

Dietary trehalose enhances virulence of epidemic *Clostridium difficile*

J. Collins¹, C. Robinson², H. Danhof¹, C. W. Knetsch³, H. C. van Leeuwen³, T. D. Lawley⁴, J. M. Auchtung¹ & R. A. Britton¹

***Clostridium difficile* disease has recently increased to become a dominant nosocomial pathogen in North America and Europe, although little is known about what has driven this emergence. Here we show that two epidemic ribotypes (RT027 and RT078) have acquired unique mechanisms to metabolize low concentrations of the disaccharide trehalose. RT027 strains contain a single point mutation in the trehalose repressor that increases the sensitivity of this ribotype to trehalose by more than 500-fold. Furthermore, dietary trehalose increases the virulence of a RT027 strain in a mouse model of infection. RT078 strains acquired a cluster of four genes involved in trehalose metabolism, including a PTS permease that is both necessary and sufficient for growth on low concentrations of trehalose. We propose that the implementation of trehalose as a food additive into the human diet, shortly before the emergence of these two epidemic lineages, helped select for their emergence and contributed to hypervirulence.**

Whole-genome sequencing analysis of *C. difficile* ribotype 027 (RT027) strains demonstrated that two independent lineages emerged in North America from 2000 to 2003 (ref. 1). Comparison with historic, pre-epidemic, RT027 strains showed that both epidemic lineages acquired a mutation in the *gyrA* gene, leading to increased resistance to fluoroquinolone antibiotics. While the development of fluoroquinolone resistance has almost certainly played a role in the spread of RT027 strains, fluoroquinolone resistance has also been observed in non-epidemic *C. difficile* ribotypes and identified in strains dating back to the mid-1980s^{2,3}. Thus, other factors probably contributed to the emergence of epidemic RT027 strains.

The prevalence of a second *C. difficile* ribotype, RT078, increased tenfold in hospitals and clinics from 1995 to 2007 and was associated with increased disease severity⁴. However, the mechanisms responsible for increased virulence remain unknown^{5–8}. It is noteworthy that RT027 and RT078 lineages are phylogenetically distant from one another (Extended Data Fig. 1), indicating that the evolutionary changes leading to concurrent increases in epidemics and disease severity might have emerged by independent mechanisms⁹.

RT027 and RT078 strains grow on low trehalose

Ribotype 027 strains exhibit a competitive advantage over non-RT027 strains *in vitro* and in mouse models of *C. difficile* infection¹⁰. To investigate potential mechanisms for increased fitness, we examined carbon source utilization in an epidemic RT027 isolate (CD2015) using Biolog 96-well Phenotype MicroArray carbon source plates (see Methods and Extended Data Table 1). Out of several carbon sources identified that supported CD2015 growth, we found the disaccharide trehalose increased the growth yield of CD2015 by approximately fivefold compared with a non-RT027 strain. To examine the specificity of enhanced growth on trehalose across *C. difficile* lineages, 21 strains encompassing 9 ribotypes were grown on a defined minimal medium (DMM) supplemented with glucose or trehalose as the sole carbon source. All *C. difficile* strains grew robustly with 20 mM glucose; however, only epidemic RT027 ($n = 8$) and RT078 ($n = 3$) strains exhibited enhanced growth on an equivalent trehalose concentration (10 mM; Fig. 1). Increasing the trehalose concentration to 50 mM enabled growth in most ribotypes (Extended Data Fig. 2a).

Molecular basis for RT027 growth on low trehalose

To identify the genetic basis for enhanced trehalose metabolism, we compared multiple *C. difficile* genomes. All *C. difficile* genomes encode a putative phosphotrehalase enzyme (TreA) preceded by a transcriptional repressor (TreR) (Fig. 2a). Phosphotrehalase enzymes metabolize trehalose-6-phosphate into glucose and glucose-6-phosphate. To test whether *treA* was essential for trehalose metabolism, we generated *treA* deletion mutants in the RT027 strain R20291 (R20291 Δ *treA*) and the RT012 strain CD630 (CD630 Δ *treA*) and grew them in DMM supplemented with 50 mM trehalose. The lack of *treA* prevented growth in both knockout strains, which could be complemented by plasmid expression of *treA* (Extended Data Fig. 2b). Thus, *treA* is required to metabolize trehalose.

We next asked whether RT027 strains have altered regulation of the *treA* gene compared with other ribotypes. To test this hypothesis and determine the minimum level of trehalose required to activate *treA* expression, we grew CD2015 (RT027) and CD2048 (RT053) and exposed them to increasing amounts of trehalose. We found that the

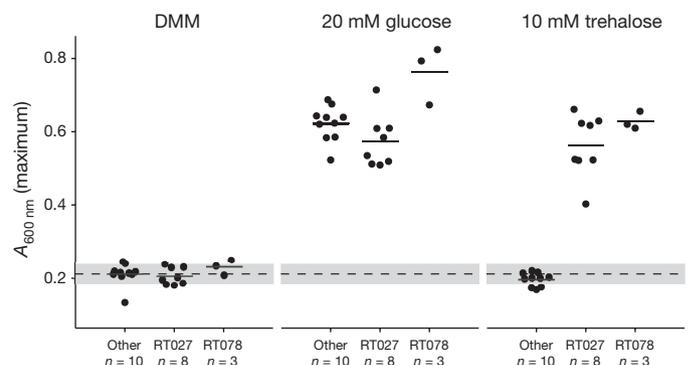


Figure 1 | Only RT027 and RT078 strains show enhanced growth on 10 mM trehalose. Dashed grey line and band indicate mean growth and s.d. in DMM without a carbon source for all samples ($n = 21$). Solid lines are mean growth yield (absorbance at 600 nm, $A_{600\text{ nm}}$) for groups: non-RT027/078 ($n = 10$), RT027 ($n = 8$), and RT078 ($n = 3$). All points represent biologically independent samples.

