bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 **A role for tetracycline selection in the evolution of *Clostridium difficile* PCR-ribotype**
2 **078**

3 Running Title: Tetracycline selection in *C. difficile* RT078

4

5 Kate E. Dingle[1,2,3], Xavier Didelot[4], T. Phuong Quan[1,2,3], David W. Eyre[1,2], Nicole
6 Stoesser[1,2], Charis A. Marwick[5], John Coia[6], Derek Brown[6], Sarah Buchanan[7], Umer Z. Ijaz[7],
7 Cosmika Goswami[7], Gill Douce[7], Warren N. Fawley[8], Mark H. Wilcox[8], Timothy E.A.
8 Peto[1,2,3], A. Sarah Walker[1,2,3], and Derrick W. Crook[1,2,3]

9

10 [1]Nuffield Department of Clinical Medicine, John Radcliffe Hospital, University of Oxford,
11 UK, OX3 9DU.
12 [2]National Institute for Health Research (NIHR) Oxford Biomedical Research Centre, John
13 Radcliffe Hospital, Oxford, UK, OX3 9DU.
14 [3]NIHR Oxford Health Protection Research Unit on Healthcare Associated Infection and
15 Antimicrobial Resistance, Oxford University, John Radcliffe Hospital, Oxford, OX3 9DU.
16 [4]Department of Infectious Disease Epidemiology, Imperial College, London, UK, SW7 2AZ.
17 [5]Population Health Sciences, School of Medicine, University of Dundee, Scotland, UK, DD1
18 4HN.
19 [6]Scottish Microbiology Reference Laboratories, Glasgow, UK, G31 2ER.
20 [7]University of Glasgow, Scotland, UK, G12 8QQ.
21 [8]Leeds Teaching Hospitals and University of Leeds, Department of Microbiology, Leeds
22 General Infirmary, Leeds, UK, LS1 3EX.

23

24 **Corresponding author**: Kate E. Dingle

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

25 **Address**: Microbiology (Level 7), Nuffield Department of Clinical Medicine, University of

26 Oxford, John Radcliffe Hospital, Oxford, UK, OX3 9DU.

27 **email**: kate.dingle@ndm.ox.ac.uk, **phone**: +44 1865 220870.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

28 **ABSTRACT**

29 Farm animals have been identified as reservoirs of *Clostridium difficile* PCR-ribotype 078
30 (RT078). Since 2005, the incidence of human clinical cases (frequently severe), with this
31 genotype has increased. We aimed to understand this change, by studying the recent
32 evolutionary history of RT078. Phylogenetic analysis of international genomes (isolates from
33 2006-2014) revealed several recent clonal expansions. A common ancestor of each expansion
34 had independently acquired different alleles of the tetracycline resistance gene *tetM*.
35 Consequently, an unusually high proportion of RT078 genomes were *tetM* positive (76.5%).
36 Additional tetracycline resistance determinants were also identified, some for the first time in
37 *C. difficile* (efflux pump *tet40*). Each *tetM*-clonal expansion lacked geographic structure,
38 indicating rapid international spread. Resistance determinants for *C. difficile*-infection-
39 triggering antimicrobials including fluoroquinolones and clindamycin were comparatively
40 rare in RT078. Tetracyclines are used intensively in agriculture; this selective pressure, plus
41 rapid spread via the food-chain may explain the increased RT078 prevalence in humans.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

42 **INTRODUCTION**

43 *Clostridium difficile* infection is a significant international challenge, affecting patients in
44 community and healthcare environments worldwide (Davies et al., 2014; Lessa et al., 2015;
45 Wilcox et al., 2008). The severity of symptoms ranges from mild diarrhoea to
46 pseudomembranous colitis and toxic megacolon. Crude 30-day mortality in the UK is 16%
47 (in an endemic setting) and can exceed 30% (Planche et al., 2013; McGowan et al., 2011),
48 while in the US it has been estimated that almost half a million CDIs caused 29,000 deaths in
49 a single year (Lessa et al., 2015).
50
51 The molecular epidemiology of *Clostridium difficile* infection varies both temporally and
52 geographically, frequently in response to local antimicrobial prescribing (Lessa et al., 2015;
53 Public Health England,2016; Wilcox et al., 2012; Dingle et al., 2017). Clinically important
54 outbreak-associated genotypes can emerge when the inherent resistance of *C. difficile* to
55 cephalosporins (Shuttleworth et al., 1980) is supplemented with acquired resistance to certain
56 high risk antimicrobials, including clindamycin (Johnson et al., 1999) and more recently
57 fluoroquinolones. The latter contributed to the emergence of multiple phylogenetically
58 unrelated outbreak-associated genotypes including the 'hypervirulent' PCR-ribotype 027
59 (Loo et al., 2005; He et al., 2017; Spigaglia., 2016; Dingle et al., 2017). However, the
60 reason(s) for the changing prevalence of other clinically important *C. difficile* genotypes is
61 frequently unknown (Belmares et al., 2009). These include RT078, which is unusual in
62 having an agricultural link (Goorhuis et al., 2008).
63
64 The increased importance of *C. difficile* RT078 as a human pathogen was first reported in
65 The Netherlands, rising from 3% to 13% of *Clostridium difficile* infection cases during 2005-
66 2008 (Goorhuis et al., 2008). Around the same time, a ten fold increase was noted in North

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

67 America (Jhung et al., 2008). Similar increases and occasional outbreaks were subsequently
68 recorded throughout Europe (Barbut et al., 2007; Bauer et al., 2011; Burns et al., 2010) and
69 *C. difficile* RT078 has recently increased to 4.4%, 9.7% and 8.1% of total *Clostridium*
70 *difficile* infection cases in North America, England and Scotland, respectively (Lessa et al.,
71 2015; Mulvey et al., 2010; Fawley et al., 2016; Health Protection Scotland (2017). Three
72 distinctive features of RT078-associated *Clostridium difficile* infection raise specific
73 concerns, namely: increased severity of disease with the highest genotype-specific mortality
74 rate (Goorhuis et al., 2008; Walker et al., 2013), a higher proportion of community-associated
75 disease, and more infections in younger age groups compared with other genotypes (Lessa et
76 al., 2015; He et al., 2013; Taori et al., 2014).
77
78 The agricultural association of *C. difficile* RT078 reflects its isolation from sick and healthy
79 animals (frequently pigs), bird droppings, vermin and the farm environment (Hensgens et al.,
80 2012; Keel et al., 2007; Bandelj et al., 2014; Burt et al., 2012). However, in common with
81 many other toxin producing *C. difficile* genotypes, ribotype RT078 can be carried
82 asymptomatically by human infants and adults (Stoesser et al., 2017; Knetsch et al., 2014).
83 This genotype has also been isolated from a variety of retail meat products including pork,
84 beef, and others (Curry et al., 2012; Songer et al., 2007). Therefore the natural reservoirs of
85 RT078 support the hypothesis that humans become colonised via the food chain and/or the
86 environment (Hensgens et al., 2012).
87
88 Whole genome sequence data have been used to study the emergence and transmission of
89 many bacterial pathogens. The international dissemination of 'hypervirulent' fluoroquinolone
90 resistant *C. difficile* 027 was revealed in this way (He et al., 2013) and its rapid localised
91 nosocomial transmission was demonstrated, as for other fluoroquinolone resistant genotypes

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

92 (Dingle et al., 2017). Here, we used whole genome sequencing and phylogenetic approaches
93 to study the recent evolutionary history of *C. difficile* RT078 and investigate the hypothesis
94 that the recent clinical prominence of this genotype is due to antimicrobial selection.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

95 **RESULTS**

96 The role of antimicrobial selection in the emergence of *C. difficile* RT078 was investigated
97 using a large, international collection of genomes from the UK, Europe and North America
98 (Table S1, n=400, Dingle et al., 2017; Stoesser et al., 2017; Knetsch et al., 2014; Louie et al.,
99 2011; Cornely et al., 2012). Virtually all RT078 *C. difficile* share the same multilocus
100 sequence type, ST11 (Dingle et al., 2011). However, this ST also includes the very closely
101 related PCR-ribotypes RT126, RT033, RT045 and RT066 (relationship to RT078 shown
102 phylogenetically, Figure S1). These ST11-associated PCR-ribotypes (RT126, RT033, RT045
103 and RT066) are relatively rare clinically. Only one RT078 genome included was a single
104 locus variant of ST11; ST317. The overall proportion of RT078 within the collections from
105 which the study genomes were sourced (see Materials and Methods) was 3-4%, because these
106 data sets are dominated by RT027. Genomes studied are referred to subsequently as RT078.
107 The presence of determinants of resistance to tetracyclines (*tetM*), fluoroquinolones (*gyrA*/*B*
108 substitutions), aminoglycosides (*aphA1* or AAC(6')-APH(2')) and clindamycin (*ermB*) within
109 these genomes was assessed (Table 1).

110

111 **Prevalence of antimicrobial resistance determinants in *C. difficile* RT078**

112 *tetM*, the presence of which leads to tetracycline resistant phenotype (Knetsch et al., 2014),
113 was by far the most prevalent antimicrobial resistance determinant in RT078, at 77.5%
114 (148/191) Oxfordshire and Leeds, 76.4% (84/110) Scottish and 75.0% (24/32) North
115 American and European genomes (Table S1). *gyrA*/*B* substitutions, which reduce
116 susceptibility to fluoroquinolones, were less prevalent at 13.0% (25/191) among Oxfordshire
117 and Leeds RT078 genomes, 10.0% (11/110) among Scottish and 18.9% (6/32) in North
118 American and European RT078s. Similarly, *aphA1* (aminoglycoside resistance) was detected
119 in 21.5% (41/191) Oxfordshire and Leeds, 9.1% (10/110) Scottish and 40.6% (13/32) North

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

120 American and European genomes. Finally, *ermB* (clindamycin resistance) occurred in 4.2%
121 (8/191) Oxfordshire and Leeds, 4.5% (5/110) Scottish and 18.8% (6/32) North American and
122 European RT078.
123

124 **Prevalence of *tetM* in the clinical *C. difficile* population**

125 To determine whether the high *tetM* prevalence in RT078 was unusual among clinical *C.*
126 *difficile*, all available additional genomes of other genotypes (defined by ST/PCR-ribotype),
127 from the same isolate collections (except those comprising only RT078 – Scottish and Dutch
128 (unpublished and Knetsch et al., 2014)), were examined for the presence of *tetM* (Figure 1A).
129 The non-RT078 genotypes were described using the notation ST37(017) to indicate, for
130 example, Multilocus Sequence Type 37 and the equivalent (PCR-ribotype 017).
131

132 Genotypes could be classified as (i) >60% *tetM* positive, (ii) >0% but <20 % *tetM* positive
133 (majority <5%) or (iii) *tetM* not detected (Figure 1A). Non-RT078 genotypes were <20%
134 *tetM* positive, with the notable exceptions of ST37(017) (52/80, 65.0%), ST54(012) (35/58,
135 60.0%), ST35(046) (27/39, 69.2%), and ST48 (8/12. 67.0%), plus non-toxigenic ST26(140)
136 (65/65, 100%) (Figure 1A). Therefore, at over 75%, RT078 was the most *tetM* positive
137 clinically relevant genotype.
138

139 **Prevalence of additional tetracycline resistance determinants in the clinical *C. difficile***
140 **population**

141 Each genotype was assessed for the presence of additional tetracycline resistance
142 determinants (Table 1, Figure 1B). The tetracycline efflux pump gene *tet40* (not previously
143 described in *C. difficile*) was present in 27.3% (109/400) RT078 (Figure 1B), and non-*tetM*
144 ribosomal protection proteins were present in 5.5% (22/400) RT078 (Figure 1B). In contrast,

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

145 only one or two genomes of other ST/PCR-ribotypes were positive for alternative tetracycline
146 resistance determinants, except ST54(012) (n=54), in which eight examples of four additional
147 tetracycline resistance determinants were found and ST33(216), which contained 10/25
148 (40%) *tetAB(P)* (Figure 1B).
149
150 **UK-representative RT078 Phylogeny**
151 A UK-specific RT078 phylogeny was constructed using genomes from clinical infections in
152 Oxfordshire (n=78), Leeds region (n=104) and Scotland (n=110) (Figure 2A, Table S1).
153 Annotation revealed minimal evidence of geographic structure, contrasting markedly with the
154 highly structured distribution of *tetM* sequences (Figure 2B, C) described in detail below.
155
156 Prior to annotation, distinct *tetM* allele sequences were assigned a number (available at
157 https://pubmlst.org/bigsdb?db=pubmlst_cdifficile_seqdef&page=downloadAlleles). Among
158 the *tetM* positive UK RT078s, three *tetM* alleles predominated; *tetM* 10 (36/292, 12.3%),
159 *tetM* 16 (101/292, 34.6%) and *tetM* 19 (78/292, 26.7%) (Figure 2B). Coloured bars (or
160 branches Figure S1) were used to identify distinct *tetM* alleles (Figure 2B).
161
162 *tetM* alleles 10, 16 and 19 were each carried by closely related, Tn916-like conjugative
163 transposons. Independent acquisition events, estimated from the phylogeny to have occurred
164 between 1995-2006, were suggested by their unique chromosomal insertion sites (Figure 2B).
165 Acquisition of *tetM* 16 or *tetM* 19 was associated with significantly shorter branch lengths
166 (confirmed by median Evolutionary Distinctiveness scores equalling 3.78 and 3.58
167 respectively, versus 7.22 for branches representing genomes lacking a ribosomal protection
168 protein gene, $p<0.001$). This observation is consistent with clonal expansion in response to
169 tetracycline-associated selection pressure (Figure 2B, C); significantly lower ED scores

