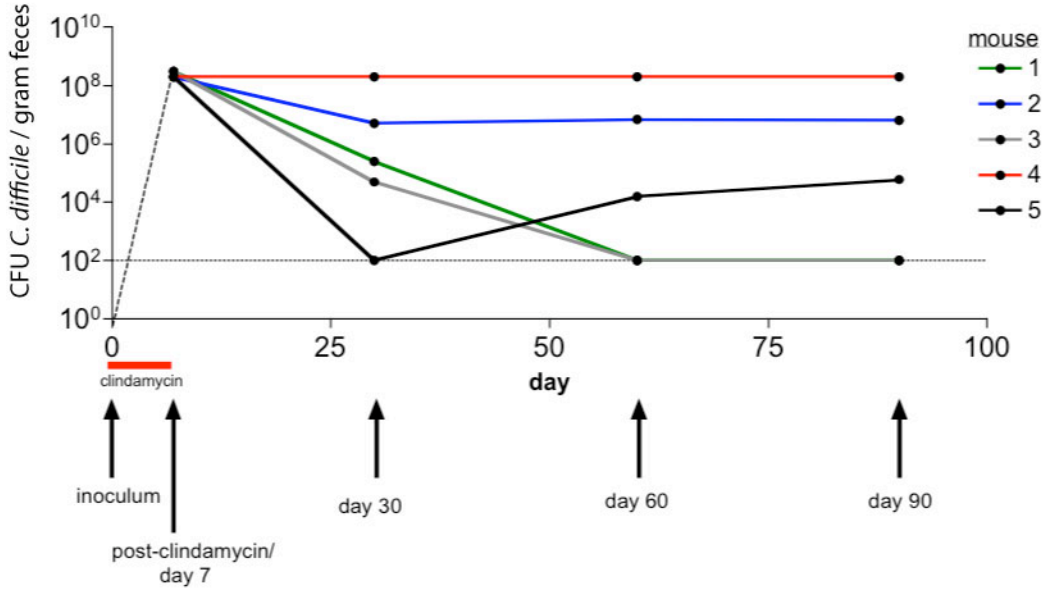

Figure 4.1: The pattern of long-term infection of C3H/HeN mice with *C. difficile* strain BI-7 (ribotype 027). Reproduced from Dr. Trevor Lawley from WTSI.

4.2.2 DNA preparation, sequencing, reads mapping and SNP detection

These steps were carried out as described in section 3.2.2. DNA preparation and sequencing were performed by Louise Ellison, Derek Pickard and the Sequencing Team at WTSI respectively.

4.2.3 Phylogenetic analysis

Phylogenetic relationships were inferred with the program PHYML (Guindon and Gascuel, 2003) with 100 bootstraps. A simple GTR model was used; the model assumes all sites evolve at the same rate, with no invariable sites.

4.3 Results

4.3.1 Genetic diversity and microevolution of hospital ribotype 027

A total of 127 ribotype 027 isolates were sampled from the Royal Liverpool University Hospital and adjacent hospitals between July 2008 and May 2010. Almost all ribotype 027 isolates underlying confirmed CDI cases during this period were included. Two isolates from 2007 were also added to the collection, making the total number 129. All isolates were sampled from patients with confirmed CDI. Ribotype 027 was the most common circulating ribotype in the sampled hospitals within this period, though its relative prevalence has dropped from 50% in 2008 to 31% in 2010 (Table 4.1). The total number of CDI cases increased from 2008 to 2009 but decreased in the later half of 2009.

	027	**106**	**001**	**002**	**014/ 020**	**015**	**078**	**005**	**023**	**All others**	**Total**
Jun/08 - Nov/08	17	8	0	0	0	0	0	1	1	7	34
Dec/08 - May/09	53	16	7	7	0	0	2	3	0	11	99
Jun/09 - Nov/09	32	14	0	2	3	7	2	3	1	8	72
Dec/09 - May/10	10	5	2	2	2	1	6	0	0	4	32

Table 4.1: Number of *C. difficile* isolates per semester by ribotypes. Data provided by Fabio Miyajima at Royal Liverpool University Hospital.

