Genomic survey of *Clostridium difficile* reservoirs in the East of England implicates environmental contamination of wastewater treatment plants by clinical lineages

Danesh Moradigaravand,[1,*] Theodore Gouliouris,[2] Catherine Ludden,[3] Sandra Reuter,[4] Dorota Jamrozy,[1] Beth Blane,[2] Plamena Naydenova,[2] Kim Judge,[1] Sani H. Aliyu,[2] Nazreen F. Hadjirin,[2] Mark A. Holmes,[2] Estée Török,[2] Nicholas M. Brown,[2] Julian Parkhill[1] and Sharon Peacock[2]

Abstract

There is growing evidence that patients with *Clostridiumdifficile*-associated diarrhoea often acquire their infecting strain before hospital admission. Wastewater is known to be a potential source of surface water that is contaminated with *C. difficile* spores. Here, we describe a study that used genome sequencing to compare *C. difficile* isolated from multiple wastewater treatment plants across the East of England and from patients with clinical disease at a major hospital in the same region. We confirmed that *C. difficile* from 65 patients were highly diverse and that most cases were not linked to other active cases in the hospital. In total, 186 *C. difficile* isolates were isolated from effluent water obtained from 18 municipal treatment plants at the point of release into the environment. Whole genome comparisons of clinical and environmental isolates demonstrated highly related populations, and confirmed extensive release of toxigenic *C. difficile* into surface waters. An analysis based on multilocus sequence types (STs) identified 19 distinct STs in the clinical collection and 38 STs in the wastewater collection, with 13 of 44 STs common to both clinical and wastewater collections. Furthermore, we identified five pairs of highly similar isolates (≤ 2 SNPs different in the core genome) in clinical and wastewater collections. Strategies to control community acquisition should consider the need for bacterial control of treated wastewater.

DATA SUMMARY

Sequence data deposited in the European Nucleotide Archive (ENA) (www.ebi.ac.uk/ena) under the individual accession numbers, along with the metadata of each isolate, are described in Table S1 (available in the online version of this article). There were restrictions on providing exact geographical locations of wastewater treatment plants. We confirm that all supporting data, code and protocols have been provided within the article or through supplementary data files.

INTRODUCTION

Clostridium difficile is a major cause of healthcare-associated infection worldwide [1–6]. Infection has been linked to nosocomial acquisition and epidemic spread of *C. difficile* [7], which has led to extensive hospital infection control measures combined with antimicrobial stewardship [8]. In the UK, there has been a major decline in incidence associated with a dramatic reduction in fluoroquinolone-resistant epidemic strains such as the hypervirulent ST1/ ribotype 027 (027) [9]. In parallel with this has been a shift towards increasing community-acquired disease, which accounts for about one-third of cases and an increase in *C. difficile* ribotype diversity [10]. Recent studies in the UK employing whole genome sequencing have found that less than half of isolates causing *C. difficile* infection are genetically linked to isolates from active cases [11, 12].

Taken together, these observations suggest that patients with *C. difficile*-associated diarrhoea increasingly acquire their infecting strain prior to hospital admission. These marked changes in clinical and bacterial epidemiology necessitate consideration of alternative reservoirs for *C. difficile* [13–17]. Wastewater treatment plants have been proposed as potential reservoirs for *C. difficile*, and *C. difficile* ribotypes associated with clinical disease have been isolated

previously from such sources [18–20], but surveillance of wastewater treatment plants and clinical reservoirs in the same geographical region using whole genome sequencing is lacking. Here, we investigate *C. difficile* in multiple wastewater treatment plants across the East of England, half of which were located downstream of hospitals, and compare these to clinical isolates from the same geographical region but sampled in a different time interval.

METHODS

Isolates from patients with *C. difficile*-associated diarrhoea

A retrospective study was performed at the Cambridge University Hospitals NHS Foundation Trust between October 2012 and April 2013. Stool samples were processed at the on-site Public Health England Clinical Microbiology and Public Health Laboratory. Samples were tested using a commercial assay (*C. DIFF QUIK CHEK COMPLETE*; Techlab), which detects glutamate dehydrogenase (GDH) antigen and toxins A and B of *C. difficile*. Positive samples were cultured anaerobically for 48 h onto Brazier's cycloserine-cefoxitin-egg yolk (CCEY) agar (BioConnections) using alcohol shock (https://www.gov.uk/government/publications/smi-b-10-processing-of-faeces-for-clostridium-difficile) and colonies with typical appearance were identified by MALDI-TOF MS (Biotyper version 3.1; Bruker Daltonics). Stool samples from which *C. difficile* was isolated were stored at −80 °C prior to repeated recovery of *C. difficile*, DNA extraction and sequencing. Clinical data were recorded on patient age, gender, date of hospital admission, discharge and ward movement, and previous admission to hospital in the preceding 6 months. Recurrent *C. difficile* infection was defined as the onset of diarrhoea 4 weeks or more after the end of the prior episode.

