Molecular typing is an essential tool to monitor *Clostridium difficile* infections and outbreaks within healthcare facilities. Molecular typing also plays a key role in defining the regional and global changes in circulating *C. difficile* types. The patterns of *C. difficile* types circulating within Europe (and globally) remain poorly understood, although international efforts are under way to understand the spatial and temporal patterns of *C. difficile* types. A complete picture is essential to properly investigate type-specific risk factors for *C. difficile* infections (CDI) and track long-range transmission. Currently, conventional agarose gel-based polymerase chain reaction (PCR) ribotyping is the most common typing method used in Europe to type *C. difficile*. Although this method has proved to be useful to study epidemiology on local, national and European level, efforts are made to replace it with capillary electrophoresis PCR ribotyping to increase pattern recognition, reproducibility and interpretation. However, this method lacks sufficient discriminatory power to study outbreaks and therefore multilocus variable-number tandem repeat analysis (MLVA) has been developed to study transmission between humans, animals and food. Sequence-based methods are increasingly being used for *C. difficile* fingerprinting/typing because of their ability to discriminate between highly related strains, the ease of data interpretation and transferability of data. The first studies using whole-genome single nucleotide polymorphism typing of healthcare-associated *C. difficile* within a clinically relevant timeframe are very promising and, although limited to select facilities because of complex data interpretation and high costs, these approaches will likely become commonly used over the coming years.

**Introduction**

*Clostridium difficile* is a gram-positive rod-shaped anaerobic bacterium that is capable of forming spores. Since its discovery as a cause of antibiotic-associated pseudomembranous colitis nearly 30 years ago [1], *C. difficile* has become the major cause of antibiotic-associated diarrhoea. Antibiotics change the protective normal gut flora, which enables *C. difficile* to colonise the colon. Clinical symptoms may range from simple diarrhoea to severe colitis which can result in death [2]. Symptoms are primarily mediated by two virulence factors, toxins A (tcdA) and B (tcdB), which are released in the gut upon colonisation by *C. difficile* [3-5]. In the past decade, the epidemiology of *C. difficile* has changed and a new type emerged: polymerase chain reaction (PCR) ribotype (RT) 027/North American pulsed (NAP)-field type 01. Besides the production of toxins A and B, the binary *C. difficile* transferase toxin A/B (cdtA and cdtB) has probably contributed to the increased virulence of this type in addition to still unknown factors [6]. Major outbreaks due to this strain were reported since 2004, first in Canada followed by North America and Europe [7-10]. In 2008, PCR R078/NAP07-08 was reported as an emerging strain [11].

To study the epidemiology of *C. difficile*, several molecular typing methods have been introduced. Ideally, a typing method must have sufficient discriminatory power, typeability (the ability to type isolates unambiguously), reproducibility and transportability (the ability to perform the method reproducibly in a fully compatible fashion in different laboratories at different times) and must be relatively easy to perform [12]. In this review, we describe the most commonly used typing methods to characterise *C. difficile*. In addition, we present the latest developments in typing of *C. difficile*. Finally, we discuss the use of typing in surveillance studies, to trace outbreaks and to study strain transmission from the environment to patients.

**Historical perspective of Clostridium difficile typing**

Molecular typing methods can be categorised into two groups, phenotypic and genotypic methods. In the 1980s only phenotypic techniques were available. Serotyping using slide agglutination was commonly used in the mid-1980s. Initially, this assay was capable to differentiate six serogroups [13], later this
was improved to 15 serogroups [14]. Other commonly used methods in this period were autoradiography polyacrylamide gel electrophoresis (radio PAGE) [15] and immunoblotting using rabbit antiserum prepared from rabbits immunised with four different *C. difficile* strains [16]. Phenotypic assays had low reproducibility, low typeability and insufficient discriminatory power to apply to epidemiological studies [12]. Genotypic techniques with better typeability and discriminatory power replaced phenotypic methods during the 1990s [12]. Genotypic methods are divided into band-based and sequence-based methods. The most commonly used band-based methods were restriction endonuclease analysis (REA), pulsed-field gel electrophoresis (PFGE), capillary or conventional PCR ribotyping and multilocus variable-number tandem repeat analysis (MLVA), whereas the most frequently used sequence-based genotyping method was multilocus sequence typing (MLST). Recently whole genome sequencing (WGS) has emerged as a promising sequence-based technique as it allows the detection of variations between *C. difficile* strains by, for example, single nucleotide polymorphisms (SNPs) analysis. Here we present a brief summary of the current performance and costs of genotyping methods (Table 1 and 2), as a detailed description is beyond our scope and can be found in three other reviews on molecular typing [12,17,18].

