



Development and Implementation of Whole Genome Sequencing-Based Typing Schemes for *Clostridioides difficile*

Sandra Janezic^{1,2*} and Maja Rupnik^{1,2}

¹ National Laboratory for Health, Environment and Food, Maribor, Slovenia, ² Medical Faculty, University of Maribor, Maribor, Slovenia

Clostridioides difficile is an important nosocomial pathogen increasingly observed in the community and in different non-human reservoirs. The epidemiology and transmissibility of *C. difficile* has been studied using a variety of typing methods, including more recently developed whole-genome sequence (WGS) analysis that is becoming used routinely for bacterial typing worldwide. Here we review the schemes for WGS-based typing methods available for *C. difficile* and their applications in the field of human *C. difficile* infection (CDI). The two main approaches to discover genomic variations are single nucleotide variant (SNV) analysis and methods based on gene-by-gene comparisons (frequently called core genome or whole genome MLST, cgMLST, or wgMLST). SNV analysis currently provides the ultimate resolution, however, typing nomenclature and standardized methodology are missing. On the other hand, gene-by-gene approaches allow portability and standardized nomenclature, and are therefore becoming increasingly popular in bacterial epidemiology and outbreak investigation. Two commercial software packages (BioNumerics and Ridom SeqSphere+) and an open source database (Enterobase) for allele and sequence type determination for *C. difficile* are currently available. Proof-of-concept WGS studies have already enabled advances in the investigation of the population structure of *C. difficile* species, microevolution within the epidemic strains, intercontinental transmission over time and in tracking of transmission events. WGS of clinical *C. difficile* isolates demonstrated a considerable genetic diversity suggesting diverse reservoirs for CDI. WGS was also shown to aid in resolving relapses and reinfections in recurrent CDI and has potential for use as a tool for assessing hospital infection prevention and control performance.

Keywords: *Clostridioides (Clostridium) difficile*, wgMLST, cgMLST, typing, CDI, SNV

OPEN ACCESS

Edited by:

Vitali Sintchenko,
University of Sydney, Australia

Reviewed by:

Valter Viana Andrade-Neto,
Oswaldo Cruz Foundation
(Fiocruz), Brazil
Daniel Raymond Knight,
Murdoch University, Australia

*Correspondence:

Sandra Janezic
sandra.janezic@nlzoh.si

Specialty section:

This article was submitted to
Infectious Diseases - Surveillance,
Prevention and Treatment,
a section of the journal
Frontiers in Public Health

Received: 31 May 2019

Accepted: 08 October 2019

Published: 24 October 2019

Citation:

Janezic S and Rupnik M (2019)
Development and Implementation of
Whole Genome Sequencing-Based
Typing Schemes for *Clostridioides*
difficile. *Front. Public Health* 7:309.
doi: 10.3389/fpubh.2019.00309

INTRODUCTION

Clostridioides (Clostridium) difficile is currently one of the most important human pathogens (1). The majority of *C. difficile* infections (CDI) is still identified or associated with the healthcare environment, though the incidence of community CDI is rapidly increasing. Because of its importance as a nosocomial pathogen, the development of different typing methods was needed to identify and control hospital transmissions and outbreaks. Several typing schemes were introduced

for *C. difficile*; among early phenotypic methods serotyping was used widely, but subsequently replaced by pulsed-field gel electrophoresis (PFGE) and finally by PCR ribotyping which is the current gold standard for *C. difficile* typing (1–3). However, apart from multi locus sequence typing (MLST), standardization of all established typing methods has been difficult and inter-laboratory comparisons hampered (2).

Although these methods have contributed greatly to understanding of the epidemiology of CDI, they usually do not have sufficient discriminatory power to distinguish between closely related strains needed for outbreak investigations and to understand transmission events. With development of new sequencing methodologies, there is now the possibility to sequence and compare whole bacterial genomes and not rely only on a single or a few genomic loci to address the genetic relatedness of strains. Therefore, the genome-wide sequence analysis is now frequently used for molecular typing to provide accurate and reproducible investigation of the relatedness of isolates with the highest level of genetic resolution (4).

