EBioMedicine 43 (2019) 347–355

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Research paper

Clostridium difficile trehalose metabolism variants are common and not associated with adverse patient outcomes when variably present in the same lineage

David W. Eyre [a,b,*], Xavier Didelot [c], Anthony M. Buckley [d], Jane Freeman [d], Ines B. Moura [d], Derrick W. Crook [b,e,f], Tim E.A. Peto [b,e,f], A. Sarah Walker [b,e,f], Mark H. Wilcox [d,1], Kate E. Dingle [b,1]

[a] Big Data Institute, University of Oxford, UK
[b] Nuffield Department of Medicine, University of Oxford, UK
[c] School of Life Sciences, Department of Statistics, University of Warwick, UK
[d] Healthcare Associated Infections Research Group, University of Leeds, Leeds, UK
[e] National Institutes of Health Research Health Protection Unit on Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, UK
[f] National Institutes of Health Research Biomedical Research Centre, University of Oxford, UK

ARTICLE INFO

Article history:
Received 7 March 2019
Received in revised form 18 April 2019
Accepted 18 April 2019
Available online 27 April 2019

Keywords:
Clostridium difficile
Whole-genome sequencing
Trehalose
Epidemiology

ABSTRACT

Background: *Clostridium difficile* ribotype-027, ribotype-078, and ribotype-017 are virulent and epidemic lineages. Trehalose metabolism variants in these ribotypes, combined with increased human trehalose consumption, have been hypothesised to have contributed to their emergence and virulence.
Methods: 5232 previously whole-genome sequenced *C. difficile* isolates were analysed. Clinical isolates were used to investigate the impact of trehalose metabolism variants on mortality. Import data were used to estimate changes in dietary trehalose. Ribotype-027 virulence was investigated in a clinically reflective gut model.
Findings: Trehalose metabolism variants found in ribotype-027 and ribotype-017 were widely distributed throughout *C. difficile* clade-2 and clade-4 in 24/29 (83%) and 10/11 (91%) of sequence types (STs), respectively. The four-gene trehalose metabolism cluster described in ribotype-078 was common in genomes from all five clinically-important *C. difficile* clades (40/167 [24%] STs).
The four-gene cluster was variably present in 208 ribotype-015 infections (98 [47%]); 27/208 (13%) of these patients died within 30-days of diagnosis. Adjusting for age, sex, and infecting ST, there was no association between 30-day all-cause mortality and the four-gene cluster (OR 0.36 [95%CI 0.09–1.34, $p = 0.13$]).
Synthetic trehalose imports in the USA, UK, Germany and the EU were < 1 g/capita/year during 2000–2006, and < 9 g/capita/year 2007–2012, compared with dietary trehalose from natural sources of ~100 g/capita/year.
Trehalose supplementation did not increase ribotype-027 virulence in a clinically-validated gut model.
Interpretation: Trehalose metabolism variants are common in *C. difficile*. Increases in total dietary trehalose during the early-mid 2000s *C. difficile* epidemic were likely relatively minimal. Alternative explanations are required to explain why ribotype-027, ribotype-078 and ribotype-017 have been successful.
Funding: National Institute for Health Research. Gut model experiments only: Hayashibara Co. Ltd.

1. Introduction

Clostridium difficile is an important cause of healthcare-associated diarrhoea. The emergence and global spread of the hypervirulent ribotype-027 strain has made control and prevention of *C. difficile* infection, CDI, a major priority for healthcare systems.

The reasons for the success of ribotype-027 are likely multifactorial, including very probably fluoroquinolone resistance [1], and possibly other adaptions to the hospital environment [2]. Acquisition of fluoroquinolone resistance, on two separate occasions, occurred around the start of the epidemic spread of ribotype-027 throughout the USA, Canada and Europe [1]. High levels of fluoroquinolone resistance are also seen in other healthcare-associated ribotypes, including 001, 018/356 and 176 [3]. Levels of ribotype-027, and other fluoroquinolone-resistant CDI fell markedly following reductions in fluoroquinolone usage in the UK [4].

* Corresponding author at: Big Data Institute, University of Oxford, UK.
E-mail address: david.eyre@bdi.ox.ac.uk (D.W. Eyre).
[1] Contributed equally.

https://doi.org/10.1016/j.ebiom.2019.04.038

Research in context

Evidence before this study

Clostridium difficile ribotype-027 and ribotype-078 are two virulent and epidemic lineages, and ribotype-017 is widely present in Asia. Acquisition of fluoroquinolone resistance by ribotype-027 likely promoted its spread. However, the reasons for the success of these ribotypes are likely multifactorial. Trehalose metabolism variants in ribotypes 027, 078 and 017, combined with recent increases in human dietary intake of the sugar trehalose, have been hypothesised to have helped select for their emergence and contributed to their virulence. We searched Pubmed and Google Scholar using the search term "*Clostridium difficile*" and "trehalose" on 22 January 2019. We also identified references citing search results using Google Scholar. Trehalose metabolism variants in *C. difficile* confer the ability to metabolise low concentrations of trehalose and that this provides a competitive advantage in growth media and laboratory animals. Enhanced trehalose metabolism in ribotype-027 has been associated with increased virulence in a mouse model of *C. difficile* infection. However, the distribution of trehalose metabolism variants within the overall genetic diversity of *C. difficile*, their role in patient outcomes and the extent of increases in dietary trehalose from synthetic production are incompletely described to date.

Added value of this study

The trehalose metabolism variant found in ribotype-027 (TreR L172I) is widely distributed throughout *C. difficile* clade-2, and the variant in ribotype-017 (TreR C171S) is found in all but one sequence type (ST) in clade-4. The four-gene trehalose metabolism cluster described in ribotype-078 is common in genomes of STs representing all five clinically important *C. difficile* clades and not restricted to clade-5 as previously proposed. Within closely related genomes from ribotype-015, with and without the four-gene trehalose metabolism cluster, there was no evidence of an association between 30-day all-cause mortality and the four-gene cluster. Annual imports per capita of synthetic trehalose into the USA, UK, Germany and the EU as a whole during 2000 to 2006, i.e. around the start of the ribotype-027 outbreak, accounted for only a minimal proportion of total dietary trehalose intake. Trehalose supplementation, at levels found in the human gut, did not increase ribotype-027 virulence in a triple phase chemostat gut model, and resulted in supressed toxin production.