A total of 70 SNPs from the non-repetitive core genome were discovered among these isolates and an un-rooted maximum likelihood phylogeny was constructed (Figure 4.2). The phylogeny suggests more than 20 distinct genotypes were circulating within the hospitals. Two isolates (Liv071 and Liv190) appear to be outliers in this phylogeny and are differentiated from the other isolates by 2 SNPs (1163835 and 1374216, positions refer to that in R20291 genome). All other isolates can be divided into three groups (A, B, C) based on their genotypes. Based on the phylogeny, group C was derived from group B. There are very few SNPs that differentiate between groups. One SNP (4150703) differentiates between groups B, C and the rest; and one SNP (2979027) differentiates between groups B and C. Despite the paucity of genetic variation that defines each group, the groupings of isolates can be considered reliable, as the divisions are based on SNPs whose alleles have been determined in all the clustered strains, except in the case of SNP 2979027, where the alleles in two isolates (Liv8 and Liv9) can not be determined. In addition, group A, three branches in group B, and four branches in group C all received >80% bootstrap support. These sub-groupings were given the names B10 – B30 and C10 – C40 (Figure 4.2).

There is very little variation within a single genotype group. The numbers of SNPs within each group are 2, 30, and 22 for A, B, and C respectively. Comparing this dataset within the global ribotype 027 dataset in Chapter 3 reveals that group C contains local samples exclusively (only samples from Royal Liverpool University Hospital and neighbouring hospitals), while group B isolates are found in the same genotype cluster as human isolates from Glasgow and Belfast, and group A isolates are found in the same genotype cluster as human isolates from Birmingham. Thus, it is possible that groups A and B reached this region independently, and the emergence of group C occurred locally in Liverpool. Apparently, all three genotypes have persisted and continued to evolve after being introduced into this area.