Isolation of *C. difficile* from wastewater treatment plants

A cross-sectional study was performed in which 20 municipal wastewater treatment plants were sampled in the East of England between June 2014 and January 2015. Ten plants were situated immediately downstream of acute NHS Hospital Trusts and ten plants were not connected to acute hospital effluent. Paired samples of untreated (at inlet after initial screening to remove large solids but prior to reaching the primary settlement tanks) and treated (effluent at the point of release) wastewater were obtained from each plant. The main septic tank at the Cambridge University Hospitals NHS Foundation Trust was also sampled in September 2014. At each wastewater sampling point, two consecutive grab samples of 0.5 litres each were collected and mixed into 1 litre sterile bottles containing 18 mg sodium thiosulphate (Sigma-Aldrich). Triplicate neat [1, 10 and 100 ml (the last only for treated)] and dilutions of untreted (1 ml of 10^{-1} and 10^{-2}) and treated (1 ml of 10^{-1}) wastewater samples were concentrated using the filtration technique onto 0.45 µm pore size, mixed cellulose ester filter membranes (S-Pak; Merck Millipore). For wastewater plants 12–20, an

IMPACT STATEMENT

There is growing evidence that patients with *Clostridium difficile*-associated diarrhoea often acquire their infecting strain before hospital admission. To understand potential environmental reservoirs of *C. difficile*, we recovered *C. difficile* from wastewater treatment plants across a wide geographical area and used whole genome sequencing to reveal that strains within each wastewater treatment plant not only are highly diverse but also that some are toxigenic. Furthermore, we showed that some clinical strains are closely related to wastewater isolates. These results demonstrate the clinical significance of *C. difficile* in wastewater treatment plants in controlling the spread of *C. difficile*.

alcohol shock step was additionally used on 1 and 10 ml neat samples (addition of 1 : 1 volume of 100 % ethanol for 30 min) prior to filtration to remove false-positive anaerobic vegetative growth. Membranes were placed aseptically onto chromID *C. difficile* selective plates (bioMérieux) and incubated anaerobically for 48 h at 37 °C. Multiple colonies with morphology consistent with *C. difficile* were picked from each positive plate and identified to species level by MALDI-TOF MS.

DNA sequencing, genome assembly and population genomics analyses

DNA libraries were prepared according to the Illumina protocol and sequenced on an Illumina HiSeq2000 device with 100-cycle paired-end runs to give an average depth of 85-fold. Taxonomic identity was assigned to all short reads and assemblies using Kraken [21]. Multilocus sequence types (STs) were derived from Illumina read data using an in-house tool, available at https://github.com/sanger-pathogens/mlst_check. *De novo* multi-contig draft assemblies were produced using an improved algorithm based on Velvet [22]. The assembly pipeline gave an average total length of 4 216 955 bp (range 3 839 400–5 306 342) from 45 contigs (range 22–1866) with an average contig length of 119 057 bp (range 2843–187 572 bp) and an average N50, which is a statistic measured such that at least half of the nucleotides in the assembly belongs to contigs with the length equal to or longer than the statistic, of 432337.9 (range 140 791–841 201). The *de novo* assemblies were annotated using Prokka [23], and these were then used in the pan-genome analysis using Roary [24]. A cut-off of 95 % was used to define the locus match in the pan-genome analysis. Scoary [25], with 50 permutation tests for the empirical *P*-value, was used to identify accessory genes in isolates from wastewater versus clinical origin. After reconstruction of the core genome alignment, SNPs in the core genome were extracted using an in-house tool, available at www.github.com/sanger-pathogens/snp-sites. The SNP alignment was used to build a neighbour-joining tree with the ape package in R. The SNP alignment was also used to estimate a maximum-

likelihood tree with FastTree 2.1.3 using the generalized time-reversible model. Phylogenetic trees aligned with the metadata and geographical data were visualized using Microreact (www.microreact.org), iTOL [16], FigTree (www.tree.bio.ed.ac.uk/software/figtree/) and Easyfig [17].

Bayesian analysis

Putative sites of recombination (regions of high SNP density) were removed using Gubbins [26] with five iterations. The program BEAST v1.7 [27] was then applied to the SNP alignment. Various models were tested, including a strict molecular clock with uniform and lognormal distribution for base frequency and a General Time Reversible (GTR) model with a gamma correction for among-size variation. We executed three independent chains for 50 million generations with sampling every 10 generations and checked the convergence by considering an effective sample size of greater than 200, after leaving out the first 50 million initial states as the burn-in phase, using the Tracer software in the BEAST package. TreeAnnotator software, which is part of the BEAST package, was then used to merge the simulated tree to estimate the Bayesian tree. The key parameters such as root age and substitution rate from different models were in agreement and therefore the choice of prior model did not seem to change the results. We presented the results for the strict clock model.

The likelihood of the ancestral state was estimated for the presence and absence of toxin genes using the rerootingMethod function (the SYM method) in the phytools package in R. The tree was visualized using the plotTree function in the same package.

Antimicrobial resistance determinants, virulence factors and plasmids

Putative virulence factors and resistance genes were identified using the srst2 package [28], and a 90 % coverage and similarity cut-off. The virulence factors database was obtained from (http://www.mgc.ac.cn/VFs/), and the resistance gene database was obtained from the srst2 package. The genomic regions encoding toxins A and B (PaLoc) and the binary toxin (CdtLoc) were initially identified using an in-house *in silico* PCR tool. The regions were then extracted and MUMmer [29] was used to map the whole genome assemblies to the identified genomic regions that contained the toxin genes. Except for two isolates where srst2 did not detect the toxin gene correctly, the results of the two methods were in agreement. Moreover, we ran BLAST on the elements against the non-redundant nucleotide NCBI database to annotate the genomic regions surrounding the genomic loci of toxin genes.