Implications of all the available evidence

Trehalose metabolism variants are common within the overall genetic diversity of clinical *C. difficile* and associated with ancient *C. difficile* population structures. Therefore, the role of trehalose metabolism variants in the rise of ribotypes 027 and 078 specifically to clinical importance is questionable. These variants are not unique to these ribotypes and increases in total dietary trehalose between 2000 and 2012 were relatively minimal as a proportion of total dietary trehalose intake. In contrast to data from a mouse model of ribotype-027 infection, the trehalose four-gene cluster is not associated with increased mortality in human infections with another commonly circulating *C. difficile* genotype, ribotype-015. Alternative explanations are required to explain why ribotype-027 and ribotype-078 have been successful.

However, sporadic fluoroquinolone resistance is seen in other less prevalent lineages and is also variably present in the successful ribotype-078 [3], which has been associated with livestock farming rather than healthcare [5]. Therefore, additional potential explanations for the success of ribotypes 027 and 078 have been explored. A recent study by Collins et al. [6] proposed that trehalose metabolism variants in ribotypes 027 and 078, combined with recent increases in human dietary trehalose, have played a significant role in the emergence of these epidemic strains. Two independent mechanisms were identified that facilitate metabolism of low concentrations of the disaccharide sugar trehalose. In ribotype-027 and closely-related ribotypes from *C. difficile* clade-2, a point mutation in the *treR* gene confers a L172I substitution in a transcriptional regulator, TreR, leading to higher expression of the phosphotrehalase enzyme TreA. Ribotype-078 and closely-related clade-5 strains have a four-gene cluster including a phosphotrehalase and a potential trehalose transporter, which Collins *et al* report is absent from other clades [6]. The same authors have since identified a second *treR* variant with a C171S amino acid substitution in TreR in ribotype-017 (from clade-4) [7]; notably these strains are also fluoroquinolone resistant [8]. Enhanced trehalose metabolism in ribotype-027 was associated with increased virulence in a mouse model of CDI [6].

We investigated the plausibility that trehalose metabolism variants conferred a selective advantage for ribotypes 027 and 078 by evaluating how common these variants are within the total genetic diversity of clinical *C. difficile* using previously sequenced isolates. Within a commonly circulating genotype, ribotype-015, we identified variable presence of the trehalose four-gene cluster, which allowed assessment of the impact of variant acquisition on virulence in patients. We also evaluated the extent to which imports of synthetic trehalose altered total dietary trehalose intake during the rise of ribotypes 027 and 078 in Europe, the USA and Canada.

2. Materials and methods

2.1. Genomic data and sequence analysis

Data from 5386 previously whole-genome sequenced *C. difficile* isolates, predominantly cultured from humans with CDI were obtained. These data include clinical isolates from hospital and community patients with CDI from Oxfordshire, UK (2006–2013) [4,9,10], Leeds, UK (2005–2013) [9,10], and the two licensing trials of fidaxomicin in Europe, USA, and Canada (2006–2009) [11,12]. Additionally, genomes from a pan-European survey of CDI [3], healthy children in the UK [13], and ribotype and toxinotype reference collections were included. Raw sequence reads were de novo assembled; 5,232 genomes passed quality control filters. Assemblies were used to determine multi-locus sequence types (STs) and associated clades. BLAST searches were used to identify *treR* variants and the presence of the four-gene cluster described by Collins *et al* [6], see supplement for details and a full list of sequences with European Nucleotide Archive identifiers (Table S1).

For phylogenetic analysis, to avoid over-representing some lineages as a result of clonal transmission, an example of each sequence type (ST) was chosen at random. Where genomes of the same ST contained different trehalose metabolism variants, one random genome with and without the variant was included. For these genomes, sequence data were mapped against the 630 reference genome as previously described [14]. Sequences were compared using single nucleotide polymorphisms, SNPs, obtaining differences between sequences from maximum likelihood phylogenies constructed from mapped read data using PhyML version 3.1 [15] (with generalized time-reversible substitution model and "BEST" tree topology search algorithm), and corrected for recombination using ClonalFrameML version 1.25 [16] (with default settings). Phylogenetic clustering of presence of the four-gene cluster

within Clade 1 was assessed using a Mantel test, based on pairwise genetic distances from the phylogeny and pairwise comparisons of the concordance of the presence or absence of the four-gene cluster. The test was performed using R 3.5.3 with 10,000 permutations. For STs with any genome containing the trehalose four-gene cluster, two further phylogenies were similarly determined using i) the mapped data and ii) the sequence of the trehalose four-gene cluster. Intact DNA sequences spanning the four-gene cluster were extracted from the assembled contigs, aligned using MUSCLE [17], and used as input for this maximum likelihood phylogeny.

2.2. Patient outcome data

Thirty-day all-cause mortality was obtained for all patients from Oxford, Leeds and the fidaxomicin trials using hospital and trial records, respectively, and for non-trial patients verified against the UK National Health Service Spine providing information on mandatory death registrations.

Multivariable logistic regression was used to determine the relationship between 30-day all-cause mortality and the presence or absence of the trehalose four-gene cluster. Short-term all-cause mortality was analysed rather than disease-specific mortality because attribution of mortality to a single cause in this setting is problematic, as patients frequently have multiple co-morbidities and prior infection(s) with pathogens other than *C. difficile*. Additional factors available in the datasets were adjusted for, including age, sex and the infecting ST, accounting for non-linearity in continuous factors using multiple fractional polynomials. Interactions between factors were included in the final model if the interaction p-value was <0.05. Stata 14.1 (Stata Corp, College Station, TX) was used.

2.3. National CDI surveillance data and trehalose import data

CDI incidence data was obtained for England, USA, and Germany. Data from England were obtained from the mandatory national surveillance undertaken by Public Health England. Mandatory reporting of all hospital and community CDI cases in patients ≥65 years was introduced in 2004 and changed to include patients ≥2 years during 2007. Reported incidence for 2004–2007 was therefore scaled up, to be representative of patients ≥2 years, using the ratio of cases in patients ≥2 years and ≥ 65 years in 2008. Voluntary reporting data were used to estimate the total number of cases prior to 2004, scaling the reported number of cases by the ratio of voluntary to scaled mandatory reports in 2004 [4]. Previously published data were used on the number of CDI cases per 1000 hospital discharges in the USA between 2001 and 2010 [18] and from the German national CDAD-KISS surveillance scheme [19]. Overall CDI incidence provides a marker of ribotype-027 and ribotype-078 incidence, in particular as ribotype-027 has been largely responsible for recent rises and falls in overall CDI incidence, e.g. in the UK [4] and USA [20]. CDI incidence was compared with trehalose imports per country (as a proxy for sales and consumption) supplied by the principal global supplier of trehalose, Hayashibara Co. Ltd.