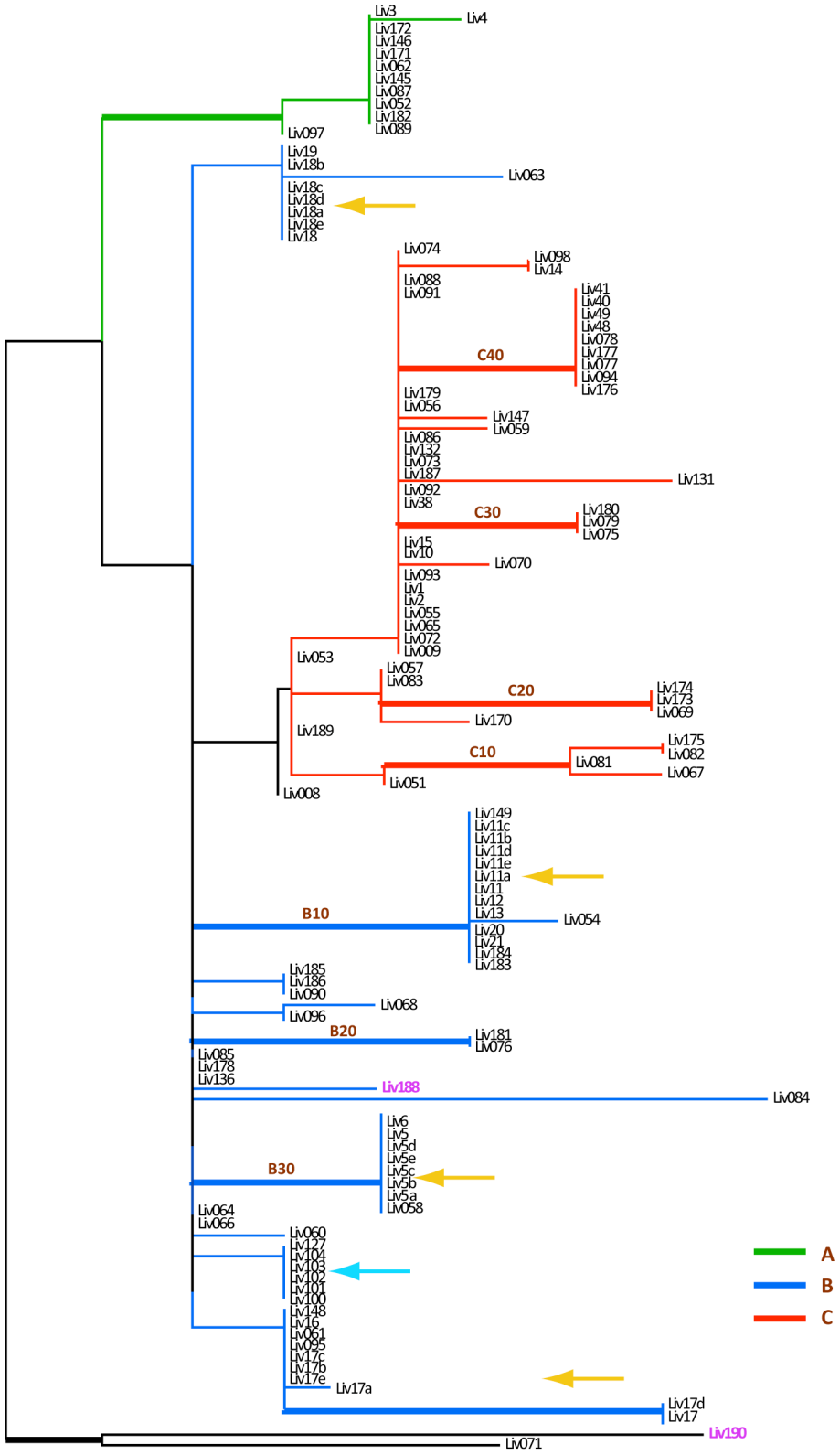

Figure 4.2: Un-rooted maximum likelihood phylogeny of Liverpool isolates. Branches are coloured according to isolate groupings. The groupings supported by > 80% bootstraps are highlighted by thick branches, with associated names assigned for each genotype subgroup labelled in brown. Yellow arrows point to multiple samples taken from the same patient at the same time point. Blue arrow denotes isolates that

exhibit spatial temporal clustering (discussed in 4.3.3). Two isolates from 2007 are shown in pink. Liv008 is labelled in black as its grouping (either B or C) cannot be determined due to missing allele information.

4.3.2 Colonization by *C. difficile* 027 in patients

This study includes samples taken from individual patients sampled at multiple time points. Here the aim was to gain insight into whether recurrent disease is caused by relapse of the same or re-infection with a genotypically distinct *C. difficile* 027 strain. It is worth noting that this collection only includes isolates confirmed to be ribotype 027. During the sampling period, twenty-six patients exhibited at least two separate infection episodes. Among these, two patients were infected with bacteria of a different ribotype at the recurrent infection point (data provided by Fabio Miyajima at Royal Liverpool University Hospital), supporting re-infection with a distinct strain.

Fourteen patients appear to be infected with ribotype 027 *C. difficile* at all monitored infection episodes. Figure 4.3 shows the sampling time points and genotype assignments of the associated *C. difficile*. In all 14 cases, isolates of the same genotype group were found during the subsequent infection period(s) within individual patients except for Patient D, whose case is difficult to determine as Liv8 cannot be assigned to a group with confidence. However, isolates from different infection period(s) of the same patient were not always identical. Sample Liv4 is differentiated from Liv3 by 1 SNP (3283560); these two samples are separated by an interval of 6 months (Figure 4.3). Strikingly, Liv17 is differentiated from Liv16 by 4 SNPs (1767576, 2384792, 2550185 and 3632130). As the two samples were separated by only 5 weeks in isolation date, it could be considered unlikely the 4 SNPs arose during this interval within the same *C. difficile* lineage as it replicated. In addition to Liv8 mentioned above, isolates Liv9, Liv14 and Liv15 also contain missing allele information. This makes it difficult to determine the genetic difference in each sample pair for Patient E (Liv9 and 10) and Patient G (Liv14 and 15).

