Genomic Epidemiology of a Protracted Hospital Outbreak Caused by a Toxin A-Negative *Clostridium difficile* Sublineage PCR Ribotype 017 Strain in London, England

M. D. Cairns, M. D. Preston, T. D. Lawley, T. G. Clark, R. A. Stabler, B. W. Wren

Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, London, United Kingdom; UCL Centre for Clinical Microbiology, University College London, London, United Kingdom; Public Health Laboratory London, Health Protection Agency, Division of Infection, The Royal London Hospital, London, United Kingdom; Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, United Kingdom; Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

*Clostridium difficile* remains the leading cause of nosocomial diarrhea worldwide, which is largely considered to be due to the production of two potent toxins: TcdA and TcdB. However, PCR ribotype (RT) 017, one of five clonal lineages of human virulent *C. difficile*, lacks TcdA expression but causes widespread disease. Whole-genome sequencing was applied to 35 isolates from hospitalized patients with *C. difficile* infection (CDI) and two environmental ward isolates in London, England. The phylogenetic analysis of single nucleotide polymorphisms (SNPs) revealed a clonal cluster of temporally variable isolates from a single hospital ward at University Hospital Lewisham (UHL) that were distinct from other London hospital isolates. De novo assembled genomes revealed a 49-kbp putative conjugal transposon exclusive to this hospital clonal cluster which would not be revealed by current typing methodologies. This study identified three sublineages of *C. difficile* RT017 that are circulating in London. Similar to the notorious RT027 lineage, which has caused global outbreaks of CDI since 2001, the lineage of toxin-defective RT017 strains appears to be continually evolving. By utilization of WGS technologies to identify SNPs and the evolution of clonal strains, the transmission of outbreaks caused by near-identical isolates can be retraced and identified.

*Clostridium difficile* causes a spectrum of disease ranging from mild diarrhea to life-threatening colitis, mainly in elderly, hospitalized patients. The disease pathogenesis is largely considered to be due to the production of two potent toxins: TcdA and TcdB (1).

The global emergence of the PCR ribotype (RT) 027 strain was responsible for multiple outbreaks and increased disease severity in Canada and the United States in 2001 (2). This strain has since spread to South America, China, Japan, Hong Kong, Korea, Singapore, Australia, and New Zealand and throughout Europe (3). Although RT027 remains the dominant clone in the United States, Europe has seen a decline in this RT. This has occurred simultaneously with an increase in other virulent RTs such as RT017 and RT078 (4–7).

The genome and phylogeny of RT027 have been well studied (2, 8–10), and multilocus sequence typing (MLST) and whole-genome sequencing (WGS) studies have confirmed the existence of at least five *C. difficile* clonal lineages where one is made up mostly of RT027 and another of RT017 (11). Pathogenic RT027 strains produce both toxins A and B and a third unrelated binary toxin (CDT) that has been implicated in virulence (12). In contrast, RT017 strains invariably lack most of the tcdA gene and completely lack the CDT gene yet have emerged worldwide, causing significant disease (6, 13). The reasons for the emergence of a less toxigenic lineage remain unclear. There is evidence that the prevalence of clinically relevant cases of CDI due to toxin A⁻B⁺ strains has increased globally (14–16). RT017 strains have been reported in The Netherlands (17), Poland (18), Ireland (14), China (19), Korea (20–22), Argentina (23), Australia (15), Israel (24), and Japan (25). Interestingly, the epidemiology of *C. difficile* in the Asia/Pacific regions and eastern parts of Europe appears to differ from that elsewhere where the prevalence of toxin A⁻B⁺ strains was higher than the prevalence other RTs, including RT027. Furthermore, given that some diagnostic laboratories rely on detecting toxin A, the incidence of A⁻B⁺ RT017 is likely to be significantly underreported.

*C. difficile* is a relatively clonal organism (8, 26, 27) and therefore amenable to WGS and subsequent single nucleotide polymorphism (SNP) analysis. WGS offers considerable advantages over traditional phenotypic and genotypic typing methods and performs a fine-grained analysis that facilitates the accurate tracing of the sources and routes of transmission (28).