RESULTS

Genomic epidemiology of *C. difficile* associated with disease and from wastewater treatment plants

We defined the genomic epidemiology of *C. difficile* associated with human disease in our region through a retrospective observational study of 65 consecutive cases of

C. difficile-associated diarrhoea over 8 months at the Cambridge University Hospitals NHS Foundation Trust, a 1100-bed acute hospital in the East of England. Four patients had more than one episode over the study period (three patients with two episodes, and one patient with three episodes, i.e. recurrent infections as defined in the Methods). A total of 70 clinical isolates were analysed. Furthermore, we examined *C. difficile* in hospital sewage and municipal wastewater treatment plants, with a particular focus on treated water released into the environment. A cross-sectional survey was conducted of 20 municipal wastewater treatment plants in the East of England (10 situated immediately downstream of acute NHS Hospital Trusts and 10 not directly connected to acute hospital effluent). Paired samples of untreated and treated wastewater were cultured from each plant. A sample from the main septic tank at the Cambridge University Hospitals NHS Foundation Trust was also cultured. *C. difficile* was isolated from the hospital sewer and from 18 treatment plants, nine of which did not directly receive hospital waste (Fig. 1). *C. difficile* was isolated from treated wastewater at the point of release onto surface waters (rivers, canals or the sea) from all 18 plants. A total of 186 colonies (hereafter termed isolates) were picked from primary wastewater culture plates and sequenced, with greater emphasis placed on treated wastewater (135/186, 73 %). The median (range) number of isolates sequenced per treatment plant was 9.8 (1–21), and 97/186 (52 %) isolates were from plants in direct receipt of hospital waste.

Initial analysis based on multilocus STs revealed 19 distinct STs in the clinical collection and 38 STs in the wastewater collection (Fig. 2a). Isolate diversity was also observed in the majority of individual treatment plants, with a median (range) number of STs per plant of 6.5 (1–13) (Fig. 2a, c). Comparison of STs in clinical versus wastewater collections revealed that 13 out of 44 STs were common to both (Table S1).
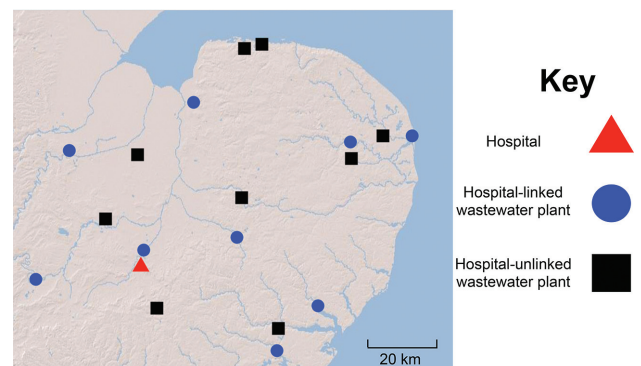

Fig. 1. Location of hospital and wastewater treatment plants in the East of England where *C. difficile* isolates were recovered. Triangle, Cambridge University Hospitals NHS Foundation Trust. Wastewater treatment plans in direct receipt of hospital effluent (circle) or not in direct receipt of hospital effluent (square).
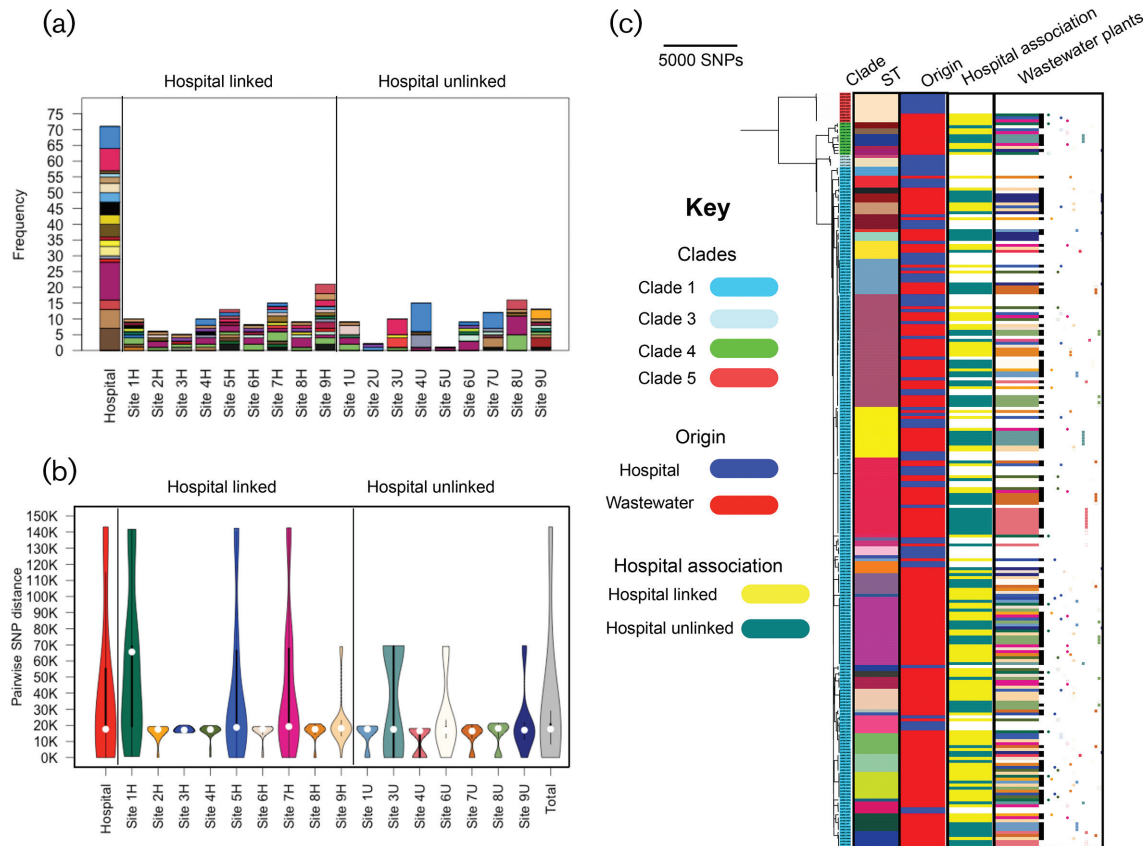
Fig. 2. Phylogenetic analysis of *C. difficile* isolates from patients and wastewater. (a) Multilocus STs for 70 clinical and 186 wastewater isolates. Clinical isolate STs are shown in a single bar, while wastewater STs are shown according to treatment plant. Each colour refers to one ST. (b) Pairwise core genome SNP distance for clinical isolates and for isolates from each treatment plant. The total histogram shows the pairwise SNP distance for the whole population. Boxes denote the interquartile range and whiskers indicate the boundary of 1.5 times the interquartile range. The white marker shows the median value. The coloured area is the probability density of the data at different values. (c) Neighbour-joining phylogenetic tree for 256 isolate genomes and the distribution of clinical and wastewater isolates. These are labelled by clade, ST, origin, link to hospital waste (only for wastewater isolates) and wastewater plants. The black bars in the wastewater plants column show the post-treatment isolates. The circle and square signs show hospital linked and hospital unlinked isolates within wastewater treatment plans, respectively. Each color corresponds to one plant. The full and empty signs in the same color show the post- and pre-treatment isolates, respectively, for each plant.

