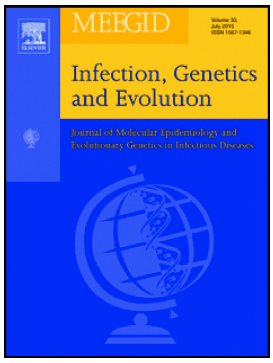

Clostridium difficile infection: Evolution, phylogeny and molecular epidemiology

Briony Elliott, Grace O. Androga, Daniel R. Knight, Thomas V. Riley

PII: S1567-1348(16)30547-0
DOI: doi: 10.1016/j.meegid.2016.12.018
Reference: MEEGID 3019

To appear in: *Infection, Genetics and Evolution*

Received date: 8 October 2016
Revised date: 19 December 2016
Accepted date: 19 December 2016

Please cite this article as: Briony Elliott, Grace O. Androga, Daniel R. Knight, Thomas V. Riley , Clostridium difficile infection: Evolution, phylogeny and molecular epidemiology. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Meegid(2016), doi: 10.1016/j.meegid.2016.12.018

Clostridium difficile infection: evolution, phylogeny and molecular epidemiology

Briony Elliott,[1] Grace O. Androga,[2] Daniel R. Knight,[2] Thomas V. Riley[1,2,3,4]

[1]School of Medical and Health Sciences, Edith Cowan University, Joondalup, Australia

[2]School of Pathology and Laboratory Medicine, The University of Western Australia, Crawley, Australia

[3]School of Veterinary and Life Sciences, Murdoch University, Murdoch, Australia

[4]Department of Microbiology, PathWest Laboratory Medicine, Perth, Australia

Corresponding author: Thomas Riley

Email: thomas.riley@uwa.edu.au

Address: 35 Stirling Hwy, Crawley, WA 6008, Australia

Keywords: *Clostridium difficile*, phylogeny, evolution, nosocomial infections, pathogenesis

ABSTRACT

Over the recent decades, *Clostridium difficile* infection (CDI) has emerged as a global public health threat. Despite growing attention, *C. difficile* remains a poorly understood pathogen, however, the exquisite sensitivity offered by next generation sequencing (NGS) technology has enabled analysis of the genome of *C. difficile*, giving us access to massive genomic data on factors such as virulence, evolution, and genetic relatedness within *C. difficile* groups. NGS has also demonstrated excellence in investigations of outbreaks and disease transmission, in both small and large-scale applications. This review summarizes the molecular epidemiology, evolution, and phylogeny of *C. difficile*, one of the most important pathogens worldwide in the current antibiotic resistance era.

INTRODUCTION

C. difficile is a spore-forming, anaerobic, Gram-positive bacterium found in the both environment and intestinal tracts of animals and humans. It is the most common cause of infectious healthcare-associated diarrhea, and has emerged as a leading nosocomial pathogen in developed countries. The increased incidence and severity of *C. difficile* infection (CDI) has led to a major economic burden on healthcare systems due to the costs associated with treatment, and extended stays of patients in hospital. This economic burden is estimated to be $5.4 billion in healthcare settings and $725 million in community settings in North America (Desai et al., 2016).

First described in 1935 following isolation from the faeces of neonates, the bacterium was named *Bacillus difficilis* to reflect the difficulty of culturing it (Hall and O'Toole, 1935). Although shown to be lethal in a number of species due to the production of an exotoxin (Hall and O'Toole, 1935; Snyder, 1937), it was only occasionally isolated in humans in whom its role as a pathogen was considered tenuous (Smith and King, 1962). Following the introduction of clindamycin, a broad-spectrum antibiotic with significant anti-anaerobic activity, colitis emerged as a serious complication associated with treatment (Tedesco et al., 1974). Prior to this, *Staphylococcus aureus* was regarded as the main etiological agent of antibiotic-associated colitis (Altemeier et al., 1963; Hummel et al., 1964; Khan and Hall, 1966; Wakefield and Sommers, 1953). When *S. aureus* was excluded as the cause of clindamycin-associated colitis, a race began to identify the organism responsible. Bartlett and colleagues were the first to suggest toxigenic *C. difficile* as the cause of clindamycin-associated colitis (Bartlett et al., 1978), while the presence of toxin similar to that of *C. sordellii* was noted by other groups (Larson and Price, 1977; Rifkin et al., 1977).

CDI is a toxin-mediated disease, characterised by diarrhoea. Symptoms range from mild to severe diarrhoea, pseudomembranous colitis (PMC), toxic megacolon, or even death. Despite the potential for severe disease, the majority of infected individuals remain asymptomatic (Donskey et al., 2015). The distinctive PMC lesions are usually limited to the colon, however, there have been cases with the small intestine involved (Jacobs et al., 2001; Keel and Songer, 2006). Extraintestinal infections are

rare (Byl et al., 1996). There are a number of risk factors involved in the development of CDI, including admission to healthcare facilities, advanced age (immune senescence), the presence of comorbidities and, principally, exposure to antibacterial agents (Bignardi, 1998). Almost all antibacterial agents have been implicated due to their effects on the intestinal microbiota. Disruption of the intestinal microbiota, typically but not only by antibiotics, is essential for the establishment of the organism and toxin production, following ingestion of spores (Kelly et al., 1994; Moore et al., 2013; Sorg and Sonenshein, 2008). Also required for disease development is the failure to mount an efficient immune response against *C. difficile* toxins (Sanchez-Hurtado et al., 2008). The primary virulence factors of *C. difficile* are toxin A (TcdA) and toxin B (TcdB), two closely related proteins belonging to the clostridial glucosylating toxin or Large Clostridial Toxin (LCT) family which target host small GTPases. At least 20% of *C. difficile* strains produce an additional binary toxin (CDT), an actin-specific ADP-ribosyltransferase. CDT has been associated with more severe disease (Barbut et al., 2005), but not proven to cause disease on its own (Eckert et al., 2015; Gerding et al., 2014). While strains that produce CDT in the absence of toxins A and B have recently been linked to symptomatic CDI in immunocompromised individuals, their pathogenicity in general still remains unclear (Androga et al., 2015; Eckert et al., 2015; Grandesso et al., 2016).

Treatment of CDI is preferably by oral administration of metronidazole or vancomycin (Kociolek and Gerding, 2016), if the inciting antibiotic cannot be stopped. These antibiotics also disrupt intestinal microbiota, leaving the patient susceptible to recurrent infection, either relapse or re-infection. Although non-antibiotic therapeutics such as toxin-binding agents showed initial promise, they have not performed as well as traditional treatments in clinical trials. One new antibiotic, fidaxomicin, has been associated with lower recurrence rates, but has not been widely adopted. Although probiotics have been investigated in preventing CDI, results have only shown limited success in preventing recurrences. There has been an increasing interest in faecal microbiota transplantation (FMT) as a means of restoring normal microbiota and preventing recurrent episodes in patients who have suffered multiple and debilitating recurrences. A number of other treatment options, such as immunological therapies, biotherapeutics, improved probiotics, phage therapy, and oral bacteriotherapy with non-

toxigenic *C. difficile* strains, may eventually offer better treatment solutions in preventing disease recurrence and treating initial episodes (Gerding, 2009, 2012; Kociolek and Gerding, 2016). Other more recent alternative therapeutic strategies for CDI include administration of monoclonal antibodies, more new antibiotics, molecular inhibitors e.g. quorum sensing and riboswitch ligands and new probiotics (Gerding, 2012; Hargreaves and Clokie, 2014; Vickers et al., 2016; Zanella Terrier et al., 2014).

Diagnosis of CDI can be quite challenging. A combination of algorithmic laboratory testing and clinical analysis is recommended. Cell cytotoxicity neutralisation assay was the initial mode of detection of *C. difficile* cytotoxin and still remains a gold standard, although laborious and time-consuming to perform (Chang et al., 1978). Enzyme immunoassay (EIA)s targeting TcdA were popular, however, had lower sensitivity and were unable to detect $A^{-}B^{+}$ strains. Although TcdA/TcdB EIAs became the norm, and in smaller laboratories remain popular, they have been largely replaced by molecular methods. There are currently numerous molecular tests to detect *C. difficile*, most of which target *tcdB*. With the exception of the Cepheid Xpert and Hain GenoType CDiff systems, however, there are currently no routine diagnostic methods for the detection of *C. difficile* strains that only produce CDT, largely because they are considered irrelevant clinically (Babady et al., 2010; Moore et al., 2013). Nonetheless, a CDT inclusive *C. difficile* detection method is relevant for surveillance purposes. The reader is directed to these reviews for an in-depth look into laboratory diagnosis of CDI (Collins et al., 2015; Rodriguez et al., 2016)

Previously a neglected pathogen, dramatic changes in the epidemiology of CDI over the last 15 years have led to a re-evaluation of this pathogen. This was largely driven by the emergence of a "hypervirulent" strain, NAP1/BI/027 (North American PFGE type 1, REA type BI, PCR ribotype 027) in North America in the 2000s associated with higher morbidity and mortality, which spread to other countries and caused outbreaks on a global scale. There have been other disturbing changes in CDI epidemiology. Although considered a nosocomial disease, CDI has also begun to emerge in the community, and in younger individuals who lack the traditional risk factors, often with a higher

incidence in females (Aronsson et al., 1985; McDonald et al., 2006; Rupnik et al., 2009). At the same time as this expansion of CDI in humans, there has also been a significant increase in animal disease caused by *C. difficile*. *C. difficile* is now the most common cause of enteritis in neonatal piglets in the USA (Songer and Uzal, 2005) as well as frequently causing diarrhoea in adult horses (Båverud et al., 2003). The increase in CDI in food animals has led to the suggestion that community-acquired CDI might be a foodborne disease (Weese, 2010) although this has yet to be proven. In support of this notion is the fact that ribotype 078, which is commonly isolated from food animals in the Northern Hemisphere, was the 3rd most frequent human isolate in a multi-country study in Europe published in 2011 (Bauer et al., 2011).

PHYLOGENY OF *C. DIFFICILE*

Taxonomy

Until a recent rearrangement of the Firmicutes, the *Clostridium* genus was paraphyletic and divided into clusters based on 16S rDNA analysis for convenience. *C. difficile* belonged to cluster XI, along with other closely related *Clostridium* spp. such as *C. sordellii* and *C. bifermentans*, as well as other species such as *Peptostreptococcus anaerobius* and *Eubacterium tenue* (Collins et al., 1994). After a brief incarnation as *Peptoclostridium difficile* (Yutin and Galperin, 2013), a name that was never "validly published", it has been renamed very recently as *Clostridioides difficile* (Lawson et al., 2016), and moved to within the *Peptostreptococcaceae* family (Ludwig et al., 2009). Whether the new name will be accepted by the *C. difficile* community around the world remains to be seen.

Clade structure

C. difficile consists of at least six phylogenetic clades: clades 1 through 5, and a sixth cryptic clade, named C-I. A recent examination of PubMLST data (http://pubmlst.org/cdifficile/) shows at least two other cryptic clades which we will term C-II and C-III (Figure 2). Clade 5, the most divergent of the non-cryptic clades is estimated to have diverged from the rest of the species between 1.1 and 85 million years ago (He et al., 2010). Given the lack of research in many geographical regions, there are

likely other clades that have yet to be discovered. Notable strains from each clade are shown in Table 1.