The University Hospital Lewisham (UHL) in South London experienced multiple, temporally variable clusters of RT017 between March 2009 and April 2011 in one elderly care ward. This study investigates the genotypic characteristics of these clusters of 18 RT017 isolates and two environmental RT017 isolates from the ward, one community RT017 isolate from a patient who had spent...
time on the ward, two RT017 isolates from patients who spent
time in other locations in UHL, and 13 human isolates from other
locations in UHL, and 13 human isolates from other London hospitals in a similar time period using WGS (Table 1).

| Isolate  | Sample date (day/mo/yr) | Hospital [date(s) (day/mo/yr) patient in elderly care ward at UHL] |
|---------|-------------------------|---------------------------------------------------------------|
| H-UHL-1 | 2005                    | UHL (ward nonexistent)                                        |
| UH-2    | 11/03/09                | UHL (time of specimen)                                        |
| UH-3    | 27/03/09                | UHL (24/02/09 – 05/03/09)                                     |
| UH-4    | 17/04/09                | UHL (time of specimen)                                        |
| UH-5    | 16/04/09                | UHL (time of specimen)                                        |
| UH-6    | 28/09/09                | UHL (time of specimen)                                        |
| UH-7    | 20/09/09                | UHL (time of specimen)                                        |
| UH-8    | 16/10/09                | UHL (time of specimen)                                        |
| UH-9    | 29/10/09                | UHL (time of specimen)                                        |
| UH-10   | 28/01/10                | UHL (time of specimen)                                        |
| UH-11   | 08/02/10                | UHL (time of specimen)                                        |
| UH-12   | 17/02/10                | UHL (time of specimen)                                        |
| UH-13   | 01/04/10                | UHL (time of specimen)                                        |
| UH-14   | 26/04/10                | UHL (time of specimen)                                        |
| UH-15   | 17/07/10                | UHL (time of specimen)                                        |
| UH-16   | 19/07/10                | UHL (time of specimen)                                        |
| UH-17   | 06/08/10                | UHL (never)                                                  |
| UH-18   | 10/08/10                | UHL (time of specimen)                                        |
| E-UHL-19 | 13/08/10             | UHL (ward: side-room toilet)                                  |
| E-UHL-20 | 13/08/10             | UHL (ward: side-room floor)                                   |
| UH-21   | 04/10/10                | UHL (never)                                                  |
| UH-22   | 07/10/10                | UHL (04/06/10 – 12/07/10)                                     |
| UH-23   | 26/04/11                | UHL (15/02/11 – 14/04/11)                                     |
| C-UHL-24 | 08/03/13             | UHL (26/12/12 – 28/12/12, 04/02/13 – 07/02/13, and 08/03/13 – 20/02/13) |
| NP-25   | 13/05/08                | Northwick Park Hospital                                       |
| B-26    | 27/02/09                | Barnet Hospital                                               |
| NM-27   | 2005                    | North Middlesex Hospital                                      |
| GOSH-28 | 22/03/10                | Great Ormond Street Hospital                                  |
| GOSH-29 | 24/03/10                | Great Ormond Street Hospital                                  |
| GOSH-30 | 27/03/10                | Great Ormond Street Hospital                                  |
| RF-31   | 09/12/10                | Royal Free Hospital                                           |
| CX-32   | 25/01/11                | Charing Cross Hospital                                        |
| B-33    | 11/05/11                | Barnet Hospital                                               |
| QM-34   | 16/07/11                | Queen Mary’s Hospital                                         |
| WX-35   | 08/12/11                | Whipp’s Cross Hospital                                        |
| WX-36   | 16/01/12                | Whipp’s Cross Hospital                                        |
| GOSH-37 | 30/11/13                | Great Ormond Street Hospital                                  |

a UHL, University Hospital Lewisham.  

b Historical isolate predating the building of the hospital ward.  

c Environmental isolate recovered from the ward.  

d Community-acquired infection isolate.

**RESULTS**

All 37 isolates were confirmed to be RT017 by repeat PCR ribotyping. Eight isolates from the UHL temporal clusters were subtyped using multilocus variable-number tandem-repeat analysis (MLVA): six were found to be indistinguishable (UH-2, UH-4, UH-5, UH-6, UH-7, and UH-9) and two differed from the other six isolates by only one locus (UHL-3 and UHL-8).