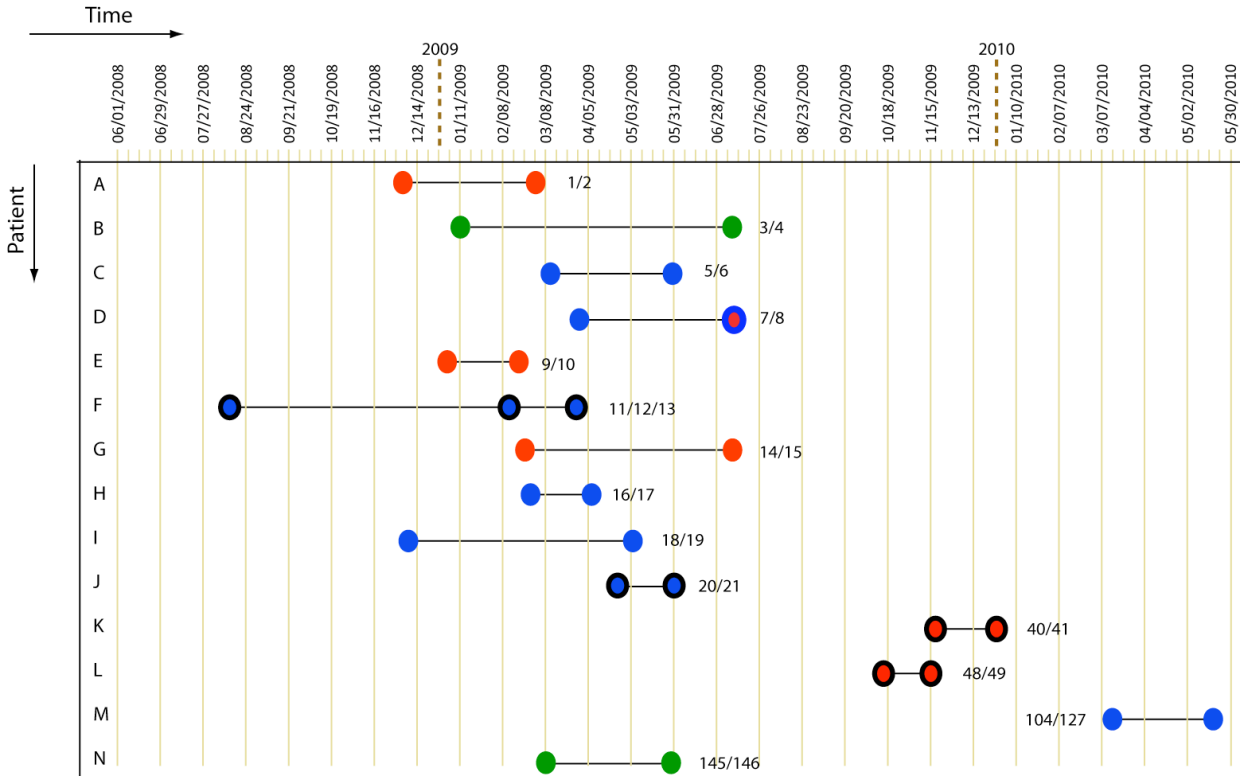

Figure 4.3: Temporal graph of *C. difficile* ribotype 027 genotype assignment from the same patient at multiple infection episodes. Each row depicts two (or three) isolates sampled from the same patient at different time points (dates labelled on top). The isolates are represented by circles, each is coloured according to the genotype group it is associated with, as in Figure 4.2. Circles with black edges represent genotype B10 (blue) and C40 (red) respectively. Liv8 is shown in red and blue as both are possible due to missing allele information. The numbers labelled next to the isolates are sample identifiers.

To investigate the possibility of multiple carriage, five more samples were collected at the first infection episodes of patients C, F, I, and at the second infection episode of patient H; these samples were named Liv5a-e, Liv11a-e, Liv17a-e and Liv18a-e (Figure 4.2), according to the original sample names Liv5, Liv11, Liv17 and Liv18, respectively.

The analysis shows that Liv5a-e, Liv11a-e, and Liv18a-e exhibit an identical genotype to their corresponding original samples. While Liv17a-c and Liv17e are identical, they are differentiated from Liv17 and Liv17d by 4 SNPs. This potentially suggests carriage of multiple lineages, or multiple variants of the same lineage by Patient H. However, a more comprehensive analysis involving a larger sample set of *C. difficile* and patients would be required to firmly resolve the relationship between isolates differentiated by a few SNPs in terms of cross infection between patients compared to evolution within the same patient.