### Currently used typing methods for *Clostridium difficile*

In Europe PCR ribotyping is presently the most frequently used typing method of *C. difficile*. This method was first applied by Gurtler et al. [21] and exploits the variability of the intergenic spacer region (ISR) between the 16S and 23S ribosomal DNA (rDNA), which is type-dependent. The variability, in combination with multiple copies of rDNA present in the genome, results in various amplicons after PCR amplification. These amplicons are separated by common agarose gel electrophoresis. The obtained banding patterns are referred to as PCR RTs. Two different sets of primers have been developed for typing of *C. difficile* [22,23]. The O’Neill primers described by Stubbs et al. [23] seem to have better discriminatory power than the Bidet primers [24]. The discriminatory power (D) of a typing method is its ability to distinguish between unrelated strains, this D-value is based on Simpson’s index of diversity [25]. PCR ribotyping is currently capable of identifying more than 400 distinct PCR RTs.

In North-America, PFGE is commonly used. PFGE of *C. difficile* involves digestion of genomic DNA with an infrequent cutting restriction enzyme, for example *SmaI* [26]. PFGE allows separation of large DNA fragments which is not possible with conventional agarose gel electrophoresis. The obtained DNA fragments are separated using agarose gel electrophoresis with an electric field orientation repeatedly switching in three different directions (pulsed-field); one direction is through the central axis of the gel, whereas the other two are at an angle of 60 degrees on either side. The pulse time of the direction is linearly increased during the run so that progressively larger fragments are able to migrate forward through the gel, resulting into separation based on fragment size. The obtained banding patterns are referred to as NAP-field types. Unfortunately, standardisation of protocols and validation of PFGE for *C. difficile* have never progressed as they did for other food-borne pathogens on PulseNet at the United States (US) Centers for Disease Control and Prevention (CDC) [27].

<table>
<thead>
<tr>
<th>Method</th>
<th>Target</th>
<th>Discriminatory power</th>
<th>Typeability</th>
<th>Reproducibility</th>
<th>Ease of interpretation</th>
<th>Technical complexity</th>
<th>Transportability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REA</td>
<td>Whole genome</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Moderate</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>PFGE</td>
<td>Whole genome</td>
<td>Moderate</td>
<td>Fair</td>
<td>Moderate</td>
<td>Fair</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>PCR ribotyping</td>
<td>16S–23S ISR</td>
<td>Good</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>Capillary PCR ribotyping</td>
<td>16S–23S ISR</td>
<td>Excellent</td>
<td>Moderate</td>
<td>Good</td>
<td>Moderate</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>MLVA</td>
<td>Whole genome, tandem repeats</td>
<td>Excellent</td>
<td>Poor</td>
<td>Moderate</td>
<td>Good</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Sequence-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLST 7HG</td>
<td>7 HG</td>
<td>Good</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Excellent</td>
<td>Moderate</td>
<td>Excellent</td>
</tr>
<tr>
<td>SNP typing</td>
<td>Whole genome, SNPs</td>
<td>Excellent</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Excellent</td>
<td>High</td>
<td>Good</td>
</tr>
</tbody>
</table>

HG: housekeeping genes; ISR: intergenic spacer region; MLST: multilocus sequence typing; MLVA: multilocus variable-number tandem repeat analysis; PCR: polymerase chain reaction; PFGE: pulsed-field gel electrophoresis; REA: restriction endonuclease analysis; SNP: single nucleotide polymorphism.

Table modified from Kuijper et al. [17].