**Whole-genome SNP and de novo genome assembly analysis.** The 37 isolates were whole-genome sequenced, and the resulting data were processed to identify and analyze high-quality SNPs (30). A total of 162 biallelic SNP loci in the samples from the 4,308,325 bp of the M68 reference were identified; the majority (79.0% [128/162]) exhibited a minor allele frequency (MAF) of <10%, including 54.3% (88/162) of loci being identified in one sample. Only 17 SNP loci (10.5%) had a nonreference allele frequency of >50%. Each isolate contained up to 46/162 (28.4%) mutations, with 64.9% (24/37) of isolates containing between 17 and 19 (10.5 to 11.7%) SNPs.

The data set revealed 23 different SNP patterns (haplotypes), 9 of which (labeled A to I) were only found in isolates H-UHL-1 and UH-2 to UHL-23 (Tables 2 and 3; Fig. 1). There are 24 SNP loci unique to these 23 UHL samples (16 nonsynonymous, 5 synonymous, and 3 nongenic). Haplotype A is the core haplotype with only the 11 SNPs common to all of these UHL isolates. Haplotypes B to H differ from haplotype A
TABLE 2 SNPs unique to cluster 1-UHL (isolates H-UHL-1 and H-UHL-2 to H-UHL-23 inclusive)

| Position | Reference residue | Alternative residue | Type          | Product/putative function                          | Haplotype |
|----------|-------------------|---------------------|---------------|---------------------------------------------------|-----------|
| 345335   | S                 | R                   | Nonsynonymous | Protein-tyrosine phosphatase reductase             | A to I    |
| 433205   | S                 | S                   | Synonymous    | Formate/nitrite transporter                        | A to I    |
| 578215   | P                 | S                   | Nonsynonymous | Iron hydrogenase                                  | A to I    |
| 707105   | F                 | L                   | Nonsynonymous | Multidrug family ABC transporter permease         | A to I    |
| 1123155  | G                 | S                   | Nonsynonymous | Putative membrane protein                         | A to I    |
| 1241002  | L                 | L                   | Nonsynonymous | NhaC family Na⁺/H⁺ antiporter                    | A to I    |
| 3136457  | A                 | A                   | Nonsynonymous | 3-Hydroxybutyrate dehydrogenase                   | A to I    |
| 2764775  | P                 | L                   | Nonsynonymous | Diguanylate kinase signaling protein              | A to I    |
| 3072208  | G                 | A                   | Nonsynonymous | Maf-like protein                                  | A to I    |
| 3202066  | L                 | F                   | Nonsynonymous | Multidrug family ABC transporter                 | A to I    |
| 4025381  |                  |                    | Intergenic    | Unknown                                           | A to I    |
| 1491685  |                  |                    | Intergenic    | Unknown                                           | B, C, and D |
| 584197   | C                 | R                   | Nonsynonymous | Response regulator (quorum-sensing system)        | E         |
| 1245898  | H                 | N                   | Nonsynonymous | Copper-sensing transcriptional repressor CooR     | H         |
| 1395682  | I                 | L                   | Nonsynonymous | Hypothetical protein                              | H         |
| 583796   | R                 | L                   | Nonsynonymous | Response regulator (quorum-sensing system)       | F         |
| 1932695  | I                 | V                   | Nonsynonymous | Putative membrane protein                         | C         |
| 3698806  |                  |                    | Intergenic    | Unknown                                           | G         |
| 3056134  | L                 | I                   | Nonsynonymous | RNaG                                              | D         |
| 34552    | S                 | Y                   | Nonsynonymous | DNA-directed RNA polymerase beta chain            | I         |
| 2744067  | E                 | E                   | Synonymous    | Putative TPR repeat-containing protein            | I         |
| 2813984  | E                 | E                   | Synonymous    | GntR family transcriptional regulator             | I         |
| 3289662  | S                 | Y                   | Nonsynonymous | ABC transporter substrate-binding protein         | I         |
| 3766047  | K                 |                    | Nonsynonymous | Two-component response regulator                  | I         |

by 1 or 2 extra SNPs. Haplotype I was distinguishable from haplotype A by 5 SNPs. A heat map of the genetic Manhattan distance (Fig. 1) and a maximum-likelihood phylogenetic tree (generated by RAXML [39]) (Fig. 2A) revealed three related groups of samples. Cluster 1-UHL was composed exclusively of haplotypes A to I, containing 23 (out of 24) isolates from UHL (Fig. 2B). Cluster 2-M68 (containing the M68 reference strain) encompasses the outer London hospitals and a UHL patient (UHL-24) who had community-acquired CDI. Cluster 3 contains all of the samples from the three inner London hospitals (GOSH, Royal Free, and Charing Cross). Twenty-four (14.8%), 61 (37.7%), and 71 (43.8%) SNP loci are found exclusively in clusters 1-UHL, 2-M68, and 3, respectively, and only six SNP loci (3.8%) have mutations in more than one cluster. Of these loci exclusive to the three clusters, 11 (out of 24 SNP loci), 1 (out of 61), and 16 (out of 72) were found in every sample in their respective clusters.