4.3.3 Spatial and temporal distribution of ribotype 027 genotypes

As a next step, the genotypes of Liverpool *C. difficile* isolates were mapped to the date and hospital location where the patient was residing when each sample was isolated. Figure 4.4 illustrates the spatial and temporal information with date of isolation indicated on the x-axis and hospital floors or neighbouring hospitals indicated on the y-axis. Coloured circles (matching Figures 4.2 and 4.3) depict sampled *C. difficile* 027 isolates from infected patients and the associated genotypes. The two isolates from 2007 (Liv188 and Liv190) were not included in Figure 4.4.

The spatial temporal graph reveals a general lack of clustering between genotype and hospital location, implying frequent transmission within the hospital and between neighbouring hospitals. However, persistence of the same genotype group can be seen within certain time periods and spatial

locations. For example, during the entire sampling period, Floors 7 (Diabetes/Endocrinology/Cardiology/Haematology) and 9 (Renal transplant/Nephrology/Rheumatology/Urology/Breast) were found to harbour group B genotypes almost exclusively (except Liv182, which belongs to group A). In addition, only group A isolates were found in wards 5XY (Gastroenterology) between November 2008 and February 2009. Group C genotypes appear to dominate between October and December of 2009 (pink shaded area in Figure 4.4). A tight cluster of C40 genotype was found in wards 2XY throughout July 2009 (pointed by an arrow in Figure 4.4). Interestingly, Liv100-102, Liv104, and Liv127, which share an identical genotype and occupy a single lineage in the phylogeny (blue arrow in Figure 4.2), form a cluster between March and May 2010 (yellow shaded area in Figure 4.4). Another observation is that isolates that were derived later in terms of their phylogeny (on branch tips) were not necessarily found at a later time in the sampling period. Isolate Liv188 from 2007 (shown in pink in Figure 4.2) is found at a branch tip in Group B. These findings seem to imply that the emergence of genotypes discovered in the collection predates the sampling period. Although genotypes B and C are more common than A in our collection, it is unclear whether this phenomenon is random or a result of difference in fitness.

