Zoonotic transfer of *Clostridium difficile* harboring antimicrobial resistance between farm animals and humans

C.W. Knetsch[1,†], N. Kumar[2,†], S.C. Forster[2,3,4], 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,*]

[1]Section Experimental Bacteriology, Department of Medical Microbiology, Leiden University Medical Center, Leiden, Netherlands.
[2]Host-Microbiota Interactions Laboratory, Wellcome Trust Sanger Institute, Hinxton, United Kingdom.
[3]Centre for Innate Immunity and Infectious Diseases, Hudson Institute of Medical Research, Clayton, Victoria, Australia.
[4]Department of Molecular and Translational Sciences, Monash University, Clayton, Victoria, Australia.
[5]Cardiff School of Biosciences, Sir Martin Evans Building, Cardiff, United Kingdom.
[6]Pathogen Genomics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom.
[7]Public Health Wales, Microbiology, Wales, United Kingdom.
[8]Institute of Translational Medicine, University of Liverpool, Liverpool, United Kingdom.
[9]Department of Veterinary Science and Microbiology, The University of Arizona, Tucson, Arizona, USA.
[10]Department of Pathobiology, Canada Veterinary College, University of Guelph, Guelph, Canada.
[11]Institute of Medical Microbiology and Hygiene, Österreichische Agentur für Gesundheit und Ernährungssicherheit (AGES), Vienna, Austria.
[12]Faculty of Medicine, University of Maribor, Maribor, Slovenia.
[13]National laboratory for Health, Environment and Food, Maribor, Slovenia.
[14]Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom.
[15]Department of Microbiology, PathWest Laboratory Medicine, Queen Elizabeth II Medical Centre, Western Australia, Australia.
[16]Microbiology & Immunology, School of Pathology & Laboratory Medicine, The University of Western Australia, Western Australia, Australia.

*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.

74 **Introduction:** Over the past decade, *Clostridium difficile* has emerged as the primary cause
75 of infectious antibiotic associated diarrhea in hospitalized patients (1). Unlike other common
76 healthcare-associated pathogens, *C. difficile* produces resistant spores that facilitate host-to-
77 host transmission and enable long term survival and dispersal in the healthcare system and
78 the wider environment (2). The emergence of epidemic *C. difficile* ribotype (RT) 027 (NAP1
79 / ST-1), responsible for many large-scale hospital outbreaks worldwide (3, 4), has been
80 linked to environmental spore contamination and the acquisition of fluoroquinolone
81 resistance (5). Enhanced research focus on *C. difficile* in the aftermath of the *C. difficile*
82 RT027 outbreaks has revealed other evolutionarily distinct *C. difficile* lineages, in particular
83 *C. difficile* RT078 (NAP07-08/ST-11), that are now emerging as significant human pathogens
84 for unknown reasons (6).

85 The "One Health" concept, connecting the health of humans to the health of animals
86 and their shared environments, represents a relevant framework for understanding the
87 emergence and spread of pathogens. *C. difficile* RT078 is commonly isolated from both
88 humans and farm animals (7) and increasingly recognized as a causative agent of both
89 healthcare and community-associated *C. difficile* infection (CDI) (8). This lineage typically
90 affects a younger population (9) and results in higher mortality than *C. difficile* RT027 (10).
91 Standard genotyping tools have highlighted genetic similarities between human and animal
92 *C. difficile* RT078 (11-13) strains raising the possibility of zoonotic transmission (14).
93 Nevertheless, the exact evolutionary and epidemiological relationships between human and
94 animal *C. difficile* RT078 strains remain unknown due to the lack of discriminatory power of
95 these typing methods and the clonal nature of *C. difficile* lineages. Recently, using whole
96 genome phylogeny, we reported that asymptomatic farmers and their pigs can be colonized
97 with clonal *C. difficile* RT078 isolates demonstrating evidence for spread between animals
98 and humans (15).

Journal of Clinical Microbiology

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) (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 $\geq$ 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

Journal of Clinical Microbiology

123 animals as found previously (15). The existence of highly related human and animal isolates
124 suggests that *C. difficile* RT078 has frequently spread between animals and humans.