---

**Table 1**

Performance characteristics of various genotyping methods for *Clostridium difficile*
It has been reported that PFGE displays better discriminatory power than PCR ribotyping with D-values of 0.843 and 0.688, respectively [18]. In contrast, preliminary results of a study comparing different typing techniques on 39 of the most frequently found PCR RTs in Europe demonstrate that only 16 NAP-field types were obtained of 39 PCR RTs (personal communications, M Mulvey and D McCannel, 2011). A common concern with all band-based typing methods is the difficult interpretation of DNA banding patterns, especially when a DNA banding pattern differs marginally from the reference patterns. Consequently, appropriate definitions are required to identify new types with both PFGE and PCR ribotyping. In Europe, the Cardiff collection of Jon Brazier and Val Hall serves as a reference collection and new PCR RTs are always validated using this database. Currently, a clinical collection of 20 different C. difficile PCR RTs (European Centre for Disease Prevention and Control (ECDC)-Brazier collection) isolated from various European countries is available to distribute among all reference laboratories in Europe who participate in the European C. difficile infection study network (ECDISnet) [28]. The usage of two different standard typing methods in Europe and America has resulted into different nomenclatures, making interlaboratory exchange of data difficult. Already in 1994 Brazier et al. [29] emphasised the need for a unified nomenclature.

In 2004, MLST was introduced to study the population structure and global epidemiology of C. difficile [30]. This sequence-based typing method relies on sequencing of DNA fragments approximately ranging between 300 and 500 bp representing seven housekeeping genes (MLST 7HG). Sequence variants for each housekeeping gene are assigned with a distinct allele number and the combination of seven allele numbers (allelic profile) provides a sequence type (ST). MLST generates high-throughput sequence data that can be uploaded from laboratories worldwide to a common web database [31]. This facilitates ST calling as well as studying the population structure and global epidemiology of C. difficile. Two different typing schemes have been proposed in literature to characterise C. difficile isolates [30,32]. Both typing schemes consist of seven housekeeping genes of which three are shared (triosephosphate isomerase (tpi), recombinase A (recA) and superoxide dismutase A (soda)). In contrast to the scheme published by Griffiths et al. [32], the MLST scheme described by Lemee et al. [30] was not widely adopted. This can be partially explained by the presence of a null allele on the D-alanine--D-alanine ligase (ddl) locus of the Lemee scheme which failed to amplify in certain strains [32]. Recently, this locus in the Lemee scheme was replaced by the groEL gene [33].

It has been reported that the discriminatory power of MLST and PCR ribotyping is comparable [18,32]. For studying outbreaks at a local level, a typing method should have higher discriminatory power than PCR ribotyping and MLST. For instance an increase in incidence of a PCR RT or MLST ST in a hospital can provide us with a clue for an outbreak and is useful data for monitoring changes in type prevalence rates, but does not necessarily proves clonal spread of one strain.

### Table 2: Techniques, time and costs associated with various genotyping methods for Clostridium difficile

<table>
<thead>
<tr>
<th>Genotyping method</th>
<th>Techniques</th>
<th>Turnaround time (post-culture)</th>
<th>Hands-on time (post-culture)</th>
<th>Equipment&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Per test&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>REA</td>
<td>DI, ER, GE</td>
<td>2 days</td>
<td>2 hours</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>PFGE</td>
<td>DI, ER, GE</td>
<td>2–4 days</td>
<td>6 hours</td>
<td>Moderate</td>
<td>Low</td>
</tr>
<tr>
<td>PCR ribotyping</td>
<td>DI, PCR, GE</td>
<td>1–1.5 days</td>
<td>2 hours</td>
<td>Low/ moderate</td>
<td>Low</td>
</tr>
<tr>
<td>Capillary ribotyping</td>
<td>DI, PCR, CE</td>
<td>1 day</td>
<td>2 hours</td>
<td>Moderate/ high</td>
<td>Low</td>
</tr>
<tr>
<td>MLVA</td>
<td>DI, PCR, CE</td>
<td>2 days</td>
<td>8 hours</td>
<td>Moderate/ high</td>
<td>Low/ moderate</td>
</tr>
<tr>
<td>MLST</td>
<td>DI, PCR, PPP, SE</td>
<td>4 days</td>
<td>8 hours</td>
<td>Moderate/ high</td>
<td>Moderate</td>
</tr>
<tr>
<td>SNP typing</td>
<td>DI, LP, TA, SE</td>
<td>5 days&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3 days&lt;sup&gt;d&lt;/sup&gt;</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