Here we will review studies on the development and implementation of typing methods based on whole genome sequencing (WGS) and their applications, focusing mainly on healthcare-associated CDI. Proof-of-concept studies have already demonstrated the general applicability of WGS-based typing for investigation of global and national surveillance of *C. difficile* epidemiology, and have expanded our understanding of transmission dynamics and recurrent infections. All these aspects will be reviewed here. However, use of WGS for strain characterization such as analysis of virulence and resistance gene pool and evolutionary aspects will not be covered in this review. *C. difficile* is commonly isolated also from animals and the environment and the paper by Knight and Riley in this special issue (5) will cover applications of comparative genomics from this perspective.

COMPARATIVE GENOMICS AND TWO DIFFERENT APPROACHES FOR WGS TYPING

For the principles of next-generation sequencing technologies and bioinformatic processes, from the raw sequence data to the genomes, the reader is referred to other recent reviews (4, 6).

To determine the genomic similarities and differences between investigated isolates (e.g., to determine which strains could be clonal) different comparative genomics approaches are available. They differ mainly in methodologies used, easiness of data sharing and their discriminatory power. Below we will briefly describe the two of most commonly used approaches for typing of isolates for epidemiological surveillance purposes. The first one is based on comparison of differences in single nucleotide polymorphisms (SNPs), also called single nucleotide variant (SNV) sites. The second approach is based on analysis of multiple genes across the whole genome, so called gene-by gene or allele-based approaches. This is also designated core genome (cg) or whole genome (wg) multi locus sequence typing (cgMLST or wgMLST) (Figure 1).

SNV Approach—When Are Two Strains Clonal?

Strain typing based on core genome SNVs (cgSNVs) is currently considered as a method with very high discriminatory power, since it allows us to distinguish between isolates if their genomes differ in a single nucleotide (7). In this approach, short reads (data generated from sequencing of short genomic fragments) or assembled contigs (longer contiguous sequences of overlapping reads) are mapped against the genome of a reference strain to identify differences in coding and non-coding regions. This process is named variant calling (8). The pipeline that has been widely used for SNV analysis of *C. difficile* includes mapping of short reads to a reference genome, variant calling, filtering of high quality SNVs, and identification and removal of putative recombination regions. The result is a concatenated set of high quality SNVs present in the core genome (part of genome that is common to all comparing isolates). The number of SNVs is subsequently used to assess genetic relatedness of isolates (9–11). Relationships between isolates can be visualized by constructing phylogenetic trees to help us understand transmission networks.

The choice of the reference strain can have significant impact, especially on the resolution of SNV-based approaches. The reference strain should be closely related to the isolates included in the comparison since only the regions present in the reference strain will be used for variant calling. Therefore, the more distant the reference sequence the more regions will be omitted from the analysis. Also, a standardized nomenclature would be difficult to adopt since there are multiple algorithms used to analyze SNVs. For this reason, SNV calling is a very useful method for local transmission analysis but not as appropriate for global strain comparisons, unless the genome sequences are made publicly available. However, in this case, the genomes still need to be (re)analyzed locally (8).

The commonly adopted way to determine relatedness of strains in the SNV approach is to count the number of SNV differences between two sequences (SNV threshold). However, it is important to note that proposed criteria of SNV relatedness should not be taken as the absolute rules but should be considered as a guide (8). To determine the SNV threshold, it is important to know the evolutionary rates, i.e., the rate at which the particular bacterial genome evolves (12). This can be estimated from longitudinal sampling from infected individuals and then assessing the number of accumulated substitutions in the genome over time (9).

By comparing genome sequences of the first and the last isolate obtained from individual patient (samples were collected at a median interval of 51 days), an evolutionary rate of 0.74 SNVs (95% confidence interval, 0.22–1.40) per genome per year and a mean within-host diversity of 0.30 SNVs (95% CI, 0.13–0.42 SNVs) were determined, in the study by Eyre et al. (10). Similar estimations of *C. difficile* evolutionary rates were obtained in other studies, either by using serial samples from the patients with recurrent or on-going CDI and/or in *in vitro* gut models (9, 10, 13). By using this prediction of evolutionary rate, the guideline for two isolates being clonal, or genetically related (are most probably a result of direct transmission), is that there are ≤ 2 SNVs between their sequenced genomes. For



genetically unrelated isolates ≥ 10 SNVs are expected (10). This SNV relatedness criterion has now been widely accepted for transmission networks and outbreak investigations, and used in several studies that will be presented later in this review.