PaLoc and evolution

Description of PaLoc

The genes for the two main virulence factors of *C. difficile*, toxins A and B, are encoded on a 19.6-kb chromosomally-located element known as the Pathogenicity Locus (PaLoc) (Braun et al., 1996). Also encoded are three accessory genes: two putative regulatory genes (*tcdC* and *tcdR*) and a holin-like gene (*tcdE*) (Hundsberger et al., 1997) required for toxin-release (Govind and Dupuy, 2012) (Figure 3). Also downstream of the *tcdE* is a partial N-acetyl-muramoyl alanine amidase of unknown significance (Monot et al., 2011). In non-toxigenic strains, the PaLoc is replaced by a 115/85-bp sequence (Braun et al., 1996; Dingle et al., 2014). Although reminiscent of a Pathogenicity Island, the PaLoc lacks any known mobility genes and repeats at its borders. Large insertions (~9kB) unrelated to the PaLoc are found at the PaLoc integration site in strains from clade C-I (Dingle et al., 2014), and one clade 5 isolate (Elliott et al., 2008).

Other LCT elements

Other LCT members include the lethal and haemorrhagic toxins (TcsL and TcsH) from *C. sordellii*, and alpha toxin (TcnA) from *C. novyi*. There has been limited study into LCT-carrying elements from other *Clostridium* spp. In *C. novyi*, TcnA is encoded by a phage situated on a plasmid (Skarin and Segerman, 2014), whereas both *C. difficile* and *C. sordellii* have two toxins arranged on very similar elements. TcsH and TcsL are homologous and antigenically cross-reactive with TcdA and TcdB, respectively. The *C. sordellii* and *C. difficile* PaLoc elements are closely related enough to undergo allelic exchange (Elliott et al., 2014), and probably share a common ancestor. The two toxins are thought to have evolved from a gene duplication event.

Variant forms

The most common variant types of the PaLoc are the toxin A-negative, toxin B-positive (A^-B^+) versions. Although both toxin genes contain repeats within the receptor-binding domain region, they are better conserved within *tcdA*, making it more susceptible to deletion due to recombination. The most common A^-B^+ strain is RT 017 (ST 37) which has a 1.8 kb deletion that abrogates that function of the repetitive receptor-binding domain (Kato et al., 1999). A second receptor-binding domain recently discovered upstream (Lambert and Baldwin, 2016) is not affected however, so the RT 017 toxin A may still be active. The toxin A gene of toxinotypes VI and VII possesses smaller deletions due to recombination of repeats, which do not affect function (Rupnik et al., 1998), but do result in the loss of an epitope and a decrease of immunogenicity *in vitro* ((Barbut et al., 2002)). The much larger deletions resulting in loss of most or all of toxin A are the result of the activity of mobile elements, as indicated by DNA remnants at deletion sites (Elliott et al., 2014; Geric Stare and Rupnik, 2010). The largest deletion known is that occurring in toxinotype XI strains, where only *tcdC* and a fragment of *tcdA* remains (Geric Stare and Rupnik, 2010). Several intact mobile elements have also been identified within the PaLoc. Toxinotypes XIVa, Xb, XXII and IXc, all belonging to clade 2, possess a large 2000 bp IStron in *tcdA* (Geric et al., 2004; Mehlig et al., 2001), a type of mobile element first discovered in *C. difficile* (Hasselmayer et al., 2004b). The clade 3 PaLoc has a stably integrated transposon Tn6218 which does not interfere with toxin production (Dingle et al., 2014). Strains producing just toxin A have only described relatively recently and carry an unusual monotoxin locus consisting of toxin A and the holin UviB instead of *tcdE*, found elsewhere in the genome than at the PaLoc integration site (Monot et al., 2015). In toxinotype XXXII strains, which belong to clade C-II, a variant version of the PaLoc lacking *tcdA* and *tcdC* is also found at an alternative location in the genome (Janezic et al., 2015b).

Toxin A and toxin B are heterogeneous at a sequence level, and such variation can give rise to changes in immunoreactivity, substrate specificity and visible cytopathic effect (Rupnik, 2008). Due to the large size of the toxins, and the difficulty in resolving the repeats in *tcdA*, the study of such variation has relied mostly on the PCR-RFLP-based method toxinotyping, which currently describes 34 toxinotypes (I-XXXIV) (Rupnik and Janezic, 2016). Variations with the most visible effects are

those that occur in a small region of the catalytic domain known as the substrate specificity region, and alter the pattern of intracellular targets. This results in two types of cytopathic effect, known as "difficile-like" and "sordellii-like". The traditional difficile-like cytopathic effect is characterised by cell rounding accompanied by long protrusions. With the sordellii-like cytopathic effect, so named because it resembles that seen with TcsL and TcsH, complete cell rounding occurs without the protrusions (Chaves-Olarte et al., 1999). The sordellii-like effect is dependent on the ability to glucosylate members of the Ras family of small GTPases in addition to those from the Rho family (Chaves-Olarte et al., 2003).

Toxigenic status and clade

Clades 1, 2, 3 and 5 consist mainly of toxigenic strains, whereas clades 4 and C-I remain largely non-toxigenic. Less is known about the other cryptic clades, but clade C-II contains ST 200, toxinotype XXXII (A^-B^+) (Janezic et al., 2015a). Clade C-III, represented by ST 204, is non-toxigenic (Kuwata et al., 2015). Clades 2, 3 and 5 strains also often carry binary toxin genes and a variant form of binary toxin has been identified in clade C-I (Monot et al., 2015).

Acquisition of the PaLoc

The PaLoc is capable of being transferred horizontally in vitro, however, this occurs via homologous recombination, typically accompanied by large flanking regions (Brouwer et al., 2013). The different clades of *C. difficile* appear to have acquired the PaLoc separately after divergence. Clade 4, which comprises mainly non-toxigenic STs, has acquired the PaLoc relatively recently, estimated at ~500 years ago (Dingle et al., 2014). The cryptic clades have largely been regarded as non-toxigenic, attributed to their lack of a PaLoc integration site, at least in clade C-I (Dingle et al., 2014). The PaLoc can also be lost via homologous recombination with non-toxigenic strains (Dingle et al., 2014).

CdtLoc

Some strains produce an additional actin-specific ADP-ribosyltransferase, known as binary toxin or CDT. The toxin consists of an enzymatic component and a binding component, CdtA and CdtB,

respectively, and belongs to the family of clostridial binary toxins that also includes *C. spiroforme* CST and *C. perfringens* iota toxin. The *cdtA* and *cdtB* genes are situated on a 6.2 kb chromosomally-located element known as the CdtLoc, along with *cdtR*, a regulatory gene required for expression (Carter et al., 2007). The CdtLoc is similar to the PaLoc in the sense that in contains no known genes associated with mobility, lacks direct repeats at its termini, and is replaced by a 68 bp sequence in strains that lack the element (Carter et al., 2007). A partial form of the CdtLoc (4.2 kb) containing an intact *cdtR* alongside fragments of *cdtA* and *cdtB* is common within clade 1, but does not occur in the other binary toxin-negative PaLoc-positive clade, clade 4 (Geric Stare et al., 2007). The presence of an intact CdtLoc is associated with changes in the PaLoc, and these occur in clades 2, 3 and 5, as well as in clade C-I (Monot et al., 2015). Interestingly, CdtR has been shown to regulate the production of toxins A and B, but only in some strains (Lyon et al., 2016).

Other virulence factors

Virulence factors of *C. difficile* other than toxins A and B, and binary toxin have had very little study, but have been recently reviewed elsewhere (Awad et al., 2014). Those that have attracted study are mostly cell-surface structures such as capsules, the S-layer, and other potential adhesins. *C. difficile* can possess an extracellular polysaccharide capsule, although its significance in the pathogenesis of disease is unclear (Baldassarri et al., 1991; Strelau, 1989). In addition to the capsule, *C. difficile* has an S-layer, a proteinaceous paracrystalline array composed of two components, a high molecular weight component, and low molecular weight component, both formed from post-transcriptional cleavage of the *slpA* gene (Calabi et al., 2001). Also present in the S-layer are up to 28 paralogous proteins (Fagan et al., 2011) which have been implicated in a number of virulence-related activities including adhesion, biofilm formation and degradation of host proteins (Cafardi et al., 2013; Pantaleon et al., 2015; Waligora et al., 2001). Flagella and motility are not essential for the pathogenesis of CDI as they are absent from clade 5 strains, but have been shown to play a role in colonisation in clade 2 strains (Stevenson et al., 2015).

Regulation of virulence

Expression of the LCTs and other virulence factors is regulated by the quorum-sensing-controlled accessory gene regulator (Agr) locus (Darkoh et al., 2016). Some *C. difficile* strains encode a second copy of the *agr* locus (*agr-2*), such as the hypervirulent NAP1/027 strain, but most strains, including str. 630, encode one copy, known as *agr-1* (Darkoh et al., 2016). The classical Agr system, best described in *S. aureus*, is a four gene locus (agrABCD). It encodes AgrD, a prepeptide that is processed by AgrB to form a mature autoinducing peptide (AIP) which is excreted into the intracellular environment. When there is a sufficient number of bacteria, and a thus a sufficiently high concentration of AIP for the histidine kinase AgrC to bind, AgrC activates the response regulator AgrA which regulates transcription. Whereas the *agr*-1 locus in *C. difficile* comprises of agrBD lacking the response genes, *agr-2* has the full complement of genes. Deletion of the *agr-1* locus results in loss of the ability to express toxin A and toxin B and the ability to cause disease in the murine model (Darkoh et al., 2016). Toxin expression occurs again upon either the restoration of *agr-1* by complementation, or by exposure to purified AIP in cell culture (Darkoh et al., 2016). The induction of toxin expression by AIP in *agr-1* mutants of strain 630, suggests that although *agr-1* lacks the response genes *agrAC*, there must exist an alternative sensor and response genes elsewhere in the genome. A number of other virulence genes are under control of the *agr-2* locus in the NAP1/027 strain, with mutants poorly flagellated and less able to colonise (Martin et al., 2013). Interestingly, alternative *agr* loci have also been detected in *C. difficile* bacteriophages, although the significance of these is unknown (Hargreaves et al., 2014b).

Mechanisms driving evolution

Genome architecture and mobile elements

C. difficile is characterised by a large open pan-genome comprised of a small core component and a large number of accessory genes. The pan-genome is the entire repertoire of genes carried by the species, comprising of the core genome, consisting of genes found in all strains, and accessory genes present in only some strains (Tettelin et al., 2008). It has been estimated that the core genome may be as low as 16%, the lowest for any bacterial species (Janvilisri et al., 2009; Scaria et al., 2010).

Using microarray and a small but diverse collection of isolates, Scaria et al. estimated the core genome was composed of 947 to 1033 genes, while the pan-genome comprised of 9640 genes (Scaria et al., 2010). The genome is also comprised of a large proportion of mobile elements. In the first *C. difficile* genome to be sequenced, str. 630 (RT 012), mobile elements accounted for ~10% of the genome and included a plasmid, prophages, transposons, IStrons, genomic islands, CRISPRs and a *skin* element (Sebaihia et al., 2006). This has remained consistent in the annotated genomes of other strains published since, including M120 (clade 5), and R20291 (clade 2) (He et al., 2010; Stabler et al., 2009).