CE: capillary electrophoresis; DI: DNA isolation; ER: enzyme restriction; GE: gel electrophoresis; LP: library preparation; MLST: multilocus sequence typing; MLVA: multilocus variable-number tandem repeat analysis; PCR: polymerase chain reaction; PFGE: pulsed-field gel electrophoresis; PPP: PCR product purification; REA: restriction endonuclease analysis; SE: sequencing; SNP: single nucleotide polymorphism; TA: template amplification.

<sup>a</sup> Cost index for the equipment set-up: low < EUR 10,000 < moderate < EUR 100,000 < high.

<sup>b</sup> Cost index per test for materials: low < EUR 10 < moderate < EUR 100 < high.

<sup>c</sup> This estimated turnaround time is based on using Illumina Miseq benchtop sequencing [19].

<sup>d</sup> The hands-on time was determined by turnaround time substracted with the average runtime of the Illumina Miseq benchtop sequencer [20].
MLST is an appropriate tool for studying the phylogeny of \textit{C. difficile}. Compared to a band-based typing method, such as PCR ribotyping, MLST is less vulnerable to recombination events. Recombination in a housekeeping gene would change the allelic profile on a single locus only. Even though the consequence would be a change of ST, this new ST would still be closely related to the original ST maintaining the phylogenetic link. Recombination of repeats present in the ISR between the 16S and 23S rDNA [34] might lead to the formation of a novel PCR RT without a clear phylogenetic link. However, the rate at which these recombination events occur and the predisposing factors are unknown. Phylogeny reconstruction with MLST revealed that \textit{C. difficile} diversified into at least five well-separated lineages during evolution [32,35,36] and possibly a sixth monophyletic lineage [37]. The majority of STs were assigned to lineage 1 with no major subdivisions (Figure 1), but this result could be due to an unfortunate choice of housekeeping genes. Changing the housekeeping genes or adding housekeeping genes to the current MLST scheme might provide a better resolution of lineage 1.

A major advantage of sequence-based typing methods like MLST is the ease of interpretation of the generated data. Sequence data are unambiguous and therefore objective, highly reproducible and easily exchangeable between laboratories. Moreover, many laboratories have submitted their sequences to a freely accessible \textit{C. difficile} MLST database [31]. Currently (last updated: 21 Nov 2012), 176 different STs have been identified. A practical disadvantage of MLST remains the relatively high cost of sequencing multiple targets, which could partially explain why MLST has not replaced conventional PCR ribotyping in many European laboratories.

MLVA is a highly discriminatory molecular typing method that has been introduced to study outbreaks and identify routes of transmission between patients and hospitals [11,38–42]. MLVA relies on the amplification of short tandem repeats that vary in size and are dispersed throughout the genome. The obtained amplicons are separated with capillary electrophoresis followed by automated fragment analysis. Initially, two different typing schemes were published which both contain seven loci of which four are identical [41,42]. Each of the seven loci is designated with a number that corresponds to the sum of repeats present on that locus. A minimum spanning tree (MST) can be constructed, in which the summed tandem repeat difference (STRD) is used as a measure of genetic difference (Figure 2). Clonal clusters are defined by an STRD of ≤2, and genetically related clusters are defined by an STRD of ≤10 [11,41]. Broukhanski et al. [43] observed that two MLVA loci (F3 and H9) were invariant, indicating that loci F3 and H9 did not contribute to the discriminatory power. In addition, Bakker et al. [44] reported that MLVA locus A6 is a null allele in PCR RT078 and that for several other loci the PCR settings had to be optimised for PCR RT078. Invariance of MLVA loci requires optimisation and validation of MLVA for individual PCR RTs. Currently, MLVA has been implemented as useful typing method to investigate \textit{C. difficile} 027 outbreaks in the Netherlands, France and the United Kingdom (UK) [38,45,46]. In England, \textit{C. difficile} infection (CDI) cases that are potentially linked, i.e. caused by isolates that share the same PCR RT and which are related in time and place, are investigated using MLVA. Notably, almost half of such presumed clusters are shown actually either to consist of unrelated isolates or a mixture of related and distinct strains [46].

