Knetsch, CW; Kumar, N; Forster, SC; Connor, TR; Browne, HP; Harmanus, C; Sanders, IM; Harris, SR; Turner, L; Morris, T; +13 more... Perry, M; Miyajima, F; Roberts, P; Pirmohamed, M; Songer, JG; Weese, JS; Indra, A; Corver, J; Rupnik, M; Wren, BW; Riley, TV; Kuijper, EJ; Lawley, TD; (2017) Zoonotic transfer of Clostridium difficile harboring antimicrobial resistance between farm animals and humans. Journal of clinical microbiology. ISSN 0095-1137 DOI: https://doi.org/10.1128/JCM.01384-17

Downloaded from: http://researchonline.lshtm.ac.uk/4645908/

DOI: https://doi.org/10.1128/JCM.01384-17

Usage Guidelines:

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
Zoonotic transfer of *Clostridium difficile* harboring antimicrobial resistance between farm animals and humans

C.W. Knetsch\(^1,\)†, N. Kumar\(^2,\)†, S.C. Forster\(^3,4\), T.R. Connor\(^5\), H.P. Browne\(^2\), C. Harmanus\(^1\), I.M. Sanders\(^1\), S.R. Harris\(^6\), L. Turner\(^7\), T. Morris\(^7\), M. Perry\(^7\), F. Miyajima\(^8\), P. Roberts\(^8\), M. Pirmohamed\(^8\), J.G. Songer\(^9\), J.S. Weese\(^10\), A. Indra\(^11\), J. Corver\(^1\), M. Rupnik\(^12,13\), B.W. Wren\(^14\), T.V. Riley\(^15,16\), E.J. Kuijper\(^1\), and T.D. Lawley\(^2\)*

\(^*\)Correspondence to: Trevor Lawley, tl2@sanger.ac.uk

†These authors contributed equally to this work

Running title: Clonal *C. difficile* infect different hosts globally.

Keywords: *Clostridium difficile*, RT078, intercontinental transmission, inter-host transmission, accessory genome, One Health concept.
Abstract. The emergence of Clostridium difficile as a significant human diarrheal pathogen is associated with the production of highly transmissible spores and the acquisition of antimicrobial resistance genes (ARGs) and virulence factors. Unlike the hospital associated C. difficile RT027 lineage, the community associated C. difficile RT078 lineage is isolated from both humans and farm animals; however, the geographical population structure and transmission networks remain unknown. Here we applied whole genome phylogenetic analysis of 248 C. difficile RT078 strains from 22 countries. Our results demonstrate limited geographical clustering for C. difficile RT078 and extensive co-clustering of human and animal strains, thereby revealing a highly-linked, inter-continental transmission network between humans and animals. Comparative whole-genome analysis reveals indistinguishable accessory genomes between human and animal strains, and a variety of antimicrobial resistance genes in the pangenome of C. difficile RT078. Thus, bi-directional spread of C. difficile RT078 between farm animals and humans may represent an unappreciated route disseminating antimicrobial resistance genes between humans and animals. These results highlight the importance of the “One Health” concept to monitor infectious disease emergence and the dissemination of antimicrobial resistance genes.
Introduction: Over the past decade, Clostridium difficile has emerged as the primary cause of infectious antibiotic associated diarrhea in hospitalized patients (1). Unlike other common healthcare-associated pathogens, C. difficile produces resistant spores that facilitate host-to-host transmission and enable long term survival and dispersal in the healthcare system and the wider environment (2). The emergence of epidemic C. difficile ribotype (RT) 027 (NAP1 / ST-1), responsible for many large-scale hospital outbreaks worldwide (3, 4), has been linked to environmental spore contamination and the acquisition of fluoroquinolone resistance (5). Enhanced research focus on C. difficile in the aftermath of the C. difficile RT027 outbreaks has revealed other evolutionarily distinct C. difficile lineages, in particular C. difficile RT078 (NAP07-08/ST-11), that are now emerging as significant human pathogens for unknown reasons (6).

The “One Health” concept, connecting the health of humans to the health of animals and their shared environments, represents a relevant framework for understanding the emergence and spread of pathogens. C. difficile RT078 is commonly isolated from both humans and farm animals (7) and increasingly recognized as a causative agent of both healthcare and community-associated C. difficile infection (CDI) (8). This lineage typically affects a younger population (9) and results in higher mortality than C. difficile RT027 (10).