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

170 indicate unexpectedly short branches (Isaac et al., 2007). It is possible that for a given branch
171 there could be some other genetic change other than *tetM* that is the cause of the clonal
172 expansion, but since we see the same pattern on several independent branches where *tetM*
173 was acquired, it seems very likely that this underlies the clonal expansion. The acquisition of
174 efflux pump *tet40* on its own was not associated with clonal expansion, with only a slightly
175 higher median Evolutionary Distinctiveness score compared with *tet40* absence (Figure 2C).
176
177 The same phylogeny was annotated for the presence of additional resistance determinants
178 (conferring aminoglycoside, fluoroquinolone or clindamycin resistance, Figure S2A-C), but
179 no evidence of associated clonal expansions was found independent of the *tetM*-associated
180 expansions (Figure S2D-F).
181
182 **RT078 Phylogenies representing UK Regions**
183 Separate phylogenies were constructed to examine the detailed evolutionary history of RT078
184 within two of the geographic regions represented in the UK-phylogeny. These were Scotland
185 (population 5.295 million, area 30,918 square miles) (Figure 3A, B) and Oxfordshire
186 (population 655,000; area 1,006 square miles) (Figure 3A, C).
187
188 The branches of the Scottish phylogeny (n=110 genomes, Figure 3B) were coloured to
189 represent geographic regions (administrative areas, or ‘health boards’, Figure 3A), thus
190 increasing the level of geographic discrimination. As before, geographic structure was absent,
191 healthcare-associated and community isolates intermingling (Figure 3B), but the distribution
192 of the *tetM* alleles 10, 16 and 19 within the phylogeny was highly structured.
193

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

194 The Oxfordshire regional phylogeny (n=94 genomes, Figure 3C), represented a more densely
195 sampled, smaller geographic area (Figure 3A). Here, the EIA-positive *C. difficile* clinical
196 isolate genomes (n=78, as in Figure 2B) were supplemented with EIA-negative clinical
197 isolates (n=9 i.e. isolates patients with diarrhoea but without evidence of toxin production
198 suggesting *C. difficile* was colonising the patient rather than causing disease), non-clinical
199 isolates from healthy infants (n=6) (Stoesser et al., 2017) and a lamb (n=1) (Table S1). All
200 genomes were pathogenicity locus ie. toxin A and B encoding sequence positive (Braun et
201 al., 1996) This regional phylogeny also lacked structure according to location or isolation
202 source, but it was again structured according to *tetM* allele (Figure 3C).
203
204 **International phylogenies confirm three *tetM* positive RT078 clades are present across**
205 **continents**
206 Two international RT078 phylogenies were constructed using genomes from clinical
207 infections in England, (Oxfordshire n=78 and Leeds n=104) supplemented firstly with
208 clinical and non-clinical isolates from The Netherlands (Knetsch et al., 2014) (Table S1,
209 human clinical n=25, human farmers n=15, pig isolates n =20) (Figure 4A, B), and secondly,
210 with genomes from clinical infections in North America (Table S1, USA n=15 and Canada
211 n=4) and Europe (five countries, n=13) (Figure 4C, D). Once again, structure according to
212 geography was absent, but structuring by *tetM* allele (Figure 4A, B) was clear.
213
214 **Tetracycline selection in other *C. difficile* genotypes**
215 Over 60% of genomes belonging to each of five non-RT078 genotypes were *tetM* positive
216 (Figure 1A). Four of these were investigated phylogenetically; ST37(017), ST54(012) and
217 ST35(046) and non-toxigenic ST26(140), (Table S2, Dingle et al., 2017; Stoesser et al.,
218 2017), but not ST48(038/104, as only 12 genomes were available. Branches were coloured

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

219 according to isolation source and geography as before, and *tetM* alleles indicated by coloured
220 bars (Figure 5A-D). As previously for RT078, additional resistance determinants, (Table 1),
221 were also highlighted (when present in five genomes or more), to reveal the possible impact
222 of selection by other antimicrobials (coloured dots, Figure 5A-D).
223
224 A number of recent *tetM* acquisition events were obvious (Figure 5). These were followed by
225 possible clonal expansions most notably within genotypes ST35(046) and non-toxigenic
226 ST26(140) (Figure 5A, B). Clonal expansion was particularly marked in ST26(140), where
227 all genomes were *tetM* positive, and clonal expansion occurred in the absence of disease-
228 causing ability, this genotype being non-toxigenic (lacking the pathogenicity locus, PaLoc, in
229 all genomes (Braun et al., 1996)). With the exception of ST35(046) where aminoglycoside
230 and clindamycin resistance determinants co-localised with *tetM* (Figure 5A), and the
231 fluoroquinolone resistant region of the ST37(017) phylogeny (Figure 6D, Dingle et al., 2017),
232 there was no clear evidence of clonal expansions which had followed the acquisition of the
233 non-*tetM* antimicrobial resistance determinants. In common with RT078, all four phylogenies
234 (Figure 5) lacked geographic structure, with the exception of the fluoroquinolone resistance
235 region of the ST37(017) phylogeny (Figure 5D) (Dingle et al., 2017).
236
237 **Sequences of RT078 tetracycline resistance determinants support its zoonotic origin**
238 The *tetM* sequences described here in *C. difficile* are typical of many Gram-positives,
239 including established zoonotic species. For example, the RT078 *tetM* 10 allele shared 100%
240 nucleotide sequence identity with *tetM* genes of *Streptococcus agalactiae*, *Enterococcus*
241 *faecalis* and *Streptococcus pneumoniae*, and 99% nucleotide sequence identity with
242 *Streptococcus suis* (a pathogen of pigs that can infect humans occupationally (Hoe at al.,
243 2011; Mancini et al., 2016)). Identical *tetM* 10 sequences have also been found in Gram-

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

244 negative bacteria including *Escherichia coli*. The RT078 *tetM* 16 allele shared >97%
245 nucleotide sequence identity with *tetM* in *Enterococcus* species (mostly *E. faecalis* and *E.*
246 *faecium*), followed by *Staphylococcus* and *Streptococcus* species, including *S. suis*.
247
248 Other RT078 tetracycline resistance determinants were also identical, or very closely related
249 to those found in zoonotic bacteria. For example, RT078 *tet40* sequences shared 99-100%
250 identity with *Streptococcus suis tet40* (GenBank KC790465.1) and the RT078 *tetO* sequences
251 shared over 99% nucleotide sequence identity with *Campylobacter jejuni*, *Campylobacter*
252 *coli* and *S. suis tetO* sequences. In addition, the RT078 *tetO/32/O* mosaic sequence shared
253 99% identity with the sequence found in the *S. suis* genome.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

254 **DISCUSSION**

255 Our time-scaled phylogenies revealed parallel *tetM*-associated RT078 clonal expansions,
256 dating from around the year 2000 (Figures 2B, 3B,C, 4B,D). These findings are consistent
257 with an evolutionary response to tetracycline selective pressure, during the timeframe of
258 increasing RT078-associated clinical cases (Goorhuis et al., 2008; Bauer et al., 2011; Burns
259 et al., 2010; Mulvey et al., 2010; Health Protection Scotland, 2017). Tetracyclines were
260 initially introduced around 60 years ago in both clinical and veterinary settings. However,
261 following the emergence of resistance they were largely replaced in human medicine by
262 fluoroquinolones (Thaker et al., 2010). By 2010-13 tetracyclines comprised <18% of total
263 antibiotics consumed in England, and 92% of this was in general practice (Public Health
264 England, 2014). Over the time period relevant to this study, tetracyclines were most
265 commonly used for the treatment of acne and chlamydial sexually transmitted diseases. It is
266 implausible that such prescribing in teenagers and young adults provided extensive selection
267 pressure for *C. difficile*, given that healthy individuals living in the community have very low
268 colonisation rates for these bacteria (Manzoor et al., 2017). This study therefore provides
269 evidence of a plausible agricultural link underlying the emergence of RT078, by presenting
270 its recent evolutionary history with respect to the acquisition of antimicrobial resistance.
271
272 In contract to humans, tetracyclines remain the most widely used antimicrobial for the
273 treatment of infections in animals (European Medicines Agency (2016). In addition, their use
274 as growth promoters (in sub-therapeutic doses) continues outside the EU, this being the only
275 region to ban the practice (Scheel 1970). During 2015, 6,880 metric tonnes of tetracyclines
276 were sold in the USA (US Food and Drug Administration 2015) (representing a 31% increase
277 from 2009), compared to 166 tonnes in the UK (Veterinary Medicines Directorate 2015). The
278 extent of agricultural tetracycline use, the prevalence of RT078 in animals used for food

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

279 (Hensgens et al., 2012; Keel et al., 2007; Curry et al., 2012; Songer et al., 2007), and the
280 timeframe of RT078 emergence all implicate tetracycline use in agriculture as a plausible
281 source of selective pressure. The global food chain provides an obvious route for rapid
282 RT078 dissemination, but indirect transmission from the agricultural environment to humans
283 via contaminated water or vegetables (Xu et al., 2014) is also possible.
284
285 The absence of geographic structure within our RT078 phylogenies is consistent with its
286 rapid international spread (Figure 4), as is the absence of large-scale, localised nosocomial
287 outbreaks (Figures 2B, 3B,C, 4B,D). Among other genotypes, such outbreaks have been
288 associated with extensive prescribing of, and resistance to, high risk antimicrobials such as
289 clindamycin, cephalosporins and fluoroquinolones. In our study, large scale clonal
290 expansions were not associated with fluoroquinolone or clindamycin resistance in RT078
291 (Figure S2). Equivalent analysis for cephalosporins cannot be performed because the genetic
292 mechanism(s) of resistance in *C. difficile* have yet to be defined and although MICs can vary,
293 *C. difficile* has typically been considered inherently cephalosporin resistant, irrespective of
294 genotype (Shuttleworth et al., 1980). The international spread of RT078 indicates that
295 changes in antimicrobial resistance phenotype could potentially impact at any location,
296 depending on local prescribing practices. Consequently, ST11(126), a frequently
297 fluoroquinolone resistant descendant of RT078 (Figure S1), which is prevalent in Italian
298 clinical settings (Spigaglia et al., 2010; Freeman et al., 2015) is of particular concern, as is the
299 epidemic multidrug resistant RT078 observed in Spanish swine (Peláez et al., 2013).
300
301 The identification of widespread tetracycline resistance (>60% *tetM* positive) in only five *C.*
302 *difficile* genotypes besides 078 (Figure 1A) is consistent with previous reports (Barbut et al.,
303 2007; Huang et al., 2009; Noren et al., 2009). Phylogenetic analysis (Figure 5) showed that

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

304 ST35(046) contained two plausible *tetM*-associated clonal expansions (Figure 5A), but the
305 relatively small numbers precluded quantitative 'Evolutionary Distinctiveness' analysis. Like
306 RT078, ST35(046) has been found in pigs and has caused human outbreaks of *C. difficile*
307 infection (Akerlund et al., 2011). This genotype also illustrates the possibility that selection
308 by one antimicrobial can drive the acquisition of further, linked antibiotic resistance genes, as
309 almost every *tetM* positive ST35(046) genome was also positive for clindamycin and
310 aminoglycoside resistance determinants (Figure 5A). Non-toxigenic ST26(140) contained
311 *tetM* in every genome examined, suggesting stable integration predating a recent clonal
312 expansion (Figure 5B), concurrent with that of RT078. ST26(140) therefore illustrates the
313 consequences of tetracycline selection in a harmless commensal organism, confirming that
314 tetracycline selection may have been sufficient to drive the emergence of RT078.
315
316 RT078 has many resistance determinants in common with zoonotic pathogens such as
317 *Streptococcus suis*, *Campylobacter jejuni* and *C. coli*. Quantitative analysis using
318 'Evolutionary Distinctiveness' (Figure 2C, 2S E-F), confirmed that *tetM* was associated with
319 *C. difficile* 078 clonal expansions. Although widespread in RT078, the tetracycline efflux
320 pump *tet40* did not on its own show such as association (Figure 2B,C). Efflux pumps often
321 confer a low-level resistance phenotype, assisting bacterial survival at sub-lethal
322 concentrations of antimicrobials (for example *tetK* in LA-MRSA CC398, Larsen et al., 2016).
323 They thereby function in promoting the acquisition of further high level resistance
324 determinants, such as *tetM*. The parallels between RT078 and zoonotic *Streptococcus suis*
325 also extend to their epidemiology. *S. suis* is a globally distributed emergent pathogen of
326 humans (Werteim et al., 2009), commonly isolated from pigs, geographic clustering of
327 subpopulations is absent (Weinert et al., 2015) and *S. suis* exhibits rapid, recent increases in
328 tetracycline resistance (Hoa et al., 2011). The emergence of human pathogens, coincident