**Recent developments in typing of \textit{Clostridium difficile}**

**Variant multilocus variable-number tandem repeat analysis typing schemes**

Recently, a modified MLVA (mMLVA) was developed, combining MLVA with PCR detection of several toxin genes (tcdA and tcdB, cdtB; and deletions in the toxin C gene (tcdC)) [37]. In addition, the number of MLVA loci was restricted to five excluding the invariable loci F3 and H9. Although the combination with toxin gene detection can be informative, it is not yet possible to correlate these data with specific \textit{C. difficile} types, like PCR RT027/NAP01. This is partially because the presence of binary toxin genes combined with the 18 bp tcdC deletion is not restricted to PCR RT027 strains [37,47].

In a study by Manzoor et al. [48] the number of MLVA loci was increased to 15. This extended MLVA (eMLVA) scheme was able to discriminate clinically significant clusters while maintaining a good concordance with PCR ribotyping. Typing schemes containing only seven loci showed in contrast poor association with PCR ribotyping [41,42]. These seven loci schemes can only be used as a subtyping method together with PCR ribotyping, whereas the extended MLVA can potentially replace both. It should be noted, however, that increasing the number of loci makes the method more laborious and increases the difficulty of data interpretation.

Wei et al. [49] screened 40 MLVA loci for developing an MLVA typing scheme that has a good concordance with PCR ribotyping and provides satisfactory data for studying outbreaks. From this study, it was concluded that typing schemes consisting of MLVA loci with low allelic diversity maintained a high correlation with PCR ribotyping, whereas typing schemes using MLVA loci with high allelic diversity were required to study outbreaks. To fulfil both purposes two different typing schemes were proposed comprising 10 loci with limited allelic diversity and four loci with highly variable allelic diversity.

**Capillary polymerase chain reaction ribotyping**

Although PCR ribotyping has become widely used in many European laboratories for \textit{C. difficile} surveillance, issues with pattern interpretation and limited access to a well standardised database are
Figure 1
Phylogenetic structure of Clostridium difficile strains

NAP: North American pulsed-field; PCR RTs: polymerase chain reaction ribotypes; UDNAP: undefined NAP field type.

The figure is modified from Knetsch et al. [37]. The phylogenetic tree (radial tree layout) was constructed using a bayesian posterior probability method based on the alignment of concatenated DNA sequences of seven housekeeping gene loci. Six major lineages are shown in colour. The PCR RTs and NAP field types of the five most frequently PCR RTs in Europe are shown between brackets and in bold.
Multilocus variable-number tandem repeat analysis (MLVA) was used to recognise three different large local outbreaks in hospital G (orange), hospital A (blue) and hospital E (brown). Smaller outbreaks are indicated for hospital C (light yellow), hospital F (green) and related isolates from hospital B (purple) and hospital D (dark yellow). Clonal clusters are defined by a STRD of ≤ 2, and genetically related clusters are defined by an STRD of ≤ 10.
important limitations. The adaptation of PCR ribotyping to high resolution capillary gel electrophoresis (CE) PCR ribotyping has greatly improved pattern reproducibility and interpretation. For instance, using conventional agarose gel-based PCR ribotyping, it is difficult to differentiate types 014 and 020. In contrast, CE-PCR ribotyping can discriminate type 014 and type 020 and distinguish subtypes within type 014 [50]. However, the need for protocol standardisation remains evident. C. difficile surveillance laboratories from the CDC in the US, Public Health Agency of Canada (PHAC) in Canada, Leiden University Medical Center (LUMC) in the Netherlands and Leeds Teaching Hospitals NHS Trust in the UK are collaborating to develop and validate a standardised protocol for the DNA extraction, primer sets, PCR cycling conditions, and reference standards for CE-PCR ribotyping. The standardised consensus protocol is tested on a well characterised collection of 70 different PCR RTs [37] distributed to each of the four laboratories. Preliminary results show consistent fingerprints between the laboratories. Peakfile-based analysis is currently being optimised and validated, with a conclusion available by mid-2013.