¹Baylor College of Medicine, Department of Molecular Virology and Microbiology, One Baylor Plaza, Houston, Texas 77030, USA. ²University of Oregon, Institute for Molecular Biology, 1318 Franklin Boulevard, Eugene, Oregon 97403, USA. ³Leiden University Medical Centre, Department of Medical Microbiology, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. ⁴Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK.

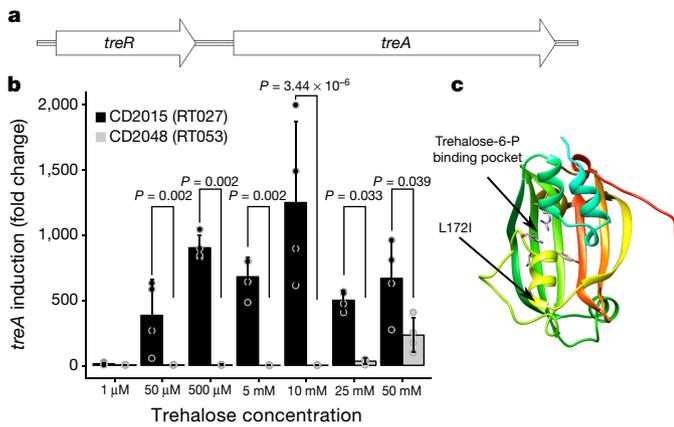


Figure 2 | The *treA* gene is responsible for trehalose metabolism.

a, Trehalose metabolism operon found in all *C. difficile* strains, consisting of a phosphotrehalase (*treA*) and its transcriptional regulator (*treR*). **b**, RT027 strains strongly induce *treA* at 50 μM trehalose and at a significantly higher level than non-RT027 strains ($n = 4$ biologically independent samples per trehalose concentration/strain). Bars are average fold increase, error bars are s.d.; P values derived from t -test (two-tailed) and Holm corrected for multiple comparisons. **c**, Structure of TreR monomer highlighting proximity of L172I mutation to trehalose-6-P binding pocket.

RT027 strain turned on *treA* expression at 50 μM trehalose, a concentration 500-fold lower than that required to turn on *treA* in RT053 (Fig. 2b). To confirm this phenotype, we took four RT027 strains and four non-RT027 strains and measured expression of *treA* in a single trehalose concentration. Again, RT027 strains exhibited significantly higher *treA* expression than all other ribotypes ($P = 0.029$; Extended Data Fig. 3). These results support the idea that RT027 strains are exquisitely sensitive to low concentrations of trehalose.

Sequence alignment of the trehalose operon across 1,010 sequenced *C. difficile* strains revealed a conserved single nucleotide polymorphism (SNP) within the *treR* gene of all RT027 strains (TreR_{RT027}) (Extended Data Fig. 4a). The SNP encodes an L172I amino acid substitution near the predicted effector (trehalose-6-phosphate) binding pocket of TreR (Fig. 2c), a site that is highly conserved across multiple species (93.9% conservation; Extended Data Fig. 4b). This SNP is found not only in every RT027 strain sequenced so far, but also in a newly isolated fluoroquinolone sensitive ribotype (RT244) that has caused community-acquired epidemic outbreaks in Australia^{11,12} and other ribotypes very closely related to RT027, such as RT176 which has caused epidemic outbreaks in the Czech Republic and Poland^{13,14}. Like RT027 strains, the RT244 strains DL3110 and DL3111 can grow on 10 mM trehalose (Extended Data Fig. 2c).

To determine the types of spontaneous mutation that lead to enhanced trehalose utilization, we cultivated several non-RT027/RT078 strains under low trehalose concentrations in minibioreactors¹⁰. After 3 days of continuous cultivation, 13 independent spontaneous mutants capable of growing on low concentrations (<10 mM) of trehalose were isolated. All 13 mutants contained either nonsense or missense mutations in the *treR* gene (Extended Data Table 2).

Effect of trehalose metabolism on disease severity

To test whether the ability of *C. difficile* RT027 strains to metabolize trehalose impacts disease severity, we performed two different experiments. In the first, humanized microbiota mice were challenged with 10^4 spores of either R20291 (RT027, $n = 27$) or R20291Δ*treA* ($n = 28$). After infection, trehalose (5 mM) was provided *ad libitum* in the drinking water and disease progression monitored. The R20291Δ*treA* mutant demonstrated a marked decrease in mortality (33.3% versus 78.6%) compared with R20291 (78% lower risk with R20291Δ*treA*; hazard ratio 0.22; 95% confidence interval 0.09–0.59; $P = 0.003$, likelihood ratio test $P = 0.002$; Fig. 3a). In the second experiment, we infected two groups of humanized microbiota mice with RT027 strain R20291.

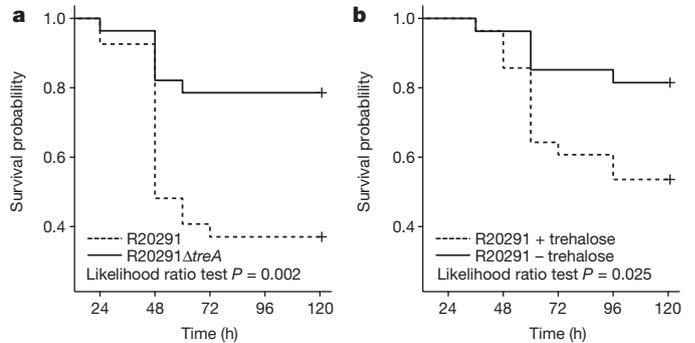


Figure 3 | Trehalose metabolism increases virulence. **a**, Mice infected with R20291Δ*treA* ($n = 27$ animals) have significantly attenuated risk of mortality compared with mice infected with R20291 ($n = 28$ animals) (78% lower risk with Δ*treA* mutant; hazard ratio 0.22; 95% confidence interval 0.09–0.59; $P = 0.003$). **b**, Mice infected with R20291 (RT027) have a significantly higher risk of mortality when trehalose is supplemented in the diet ($n = 28$ animals) than those with no trehalose supplementation ($n = 27$ animals) (threefold increased risk with trehalose; hazard ratio 3.20; 95% confidence interval 1.09–9.42; $P = 0.035$). Experiments were repeated twice. All statistical tests were two-sided.