2.4. In vitro gut model

We determined the effect of trehalose, glucose or saline supplementation on CDI, using a well validated and clinically reflective gut model, previously described in detail [21,22]. Triple stage chemostat models were seeded with 30 g of the same pooled faecal slurry from five healthy human donors and allowed to equilibrate before adding *C. difficile* ribotype-027 strain 210 (CD027) [23] spores at days 14 and 24. From previous whole-genome sequencing [24], this strain, in common with other ribotype-027 isolates, contains the L172I substitution in TreR. Sugar/saline supplementation was undertaken for 28 days from day 14 in three separate parallel models, each run once: trehalose (560 mM), glucose (1120 mM) or saline three times daily. The dosing regimens were sufficient to achieve final trehalose and glucose concentrations in the model of 10 mM and 20 mM, respectively, i.e. consistent with levels observed in humans consuming trehalose [25,26]. Clindamycin was instilled from day 24 for 8 days to create the microbial niche to allow CDI. Total *C. difficile* and *C. difficile* spore counts, and toxin levels were obtained on day 3, 10 and daily from day 14 in triplicate until day 55. Presence of any *C. difficile* toxin was detected using a vero cell cytotoxicity assay [22].

2.5. Role of the funding source

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

3. Results

3.1. Sequence data and sequence analysis

Data from 5232 previously whole-genome sequenced *C. difficile* isolates were assembled and analysed (Table 1). The *treR* mutant conferring a L172I substitution was restricted to clade-2 as previously described [6]. However it was widely present within clade-2, in 24/29 (83%) STs (both clinically important and rarely isolated), and therefore does not confer a selective advantage unique to ribotype-027 or specific to other successful lineages from clade-2 such as ribotype-176 and ribotype-244 [27]. Similarly the second *treR* variant conferring a C171S amino acid substitution, described in ribotype-017 [7], was a clade rather than ribotype-specific variant, with only one of 11 STs in clade-4 lacking the substitution. The C171S substitution was also found in a subset of ST41 isolates from clade-2 (while ST41 includes the ribotype-244 lineage that has recently spread across Australia, these outbreak isolates [28] contained the L172I variant rather than C171S). Thus, variants in *treR* appear to be clade-specific, i.e. ancient polymorphisms, that are present in nearly all STs in two of the five *C. difficile* clades.

In contrast to Collins et al [6], the four-gene cluster was found to be variably present in genomes representing all five *C. difficile* clades, and not restricted to clade-5 as proposed previously (Fig. 1). Instead of conferring a selective advantage limited to ribotype-078/ST11 and related genomes, 40/167 (24%) STs represented contained the four-gene cluster, including 33 STs from outside clade-5. Despite the presence of the

Table 1
Prevalence of trehalose metabolism variants in 5232 previously sequenced *Clostridium difficile* genomes.

Clade	*treR*				Four-gene insertion	
	Wildtype	L172I substitution	C171S substitution	*treR* gene not detected	Absent	Present
1	3310			2	2985	327
2	6	1289	2	1	1285	13
3				145		145
4	1[a]		96[a]		94	4
5	374				4	370
Novel/not determined	5				5	

[a] A single clade-4 genome contained nucleotides consistent with a mix of wildtype and the C171S substitution; this genome is not included in the *treR* columns of the table.

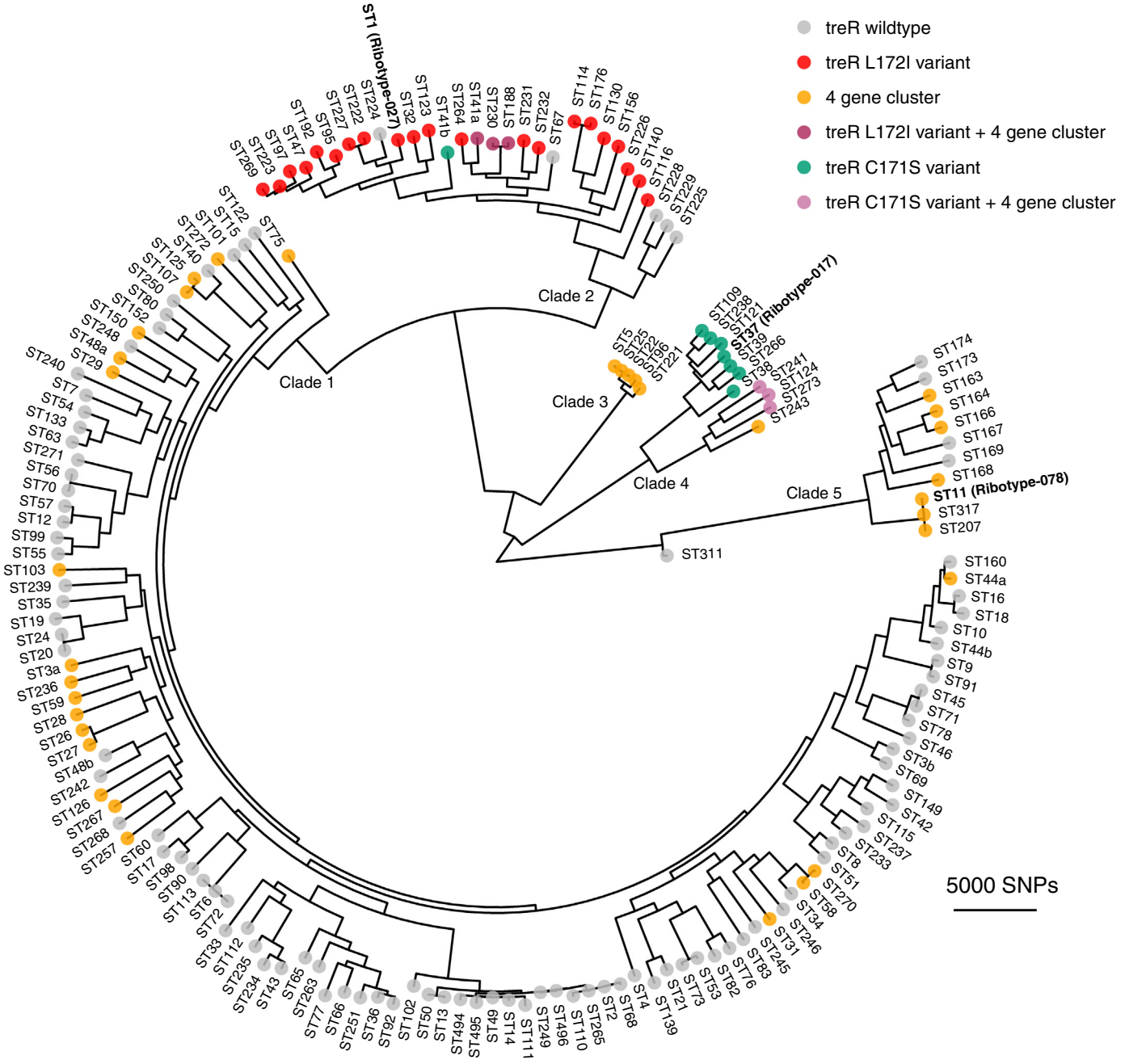

Fig. 1. Phylogenetic distribution of trehalose metabolism variants. ST, muti-locus sequence type. One representative of each ST is shown. The most common ribotype / ST equivalents for ribotypes 027, 078 and 017 are shown. Where different isolates of the same ST contained sequences with differing trehalose variants, (e.g. different examples of the ST were positive or negative for the four-gene cluster) one example of each is shown, denoted, e.g. ST44a, ST44b. The tree depicts a maximum likelihood phylogeny corrected for recombination and with branch lengths measured in units of single nucleotide polymorphisms, SNPs.