De novo assembly of each isolate and comparison to the M68 reference strain revealed a 49-kbp, chromosomal, genetic region exclusive to cluster 1-UHL (23/37 isolates) (Fig. 3). Programmatic and visual inspection of the comparisons revealed no other large structural variations between samples. Using Prokka (38) to annotate this 49-kbp putative transposon with predicted protein coding regions, we identified 45 sequences that have an ortholog to known gene sequences (see Table S1 in the supplemental material). Predicted genes include sortase (sortase B), sporulation (Spo0J), and collagen-binding proteins with known links to phenotypic and virulence markers. These 45 predicted genes are highly conserved with 41 being amino-acid identical across all 23 samples in cluster 1-UHL.

**DISCUSSION**

Between 2009 and 2011, an elderly care ward at UHL experienced multiple, temporally variable clusters of CDI caused by RT017. This study analyzed the genotypic characteristics of these 20 RT017 isolates along with a further 17 RT017 isolates: 2 isolates from patients nursed in other wards in UHL, 1 community-acquired isolate, 1 historical isolate from UHL that predates the building of the hospital ward, and 12 isolates from other London hospitals.

Phylogenetic analysis of the SNPs identified through WGS of the isolates revealed three related groups. Twenty-three UHL isolates (including the historical H-UHL isolate) formed a tight cluster with only nine haplotypes and few SNP differences between them. The 13 non-UHL isolates, the community-acquired isolate, 1 historical isolate from UHL that predates the building of the hospital ward, and 12 isolates from other London hospitals.

Persistence of a C. difficile PCR Ribotype 017 Strain
(E-UHL-19 and E-UHL-20) recovered from the toilet and floor of the elderly care ward side room were indistinguishable from the cluster 1-UHL isolates; the environment was therefore contaminated with this strain. Two UHL isolates (UHL-17 and UHL-21) from patients who were never admitted to the elderly care ward were found to be part of cluster 1-UHL, strongly suggesting interward transmission. One isolate (C-UHL-24) recovered from a patient specimen taken on the day of admittance to the elderly care ward was distinguishable from the cluster 1-UHL isolates; however, this CDI was defined as community acquired with symptoms occurring \(\leq 48\) h after hospital admission.

On 30 September 2010, hydrogen peroxide vapor (HPV) decontamination was performed in the elderly care ward at UHL. Subsequent to this, three isolates of RT017 indistinguishable from the cluster 1-UHL isolates were recovered from patients in October 2010 and April 2011 (Table 1). This strain has persisted at UHL with a possible internal reservoir that was never eliminated during the HPV decontamination. With the historical isolate from 2005 being indistinguishable from the more recent clonal isolates, the possibility of an external source that has reintroduced this strain to the elderly care ward cannot be excluded.

These data suggest that a clonal strain has persisted in one London hospital for at least 5 years and is different from other RT017 strains that are circulating in London hospitals. Our study is a snapshot of isolates from hospitalized patients with CDI from London, England, and demonstrates the persistence of a toxin \(A^-B^+\) RT017 strain in a ward at UHL since at least 2005. We identify three lineages, one harboring a large putative conjugate transposon found exclusively in the cluster 1-UHL isolates.

This lineage of toxin \(A^-\) and CDT-defective RT017 strains appears to be evolving similar to the RT027 counterpart, which has caused global outbreaks since 2001. The RT017 lineage, with its unique toxin profile and unusual global prevalence, is understudied. Our data show that there are existing questions about the biology, population structure, and epidemiology of toxin \(A^-B^+\)
RT017. Answering these will contribute to our understanding of the evolution of *C. difficile* across all lineages and help the diagnosis and treatment of CDI. With utilization of WGS technologies to identify SNPs and the evolution of clonal strains, the transmission of outbreaks caused by near-identical isolates can be retraced and identified.