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

329 with tetracycline resistance acquisition, has also been noted among other bacterial species. In
330 Group B streptococcus, tetracycline resistance may have contributed to its emergence as a
331 leading cause of human neonatal infections (da Cunha et al., 2014). Among livestock-
332 associated methicillin-resistant *Staphylococcus aureus* CC398, almost all isolates are *tetM*
333 positive, and often also carry the *tetK* (efflux pump) (Larsen et al., 2016).
334
335 Although multiple lines of evidence indicate a role for tetracycline selection in the recent
336 evolutionary history of RT078, the possibility exists that further genetic changes (unrelated to
337 tetracycline resistance) contributed to its *tetM*-associated clonal expansions (Figures 2-4,
338 Figure S1). The role of selection by other antimicrobials (fluoroquinolones, clindamycin,
339 aminoglycosides) was investigated (Figure S2A-F), and a small potential contribution by
340 aminoglycosides was indicated by the presence of the resistance gene *aphA1* in a minority of
341 RT078 (Figure S2A). However, *aphA1* could not be assessed independent of *tetM* (Figure S1)
342 because the two genes co-localised. Further work would be required to compare the total
343 gene content of *tetM* positive RT078 isolates with older *tetM* negatives, to identify further
344 potentially relevant genetic differences that could explain the clonal expansions. The
345 identification of *tetM*-associated clonal expansions in genetically divergent *C. difficile*
346 genotypes; ST35(046) (together with clindamycin and aminoglycoside resistance
347 determinants, Figure 5A) and non-toxigenic ST26(140) (Figure 5B), serve to further highlight
348 *tetM* as a factor common to recent clonal expansions within distinct *C. difficile* genetic
349 backgrounds. To further confirm the zoonotic origin of RT078, additional data on the
350 changing use of tetracycline over time, and concurrent isolates from clinical cases and farm
351 animals would be useful, although challenging to source retrospectively.
352

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

353 In summary, numerous lines of evidence support the hypothesis that tetracycline use,
354 plausibly in agriculture, has provided recent selection pressure which has impacted on the
355 evolution of tetracycline resistant RT078. The major *C. difficile* RT078 transmission routes to
356 humans are consequently more likely to be related to agriculture and international food
357 chains than nosocomial. Our work strongly suggests that the use of antimicrobials outside the
358 healthcare environment can not only select for resistant organisms, but can contribute to the
359 emergence of human pathogens. Furthermore, our findings add to the body of evidence
360 supporting initiatives such as 'One Health' (One Health Initiative, 2018), which aims to
361 expand interdisciplinary collaboration in all aspects of health care for humans, animals and
362 the environment.
363

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

364 **METHODS**

365 ***C. difficile* Whole Genome Sequences**

366 *C. difficile* genomes derived from isolates of either RT078 or ST11 (n=400) were sourced
367 from several published collections (Dingle et al., 2017; Stoesser et al., 2017; Knetsch et al.,
368 2014; Louie et al., 2011; Cornely et al., 2012), as well as from an unpublished Scottish
369 collection and the Oxford University farm, Wytham, UK; Table S1 and items (i) to (v) below.
370 Each isolate was obtained from a distinct sample. EIA-negative isolates were inferred to be
371 toxigenic or non-toxigenic, depending on the presence/absence of the toxin encoding
372 pathogenicity locus (PaLoc) (Braun et al., 1996) (Tables S1, S2). Complete collections in (i),
373 (iii) and (iv) below have been described previously (Dingle et al., 2017; Stoesser et al., 2017;
374 Knetsch et al., 2014).
375

376 **(i) Clinical *C. difficile*: Oxfordshire and Leeds, UK**

377 Genomes were available for 87 RT078 or ST11 *C. difficile* isolates cultured from
378 symptomatic Oxfordshire patients by the Clinical Microbiology Laboratory, Oxford
379 University Hospitals NHS Trust, Oxford between September 2006 - April 2013 (Dingle et al.,
380 2017). Seventy-eight isolates were derived from toxin immunoassay (EIA)-positive stools
381 (initially the Meridian Premier Toxins A&B Enzyme Immunoassay [Meridian Bioscience
382 Europe, Milan, Italy], until April 2012 and subsequently the TechLab Tox A/B II assay
383 [TechLab Inc, Blacksburg, VA, USA]), and nine from EIA-negative, but glutamate
384 dehydrogenase (GDH)-positive (by the Premier *C. difficile* GDH EIA, Meridian Bioscience
385 Europe, Milan, Italy) (Dingle et al., 2017). Genomes were also available for 104 RT078 *C.*
386 *difficile* positive stool samples (identified using cytotoxin testing) obtained from routinely
387 examined, diarrhoeal faecal samples at the Leeds Teaching Hospitals NHS Trust, between
388 August 2010 - April 2013 (Table S1).

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

389

390 **(ii) Clinical *C. difficile*: Scotland, UK**

391 Isolates from Scotland, UK, included 109 isolates of RT078, and one closely related RT066
392 isolate (Table S1). These isolates form part of a collection stored at the Scottish Microbiology
393 Reference Laboratory (Glasgow). Cultures are provided by all twelve Scottish regional NHS
394 Healthcare Boards in the event of a severe/fatal case, a suspected outbreak or a suspected
395 ribotype 027 infection. In addition each Health Board provides a fixed number of samples
396 based on the rates of infection/population. This allows surveillance of prevalent circulating
397 strains to be assessed. RT078 isolates for this study were selected based on the ribotypes from
398 samples referred to the Reference laboratory between November 2007 - October 2014, and
399 aiming to provide the widest temporal and geographical representation. Locally, positive
400 faecal stool samples were identified prior to 2009 using a toxin specific EIA (or cell
401 cytotoxicity), and post 2009, using a two step algorithm requiring GDH detection, followed
402 by toxin assessment. These samples were from patients located in healthcare (n=99) and
403 community settings (n=10) (unassigned n=1) from twelve of fourteen Health Boards (as per
404 the map, there were no relevant samples from two small island Health Boards not shown on
405 the map, Figure 3A).

406

407 **(iii) Clinical *C. difficile*: North America and Europe**

408 Thirty-two ST11 genomes from North American (Canada n=4, USA n=15) and European
409 (n=13) *C. difficile* isolates cultured from clinical infections between November 2006 - June
410 2009 were available from a variety of locations from two clinical trials of fidaxomicin (Table
411 S1, showing city and country (Louie et al., 2011; Cornely et al., 2012)). Previously published
412 RT078 genomes from human clinical cases (n=25, 2002-2011) in The Netherlands were also
413 included (Knetsch et al., 2014) (Table S1).

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

414

415 **(iv) Non-Clinical *C. difficile***

416 Six ST11 isolates were cultured from the stools of healthy, asymptomatic Oxfordshire infants
417 (Stoesser et al., 2017) between April - October 2012 (Table S1). A single ST11 isolate was
418 isolated from a lamb at Oxford University Farm, Oxfordshire, UK, June 2009. Previously
419 published *C. difficile* RT078 genomes from farmers (n=15, 2009-2011) and pigs (n=20, 2008-
420 2011) in The Netherlands were included (Knetsch et al., 2014) (Table S1).

421

422 **(v) PCR-Ribotype Reference Isolates**

423 For additional context, five genomes from PCR-ribotype reference *C. difficile* representing
424 RT078, RT126, RT033, RT045 and RT066 were included, all of which are genetically very
425 closely related, sharing the same multilocus sequence type, ST11 (Table S1, Figure S1).

426

427 **(vi) Four additional genotypes**

428 Four additional *C. difficile* genotypes were also analysed phylogenetically to contextualise
429 findings in RT078; these were ST35(046) (n=34), ST54(012) (n=54), ST37(017) (n=64) and
430 non-toxigenic ST26(140) (n=65) (Table S2). These four genotypes underwent detailed study
431 because they had the highest *tetM* prevalence after RT078. The isolates came from
432 Oxfordshire EIA positive and negative, Oxfordshire Infant, and Leeds region isolate
433 collections described above (Dingle et al., 2017; Stoesser et al., 2017). ST26(140) lacks the
434 toxin-encoding pathogenicity locus and is therefore carried asymptomatically, providing a
435 naturally occurring 'control' for the impact of antimicrobial selection in the absence of
436 disease.

437

438 **Genome Assemblies**

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

439 *C. difficile* genomes were assembled from short reads generated using Illumina technology
440 (Bentley et al., 2008). Reference-based assemblies were made for genomes belonging to *C.*
441 *difficile* clade 5 (ie RT078 and close relatives) as described (Eyre et al., 2013), by mapping
442 reads to the *C. difficile* M120 reference genome (He et al., 2010) and for non-clade 5
443 genomes by mapping to the CD630 reference genome (GenBank AM180355.1) (He et al.,
444 2010) (clades as defined, Dingle et al., 2011).
445 *De novo* assembly was performed using Velvet (version 1.0.7–1.0.18) (Zerbino and Birney.,
446 2008) and VelvetOptimiser 2.1.7 (Gladman and Seeman, 2008), optimising kmer size (k),
447 expected coverage (average kmer coverage of contigs) and coverage cutoff (kmer coverage
448 threshold) to achieve the highest assembly N50 value (length of the smallest contig such that
449 all contigs of that length or less form half of the final assembly).
450 Reads for unassembled genomes have been submitted to NCBI, BioProject ID number
451 PRJNA304087 (Dingle et al., 2017) and PRJNA381384 for the Scottish isolates (accession
452 numbers provided in Tables S1 and S2).
453

454 **Identification of Antimicrobial Resistance Determinants**

455 The *de novo* assemblies were queried using the BLAST function of BIGSdb (Jolley and
456 Maiden 2010) to determine whether genes or non-synonymous point mutations known to
457 confer resistance to antimicrobials including fluoroquinolones, tetracyclines, clindamycin and
458 aminoglycosides were present, and to extract the sequences of interest for further analysis. A
459 list of the resistance gene sequences (and GenBank accession numbers) used to perform the
460 BLAST search is provided (Table 1). For acquired resistance genes, a minimum level of 90%
461 nucleotide sequence identity and gene coverage was required. Each unique *tetM* allele was
462 assigned a number (allele nucleotide sequences available at
463 https://pubmlst.org/bigsdb?db=pubmlst_cdifficile_seqdef&page=downloadAlleles).

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

464

465 **Phylogenetic Analyses**

466 Phylogenetic trees were built, based on the *C. difficile* ST11 reference M120-mapped
467 assemblies, using the maximum likelihood approach implemented in PhyML version 3.1.17
468 (using a generalized time-reversible substitution model and the "BEST" tree topology search
469 algorithm) (Guindon et al., 2010). The trees were then corrected to account for recombination
470 events using ClonalFrameML (Didelot and Wilson 2015) version 1.11 (using default
471 settings). The nodes of the trees were dated using the previously estimated *C. difficile*
472 evolutionary rate of 1.1 mutation per year (Knetsch et al., 2014) for clade 5 STs (including
473 RT078), and 1.4 mutation per year for all other genotypes (Didelot et al., 2012). The main
474 period of particular interest from 1990 to 2015 was allocated most horizontal space in
475 graphical tree representations by compressing the period pre-1990, making trees directly
476 comparable post-1990. Events before 1990 are not shown since dating older nodes using a
477 short-term evolutionary rate is problematic due to the time-dependency of evolutionary rates
478 (Biek et al., 2015). Graphical representations of trees were made using FigTree version 1.4.2
479 (Rambaut 2016).

480

481 A quantitative assessment of clonal expansion(s) within a phylogeny was performed as
482 described (Isaac et al., 2007) and implemented (Dingle et al., 2017) previously. The
483 Evolutionary Distinctiveness (ED) score of each isolate was calculated, equal to the sum, for
484 all branches on the path from the root to the leaf (isolate), of the length of the branch divided
485 by the number of leaves it supports (Isaac et al., 2007). For a given isolate, a low ED score
486 indicated the presence of close relatives in the tree, whereas a high ED score indicated their
487 relative absence. ED scores were compared across various factors using quantile regression
488 statistics, performed using Stata version 14.1 (College Station, TX, USA).

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

489 **ACKNOWLEDGEMENTS**

490 This project was funded by the National Institute for Health Research Health Protection
491 Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial
492 Resistance at Oxford University in partnership with Public Health England (PHE). The views
493 expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the
494 Department of Health or Public Health England.
495
496 This study was also supported by the UK Clinical Research Collaboration (Wellcome Trust
497 [grant 087646/Z/08/Z], Medical Research Council, National Institute for Health Research
498 [NIHR grant G0800778]); NIHR Oxford Biomedical Research Centre, NIHR Oxford Health
499 Protection Research Units on Healthcare Associated Infection and Antimicrobial Resistance
500 (grant HPRU-2012-10041) and on Modelling Methodology (grant HPRU-2012-10080), and
501 the Health Innovation Challenge Fund (a parallel funding partnership between the Wellcome
502 Trust [grant WT098615/Z/12/Z] and the Department of Health [grants WT098615 and HICF-
503 T5-358]).
504
505 The identification, isolation and analysis of the Scottish RT078 isolates was funded by a
506 Scottish Infection Research Network (SIRN) grant (ref SIRN/09), awarded to GD and CAM
507 (PIs), JC, DB and UI (co-applicants).
508
509 KED was supported by NIHR Oxford Biomedical Research Centre, DWC and TEAP are
510 NIHR senior investigators.
511
512 SB was supported as a Daphne Jackson Fellow funded by Medical Research Scotland. UZI
513 was supported by a NERC independent research fellowship (NE/LO11956/1).