The genome of str. 630 contained seven conjugative transposons (CTn1–CTn7), and one mobilisable transposon Tn*5398* (Sebaihia et al., 2006). The conjugative transposons are divided into two main families based on their conjugation modules, the Tn*916* family, and the Tn*1549* family (Sebaihia et al., 2006). CTn3 (Tn*5397*) and Tn*5398* mediate tetracycline and macrolide-lincosamide-streptogramin B (MLS_B) resistance, respectively. Another transposon, Tn*4453*, has been found to mediate chloramphenicol resistance in *C. difficile* (Lyras et al., 1998). Other than resistance arising from mutations in target genes, transposons appear to be the most common mechanism of acquired antibiotic resistance in *C. difficile*. Numerous other transposons have been identified in *C. difficile* genomes (Brouwer et al., 2012; Brouwer et al., 2011), most of which are cryptic in function, with some capable of intra- and interspecies transfer, most notably the *vanB2* Tn*1549*-like element in strain AI0499 (Knight et al., 2016a)

Most investigation of plasmids in *C. difficile* has been for epidemiological investigations and typing purposes. Although early typing schemes based on plasmids found them to be plentiful, with anywhere up to six plasmids per strain (Mulligan et al., 1988), this has not been borne out by genome sequencing. The very limited number of plasmids that have been sequenced, such as the plasmid from str. 630, pCD630, have all been cryptic in function (Sebaihia et al., 2006). Genes mediating virulence

or antibiotic resistance have not been identified on *C. difficile* plasmids as they have in other *Clostridium* spp.

Prophages reported in *C. difficile* so far belong to the Myoviridae or Siphoviridae family, although myoviruses tend to dominate (Hargreaves et al., 2013; Shan et al., 2012). There has been some evidence that prophages are involved in regulation of toxin production, possibly via quorum sensing loci (Goh, 2005; Sekulovic et al., 2011), but they do not carry virulence genes. They can be involved in the transduction of other elements, as phiC2 is with the *ermB*-carrying Tn*6215* (Goh et al., 2013). Nearly all detected prophages have a GC content not dissimilar to that of the *C. difficile* genome (28-30%), and putative integrase genes suggesting they have access to the lysogenic lifestyle (Knight et al., 2015). Many phages also have CRISPR arrays of their own suggesting long term evolution with the host *C. difficile* and defence against secondary phage infection (Hargreaves et al., 2014a).The IStron, is a hybrid of a group I intron and an insertion element (Hasselmayer et al., 2004a; Hasselmayer et al., 2004b) apparently particular to *C. difficile*. Insertion in genes does not result in their disruption, as it excises from the transcript.

Rate of evolution

The molecular clock, the rate at which an organism mutates, is essential for examining the relatedness of isolates, and investigating transmission. The within-host evolutionary rate of *C. difficile* has been calculated at 3.2×10^{-7} mutations per nucleotide per year (95% CI, 1.3×10^{-7} to 5.3×10^{-7}), equating to approximately 1.4 mutations per genome annually (Didelot et al., 2012). Similar rates have been calculated for RT 078 and RT 027 (He et al., 2013; Knetsch et al., 2014). These rates are much higher than the estimated long-term evolutionary rate (He et al., 2010). This difference may be due to the effects of purifying selection purging deleterious mutations over time. Given that *C. difficile* can exist in the environment as a dormant endospore for long periods of time, a rate calculated from strains replicating in hosts may not be relevant over larger time scales.

Homologous recombination

The rate of homologous recombination (r) in *C. difficile* is estimated to be relatively low, and is estimated to have effect ~4 times lower than point mutation (m), (r/m=0.25) (Dingle et al., 2011; Vos and Didelot, 2009). This rate could be an underestimation, however, if barriers between lineages existed, such as geographical separation. The PaLoc itself shows evidence of mosaicism resulting from homologous recombination between lineages of *C. difficile*, and even with *C. sordellii* (Dingle et al., 2014; Elliott et al., 2014). Investigation of PaLoc gain/loss events shows that very large regions (>200 kb) of DNA can be involved in homologous recombination (Dingle et al., 2014), and probably utilise the conjugative machinery of other elements (Brouwer et al., 2013). The ST-122 lineage is the result of a massive recombination event between a clade 1 strain and a clade 2 strain, as is seen in the combination of a non-variant PaLoc alongside CdtLoc, and the combination of MLST alleles present (Dingle et al., 2014). It is thus not a true independent clade as some have suggested (Knetsch et al., 2012), but a hybrid clade. Homologous recombination is an important driver of diversity in the S-layer cassette, which consists of *slpA* and 11 of the 28 S-layer paralogs (Dingle et al., 2013). Recombination is so common in this cassette that the S-layer type does not correlate with core genome, or even phylogenetic clade (Dingle et al., 2013). This S-layer is dominant at the host-bacteria interface, and likely under great selective pressure, with switching analogous to polysaccharide switching seen in other species.

EPIDEMIOLOGY OF *C. DIFFICILE* INFECTION

Typing methods

Typing is not only important for understanding the epidemiology of CDI on a larger scale, but investigating outbreaks and transmission in general. A wide array of methods has been used to type *C. difficile* isolates, and many remain in routine use around the world. Initial methods for *C. difficile* strain typing and outbreak investigations were based on phenotypic characteristics. They included antibiogram patterns, serotyping, immunoblotting, plasmid analysis, bacteriophage and bacteriocin susceptibility patterns, soluble protein pattern using polyacrylamide gel electrophoresis (SDS-PAGE), autoradiography (Radio PAGE) and pyrolysis mass spectroscopy (PyMs) (Brazier, 1998; Cohen et al., 2001; Delmée et al., 1985; Tabaqchali et al., 1986; Toma et al., 1988; Wust et al., 1982). These

methods were pivotal in determining the epidemiology of local outbreaks of CDI, however, they did not have the capacity for investigating epidemiology at a larger scale (Brazier, 1998). Genetic fingerprinting of chromosomal *C. difficile* DNA became a substitute for phenotypically indistinguishable isolates (Kuijper et al., 1987). Restriction endonuclease analysis (REA), which uses different restriction enzymes, became a popular method for demonstrating strain diversity and patient cross-infection and/or transmission (Devlin et al., 1987; Johnson et al., 1990; Kuijper et al., 1987; Peerbooms et al., 1987; Wren and Tabaqchali, 1987). Although highly discriminatory and reproducible, REA is complex to interpret due to the large number of bands produced. Restriction length polymorphism (RFLP) of amplified 16S rDNA, also known as ribotyping, was not as laborious as REA, but was less discriminatory (Bowman et al., 1991; O'Neill et al., 1993). In 1992, McMillin & Muldrow recommended the use of arbitrarily primed PCR (AP-PCR), also known as random amplified polymorphic DNA (RAPD)-PCR, to type *C. difficile* isolates and this was successfully adopted by numerous investigators (Barbut et al., 1993; Killgore and Kato, 1994; McMillin and Muldrow, 1992; Wilks and Tabaqchali, 1994).

During the same period PCR ribotyping involving amplification of the 16S-23S intergenic spacer region (ISR) was introduced (Gurtler, 1993). This was then modified by Cartwright *et al* and O'Neil *et al* for easier routine use and is still widely used for *C. difficile* typing in Europe, Australia and Asia (Cartwright et al., 1995; O'Neill et al., 1996). Currently, Public Health England's *C. difficile* Ribotyping Network (CDRN) maintains the reference library, but patterns are not readily available. A move away from agarose-based resolution to capillary-based resolution on sequencers has made the exchange of patterns easier, but assigning a universally recognized type is still hampered by access to reference patterns and strains.

PFGE is the primary method used for *C. difficile* typing in North America (Knetsch et al., 2013). Although initially plagued by untypeable strains due to the instability of *C. difficile* DNA, this has now largely resolved by the inclusion of thiourea (Corkill et al., 2000). Multilocus variable number of tandem repeats (VNTR), also known as multilocus variable analysis (MLVA), has been used as a

method of discriminating strains with identical PFGE types or PCR ribotypes, but has been largely superseded by sequencing-based technologies.

MLST is based on the assignment of alleles based on nucleotide sequence of multiple housekeeping genes. It is best used to represent phylogenetic diversity and genetic relationships e.g. potential zoonosis (Maiden et al., 1998). An initial *C. difficile* MLST scheme (Lemee et al., 2004) was later revised and applied to several large-scale *C. difficile* epidemiological studies (Griffiths et al., 2010). MLST has the advantage that it can easily be performed *in silico* from whole genome sequence data.

With the advent of more affordable sequencing, typing methods that utilise whole genome sequencing data have become feasible for routine use. Analysis of single nucleotide polymorphisms, or SNP typing, allows the high resolution of relationships between strains. Whole genome sequencing also generates a large amount of data that can be used to assess the carriage of genes involved in antibiotic resistance and virulence. Current challenges are the expertise required to process the data appropriately and interpret it correctly. Another issue is back-translating it to current typing systems. Unfortunately PCR ribotyping, the most widely used typing method currently in use, is based on a repetitive element (the 16S-23S intergenic region) that cannot be resolved readily with current read lengths. With the array of typing methods available for *C. difficile*, there is need for a universal standardized typing method that allows easy exchange or comparison of types between laboratories, improving our knowledge regarding the epidemiology of *C. difficile* on a global scale. Whole genome sequencing is the future for *C. difficile* typing for both evolutionary and epidemiological investigations, and for routine clinical pathology laboratories as well. It can still be improved with a more robust molecular clock, as estimates of mutation rates are needed for each lineage and different hosts. Whole genome sequencing will only improve with the introduction of long read sequencing chemistries such as the Nanopore and PacBio (Hargreaves et al., 2016).

Global epidemiology

Most of the data on the molecular epidemiology of *C. difficile* and CDI has been derived from the United Kingdom, North America, and Australia. Thus, there is lack of representativeness in many of the strain collections that have been used to examine the epidemiology of *C. difficile* at a global level, with many areas of the world lacking isolates let alone typing data. The use of different typing methods (e.g. PCR ribotyping in Europe and Australia, and PFGE in North America) makes the comparison of typing data difficult. Even with PCR ribotyping, it is difficult to compare data between laboratories without reference strains. PCR ribotyping data from the UK shows that circulating strains change over time (Health Protection Agency, 2011; Public Health England, 2016). Much of the systematically collected typing data in the UK was collected following the introduction of RT 027 into the UK (Brazier et al., 2007), where it dominated in the following years. A small collection of strains from between 1979 and 2004 in South East Scotland found RTs 002, 014, 012, 015, 020 and 001 were the most common strains (Taori et al., 2010). A month-long survey in 2008 of 34 countries found that RTs 014/020, 001, 078, 018, 106, 027 and 002 were the most common types circulating in Europe at the time (Bauer et al., 2011). Because of the difficulties with ribotyping, there has been a move towards sequencing and *in silico* MLST as a substitute (Dingle et al., 2011). The limited amount of whole genome sequencing-based epidemiological data currently available nearly all comes from Oxfordshire, UK (Didelot et al., 2012; Dingle et al., 2014; Dingle et al., 2011; Eyre et al., 2013), although there have been smaller scale studies performed elsewhere in the world.

Much of the typing data from North America is PFGE or REA-based, however, a study over 2011–2013 found the most common ribotypes among 32 USA hospitals to be RTs 027, 014/020, 106, 001, 053 and 002 (Tickler et al., 2014). No large-scale study of Asian typing data has yet to be published, but a comprehensive review of the epidemiology of CDI in Asia found RT 017 to be predominant in many countries, with RTs 018, 002 and 001 also common (Collins et al., 2013). In an Australian survey completed in 2010, the five most common ribotypes were RTs 014/020, 002, 054, 056 and 070 (Cheng et al., 2016). Data from other regions such as South America, Africa, and the Middle East is largely limited to single hospital studies or small multicentre studies.