four-gene cluster, many of these STs are not commonly isolated from human CDI cases (see Table S2 in supplement). Where present, the four-gene cluster was conserved in a form likely to be functional, with a different *treX* starting position to that previously reported (see supplement, Fig. S1).

Additionally, genomes from clade-3 consistently lacked the canonical *treR* gene, but all contained the four-gene cluster, suggesting strong historical evolutionary pressure to retain trehalose metabolism (Table 1). Three STs within clade-2 and three within clade-4 were identified, which contained both a *treR* variant (L172I or C171S, respectively) and the four-gene insertion (Fig. 1). It may be noteworthy that even with the genetic potential to encode both mechanisms of enhanced trehalose metabolism, these genotypes are clinically rare, although whether the two variants act in an additive manner remains to be demonstrated.

It is likely therefore that trehalose metabolism variants have been established in *C. difficile* long before recent rises in dietary trehalose intake since 2000. Collins et al. [6] report that ribotype-027 isolates from 1985 and 1988 both contain the *treR* L172I mutation. A phylogeny based on concatenated sequence data from the four-gene cluster follows a very similar structure to the phylogeny based on the whole genome (Fig. 2), providing further evidence that the four-gene insertion predates recent rises in ribotype-078. Additionally, there is clustering of the four-gene insertion within Clade 1 (Fig. 1, Mantel test $p < 0.001$), with the clusters sharing common ancestors that are several thousand SNPs distinct from the tips of each cluster. The most

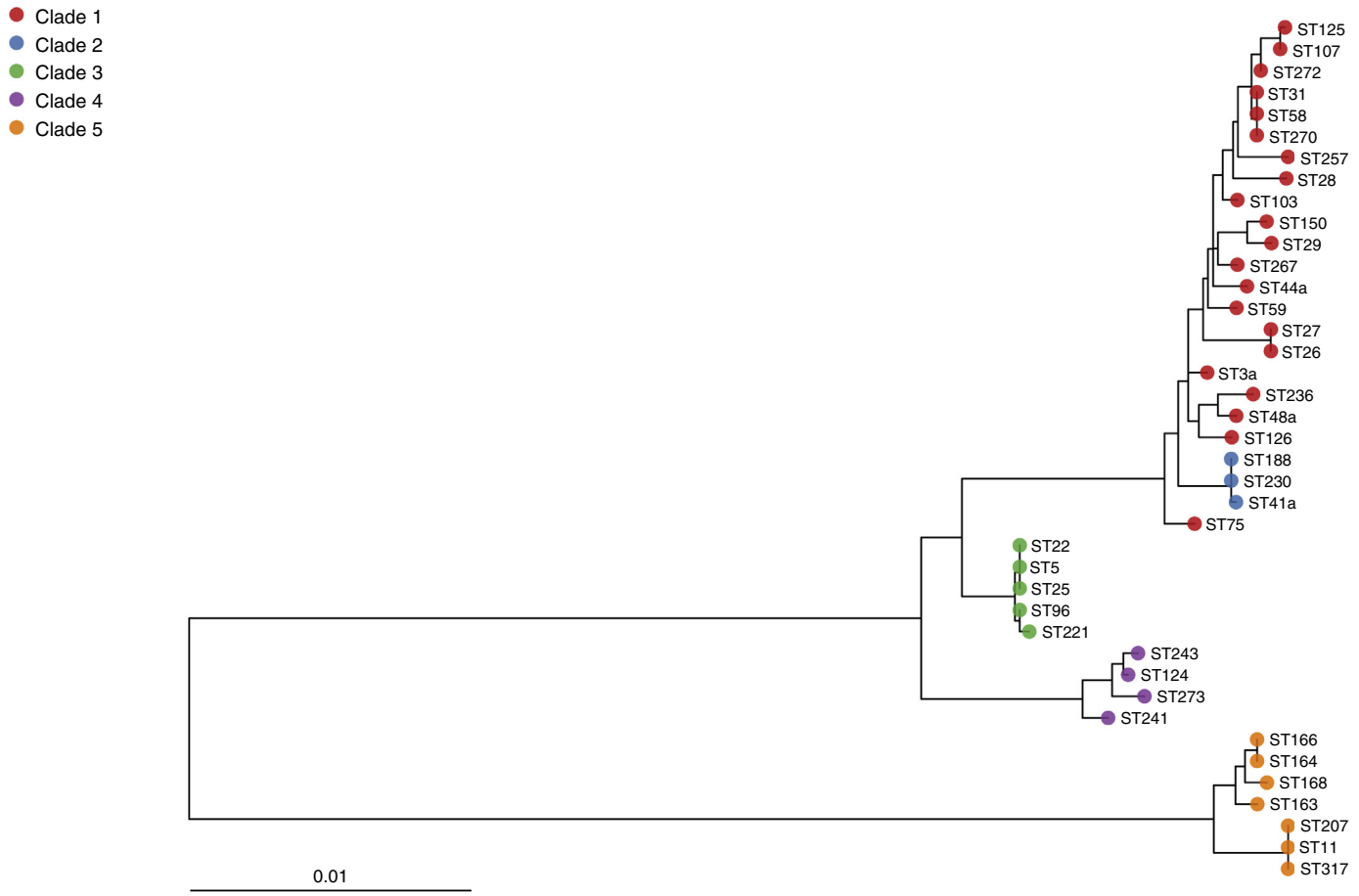

Fig. 2. Phylogeny based on the four-gene trehalose metabolism cluster sequences. Intact DNA sequences spanning the four-gene cluster were extracted from the assembled contigs, aligned and used as input for this maximum likelihood phylogeny. Sequence tips are labelled by their ST, and coloured by clade demonstrating that the phylogeny derived from the four-gene cluster sequence alone follows the population structure of the whole genome sequences.