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

514 The authors thank the Food Animal Initiative in Oxford for access to animals to sample for *C.*
515 *difficile* at the Oxford University Farm, Wytham.
516
517 The authors acknowledge the contribution of the Modernising Medical Microbiology
518 Informatics Group comprising: Carlos Del Ojo Elias, Charles Crichton, Vasiliki Kostiou, and
519 Adam Giess (Nuffield Department of Clinical Medicine, University of Oxford, UK), Jim
520 Davies , (Department of Computer Science, University of Oxford, UK).
521
522 The funders had no role in the writing of the manuscript or the decision to submit it for
523 publication.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

524 **AUTHOR CONTRIBUTIONS**

525 KED, XD, TEAP, ASW, and DWC conceived and designed the study; KED, DWE, NS,
526 CAM, JC, DB, SB, UZI, CG, GD, WNF, MHW, TEAP, ASW, and DWC contributed isolates
527 and DNA sequence data; XD and DWE Performed phylogenetic analysis; XD and DWE
528 provided bioinformatics support; TPQ and ASW Performed Evolutionary Distinctiveness
529 analysis; KED analysed DNA sequence data for antimicrobial resistance determinants; KED,
530 XD, DWE, MHW, TEAP, ASW, DWC wrote the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

531 **COMPETING INTERESTS**

532 MHW has received: Consulting fees from Abbott Laboratories, Actelion, Aicuris, Astellas,
533 Astra-Zeneca, Bayer, Biomèrieux, Cambimune, Cerexa, Da Volterra, The European Tissue
534 Symposium, The Medicines Company, MedImmune, Menarini, Merck, Meridian, Motif
535 Biosciences, Nabriva, Paratek, Pfizer, Qiagen, Roche, Sanofi-Pasteur, Seres, Summit,
536 Synthetic Biologics and Valneva; lecture fees from Abbott, Alere, Allergan, Astellas, Astra-
537 Zeneca, Merck, Pfizer, Roche & Seres; grant support from Abbott, Actelion, Astellas,
538 Biomèrieux, Cubist, Da Volterra, MicroPharm, Morphochem AG, Sanofi-Pasteur, Seres,
539 Summit and The European Tissue Symposium, Merck.
540 The remaining authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

541 **REFERENCES**

542 1. **Geographical clustering of cases of infection with moxifloxacin-resistant**
543 ***Clostridium difficile* PCR-ribotypes 012, 017 and 046 in Sweden, 2008 and 2009.**
544 T Akerlund, I Alefjord , U Dohnhammar, J Struwe, T Noren, K Tegmark-Wisell, Swedish C.
545 difficile Study Group (2011)
546 *Euro. Surveill.* **16** pii:19813.
547 2. **Prevalence and molecular characterization of *Clostridium difficile* isolated from**
548 **European Barn Swallows (*Hirundo rustica*) during migration.**
549 P Bandelj, T Trilar, R Blagus, M Ocepek, J Rousseau, JS Weese, M Vengust M (2014)
550 *BMC Vet. Res.* 10:40. doi: 10.1186/1746-6148-10-40.
551 3. **Prospective study of *Clostridium difficile* infections in Europe with phenotypic**
552 **and genotypic characterisation of the isolates.**
553 F Barbut, P Mastrantonio, M Delmée, J Brazier, E Kuijper, I Poxton, European Study Group
554 on *Clostridium difficile* (ESGCD) (2007)
555 *Clin. Microbiol. Infect.* **13**:1048-1057.
556 4. ***Clostridium difficile* infection in Europe: a hospital-based survey.**
557 MP Bauer, DW Notermans, BH van Benthem, JS Brazier, MH Wilcox, M Rupnik, DL
558 Monnet, JT van Dissel, EJ Kuijper, ECDIS Study Group (2011)
559 *Lancet* **377**:63-73.
560 5. **Molecular epidemiology of *Clostridium difficile* over the course of 10 years in a**
561 **tertiary care hospital.**
562 J Belmares, S Johnson, JP Parada, MM Olson, CR Clabots, KM Bettin, LR Peterson, DN
563 Gerding (2009)
564 *Clin. Infect. Dis.* **49**:1141-1147.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

565 6. **Accurate whole human genome sequencing using reversible terminator**
566 **chemistry.**
567 DR Bentley et al. (2008)
568 *Nature* **456**:53-59.
569 7. **Measurably evolving pathogens in the genomic era.**
570 R Biek, OG Pybus, JO Lloyd-Smith, X Didelot (2015)
571 *Trends Ecol. Evol.* **30**:306-313.
572 8. **Definition of the single integration site of the pathogenicity locus in *Clostridium***
573 ***difficile*.**
574 V Braun, T Hundsberger, P Leukel, M Sauerborn, C Von Eichel-Streiber (1996)
575 *Gene* **27**:29-38.
576 9. **Infection due to *C. difficile* ribotype 078: first report of cases in the Republic of**
577 **Ireland.**
578 K Burns, M Morris-Downes, WN Fawley, E Smyth, MH Wilcox, F Fitzpatrick (2010)
579 *J. Hosp. Infect.* **75**:287-291.
580 10. **Vermin on pig farms are vectors for *Clostridium difficile* PCR ribotypes 078 and**
581 **045.**
582 SA Burt, L Siemeling, EJ Kuijper, LJ Lipman (2012)
583 *Vet. Microbiol.* **160**:256-258.
584 11. **Antimicrobial susceptibility, tetracycline and erythromycin resistance genes, and**
585 **multilocus sequence typing of *Streptococcus suis* isolates from diseased pigs in China.**
586 L Chen, Y Song, Z Wei, H He, A Zhang, M Jin (2013)
587 *J. Vet. Med. Sci.* **75**:583-587.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

588 12. **Fidaxomicin versus vancomycin for infection with *Clostridium difficile* in**
589 **Europe, Canada, and the USA: a double-blind, non-inferiority, randomised controlled**
590 **trial.**
591 OA Cornely, DW Crook, R Esposito, A Poirier, MS Somero, K Weiss, P Sears, S Gorbach,
592 OPT-80-004 Clinical Study Group (2012)
593 *Lancet Infect. Dis.* **12**:281-289
594 13. **Prevalence of *Clostridium difficile* in uncooked ground meat products from**
595 **Pittsburgh, Pennsylvania.**
596 SR Curry, JW Marsh, JL Schlackman, LH Harrison (2012)
597 *Appl Environ. Microbiol.* **78**:4183-4186.
598 14. ***Streptococcus agalactiae* clones infecting humans were selected and fixed through**
599 **the extensive use of tetracycline.**
600 V Da Cunha, MR Davies, PE Douarre, I Rosinski-Chupin, I Margarit, S Spinali, T Perkins, P
601 Lechat, N Dmytruk, E Sauvage, L Ma, B Romi, M Tichit, MJ Lopez-Sanchez, S Descorps-
602 Declere, E Souche, C Buchrieser, P Trieu-Cuot, I Moszer, D Clermont, D Maione, C
603 Bouchier, DJ McMillan, J Parkhill, JL Telford, G Dougan, MJ Walker, DEVANI
604 Consortium, MTG Holden, C Poyart, P Glaser (2014)
605 *Nat. Commun.* **5**:4544.
606 15. **Underdiagnosis of *Clostridium difficile* across Europe: the European,**
607 **multicentre, prospective, biannual, point-prevalence study of *Clostridium difficile***
608 **infection in hospitalised patients with diarrhoea (EUCLID).**
609 KA Davies, CM Longshaw, GL Davis, E Bouza, F Barbut, Z Barna, M Delmée, F
610 Fitzpatrick, K Ivanova, E Kuijper, IS Macovei, S Mentula, P Mastrantonio, L von Müller, M
611 Oleastro, E Petinaki, H Pituch, T Norén, E Nováková, O Nyč, M Rupnik, D Schmid, MH
612 Wilcox (2014)

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

613 *Lancet Infect. Dis.* **14**:1208-1219.

614 16. **ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial**
615 **Genomes.**
616 X Didelot, DJ Wilson (2015)
617 *PLoS. Comput. Biol.* **11**:e1004041.

618 17. **Microevolutionary analysis of *Clostridium difficile* genomes to investigate**
619 **transmission.**
620 X Didelot, DW Eyre, M Cule, CL Ip, MA Ansari, D Griffiths, A Vaughan, L O'Connor, T
621 Golubchik, EM Batty, P Piazza, DJ Wilson, R Bowden, PJ Donnelly, KE Dingle, M Wilcox,
622 AS Walker, DW Crook, TE Peto, RM Harding (2012)
623 *Genome Biol.* **13**:R118.

624 18. **Clinical *Clostridium difficile*: clonality and pathogenicity locus diversity.**
625 KE Dingle, D Griffiths, X Didelot, J Evans, A Vaughan, M Kachrimanidou, N Stoesser, KA
626 Jolley, T Golubchik, RM Harding, TE Peto, W Fawley, AS Walker, M Wilcox, DW Crook
627 (2011)
628 *PLoS One* **6**:e19993.

629 19. **Effects of control interventions on *Clostridium difficile* infection in England: an**
630 **observational study.**
631 KE Dingle, X Didelot, TP Quan, DW Eyre, N Stoesser, T Golubchik, RM Harding, DJ
632 Wilson, D Griffiths, A Vaughan, JM Finney, DH Wyllie, SJ Oakley, WN Fawley, J Freeman,
633 K Morris, J Martin, P Howard, S Gorbach, EJC Goldstein, DM Citron, S Hopkins, R Hope,
634 AP Johnson, MH Wilcox, TEA Peto, AS Walker, DW Crook DW, Modernising Medical
635 Microbiology Informatics Group (2017)
636 *Lancet Infect. Dis.* **17**:411-421.

637 20. **Sales of veterinary antimicrobial agents in 29 European countries in 2014.**

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

638 European Medicines Agency (2016)
639 http://www.ema.europa.eu/docs/en_GB/document_library/Report/2016/10/WC500214217.pd
640 f Accessed 26th January 2018.
641 21. **Diverse sources of *C. difficile* infection identified on whole-genome sequencing.**
642 DW Eyre, ML Cule ML, DJ Wilson, D Griffiths, A Vaughan, L O'Connor, CLC Ip, T
643 Golubchik, EM Batty, JM Finney, DH Wyllie, X Didelot, P Piazza, R Bowden, KE Dingle,
644 RM Harding, DW Crook, MH Wilcox, TEA Peto, AS Walker (2013)
645 *N. Engl. J. Med.* **369**:1195-1205.
646 22. **Enhanced surveillance of *Clostridium difficile* infection occurring outside**
647 **hospital, England, 2011 to 2013.**
648 WN Fawley, KA Davies, T Morris, P Parnell, R Howe, MH Wilcox, *Clostridium difficile*
649 Ribotyping Network (CDRN) Working Group (2016)
650 *Euro Surveill* **21**(29). doi: 10.2807/1560-7917.ES.2016.21.29.30295.
651 23. **The Pan-European Longitudinal Surveillance of Antibiotic Resistance among**
652 **Prevalent *Clostridium difficile* Ribotypes' Study Group. Pan-European longitudinal**
653 **surveillance of antibiotic resistance among prevalent *Clostridium difficile* ribotypes.**
654 J Freeman, J Vernon, K Morris, S Nicholson, S Todhunter, C Longshaw, MH Wilcox, Pan-
655 European Longitudinal Surveillance of Antibiotic Resistance among Prevalent *Clostridium*
656 *difficile* Ribotypes Study Group (2015)
657 *Clin Microbiol. Infect.* **21**:248.e9-248.e16.
658 24. **VelvetOptimiser, 2.1.7.**
659 S Gladman, T Seemann (2008)
660 Monash University, Victoria, Australia.
661 25. **Emergence of *Clostridium difficile* infection due to a new hypervirulent strain,**
662 **polymerase chain reaction ribotype 078.**