While there are some strains that seem to be universally successful, particularly in hospitals (RTs 014, 020, 002), there do seem to be regional differences. Certain phylogenetic clades appear to be associated with particular regions of the world in a type of phylogeographic tropism. Although RT 017, one of the few toxigenic strains within clade 4, has been associated with outbreaks across the globe (van den Berg et al., 2004), it has been consistently found in Asia. More tellingly, non-toxigenic strains of clade 4, which lack the selective mechanisms of disease-causing strains, are almost without exception reported in Asia. Interestingly, toxinotype XXI (A^+B^+) which is apparently the parent strain of RT 017, and the first of clade 4 to undergo recombination, is still in existence (Janezic et al., 2015a), but has been nowhere as successful as its progeny.

Clade 5 is strongly associated with Australia, with the greatest diversity of clade 5 strains found there. The most prominent ribotype 078 (ST 11), is a global strain however, particularly associated with animals, but emerging in humans (Bauer et al., 2011). In other areas, RT 078 related strains such as RTs 033, 126 and 127 are commonly found in animals (Janezic et al., 2014). It is possible that the divergence of the clades occurred following continental separation after the breakup of Pangaea. *C. difficile* is an ancient species that evolved prior to the evolution of animals let alone humans. Although *C. difficile* does colonise human neonates, such a transient relationship is unlikely to have favoured its spread along ancient human migration patterns. Until the introduction of broad-spectrum antibiotics and its emergence of a pathogen in the context of global travel, the movement of the clades was probably quite restricted. However, the acquisition of the PaLoc by clade 4 from another clade approximately 500 years ago (Dingle et al., 2014) does demonstrate that some spread was occurring in the past.

Emerging strains

In the early 2000s, an increase in the severity and incidence of CDI in North America was attributed to the emergence of a previously rare strain NAP1/BI/027 (Loo et al., 2005; McDonald et al., 2005). Although this strain was first detected in Quebec, Canada, it had likely originated from across the border in the USA at about the same time (Pepin et al., 2004). It quickly spread globally via two main

lineages causing outbreaks in diverse geographic areas (He et al., 2013). Unlike historic RT 027 isolates, this strain was fluoroquinolone-resistant, and its association with higher morbidity and mortality led to being labelled "hypervirulent" (Loo et al., 2005). Because the strain possessed a variant *tcdC*, the putative negative regulator of toxin production, that had a nonsense mutation as well as distinctive 18bp deletion (Loo et al., 2005), it was initially assumed that increased toxin production was responsible for the apparent increase in virulence. Although initial data did show an increase in toxin production (Warny et al., 2005), these results were eventually disproved by allelic exchange experiments showing *tcdC* had no effect on toxin production (Cartman et al., 2012). A microarray study identified a region that was either absent or divergent in the receptor-binding domain of the RT 027 strain (Stabler et al., 2006), which was shown to be the result of significant polymorphisms as opposed to a deletion (Stabler et al., 2008). Lanis et al. demonstrated that this variant TcdB displayed increased toxicity *in vivo* due to a broader cell tropism and an ability to enter the cytosol of cells at an earlier stage of endocytosis (Lanis et al., 2010; Lanis et al., 2013). The use of third-generation fluoroquinolones was the likely driving factor behind the emergence of fluoroquinolone-resistant RT 027 (Clements et al., 2010). In Australia, where such drugs are used infrequently, the strain did not become established, with only a limited number of imported cases reported (Riley et al., 2009). In the wake of RT 027, several other clade 2 strains have emerged, some of which have also been associated with increased disease severity. These include the emergence of RT 244 in Australia and New Zealand (De Almeida et al., 2013; Eyre et al., 2015; Lim et al., 2014), RT 176 in parts of Europe (Polivkova et al., 2016), and RT 251 in Australia (unpublished data). What is driving the emergence of these strains is unclear, however, their increased virulence may also be due to polymorphisms in the *tcdB* RBD, in which clade 2 has the highest diversity of all the clades (Dingle et al., 2014).

Another emerging strain, long associated with animals is RT 078 (Bakker et al., 2010; Rupnik et al., 2008). It has been associated with a wide range of species, including several production animals (Moono et al., 2016). Previously rare in humans (Health Protection Agency, 2011), it has risen to become the third most common strain in Europe (Bauer et al., 2011) and linked to increased mortality (Goorhuis et al., 2008). The strong association with animals has led to the suggestion of zoonotic

transmission (Hensgens et al., 2012; Lund and Peck, 2015). Detection of *C. difficile* in meat is rare (Knight et al., 2016b; Lund and Peck, 2015), however, and it may simply be that the rise of this strain is due to amplification in production herds and contamination of the environment (Casey et al., 2015).

CONCLUSIONS

C. difficile remains a formidable and poorly understood pathogen, with pathogenesis of infection the result of a complex interplay between host and bacterium. Although there has been a large amount of genomic data generated in recent years, allowing insight into the evolution and phylogeny of *C. difficile*, it is not representative of the global population. The species has diversified into at least eight lineages, with others probably yet to be discovered. The genome of *C. difficile* is characterised by a small core component, and a large proportion of mobile elements, giving the organism potential access to a large repertoire of genes. The emergence of new multidrug-resistant hypervirulent strains highlights the need for further work and constant surveillance of *C. difficile*.

ACKNOWLEDGEMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Altemeier, W.A., Hummel, R.P., Hill, E.O., 1963. Staphylococcal enterocolitis following antibiotic therapy. Ann Surg 157, 847-858.

Androga, G.O., Hart, J., Foster, N.F., Charles, A., Forbes, D., Riley, T.V., 2015. Infection with toxin A-negative, toxin B-negative, binary toxin-positive *Clostridium difficile* in a young patient with ulcerative colitis. J Clin Microbiol 53, 3702-3704.

Aronsson, B., Mollby, R., Nord, C.E., 1985. Antimicrobial agents and *Clostridium difficile* in acute enteric disease: epidemiological data from Sweden, 1980-1982. J Infect Dis 151, 476-481.

Awad, M.M., Johanesen, P.A., Carter, G.P., Rose, E., Lyras, D., 2014. Clostridium difficile virulence factors: Insights into an anaerobic spore-forming pathogen. Gut Microbes 5, 579-593.

Babady, N.E., Stiles, J., Ruggiero, P., Khosa, P., Huang, D., Shuptar, S., Kamboj, M., Kiehn, T.E., 2010. Evaluation of the Cepheid Xpert *clostridium difficile* Epi assay for diagnosis of *Clostridium difficile* infection and typing of the NAP1 strain at a cancer hospital. J Clin Microbiol 48, 4519-4524.

Bakker, D., Corver, J., Harmanus, C., Goorhuis, A., Keessen, E.C., Fawley, W.N., Wilcox, M.H., Kuijper, E.J., 2010. Relatedness of human and animal *Clostridium difficile* PCR ribotype 078 isolates determined on the basis of multilocus variable-number tandem-repeat analysis and tetracycline resistance. J Clin Microbiol 48, 3744-3749.

Baldassarri, L., Donelli, G., Cerquetti, M., Mastrantonio, P., 1991. Capsule-like structures in *Clostridium difficile* strains. Microbiologica 14, 295-300.

Barbut, F., Decre, D., Lalande, V., Burghoffer, A., Noussair, L., Gigandon, A., Espinasse, F., Raskine, L., Robert, J., Mangeol, A., Branger, C., Petit, J.C., 2005. Clinical features of *Clostridium difficile*-associated diarrhoea due to binary toxin (actin-specific ADP-ribosyltransferase)-producing strains. J Med Microbiol 54, 181-185.

Barbut, F., Lalande, V., Burghoffer, B., Thien, H.V., Grimprel, E., Petit, J.C., 2002. Prevalence and genetic characterization of toxin A variant strains of *Clostridium difficile* among adults and children with diarrhea in France. J Clin Microbiol 40, 2079-2083.

Barbut, F., Mario, N., Delmee, M., Gozian, J., Petit, J.C., 1993. Genomic fingerprinting of *Clostridium difficile* isolates by using a random amplified polymorphic DNA (RAPD) assay. FEMS Microbiol Lett 114, 161-166.

Bartlett, J.G., Chang, T.W., Gurwith, M., Gorbach, S.L., Onderdonk, A.B., 1978. Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia. N Engl J Med 298, 531-534.

Bauer, M.P., Notermans, D.W., van Benthem, B.H., Brazier, J.S., Wilcox, M.H., Rupnik, M., Monnet, D.L., van Dissel, J.T., Kuijper, E.J., 2011. *Clostridium difficile* infection in Europe: a hospital-based survey. Lancet 377, 63-73.

Båverud, V., Gustafsson, A., Franklin, A., Aspan, A., Gunnarsson, A., 2003. *Clostridium difficile*: prevalence in horses and environment, and antimicrobial susceptibility. Equine Vet J 35, 465-471.

Bignardi, G.E., 1998. Risk factors for *Clostridium difficile* infection. J Hosp Infect 40, 1-15.

Bowman, R.A., O'Neill, G.L., Riley, T.V., 1991. Non-radioactive restriction fragment length polymorphism (RFLP) typing of *Clostridium difficile*. FEMS Microbiol Lett 79, 269-272.

Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M., von Eichel-Streiber, C., 1996. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. Gene 181, 29-38.

Brazier, J.S., 1998. The epidemiology and typing of *Clostridium difficile*. J Antimicrob Chemother 41 (Suppl 3), S47-57.

Brazier, J.S., Patel, B., Pearson, A., 2007. Distribution of *Clostridium difficile* PCR ribotype 027 in British hospitals. Euro Surveill 12, E070426 070422.

Brouwer, M.S., Roberts, A.P., Hussain, H., Williams, R.J., Allan, E., Mullany, P., 2013. Horizontal gene transfer converts non-toxigenic *Clostridium difficile* strains into toxin producers. Nat Commun 4, 2601.

Brouwer, M.S., Roberts, A.P., Mullany, P., Allan, E., 2012. In silico analysis of sequenced strains of *Clostridium difficile* reveals a related set of conjugative transposons carrying a variety of accessory genes. Mob Genet Elements 2, 8-12.

Brouwer, M.S., Warburton, P.J., Roberts, A.P., Mullany, P., Allan, E., 2011. Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of *Clostridium difficile*. PLoS One 6, e23014.

Byl, B., Jacobs, F., Struelens, M.J., Thys, J.P., 1996. Extraintestinal *Clostridium difficile* infections. Clin Infect Dis 22, 712.

Cafardi, V., Biagini, M., Martinelli, M., Leuzzi, R., Rubino, J.T., Cantini, F., Norais, N., Scarselli, M., Serruto, D., Unnikrishnan, M., 2013. Identification of a novel zinc metalloprotease through a global analysis of *Clostridium difficile* extracellular proteins. PLoS One 8, e81306.

Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G., Fairweather, N., 2001. Molecular characterization of the surface layer proteins from *Clostridium difficile*. Mol Microbiol 40, 1187-1199.

Carter, G.P., Lyras, D., Allen, D.L., Mackin, K.E., Howarth, P.M., O'Connor, J.R., Rood, J.I., 2007. Binary toxin production in *Clostridium difficile* is regulated by CdtR, a LytTR family response regulator. J Bacteriol 189, 7290-7301.

Cartman, S.T., Kelly, M.L., Heeg, D., Heap, J.T., Minton, N.P., 2012. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype and toxin production. Appl Environ Microbiol 78, 4683-4690.

Cartwright, C.P., Stock, F., Beekmann, S.E., Williams, E.C., Gill, V.J., 1995. PCR amplification of rRNA intergenic spacer regions as a method for epidemiologic typing of *Clostridium difficile*. J Clin Microbiol 33, 184-187.