Table 2
Relationship between trehalose four-gene cluster presence and 30-day mortality from ribotype-015 (ST10/ST44) *C. difficile* infection. There was no evidence for a non-linear relationship between mortality and age using multiple fractional polynomials ($p = 0.80$). There were no significant interactions (all $p > 0.05$).

Factor	Alive, n (%) / median (IQR)	Died, n (%) / median (IQR)	Univariable OR (95% CI)	*p* value	Multivariable OR (95% CI)	*p* value
n	181	27				
Age, years OR per 10-year increase	73 (59–82)	84 (69–88)	1.42 (1.06–1.90)	0.02	1.45 (1.08–1.34)	0.01
Sex						
Female	110 (88%)	15 (12%)				
Male	71 (86%)	12 (14%)	1.23 (0.55–2.80)	0.61	1.27 (0.55–2.95)	0.58
ST						
ST44	103 (88%)	14 (12%)				
ST10	78 (86%)	13 (14%)	1.23 (0.55–2.76)	0.62	0.55 (0.15–2.00)	0.36
Four-gene cluster						
Absent	93 (85%)	17 (15%)				
Present	88 (90%)	10 (10%)	0.62 (0.27–1.43)	0.26	0.36 (0.09–1.34)	0.13

parsimonious explanation for this is that the four-gene insertion was acquired a limited number of times, and lost more recently by STs within these clusters without the insertion. These results, together with rates of *C. difficile* evolution (~1 SNP / genome / year) [14], suggest that the four-gene insertion, along with the L172I variant in clade-2 and C171S in clade-4, are likely to have been stably present in *C. difficile* for thousands of years with variable presence being caused by loss rather than gain over time.

3.2. Patient outcome data

The variable presence and absence of the four-gene cluster in otherwise highly related strains from patients with disease provides an opportunity to test the hypothesis that trehalose metabolism variation results in virulence that is responsible for the adverse outcomes seen in patients with ribotypes 027 and 078. Ribotype-015, a commonly circulating lineage [29], equivalent to ST10 and ST44, variably contains the four-gene cluster. From previous hospital and community patients with ribotype-015 CDI we obtained 208 whole-genome sequences with matched outcome data, 98 (47%) with the four-gene cluster. A total of 27 (13%) died within 30 days of diagnosis. Adjusting for patient age, sex, and infecting ST, there was no evidence of any association between 30-day all-cause mortality and the presence of the four-gene cluster, multivariable odds ratio 0.36 (95% confidence interval 0.09–1.34, $p = 0.13$, Table 2).

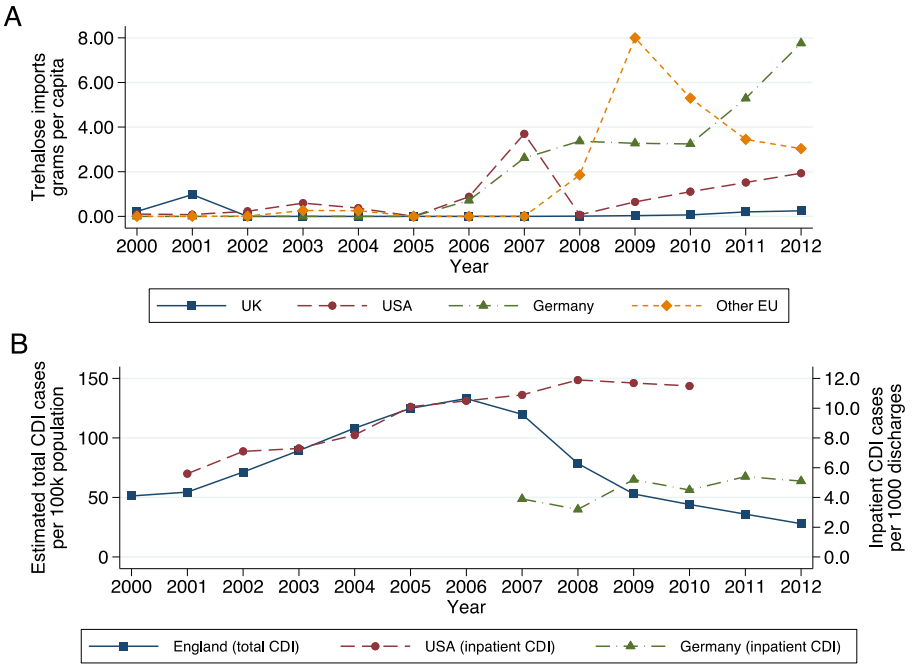

Fig. 3. Temporal relationship between *C. difficile* incidence and trehalose imports. Panel A shows imports of trehalose per capita by year. Trehalose import data were provided by Hayashibara Co. Ltd. Population denominators were obtained from the UK Office of National Statistics, the United States Census Bureau and the European Commission's Eurostat. Panel B shows estimated total CDI cases per 100 k population in England by year and rates of inpatient CDI per 1000 discharges in the USA and Germany.