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

663 A Goorhuis, D Bakker, J Corver, SB Debast, C Harmanus, DW Notermans, AA Bergwerff,
664 FW Dekker, EJ Kuijper (2008)
665 *Clin. Infect. Dis.* **47**:1162-1170.
666 26. **New algorithms and methods to estimate maximum-likelihood phylogenies:
667 assessing the performance of PhyML3.0.**
668 S Guindon, JF Dufayard, V Lefort, M Anisimova, W Hordijk, O Gascuel (2010)
669 *Syst. Biol.* **59**:307-321.
670 27. **Evolutionary dynamics of *Clostridium difficile* over short and long time scales.**
671 M He, M Sebaihia, TD Lawley, RA Stabler, LF Dawson, MJ Martin, KE Holt, HM Seth-
672 Smith, MA Quail, R Rance, K Brooks, C Churcher, D Harris, SD Bentley, C Burrows, L
673 Clark, C Corton, V Murray, G Rose, S Thurston, A van Tonder, D Walker, BW Wren, G
674 Dougan, J Parkhill (2010)
675 *Proc. Natl. Acad. Sci. U S A* **107**:7527-7532.
676 28. **Emergence and global spread of epidemic healthcare-associated *Clostridium*
677 *difficile*.**
678 M He, F Miyajima, P Roberts, L Ellison, DJ Pickard, MJ Martin, TR Connor, SR Harris, D
679 Fairley, KB Bamford, S D'Arc, J Brazier, D Brown, JE Coia, G Douce, D Gerding, HJ Kim,
680 TH Koh, H Kato, M Senoh, T Louie, S Michell, E Butt, SJ Peacock, NM Brown, T Riley, G
681 Songer, M Wilcox, M Pirmohamed, E Kuijper, P Hawkey, BW Wren, G Dougan, J Parkhill,
682 TD Lawley (2013)
683 *Nat. Genet.* **45**:109-113.
684 29. **Healthcare Associated Infection, Annual Report 2016.**
685 Health Protection Scotland (2017)
686 http://www.hps.scot.nhs.uk/resourcedocument.aspx?id=5934 Accessed 26th January 2018.
687 30. ***Clostridium difficile* infection in the community: a zoonotic disease?**

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

688 MP Hensgens, EC Keessen, MM Squire, TV Riley, MG Koene, E de Boer, LJ Lipman, EJ
689 Kuijper, European Society of Clinical Microbiology and Infectious Diseases Study Group for
690 *Clostridium difficile* (ESGCD) (2012)
691 *Clin. Microbiol. Infect.* **18**:635-645.
692 31. **The antimicrobial resistance patterns and associated determinants in**
693 ***Streptococcus suis* isolated from humans in southern Vietnam, 1997-2008.**
694 NT Hoa, TT Chieu, HD Nghia, NT Mai, PH Anh, M Wolbers, S Bake, JI Campbell, NV
695 Chau NV, TT Hien, J Farrar, C Schultsz (2011)
696 *BMC Infect. Dis*:**11**:6.
697 32. **Distinct ribotypes and rates of antimicrobial drug resistance in *Clostridium***
698 ***difficile* from Shanghai and Stockholm.**
699 H Huang, H Fang, A Weintraub, CE Nord (2009)
700 *Clin. Microbiol. Infect.* **15**:1170-1173.
701 33. **Mammals on the EDGE: Conservation priorities based on threat and phylogeny.**
702 NJB Isaac, ST Turvey, B Collen, C Waterman, JEM Baillie (2007)
703 *PLoS One* **2**:e296.
704 34. **Toxinotype V Clostridium difficile in humans and food animals.**
705 MA Jhung, AD Thompson, GE Killgore, WE Zukowski, G Songer, M Warny, S Johnson,
706 DN Gerding, LC McDonald, BM Limbago (2008)
707 Emerg Infect Dis **14**:1039-1045.
708 35. **Epidemics of diarrhea caused by a clindamycin-resistant strain of *Clostridium***
709 ***difficile* in four hospitals.**
710 S Johnson, MH Samore, KA Farrow, GE Killgore, FC Tenover, D Lyras, JI Rood, P
711 DeGirolami, AL Baltch, ME Rafferty, SM Pear, DN Gerding (1999)
712 *N. Engl. J. Med.* **341**:1645-1651.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

713 36. **BIGSdb: Scalable analysis of bacterial genome variation at the population level.**
714 KA Jolley, MC Maiden (2010)
715 *BMC Bioinformatics* **11**:595.
716 37. **Prevalence of PCR ribotypes among *Clostridium difficile* isolates from pigs,**
717 **calves, and other species.**
718 K Keel, JS Brazier, KW Post, S Weese, JG Songer (2007)
719 *J. Clin. Microbiol.* **45**:1963-1964.
720 38. **Whole genome sequencing reveals potential spread of *Clostridium difficile***
721 **between humans and farm animals in the Netherlands, 2002 to 2011.**
722 CW Knetsch, TR Connor, A Mutreja, SM van Dorp, IM Sanders, HP Browne, D Harris, L
723 Lipman, EC Keessen, J Corver, EJ Kuijper, TD Lawley (2014)
724 *Euro Surveill* **19**:20954.
725 39. **Copresence of *tet*(K) and *tet*(M) in livestock-associated methicillin-resistant**
726 ***Staphylococcus aureus* clonal complex 398 is associated with increased fitness during**
727 **exposure to sublethal concentrations of tetracycline.**
728 J Larsen, J Clasen, JE Hansen, W Paulander, A Petersen, AR Larsen, D Frees (2016)
729 *Antimicrob. Agents Chemother.* **60**:4401-4403.
730 40. **Burden of *Clostridium difficile* infection in the United States.**
731 FC Lessa, LG Winston, LC McDonald, Emerging Infections Program *C. difficile*
732 Surveillance Team (2015)
733 *N. Engl. J. Med.* **372**:825-834.
734 41. **A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-**
735 **associated diarrhea with high morbidity and mortality.**

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

736 VG Loo, L Poirier, MA Miller, M Oughton, MD Libman, S Michaud, AM Bourgault, T
737 Nguyen, C Frenette, M Kelly, A Vibien, P Brassard, S Fenn, K Dewar, TJ Hudson, R Horn, P
738 René, Y Monczak, A. Dascal (2005)
739 *N. Engl. J. Med.* **353**:2442-2249.
740 42. **Fidaxomicin versus vancomycin for *Clostridium difficile* infection.**
741 TJ Louie, MA Miller, KM Mullane, K Weiss, A Lentnek, Y Golan, S Gorbach, P Sears, YK
742 Shue, OPT-80-003 Clinical Study Group (2011)
743 *N. Engl. J. Med.* **364**:422-431.
744 43. **A fatal case of streptococcal toxic shock syndrome caused by *Streptococcus suis***
745 **carrying tet (40) and tet (O/W/32/O), Italy.**
746 F Mancini, F Adamo, R Creti, M Monaco, G Alfarone, A Pantosti, A Ciervo (2016)
747 *J. Infect. Chemother.* **22**:774-776.
748 44. **Investigation of community carriage rates of *Clostridium difficile* and *Hungatella***
749 ***hathewayi* in healthy volunteers from four regions of England.**
750 SE Manzoor, CAM McNulty, D Nakiboneka-Ssenabulya, DM Lecky, KJ Hardy, PM Hawkey
751 (2017)
752 *J Hosp Infect*. **97**:153-155.45.**Thirty-day mortality of *Clostridium difficile* infection in a**
753 **UK National Health Service Foundation Trust between 2002 and 2008.**
754 AP McGowan, LC Lalayiannis, JB Sarma, B Marshall, KE Martin, MR Welfare (2011)
755 *J. Hosp. Infect*. **77**:11-5.
756 46. **Hypervirulent *Clostridium difficile* strains in hospitalized patients, Canada.**
757 MR Mulvey, DA Boyd, D Gravel, J Hutchinson, S Kelly, A McGeer, D Moore, A Simor, KN
758 Suh, G Taylor, JS Weese, M Miller, Canadian Nosocomial Infection Surveillance Program
759 (2010)
760 *Emerg. Infect. Dis.* **16**:678-681.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

761 47. **Tetracycline antibiotics and resistance mechanisms.**
762 F Nguyen, AL Starosta, S Arenz, D Sohmen, A Dönhöfer, DN Wilson. (2014)
763 Biol Chem **395**:559-575.
764 48. ***In vitro* susceptibility to 17 antimicrobials among clinical *Clostridium difficile***
765 **isolates collected 1993-2007 in Sweden.**
766 T Noren, I Alriksson, T Akerlund, LG Burman, M Unemo (2009)
767 *Clin Microbiol Infect* **16**:1104-1110.
768 49. **One Health Initiative will unite human and veterinary medicine.**
769 One Health Initiative (2018)
770 http://www.onehealthinitiative.com/index.php Accessed 26th January 2018.
771 50. **Differences in outcome according to *C. difficile* testing method: a prospective**
772 **multicentre diagnostic validation study of *C. difficile* infection.**
773 TD Planche, KA Davies, PG Coen, JM Finney, IM Monahan, KA Morris, L O'Connor, SJ
774 Oakley, CF Pope, MW Wren, NP Shetty, DW Crook, MH Wilcox (2013)
775 *Lancet Infect. Dis.* **13**:936-945.
776 51. **Characterization of swine isolates of *Clostridium difficile* in Spain: a potential**
777 **source of epidemic multidrug resistant strains?**
778 T Peláez, L Alcalá, JL Blanco, S Álvarez-Pérez, M Marín, A Martín-López, P Catalán, E
779 Reigadas, ME García, E Bouza (2013)
780 *Anaerobe* **22**:45-49.
781 52. **English surveillance programme for antimicrobial utilisation and resistance**
782 **(ESPAUR) Report 2014**
783 Public Health England **(2014)**
784 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/362374/ESPA
785 UR_Report_2014__3_.pdf Accessed 26th January 2018.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

786 53. ***Clostridium difficile* Ribotyping Network (CDRN) for England and Northern**
787 **Ireland. Biennial Report (2013-2015).**
788 Public Health England. (2016)
789 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/491253/CDRN
790 _2013-15_Report.pdf Accessed 26th January 2018.
791 54. **Figtree**
792 Rambaut, A. (2016)
793 Figtree http://tree.bio.ed.ac.uk/software/figtree/ Accessed 26th January 2018.
794 55. **Official Journal of the European Communities; Council Directive of 23**
795 **November 1970 concerning additives in feed-stuffs.**
796 Scheel W (1970) https://publications.europa.eu/en/publication-detail/-/publication/de2f6bfa-
797 e0e7-4383-af1f-3406fa00a4f1/language-en Accessed 26th January 2018.
798 56. **Antimicrobial susceptibilities of *Clostridium difficile*.**
799 R Shuttleworth, M Taylor, DM Jones (1980)
800 *J Clin Pathol* **33**:1002-1005.
801 57. ***Clostridium difficile* in retail meat products, USA, 2007.**
802 JG Songer, HT Trinh, GE Killgore, AD Thompson, LC McDonald, BM Limbago (2007)
803 *Emerg. Infect. Dis.* **15**:819-821
804 58. **Epidemiology of *Clostridium difficile* in infants in Oxfordshire, UK: risk factors**
805 **for colonization and carriage, and genetic overlap with regional *C. difficile* infection**
806 **strains.**
807 N Stoesser, DW Eyre, TP Quan, H Godwin, G Pill, E Mbuvi, A Vaughan, D Griffiths, J
808 Martin, W Fawley, KE Dingle, S Oakley, K Wanelik, JM Finney, M Kachrimanidou, CE
809 Moore, S Gorbach, TV Riley, DW Crook, TEA Peto, MH Wilcox, AS Walker, Modernising
810 Medical Microbiology Informatics Group (MMMIG) (2017)

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

811 *PLoS One* **12**:e0182307.
812 59. ***Clostridium difficile* isolates resistant to fluoroquinolones in Italy: emergence of**
813 **PCR ribotype 018.**
814 P Spigaglia, F Barbanti, AM Dionisi, P Mastrantonio (2010)
815 *J. Clin. Microbiol.* **48**:2892-2896.
816 60. **Recent advances in the understanding of antibiotic resistance in *Clostridium***
817 ***difficile* infection.**
818 Spigaglia P (2016)
819 *Ther. Adv. Infect. Dis.* **3**:23-42.
820 61. **A prospective study of community-associated *Clostridium difficile* infections: the**
821 **role of antibiotics and co-infections.**
822 SK Taori, A Wroe, A Hardie, AP Gibb, IR Poxton (2014)
823 *J. Infect.* **69**:134-144.
824 62. **The tetracycline resistome.**
825 M Thaker, P Spanogiannopoulos, GD Wright (2010)
826 *Cell. Mol. Life Sci.* **67**:419-431.
827 63. **Center for Veterinary Medicine Summary report on Antimicrobials Sold or**
828 **Distributed for Use in Food-Producing Animals.**
829 US Food and Drug Administration (2016)
830 https://www.fda.gov/downloads/ForIndustry/UserFees/AnimalDrugUserFeeActADUFA/UC
831 M534243.pdf Accessed 26th January 2018.
832 64. **UK Veterinary Antibiotic Resistance and Sales Surveillance Report.**
833 Veterinary Medicines Directorate (2015)
834 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/582341/10517
835 28-v53-UK-VARSS_2015.pdf Accessed 26th January 2018.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

836 65. **Relationship between bacterial strain type, host biomarkers, and mortality in**
837 ***Clostridium difficile*** **infection.**
838 AS Walker, DW Eyre, DH Wyllie, KE Dingle, D Griffiths, B Shine, S Oakley, L O'Connor, J
839 Finney, A Vaughan, DW Crook, MH Wilcox, TE Peto, Infections in Oxfordshire Research
840 Database (2013)
841 *Clin. Infect. Dis.* **56**:1589-1600.
842 66. **Mosaic tetracycline resistance genes encoding ribosomal protection proteins.**
843 PJ Warburton, N Amodeo, AP Roberts (2016)
844 *J Antimicrob Chemother* **71**:3333-3339.
845 67. ***Streptococcus suis*****: an emerging human pathogen.**
846 HF Wertheim, HD Nghia, W Taylor, C Schultsz (2009)
847 *Clin. Infect. Dis.* **48**:617-625.
848 68. **Genomic signatures of human and animal disease in the zoonotic pathogen**
849 ***Streptococcus suis*****.**
850 LA Weinert, RR Chaudhuri, J Wang, SE Peters, J Corander, T Jombart, A Baig, KJ Howell,
851 M Vehkala, N Välimäki, D Harris, TT Chieu, N Van Vinh Chau, J Campbell, C Schultsz, J
852 Parkhill, SD Bentley, PR Langford, AN Rycroft, BW Wren, J Farrar, S Baker, NT Hoa, MT
853 Holden, AW Tucker, DJ Maskell; BRaDP1T Consortium (2015)
854 *Nat. Commun.* **6**:6740. doi: 10.1038/ncomms7740.
855 69. **Identification of a new ribosomal protection type of tetracycline resistance gene,**
856 **tet(36), from swine manure pits.**
857 G Whittle, TR Whitehead, N Hamburger, NB Shoemaker, MA Cotta, AA Salyers (2003)
858 *Appl Environ Microbiol* **69**:4151-4158.
859 70. **A case-control study of community-associated** ***Clostridium difficile*** **infection.**
860 MH Wilcox, L Mooney, R Bendall, CD Settle, WN Fawley (2008)

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

861 *J. Antimicrob. Chemother.* **62**:388-396.