One group received 5 mM trehalose in water as well as a daily gavage of 300 mM trehalose ($n = 28$) to mimic a dose expected in a meal for humans, whereas the control group ($n = 27$) received a water control. Trehalose addition was found to cause increased mortality compared with the RT027-infected mice without dietary trehalose (threefold increased risk with trehalose; hazard ratio, 3.20; 95% confidence interval 1.09–9.42; $P = 0.035$, likelihood ratio test $P = 0.026$; Fig. 3b). Combined, these results show that metabolism of dietary trehalose can contribute to disease severity of RT027 *C. difficile* strains.

To identify the cause of increased disease severity when trehalose is present, we challenged mice with either R20291 or R20291Δ*treA* and provided 5 mM trehalose *ad libitum* in the drinking water. Forty-eight hours after challenge, *C. difficile* load and toxin levels were measured. Over two independent experiments, no significant difference in *C. difficile* numbers was observed; however, a significant increase in the relative levels of toxin B was detected (median 9.2×10^4 , interquartile range (IQR) 5.1×10^4 to 1.0×10^5 versus median 4.1×10^4 , IQR 2.3×10^4 to 4.6×10^4 , $P = 0.0268$; Extended Data Fig. 5). This increased toxin production could contribute to increased disease severity.

Molecular basis for RT078 growth on low trehalose

Unlike RT027, RT078 strains do not possess the TreR L172I substitution or other conserved SNPs in the *treRA* operon. To identify sequences of potential relevance to trehalose metabolism, we performed whole-genome comparisons. A four-gene insertion was found in all RT078 strains sequenced so far, annotated to encode a second copy of a phosphotrehalase (TreA2, sharing 55% amino acid identity with TreA), a potential trehalose specific PTS system IIBC component transporter (PtsT), a trehalase family protein that is a putative glycan debranching enzyme (TreX), and a second copy of a TreR repressor protein (TreR2, sharing 44% amino acid identity with TreR) (see Fig. 4a). Genomic comparison of publicly available *C. difficile* genomes revealed the four-gene insertion was present in RT078 and the closely related RT033, RT045, RT066 and RT126 ribotypes and absent from reference genomes of any other *C. difficile* lineage (Extended Data Fig. 6).

To test whether the newly acquired transporter (*ptsT*) was responsible for enhanced trehalose metabolism, a *ptsT* deletion mutant was constructed in a RT078 (CD1015) strain. This strain was unable to grow on DMM supplemented with 10 mM trehalose (Fig. 4b), but retained the ability to grow in medium supplemented with 50 mM trehalose (Extended Data Fig. 2d). The growth defect in this deletion mutant (CD1015Δ*ptsT*) was directly due to the lack of *ptsT* since expression of *ptsT* from an inducible promoter could complement growth on 10 mM trehalose (Fig. 4b).

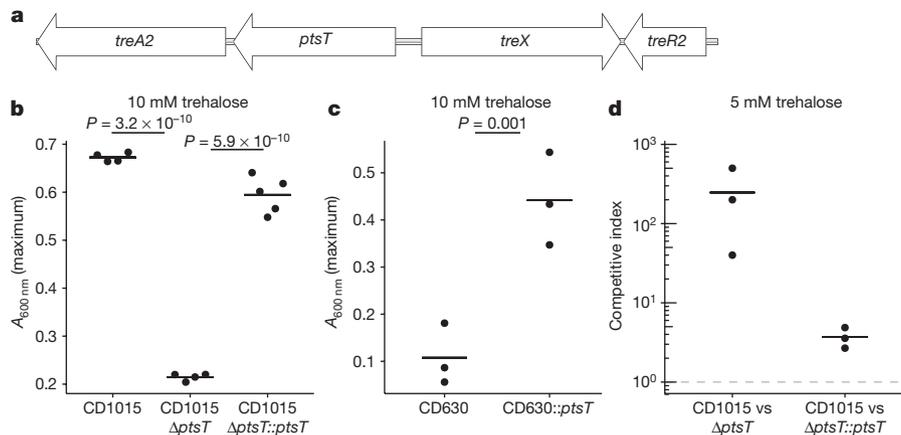


Figure 4 | The *ptsT* gene enables enhanced trehalose metabolism. **a**, Structure of horizontally acquired trehalose metabolism module found in RT078 and closely related strains. **b**, Deletion of the trehalose transporter from a clinical RT078 strain (CD1015) ablates its ability to grow on 10 mM trehalose. Expression of *ptsT* from an inducible plasmid restores growth of CD1015 Δ *ptsT* on 10 mM trehalose (CD1015, $n = 4$; CD1015 Δ *ptsT*, $n = 4$; CD1015 Δ *ptsT*::*ptsT*, $n = 5$). **c**, Expressing *ptsT* from an inducible plasmid enables enhanced growth of CD630 (RT012) on

10 mM trehalose (CD630 $n = 3$; CD630::*ptsT* $n = 3$). **d**, The *ptsT* provides a competitive advantage in complex microbial communities. Dashed grey line (competitive index = 1) indicates equal fitness of the competing strains, points above this line represent out-competition by CD1015. All points (Fig. 4b–d) represent biologically independent samples, bars are mean, P values derived from t -test (two-tailed) and Holm corrected for multiple comparisons where appropriate.