125 Next, a detailed analysis of the accessory genome, including mobile genetic
126 elements, was performed to further explore the genomic similarities between human and
127 animal strains. Of the 6,239 unique genes present across our genome collection, 3,368 genes
128 (54.0%) were assigned to the core genome leaving 2,871 genes (46.0%) present in the
129 accessory genome (Fig. S2). Considering only the human and animal isolates, 2,859
130 accessory genes were identified. The vast majority of human and animal specific accessory
131 genes were found at low frequencies in the population (Fig. 2A). We observed no statistically
132 significant difference in the number of strains carrying accessory genes exclusive to either the
133 human or the animal population ($\chi 2$ p-value of 0.39). Considering only those accessory genes
134 present in at least 10% of isolates (n= 465), 461 (99.1%) were identified in both human and
135 animal isolates. The absence of accessory genes unique to either group demonstrates that
136 either *C. difficile* has a stable accessory genome, which is host independent or provides
137 further support for the frequent transmission of *C. difficile* between host populations.

138 Given the high percentage of mobile elements including antimicrobial resistance
139 genes harbored by *C. difficile* genomes (5, 6), we next sought to analyze distribution of
140 different ARGs in the pangenome of human and animal strains. In total, 22 different putative
141 ARGs are present in the 243 *C. difficile* RT078 genomes (Fig. 2B). The most common ARG
142 was the chromosome encoded *cdeA*, a well-known multidrug transporter that was detected in
143 all strains; however, other common genes included those encoding resistance to
144 aminoglycosides, tetracycline and erythromycin (Fig. 2B). Importantly, no specific ARGs
145 were statistically enriched in the animal isolates; however, the *ermB* (erythromycin resistance
146 methylase) gene was identified in the human isolates (Fisher's-exact test, q value = 1.25E-
147 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

Journal of Clinical Microbiology

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

205 **Construction and analysis of the Pan genome**

206 We used the pan genome pipeline Roary (26), to identify the *C. difficile* RT078 pan
207 genome. Roary takes annotated draft assemblies in GFF3 format which were produced by
208 Prokka (25). Predicated coding regions were extracted from the input and converted to
209 protein sequences. Partial sequences (>5% nucleotides unknown or sequence length less than
210 120 nucleotides) were filtered and the remaining sequences were iteratively clustered with
211 CD-HIT beginning with a sequence identity of 100% and matching length of 100% down to a
212 default sequence identity of 98%. One final clustering step was performed again with CD-hit,
213 with a sequence identity of 100% leaving one representative sequence for each cluster in a
214 protein FASTA file. This was followed by a comprehensive, pairwise comparison with
215 BLASTP on the reduced sequences with a default sequence identity percentage of 95% and
216 matching length of 100%. The pan genome embodies the core genome, defined as those
217 genes present in at least 90% of the genomes, and the accessory genome, defined as those
218 genes present in between 10% and 90% of the genomes. Rare variant genes, found in less
219 than 10% of genomes, were discarded.

220 Core genes (n=3,368) alignment, an output from Roary, was used to construct
221 phylogenetic structure of 248 *C. difficile* strains. Single nucleotide polymorphisms (SNPs)
222 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).

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).

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/ariba).

Acknowledgments:

We acknowledge the Wellcome Trust [098051]; the United Kingdom Medical Research Council [PF451 to TL, MR/L015080/1 to TC and MR/K000551/1to 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.