862 71. **Changing epidemiology of *Clostridium difficile* infection following the**
863 **introduction of a national ribotyping-based surveillance scheme in England.**
864 MH Wilcox, N Shetty, WN Fawley, M Shemko, P Coen, A Birtles, M Cairns, MD Curran,
865 KJ Dodgson, SM Green, KJ Hardy, PM Hawkey, JG Magee, AD Sails, MW Wren (2012)
866 *Clin. Infect. Dis.* **55**:1056-1063.

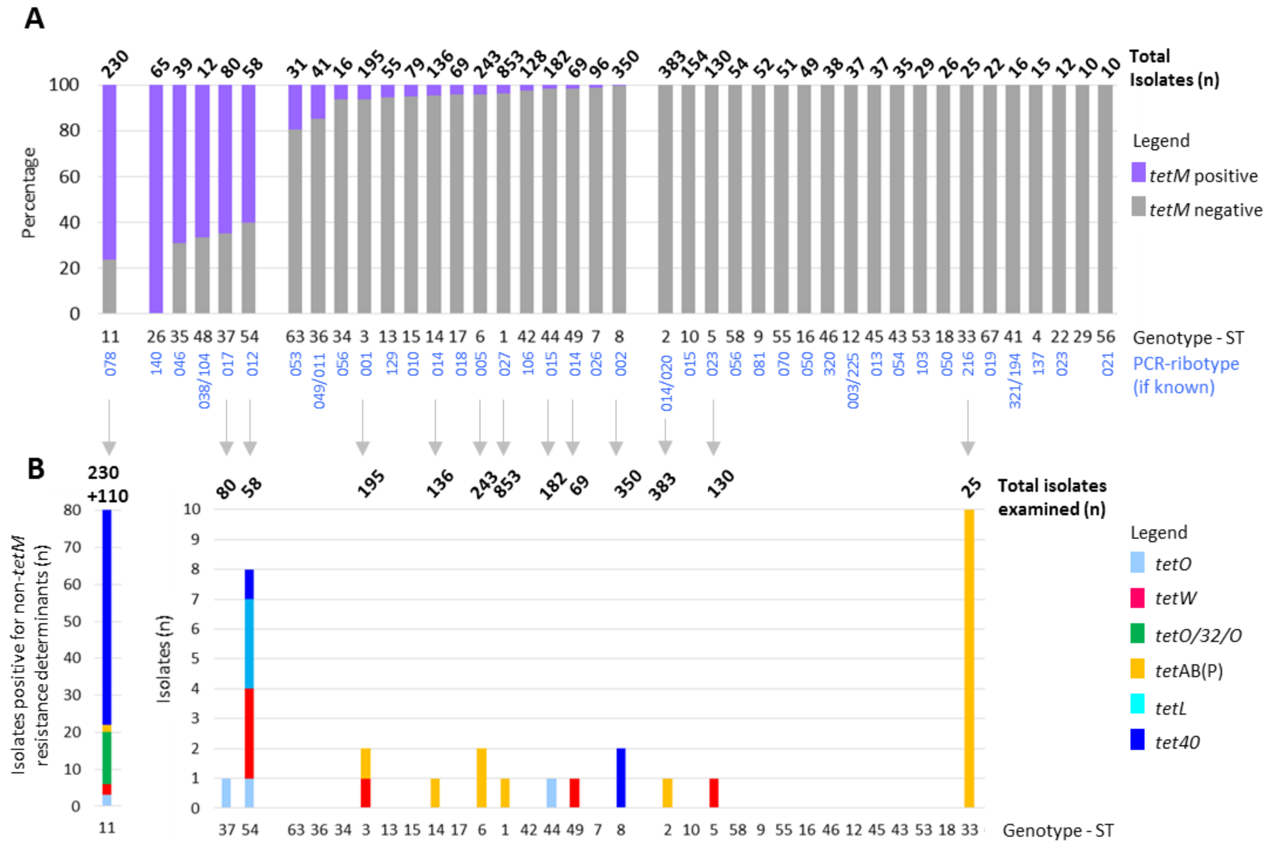
867 72. **Fate of *Clostridium difficile* during wastewater treatment and incidence in**
868 **Southern Ontario watersheds.**
869 C Xu, JS Weese, C Flemming, J Odumeru, K Warriner (2014)
870 *J. Appl. Microbiol.* **117**:891-904.

871 73. **Velvet: algorithms for de-novo short read assembly using de Bruijn graphs.**
872 DR Zerbino, E Birney (2008)
873 *Genome Res.* **18**:821-829.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

874 **FIGURE LEGENDS**

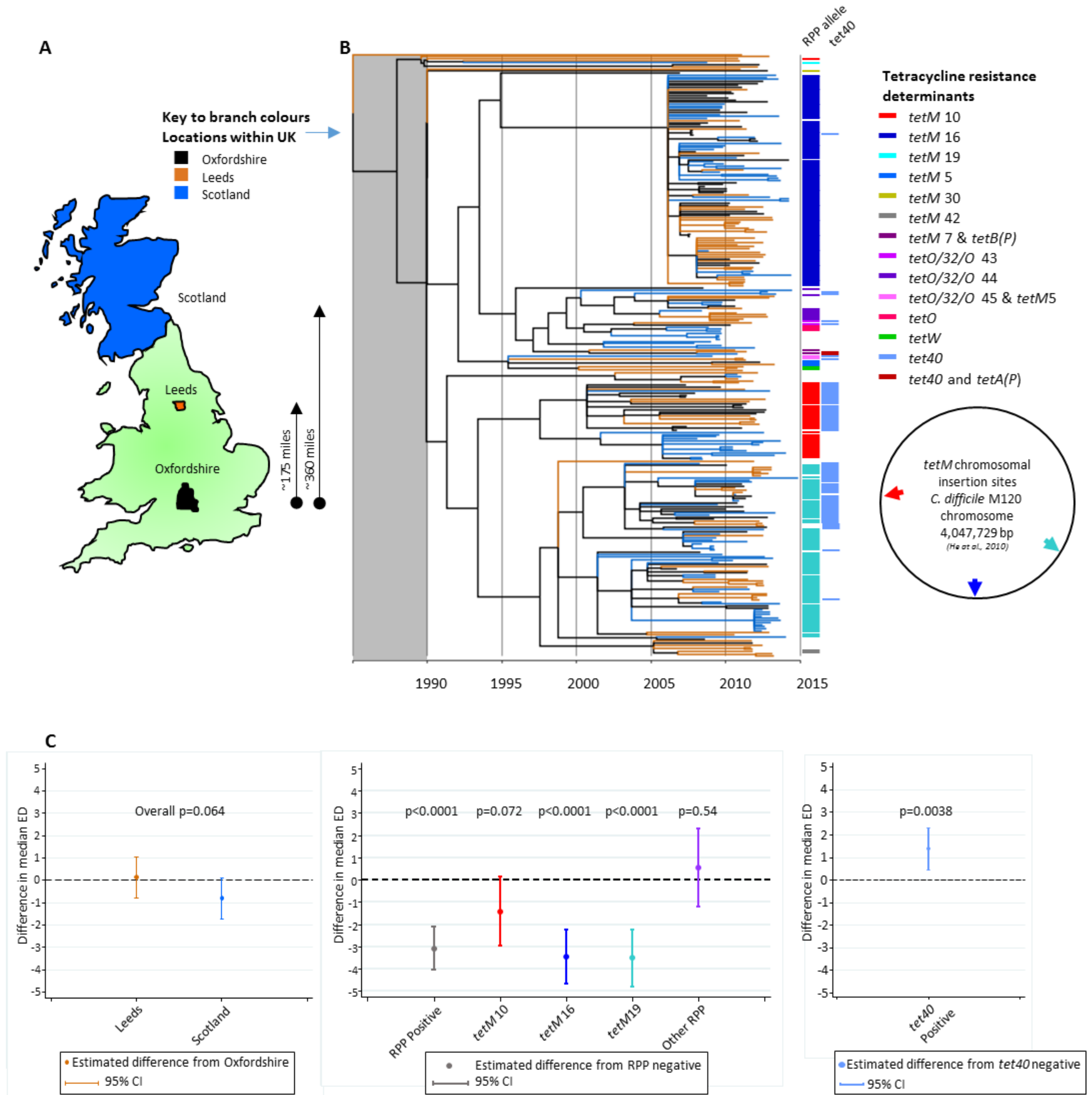
875 **Figure 1**
876 **Prevalence of tetracycline resistance determinants in RT078 and other clinically**
877 **relevant *C. difficile* genotypes.**
878



879
880
881
882
883 **(A) The proportion (%) of each clinically important genotype that was positive for the**
884 **ribosomal protection protein (RPP) gene *tetM*.** Data are shown for genotypes having 10
885 genomes or more, from isolate collections representing Oxfordshire (EIA positives,
886 negatives, infant and farm), Leeds, North America and Europe (Optimer clinical trial)
887 (Dingle et al., 2017; Stoesser et al., 2017; Knetsch et al., 2014; Louie et al., 2011; Cornely et
888 al., 2012). The total number of isolates of each genotype is shown above the bar.
889
890 **(B) The numbers of genomes in the collections above which contained additional non-**
891 ***tetM* tetracycline resistance determinants.** For the ST11(078) genotype, the additional
892 Scottish (n=110) isolate collection was also included (indicated by +110 above the bar).
893 Therefore, the total ST11(078) isolates examined was 340 isolates; (the n=230 included in
894 (A) above; plus additional Scottish ST11s (n=110)) the aim being to illustrate the overall
895 prevalence of 'non-*tetM*' tetracycline resistance determinants within this genotype.
896

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

897 **Figure 2**
898 **UK-representative, time-scaled RT078 phylogeny revealing a lack of geographic**
899 **structure, but strong structuring of tetracycline resistance.**