Standard genotyping tools have highlighted genetic similarities between human and animal C. difficile RT078 (11-13) strains raising the possibility of zoonotic transmission (14). Nevertheless, the exact evolutionary and epidemiological relationships between human and animal C. difficile RT078 strains remain unknown due to the lack of discriminatory power of these typing methods and the clonal nature of C. difficile lineages. Recently, using whole genome phylogeny, we reported that asymptomatic farmers and their pigs can be colonized with clonal C. difficile RT078 isolates demonstrating evidence for spread between animals and humans (15).
Results and Discussion: Here we assess the broad genetic diversity of *C. difficile* RT078, by performing whole genome sequence analysis of 247 strains isolated predominantly from humans and animals that were collected from 22 countries across North America, Europe, Australia and Asia between 1996 and 2012 ([https://microreact.org/project/rJs-SYGMe](https://microreact.org/project/rJs-SYGMe)) (Table S1). We explored the phylogenetic structure of *C. difficile* RT078 by generating a core genome maximum likelihood phylogeny that included the 247 *C. difficile* RT078 strains and the reference genome of *C. difficile* M120 (n=248) (Fig. 1). Superimposing the geographic origin of strains revealed considerable co-clustering of European (dark green) and North American (purple) strains across the phylogeny (Fig. 1). Permutation analysis on randomly generated, equalized subsets of European (dark green) and North American (purple) genomes confirmed co-clustering of geographically diverse strains (Fig. S1). In addition, the absence of a single clade of *C. difficile* RT078 isolated in Australia (light green) is also suggestive of sporadic transmission between Europe and Australia (Fig. 1). Overall, the observed lack of geographic clustering is characteristic of repeated, international transmission.

We next examined the phylogenetic distribution of strains isolated from humans (n=184) and animals (n=59) to understand the potential for zoonotic spread. This analysis identified examples of human to human and animal to animal spread and strong evidence of bi-directional spread of *C. difficile* RT078 between animals and humans across the phylogeny. These observations are supported by the extensive co-clustering of human (blue lines) and animal strains (red lines) (Fig. 1). Focused analysis of closely related *C. difficile* RT078 strains identified 6 clusters containing both animal and human isolates with identical core genome and highly similar whole genomes (ANI ≥ 99.73%; Table 1). Surprisingly, Cluster 1 consists of an animal strain from Canada and human strains from UK indicating that zoonotic spread of *C. difficile* is not confined to a local population of humans and...
animals as found previously (15). The existence of highly related human and animal isolates suggests that C. difficile RT078 has frequently spread between animals and humans.

Next, a detailed analysis of the accessory genome, including mobile genetic elements, was performed to further explore the genomic similarities between human and animal strains. Of the 6,239 unique genes present across our genome collection, 3,368 genes (54.0%) were assigned to the core genome leaving 2,871 genes (46.0%) present in the accessory genome (Fig. S2). Considering only the human and animal isolates, 2,859 accessory genes were identified. The vast majority of human and animal specific accessory genes were found at low frequencies in the population (Fig. 2A). We observed no statistically significant difference in the number of strains carrying accessory genes exclusive to either the human or the animal population ($\chi^2$ p-value of 0.39). Considering only those accessory genes present in at least 10% of isolates ($n= 465$), 461 (99.1%) were identified in both human and animal isolates. The absence of accessory genes unique to either group demonstrates that either C. difficile has a stable accessory genome, which is host independent or provides further support for the frequent transmission of C. difficile between host populations.

Given the high percentage of mobile elements including antimicrobial resistance genes harbored by C. difficile genomes (5, 6), we next sought to analyze distribution of different ARGs in the pan-genome of human and animal strains. In total, 22 different putative ARGs are present in the 243 C. difficile RT078 genomes (Fig. 2B). The most common ARG was the chromosome encoded cdeA, a well-known multidrug transporter that was detected in all strains; however, other common genes included those encoding resistance to aminoglycosides, tetracycline and erythromycin (Fig. 2B). Importantly, no specific ARGs were statistically enriched in the animal isolates; however, the ermB (erythromycin resistance methylase) gene was identified in the human isolates (Fisher’s-exact test, q value = 1.25E-07). These results provide further support that a clonal C. difficile RT078 population...
containing a broad array of ARGs is spreading between humans and farm animals except ermB, which has signs of unknown selective pressure in the human isolates.

*C. difficile* is an ancient, genetically diverse species that has only emerged as a significant human pathogen over the past four decades. It remains to be determined why evolutionary distinct lineages such as *C. difficile* RT027 and RT078 (6) are simultaneously emerging to cause disease in the human population. Previously we have demonstrated that *C. difficile* RT027 acquired fluoroquinolone resistance during the 1990s in North America and rapidly spread through the global healthcare system (5). Here we demonstrated that *C. difficile* RT078 has spread multiple times between continents, in particular North America and Europe, highlighting that *C. difficile* emergence and spread is a global issue. In contrast to the distinct animal- and human-associated populations observed for the multidrug-resistant enteric pathogen *Salmonella* Typhimurium DT104 (16), we demonstrated that *C. difficile* RT078 is a clonal population moving frequently between livestock and human hosts with no geographical barriers. Although the original reservoir remains unknown, the reciprocal transmission between humans and farm animals emphasizes the importance of a comprehensive One Health perspective in managing and controlling *C. difficile* RT078.
Materials and Methods