We next tested whether *ptsT* was sufficient to confer enhanced trehalose utilization in a non-ribotype 078 strain, which fails to grow under low trehalose concentrations. To do this, *ptsT* was expressed from an inducible promoter in strain CD630 (RT012). Expression of *ptsT* was sufficient to allow growth of CD630 in DMM supplemented with 10 mM trehalose (Fig. 4c). Taken together, we conclude that *ptsT* is both necessary and sufficient to support growth on low concentrations of trehalose.

To test whether the expression of *ptsT* could confer a fitness advantage, CD1015 (RT078) was competed against its isogenic CD1015 Δ *ptsT* mutant in a human faecal minibioreactor model of *C. difficile* infection¹⁰. After clindamycin treatment of minibioreactor communities to enable infection, CD1015 and CD1015 Δ *ptsT* strains were added together to each reactor and levels monitored over time. Remarkably, the CD1015 strain was found to be significantly more efficient at competing *in vivo* in the presence of a complex microbiota than the CD1015 Δ *ptsT* mutant (mean competitive index of 246 on day 7). To ensure the CD1015 Δ *ptsT* loss was due to the absence of *ptsT*, CD1015 was competed against the CD1015 Δ *ptsT* mutant complemented with *ptsT* from an inducible vector. After 5 days of continuous competition, the wild-type RT078 had a mean competitive index of just 3.7 (Fig. 4d). Hence, *ptsT* provides a competitive fitness advantage to RT078 strains.

Trehalose is observed in the distal gut

Despite the presence of a localized brush border trehalase enzyme in the small intestine, human studies suggest that high levels of trehalose consumption can result in significant amounts reaching the distal ileum and colon^{15–17}. To demonstrate that a significant amount of dietary trehalose can survive transit through the small intestine, we gavaged mice with 100 μ l (300 mM) trehalose (equivalent to the suggested concentration in ice cream) and measured trehalose levels in the caecum over time. Using clinical *C. difficile* strains as biosensors, we found the level of trehalose to be sufficient to activate *treA* gene expression in the RT027 strain CD2015 but not in RT053 strain CD2048 (Fig. 5a). To test whether we could detect a low dietary amount of trehalose, we gavaged antibiotic-treated mice with 100 μ l (5 mM) trehalose and measured *treA* activation in these same strains. Again, the RT027 strain showed significant *treA* activation (Fig. 5b). Finally, to determine whether trehalose is bioavailable in humans at sufficient levels to be used by epidemic *C. difficile* isolates, we tested ileostomy effluent from three anonymous donors consuming their normal diets. In two of three samples, *treA* expression was strongly induced in the RT027 strain CD2015 but

not in the RT053 strain CD2048 (Fig. 5c), supporting the notion that levels of trehalose found in food are sufficient to be used by epidemic *C. difficile* strains.

Discussion

Containing an α,α -1,1-glucoside bond between two α -glucose units, trehalose is a non-reducing and extremely stable sugar, resistant both to high temperatures and to acid hydrolysis. Although considered an ideal sugar for use in the food industry, the use of trehalose in the United States and Europe was limited before 2000 owing to the high cost of production (approximately US\$700 per kilogram). The innovation of a novel enzymatic method for low-cost production from starch made it commercially viable as a food supplement (approximately US\$3 per kilogram)¹⁸. Granted ‘generally recognized as safe’ status by the US Food and Drug Administration in 2000 and approved for use in food

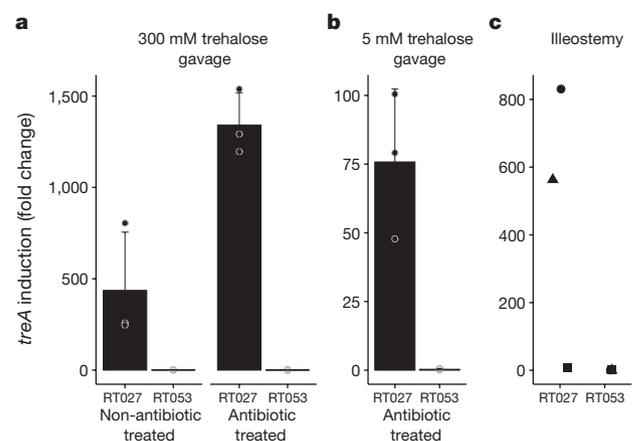


Figure 5 | Trehalose can be detected in mouse caecum and human ileostomy fluid. **a**, Twenty minutes after gavage, trehalose reaches high enough levels in the mouse caecum to turn on expression of *treA* in RT027 but not non-RT027 in both non-antibiotic- and antibiotic-treated mice ($n = 3$ animals per trehalose concentration/strain). **b**, Trehalose can be detected by RT027 but not non-RT027 in the caecum of antibiotic-treated mice gavaged with just 100 μ l of 5 mM trehalose ($n = 3$ animals per group). **c**, RT027 strains can detect trehalose in two out of three human ileostomy fluid samples tested from patients eating a normal (no deliberate trehalose addition) diet. Points represent biologically independent replicates, bars are average fold increase, error bars are s.d.