900

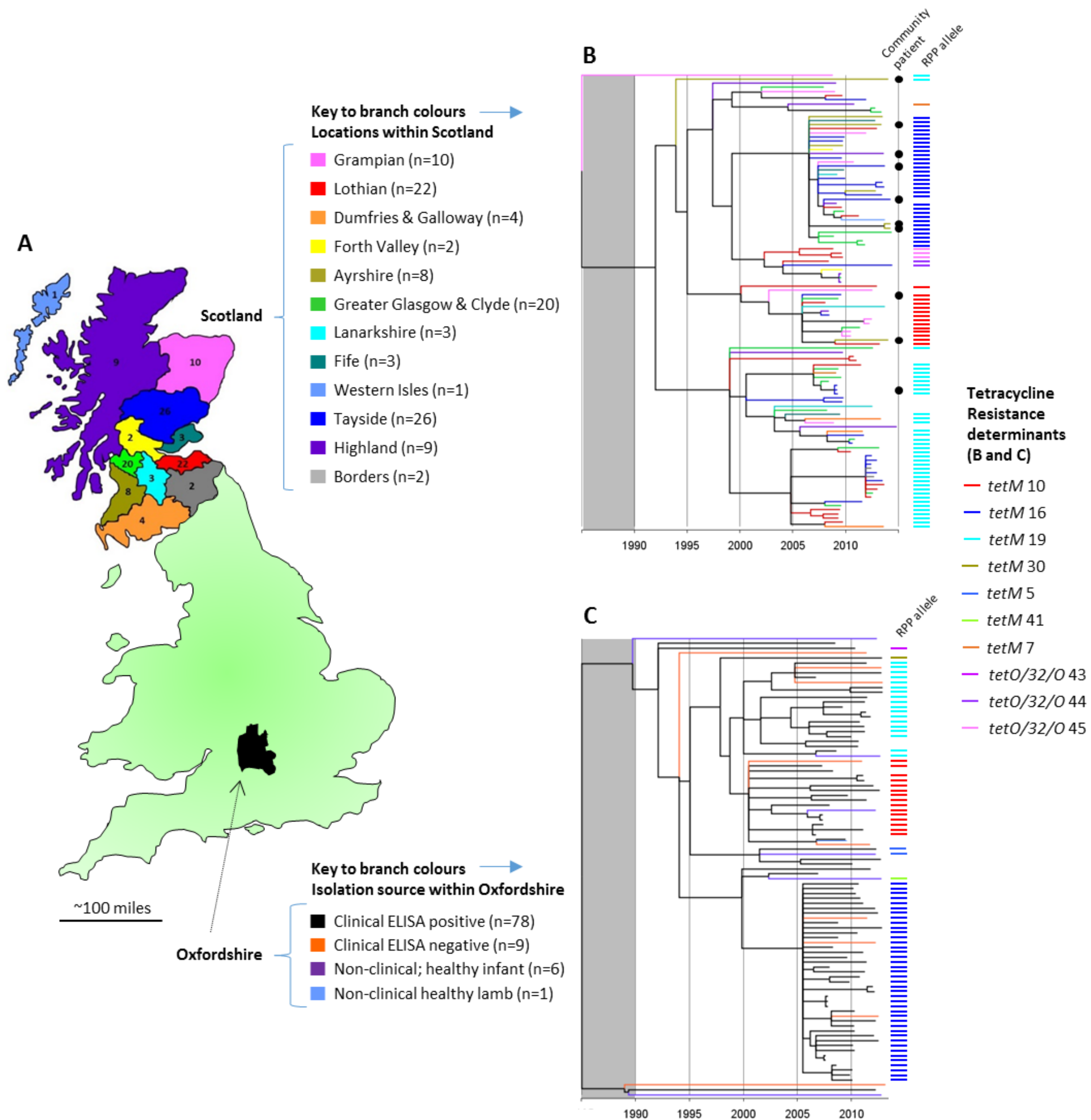
901 **(A)** Map showing areas of the UK from which the RT078 *C. difficile* genomes were obtained.
902
903 **(B)** Time-Scaled ClonalFrameML phylogeny constructed using genomes from UK *C. difficile*
904 isolates; Oxfordshire (n=78), Leeds (n=104), and Scottish (n=110). Branch colours, as per
905 (A), denote the location of each genome. Coloured bars to the right of the phylogeny indicate
906 the presence of tetracycline resistance determinants; ribosomal protection protein allele
907 sequences detected within each genome, were assigned numbers to identify distinct
908 nucleotide sequences of *tetM*, *tetO/32/O*, *tetO* or *tetW*. To the right of the phylogeny, the
909 chromosomal location of the three most prevalent *tetM* alleles (designated *tetM* 10, 16 and

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

910 19) relative to the RT078 M120 genome (NCBI Reference Sequence NC_017174.1) are
911 shown.
912 All phylogenies included in this study are directly comparable post 1990 ie the timeframe of
913 RT078 emergence; the grey shaded block over the region prior to this date indicates the
914 region which is not scaled identically and should not be compared.
915
916 **(C)** Extent to which RT078 clonal expansions are associated with geographic structure and
917 tetracycline resistance (ribosomal protection proteins and efflux pumps), using two-sided
918 quantile regression.
919 (i) Difference in median Evolutionary Distinctiveness score compared to Oxfordshire
920 samples. A lower Evolutionary Distinctiveness value indicates a larger proportion of close
921 relatives in the tree. The p-value measures the overall significance of geographic location on
922 Evolutionary Distinctiveness score.
923 (ii) Difference in median Evolutionary Distinctiveness score for samples with ribosomal
924 protection proteins detected compared to ribosomal protection protein-negative samples,
925 overall and for each of the three putative *tetM*-associated clonal expansions. A lower
926 Evolutionary Distinctiveness value indicates a larger proportion of close relatives in the tree.
927 The p-values measure the significance of gene presence on Evolutionary Distinctiveness
928 score.
929 (iii) Difference in median Evolutionary Distinctiveness score for samples with tetracycline
930 efflux pumps (*tet40* and *tetA*(*P*)) detected compared to efflux pump negative samples. A
931 lower Evolutionary Distinctiveness value indicates a larger proportion of close relatives in the
932 tree. The p-value measures the significance of gene presence on Evolutionary Distinctiveness
933 score.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

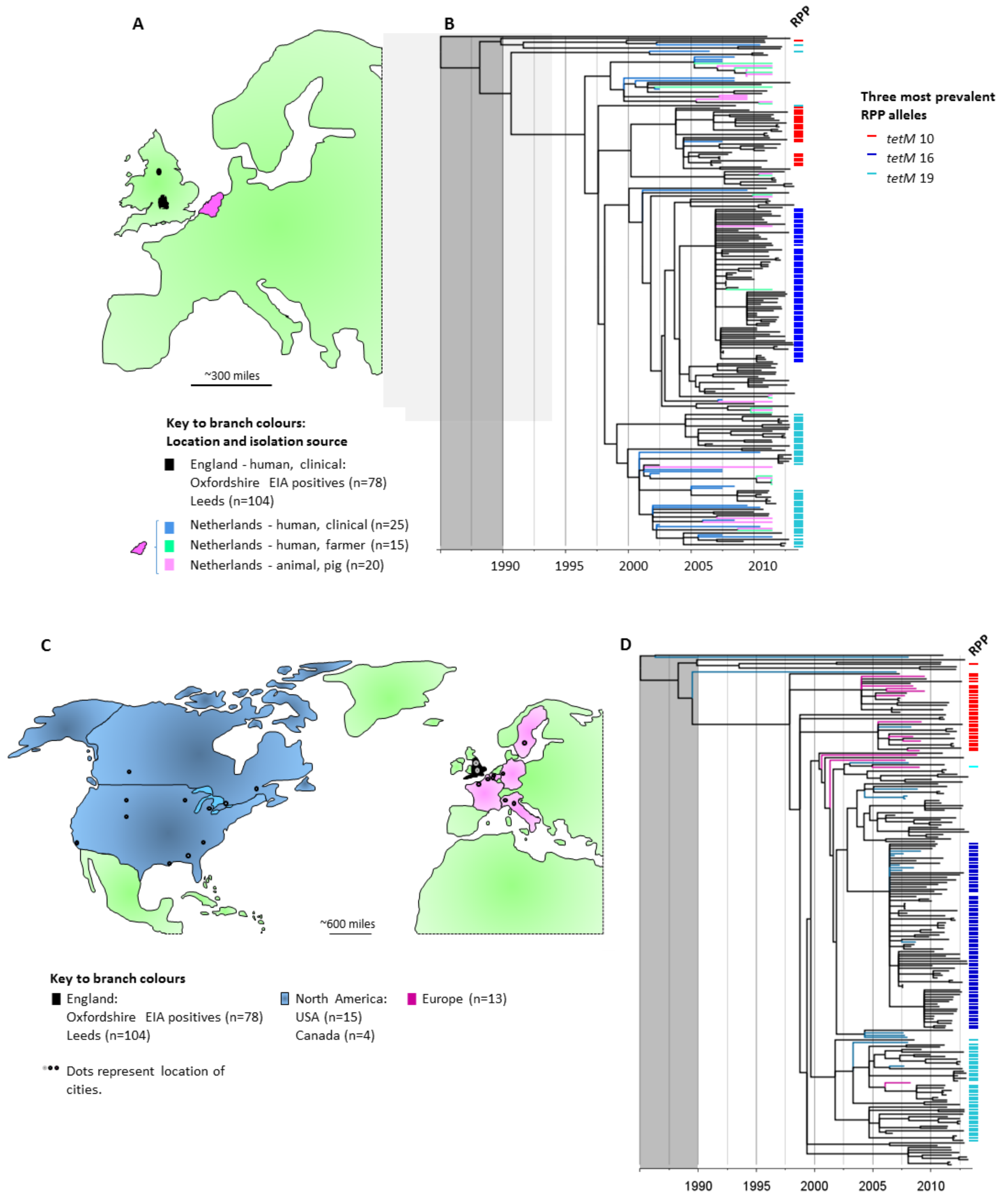
934 **Figure 3**
935 **UK Regional RT078 Phylogenies; Scotland and Oxfordshire**



936
937
938 **(A)** Map and legend indicating the regions of Scotland and Oxfordshire from which genomes
939 originate. Scottish regions correspond to administrative areas known as 'health boards'.
940
941 **(B)** Time-Scaled RT078 phylogeny for Scotland. Branch colours, as per map (A). Coloured
942 bars to the right of the phylogeny denote the ribosomal protection protein (RPP) allele
943 sequences detected within each genome (as Figure 2), numbers being assigned to identify
944 distinct nucleotide sequences of *tetM* or *tetO/32/O*. Isolates were cultured from human
945 clinical samples received from both hospital and community patients, the latter indicated by a
946 black dot. The grey shaded block over the region prior to 1990 indicates the region which is
947 not scaled identically for different phylogenies and should not be compared.
948
949 **(C)** Time-Scaled RT078 phylogeny for Oxfordshire clinical and non-clinical isolates. Branch
950 colours as per map (A). Coloured bars indicate ribosomal protection protein alleles as above.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

951 **Figure 4**
952 **International phylogenies confirm three major *tetM* positive RT078 clades are present**
953 **across continents.**



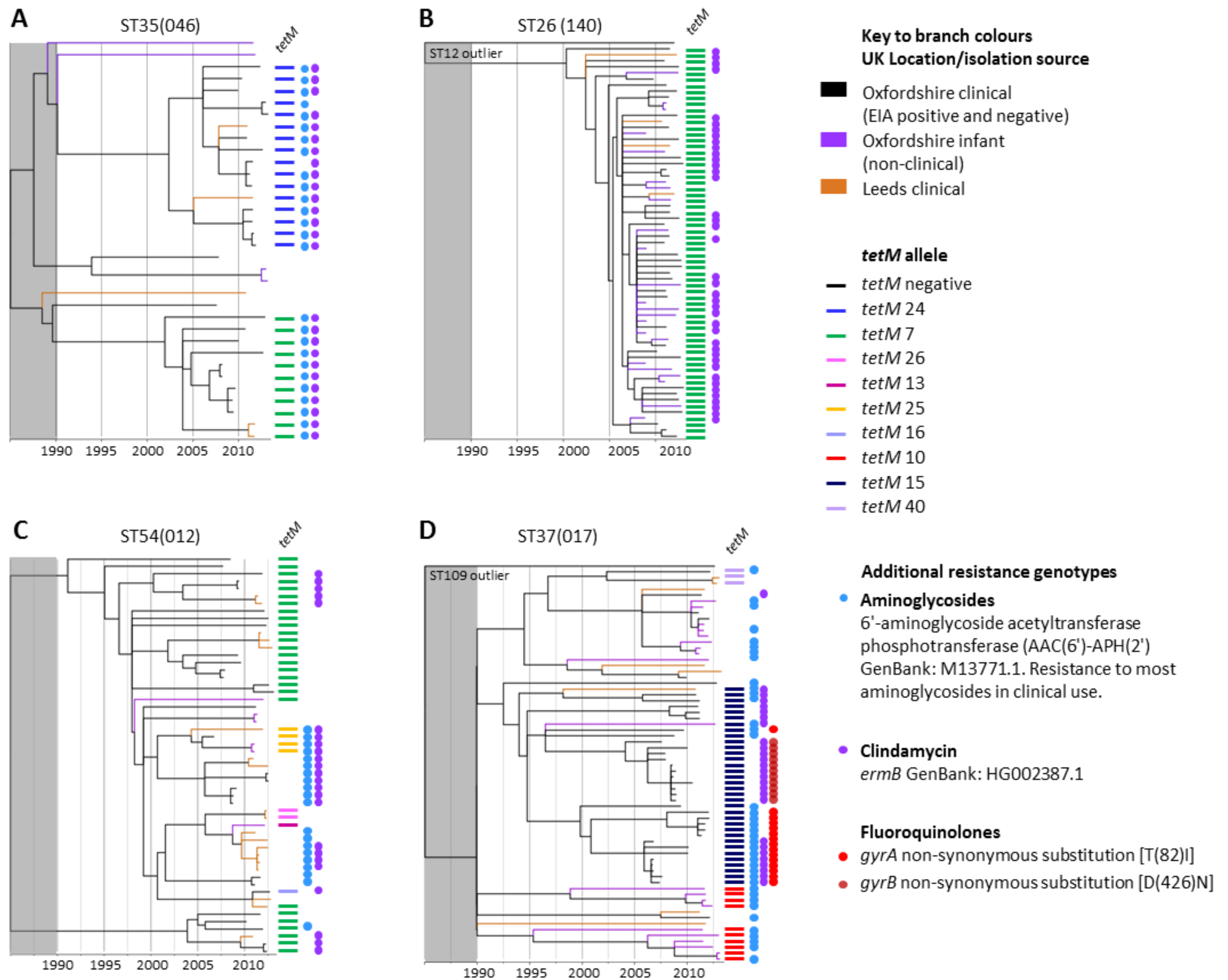
954
955 **(A)** Map of Western Europe; regions of England and The Netherlands, from which the
956 genomes included in (B) originate are highlighted (black and pink respectively).
957