Acknowledgments
We acknowledge Debbie Flaxman and Maria Osman from the infection control team at UHL for help with the patient epidemiological data. We acknowledge the London CDRN laboratory for support with PCR ribotyping and the lead CDRN laboratory in Leeds for MLVA testing. We also thank David Harris at the WTSI for assistance with DNA sequencing. We declare no conflicts of interest.

The work was supported by the National Institute for Health Research (NIHR), the Wellcome Trust (grants 086418 and 098051), and the Medical Research Council (grants PF451 and MR/K000551/1). M.D.C. is funded by a Doctoral Research Fellowship award from the NIHR. This report is independent research arising from a CSO Healthcare Scientist Award supported by the NIHR and the CSO. The views expressed in this

**FIG 2** (A) Maximum-likelihood phylogenetic analysis of all 37 London isolates based on core genome SNPs against the M68 reference. (B) Cluster 1-UHL haplotypes.

**FIG 3** Artemis Comparison Tool (ACT) comparison between *C. difficile* strains M68 and H-UHL-19, illustrating the putative chromosomal transposon exclusive to H-UHL-19 and other members of the cluster 1-UHL.
publication are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Study design, data collection and analysis, decision to publish, or preparation of the manuscript are those of the author(s) and not necessarily those of the

Cairns et al.

Gerding Elliott Wren MW, Kinson R, Sivapalan M, Shemko M, Shetty NR.

Emergence of new PCR ribotypes from the hypervirulent

difficile

Cairns MD, Stabler RA, Wren BW. 2012. Molecular characterization of toxin A-negative, toxin B-positive variant strains of Clostridium difficile isolated in Korea. Diagn Microbiol Infect Dis 67:198–201. http://dx.doi.org/10.1016/j.diagmicrobio.2010.01.007.

Goorhuis A, Legaria MC, van den Berg RJ, Harmanus C, Klaassen CH, Brazier JS, Lumelsky G, Kuijper EJ. 2009. Application of multiple-locus variable-number tandem-repeat analysis to determine clonal spread of toxin A-negative Clostridium difficile in a general hospital in Buenos Aires, Argentina. Clin Microbiol Infect 15:1080–1086. http://dx.doi.org/10.1111/j.1469-0691.2009.02759.x.

Samra Z, Talmon S, Bahar J. 2002. High prevalence of toxin A-negative toxin B-positive Clostridium difficile in hospitalized patients with gastrointestinal disease. Diagn Microbiol Infect Dis 43:189–192. http://dx.doi.org/10.1016/S0732-8893(02)00400-5.

Komatsu M, Kato H, Aihara M, Shimakawa K, Iwaski M, Nagasaka Y, Fukuda S, Matsuo S, Arakawa Y, Watanabe M, Iwatani Y. 2003. High frequency of antibiotic-associated diarrhea due to toxin A-negative, toxin B-positive Clostridium difficile in a hospital in Japan and risk factors for infection. Eur J Clin Microbiol Infect Dis 22:525–529. http://dx.doi.org/10.1007/s10096-003-0992-5.

Elliott B, Squire MM, Thane S, Chang BJ, Brazier JS, Rupnik M, Riley TV. 2011. New types of toxin A-negative, toxin B-positive Clostridium difficile. Infect Control Hosp Epidemiol 28:932–940. http://dx.doi.org/10.1086/619181.

Collins DA, Hawkey PM, Riley TV. 2013. Epidemiology of Clostridium difficile infection in Asia. Antimicrob Resist Infect Control 2:21. http://dx.doi.org/10.1186/2047-2994-2-21.

Kuijper EJ, de Weerdt J, Kato H, Kato N, van Dam AP, van der Vorm ER, Weel J, van Rheenen C, Dankert J. 2001. Nosocomial outbreak of Clostridium difficile-associated diarrhea due to a clindamycin-resistant enterotoxin A-negative strain. Eur J Clin Microbiol Infect Dis 20:528–534. http://dx.doi.org/10.1007/s10096-000-0550-5.

Pituch H, van den Braak N, van Leeuwen W, van Belkum A, Martirosian G, Obuch-Woszczyński P, Luczak M, Meisel-Mikolajczyk F. 2001. Clonal dissemination of a toxin-A-negative/toxin-B-positive Clostridium difficile strain from patients with antibiotic-associated diarrhea in Poland. Clin Microbiol Infect 7:442–446. http://dx.doi.org/10.1046/j.1469-0691.2001.00312.x.