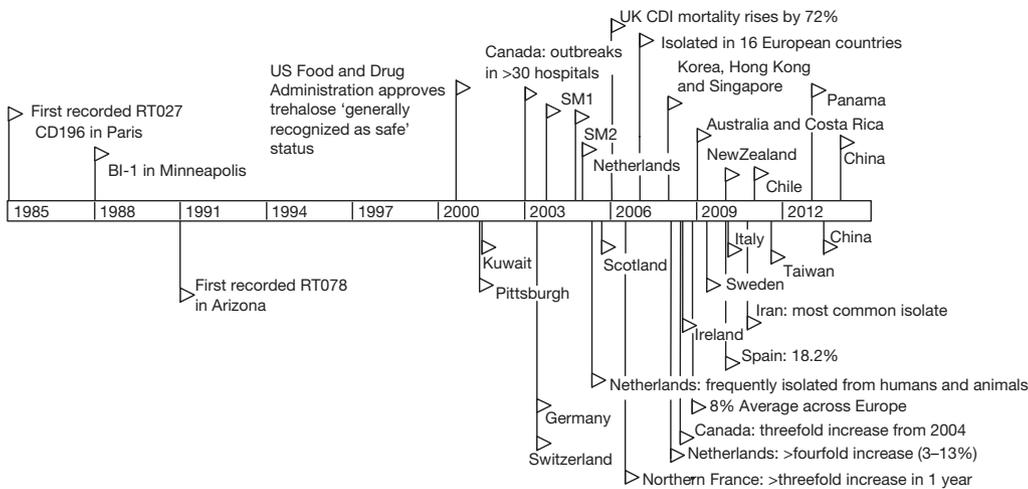


Figure 6 | Timeline of trehalose adoption and spread of RT027 and RT078 lineages. Flags indicate reported outbreaks or first reports of RT027 (top) or RT078 (bottom) in PubMed. SM1 and SM2, outbreaks at Stoke Mandeville Hospital, Buckinghamshire, UK. CDI, *C. difficile* infection.

in Europe in 2001, reported expected usage ranges from concentrations of 2% to 11.25% for foods including pasta, ground beef, and ice cream. The widespread adoption and use of trehalose in the diet coincides with the emergence of both RT027 and RT078 outbreaks (Fig. 6).

Several lines of evidence support the idea that dietary trehalose has participated in the spread of epidemic *C. difficile* ribotypes. First, the ability of RT027 and RT078 strains to metabolize trehalose was present before epidemic outbreaks. The earliest retrospectively recorded RT027 isolate was the non-epidemic strain CD196, isolated in 1985 in a Paris hospital¹⁹. Three years later in 1988, another non-epidemic strain RT027 (BI1) was isolated in Minneapolis, Minnesota. Both isolates, in addition to every RT027 strain sequenced so far, contain the L172I substitution in TreR. RT078 strains were also present in humans before 2001, but epidemic outbreaks were not reported until 2003 (ref. 4). Second, RT027 and RT078 lineages are phylogenetically distant clades of *C. difficile*, yet have convergently evolved distinct mechanisms to metabolize low levels of trehalose. Third, increased disease severity of a RT027 strain that can metabolize trehalose in our mouse model of *C. difficile* infection is consistent with increased virulence of RT027 and RT078 ribotypes observed in patients. Fourth, the ability to metabolize trehalose at lower concentrations confers a competitive growth advantage in the presence of a complex intestinal community. Finally, levels of trehalose in ileostomy fluid from patients eating a normal diet are sufficiently high to be detected by RT027 strains. On the basis of these observations, we propose that the widespread adoption and use of the disaccharide trehalose in the human diet has played a significant role in the emergence of these epidemic and hypervirulent strains²⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 May; accepted 28 November 2017.

Published online 3 January 2018.

1. He, M. *et al.* Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat. Genet.* **45**, 109–113 (2013).
2. Spigaglia, P. *et al.* Fluoroquinolone resistance in *Clostridium difficile* isolates from a prospective study of *C. difficile* infections in Europe. *J. Med. Microbiol.* **57**, 784–789 (2008).
3. Spigaglia, P., Barbanti, F., Dionisi, A. M. & Mastrantonio, P. *Clostridium difficile* isolates resistant to fluoroquinolones in Italy: emergence of PCR ribotype O18. *J. Clin. Microbiol.* **48**, 2892–2896 (2010).
4. Jhung, M. A. *et al.* Toxinotype V *Clostridium difficile* in humans and food animals. *Emerg. Infect. Dis.* **14**, 1039–1045 (2008).
5. Goorhuis, A. *et al.* Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. *Clin. Infect. Dis.* **47**, 1162–1170 (2008).
6. Gupta, A. & Khanna, S. Community-acquired *Clostridium difficile* infection: an increasing public health threat. *Infect. Drug Resist.* **7**, 63–72 (2014).
7. Limbago, B. M. *et al.* *Clostridium difficile* strains from community-associated infections. *J. Clin. Microbiol.* **47**, 3004–3007 (2009).