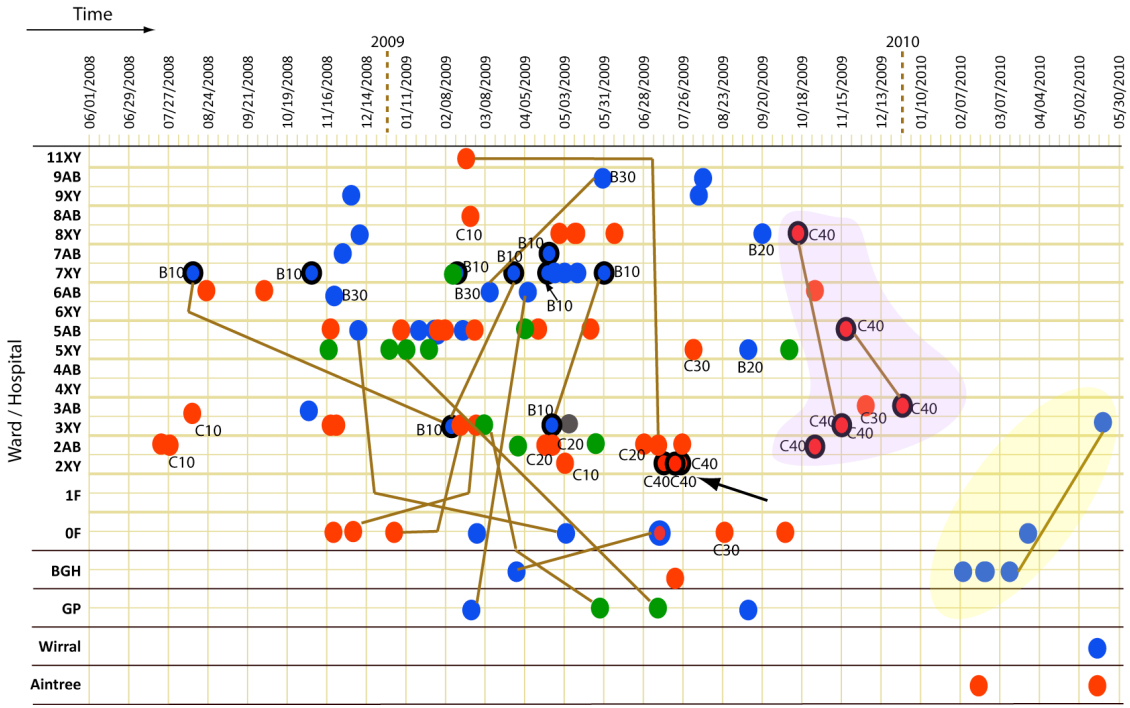

Figure 4.4: Temporal and spatial graph of isolates from local hospitals in Liverpool region. Rows depict hospitals (BGH, Wirral, Aintree), visiting general practitioners (GP) or wards from a single large hospital (0F – 11XY). Time is depicted on the x-axis, with dates labelled on top. Every four weeks is represented by a major division, shown with yellow lines. Minor divisions on the x-axis represent 1 week. The isolates are represented by circles, each is coloured according to the genotype group it is associated with, as in Figure 4.2. The uncoloured circle represents Liv071, which belongs to none of groups A-C. The numbers labelled next to the isolates match genotype names from Figure 4.2. Circles with black edges represent genotype B10 (blue) and C40 (red) respectively. The isolate shown in red and blue is Liv8, which cannot be assigned with confidence to either B or C group due to missing allele information. Isolates from the same patient are connected by brown lines. Shaded areas and the arrow highlight three genotype clusters.

4.3.4 Genetic diversity of ribotype 027 during colonization of mice

An experiment to study aspects of the persistence of ribotype 027 colonization in mice was carried out by Drs. Trevor Lawley and Simon Clare at WTSI. These results showed that five individual mice exhibited different outcomes in terms of their respective colonization levels and disease symptoms following infection with the same dose of *C. difficile* ribotype 027 strain (BI-7) and the same antibiotic treatment (Figure 4.1). This experiment provided an opportunity to sample the genomes of individual *C. difficile* lineages as they persisted *in vivo* in the mouse over time. How quickly does the genome change, if at all? These data could also be used to compare with the hospital patient dataset. Consequently, *C. difficile* samples were collected at three time points: the input inoculum (Day 0), immediately after clindamycin treatment (Day 7), and at 90 days post-infection (Day 90). For Day 7 and Day 90, samples were collected from each of mouse 3, 4, and 5 (more details in 4.2.1.2). Twelve colonies were collected for each time point for each mouse, resulting in 84 isolates in total. The samples were named Day0-1 to 12; Day7-M3/4/5-1 to 12, and Day90-M3/4/5-1 to 12.

Sample Day90-M5-7 yielded little data and was excluded from the analysis. Five samples (Day7-M3-1, Day7-M3-2, Day7-M5-9, Day7-M5-12, and Day7-M4-1) were found to be non-027 *C. difficile*, as shown by >22,000 SNPs when compared to the genome of R20291. Day7-M3-1, Day7-M3-2, Day7-M5-9, Day7-M5-12 were discovered to be highly similar to ribotype 002 isolates, and Day7-M4-1 to M120 (ribotype 078). These were considered to be contamination, most likely occurring within the anaerobic cabinet at the time (subsequently containment protocols have been tightened considerably to control this highly transmissible organism). In the remaining 78 isolates, only 1 intergenic SNP (a G->C change at position 3876023) was found in Day7-M5-5, while all the other samples were identical. The fact that this SNP is present

in one sample at Day 7 but absent from all samples at Day 90 suggests it was not fixed in the population.

4.4 Discussion

4.4.1 Strengths and limitations of spatial temporal genotype analysis

In this chapter, genetic diversity was analyzed for *C. difficile* ribotype 027 isolates from infected patients within a single hospital and its local capture areas. This analysis shows that the *C. difficile* 027 circulating in the sampling area have highly similar but in some cases distinguishable genotypes. Phylogenetic analysis can be used to divide these genotypes into three main groups, which differ in terms of their dominance over time and space. Interestingly, although not dominant, genotype clustering is more apparent during the latter part of the sampling period (October 2009 to May 2010), while a large number of genotypes co-exist earlier, particularly in the first half of 2009. This could be related to the fact that there were fewer reported CDI cases in late 2009 and 2010 (Table 4.1), which is concurrent with fewer local transmission events, therefore the same genotype is more likely to persist longer.