Casey, J.A., Kim, B.F., Larsen, J., Price, L.B., Nachman, K.E., 2015. Industrial food animal production and community health. Curr Environ Health Rep 2, 259-271.

Chang, T.W., Bartlett, J.G., Gorbach, S.L., Onderdonk, A.B., 1978. Clindamycin-induced enterocolitis in hamsters as a model of pseudomembranous colitis in patients. Infect Immun 20, 526-529.

Chaves-Olarte, E., Freer, E., Parra, A., Guzman-Verri, C., Moreno, E., Thelestam, M., 2003. R-Ras glucosylation and transient RhoA activation determine the cytopathic effect produced by toxin B variants from toxin A-negative strains of *Clostridium difficile*. J Biol Chem 278, 7956-7963.

Chaves-Olarte, E., Low, P., Freer, E., Norlin, T., Weidmann, M., von Eichel-Streiber, C., Thelestam, M., 1999. A novel cytotoxin from *Clostridium difficile* serogroup F is a functional hybrid between two other large clostridial cytotoxins. J Biol Chem 274, 11046-11052.

Cheng, A.C., Collins, D.A., Elliott, B., Ferguson, J.K., Thean, S., Paterson, D.L., Riley, T.V., 2016. Laboratory-based surveillance of *Clostridium difficile* circulating in Australia, September - November 2010. Pathology 48, 257-260.

Clements, A.C., Magalhaes, R.J., Tatem, A.J., Paterson, D.L., Riley, T.V., 2010. *Clostridium difficile* PCR ribotype 027: assessing the risks of further worldwide spread. Lancet Infect Dis 10, 395-404.

Cohen, S.H., Tang, Y.J., Silva, J., Jr., 2001. Molecular typing methods for the epidemiological identification of *Clostridium difficile* strains. Expert Rev Mol Diagn 1, 61-70.

Collins, D.A., Elliott, B., Riley, T.V., 2015. Molecular methods for detecting and typing *Clostridium difficile*. Pathology 47, 211-218.

Collins, D.A., Hawkey, P.M., Riley, T.V., 2013. Epidemiology of *Clostridium difficile* infection in Asia. Antimicrob Resist Infect Control 2, 21.

Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., Farrow, J.A., 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int J Syst Bacteriol 44, 812-826.

Corkill, J.E., Graham, R., Hart, C.A., Stubbs, S., 2000. Pulsed-field gel electrophoresis of degradation-sensitive DNAs from *Clostridium difficile* PCR ribotype 1 Strains. J Clin Microbiol 38, 2791-2792.

Darkoh, C., Odo, C., DuPont, H.L., 2016. Accessory gene regulator-1 locus is essential for virulence and pathogenesis of *Clostridium difficile*. MBio 7, e01237-01216.

De Almeida, M.N., Heffernan, H., Dervan, A., Bakker, S., Freeman, J.T., Bhally, H., Taylor, S.L., Riley, T.V., Roberts, S.A., 2013. Severe *Clostridium difficile* infection in New Zealand associated with an emerging strain, PCR-ribotype 244. N Z Med J 126, 9-14.

Delmée, M., Homel, M., Wauters, G., 1985. Serogrouping of *Clostridium difficile* strains by slide agglutination. J Clin Microbiol 21, 323-327.

Desai, K., Gupta, S.B., Dubberke, E.R., Prabhu, V.S., Browne, C., Mast, T.C., 2016. Epidemiological and economic burden of *Clostridium difficile* in the United States: estimates from a modeling approach. BMC Infect Dis 16, 303.

Devlin, H.R., Au, W., Foux, L., Bradbury, W.C., 1987. Restriction endonuclease analysis of nosocomial isolates of *Clostridium difficile*. J Clin Microbiol 25, 2168-2172.

Didelot, X., Eyre, D.W., Cule, M., Ip, C.L., Ansari, M.A., Griffiths, D., Vaughan, A., O'Connor, L., Golubchik, T., Batty, E.M., Piazza, P., Wilson, D.J., Bowden, R., Donnelly, P.J., Dingle, K.E., Wilcox, M., Walker, A.S., Crook, D.W., TE, A.P., Harding, R.M., 2012. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. Genome Biol 13, R118.

Dingle, K.E., Didelot, X., Ansari, M.A., Eyre, D.W., Vaughan, A., Griffiths, D., Ip, C.L., Batty, E.M., Golubchik, T., Bowden, R., Jolley, K.A., Hood, D.W., Fawley, W.N., Walker, A.S., Peto, T.E., Wilcox, M.H., Crook, D.W., 2013. Recombinational switching of the *Clostridium difficile* S-layer and a novel glycosylation gene cluster revealed by large-scale whole-genome sequencing. J Infect Dis 207, 675-686.

Dingle, K.E., Elliott, B., Robinson, E., Griffiths, D., Eyre, D.W., Stoesser, N., Vaughan, A., Golubchik, T., Fawley, W., Wilcox, M.H., Peto, T.E., Walker, A.S., Riley, T.V., Crook, D.W., Didelot, X., 2014. Evolutionary history of the *Clostridium difficile* Pathogencity Locus. Genome Biol Evol 6, 36-52.

Dingle, K.E., Griffiths, D., Didelot, X., Evans, J., Vaughan, A., Kachrimanidou, M., Stoesser, N., Jolley, K.A., Golubchik, T., Harding, R.M., Peto, T.E., Fawley, W., Walker, A.S., Wilcox, M., Crook, D.W., 2011. Clinical *Clostridium difficile*: clonality and pathogenicity locus diversity. PLoS ONE 6, e19993.

Donskey, C.J., Kundrapu, S., Deshpande, A., 2015. Colonization versus carriage of *Clostridium difficile*. Infect Dis Clin North Am 29, 13-28.

Eckert, C., Emirian, A., Le Monnier, A., Cathala, L., De Montclos, H., Goret, J., Berger, P., Petit, A., De Chevigny, A., Jean-Pierre, H., Nebbad, B., Camiade, S., Meckenstock, R., Lalande, V., Marchandin, H., Barbut, F., 2015. Prevalence and pathogenicity of binary toxin-positive *Clostridium difficile* strains that do not produce toxins A and B. New Microbes New Infect 3, 12-17.

Elliott, B., Dingle, K.E., Didelot, X., Crook, D.W., Riley, T.V., 2014. The complexity and diversity of the Pathogenicity Locus in *Clostridium difficil*e clade 5. Genome Biol Evol 6, 3159-3170.

Elliott, B., Read, R., Chang, B.J., Riley, T.V., 2008. Toxins A and B negative, binary toxin positive *Clostridium difficile* isolated from a case of fatal bacteraemia, Anaerobe Society of Americas 9th Biennial Congress, Long Beach, USA.

Eyre, D.W., Cule, M.L., Wilson, D.J., Griffiths, D., Vaughan, A., O'Connor, L., Ip, C.L., Golubchik, T., Batty, E.M., Finney, J.M., Wyllie, D.H., Didelot, X., Piazza, P., Bowden, R., Dingle, K.E., Harding, R.M., Crook, D.W., Wilcox, M.H., Peto, T.E., Walker, A.S., 2013. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med 369, 1195-1205.

Eyre, D.W., Tracey, L., Elliott, B., Slimings, C., Huntington, P.G., Stuart, R., Korman, T., Kotsiou, G., McCann, R., Griffiths, D., Fawley, W.N., Armstrong, P., Dingle, K.E., Walker, A.S., Peto, T.E.A., Cook, D.W., Wilcox, M.H., Riley, T.V., 2015. Emergence and spread of predominantly community-onset *Clostridium difficile* PCR ribotype 244 infection in Australia. Euro Surveill 20, 21059.

Fagan, R.P., Janoir, C., Collignon, A., Mastrantonio, P., Poxton, I.R., Fairweather, N.F., 2011. A proposed nomenclature for cell wall proteins of *Clostridium difficile*. J Med Microbiol 60, 1225-1228.

Gerding, D.N., 2009. *Clostridium difficile* 30 years on: what has, or has not, changed and why? Int J Antimicrob Agents 33 Suppl 1, S2-8.

Gerding, D.N., 2012. *Clostridium difficile* infection prevention: biotherapeutics, immunologics, and vaccines. Discov Med 13, 75-83.

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

Geric, B., Rupnik, M., Gerding, D.N., Grabnar, M., Johnson, S., 2004. Distribution of *Clostridium difficile* variant toxinotypes and strains with binary toxin genes among clinical isolates in an American hospital. J Med Microbiol 53, 887-894.

Geric Stare, B., Delmee, M., Rupnik, M., 2007. Variant forms of the binary toxin CDT locus and *tcdC* gene in *Clostridium difficile* strains. J Med Microbiol 56, 329-335.

Geric Stare, B., Rupnik, M., 2010. *Clostridium difficile* toxinotype XI (A^-B^-) exhibits unique arrangement of PaLoc and its upstream region. Anaerobe 16, 393-395.

Goh, S., 2005. Effect of phage infection on toxin production by *Clostridium difficile*. J Med Microbiol 54, 129-135.

Goh, S., Hussain, H., Chang, B.J., Emmett, W., Riley, T.V., Mullany, P., 2013. Phage C2 mediates transduction of Tn6215, encoding erythromycin resistance, between *Clostridium difficile* strains. MBio 4, e00840-00813.

Goorhuis, A., Bakker, D., Corver, J., Debast, S.B., Harmanus, C., Notermans, D.W., Bergwerff, A.A., Dekker, F.W., Kuijper, E.J., 2008. Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clin Infect Dis 47, 1162-1170.

Govind, R., Dupuy, B., 2012. Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. PLoS Pathog 8, e1002727.

Grandesso, S., Arena, F., Eseme, F.E., Panese, S., Henrici De Angelis, L., Spigaglia, P., Barbanti, F., Rossolini, G.M., 2016. *Clostridium difficile* ribotype 033 colitis in a patient following broad-spectrum antibiotic treatment for KPC-producing *Klebsiella pneumoniae* infection, Italy. New Microbiol 39, 235-236.

Griffiths, D., Fawley, W., Kachrimanidou, M., Bowden, R., Crook, D.W., Fung, R., Golubchik, T., Harding, R.M., Jeffery, K.J., Jolley, K.A., Kirton, R., Peto, T.E., Rees, G., Stoesser, N., Vaughan, A., Walker, A.S., Young, B.C., Wilcox, M., Dingle, K.E., 2010. Multilocus sequence typing of *Clostridium difficile*. J Clin Microbiol 48, 770-778.

Gurtler, V., 1993. Typing of *Clostridium difficile* strains by PCR-amplification of variable length 16S-23S rDNA spacer regions. J Gen Microbiol 139, 3089-3097.

Hall, J.C., O'Toole, E., 1935. Intestinal flora in new-born infants with a description of a new pathogenic anaerobe, *Bacillus difficilis*. Am J Dis Child 49, 390-342.

Hargreaves, K.R., Clokie, M.R., 2014. *Clostridium difficile* phages: still difficult? Front Microbiol 5, 184.

Hargreaves, K.R., Colvin, H.V., Patel, K.V., Clokie, J.J., Clokie, M.R., 2013. Genetically diverse *Clostridium difficile* strains harboring abundant prophages in an estuarine environment. Appl Environ Microbiol 79, 6236-6243.

Hargreaves, K.R., Flores, C.O., Lawley, T.D., Clokie, M.R., 2014a. Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium difficile* strains and prophages target multiple phage types within this pathogen. MBio 5, e01045-01013.