Whole-genome single nucleotide polymorphism typing

High-throughput, WGS of bacterial pathogens has reached a scale and reliability to accurately define the natural history and global population structures of virulent and epidemic lineages [51–55]. Phylogenetic and comparative genome analysis of hundreds (soon to be thousands) of genomes can identify precise genetic changes, often linked to virulence and antibiotic resistance phenotypes, that can quickly inform about the pathogen’s biology. Whole genome sequencing can also distinguish between strains at the single nucleotide level, by comparing genomes in terms of single nucleotide polymorphisms, and therefore drastically improves the discriminatory power over conventional genetic typing methods. Thus, WGS has also (i.e. besides phylogeny) practical value for clinical microbiology and public health epidemiology by defining the selective forces that precipitate pathogen emergence and also by tracking transmission events ([56], Figure 3).

WGS approaches represent the ultimate pathogen typing method and, although its use and application remains limited to select facilities, we believe WGS will become a commonly used tool for C. difficile surveillance and epidemiology in the coming years. Although the cost of WGS is relatively high compared to traditional typing methods, sequencing costs are falling rapidly [19,57]. In addition, the ability to extrapolate MLST, PFGE, resistance gene, toxin gene sequence and other data from the same test could balance the cost-benefit analysis. Standardised computational pipelines are emerging for C. difficile genome data quality control and subsequent downstream analysis associated with informatics, phylogeny and phylogeography (Figure 3). Improved high-quality draft genomes [58] for the most

**Figure 3**

General sequencing and analysis strategy used to track genomic variants of Clostridium difficile at local and global levels

SNPs: single nucleotide polymorphisms.

Genomic DNA derived from Clostridium difficile isolates under study are subjected to sequencing with next generation sequencing technologies. Short read data from next generation sequencing platforms are mapped to reference genomes to determine the population level genome variation, such as SNPs, mobile element or other signatures of selection. Isolate sequences of interest are phylogenetically analysed. Combining phylogeny to epidemiological sequence data allows for inferences to be made about pathogen evolution and transmission events at healthcare and global level.
common *C. difficile* variants causing disease in human and animal populations [59] serve as references to map next generation sequence data in order to detect variation within the core genome (genes shared by all organisms) or the accessory genome (genes present in only some organisms) [60].

The first description of *C. difficile* PCR RT027 phylodynamics using high-throughput WGS demonstrated that 25 PCR RT027 isolates from the US and Europe could be further discriminated into 25 distinct genotypes based on SNP analysis [54]. Furthermore, this study demonstrated that isolates from different regions of the US and Europe occupy distinct evolutionary lineages and harbour unique antibiotic resistance genes. More recently, it was demonstrated that PCR RT027 isolates emerged through two distinct epidemic lineages after acquiring the same antibiotic resistance mutation; moreover these two lineages displayed different patterns of global spread [61]. The routine use of WGS in diagnostics and epidemiology is nicely reflected by the study of Kosor et al. [62]. In this study it was reported that whole-genome SNP typing can be mainly used for monitoring outbreaks and recognition of pathogen transmission pathways. Current methods for monitoring *C. difficile* hospital associated outbreaks, such as PCR ribotyping, have too limited discriminatory power to characterise potential outbreak strains as the same bacterial clone. Sequencing of whole genomes offers the optimal discriminatory power allowing laboratories to detect transmission pathways between hospitals, hospital wards and patients on the same ward.