C. difficile consists of six different clades termed 1–5 and C-I, of which clade 2 has been characterized as being associated with more severe disease (hypervirulent) [30]. A phylogenetic analysis based on 112378 SNPs in the core genomes of the 70 clinical and 186 environmental isolates showed that our collection lacked isolates belonging to clade 2 (consistent with a previous study of clinical isolates from across Europe [31]) and clade C-I, but included isolates belonging to the other major *C. difficile* clades. The majority of isolates belonged to clade 1 (231/256), the remainder being distributed across other clades (4/256 clade 3, 11/256 clade 4 and 10/256 clade 5) [30] (Fig. 2c). Clade 4 was entirely composed of environmental isolates, two of which were assigned to ST37.

The phylogenetic tree demonstrated extensive mixing between the two groups (Fig. 2c). Genetic relatedness of the

70 clinical isolates was defined by a pairwise comparison of core genomes to identify the difference in SNPs between each pair. Isolates were classified as genetically related if they were ≤ 2 SNPs different, genetically distinct if >10 SNPs different and indeterminate if 3–10 SNPs different, as previously described [11]. This analysis showed that 26 isolates were genetically related to at least one other isolate, five isolates were indeterminate and 39 isolates were genetically distinct (median 6673 SNPs, range 11–65104). A phylogenetic tree based on SNPs in the core genome of the 70 isolates demonstrated considerable diversity overall (Fig. S1). Analysis of the genetically related isolates confirmed that two of four patients with recurrence had infectious episodes caused by the same strain. Epidemiological analysis of patients associated with the 26 genetically related isolates revealed 16 transmission events, of which seven were associated with direct ward contact, five with indirect ward

contact and four with hospital-wide contact. These data confirm that the majority of *C. difficile*-associated diarrhoea was unlinked to other active cases in the hospital, suggesting that the patient had acquired the infecting organism prior to admission or from unrecognized cases in the hospital.

On pairwise core genome comparison, the most closely related clinical–wastewater isolate pair differed by 1 SNP for treatment plants that directly received hospital effluent and 4 SNPs for plants that did not. Within the wastewater collection, we identified 26 pairs of closely related isolates ($\leq$2 SNPs difference) in which each isolate of the pair was from a different treatment plant (indicative of dissemination of the same strain), together with 92 pairs of highly similar isolates ($\leq$2 SNPs difference) in which each isolate in the pair was from the same plant (indicative of multiple colony picks of the same strain). Isolate relatedness based on place of origin was identified through a network analysis (Fig. S2). This relatedness was nested within a context of genetic diversity, in which the mean pairwise core genome SNP distance of isolates cultured from 15 individual plants was largely similar (~20 000 SNPs) to that of the entire clinical and wastewater collections (Fig. 2b).

ST11 (ribotype 078) has been recovered from animals and wastewater and recognized over the last decade as a pathogenic strain associated with community- and hospital-acquired disease in Europe [10, 18, 32–34]. In light of this, we used Bayesian analysis to compare the 10 ST11 isolates in our collection (seven clinical isolates, three wastewater isolates each from a different plant in receipt of hospital effluent) with the genomes of 67 ST11 isolates recovered in the Netherlands from hospitals and pig farms (Table S2), where transmission between farmers and pigs has been observed [33]. This demonstrated recent mixing of isolates in our collection with isolates from animals, as well as healthy and diseased humans in the Netherlands, with four clades diverging within the past 20 years (Fig. S3).