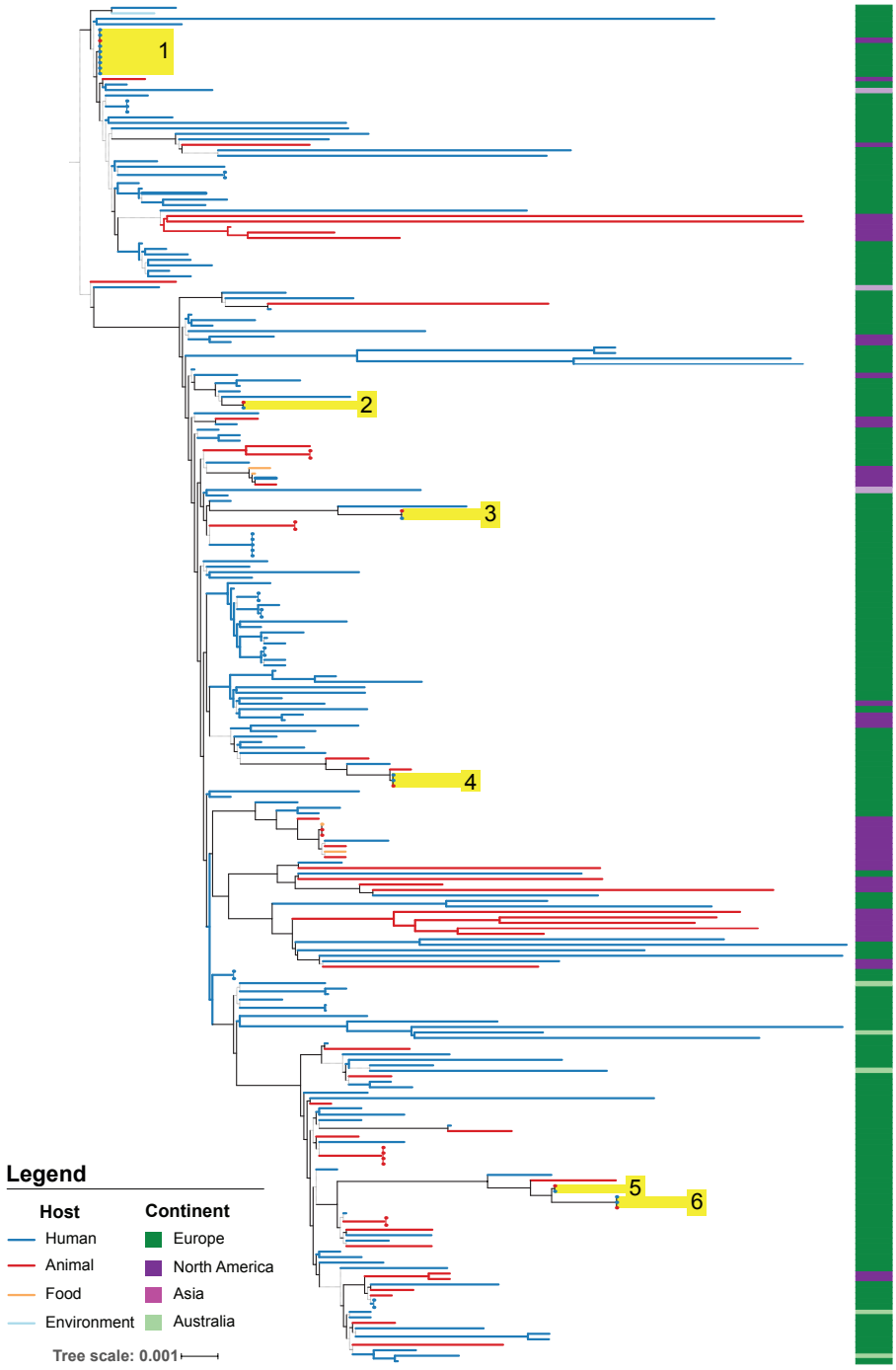
References:

1. Lessa FC, Winston LG, McDonald LC, Emerging Infections Program CdST. 2015. Burden of *Clostridium difficile* infection in the United States. N Engl J Med 372:2369-2370.
2. Leffler DA, Lamont JT. 2015. *Clostridium difficile* Infection. N Engl J Med 373:287-288.
3. Kuijper EJ, Coignard B, Tull P, difficile ESGfC, States EUM, European Centre for Disease P, Control. 2006. Emergence of *Clostridium difficile*-associated disease in North America and Europe. Clin Microbiol Infect 12 Suppl 6:2-18.
4. Loo VG, Poirier L, Miller MA, Oughton M, Libman MD, Michaud S, Bourgault AM, Nguyen T, Frenette C, Kelly M, Vibien A, Brassard P, Fenn S, Dewar K, Hudson TJ, Horn R, Rene P, Monczak Y, Dascal A. 2005. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. N Engl J Med 353:2442-2449.
5. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, Connor TR, Harris SR, Fairley D, Bamford KB, D'Arc S, Brazier J, Brown D, Coia JE, Douce G, Gerding D, Kim HJ, Koh TH, Kato H, Senoh M, Louie T, Michell S, Butt E, Peacock SJ, Brown NM, Riley T, Songer G, Wilcox M, Pirmohamed M, Kuijper E, Hawkey P, Wren BW, Dougan G, Parkhill J, Lawley TD. 2013. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. Nat Genet 45:109-113.
6. He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, Holt KE, Seth-Smith HM, Quail MA, Rance R, Brooks K, Churcher C, Harris D, Bentley SD, Burrows C, Clark L, Corton C, Murray V, Rose G, Thurston S, van Tonder A, Walker D, Wren BW, Dougan G, Parkhill J. 2010. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. Proc Natl Acad Sci U S A 107:7527-7532.
7. Rupnik M, Widmer A, Zimmermann O, Eckert C, Barbut F. 2008. *Clostridium difficile* toxinotype V, ribotype 078, in animals and humans. J Clin Microbiol 46:2146.
8. Bauer MP, Notermans DW, van Benthem BH, Brazier JS, Wilcox MH, Rupnik M, Monnet DL, van Dissel JT, Kuijper EJ, Group ES. 2011. *Clostridium difficile* infection in Europe: a hospital-based survey. Lancet 377:63-73.
9. Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, Bergwerff AA, Dekker FW, Kuijper EJ. 2008. Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clin Infect Dis 47:1162-1170.
10. Walker AS, Eyre DW, Wyllie DH, Dingle KE, Griffiths D, Shine B, Oakley S, O'Connor L, Finney J, Vaughan A, Crook DW, Wilcox MH, Peto TE, Infections in Oxfordshire Research D. 2013. Relationship between bacterial strain type, host biomarkers, and mortality in *Clostridium difficile* infection. Clin Infect Dis 56:1589-1600.
11. Bakker D, Corver J, Harmanus C, Goorhuis A, Keessen EC, Fawley WN, Wilcox MH, Kuijper EJ. 2010. Relatedness of human and animal *Clostridium difficile* PCR ribotype 078 isolates determined on the basis of multilocus variable-number tandem-repeat analysis and tetracycline resistance. J Clin Microbiol 48:3744-3749.
12. Keel K, Brazier JS, Post KW, Weese S, Songer JG. 2007. Prevalence of PCR ribotypes among *Clostridium difficile* isolates from pigs, calves, and other species. J Clin Microbiol 45:1963-1964.
13. Keessen EC, Gaastra W, Lipman LJ. 2011. *Clostridium difficile* infection in humans and animals, differences and similarities. Vet Microbiol 153:205-217.

Journal of Clinical Microbiology

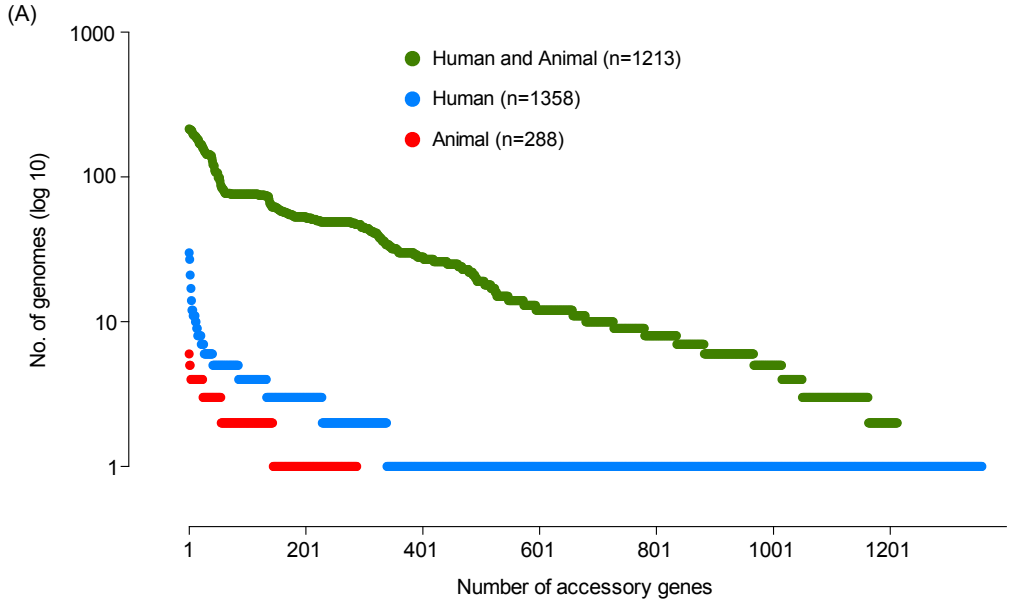
14. Squire MM, Riley TV. 2013. *Clostridium difficile* infection in humans and piglets: a 'One Health' opportunity. Curr Top Microbiol Immunol 365:299-314.
15. Knetsch CW, Connor TR, Mutreja A, van Dorp SM, Sanders IM, Browne HP, Harris D, Lipman L, Keessen EC, Corver J, Kuijper EJ, Lawley TD. 2014. Whole genome sequencing reveals potential spread of *Clostridium difficile* between humans and farm animals in the Netherlands, 2002 to 2011. Euro Surveill 19:20954.
16. Mather AE, Reid SW, Maskell DJ, Parkhill J, Fookes MC, Harris SR, Brown DJ, Coia JE, Mulvey MR, Gilmour MW, Petrovska L, de Pinna E, Kuroda M, Akiba M, Izumiya H, Connor TR, Suchard MA, Lemey P, Mellor DJ, Haydon DT, Thomson NR. 2013. Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. Science 341:1514-1517.
17. Stubbs SL, Brazier JS, O'Neill GL, Duerden BI. 1999. 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 37:461-463.
18. Argimon S, Abudahab K, Goater RJ, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MT, Yeats CA, Grundmann H, Spratt BG, Aanensen DM. 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom 2:e000093.
19. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD. 2011. Rapid pneumococcal evolution in response to clinical interventions. Science 331:430-434.
20. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science 327:469-474.
21. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ. 2008. A large genome center's improvements to the Illumina sequencing system. Nat Methods 5:1005-1010.
22. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821-829.
23. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578-579.
24. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. Genome Biol 13:R56.
25. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069.
26. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691-3693.
27. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial Genomics 2.
28. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26:1641-1650.
29. Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39:W475-478.
30. Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res 30:2478-2483.