Collection of C. difficile strains

C. difficile laboratories worldwide were asked to send a diverse representation of their C. difficile 078 collections to the Lawley laboratory hosted at the Wellcome Trust Sanger Institute. Sample shipping was coordinated by the Lawley laboratory. After receiving all shipped samples the DNA extraction was performed batch wise by one person using the same protocol and reagents to minimize bias. Phenol-Chloroform was the preferred method for extraction since it provides high DNA yield and intact chromosomal DNA. The genomes of 182 strains designated as C. difficile RT078 (/NAP07-08/ST-11), by PCR ribotyping (17) were sequenced and combined with our previous collection of 65 strains of C. difficile RT078 (12) making a total of 247 strains analyzed in this study. These 247 strains were collected between 1996 and 2012 and are comprised of representative strains from 4 continents (North America, Europe, Australia and Asia). Of these strains, 183 were derived from humans, 59 from animals (pigs, cattle, horses and poultry), 4 foods and 1 environmental sample. Details of all sequenced strains are listed in Table S1, including the European Nucleotide Archive (ENA) sample accession numbers. Metadata of the C. difficile RT078 strains has been made freely publicly available through Microreact (18) (https://microreact.org/project/rJs-SYgMe).

Bacterial culture and genomic DNA preparation

C. difficile strains were cultured on blood agar plates (bioMérieux, the Netherlands) for 48 hours, inoculated into liquid medium (brain–heart infusion (BHI) broth supplemented with yeast extract and cysteine) and grown over night (ca 16 hours) anaerobically at 37 °C. Cells were pelleted, washed with phosphate-buffered saline (PBS), and genomic DNA preparation was performed using a phenol–chloroform extraction as previously described (19).

DNA sequencing, assembly and annotation
Paired-end multiplex libraries were prepared and sequenced using Illumina Hi-Seq platform with fragment size of 200-300bp and a read length of 100bp, as previously described (20, 21). An in-house pipeline developed at the Wellcome Trust Sanger Institute (https://github.com/sanger-pathogens/Bio-AutomatedAnnotation) was used for bacterial assembly and annotation. It consisted of de novo assembly for each sequenced genome using Velvet v. 1.2.10 (22), SSPACE v. 2.0 (23) and GapFiller v 1.1 (24) followed by annotation using Prokka v. 1.5-1 (25).

**Construction and analysis of the Pan genome**

We used the pan genome pipeline Roary (26), to identify the *C. difficile* RT078 pan genome. Roary takes annotated draft assemblies in GFF3 format which were produced by Prokka (25). Predicated coding regions were extracted from the input and converted to protein sequences. Partial sequences (>5% nucleotides unknown or sequence length less than 120 nucleotides) were filtered and the remaining sequences were iteratively clustered with CD-HIT beginning with a sequence identity of 100% and matching length of 100% down to a default sequence identity of 98%. One final clustering step was performed again with CD-hit, with a sequence identity of 100% leaving one representative sequence for each cluster in a protein FASTA file. This was followed by a comprehensive, pairwise comparison with BLASTP on the reduced sequences with a default sequence identity percentage of 95% and matching length of 100%. The pan genome embodies the core genome, defined as those genes present in at least 90% of the genomes, and the accessory genome, defined as those genes present in between 10% and 90% of the genomes. Rare variant genes, found in less than 10% of genomes, were discarded.

Core genes (n=3,368) alignment, an output from Roary, was used to construct phylogenetic structure of 248 *C. difficile* strains. Single nucleotide polymorphisms (SNPs) were extracted from the core gene alignment using SNP-sites (27). Maximum likelihood tree
based on SNPs alignment was constructed using FastTree with –gamma –gtr settings (28) and
tree was visualized with iTOL (29).

226 **Average Nucleotide Identity (ANI) analysis**

Using Roary analysis, *C. difficile* RT078 strains isolated from humans and animals
with identical core genome were extracted using an in-house R script. ANI was calculated by
performing pairwise comparison of genome assemblies of these *C. difficile* RT078 strains
using MUMmer (30).

231 **Identification of antimicrobial resistance gene sequence**

Antimicrobial resistance genes were identified within the *C difficile* RT078 genomes through
comparison to the CARD database with the ARIBA software (https://github.com/sanger-
pathogens/ariab).
Acknowledgments:

We acknowledge the Wellcome Trust [098051]; the United Kingdom Medical Research Council [PF451 to TL, MR/L015080/1 to TC and MR/K000551/1 to FM, MP, BW and TL]; the Australian National Health and Medical Research Council [1091097 to SF] and the Victorian Government’s Operational and Infrastructure Support Program [SF] for financial support. CK has been supported by a ZonMw grant (125020004). We are grateful to A. Neville and the Wellcome Trust Sanger Institute Pathogen Informatics team for their input.