Distribution of toxin genes in *C. difficile* associated with disease and from wastewater treatment plants

Having established that wastewater treatment plants in the East of England release *C. difficile* into the environment, we used whole genome sequencing to evaluate the clinical relevance of isolates destined for release. This analysis considered the presence of genes encoding putative virulence determinants (primarily toxin genes) and antibiotic resistance determinants.

The detection of toxin A and B production by clinical isolates is integral to our diagnostic pathway (see Methods), and thus all our clinical isolates were toxigenic; 68 isolates carried both PaLoc and CdtLoc (tcdA+/tcdB+/cdtA+/cdtB+) and two isolates carried PaLoc alone (tcdA+/tcdB+/cdtA−/cdtB−). Screening of 186 wastewater isolates showed that 106 (56 %) were tcdA+/tcdB+/cdtA+/cdtB+, 13 (8 %) were tcdA+/tcdB+/cdtA−/cdtB−, five (2 %) were tcdA−/tcdB−/cdtA+/cdtB+, two (1 %) were tcdA−/tcdB+/cdtA−/

cdtB− and 60 (32 %) were tcdA−/tcdB−/cdtA−/cdtB− (Fig. 3a). Two ST37 (ribotype 17) genomes contained full-length *tcdB* and mutated *tcdA* genes with a premature stop codon, which has been associated with outbreaks worldwide [35–38]. Furthermore, five isolates were tcdA−/tcdB−/cdtA+/cdtB+, all of which were ST48 isolated from three wastewater treatment plants. There was a complete absence of PaLoc in the five isolate genomes, and no replacement by an alternative element.

Toxigenic isolates were isolated from all water treatment plants, including all treated wastewater samples in these plants (Fig. 3b). The proportion of toxigenic to non-toxigenic isolates was comparable for untreated and treated wastewater, and between plants that directly received hospital waste versus those that did not (Fig. 3b, c). There was a strong correlation between phylogeny and the presence of the PaLoc and CdtLoc (Pearson correlation test: $P<10^{-3}$). This is consistent with stable integration into the genome following acquisition (Figs S4 and S5). We investigated the ancestral state for the presence of the CDT locus and found that this was likely to have been present in ancestral strains (Fig. S5b) and to have been lost from the same genomic location (Fig. S5a) in non-toxigenic lineages.

Analysis of 78 additional putative virulence genes revealed that genes in the flagellar operon (implicated in adherence and linked to toxin production [39]) were over-represented in clinical isolates (Student's *t*-test $P<10^{-3}$), but no other genes were over-represented in either clinical or wastewater isolates (Fig. S6).

Distribution of antibiotic resistance determinants in *C. difficile* associated with disease and from wastewater treatment plants

Screening of all 256 isolates for genes encoding antibiotic resistance demonstrated these to be present at relatively low prevalence overall. The two most common were *ermB* encoding erythromycin resistance [33 (13 %) isolates (seven clinical, 26 wastewater)] and *tetM* encoding tetracycline resistance [18 (7 %) isolates (four clinical, 14 wastewater)] (Fig. S7). The distribution of *ermB*- and *tetM*-positive isolates across the whole collection phylogeny indicated that these genes had been gained on multiple occasions by different lineages (Fig. S7).