Gene-by-Gene Comparison, cgMLST, and wgMLST

Cg- or wgMLST typing works on the same principles as the classical MLST, described by Maiden et al. (14), a comparison of strains based on sequence differences in a pre-defined set of housekeeping genes/loci. Usually seven housekeeping genes/loci are included in MLST schemes for most bacteria, including *C. difficile* (15). For each of the seven loci, the different sequences are assigned distinct allele numbers and the alleles at each genes are described as the allelic profile (Figure 1). Finally, for each allelic profile (the series of seven allele numbers) a unique sequence type (ST) is determined (14).

Because only a small number of genes are included in the analysis, MLST usually does not have sufficient discriminatory power to differentiate between closely related strains, e.g., strains that belong to the same PCR ribotypes, which makes this method insufficient for investigations of transmission events.

To overcome this, an extension of MLST using a genome-wide gene-by-gene allele calling of hundreds or thousands of loci, so-called cgMLST and wgMLST was developed (16). The cgMLST scheme utilizes comparison of the non-repetitive genes that are conserved in all the members of a species, so called core genes. On the other hand, wgMLST examines a greater number of loci, and includes accessory genes as well as the core genes; these are genes that are variably present across a species (Figure 1), including also repetitive genes and pseudogenes (4).

Available Schemes for *C. difficile* WGS-Based Typing

For *C. difficile*, three publicly available schemes are available for cg- and/or wgMLST typing, and analysis can be performed either by using commercial software (BioNumerics, Applied Maths or SeqSphere+, Ridom) or by a freely accessible online resource (Enterobase). Additional new schemes are being developed (<https://www.biorxiv.org/content/10.1101/686212v1?rss=1>). The cgMLST scheme for *C. difficile* include 2270 loci (60.4% of the genes present in strain 630; SeqSphere; <https://www.cgmlst.org/ncs/schema/3560802/>) (17). wgMLST is available in Enterobase and in BioNumerics where, together with 1,999

core genes, another 6,713 accessory genes are included in the analysis <http://www.applied-maths.com/sites/default/files/extra/Release-Note-Clostridium-difficile-schema.pdf>.

The advantage of cgMLST and MLST is that sequences and allelic profiles of strains can be compared via the internet with central databases enabling uniform typing nomenclature that facilitate international comparability of typing data (16, 17). On the other hand, wgMLST might offer greater resolution between closely related strains, but the nomenclature is not standardized. However, EnteroBase contains all publically available genomic sequences (uploaded from public archives and assembled into annotated draft genomes), and therefore wgMLST data can be compared to all previously published *C. difficile* genomes and interpreted within a global context (18).

In contrast to SNV, the allele-based approaches do not need the genome of a closely related reference strain for the initial alignment of reads or contigs. Also, in the allele-based approach, both mutation (usually resulting in a single SNV) and recombination (that is more likely to introduce multiple deletions or insertions within allele) are counted as a single evolutionary event, meaning that there is no need to apply additional steps to identify and remove putative recombination regions (9, 19).

To test the discriminatory power and applicability of cgMLST to differentiate closely related strains, Bletz et al. (17) reanalyzed data from published outbreak investigations. With cgMLST they were able to differentiate among epidemiologically related strains and the conclusions were in concordance with the published SNV analysis. By re-analyzing two different outbreak investigations and considering the guide for number of SNV expected in genetically unrelated and related isolates (≥ 10 SNV and < 2 SNVs, respectively) (10), the authors proposed a threshold of ≥ 7 alleles difference for strains being unrelated and ≤ 6 alleles for strains that are likely to be clonal. With this threshold, the cgMLST predicted the same clusters of related strains as SNV analysis. All strains within the defined threshold were assigned to the same cgMLST cluster type (CT) (17).

CURRENT IMPLEMENTATION OF WGS TYPING IN HUMAN CDI

The feasibility of using WGS of *C. difficile* genomes on benchtop sequencing platforms for transmission investigation to rapidly distinguish between outbreak and non-outbreak cases in a clinically relevant timescale was first demonstrated in 2012 in a pilot study conducted by Eyre et al. (20). Since then SNV-based analysis has been widely adopted for CDI surveillance and has revealed some novel understandings about transmission dynamics and recurrent infections (Table 1).