8. Walker, A. S. *et al.* Relationship between bacterial strain type, host biomarkers, and mortality in *Clostridium difficile* infection. *Clin. Infect. Dis.* **56**, 1589–1600 (2013).
9. He, M. *et al.* Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc. Natl Acad. Sci. USA* **107**, 7527–7532 (2010).
10. Robinson, C. D., Auchung, J. M., Collins, J. & Britton, R. A. Epidemic *Clostridium difficile* strains demonstrate increased competitive fitness compared to nonepidemic isolates. *Infect. Immun.* **82**, 2815–2825 (2014).
11. Lim, S. K. *et al.* 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 (2014).
12. Eyre, D. W. *et al.* Emergence and spread of predominantly community-onset *Clostridium difficile* PCR ribotype 244 infection in Australia, 2010 to 2012. *Euro Surveill.* **20**, 21059 (2015).
13. Polivkova, S., Krutova, M., Petrova, K., Benes, J. & Nyc, O. *Clostridium difficile* ribotype 176 – a predictor for high mortality and risk of nosocomial spread? *Anaerobe* **40**, 35–40 (2016).
14. Rupnik, M. *et al.* Distribution of *Clostridium difficile* PCR ribotypes and high proportion of 027 and 176 in some hospitals in four South Eastern European countries. *Anaerobe* **42**, 142–144 (2016).
15. Bergoz, R. Trehalose malabsorption causing intolerance to mushrooms. Report of a probable case. *Gastroenterology* **60**, 909–912 (1971).
16. Bergoz, R., Bolte, J. P. & Meyer zum Bueschenfelde, K.-H. Trehalose tolerance test. Its value as a test for malabsorption. *Scand. J. Gastroenterol.* **8**, 657–663 (1973).
17. Oku, T. & Nakamura, S. Estimation of intestinal trehalase activity from a laxative threshold of trehalose and lactulose on healthy female subjects. *Eur. J. Clin. Nutr.* **54**, 783–788 (2000).
18. Higashiyama, T. Novel functions and applications of trehalose. *Pure Appl. Chem.* **74**, 1263–1269 (2002).
19. Stabler, R. A. *et al.* Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* **10**, R102 (2009).
20. Leffler, D. A. & Lamont, J. T. *Clostridium difficile* infection. *N. Engl. J. Med.* **372**, 1539–1548 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the National Institutes of Health (U01AI124290-01 and 5U19AI09087202). We thank the members of the Ostomy Association of Greater Lansing for anonymously donating samples, and D. Lyras for the RT244 strains. We thank V. Young, D. Mills, J. Walter and M. Costa-Mattioli for comments on the manuscript.

Author Contributions Concept and design of study: R.A.B., J.M.A., J.C. and C.R. Experiments: *C. difficile* growth, J.C. and C.R.; identification of L172I SNP and comparative analysis, C.R. and J.C.; *treA* RT-qPCR, H.D.; mouse infection model, J.C.; genetic manipulation of *C. difficile* strains, J.C. and C.R.; identification of RT078 trehalose insertion, C.W.K., H.C.L. and T.D.L.; faecal mini-bioreactor competitions, J.M.A.; spontaneous *C. difficile* mutant identification, H.D.; analysis, J.C., C.R., H.D., J.M.A. and R.B. The manuscript was drafted by J.C., J.M.A., and R.A.B., and revised by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.A.B. (robert.britton@bcm.edu).

Reviewer Information *Nature* thanks J. Ballard, E. Pamer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Bacterial strains and growth. A full list of strains can be found in Extended Data Table 3. Carbon source utilization of CD630 (RT012) and CD2015 (clinical RT027) was performed using Biolog Phenotypic Microarray plates. Growth studies were performed under anaerobic conditions (5% hydrogen, 90% nitrogen, 5% carbon dioxide). Strains were cultured overnight in BHI media (Difco) supplemented with 0.5% (w/v) yeast extract. Growth assays used a DMM as described previously²¹ supplemented with either trehalose or glucose as indicated. Anhydrous tetracycline was used at 500 ng ml⁻¹ to induce expression of *ptsT* or *treA* from ectopic expression vectors.

Comparative genomics. To identify unique functional features in RT078 strains, we reviewed publicly available *C. difficile* genomes covering all phylogenetic lineages⁹ using a tool based on BLASTX comparisons of protein annotations²². The genomes included in the analysis were PCR RT012 (strain 630, lineage I), RT027 (R20291, lineage II), PCR RT017 (CF5, M68 lineage IV), and RT078 (QCD-23m63, CDM120 lineage V).

Genetic manipulation of *C. difficile*. Inactivation of *treA* in CD630 was accomplished by group-II intron-directed insertion as previously described²³. Primers were designed to target intron to insert at base pair 177 of *treA* of CD630 (IBS1.2, EBS1, and EBS2; all primers are described in Extended Data Table 3). The resulting *treA* insertion–deletion mutant was verified by PCR using a primer pair (CR064–CR065) designed to flank the *treA* insertion site, resulting in a 350 bp product for the wild-type gene and a 2.4 kbp product for the gene knockout.

Clean deletions in R20291 and CD1015 were performed using a *pyrE* allelic exchange system as described previously²⁴. This is the first case of the *pyrE* allelic exchange system being used in the RT078 lineage, which required generation of CD1015Δ*pyrE* before further deletions. Complementation of *treA* and *ptsT* was performed using an anhydrous tetracycline-inducible system as described previously²⁵. All plasmid conjugations into *C. difficile* strains were performed with *Escherichia coli* SD46. Cloning was accomplished with a combination of restriction digest and ligase cycling reactions as described previously²⁶. Primers and detailed plasmid maps for construction of knockout strains are available at the links provided in Extended Data Table 3.

Quantitative PCR with reverse transcription. Strains were grown overnight and subcultured 1:50 into DMM supplemented with 20 mM succinate. Upon reaching an $A_{600\text{nm}}$ of 0.2–0.3, indicated concentrations of trehalose were added to the culture. After incubation for 30 min, *C. difficile* cells were collected by centrifugation, resuspended in RNALater solution (Invitrogen), and stored at –80°C. Cells were resuspended in 1 ml RLT buffer (Qiagen RNeasy Kit) and lysed by bead beating (2 × 1 min) at 4°C followed by RNA extraction according to the manufacturer's instructions. cDNA was synthesized using Invitrogen Superscript III reverse transcriptase following the recommended protocol. Quantitative PCR reactions were performed in triplicate using Power SYBR Green PCR Master Mix (ABI) with either *C. difficile* 16s (JP048–JP049)- or *treA* (CR045–CR046)-specific primers. Standard curves of cDNA were run to determine primer efficiencies and calculated as in ref. 27. Expression of *treA* was determined using an average of triplicate C_T values from each biological sample.