The mobile genetic elements (MGEs) on which *tetM* and *ermB* resided were investigated. *tetM* was identified on four distinct MGEs. The majority of *tetM*-positive isolates (13/18, 72 %) carried an accessory region (Acc1tet) that shared 99 % sequence identity with the *C. difficile* reference genome strain M120 (bases 2 171 636-2 185 712), isolated in the UK in 2007 from a human source [40]. Isolates carrying this element belonged to numerous lineages (ST11, ST29, ST35 and ST48), the last containing clinical and wastewater isolates. Two Acc1tet-positive isolates (one wastewater, one clinical; both ST11) carried an additional *tetM* accessory region (Acc2tet) which were also a close match to the genome of *C. difficile* strain MG120 (bases 416 096–
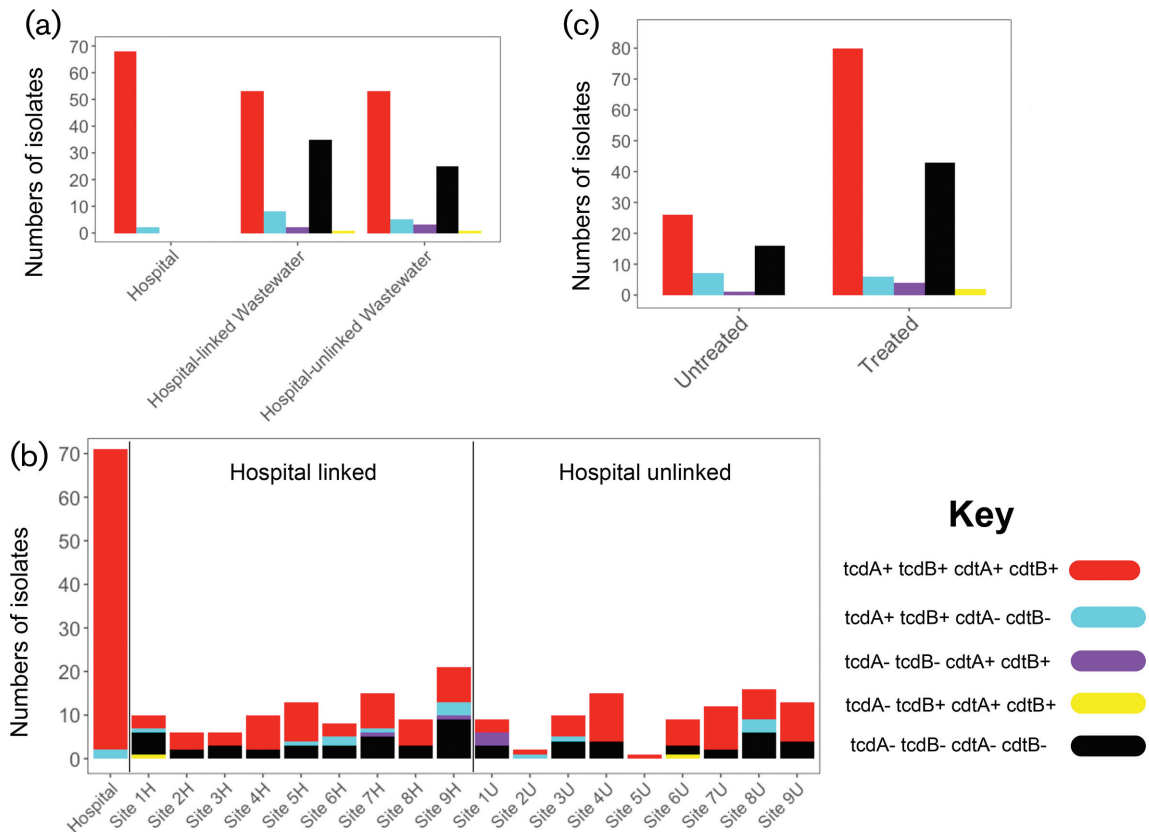
Fig. 3. Presence of two *C. difficile* pathogenicity loci in different sub-populations. PaLoc, pathogenicity locus carrying genes encoding toxin A and toxin B. CdtLoc, pathogenicity locus carrying genes encoding the CDT binary toxin. Distribution of pathogenicity loci or toxin A and toxin B genes according to: (a) source, clinical isolates associated with *C. difficile*-associated diarrhoea, or wastewater from treatment plants that were or were not linked to hospital waste; (b) individual wastewater treatment plant; and (c) treated and untreated wastewater.

485 371). A further two *tetM*-positive isolates (both ST54) carried a transposon that was highly related to Tn*6086*, and three isolates (each a different ST) carried a *tetM*-associated accessory fragment (Acc3tet), which revealed a partial match to the *Enterococcus faecium* Tn*6085* transposon. *ermB* was also associated with multiple accessory elements, although a portion of isolates (11/33, 33 %, all ST15 and including clinical and environmental isolates) carried an MGE that was closely related to the integrative mobile element Tn*6215*, identified previously in *C. difficile* strain CD80 [41]. We also identified two additional *ermB*-carrying elements that have been reported previously, namely Tn*6194* identified in *C. difficile* strain CII7 (accession HG475346) and Tn*6218* identified in *C. difficile* strain Ox42 (accession HG002387). A proportion of isolates belonging to different STs carried the *ermB* gene on a small novel accessory element of around 3.3 kb (Acc4erm). Besides antibiotic resistance genes, we identified known mutations in *gyrA* (T82I in seven clinical isolates and four wastewater isolates) and *gyrB* (S416A in seven clinical and three wastewater isolates, S366V in nine clinical isolates and one

wastewater isolate, S366A in eight clinical and 11 wastewater isolates) associated with fluoroquinolone resistance [42, 43].

DISCUSSION

We found that *C. difficile* was extensively released into surface environmental waters, which is consistent with an earlier study across South Wales in which more than 2500 environmental water samples were tested. *C. difficile* was recovered from water samples taken from rivers, lakes and the sea [44]. Previous studies have reported that *C. difficile* survive wastewater treatment processes and have found ribotypes that were common to wastewater and clinical isolates [18–20]. However, ribotyping is known to lack sufficient genomic resolution to indicate close relatedness [11]. The use of genome sequencing allowed us to characterize and compare isolates destined for release into surface waters and isolates associated with human disease. Our study revealed multiple instances of genetic relatedness between wastewater and clinical isolates, providing strong evidence for high prevalence of *C. difficile* strains in

treatment plants and overlap between these and clinically relevant strains, indicating transmission between patients and environmental waters and persistence of some genotypes in humans and the environment.

Our results suggest that circulation of *C. difficile* may include other potential reservoirs such as animal sources. For instance, the recent emergence of *C. difficile* ribotype 078 as a cause of diarrhoea in humans and its isolation from livestock farms in the Netherlands and elsewhere [33, 45] suggests that this lineage has an animal reservoir in some settings. The lineage has been increasingly identified as an emerging type with reported outbreaks in the UK [34, 46]. We isolated *C. difficile* ST11/ribotype 078 from wastewater that was genetically related to isolates from Oxford and the Netherlands, with evidence for genetic divergence over the last 20 years. This suggests inter-country as well as inter-host transmission of this important lineage.