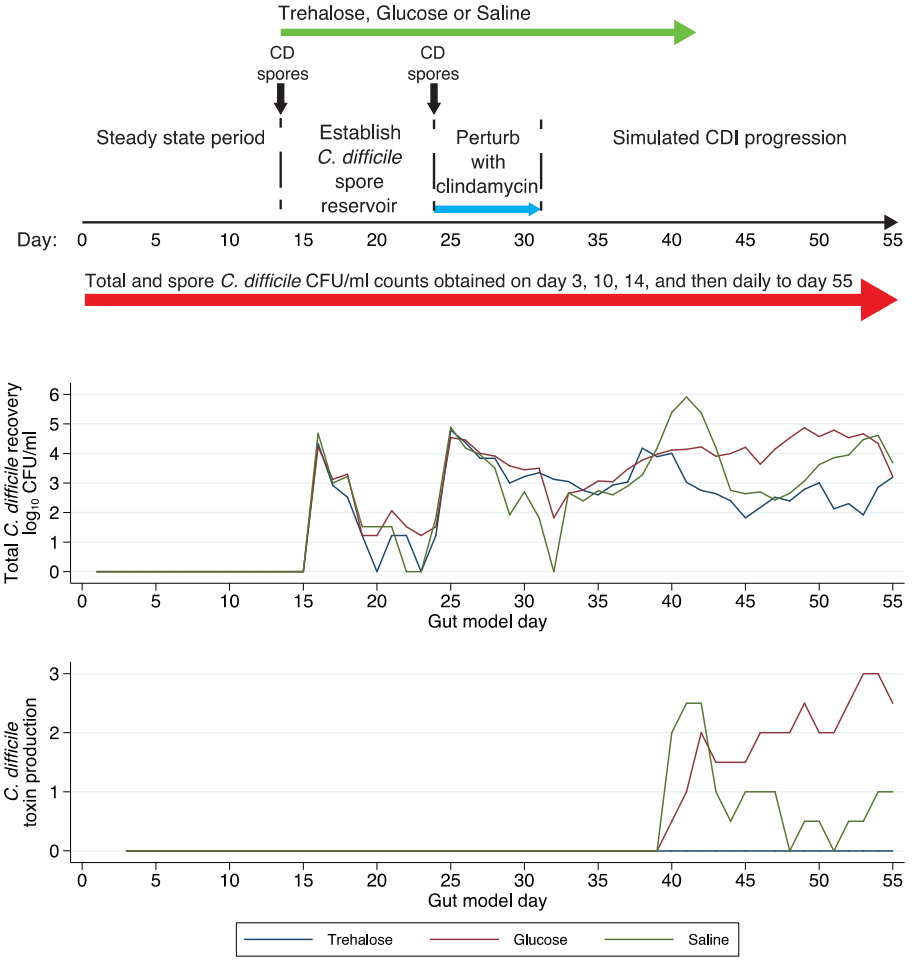

Fig. 4. Gut model timeline, total *C. difficile* counts and toxin detection. CD, *C. difficile*; CFU, colony forming units. Toxin production is a semi-quantitative assay; the scale provided is in log arbitrary units.

3.3. National CDI surveillance data and trehalose import data

To assess the plausibility of synthetic trehalose having substantially altered diets in the USA and Europe from 2000 onwards, i.e. at the same time as CDI incidence increased, we obtained the imported volumes of trehalose and combined these with available population statistics. Fig. 3A shows imports of trehalose per capita per year between 2000 and 2012 in the USA, UK, Germany and the rest of the European Union (EU). Annual imports of trehalose in the USA, UK, Germany and the EU as a whole were < 1 g per capita per year from 2000 to 2006, and increased subsequently until 2012, variably across countries, but remained <9 g per capita per year. Fig. 3B shows the estimated annual total incidence of CDI in England (representing 84% of the UK population) and incidence per 1000 admissions/discharges in hospital inpatients in Germany and the USA. Data from England and the USA [18] show a rise in incidence peaking in mid 2000s, followed by a fall in the UK and stable levels in the USA. In Germany, a smaller rise in incidence is seen in the later 2000s. Comparison of the temporal trends in Fig. 3A and B shows that the increase in trehalose imports postdates the start of the CDI epidemic by at least 5 years.

3.4. In vitro gut model

Total *C. difficile* and spore recovery rates were similar in the three gut models, instilled with trehalose, glucose or saline (Fig. 4, Fig. S2). Spikes in *C. difficile* recovery were seen after addition of spores at day 14 and 24 but were only sustained after induction of simulated CDI with clindamycin. Toxin production was detected from day 40 in the saline and glucose model; however, no toxin production was detected in the trehalose model, despite 42 daily measurements (days 14–55). Post-clindamycin changes in the microbial populations showed quicker recoveries in the glucose/trehalose models of key bacterial populations, such as *Bifidobacterium* spp. (peak recovery at day 36 and 38, respectively) and *Bacteroides* spp. (day 38 and 35), compared with recoveries from saline (day 41 and 39) supplemented model (Fig. S3).

4. Discussion

Convincing evidence, including molecular experiments using gene deletions and recombinants, demonstrates that trehalose metabolism variants in *C. difficile* confer the ability to metabolise low concentrations of trehalose and that this provides a competitive advantage in growth media and laboratory animals [6]. However, we have shown here that trehalose metabolism variants are widespread in multiple distinct *C. difficile* genotypes causing human infections, representing genetically divergent *C. difficile* clades. The *treR* mutations conferring the L172I and C171S substitutions are ancient and widely conserved within clades 2 and 4 respectively. Similarly, the genetic diversity within the four-gene cluster follows the *C. difficile* population structure. While the importance of trehalose metabolism variants in the success of ribotype-027 and ribotype-078 cannot be discounted, other hypotheses are required to explain why these lineages have been so successful when many other lineages with the same trehalose metabolism variants, many from the same clades, have not been.

In contrast, acquisition of antimicrobial resistance does distinguish many of the successful lineages, in particular fluoroquinolone resistance in ribotype 027, 001, 176, 018/356, which have all been associated with healthcare, and local and national clustering of *C. difficile* [3]. Although ribotype-078 lacks consistent fluoroquinolone resistance, this may be less relevant to its success, which has been supported by a strong agricultural niche. Ribotype-078 is widely tetracycline resistant and animal tetracycline administration may be important in promoting its spread [30]. Notably, unlike the healthcare-associated ribotypes above, it does not exhibit phylogeographic structure, but instead is widely-distributed suggesting dissemination via the food chain or another similar route [3].

Within ribotype-027, the TreR L172I variant is associated with increased mortality in the presence of dietary trehalose in a mouse model of CDI [6]. Both ribotype-027 (ST1) and ribotype-078 (ST11) have been associated with increased patient mortality [31] and contain different trehalose metabolism variants. However, these lineages differ by 10,000 s to 100,000 s of SNPs from other control lineages used in outcome studies. Therefore, distinguishing which mutation, gene, or genes are responsible for this phenotype is challenging. An attempt to associate one mutation, e.g. that leading to the TreR L172I substitution, with outcome will be potentially confounded by the many other SNPs that define each clade. Within *C. difficile* ribotype-015, the trehalose four-gene cluster is variably present and absent, providing an opportunity to test for an association with mortality in patients with considerably fewer potentially confounding SNPs. In a population of patients likely to be widely representative of those with CDI, i.e. from consecutive unselected cases in two UK centres and from the two fidaxomicin licensing trials in the USA, Canada and Europe, we found no evidence that the four-gene cluster was associated with increased 30-day all-cause mortality following CDI with ribotype-015. We did not find other examples of common ribotypes with and without trehalose metabolism variants and so were unable to test this finding in other lineages. It is therefore possible that differences in strain background or species differences in intestinal microbiomes could explain the discrepancies between these human findings and previous experimental mouse models.