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

958 **(B)** Time-Scaled RT078 phylogeny constructed using genomes of clinical isolates from
959 England (Oxfordshire and Leeds), supplemented with genomes from the Netherlands (human
960 clinical, farmer and pig isolates (Knetsch et al., 2014). Branch colours as per map (A). The
961 presence of the three predominant ribosomal protection protein (RPP) *tetM* alleles (*tetM* 10,
962 16 and 19) is indicated by the coloured bars to the right of the tree. The grey shaded block
963 over the region prior to 1990 indicates the region which is not scaled identically for different
964 phylogenies and should not be compared.
965
966 **(C)** Map highlighting North America and Western Europe, from which the genomes included
967 in (D) originate.
968
969 **(D)** Time -Scaled RT078 phylogeny constructed using genomes of clinical isolates from
970 England supplemented with clinical isolates from North America and Europe (distinct from
971 the isolates used above in (B), from two clinical trials of the drug fidaxomicin (Table S1)
972 (Louie et al., 2011; Cornely et al., 2012).
973

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

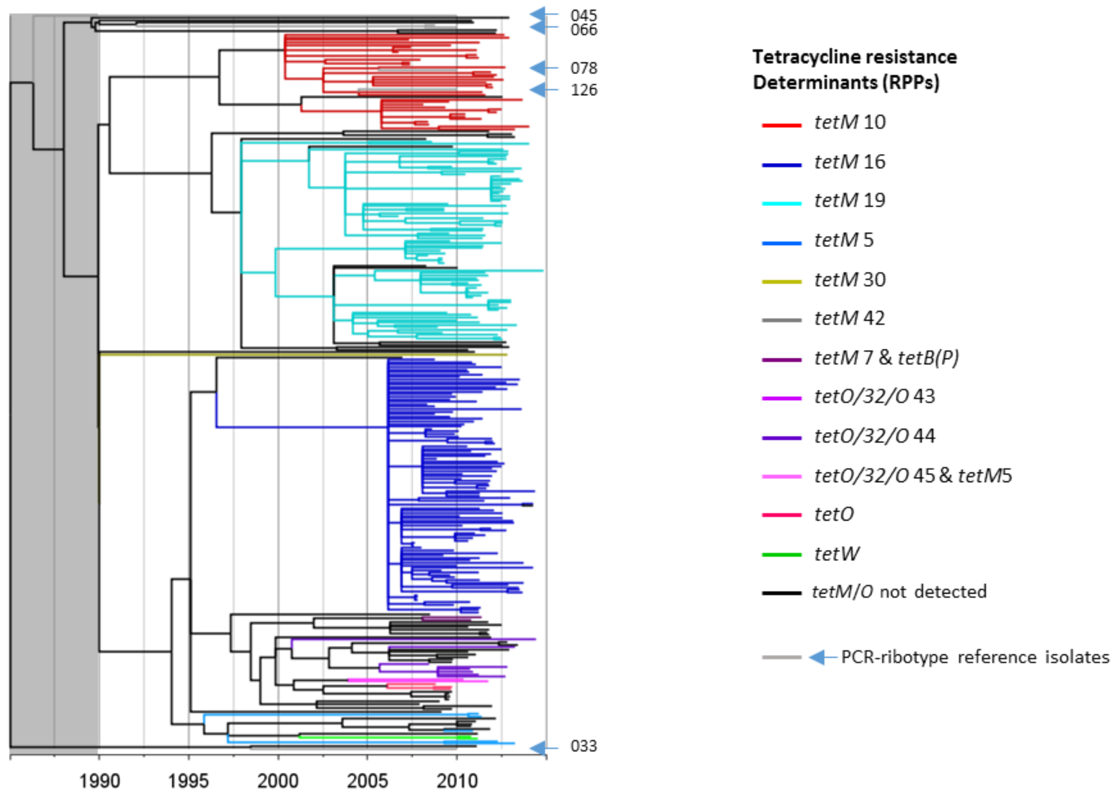
974 **Figure 5**
975 **Phylogenetic analysis of additional *tetM* positive *C. difficile* genotypes**



976
977
978 Time-scaled phylogenies were constructed representing four non-RT078 genotypes with
979 >60% *tetM* prevalence; **(A)** ST35(046), **(B)** ST26(140), **(C)** ST54(012) and **(D)** ST37(017).
980
981 In phylogenies (B) and (D) a single closely related genome of a distinct genotype (ST12 and
982 ST109 respectively) was included to ensure that the tree was rooted pre-1990 and that the
983 four phylogenies could therefore be compared post 1990. The grey shaded block over the
984 region prior to 1990 indicates the region which is not scaled identically for different
985 phylogenies and should not be compared. Genomes were from Oxfordshire (clinical EIA
986 positives and negatives plus non-clinical, healthy infants) and Leeds (clinical isolates);
987 branch colours indicate location/isolation source as before. Coloured bars to the right of each
988 phylogeny indicate the presence of tetracycline resistance determinants. Coloured dots
989 represent additional genetic determinants identified, conferring resistance to
990 fluoroquinolones, rifampicin, clindamycin and aminoglycosides (Table 1), which are shown
991 if five or more positive genomes were identified per genotype.
992

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

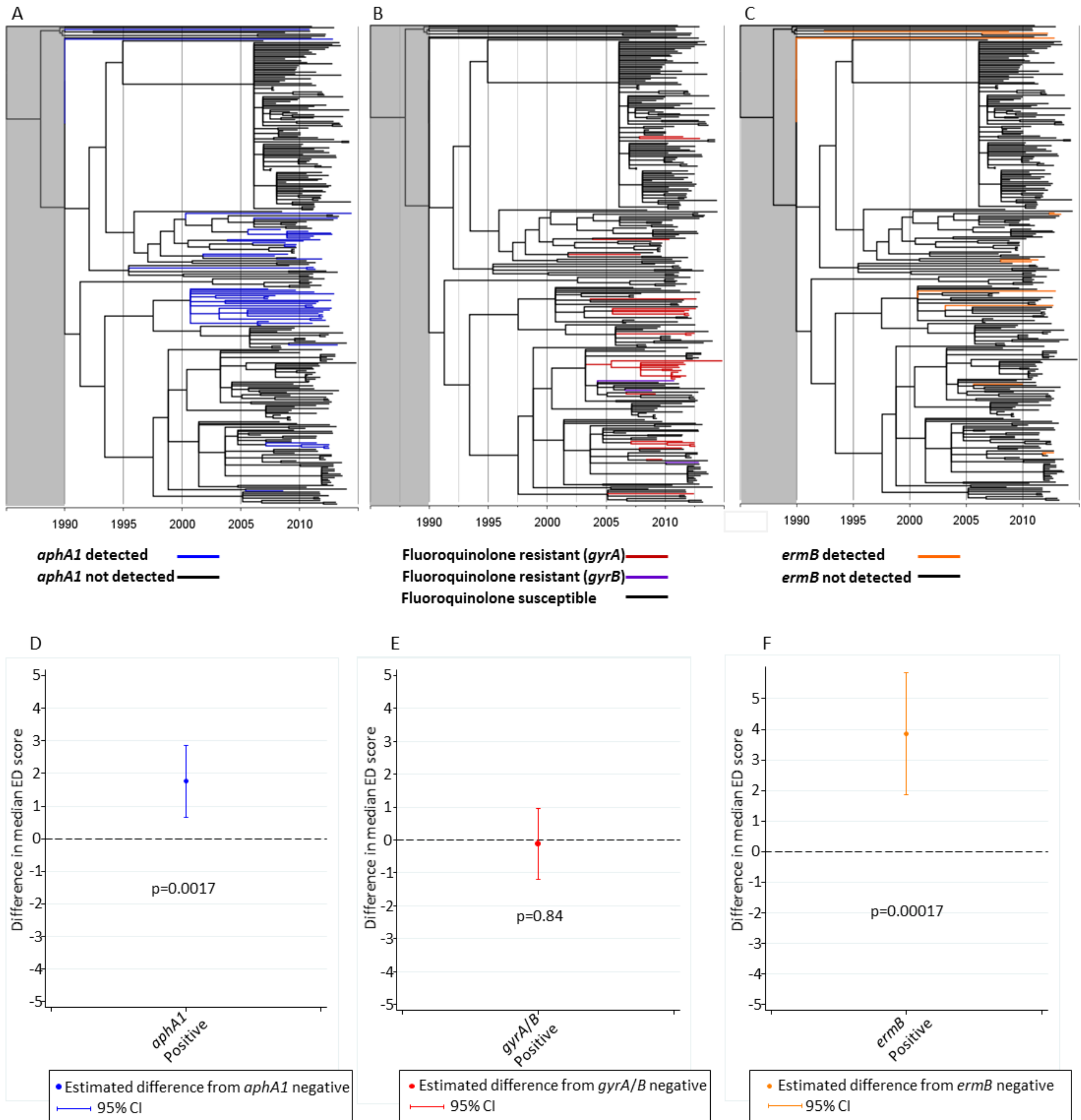
993 **Figure S1**
994 **UK-representative, time-scaled ST11/RT078 phylogeny**
995



996
997
998 The phylogeny was constructed using the same genomes as Figure 1, but with an additional
999 five genomes representing five closely related PCR-ribotype reference isolates, all of which
1000 are ST11.
1001 In contrast to Figure 1 where the branches were coloured for geographic location, here the
1002 branch colours indicate presence of tetracycline resistance determinants to highlight the
1003 associated clonal expansions.
1004

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1005 **Figure S2**
1006 **UK-representative, time-scaled RT078 phylogeny.**
1007



1008
1009
1010 The same phylogeny as Figure 1, however here the branch colours indicate presence of the
1011 following antimicrobial resistance determinants:
1012 **(A)** Aminoglycoside resistance; *aphA1* (aminoglycoside 3'-phosphotransferase) associated
1013 with resistance to streptomycin28.
1014 **(B)** Fluoroquinolone resistance conferred by non-synonymous *gyrA* or *gyrB* substitutions, (as
1015 cited14).
1016 **(C)** Clindamycin resistance; *ermB*.
1017 **(D) - (F)** Extent to which 078 clonal expansions are associated with antimicrobial resistance
1018 determinants *aphA1*, *gyrA/B* substitutions, and *ermB*, using two-sided quantile regression.

bioRxiv preprint doi: https://doi.org/10.1101/262352; this version posted February 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1019 Difference in median ED (Evolutionary Distinctiveness) score (Isaac et al., 2007) for isolates
1020 with the resistance determinant compared to isolates without. A lower ED value indicates a
1021 larger proportion of close relatives in the tree. The p-value measures the significance of the
1022 presence of the resistance determinant on ED score.
1023

1024 **TABLE 1. Antimicrobial resistance genes used to search *C. difficile* whole genome sequences**

1025

Accessory gene	Reference Sequence (GenBank)	Protein Encoded	Antimicrobial Resistance Phenotype (Predicted)
tetM	NG_048243.1	Ribosomal Protection Protein	Tetracycline
tetO	AY394561.1	Ribosomal Protection Protein	Tetracycline
tetW	FR838948.1	Ribosomal Protection Protein	Tetracycline
*tetO/32/O**	AJ295238.3	Ribosomal Protection Protein	Tetracycline
tetB(P)	NG_048319.1	Ribosomal Protection Protein	Tetracycline
tet40	JQ280445.2	Efflux pump	Tetracycline
tetA(P)	AB054980.1	Efflux pump	Tetracycline
tetL[†]	NG_048203.1	Efflux pump	Tetracycline
ermB	HG002387.1	rRNA adenine N-6-methyltransferase	Macrolide-Lincosamide-Streptogramin B (MLS_B) antibiotics including clindamycin
aphA1	M26832.1	Aminoglycoside 3'-phosphotransferase	Aminoglycoside (streptomycin)
AAC(6')-APH(2')	M13771.1	6'-N-acetyltransferase and 2"-O-phosphotransferase activities, bifunctional	Most clinically important aminoglycosides,
Housekeeping gene	**Non-synonymous substitution**	**Protein Modified**	**Antimicrobial Resistance Phenotype**
gyrA	T(82)I	DNA gyrase subunit A	Fluoroquinolones
gyrB	D(426)N	DNA gyrase subunit B	Fluoroquinolones
rpoB	R(505)K	β subunit RNA polymerase	Rifampicin

1026

1027 * Ribosomal protection protein gene mosaic (Warburton et al., 2016).
1028 [†] Only found in ST54(012) (not RT078).
1029 Additional classes of tetracycline resistance ribosomal protection proteins searched, but not found; *tet*, *otrA*, *tetS*, *tetQ*, *tet36*, *tetT, tet44* (Whittle
1030 et al., 2003; Nguygen et al., 2014).
1031 Additional tetracycline efflux pumps searched but not found: *tetA*, *tetB*, *tetC*, *tetD*, *tetE*, *tetG*, *tetH*, *tetJ*, *tetV*, *tetY*, *tetZ*, *tet30* (Nguyen et al.,
1032 2014).