Genetic relatedness among clinical *Clostridium difficile* strains in one Scottish region

Nicosha De Souza[1], Gill Douce[2], Cosmika Goswami[2], Umer Ijaz[2], Derek Brown[3], John Coia[3], Peter Donnan[1], Camilla Wiuff[4], Peter Davey[1], Marion Bennie[4], Jan Lindstrom[2], Charis Marwick[1]

University of Dundee[1], University of Glasgow[2], NHS Greater Glasgow and Clyde[3], NHS National Services Scotland[4]

INTRODUCTION & PURPOSE

Clostridium difficile infection (CDI) is primarily a healthcare associated infection particularly affecting mortality and morbidity in the elderly. Investigation of the molecular epidemiology of CDI in Scotland is a interdisciplinary collaboration between various institutions involving anonymising patient level data, sequencing the bacterial genome isolated from the patient and linking the two. **This study analysed linked data with the aim of investigating the genetic relatedness between the bacterial strains and to decipher any plausible transmission of infection between patients within one geographical region in Scotland.**

METHODS

PATIENT DATA AND GENOME SEQUENCE

Isolates identified from the Scottish National Microbiology Reference Laboratory → Isolate genotyped by University of Glasgow & Patient data obtained from Electronic Data Research and Innovation Service (eDRIS), NHS National Services Scotland → Data anonymised and linked by health informatics (HIC), University of Dundee

DATA ANALYSIS

Genomes between sample pairs are compared and ≤ 10 single nucleotide variant (SNVs) difference is considered related → Comparisons classed as 0-2, 0-6 and 0-10 based on previous publications (Eyre *et.al.*, 2013) and plotted. → Pairs inspected to detect epidemiological links using hospital information → Potential transmissions identified and classified as direct hospital contact, hospital contamination or community contact

Epidemiological definitions from Eyre *et.al.*, 2013

- Infectious period – One week before diagnosis until eight weeks after.
- Incubation period – Twelve weeks before diagnosis up to diagnosis.
- Direct hospital contact – Same hospital at same time with infectious and incubation periods overlapping.
- Hospital contamination – Same hospital with time from discharge (or end of infectious period) of source patient to start of incubation period of recipient patient ≤28 days, but no overlap in infection and incubation periods.
- Community contact – Same general practice and postal district.

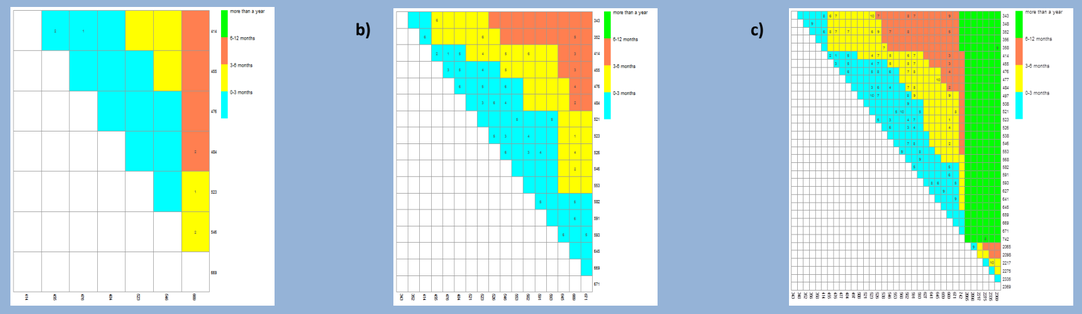
- All data analyses was conducted in the HIC safe haven.

Eyre et. al. (2013),Diverse Sources of *C. difficile* Infection Identified on Whole-Genome Sequencing. N Engl J Med 2013; 369:1195-1205

RESULTS

Figure 1: Grid plots for *C. difficile* genome comparisons differing by a) 0-2, b) 0-6, c) 0-10 SNVs in one postcode region of Scotland

a) b) c)



A total 35 (29 %) of 120 episodes in this region had ≤ 10 SNVs different from at least one other episode (total 87 pair matches). The figures are a distribution of the samples that were related to each other by 0-2, 0-6 and 0-10 SNVs difference. Each episode is represented on both axes as the number of days from a fixed time point to CDI. Each grid in the plot is a comparison between samples and the number printed in the grid is the number of SNVs that are different between the two episodes intersecting at that point, if it falls within the specified range. The colour of the grid represents time separating the two samples.

Table 1: Epidemiological links for various combination of SNV groups (0-2, 0-6, 0-10)

SNVs between pairs	0-2	0-6	0-10
Total pairs involved	5	40	87
Unique episodes involved	7	17	35
	No. Of Pairs (Unique Episodes)	No. Of Pairs (Unique Episodes)	No. Of pairs (Unique Episodes)
Direct hospital contact	1 (2)	7 (9)	13 (16)
Hospital contamination	1 (2)	3 (6)	5 (8)
Same general practice (GP)	0	0	3 (6)
Same postcode district (*e.g.*- DD1)	0	10 (9)	16 (9)

*** the epidemiological links are mutually excusive with the preceding entry given more priority compared to the proceeding ones. *i.e.* if the episode pair has both a potential direct hospital contact and hospital contamination, the potential transmission by direct hospital contact will be considered and transmission by hospital contamination will be discarded.**

CONCLUSION

Whole genome sequencing permits the identification of closely related sequences that can shed light on the transmission and sources of infection. Identifying linked cases and transmission routes using healthcare data can inform interventions to prevent spread. This collaborative group plans to prospectively sample a broader Scottish population and generate more conclusive results.