Source Identification for Hospital CDI Cases

Traditionally, most cases of CDI have been thought to be acquired within the hospital environment, where transmissions occurs by horizontal spread from

symptomatic patients (39, 40). However, assessment of CDI transmission in hospital settings by classical genotyping approaches was hampered by the low discriminatory power of used methods and by the number of patients that carry endemic genotypes, either PCR ribotypes or STs (9).

To assess the role of symptomatic patients in the transmission of *C. difficile* in the hospital environment Eyre et al. (10), sequenced genomes of *C. difficile* isolates from 1,223 patients with CDI. In this study, only 35% ($n = 333$) of isolates could be genetically linked (had ≤ 2 SNV) to at least one other isolate from a symptomatic patient and for 36% ($n = 120$) of these cases no plausible epidemiological link could be identified. Isolates from almost half (45%) the patients were genetically unrelated (≥ 10 SNPs) to any other previous case, meaning that these patients had likely acquired *C. difficile* from sources other than symptomatic patients. These findings suggest that there are rather diverse reservoirs of *C. difficile* and that transmissions other than those occurring between symptomatic patients within the hospital settings should be considered (e.g., asymptomatic patients, animals, households, and environmental sources) (10).

The role of asymptomatic patients in the transmission of *C. difficile* was explored by WGS in another study conducted by Eyre et al. (22), which demonstrated that although asymptomatic carriage is common, transmission from asymptomatic carriers is likely to be infrequent. In a similar Canadian study, slightly higher linkage rates were reported, where 46 and 52% of CDI cases could be linked to previous symptomatic and infected or colonized patients, respectively (36).

A study conducted in a single hospital demonstrated that a diverse set of isolates can be found also among children with CDI and that *C. difficile* transmissions, direct or indirect, between children with CDI are even less frequent (12.5%) than transmissions among adult CDI patients (35).

Several other studies have also addressed the questions of importance of other non-hospital reservoirs in *C. difficile* transmission and are reviewed in more details by Knight and Riley in this issue (5).

Use of WGS for Study of CDI Recurrences

Within 2 months after treatment of an initial CDI episode, up to 25% of patients develop recurrent infection (41). Recurrent infection can be due to reinfection (CDI caused by newly acquired strain) or relapse (CDI caused by the original strain). Discrimination between relapses and reinfection usually does not have direct clinical implications and will not affect treatment. However, it might be important for controlling CDI, either through interventions to manage *C. difficile* transmission, or treatment policies (25). Several studies have already demonstrated usefulness of WGS comparisons in understanding the epidemiology of CDI recurrences (23–26). In these studies, the authors used similar approaches as described for transmission studies. In case of reinfections, isolates from the initial and following episodes were expected to be genetically unrelated, differing ≥ 10 SNVs, and in case of relapses, the isolates would be clonal, differing in ≤ 2 SNVs (23). All studies that explored the source of recurrent infection demonstrated

TABLE 1 | WGS-based studies of *C. difficile* transmissions, outbreaks, or recurrences.