Huang H, Weintrob A, Fang H, Wu S, Zhang Y, Nord CE, 2010. Antimicrobial susceptibility and heteroresistance in Chinese Clostridium difficile strains. Anaerobe 16:633–635. http://dx.doi.org/10.1016/j.anaero.2010.09.002.

Kim H, Jeong SH, Roh KH, Hong SG, Kim JW, Shin MG, Kim MN, Shin BH, Uh Y, Lee H, Lee K. 2010. Investigation of toxin gene diversity, molecular epidemiology, and antimicrobial resistance of Clostridium difficile isolated from 12 hospitals in South Korea. J Korean J Lab Med 30:491–497. http://dx.doi.org/10.5334/jkjm.2010.35.5.491.

Kim H, Riley TV, Kim M, Kim CK, Yong D, Lee K, Chong Y, Park JW. 2008. Increasing prevalence of toxin A-negative, toxin B-positive isolates of Clostridium difficile in Korea: impact on laboratory diagnosis. J Clin Microbiol 46:1116–1117. http://dx.doi.org/10.1128/JCM.01818-08.

Kuijper EJ, de Weerdt J, Kim H, Seo Y, Joo BS, Jeong SH, Chong Y, Lee K. 2010. Molecular characterization of toxin A-negative, toxin B-positive variant strains of Clostridium difficile isolated in Korea. Diagn Microbiol Infect Dis 67:198–201. http://dx.doi.org/10.1016/j.diagmicrobio.2010.01.007.

Roper C, Clark TG, Djimde AA, Doumbo OK, Nosten F, Fairhurst RM, Conway DJ, Komatsu M, Kato H, Aihara M, Shimakawa K, Iwasaki M, Nagasaka Y, Fukuda S, Matsuo S, Arakawa Y, Watanabe M, Iwatani Y. 2003. Importance of toxin A, toxin B, and CDT in virulence of an enterotoxin A-negative strain. J Med Microbiol 52:1108–1111. http://dx.doi.org/10.1099/jmm.0.031062-0.

Drudy D, Harnedy N, Fanning S, Hanrahan M, Kato H, Holt KE, Seth-Smith HM, Quail MA, Rance R, Brooks K, Churcher C, Harris D, Bentley SD, Burrows C, Clark L, Corton C, Murray V, Rose G, Thurston S, van Tonder A, Walker D, Wren BW, Dougan G, Parkhill J. 2010. Evolutionary dynamics of Clostridium difficile over short and long time scales. Proc Natl Acad Sci U S A 107:7527–7532. http://dx.doi.org/10.1073/pnas.0914322107.

Parkhill J, Wren BW. 2011. Bacterial epidemiology and biology-lessons from genome sequencing. Genome Biol 12:230. http://dx.doi.org/10.1186/gb-2011-12-10-230.

Harris SR, Fel IJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastr H, Parkhill J, Peacock SJ, Conway DJ, Roper C, Clark TG. 2014. PlasmidView: a web-based resource to visualise global Plasmidis salpinis genome variants. J Infec Dis 209:1808–1813. http://dx.doi.org/10.1093/infdis/jit812.
for Illumina sequence data. Bioinformatics 30:2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.
32. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.
33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
34. Zerbino DR. 2010. Using the Velvet de novo assembler for short-read sequencing technologies. Curr Protoc Bioinformatics Chapter 11:unit 11.5.
35. Seemann T, Gladman S. 2012. VelvetOptimiser. Victorian Bioinformatics Consortium. http://bioinformatics.net.au/software.velvetoptimiser.shtml.
36. The R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
37. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics 25:1968–1969. http://dx.doi.org/10.1093/bioinformatics/btp347.
38. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.
39. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. http://dx.doi.org/10.1093/bioinformatics/btu033.
40. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O’Connor L, Ip CL, Golubchik T, Batty EM, Finney JM, Wylie DH, Didelot X, Piazza P, Bowden R, Dingle KE, Harding RM, Crook DW, Wilcox MH, Peto TE, Walker AS. 2013. Diverse sources of C. difficile identified on whole-genome sequencing. N Engl J Med 369:1195–1205. http://dx.doi.org/10.1056/NEJMoA1216064.