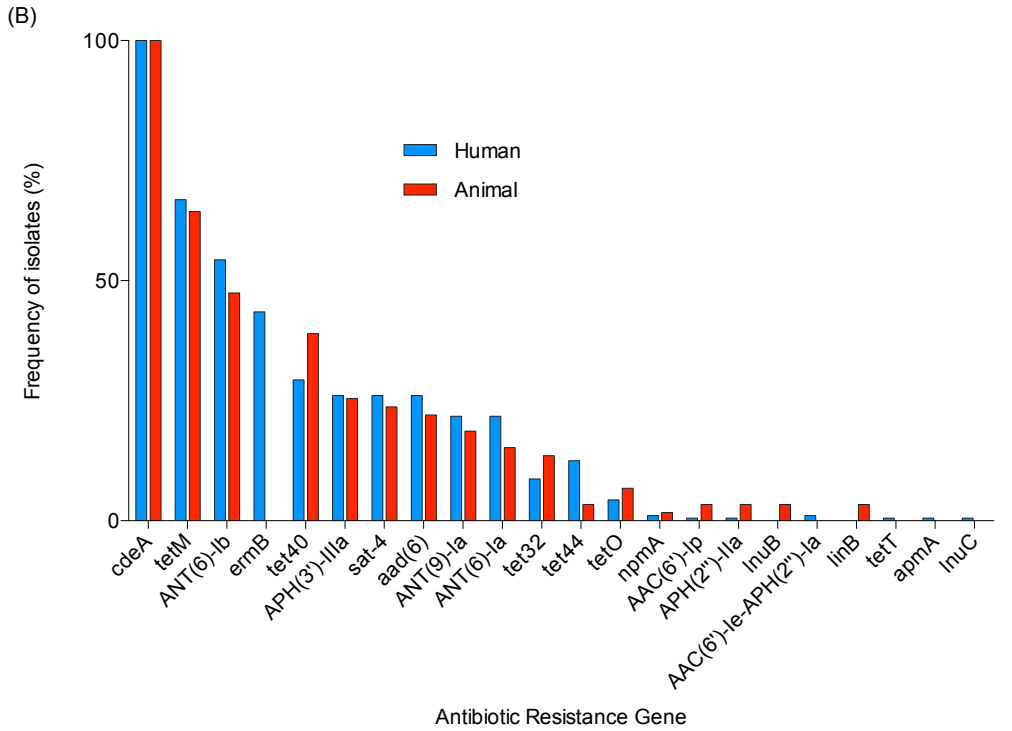
389 **Figures**



390

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.





406

407

408

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.

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