In addition, Eyre et al. [19] demonstrated that WGS can produce practical, clinically relevant data in a time frame that can influence patient management and infection control practice during an outbreak. Moreover, this study demonstrated that a cluster of healthcare-associated *C. difficile* cases caused by the same ST was in fact a number of unrelated sub-lineages, therefore allowing to rule out in patient-to-patient transmission. Furthermore, WGS combined with comparative genomics is an effective approach to identify novel genetic markers that are potentially linked to virulence. This is an important advantage above conventional typing methods that use existing markers for characterisation of isolates. Whole genome sequencing is not likely to replace routine diagnostic techniques in reference laboratories. For example, matrix-assisted laser desorption/ionisation (MALDI) time-of-flight (TOF), which is rapid and easy to perform, is currently used in the Dutch reference laboratory for primary detection of pathogens.

In order to determine whether sequenced isolates are part of an outbreak, it must be defined how many SNP differences still represent ‘related’ isolates. For that reason, we should be informed on the rate of SNP accumulation in *C. difficile* lifecycle (molecular clock), although bacterial isolates with a hypermutator phenotype could complicate the determination of such a threshold [56]. The molecular clock rate of *C. difficile* was reported at 2.3 SNPs/genome/year in the study done by Eyre et al. [19]. Further study is necessary to confirm this rate of *C. difficile* evolution.

**Application of typing methods to study the epidemiology of *Clostridium difficile* infections**

An obvious reason to type *C. difficile* isolates is to early detect and investigate outbreaks, which can be defined as ‘a temporal increase in the incidence of a bacterial species caused by transmission of a certain strain’ [63]. In addition, typing methods contribute to epidemiological surveillance on national, European or worldwide level and can be used to report the incidence of various *C. difficile* types and recognise newly emerging virulent types [63]. Typing might also establish the local and global spread of bacteria and elucidate routes of transmission.

In the beginning of the 21st century, a worldwide increase in the incidence of CDI was seen. Soon thereafter, it was recognised that a specific type of *C. difficile*, PCR RT027, was linked to this increase of incidence [7,9]. PCR RT027 was associated with specific predisposing factors, course and outcome of CDI. In a large Canadian outbreak, fluoroquinolones were associated with PCR RT027 and mortality rates among patients with this type increased to 23% within 30 days of diagnosis [9,64]. In the Netherlands, molecular typing of *C. difficile* using PCR ribotyping contributed to recognition of an outbreak of two simultaneously occurring PCR RTs (027 and 017) [45]. Again, patients had PCR RT-specific risk factors and mortality rates. Numerous studies demonstrated the increased virulence of PCR RT027 [6–10] and found that other emerging types, such as PCR RT078, were also associated with specific risk factors or complicated clinical course [11]. Without results from typing methods, these associations would have stayed unrecognised.

Molecular typing results can also be used to compare the distribution of various *C. difficile* types isolated from animals, humans and food, which can hint towards food-borne disease or zoonotic potential of specific PCR RTs. The emerging *C. difficile* PCR RT078 in humans is found in high numbers in animals, especially piglets and calves [11,65–67]. Koene et al. [68] investigated the presence and characteristics of *C. difficile* in seven different animal species. PCR RTs 012, 014 and 078 were most frequently isolated among these Dutch animals, similar types were found among hospitalised patients in the Netherlands in 2009/2010. Meat consumption has also been suspected to contribute to transmission of *C. difficile*. PCR RTs 001, 017, 012 and 087 have been isolated from meat in Europe, however, isolation rates are low and might not be high enough to exceed the infectious dose [65–69]. Although PCR RTs in animals, meat and humans overlap, PCR ribotyping lacks discriminatory power to show clonal spread of *C. difficile* isolates from humans to animals. New
molecular methods should be developed and applied. The optimised MLVA scheme developed by Bakker et al. [44] showed relatedness between human and porcine PCR RT078 strains, although this could not always be confirmed with epidemiological data. Hopefully, highly discriminative typing methods such as whole-genome SNP typing can provide us with novel insights on zoonotic transmission.

Importance of molecular typing for national surveillance by reference laboratories

In Europe and North America, surveillance studies to monitor the incidence of CDI and the spread of hypervirulent strains have been established at regional and national levels since 2007 although reporting of CDI is not mandatory in all European Union (EU) countries. To enhance surveillance for CDI, the ECDC and the US CDC advised to widely launch surveillance programmes for CDI [28]. Consequently, a European network to support capacity building for standardised surveillance of CDI was initiated by the ECDC [28].