References	Aim	Country	Description
Didelot et al. (9)	Transmission	UK	Microevolutionary analysis of <i>C. difficile</i> (assessment of within-host evolutionary rate) and use of whole-genome sequencing for studying <i>C. difficile</i> transmission.
Eyre et al. (20)	Transmission	UK	A proof-of-principle study to investigate potentials of benchtop sequencers in routine clinical practice to investigate transmissions. Example of small cluster of genetically (MLST) identical <i>C. difficile</i> strains that could be differentiated with WGS.
Eyre et al. (10)	Transmission	UK	Investigating the role of symptomatic patients in the transmission of <i>C. difficile</i> . Study also demonstrates that in the settings with standard infection control most cases of infections are acquired from other sources, not symptomatic cases.
Eyre et al. (21)	Mixed infections	UK	Describing new algorithm for detection of mixed CDI in clinical samples from whole genome sequencing data.
Eyre et al. (22)	Transmission	UK	Investigating the role of asymptomatic patients in the transmission of <i>C. difficile</i> .
Eyre et al. (23)	Recurrence	UK	Use of WGS to determine if the reductions in recurrence of CDI observed with fidaxomicin occurred by preventing relapse, reinfection or both. Study demonstrated that fidaxomicin was superior to vancomycin in treating recurrent CDI.
Mac Aoga'in et al. (24)	Recurrence	Ireland	Use of WGS of <i>C. difficile</i> to discriminate between relapses and reinfections, and putative patient-patient transmission events in Ireland.
Kumar et al. (25)	Transmission	UK	A WGS to track the transmission of <i>C. difficile</i> PCR ribotype 027 within single hospital in UK, and to distinguish between the relapses and reinfections.
Sim et al. (26)	Recurrence	USA	Use of WGS to determine the rate of relapse and reinfection in patients with recurrent CDI.
Mawer et al. (27)	Transmission	UK	Exploring the role of symptomatic patients that are toxigenic strain positive but fecal toxin negative in transmissions of <i>C. difficile</i> .
Eyre et al. (28)	Transmission	UK	Use of WGS as surveillance tool to assess infection control effectiveness in hospitals by identifying the extent of hospital-acquired CDI transmissions within hospitals.
Stoesser et al. (29)	Transmission	UK	Investigation of genetic overlap of infant and regional <i>C. difficile</i> strains in Oxfordshire.
Donskey et al. (30)	Transmission	USA	Transmission of <i>C. difficile</i> from colonized or infected long-term care facility residents.
Endres et al. (31)	Outbreak	USA	Environmental transmission of <i>C. difficile</i> PCR ribotype 027 at a long-term care facility.
Eyre et al. (32)	Transmission	UK	WGS to analyze distinct patterns of <i>C. difficile</i> PCR ribotype spread across Europe.
Halstead et al. (33)	Transmission	UK	WGS to investigate if asymptomatic carriers contribute to nosocomial CDI.
Isidro et al. (34)	Outbreak	Portugal	Genomic investigation of <i>C. difficile</i> PCR ribotype 017 outbreak strains.
Kociolek et al. (35)	Transmission	USA	Transmission of CDI among symptomatic children.
Kong et al. (36)	Transmission	Canada	Investigation of transmission patterns between infected and colonized patients.
Williamson et al. (37)	Transmission	USA	Transmission of PCR ribotype 027 within healthcare facility and comparison to global collection of ribotype 027 isolates.
García-Fernández et al. (38)	Transmission	Spain	Routes and frequencies of transmission of <i>C. difficile</i> in a tertiary-care hospital in Madrid.

that the majority of recurrent episodes are caused by primarily infecting strain, meaning that relapses are more common than reinfections (23–26).

BACKWARD COMPATIBILITY BETWEEN WGS AND MLST

Currently, an assortment of classical and WGS-based typing methods is used for investigations of *C. difficile* epidemiology (2). Reverse compatibility of WGS with traditional typing methods is therefore important to compare the genotypes obtained with different approaches and to compare newly sequenced strains to existing and historical strains (42). From WGS data, seven MLST loci can be easily extracted to determine the allelic profile and ST. For ST calling directly from draft genomes a publically available PubMLST.org database can be used (43). SeqSphere and BioNumerics also enable ST determination directly from WGS data.

WHY CAN PCR RIBOTYPE NOT BE DETERMINED WITH WGS

PCR ribotyping has become a method of choice for typing of *C. difficile* in the majority of laboratories (2, 44). The method is based on analysis of banding patterns of PCR-amplified internal transcribed spacers (ITS) located between 16S and 23S rRNA genes in ribosomal operon. In *C. difficile*, as in many other bacteria, the ribosomal operon is present in several copies in the genome and different copies differ in the length of ITS (45) and, due to intraspecific diversity of ITS, PCR ribotyping is a good method for *C. difficile* genotyping (2).

In contrast to MLST-ST, PCR ribotype cannot be directly determined from WGS. Regions that are amplified in PCR ribotyping are repetitive and it is not possible to map short sequence reads generated by NGS correctly to such repetitive and modular regions (45, 46). To assign a PCR ribotype to a new ST or cgMLST cluster type, a representative strain would still need to be PCR ribotyped. But with the advances in NGS technologies (e.g.,

PacBio and Nanopore), read lengths are continually increasing (4). The availability of very long and very precise sequences will ultimately enable the *in silico* PCR ribotyping.