The co-existence of multiple genotypes and the overall lack of general clustering implies frequent transmission, underpinning the importance of applying frequent deep and effective disinfection regimes. However, this analysis did not identify obvious significant transmission events within the hospital. One reason could be the absence of carriers in the sampling. This collection only includes isolates from patients with clinical disease. Further, environmental samples were not included. Asymptomatic carriers of *C. difficile* can also excrete spores, which become potential sources of infection (Peach *et al.*, 1986). The percentage of asymptomatic carriers of *C. difficile* was found in an independent study to be 32% in patients with cystic fibrosis (Peach *et*

al., 1986). It is possible that data from such carriers will help construct transmission chains.

The hospital isolates analyzed in this chapter are highly similar, with many sharing identical genotypes, making them difficult to differentiate even by whole genome sequencing. Although the genotype groups A – C were supported by almost all alleles in the relevant strains, verifying the alleles using an independent genotyping platform would be useful.

4.4.2 Relapse, re-infection and multiple strain carriage

The analysis presented here suggests that the majority of recurrent CDI cases in this dataset are due to relapse rather than re-infection. However, this conclusion should at this stage be regarded as not being conclusive. Nine out of fourteen patients studied demonstrate *C. difficile* isolates of identical genotype for different infection episodes, which can be interpreted as evidence for relapse. However, it is also arguable that these patients could have acquired isolates of the same genotype from exogenous sources after the initial infections, since these genotypes are fairly prevalent during the study period. Patient H shows evidence for harbouring multiple strains, as the two genotypes differ by 4 SNPs, which is considerable divergence within this dataset. However, even here we cannot rule out the unusual accumulation of these SNPs by the same lineage within this patient. It is also intriguing that the genotype of Liv17/Liv17d (Figure 4.2) was unique to this patient. It is possible that the 4 SNPs occurred long before this study period (within Patient H or not), and both genotypes have been carried by Patient H since. The missing data in Liv8, Liv9, Liv14 and Liv15 makes it difficult to judge whether relapse or re-infection is true for Patient D, E and G, while the single SNP difference between Liv3 and Liv4 (isolates from two infection episodes of Patient B, collected 6 months apart) could be interpreted as either bacterial evolution within the host or re-infection by a new strain. The finding that the

Liv4 genotype is unique in our dataset implies that the former interpretation is more likely. All this analysis emphasises the importance of collecting larger comparative datasets.

The lack of obvious evidence for re-infection by a new strain in our dataset is in contrast with the early reports (Johnson *et al.*, 1989; ONeill *et al.*, 1991). However, as this study only focused on ribotype 027 isolates, the proportion of this type of re-infection might be expected to be lower. Multiple strain carriage was found in 1 out of 4 cases tested in this study, more common than suggested by O'Neill *et al.* (ONeill *et al.*, 1991). However, a recent report focused on ribotype 027 found more than one MLVA profile in 5 out of 39 faecal specimens, and suggested multiple strain carriage and rapid evolution as possible reasons (Tanner *et al.*, 2010). The two patients who exhibit *C. difficile* of different ribotype at two infection points strongly support exogenous re-infection, if not carriage of multiple strains. Sequencing multiple samples from patients will help us to understand this more clearly. It would also have been potentially more informative if the original sampling study was not limited to ribotype 027.

4.4.3 Insights from *C. difficile* colonization in mice

During a 90-day colonization period in mice, the *C. difficile* 027 strain BI-7 did not exhibit any detectable change in the core genome. One mutation arose but did not become a fixed variant. It is worth noting that the 3 mice showed different colonization levels of *C. difficile* and disease pathology, even though no genetic difference was discovered between the isolates they carried. This implies the colonization outcome can be more attributed to host factors, such as the immune system or the intestinal microbiota composition of the individual mouse, rather than the bacterial genotype.

Finally, the mutation rate of *C. difficile* ribotype 027 was estimated to be 1-2 mutations per genome per year in Chapter 3 and the current data are generally in agreement with this estimate.