Hargreaves, K.R., Kropinski, A.M., Clokie, M.R.J., 2014b. What does the talking?: quorum sensing signalling genes discovered in a bacteriophage genome. PLoS ONE 9, e85131.

Hargreaves, K.R., Thanki, A.M., Jose, B.R., Oggioni, M.R., Clokie, M.R., 2016. Use of single molecule sequencing for comparative genomics of an environmental and a clinical isolate of *Clostridium difficile* ribotype 078. BMC Genomics 17, 1020.

Hasselmayer, O., Braun, V., Nitsche, C., Moos, M., Rupnik, M., von Eichel-Streiber, C., 2004a. Clostridium difficile IStron CdISt1: discovery of a variant encoding two complete transposase-like proteins. Journal of Bacteriology 186, 2508-2510.

Hasselmayer, O., Nitsche, C., Braun, V., von Eichel-Streiber, C., 2004b. The IStron Cd*ISt1* of *Clostridium difficile*: molecular symbiosis of a group I intron and an insertion element. Anaerobe 10, 85-92.

He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D.J., Martin, M.J., Connor, T.R., Harris, S.R., Fairley, D., Bamford, K.B., D'Arc, S., Brazier, J., Brown, D., Coia, J.E., Douce, G., Gerding, D., Kim, H.J., Koh, T.H., Kato, H., Senoh, M., Louie, T., Michell, S., Butt, E., Peacock, S.J., Brown, N.M., Riley, T., Songer, G., Wilcox, M., Pirmohamed, M., Kuijper, E., Hawkey, P., Wren, B.W., Dougan, G., Parkhill, J., Lawley, T.D., 2013. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. Nat Genet 45, 109-113.

He, M., Sebaihia, M., Lawley, T.D., Stabler, R.A., Dawson, L.F., Martin, M.J., Holt, K.E., Seth-Smith, H.M., Quail, M.A., Rance, R., Brooks, K., Churcher, C., Harris, D., Bentley, S.D., Burrows, C., Clark, L., Corton, C., Murray, V., Rose, G., Thurston, S., van Tonder, A., Walker, D., Wren,

B.W., 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.

Health Protection Agency, 2011. *Clostridium difficile* Ribotyping Network (CDRN) for England and Northern Ireland: 2009/10 Report.

Hensgens, M.P., Keessen, E.C., Squire, M.M., Riley, T.V., Koene, M.G., de Boer, E., Lipman, L.J., Kuijper, E.J., 2012. *Clostridium difficile* infection in the community: a zoonotic disease? Clin Microbiol Infect 18, 635-645.

Hummel, R.P., Altemeier, W.A., Hill, E.O., 1964. Iatrogenic staphylococcal enterocolitis. Ann Surg 160, 551-560.

Hundsberger, T., Braun, V., Weidmann, M., Leukel, P., Sauerborn, M., von Eichel-Streiber, C., 1997. Transcription analysis of the genes *tcdA-E* of the pathogenicity locus of *Clostridium difficile*. Eur J Biochem 244, 735-742.

Jacobs, A., Barnard, K., Fishel, R., Gradon, J.D., 2001. Extracolonic manifestations of *Clostridium difficile* infections. Presentation of 2 cases and review of the literature. Medicine (Baltimore) 80, 88-101.

Janezic, S., Dingle, K.E., Didelot, X., Rikanovic, T., Crook, D.W., Rupnik, M., 2015a. Comparative genomics of *Clostridium difficile* toxinotypes, 5th International Conference on *Clostridium difficile*, Bled, Slovenia.

Janezic, S., Marin, M., Martin, A., Rupnik, M., 2015b. A new type of toxin A-negative, toxin B-positive *Clostridium difficile* strain lacking a complete *tcdA* gene. J Clin Microbiol 53, 692-695.

Janezic, S., Zidaric, V., Pardon, B., Indra, A., Kokotovic, B., Blanco, J.L., Seyboldt, C., Diaz, C.R., Poxton, I.R., Perreten, V., Drigo, I., Jiraskova, A., Ocepek, M., Weese, J.S., Songer, J.G., Wilcox, M.H., Rupnik, M., 2014. International *Clostridium difficile* animal strain collection and large diversity of animal associated strains. BMC Microbiol 14, 173.

Janvilisri, T., Scaria, J., Thompson, A.D., Nicholson, A., Limbago, B.M., Arroyo, L.G., Songer, J.G., Grohn, Y.T., Chang, Y.F., 2009. Microarray identification of *Clostridium difficile* core components and divergent regions associated with host origin. J Bacteriol 191, 3881-3891.

Johnson, S., Clabots, C.R., Linn, F.V., Olson, M.M., Peterson, L.R., Gerding, D.N., 1990. Nosocomial *Clostridium difficile* colonisation and disease. Lancet North Am Ed 336, 97.

Kato, H., Kato, N., Katow, S., Maegawa, T., Nakamura, S., Lyerly, D.M., 1999. Deletions in the repeating sequences of the toxin A gene of toxin A-negative, toxin B-positive *Clostridium difficile* strains. FEMS Microbiol Lett 175, 197-203.

Keel, M.K., Songer, J.G., 2006. The comparative pathology of *Clostridium difficile*-associated disease. Vet Pathol 43, 225-240.

Kelly, C.P., Pothoulakis, C., LaMont, J.T., 1994. *Clostridium difficile* colitis. N Engl J Med 330, 257-262.

Khan, M.Y., Hall, W.H., 1966. Staphylococcal enterocolitis—treatment with oral vancomycin. Ann Intern Med 65, 1-8.

Killgore, G.E., Kato, H., 1994. Use of arbitrary primer PCR to type *Clostridium difficile* and comparison of results with those by immunoblot typing. J Clin Microbiol 32, 1591-1593.

Knetsch, C.W., Connor, T.R., Mutreja, A., van Dorp, S.M., Sanders, I.M., Browne, H.P., Harris, D., Lipman, L., Keessen, E.C., Corver, J., Kuijper, E.J., Lawley, T.D., 2014. Whole genome sequencing reveals potential spread of Clostridium difficile between humans and farm animals in the Netherlands, 2002 to 2011. Euro Surveill 19, 20954.

Knetsch, C.W., Lawley, T.D., Hensgens, M.P., Corver, J., Wilcox, M.W., Kuijper, E.J., 2013. Current application and future perspectives of molecular typing methods to study *Clostridium difficile* infections. Euro Surveill 18, 20381.

Knetsch, C.W., Terveer, E.M., Lauber, C., Gorbalenya, A.E., Harmanus, C., Kuijper, E.J., Corver, J., van Leeuwen, H.C., 2012. Comparative analysis of an expanded *Clostridium difficile* reference strain collection reveals genetic diversity and evolution through six lineages. Infect Genet Evol 12, 1577-1585.

Knight, D.R., Androga, G.O., Ballard, S.A., Howden, B.P., Riley, T.V., 2016a. A phenotypically silent *vanB2* operon carried on a Tn1549-like element in *Clostridium difficile*. mSphere 1, e00177-00116.

Knight, D.R., Elliott, B., Chang, B.J., Perkins, T.T., Riley, T.V., 2015. Diversity and evolution in the genome of *Clostridium difficile*. Clin Microbiol Rev 28, 721-741.

Knight, D.R., Putsathit, P., Elliott, B., Riley, T.V., 2016b. Contamination of Australian newborn calf carcasses at slaughter with *Clostridium difficile*. Clin Microbiol Infect 22, 266.e261-267.

Kociolek, L.K., Gerding, D.N., 2016. Breakthroughs in the treatment and prevention of *Clostridium difficile* infection. Nat Rev Gastroenterol Hepatol 13, 150-160.

Kuijper, E.J., Oudbier, J.H., Stuifbergen, W.N., Jansz, A., Zanen, H.C., 1987. Application of whole-cell DNA restriction endonuclease profiles to the epidemiology of *Clostridium difficile*-induced diarrhea. J Clin Microbiol 25, 751-753.

Kuwata, Y., Tanimoto, S., Sawabe, E., Shima, M., Takahashi, Y., Ushizawa, H., Fujie, T., Koike, R., Tojo, N., Kubota, T., Saito, R., 2015. Molecular epidemiology and antimicrobial susceptibility of *Clostridium difficile* isolated from a university teaching hospital in Japan. Eur J Clin Microbiol Infect Dis 34, 763-772.

Lambert, G.S., Baldwin, M.R., 2016. Evidence for dual receptor-binding sites in *Clostridium difficile* toxin A. FEBS Lett.

Lanis, J.M., Barua, S., Ballard, J.D., 2010. Variations in TcdB activity and the hypervirulence of emerging strains of *Clostridium difficile*. PLoS Pathog 6, e1001061.

Lanis, J.M., Heinlen, L.D., James, J.A., Ballard, J.D., 2013. *Clostridium difficile* 027/BI/NAP1 encodes a hypertoxic and antigenically variable form of TcdB. PLoS Pathog 9, e1003523.

Larson, H.E., Price, A.B., 1977. Pseudomembranous colitis: presence of clostridial toxin. Lancet 310, 1312-1314.

Lawson, P.A., Citron, D.M., Tyrrell, K.L., Finegold, S.M., 2016. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prevot 1938. Anaerobe 40, 95-99.

Lemee, L., Dhalluin, A., Pestel-Caron, M., Lemeland, J.F., Pons, J.L., 2004. Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. J Clin Microbiol 42, 2609-2617.

Lim, S.K., Stuart, R.L., Mackin, K.E., Carter, G.P., Kotsanas, D., Francis, M.J., Easton, M., Dimovski, K., Elliott, B., Riley, T.V., Hogg, G., Paul, E., Korman, T.M., Seemann, T., Stinear, T.P., Lyras, D., Jenkin, G.A., 2014. Emergence of a ribotype 244 strain of *Clostridium difficile* associated with severe disease and related to the epidemic ribotype 027 strain. Clin Infect Dis 58, 1723-1730.

Loo, V.G., Poirier, L., Miller, M.A., Oughton, M., Libman, M.D., Michaud, S., Bourgault, A.-M., Nguyen, T., Frenette, C., Kelly, M., Vibien, A., Brassard, P., Fenn, S., Dewar, K., Hudson, T.J., Horn, R., Rene, P., Monczak, Y., Dascal, A., 2005. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. N Engl J Med 353, 2442-2449.

Ludwig, W., Schleifer, K.-H., Whitman, W.B., 2009. Revised road map to the phylum Firmicutes, in: De Vos, P., Garrity, G., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K.H., Whitman, W.B. (Eds.), Bergey's Manual of Systematic Bacteriology, 2nd ed. Springer, New York, pp. 1-14.

Lund, B.M., Peck, M.W., 2015. A possible route for foodborne transmission of *Clostridium difficile*? Foodborne Pathog Dis 12, 177-182.

Lyon, S.A., Hutton, M.L., Rood, J.I., Cheung, J.K., Lyras, D., 2016. CdtR Regulates TcdA and TcdB Production in *Clostridium difficile*. PLoS Pathog 12, e1005758.

Lyras, D., Storie, C., Huggins, A.S., Crellin, P.K., Bannam, T.L., Rood, J.I., 1998. Chloramphenicol resistance in *Clostridium difficile* is encoded on Tn*4453* transposons that are closely related to Tn*4451* from *Clostridium perfringens*. Antimicrob Agents Chemother 42, 1563-1567.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95, 3140-3145.