The virulence of *C. difficile* is primarily attributed to toxin A and toxin B, located in a pathogenicity locus (PaLoc) [47], and a binary toxin encoded by the CDT locus (CdtLoc), reported to be associated with more severe disease [48]. Toxigenic *C. difficile* was extensively represented, with two-thirds of isolates positive for genes encoding toxin A, toxin B and a binary toxin encoded by the CDT locus, or genes for toxins A and B without CDT (56 and 8 %, respectively). Our findings indicate a high prevalence of toxigenic strains in wastewater and confirm that toxigenic *C. difficile*, including isolates closely related to those causing human disease, are widely dispersed across the wastewater system and are released into the environment. Human acquisition of *C. difficile* from downstream surface waters may be possible directly through recreational activities or indirectly through the use of water for irrigation of crops.

C. difficile carried in the human gut may act as a donor of genes encoding antibiotic resistance to other bacterial species of clinical importance. Furthermore, sewage provides a milieu in which mobile elements can be transferred within and between bacterial species. Screening of study isolates for carriage of genes encoding resistance to commonly used antibiotics suggested that these were at a relatively low prevalence. Further investigation of the genetic context of the two most frequent resistance genes (*tetM* and *ermB*) suggested that each was associated with multiple accessory elements. Taken together, our findings indicate that resistance genes were carried by diverse MGEs. However, *C. difficile* isolates belonging to the same lineage but derived from distinct sources (wastewater and clinical) carried the same antimicrobial resistance-associated MGEs. One explanation for this is sharing of MGEs related to common ancestry.

A limitation of our study was that recovery of wastewater isolates post-dated clinical isolates by 1 year. Contemporaneous data would allow us to infer directionality of transmission between the clinical and environment settings, but we propose that our findings still hold because the prevalence of *C. difficile* ribotypes was largely unchanged between 2014 and 2015 [*C. difficile* ribotyping network for England and Northern Ireland, Biennial report (2013–2015), www.gov.uk/government/uploads/system/uploads/attachment_data/file/491253/CDRN_2013-15_Report.pdf].

In conclusion, this study of the reservoirs and sources for human acquisition of *C. difficile* demonstrated extensive release of toxigenic *C. difficile* into the environment, including toxigenic isolates that were highly related to those causing clinical disease. The proportion of human disease based on the acquisition of *C. difficile* carriage from the natural environment is difficult to estimate, but the genetic similarity between isolates associated with disease and released into surface waters calls into question whether strategies are required to improve bacterial control in treated wastewater.

References
1. Heinlen L, Ballard JD. *Clostridium difficile* infection. *Am J Med Sci* 2010;340:247–252.
2. Bartlett JG. *Clostridium difficile*: history of its role as an enteric pathogen and the current state of knowledge about the organism. *Clin Infect Dis* 1994;18:S265–272.
3. Rupnik M, Wilcox MH, Gerding DN. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol* 2009;7:526–536.
4. Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK *et al.* Burden of *Clostridium difficile* infection in the United States. *N Engl J Med* 2015;372:825–834.
5. Leffler DA, Lamont JT. *Clostridium difficile* infection. *N Engl J Med* 2015;372:1539–1548.

6. Martin JS, Monaghan TM, Wilcox MH. Clostridium difficile infection: epidemiology, diagnosis and understanding transmission. *Nat Rev Gastroenterol Hepatol* 2016;13:206–216.

7. Kumar N, Miyajima F, He M, Roberts P, Swale A *et al.* Genome-based infection tracking reveals dynamics of *Clostridium difficile* transmission and disease recurrence. *Clin Infect Dis* 2016;62:746–752.

8. Cohen SH, Gerding DN, Johnson S, Kelly CP, Loo VG *et al.* Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). *Infect Control Hosp Epidemiol* 2010;31:431–455.

9. Dingle KE, Didelot X, Quan TP, Eyre DW, Stoesser N *et al.* Effects of control interventions on *Clostridium difficile* infection in England: an observational study. *Lancet Infect Dis* 2017;17:411–421.

10. Fawley WN, Davies KA, Morris T, Parnell P, Howe R *et al.* *Clostridium difficile* ribotyping network working G: enhanced surveillance of *Clostridium difficile* infection occurring outside hospital, England, 2011 to 2013. *Euro Surveill* 2016;21:29.

11. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A *et al.* Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med* 2013;369:1195–1205.

12. Mawer DPC, Eyre DW, Griffiths D, Fawley WN, Martin JSH *et al.* Contribution to *Clostridium Difficile* transmission of symptomatic patients with toxigenic strains who are fecal toxin negative. *Clin Infect Dis* 2017;64:1163–1170.

13. Songer JG, Anderson MA. *Clostridium difficile*: an important pathogen of food animals. *Anaerobe* 2006;12:1–4.

14. Rodriguez-Palacios A, Staempfli HR, Duffield T, Weese JS. *Clostridium difficile* in retail ground meat, Canada. *Emerg Infect Dis* 2007;13:485–487.

15. Knight DR, Squire MM, Collins DA, Riley TV. Genome analysis of *Clostridium difficile* PCR ribotype 014 lineage in Australian pigs and humans reveals a diverse genetic repertoire and signatures of long-range interspecies transmission. *Front Microbiol* 2016;7:2138.

16. Stone NE, Sidak-Loftis LC, Sahl JW, Vazquez AJ, Wiggins KB *et al.* More than 50% of *Clostridium difficile* isolates from pet dogs in Flagstaff, USA, carry toxigenic genotypes. *PLoS One* 2016;11:e0164504.

17. Moono P, Lim SC, Riley TV. High prevalence of toxigenic *Clostridium difficile* in public space lawns in Western Australia. *Sci Rep* 2017;7:41196.