Given trehalose metabolism variants have likely been stably present within *C. difficile* for thousands of years, recent changes in dietary trehalose have been hypothesised to have enabled trehalose metabolism to play significant role in the emergence of ribotypes 027 and 078 [6]. Low cost production methods for synthesis of trehalose were developed in the 1990s by Hayashibara Co. Ltd., with approval for use in food granted first in Japan in 1995, and then in the USA in 2000, Europe in 2001 and Canada in 2005. Collins *et al* estimated that 21 g of trehalose is consumed per person annually from naturally occurring trehalose in mushrooms and honey, two trehalose rich foods, based on consumption figures from the USA. This was extrapolated to estimate that the total naturally occurring dietary trehalose intake may be ~100 g per capita per year [7]. The Food and Drug Administration GRAS (generally recognised as safe) submission from Hayashibara International Inc. in 2000, notes potential for consumption of up to 34 g of trehalose a day based on levels of trehalose use in foods in Japan and consumption levels of these foods in the USA. However, as shown in Fig. 3A, imports of trehalose between 2000 and 2012 in the USA and Europe were only a fraction of these levels, <1 g per capita per year from 2000 to 2006, and < 9 g per capita per year until 2012. Therefore, synthetic trehalose imports did not significantly alter the total dietary intake of trehalose, particularly during the 2000–2006 when overall incidence of CDI rose substantially in the USA and England, driven by the spread of ribotype-027 [4] (Fig. 3B).

A potential weakness of using import data is that trehalose may be used in food production and then exported and consumed in other countries. Therefore, we explored the example of England in more detail. England contains the 84% of the UK population. The large majority of food consumed in the UK is sourced from the UK and EU, with only 4% from North America [32]. If we make the unlikely, but conservative assumption that all trehalose imported into the EU was consumed in the UK, rather than elsewhere or used for manufacturing development, then <1 g of trehalose per capita per year was consumed in between 2000 and 2005 inclusive, i.e. <1% of naturally occurring consumption [7]. During this time large outbreaks of ribotype-027 occurred [33]. Similarly, up until 2012 the per capita imports to the US only reached a peak of 3.7 g per year in 2007, and were < 1 g per year until 2006, while rises in ribotype-027 cases [34] and total CDI incidence [18] were reported from 2000 onwards. In Canada, trehalose was only licenced for use in 2005; however, a large outbreak of ribotype-027 was reported in Quebec Province from 2003 onwards [35]. Although changes in CDI incidence have been largely driven by changes in ribotype-027, a

limitation of our approach is that we have not been able to specifically analyse temporal trends in lineages with and without trehalose metabolism variants as these are not captured by the national surveillance data used. Another limitation of our approach is that we assume synthetic trehalose is consumed uniformly by a population, however supporting this, we are not aware of any evidence to suggest synthetic trehalose is consumed preferentially by those most at risk of CDI.

In a chemostat gut model of CDI representative of human CDI, supplementation of trehalose did not result in increased levels of *C. difficile* or its spores compared with glucose or saline supplementation. Additionally, supplementation with trehalose actually reduced toxin detection to undetectable levels. To incorporate the many gut microbial families identified in humans, we used a pooled slurry from 5 healthy donors in our gut model. However, we acknowledge the limitation that our model experiment was only run once for each set of test conditions, which limits the generalisability of these data. However, the multiple longitudinal measurements over 8 weeks, from this well studied model, represent replicates of the measurements obtained. The data from these preliminary models highlight several potential areas for further study. Trehalose is readily metabolised into two glucose molecules by the human intestinal flora through production of trehalase. High glucose levels have been shown to repress *C. difficile tcdA* and *tcdB* expression, via the regulator CcpA [36], which could explain the low levels of toxin detected during glucose supplementation, subsequent increases after dosing stopped, and also the absence of toxin production during/after trehalose administration. The difference in the effect of trehalose in our model compared with previously published findings [6] could be attributed to, either the difference in the intestinal flora between humans and mice, or as part of an immune system response, not accounted for in our system. Here, we did not measure expression of *treA* from the models, as any results obtained could not be assigned to our *C. difficile* strain given that many other human intestinal bacterial species also express *treA*. Our dosing schedule allowed us to set the concentration of trehalose and glucose in the first vessel of the model, however we do not have data available on the levels achieved in the second and third vessels.

The main part of our study is based on genomic analysis and does not include functional assessment of the relative fitness of the strains studied. However, it provides a basis for such work to be carried out in the future. Additionally, the power of our study to detect a mortality difference in the ribotype-015 infections with and without the four-gene cluster is limited by the relatively low overall rate of mortality in this patient group, 15%, and the sample size. However, the 30-day all-cause mortality for ribotype-027 and ribotype-078 is approximately 30% [31]; a study of a similar size to ours ($n = 200$, with equal proportions with and without the four-gene cluster) would have 95% power to detect a difference in outcome of for example 10% versus 30%.

In conclusion, we find that trehalose metabolism variants are common within the overall genetic diversity of *C. difficile*. Consequently, additional explanations are required to explain why ribotype-027 and ribotype-078 have been successful when many other lineages sharing the same variants have not. Increases in total dietary trehalose were minimal as a proportion of total dietary trehalose intake, and therefore very unlikely to have contributed significantly to the emergence and spread of ribotype-027 and ribotype-078.

Funding sources

This work was supported by the National Institute for Health Research Oxford Biomedical Research Centre and the Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford in partnership with Public Health England (PHE) [HPRU-2012-10041]. DWC, TEAP and ASW are National Institute for Health Research Senior Investigators. DWE is a National Institute for Health Research Clinical Lecturer and Big Data Institute Robertson Fellow. Funding for the gut model experiments only was provided by Hayashibara Co. Ltd.

Declaration of interests

MHW has received consulting fees from Actelion, Astellas, MedImmune, Merck, Pfizer, Sanofi-Pasteur, Seres, Summit, and Synthetic Biologics; lecture fees from Alere, Astellas, Merck & Pfizer; and grant support from Actelion, Astellas, bioMerieux, Da Volterra, Merck and Summit. SDG has received consulting fees from Abbott, Aquarius Population Health, Astellas and MSD; lecture fees from Astellas, MSD and Orion Diagnostics; and grant support from Astellas. No other author has a conflict of interest to declare.

Author contributions

DWE, DWC, TEAP, ASW, MHW and KED conceived the study. DWE, XD and KED performed the genomic analysis. DWE performed the patient outcome and trehalose import analysis. AB, JF, IM and MHW designed, undertook and analysed the gut model experiments. DWE wrote the first draft of the manuscript. All authors revised the manuscript.

Acknowledgements

We thank Emma Clark, Duncan Ewin, William Spittal and Karen Bentley for technical support with the gut model experiments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.04.038.