Martin, M.J., Clare, S., Goulding, D., Faulds-Pain, A., Barquist, L., Browne, H.P., Pettit, L., Dougan, G., Lawley, T.D., Wren, B.W., 2013. The *agr* locus regulates virulence and colonization genes in *Clostridium difficile* 027. J Bacteriol 195, 3672-3681.

McDonald, L.C., Killgore, G.E., Thompson, A., Owens, R.C., Jr., Kazakova, S.V., Sambol, S.P., Johnson, S., Gerding, D.N., 2005. An epidemic, toxin gene-variant strain of *Clostridium difficile*. N Engl J Med 353, 2433-2441.

McDonald, L.C., Owings, M., Jernigan, D.B., 2006. *Clostridium difficile* infection in patients discharged from US short-stay hospitals, 1996-2003. Emerg Infect Dis 12, 409-415.

McMillin, D.E., Muldrow, L.L., 1992. Typing of toxic strains of *Clostridium difficile* using DNA fingerprints generated with arbitrary polymerase chain reaction primers. FEMS Microbiol Lett 71, 5-9.

Mehlig, M., Moos, M., Braun, V., Kalt, B., Mahony, D.E., von Eichel-Streiber, C., 2001. Variant toxin B and a functional toxin A produced by *Clostridium difficile* C34. FEMS Microbiol Lett 198, 171-176.

Monot, M., Boursaux-Eude, C., Thibonnier, M., Vallenet, D., Moszer, I., Medigue, C., Martin-Verstraete, I., Dupuy, B., 2011. Reannotation of the genome sequence of *Clostridium difficile* strain 630. J Med Microbiol 60, 1193-1199.

Monot, M., Eckert, C., Lemire, A., Hamiot, A., Dubois, T., Tessier, C., Dumoulard, B., Hamel, B., Petit, A., Lalande, V., Ma, L., Bouchier, C., Barbut, F., Dupuy, B., 2015. *Clostridium difficile*: new insights into the evolution of the Pathogenicity Locus. Sci Rep 5, 15023.

Moono, P., Foster, N.F., Hampson, D.J., Knight, D.R., Bloomfield, L.E., Riley, T.V., 2016. *Clostridium difficile* infection in production animals and avian species: a review. Foodborne Pathog Dis, http://dx.doi.org/10.1089/fpd.2016.2181.

Moore, P., Kyne, L., Martin, A., Solomon, K., 2013. Germination efficiency of clinical *Clostridium difficile* spores and correlation with ribotype, disease severity and therapy failure. J Med Microbiol 62, 1405-1413.

Mulligan, M.E., Peterson, L.R., Kwok, R.Y.Y., Clabots, C.R., Gerding, D.N., 1988. Immunoblots and plasmid fingerprints compared with serotyping and polyacrylamide gel electrophoresis for typing *Clostridium difficile*. J Clin Microbiol 26, 41-46.

O'Neill, G.L., Adams, J.E., Bowman, R.A., Riley, T.V., 1993. A molecular characterization of *Clostridium difficile* isolates from humans, animals and their environments. Epidemiol Infect 111, 257-264.

O'Neill, G.L., Ogunsola, F.T., Brazier, J.S., Duerden, B.I., 1996. Modification of a PCR ribotyping method for application as a routine typing scheme for *Clostridium difficile*. Anaerobe 2, 205-209.

Pantaleon, V., Soavelomandroso, A.P., Bouttier, S., Briandet, R., Roxas, B., Chu, M., Collignon, A., Janoir, C., Vedantam, G., Candela, T., 2015. The *Clostridium difficile* protease Cwp84 modulates both biofilm formation and cell-surface properties. PLoS One 10, e0124971.

Peerbooms, P.G., Kuijt, P., Maclaren, D.M., 1987. Application of chromosomal restriction endonuclease digest analysis for use as typing method for *Clostridium difficile*. J Clin Pathol 40, 771-776.

Pepin, J., Valiquette, L., Alary, M.E., Villemure, P., Pelletier, A., Forget, K., Pepin, K., Chouinard, D., 2004. *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. CMAJ 171, 466-472.

Polivkova, S., Krutova, M., Petrlova, K., Benes, J., Nyc, O., 2016. *Clostridium difficile* ribotype 176 - A predictor for high mortality and risk of nosocomial spread? Anaerobe 40, 35-40.

Public Health England, 2016. *Clostridium difficile* Ribotyping Network (CDRN) for England and Northern Ireland - Biennial report (2013-2015), London.

Rifkin, G.D., Fekety, F.R., Silva, J., Jr., 1977. Antibiotic-induced colitis implication of a toxin neutralised by *Clostridium sordellii* antitoxin. Lancet 310, 1103-1106.

Riley, T.V., Thean, S., Hool, G., Golledge, C.L., 2009. First Australian isolation of epidemic *Clostridium difficile* PCR ribotype 027. Med J Aust 190, 706-708.

Rodriguez, C., Van Broeck, J., Taminiau, B., Delmee, M., Daube, G., 2016. *Clostridium difficile* infection: early history, diagnosis and molecular strain typing methods. Microb Pathog 97, 59-78.

Rupnik, M., 2008. Heterogeneity of large clostridial toxins: importance of *Clostridium difficile* toxinotypes. FEMS Microbiol Rev 32, 541-555.

Rupnik, M., Avesani, V., Janc, M., von Eichel-Streiber, C., Delmée, M., 1998. A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. J Clin Microbiol 36, 2240-2247.

Rupnik, M., Janezic, S., 2016. An update on *Clostridium difficile* toxinotyping. J Clin Microbiol 54, 13-18.

Rupnik, M., Widmer, A., Zimmermann, O., Eckert, C., Barbut, F., 2008. *Clostridium difficile* toxinotype V, ribotype 078, in animals and humans. J Clin Microbiol 46, 2146.

Rupnik, M., Wilcox, M.H., Gerding, D.N., 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. Nat Rev Microbiol 7, 526-536.

Sanchez-Hurtado, K., Corretge, M., Mutlu, E., McIlhagger, R., Starr, J.M., Poxton, I.R., 2008. Systemic antibody response to *Clostridium difficile* in colonized patients with and without symptoms and matched controls. J Med Microbiol 57, 717-724.

Scaria, J., Ponnala, L., Janvilisri, T., Yan, W., Mueller, L.A., Chang, Y.F., 2010. Analysis of ultra low genome conservation in *Clostridium difficile*. PLoS ONE 5, e15147.

Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeno-Tarraga, A.M., Wang, H., Holden, M.T., Wright, A., Churcher, C., Quail, M.A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B., Parkhill, J., 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet 38, 779-786.

Sekulovic, O., Meessen-Pinard, M., Fortier, L.C., 2011. Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. J Bacteriol 193, 2726-2734.

Shan, J., Patel, K.V., Hickenbotham, P.T., Nale, J.Y., Hargreaves, K.R., Clokie, M.R., 2012. Prophage carriage and diversity within clinically relevant strains of *Clostridium difficile*. Appl Environ Microbiol 78, 6027-6034.

Skarin, H., Segerman, B., 2014. Plasmidome Interchange between *Clostridium botulinum*, *Clostridium novyi* and *Clostridium haemolyticum* converts strains of independent lineages into distinctly different pathogens. PLoS ONE 9, e107777.

Smith, L.D.S., King, E.O., 1962. Occurrence of *Clostridium difficile* in infections of man. J Bacteriol 84, 65-67.

Snyder, M.L., 1937. Further studies on *Bacillus difficilis* (Hall and O'Toole). J Infect Dis 60, 223-231.

Songer, J.G., Uzal, F.A., 2005. Clostridial enteric infections in pigs. J Vet Diagn Invest 17, 528-536.

Sorg, J.A., Sonenshein, A.L., 2008. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. J Bacteriol 190, 2505-2512.

Stabler, R.A., Dawson, L.F., Phua, L.T., Wren, B.W., 2008. Comparative analysis of BI/NAP1/027 hypervirulent strains reveals novel toxin B-encoding gene (*tcdB*) sequences. J Med Microbiol 57, 771-775.

Stabler, R.A., Dawson, L.F., Valiente, E., Cairns, M.D., Martin, M.J., Donahue, E.H., Riley, T.V., Songer, J.G., Kuijper, E.J., Dingle, K.E., Wren, B.W., 2012. Macro and micro diversity of *Clostridium difficile* isolates from diverse sources and geographical locations. PLoS ONE 7, e31559.

Stabler, R.A., Gerding, D.N., Songer, J.G., Drudy, D., Brazier, J.S., Trinh, H.T., Witney, A.A., Hinds, J., Wren, B.W., 2006. Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. J Bacteriol 188, 7297-7305.

Stabler, R.A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T.D., Sebaihia, M., Quail, M.A., Rose, G., Gerding, D.N., Gibert, M., Popoff, M.R., Parkhill, J., Dougan, G., Wren, B.W., 2009. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol 10, R102.

Stevenson, E., Minton, N.P., Kuehne, S.A., 2015. The role of flagella in *Clostridium difficile* pathogenicity. Trends Microbiol 23, 275-282.

Strelau, E., 1989. Demonstration of capsules in *Clostridium difficile*. Zentralbl Bakteriol Mikrobiol Hyg A 270, 456-461.

Tabaqchali, S., O'Farrell, S., Holland, D., Silman, R., 1986. Method for the typing of *Clostridium difficile* based on polyacrylamide gel electrophoresis of [35S]methionine-labeled proteins. J Clin Microbiol 23, 197-198.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30, 2725-2729.

Taori, S.K., Hall, V., Poxton, I.R., 2010. Changes in antibiotic susceptibility and ribotypes in *Clostridium difficile* isolates from southern Scotland, 1979-2004. J Med Microbiol 59, 338-344.

Tedesco, F.J., Barton, R.W., Alpers, D.H., 1974. Clindamycin-associated colitis. A prospective study. Ann Intern Med 81, 429-433.

Tettelin, H., Riley, D., Cattuto, C., Medini, D., 2008. Comparative genomics: the bacterial pan-genome. Current Opinion in Microbiology 11, 472-477.

Tickler, I.A., Goering, R.V., Whitmore, J.D., Lynn, A.N., Persing, D.H., Tenover, F.C., 2014. Strain types and antimicrobial resistance patterns of *Clostridium difficile* isolates from the United States, 2011 to 2013. Antimicrob Agents Chemother 58, 4214-4218.

Toma, S., Lesiak, G., Magus, M., Lo, H.L., Delmee, M., 1988. Serotyping of *Clostridium difficile*. J Clin MIcrobiol 26, 426-428.

van den Berg, R.J., Claas, E.C.J., Oyib, D.H., Klaassen, C.H.W., Dijkshoorn, L., Brazier, J.S., Kuijper, E.J., 2004. Characterization of toxin A-negative, toxin B-positive *Clostridium difficile* isolates from outbreaks in different countries by amplified fragment length polymorphism and PCR ribotyping. J Clin Microbiol 42, 1035-1041.

Vickers, R.J., Tillotson, G., Goldstein, E.J., Citron, D.M., Garey, K.W., Wilcox, M.H., 2016. Ridinilazole: a novel therapy for *Clostridium difficile* infection. Int J Antimicrob Agents 48, 137-143.

Vos, M., Didelot, X., 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J 3, 199-208.

Wakefield, R.D., Sommers, S.D., 1953. Fatal membranous staphylococcal enteritis in surgical patients. Ann Surg 138, 249-252.

Waligora, A.J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A., Karjalainen, T., 2001. Characterization of a cell surface protein of Clostridium difficile with adhesive properties. Infect Immun 69, 2144-2153.