When methods and data on existing national CDI surveillance systems in Europe were reviewed (personal communication, A Kola, 2012), surveillance of CDI was reported in 45% (14/31) of the European countries. Active surveillance of CDI is performed in Austria, Norway, Belgium, Denmark, France, Germany, Ireland, Hungary, the Netherlands, Spain, Sweden, Luxembourg and the UK [46,70–79]. Surveillance was mostly continuous and prospective, but only four surveillance systems combined microbiological and epidemiological data (typing and susceptibility testing results) on a regular basis. A second recently completed survey in Europe (personal communication, D W Notermans, 2012) demonstrated that the majority of the laboratories were able to culture, but only half had access to typing. This limited typing capacity demonstrates the uncertainty of the true incidence levels of C. difficile types across Europe and hampers recognition of new emerging C. difficile types.

The contribution of national reference laboratories to survey CDI on a national level is illustrated by examples from the Netherlands and the UK. In 2005, soon after the emergence of C. difficile PCR RT027, the Center for Infectious Disease Control (Cib) of the National Institute for Public Health and the Environment (RIVM) in the Netherlands started a national Reference Laboratory for C. difficile. In 2009, this laboratory noticed an emergence of a new virulent PCR RT078, which was the third most frequently found type in the Netherlands among humans and was present in nearly all pig farms investigated [11,67]. Subsequently, this type was also found emerging in other European countries [80]. Recently, the reference laboratory noticed a re-emergence of C. difficile PCR RT027 since 2010. In the period between May 2011 and May 2012, 289 samples from 26 healthcare facilities and laboratories in the Netherlands were submitted because of severe CDI cases or outbreaks. PCR RTs 001 and 027 were the most commonly found (both 15.0%). Interestingly, in contrast to a previous report of declining PCR RT027 in hospitals in the Netherlands [81], type 027 was frequently identified in long-term care facilities associated with exchange of patients to neighbouring hospitals.

In the UK, the C. difficile Ribotyping Network (CDRN) was established in 2007, as part of improved CDI surveillance, to facilitate the detection and control of epidemic strains. Between 2007 and 2010, the CDRN received a large number of isolates (n=11,294) for PCR ribotyping. Typing results indicated that almost all of the 10 most common PCR RTs changed significantly during this time period [79]. As the proportion of CDI caused by PCR RT027 declined (from 55% to 21%), significant increases were observed in the prevalence of other C. difficile types, especially PCR RTs 014/020, 015, 002, 078, 005, 023, and 016. In addition, there was a 61% reduction in reports of C. difficile in England from 2008 to 2011, which occurred coincidently as the proportion of CDI caused by C. difficile PCR RT027 declined. Notably, the large reduction in incidence of C. difficile PCR RT027 cases has been paralleled by decreases in CDI related mortality [82]. The perceived success of the surveillance programme means that currently approximately a third of all CDI cases in England are referred to CDRN. CDI control programs should ideally include prospective access to C. difficile typing and analysis of risk factors for CDI and outcomes.

Future perspective

In the last fifteen years molecular genotyping methods have replaced some of the more traditional typing methods. WGS will dominate the field of molecular typing in the next decade. However, before WGS can be used as a routine tool for molecular typing some requirements need to be fulfilled. First, WGS needs to be fast, preferably within 48 hours. Furthermore, the technical workflow including data analysis needs to be simplified into an automatic pipeline. Finally, the costs for acquiring the technical and organisational platform needed to perform WGS must be reduced. Fulfilling these requirements, which is in our opinion a matter of time, would greatly increase the use of WGS worldwide.

Acknowledgments

This work was supported by ZonMw Grant 50-50800-98-079 from the Netherlands Organization for Scientific Research (NWO). We would like to acknowledge the European Study group of Clostridium difficile on behalf of the European Society for Clinical Microbiology and Infectious Diseases (E.J.K.) for their contribution. This work was funded by the European Centre for Disease Prevention and Control (ECDC) through the call for tender OJ/2010/07/09-PROC/2010/035.