Conflict of interests

The authors declare no competing financial interests.
References:


Figure 1. Phylogeography of human and animal *Clostridium difficile* RT078. Maximum likelihood, midpoint rooted phylogenetic tree of 248 genomes, representing strains isolated from human (dark blue), animal (red), food (orange) and environment (light blue) and collected from Europe (dark green), North America (purple), Asia (pink) and Australia (light green). Branches with bootstrap confidence values above 0.7 are shown as solid lines. The phylogeny demonstrates clear mixing of European and North American strain indicating multiple transmission events between continents and mixing of human and animal strains indicating multiple transmissions events between these hosts. Closely related clusters (see Table 1) containing both human and animal isolates are labeled 1–6 and highlighted in yellow.
Number of accessory genes

Antibiotic Resistance Gene

Frequency of isolates (%) Human and Animal (n=1213) Human (n=1358) Animal (n=288)

Antibiotic Resistance Gene

Human
Animal
Figure 2. Indistinguishable accessory genome of C. difficile RT078 harbours a variety of antimicrobial resistance genes.

A. The accessory genes (n=2,859) categorized according to host origin. The number of accessory genes (x-axis) only found in human genomes (dark blue), only found in animal genomes (red) or found in both human and animal genomes (green) is plotted against the number of genomes in which these genes are present (y-axis).

B. The frequency of predicted antimicrobial resistances genes (ARGs) within the 243 C. difficile RT078 strains. Human (dark blue) and animal (red) isolation sources are shown by color.
Table 1. Table of 6 highly similar *C. difficile* RT078 clusters identified as identical through core genome analysis, where isolates from both human and animal are present. Average Nucleotide Identity (ANI) for human isolate compared to the animal isolate is also shown.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>ENA ID</th>
<th>Year</th>
<th>Continent</th>
<th>Country</th>
<th>Host</th>
<th>ANI (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ERR171209</td>
<td>2004</td>
<td>North America</td>
<td>Canada</td>
<td>Animal</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ERR171230</td>
<td>2010</td>
<td>Europe</td>
<td>UK</td>
<td>Human</td>
<td>99.93</td>
</tr>
<tr>
<td></td>
<td>ERR256911</td>
<td>2011</td>
<td>Europe</td>
<td>UK</td>
<td>Human</td>
<td>99.91</td>
</tr>
<tr>
<td></td>
<td>ERR171303</td>
<td>2008</td>
<td>Europe</td>
<td>UK</td>
<td>Human</td>
<td>99.90</td>
</tr>
<tr>
<td></td>
<td>ERR256986</td>
<td>2012</td>
<td>Europe</td>
<td>UK</td>
<td>Human</td>
<td>99.84</td>
</tr>
<tr>
<td></td>
<td>ERR256910</td>
<td>2011</td>
<td>Europe</td>
<td>UK</td>
<td>Human</td>
<td>99.83</td>
</tr>
<tr>
<td></td>
<td>ERR1910469</td>
<td>1997</td>
<td>Europe</td>
<td>UK</td>
<td>Human</td>
<td>99.82</td>
</tr>
<tr>
<td></td>
<td>ERR1910468</td>
<td>1997</td>
<td>Europe</td>
<td>UK</td>
<td>Human</td>
<td>99.80</td>
</tr>
<tr>
<td></td>
<td>ERR256981</td>
<td>2008</td>
<td>Europe</td>
<td>UK</td>
<td>Human</td>
<td>99.75</td>
</tr>
<tr>
<td>2</td>
<td>ERR257071</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Animal</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ERR257072</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Human</td>
<td>99.94</td>
</tr>
<tr>
<td>3</td>
<td>ERR257053</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Animal</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ERR257057</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Human</td>
<td>99.77</td>
</tr>
<tr>
<td>4</td>
<td>ERR257067</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Animal</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ERR171352</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Human</td>
<td>99.97</td>
</tr>
<tr>
<td></td>
<td>ERR257052</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Human</td>
<td>99.91</td>
</tr>
<tr>
<td>5</td>
<td>ERR257046</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Animal</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ERR257061</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Human</td>
<td>99.82</td>
</tr>
<tr>
<td>6</td>
<td>ERR257065</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Animal</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ERR257044</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Human</td>
<td>99.80</td>
</tr>
<tr>
<td></td>
<td>ERR257050</td>
<td>2011</td>
<td>Europe</td>
<td>Netherlands</td>
<td>Human</td>
<td>99.73</td>
</tr>
</tbody>
</table>