The ability to predict PCR ribotypes from whole genome sequencing data remains controversial. While the genome sequences of strains belonging to the same PCR ribotype mostly group together, it is important to appreciate the differences between a true 'PCR ribotype determination' and ribotype inferred from genome sequencing data. Firstly, while grouping of strains with identical PCR ribotype is to be expected, there are exceptions and similarity of genome sequences of two different PCR ribotypes has been documented (36 and unpublished data). Secondly, due to limitations of short read sequencing explained above, comparison of two genomes shows only similarities in large part of genome, but not necessarily in the regions that are actually used for PCR ribotyping (i.e. ITS). Therefore, it is important to differentiate between ribotypes determined by actual PCR ribotyping and putative PCR ribotypes based on genome similarity, but excluding rDNA regions.

REFERENCES

- Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. *Clostridium difficile* infection. *Nat Rev Dis Primer.* (2016) 2:16020. doi: 10.1038/nrdp.2016.20
- Huber CA, Foster NF, Riley TV, Paterson DL. Challenges for standardization of *Clostridium difficile* typing methods. *J Clin Microbiol.* (2013) 51:2810–4. doi: 10.1128/JCM.00143-13
- Knetsch CW, Lawley TD, Hensgens MP, Corver J, Wilcox MW, Kuijper EJ. Current application and future perspectives of molecular typing methods to study *Clostridium difficile* infections. *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull.* (2013) 18:20381. doi: 10.2807/ese.18.04.20381-en
- Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, et al. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. *Clin Microbiol Rev.* (2017) 30:1015–63. doi: 10.1128/CMR.00016-17
- Knight DR, Riley TV. Genomic delineation of zoonotic origins of *Clostridium difficile*. *Front Public Health.* (2019) 7:164. doi: 10.3389/fpubh.2019.00164
- Carriço JA, Rossi M, Moran-Gilad J, Van Domselaar G, Ramirez M. A primer on microbial bioinformatics for nonbioinformaticians. *Clin Microbiol Infect.* (2018) 24:342–9. doi: 10.1016/j.cmi.2017.12.015
- Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet.* (2012) 13:601–12. doi: 10.1038/nrg3226
- Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect.* (2018) 24:350–4. doi: 10.1016/j.cmi.2017.12.016
- Didelot X, Eyre DW, Cule M, Ip CL, Ansari MA, Griffiths D, et al. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biol.* (2012) 13:R118. doi: 10.1186/gb-2012-13-12-r118
- Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med.* (2013) 369:1195–205. doi: 10.1056/NEJMoa1216064
- 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.* (2017) 7:2138. doi: 10.3389/fmicb.2016.02138
- Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, et al. Genome-scale rates of evolutionary change in bacteria. *Microb Genomics.* (2016) 2:e000094. doi: 10.1099/mgen.0.000094
- Eyre DW, Walker AS, Freeman J, Baines SD, Fawley WN, Chilton CH, et al. Short-term genome stability of serial *Clostridium difficile* ribotype 027 isolates in an experimental gut model and recurrent human disease. *PLoS ONE.* (2013) 8:e63540. doi: 10.1371/journal.pone.0063540
- Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA.* (1998) 95:3140–5. doi: 10.1073/pnas.95.6.3140
- Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, et al. Multilocus sequence typing of *Clostridium difficile*. *J Clin Microbiol.* (2010) 48:770–8. doi: 10.1128/JCM.01796-09
- Maiden MCJ, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol.* (2013) 11:728–36. doi: 10.1038/nrmicro3093
- Bletz S, Janezic S, Harmsen D, Rupnik M, Mellmann A. Defining and evaluating a core genome multilocus sequence typing scheme for genome-wide typing of *Clostridium difficile*. *J Clin Microbiol.* (2018) 56:e01987–17. doi: 10.1128/JCM.01987-17
- Zhou Z, Alikhan N-F, Mohamed K, the Agama Study Group, Achtman M. The user's guide to comparative genomics with Enterobase. Three case studies: micro-clades within *Salmonella enterica* serovar Agama, ancient and modern populations of *Yersinia pestis*, and core genomic diversity of all *Escherichia*. *bioRxiv.* (2019) 613554. doi: 10.1101/613554
- He M, Sebahia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, et al. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci USA.* (2010) 107:7527–32. doi: 10.1073/pnas.0914322107
- Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, et al. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open.* (2012) 2:e001124. doi: 10.1136/bmjopen-2012-001124
- Eyre DW, Cule ML, Griffiths D, Crook DW, Peto TEA, Walker AS, et al. Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in *Clostridium difficile* transmission. *PLoS Comput Biol.* (2013) 9:e1003059. doi: 10.1371/journal.pcbi.1003059

CONCLUSION

WGS-based typing methods offer an excellent platform with high resolution and reproducibility that enable studies of both transmission and epidemiology of CDI, as well as positioning strains within the global population. However, especially for the understanding of global CDI epidemiology, whole genome data availability, either by sharing raw data or allelic profiles through freely accessible databases that support direct comparison of isolates is of paramount importance.