Warny, M., Pepin, J., Fang, A., Killgore, G., Thompson, A., Brazier, J., Frost, E., McDonald, L.C., 2005. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. The Lancet 366, 1079-1084.

Weese, J.S., 2010. *Clostridium difficile* in food--innocent bystander or serious threat? Clin Microbiol Infect 16, 3-10.

Wilks, M., Tabaqchali, S., 1994. Typing of *Clostridium difficile* by polymerase chain reaction with an arbitrary primer. J Hosp Infect 28, 231-234.

Wren, B.W., Tabaqchali, S., 1987. Restriction endonuclease DNA analysis of *Clostridium difficile*. J Clin Microbiol 25, 2402-2404.

Wust, J., Sullivan, N.M., Hardegger, U., Wilkins, T.D., 1982. Investigation of an outbreak of antibiotic-associated colitis by various typing methods. J Clin Microbiol 16, 1096-1101.

Yutin, N., Galperin, M.Y., 2013. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. Environ Microbiol 15, 2631-2641.

Zanella Terrier, M.C., Simonet, M.L., Bichard, P., Frossard, J.L., 2014. Recurrent *Clostridium difficile* infections: the importance of the intestinal microbiota. World J Gastroenterol 20, 7416-7423.

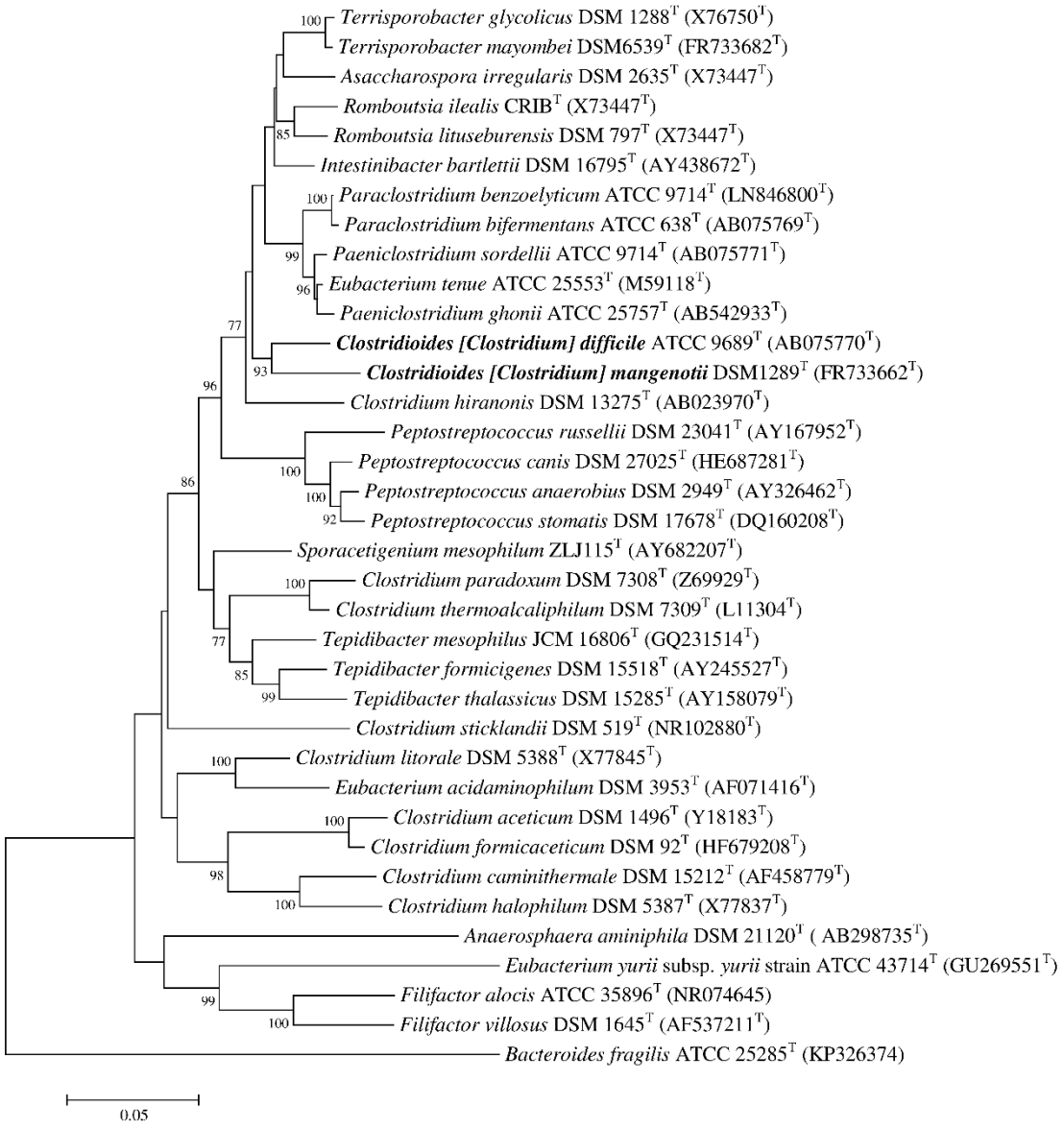

Fig. 1. Evolutionary relationships of taxa. Neighbor-Joining tree based on 16S rRNA gene sequence data, showing the phylogenetic position of members of the family *Peptostreptococcaceae*. The tree was constructed using the Neighbor-Joining method using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair. There were a total of 1425 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The outgroup used was *Bacteroides fragilis*. Figure from (Lawson et al., 2016).

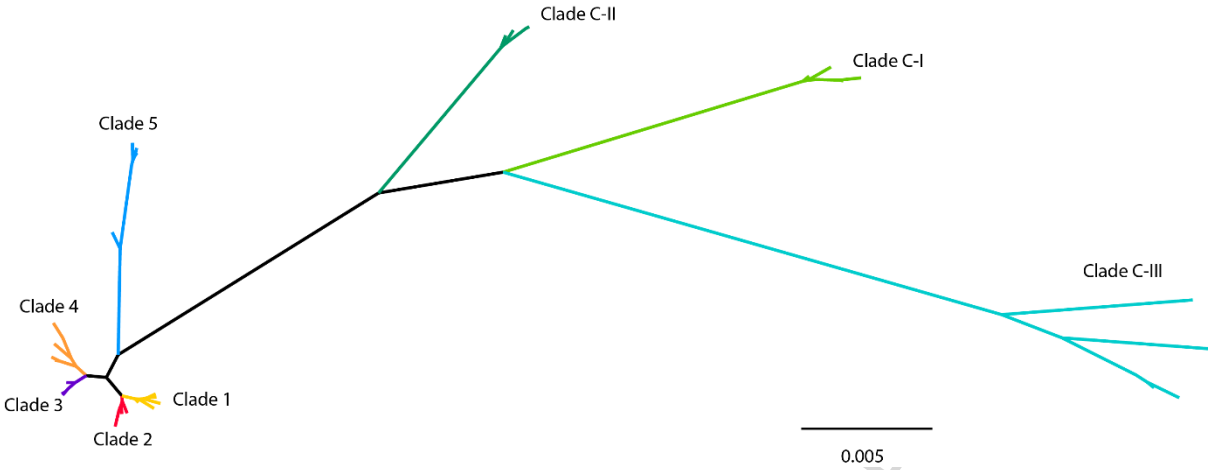

Figure 2. Maximum likelihood tree generated using MLST data from all known clades present in the PubMLST database. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The tree with the highest log likelihood (-8568.3698) is shown. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013).

TABLE 1. Notable types within the phylogenetic clades

Clade	Notable types
1	UK 014 (ST-14),[†] UK 020 (ST-11)[†], UK 002 (ST-8),[†] UK 015 (ST-44),[†] UK 018 (ST-17)[†]
2	UK 027 (ST-1),[‡] UK 244 (ST-41),[‡] UK 176 (ST-1),[‡] UK 251 (ST-231)[‡]
3	UK 023 (ST-5) [‡]
4	UK 017 (ST-37)[§], UK 047 (ST-37)[§]
5	UK 078 (ST-11),[‡] UK 126 (ST-11),[‡] UK 033 (ST-11),[¦] UK 237 (ST-167)[¶]
C-I	ST-181,[¶] ST-206[¶]
C-II	ST-200,[§ #]
C-III	Unknown[*]

NB: Where ribotypes are associated with multiple sequence types, the most common sequence types is given (Dingle et al., 2011; Stabler et al., 2012).

[†] Toxin profile: A+B+CDT−

[‡] Toxin profile: A+B+CDT+

[§] Toxin profile: A−B+CDT−

[¦] Toxin profile: A−B−CDT+

[¶] Toxin profile: A−B+CDT+

[#] Toxin profile: A+B−CDT− (unusual monotoxin locus)

[*] Very little is known about the three cryptic clades. Although some do contain unusual toxigenic strains, the majority are apparently non-toxigenic.

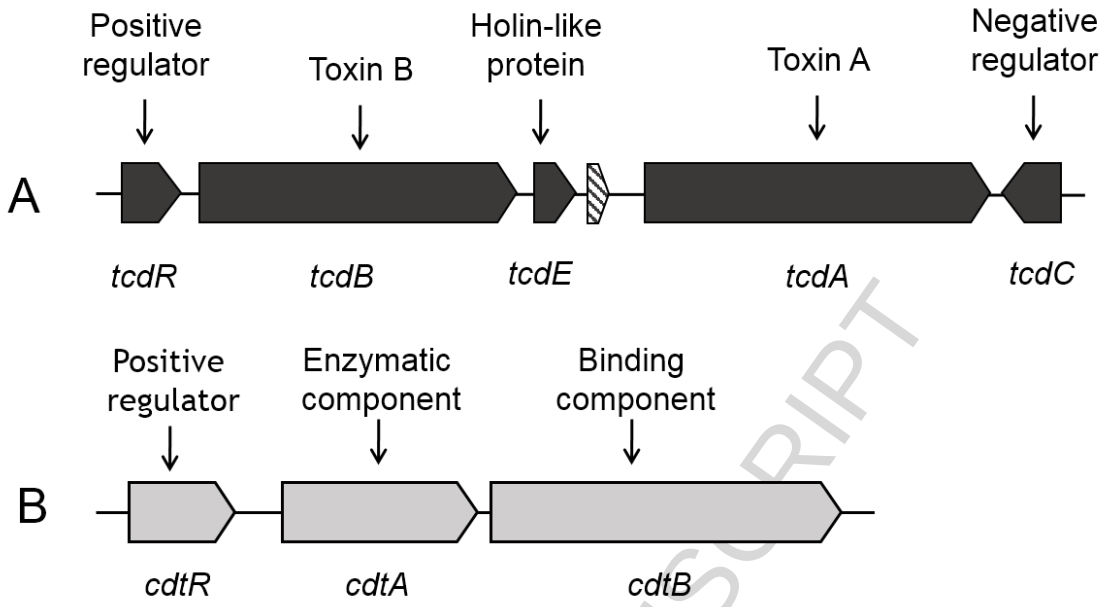

Figure 3. The two main toxin loci of *C. difficile*. (A) The PaLoc which encodes the toxin A and toxin B, as well as three accessory proteins. A partial pseudogene is present downstream of *tcdE*, shown with a dashed fill. (B) The CdtLoc which encodes the two binary toxin genes and one accessory gene.

Article highlights

- Whole genome sequencing of *C. difficile* in recent years has generated a large amount of data giving greater insight into the species.
- The *C. difficile* population consists of at least eight phylogenetic lineages.
- The *C. difficile* genome is characterised by a small core genome, and a large proportion of mobile elements.