18. Romano V, Pasquale V, Krovacek K, Mauri F, Demarta A *et al.* Toxigenic *Clostridium difficile* PCR ribotypes from wastewater treatment plants in southern Switzerland. *Appl Environ Microbiol* 2012;78:6643–6646.

19. Xu C, Weese JS, Flemming C, Odumeru J, Warriner K. Fate of *Clostridium difficile* during wastewater treatment and incidence in Southern Ontario watersheds. *J Appl Microbiol* 2014;117:891–904.

20. Steyer A, Gutiérrez-Aguirre I, Rački N, Beigot Glaser S, Brajer Humar B *et al.* The detection rate of enteric viruses and *Clostridium difficile* in a waste water treatment plant effluent. *Food Environ Virol* 2015:164–172.

21. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 2014;15:R46.

22. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008;18:821–829.

23. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.

24. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.

25. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 2016;17:238.

26. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;43:e15.

27. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7:214.

28. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014;6:90.

29. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* 2002;30:2478–2483.

30. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and evolution in the genome of *Clostridium difficile*. *Clin Microbiol Rev* 2015;28:721–741.

31. Davies KA, Ashwin H, Longshaw CM, Burns DA, Davis GL *et al.* Diversity of *Clostridium difficile* PCR ribotypes in Europe: results from the European, multicentre, prospective, biannual, point-prevalence study of *Clostridium difficile* infection in hospitalised patients with diarrhoea (EUCLID), 2012 and 2013. *Euro Surveill* 2016;21.

32. Cairns MD, Stabler RA, Shetty N, Wren BW. The continually evolving *Clostridium difficile* species. *Future Microbiol* 2012;7:945–957.

33. Knetsch CW, Connor TR, Mutreja A, van Dorp SM, Sanders IM *et al.* Whole genome sequencing reveals potential spread of *Clostridium difficile* between humans and farm animals in the Netherlands, 2002 to 2011. *Euro Surveill* 2014;19:20954.

34. Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C *et al.* Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. *Clin Infect Dis* 2008;47:1162–1170.

35. Shin BM, Kuak EY, Yoo SJ, Shin WC, Yoo HM. Emerging toxin A-B+ variant strain of *Clostridium difficile* responsible for pseudomembranous colitis at a tertiary care hospital in Korea. *Diagn Microbiol Infect Dis* 2008;60:333–337.

36. Huang H, Weintraub A, Fang H, Wu S, Zhang Y *et al.* Antimicrobial susceptibility and heteroresistance in Chinese *Clostridium difficile* strains. *Anaerobe* 2010;16:633–635.

37. Kuijper EJ, de Weerdt J, Kato H, Kato N, van Dam AP *et al.* Nosocomial outbreak of *Clostridium difficile*-associated diarrhoea due to a clindamycin-resistant enterotoxin A-negative strain. *Eur J Clin Microbiol Infect Dis* 2001;20:528–534.

38. Cairns MD, Preston MD, Lawley TD, Clark TG, Stabler RA *et al.* Genomic epidemiology of a protracted hospital outbreak caused by a toxin A-negative *Clostridium difficile* sublineage PCR ribotype 017 strain in London, England. *J Clin Microbiol* 2015;53:3141–3147.

39. Stevenson E, Minton NP, Kuehne SA. The role of flagella in *Clostridium difficile* pathogenicity. *Trends Microbiol* 2015;23:275–282.

40. He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF *et al.* Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci U S A* 2010;107:7527–7532.

41. Goh S, Hussain H, Chang BJ, Emmett W, Riley TV *et al.* Phage φC2 mediates transduction of Tn6215, encoding erythromycin resistance, between *Clostridium difficile* strains. *MBio* 2013;4:e00840-13.

42. Spigaglia P, Barbanti F, Mastrantonio P, Brazier JS, Barbut F *et al.* European Study Group on *Clostridium difficle*: Fluoroquinolone resistance in *Clostridium difficile* isolates from a prospective study of *C. difficile* infections in Europe. *J Med Microbiol* 2008;57:784–789.

43. Drudy D, Kyne L, O'Mahony R, Fanning S. gyrA mutations in fluoroquinolone-resistant *Clostridium difficile* PCR-027. *Emerg Infect Dis* 2007;13:504–505.

44. Al Saif N, Brazier JS. The distribution of *Clostridium difficile* in the environment of South Wales. *J Med Microbiol* 1996;45:133–137.

45. Bakker D, Corver J, Harmanus C, Goorhuis A, Keessen EC *et al.* Relatedness of human and animal *Clostridium difficile* PCR ribotype 078 isolates determined on the basis of multilocus variable-

number tandem-repeat analysis and tetracycline resistance. *J Clin Microbiol* 2010;48:3744–3749.

46. Eyre DW, Fawley WN, Rajgopal A, Settle C, Mortimer K *et al.* Comparison of control of *Clostridium difficile* infection in six english hospitals using whole-genome sequencing. *Clin Infect Dis* 2017;65: 433–441.

47. Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW *et al.* Evolutionary history of the *Clostridium difficile* pathogenicity locus. *Genome Biol Evol* 2014;6:36–52.

48. Gerding DN, Johnson S, Rupnik M, Aktories K. *Clostridium difficile* binary toxin CDT: mechanism, epidemiology, and potential clinical importance. *Gut Microbes* 2014;5:15–27.