AUTHOR CONTRIBUTIONS

SJ and MR both contributed to the conception and writing of the paper.

FUNDING

This work was in part supported by the Slovenian Research Agency grant J4-8224.

22. Eyre DW, Griffiths D, Vaughan A, Golubchik T, Acharya M, O'Connor L, et al. Asymptomatic *Clostridium difficile* colonisation and onward transmission. *PLoS ONE*. (2013) 8:e78445. doi: 10.1371/journal.pone.0078445
23. Eyre DW, Babakhani F, Griffiths D, Seddon J, Del Ojo Elias C, Gorbach SL, et al. Whole-genome sequencing demonstrates that fidaxomicin is superior to vancomycin for preventing reinfection and relapse of infection with *Clostridium difficile*. *J Infect Dis*. (2014) 209:1446–51. doi: 10.1093/infdis/jit598
24. Mac Aogáin M, Moloney G, Kilkenny S, Kelleher M, Kelleghan M, Boyle B, et al. Whole-genome sequencing improves discrimination of relapse from reinfection and identifies transmission events among patients with recurrent *Clostridium difficile* infections. *J Hosp Infect*. (2015) 90:108–16. doi: 10.1016/j.jhin.2015.01.021
25. Kumar N, Miyajima F, He M, Roberts P, Swale A, Ellison L, et al. Genome-based infection tracking reveals dynamics of *Clostridium difficile* transmission and disease recurrence. *Clin Infect Dis Off Publ Infect Dis Soc Am*. (2016) 62:746–52. doi: 10.1093/cid/civ1031
26. Sim JHC, Truong C, Minot SS, Greenfield N, Budvytiene I, Lohith A, et al. Determining the cause of recurrent *Clostridium difficile* infection using whole genome sequencing. *Diagn Microbiol Infect Dis*. (2017) 87:11–6. doi: 10.1016/j.diagmicrobio.2016.09.023
27. Mawer DPC, Eyre DW, Griffiths D, Fawley WN, Martin JSH, Quan TP, et al. Contribution to *Clostridium Difficile* transmission of symptomatic patients with toxigenic strains who are fecal toxin negative. *Clin Infect Dis Off Publ Infect Dis Soc Am*. (2017) 64:1163–70. doi: 10.1093/cid/cix079
28. Eyre DW, Fawley WN, Rajgopal A, Settle C, Mortimer K, Goldenberg SD, et al. Comparison of control of *Clostridium difficile* infection in six English hospitals using whole-genome sequencing. *Clin Infect Dis Off Publ Infect Dis Soc Am*. (2017) 65:433–41. doi: 10.1093/cid/cix338
29. Stoesser N, Eyre DW, Quan TP, Godwin H, Pill G, Mbuvi E, et al. Epidemiology of *Clostridium difficile* in infants in Oxfordshire, UK: Risk factors for colonization and carriage, and genetic overlap with regional *C. difficile* infection strains. *PLoS ONE*. (2017) 12:e0182307. doi: 10.1371/journal.pone.0182307
30. Donskey CJ, Sunkesula VCK, Stone ND, Gould CV, McDonald LC, Samore M, et al. Transmission of *Clostridium difficile* from asymptotically colonized or infected long-term care facility residents. *Infect Control Hosp Epidemiol*. (2018) 39:909–16. doi: 10.1017/ice.2018.106
31. Endres BT, Dotson KM, Poblete K, McPherson J, Lancaster C, Bassères E, et al. Environmental transmission of *Clostridioides difficile* ribotype 027 at a long-term care facility; an outbreak investigation guided by whole genome sequencing. *Infect Control Hosp Epidemiol*. (2018) 39:1322–9. doi: 10.1017/ice.2018.230
32. Eyre DW, Davies KA, Davis G, Fawley WN, Dingle KE, De Maio N, et al. Two distinct patterns of *Clostridium difficile* diversity across Europe indicating contrasting routes of spread. *Clin Infect Dis Off Publ Infect Dis Soc Am*. (2018) 67:1035–44. doi: 10.1093/cid/ciy252