Mouse model of *C. difficile* infection. Humanized microbiota mice were derived from an initial population of germ-free C57BL/6 mice stably colonized with human gut microbiota and validated for use as a model of *C. difficile* infection²⁸. Humanized microbiota mice aged 6–8 weeks of both sexes were treated with a five-antibiotic cocktail consisting of kanamycin (0.4 mg ml⁻¹), gentamicin (0.035 mg ml⁻¹), colistin (850 U ml⁻¹), metronidazole (0.215 mg ml⁻¹), and vancomycin (0.045 mg ml⁻¹) administered *ad libitum* in drinking water for 4 days. Water was switched to antibiotic-free sterile water and 24 h later mice were administered an intraperitoneal injection of clindamycin (10 mg per kg (body weight)). After a further 24 h, mice were challenged with 10⁴ *C. difficile* spores by oral gavage. Sterile drinking water containing 5 mM trehalose was provided *ad libitum* (for the with or without trehalose study, mice were administered an additional 100 μl oral gavage of 300 mM trehalose daily) and mice were monitored for signs of disease.

In a separate experiment to determine *C. difficile* colonization load and toxin production, mice were euthanized 48 h after challenge with either R20291 or R20291Δ*treA*. *C. difficile* levels in caecal contents were determined by qPCR of toxin genes¹⁰. Relative toxin levels were assessed using a Vero Cell rounding assay¹⁰. Sample sizes for all experiments were determined using power analysis based upon previous experimental data. No randomization of animals was performed; however, all groups were checked to ensure no significant difference in the age, weight, or sex of mice between groups before starting experiments. All animal use was approved by the Animal Ethics Committee of Baylor College of Medicine (protocol number AN-6675). The investigators were not blinded to allocation during experiments and outcome assessment.

Detection of trehalose in caecal contents and human ileostomy fluid. Antibiotic-treated groups were pre-treated with the five-antibiotic cocktail for 3 days. Mice

were gavaged with 100 μl of 5 mM trehalose, 300 mM trehalose, or water. Twenty minutes after gavage, mice were euthanized, and caecal contents harvested and vigorously mixed with two volumes/weight ice cold DMM (no carbohydrate). Supernatant was separated by centrifugation, filter sterilized, and reduced in an anaerobic chamber overnight before use. Ileostomy effluent from three anonymous donors was self-collected into sterile containers and stored at –20°C until thawed, filter sterilized, and used for assay.

Strains were grown overnight and subcultured 1:50 into DMM supplemented with 20 mM succinate. Upon reaching an $A_{600\text{nm}}$ of 0.2–0.3, cells were collected by centrifugation and resuspended in approximately 300 μl caecal or ileostomy fluid and incubated anaerobically for 30 min. Cells were then centrifuged and resuspended in RNALater (Invitrogen) before qRT-PCR analysis.

Bioreactor model for RT078 Δ*ptsT* competition. Faecal communities were established in continuous-flow mini-bioreactor arrays as previously described¹⁰ using bioreactor defined medium²⁹ without starch (BDM4). Communities were disrupted by addition of clindamycin (250 μg ml⁻¹) continuously supplied in the medium for 4 days. After clindamycin treatment, communities were supplied BDM4 without clindamycin supplemented with trehalose (5 mM final concentration, BDM4_{tre}). After 1 day of growth in BDM4_{tre}, to allow washout of clindamycin, communities were challenged with a mixture of exponentially growing CD1015 strains (RT078 wild type and Δ*ptsT*). The competitive index was determined by dividing the proportion of wild-type cells at the end of the competition by the proportion at the start. The competitive index of wild type:Δ*ptsT* strains was determined by qPCR. The competitive index of wild type versus CD1015Δ*ptsT*::ahTCptsT was calculated by selective plating.

Isolation of spontaneous *treR* mutants. *C. difficile* strains were inoculated into continuous-flow mini-bioreactor arrays as previously described¹⁰ using bioreactor defined medium²⁹ without starch (BDM4) supplemented with 5 mM trehalose (BDM4_{tre}). Every 24 h after the start of the experiment, 200 μl PBS containing 100 mM trehalose was spiked into each mini-bioreactor. The reactors were sampled daily, serially diluted, and plated to DMM agar supplemented with 10 mM trehalose. Resulting colonies were streak purified, and the ability to grow on low trehalose (10 mM) verified on plates and in broth culture. The *treR* gene was sequenced and compared with the isogenic parent strain.

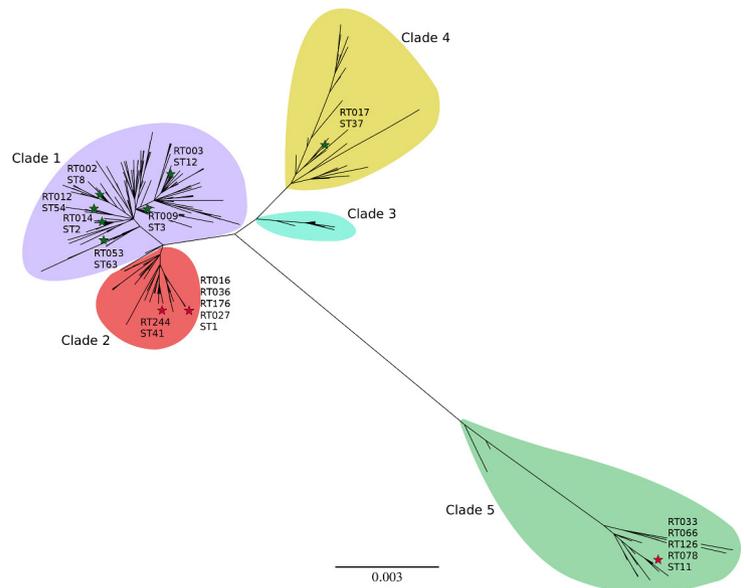
Statistics. Statistical analyses were performed using R (version 3.3.2). A Student's two-sample *t*-test (two-tailed) was used for comparisons of continuous variables between groups with similar variances; Welch's two-sample *t*-test (two-tailed) was used for comparisons of continuous variables between groups with dissimilar variances. *P* values from multiple comparisons were corrected using the Holm method³⁰. A Wilcoxon rank-sum test with continuity correction was used for the toxin assay where data were non-normal. Fold-change data from *treA* gene expression experiments were log-normalized before statistical analysis. Data were visualized using individual data points and group means. Cox proportional hazards models and likelihood ratio tests were used to test significant differences in survival distributions among *C. difficile*-challenged groups of animals.