References

[1] He M, Miyajima F, Roberts P, et al. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. Nat Genet 2013;45:109–13.
[2] Van Kleef E, Luangasanatip N, Bonten MJ, Cooper BS. Why sensitive bacteria are resistant to hospital infection control. Wellcome Open Res 2017;2:16.
[3] Eyre DW, Davies KA, Davis G, et al. Two distinct patterns of *Clostridium difficile* diversity across Europe indicating contrasting routes of spread. Clin Infect Dis 2018;67:1035–44.
[4] Dingle KE, Didelot X, Quan TP, et al. Effects of control interventions on *Clostridium difficile* infection in England: an observational study. Lancet Infect Dis 2017;17:411–21.
[5] Knetsch CW, Connor TR, Mutreja A, et al. Whole genome sequencing reveals potential spread of *Clostridium difficile* between humans and farm animals in the Netherlands, 2002 to 2011. Euro Surveill 2014;19:20262–954.
[6] Collins J, Robinson C, Danhof H, et al. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. Nature 2018;553:291–4.
[7] Collins J, Danhof H, Britton RA. The role of trehalose in the global spread of epidemic *Clostridium difficile*. Gut Microbes 2018;56:1–6.
[8] Drudy D, Harnedy N, Fanning S, Hannan M, Kyne L. Emergence and control of fluoroquinolone-resistant, toxin A–negative, toxin B–positive *Clostridium difficile*. Infect Control Hosp Epidemiol 2007;28:932–40.
[9] Mawer DPC, Eyre DW, Griffiths D, et al. Contribution to *Clostridium difficile* transmission of symptomatic patients with toxigenic strains who are fecal toxin negative. Clin Infect Dis 2017;64:1163–70.
[10] Martin JSH, Eyre DW, Fawley WN, et al. Patient and strain characteristics associated with *Clostridium difficile* transmission and adverse outcomes. Clin Infect Dis 2018;67:1379–87.
[11] Louie TJ, Miller MA, Mullane KM, et al. Fidaxomicin versus vancomycin for *Clostridium difficile* infection. N Engl J Med 2011;364:422–31.
[12] Cornely OA, Crook DW, Esposito R, et al. Fidaxomicin versus vancomycin for infection with *Clostridium difficile* in Europe, Canada, and the USA: a double-blind, non-inferiority, randomised controlled trial. Lancet Infect Dis 2012;12:281–9.
[13] Stoesser N, Eyre DW, Quan TP, et al. Epidemiology of *Clostridium difficile* in infants in Oxfordshire, UK: risk factors for colonization and carriage, and genetic overlap with regional C. difficile infection strains. PLoS One 2017;12:e0182307.
[14] Eyre DW, Cule ML, Wilson DJ, et al. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med 2013;369:1195–205.
[15] Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010;59:307–21.
[16] Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol 2015;11:e1004041.

[17] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–7.
[18] Steiner C, Barrett M, Terrel L. HCUP projections: *Clostridium difficile* Infection 2011 to 2012. U.S. Agency for Healthcare Research and Quality; 2012, https://www.hcup-us.ahrq.gov/reports/projections/CDI_Regional_projections_Final.pdf (published online Jan 1, accessed Nov 23, 2018).
[19] Nationales Referenzzentrum für Surveillance von nosokomialen Infektionen. KISS Krankenhaus-Infektions-Surveillance-System: Modul CDAD-KISS Referenzdaten; 2018 (accessed July 17, 2018) http://www.nrz-hygiene.de/fileadmin/nrz/module/cdad/201701_201712_CDADRef.pdf.
[20] Lessa FC, Mu Y, Bamberg WM, et al. Burden of *Clostridium difficile* infection in the United States. N Engl J Med 2015;372:825–34.
[21] Macfarlane GT, Macfarlane S, Gibson GR. Validation of a three-stage compound continuous culture system for investigating the effect of retention time on the ecology and metabolism of Bacteria in the human colon. Microb Ecol 1998;35:180–7.
[22] Freeman J, O'Neill FJ, Wilcox MH. Effects of cefotaxime and desacetylcefotaxime upon *Clostridium difficile* proliferation and toxin production in a triple-stage chemostat model of the human gut. J Antimicrob Chemother 2003;52:96–102.
[23] Baines SD, Saxton K, Freeman J, Wilcox MH. Tigecycline does not induce proliferation or cytotoxin production by epidemic *Clostridium difficile* strains in a human gut model. J Antimicrob Chemother 2006;58:1062–5.
[24] Eyre DW, Walker AS, Freeman J, et al. Short-term genome stability of serial *Clostridium difficile* ribotype 027 isolates in an experimental gut model and recurrent human disease. PLoS One 2013;8:e63540.
[25] Australia New Zealand Food Authority. Predicted Trehalose Intakes based on National Food Consumption Data; 2000.
[26] Abbott PJ, Chen J. WHO food additives series 46: Trehalose. published online Jan 1 http://www.inchem.org/documents/jecfa/jecmono/v46je05.htm; 2001.
[27] Valiente E, Dawson LF, Cairns MD, Stabler RA, Wren BW. Emergence of new PCR ribotypes from the hypervirulent *Clostridium difficile* 027 lineage. J Med Microbiol 2012;61:49–56.
[28] Eyre DW, Tracey L, Elliott B, et al. Emergence and spread of predominantly community-onset *Clostridium difficile* PCR ribotype 244 infection in Australia, 2010 to 2012. Euro Surveill 2015;20:21059.
[29] Wilcox MH, Shetty N, Fawley WN, et al. Changing epidemiology of *Clostridium difficile* infection following the introduction of a National Ribotyping-Based Surveillance Scheme in England. Clin Infect Dis 2012;55:1056–63.
[30] Dingle KE, Didelot X, Quan TP, et al. A role for tetracycline selection in recent evolution of agriculture-associated *Clostridium difficile* PCR Ribotype 078. MBio 2019. https://doi.org/10.1128/mbio.02790-18.
[31] Walker AS, Eyre DW, Wyllie DH, et al. Relationship between bacterial strain type, host biomarkers, and mortality in *Clostridium difficile* infection. Clin Infect Dis 2013;56:1589–600.
[32] Department for Environment, Food & Rural Affairs. Food statistics in your pocket 2017 - global and UK supply; 2017 published online Jan 1.
[33] Healthcare Commission. Investigation into outbrekas of Clostridium difficile at Stoke Madeville Hospitals, Buckinghamshire Hospitals NHS Trust. buckshealthcare.nhs.uk; 2006 (accessed Nov 28, 2018) https://www.buckshealthcare.nhs.uk/Downloads/healthcarecommision/HCC-Investigation-into-the-Outbreak-of-Clostridium-Difficile.pdf.
[34] McDonald LC, Killgore GE, Thompson A, et al. An epidemic, toxin gene-variant strain of *Clostridium difficile*. N Engl J Med 2005;353:2433–41.
[35] Loo VG, Poirier L, Miller MA, et al. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. N Engl J Med 2005;353:2442–9.
[36] Bouillaut L, Dubois T, Sonenshein AL, Dupuy B. Integration of metabolism and virulence in *Clostridium difficile*. Res Microbiol 2015;166:375–83.