33. Halstead FD, Ravi A, Thomson N, Nuur M, Hughes K, Brailey M, Oppenheim BA. Whole genome sequencing of toxigenic *Clostridium difficile* in asymptomatic carriers: insights into possible role in transmission. *J Hosp Infect*. (2018) 102:125–34. doi: 10.1016/j.jhin.2018.10.012
34. Isidro J, Menezes J, Serrano M, Borges V, Paixão P, Mimoso M, et al. Genomic study of a *Clostridium difficile* multidrug resistant outbreak-related clone reveals novel determinants of resistance. *Front Microbiol*. (2018) 9:2994. doi: 10.3389/fmicb.2018.02994
35. Kociolek LK, Gerding DN, Espinosa RO, Patel SJ, Shulman ST, Ozer EA. *Clostridium difficile* whole genome sequencing reveals limited transmission among symptomatic children: a single-center analysis. *Clin Infect Dis*. (2018) 67:229–34. doi: 10.1093/cid/ciy060
36. Kong LY, Eyre DW, Corbeil J, Raymond F, Walker AS, Wilcox MH, et al. *Clostridium difficile*: investigating transmission patterns between infected and colonized patients using whole genome sequencing. *Clin Infect Dis*. (2019) 68:204–9. doi: 10.1093/cid/ciy457
37. Williamson CHD, Stone NE, Nunnally AE, Hornstra HM, Wagner DM, Roe CC, et al. A global to local genomics analysis of *Clostridioides difficile* ST1/RT027 identifies cryptic transmission events in a northern Arizona healthcare network. *Microb Genom*. (2019) 5:e000271. doi: 10.1099/mgen.0.000271
38. García-Fernández S, Frentrup M, Steglich M, Gonzaga A, Cobo M, López-Fresneña N, et al. Whole-genome sequencing reveals nosocomial *Clostridioides difficile* transmission and a previously unsuspected epidemic scenario. *Sci Rep*. (2019) 9:6959. doi: 10.1038/s41598-019-43464-4
39. Vonberg R-P, Kuijper EJ, Wilcox MH, Barbut F, Tüll P, Gastmeier P, et al. Infection control measures to limit the spread of *Clostridium difficile*. *Clin Microbiol Infect*. (2008) 14:2–20. doi: 10.1111/j.1469-0691.2008.01992.x
40. Khanna S, Pardi DS. *Clostridium difficile* infection: new insights into management. *Mayo Clin Proc*. (2012) 87:1106–17. doi: 10.1016/j.mayocp.2012.07.016
41. Kelly CP. Can we identify patients at high risk of recurrent *Clostridium difficile* infection? *Clin Microbiol Infect*. (2012) 18:21–7. doi: 10.1111/1469-0691.12046
42. Rossen JWA, Friedrich AW, Moran-Gilad J, ESCMID Study Group for Genomic and Molecular Diagnostics (ESGMD). Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis*. (2018) 24:355–60. doi: 10.1016/j.cmi.2017.11.001
43. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res*. (2018) 3:124. doi: 10.12688/wellcomeopenres.14826.1
44. Fawley WN, Knetsch CW, MacCannell DR, Harmanus C, Du T, Mulvey MR, et al. Development and validation of an internationally-standardized, high-resolution capillary gel-based electrophoresis PCR-ribotyping protocol for *Clostridium difficile*. *PLoS ONE*. (2015) 10:e0118150. doi: 10.1371/journal.pone.0118150
45. Sadeghifard N, Gürtler V, Beer M, Seviour RJ. The mosaic nature of intergenic 16S-23S rRNA spacer regions suggests rRNA operon copy number variation in *Clostridium difficile* strains. *Appl Environ Microbiol*. (2006) 72:7311–23. doi: 10.1128/AEM.01179-06
46. Janezic S, Indra A, Rattai T, Weinmaier T, Rupnik M. Recombination drives evolution of the *Clostridium difficile* 16S-23S rRNA intergenic spacer region. *PLoS ONE*. (2014) 9:e106545. doi: 10.1371/journal.pone.0106545

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Janezic and Rupnik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.