Collection of human bio-specimens. For faecal samples, live participants who were self-described as healthy and had not consumed antibiotics within the previous 2 months were recruited to provide faecal samples for human faecal bioreactor experiments. Informed consent was obtained before collection of samples and no identifying information was obtained along with the sample. Faecal samples were collected in sterile containers, transported to the laboratory on ice in the presence of anaerobic gas packs (BD Biosciences) within 16 h of collection, manually homogenized in an anaerobic environment, aliquoted into anaerobic tubes, and sealed and stored at –80°C until use. Participants who had ileostomies placed owing to previous, undisclosed illnesses were recruited to provide ileostomy effluents. Informed consent was obtained before collection of samples and no identifying information was obtained with the sample. After transfer from the ostomy bag to a sterile collection container, ileostomy samples were transported to the laboratory on ice within 12 h of collection. Upon receipt, samples were stored at –20°C. Ileostomy donors were recruited through the Ostomy Association of Greater Lansing, and were most probably residents of Lansing, Michigan, USA, and its surrounding counties. Samples were stored at –80°C or –20°C for 3–4 years before use. Samples were randomly selected for testing from a bank of available samples. Samples were collected according to a protocol approved by the Institutional Review Board of Michigan State University (protocol number 10-736SM).

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request. Source data for Figs 1–5 and Extended Data Figs 2, 3 and 5 are provided with the paper.

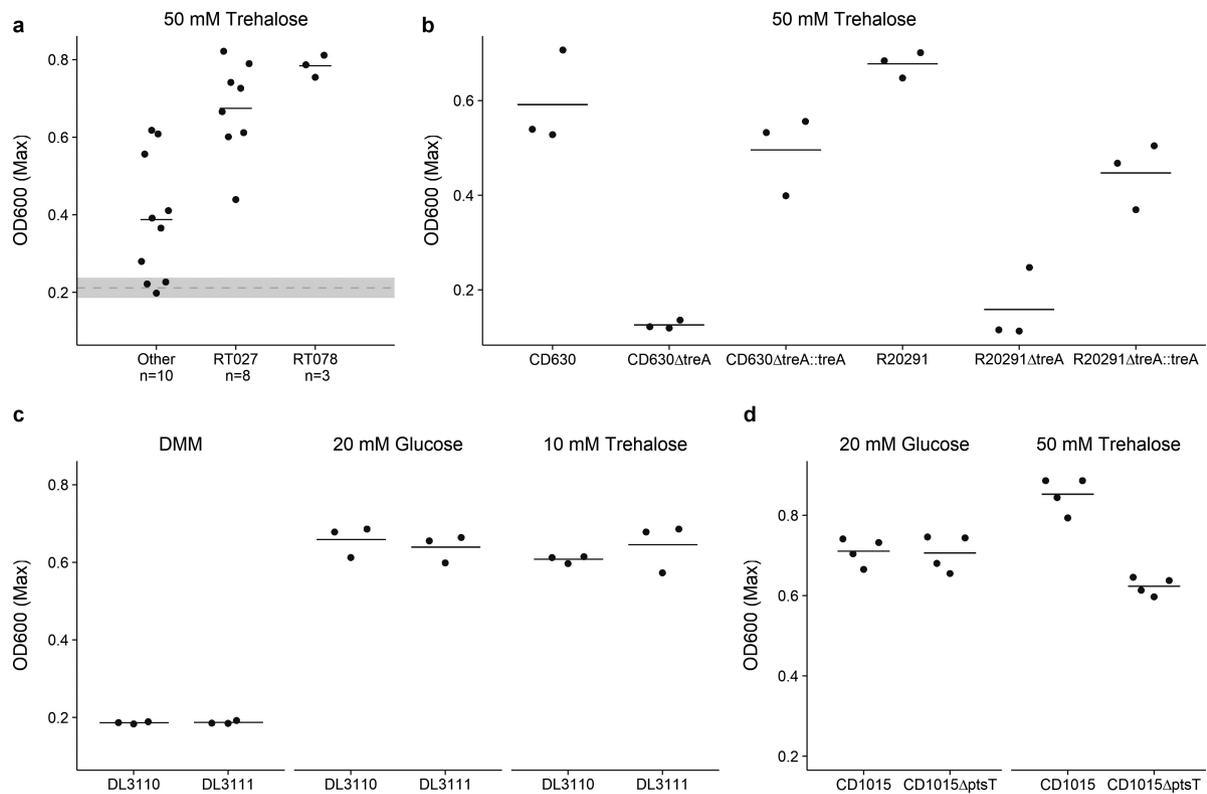
21. Theriot, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nat. Commun.* **5**, 3114 (2014).

22. Knetsch, C. W. *et al.* Genetic markers for *Clostridium difficile* lineages linked to hypervirulence. *Microbiology* **157**, 3113–3123 (2011).
23. Bouillaut, L., Self, W. T. & Sonenshein, A. L. Proline-dependent regulation of *Clostridium difficile* Stickland metabolism. *J. Bacteriol.* **195**, 844–854 (2013).
24. Ng, Y. K. *et al.* Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using pyrE alleles. *PLoS ONE* **8**, e56051 (2013).
25. Fagan, R. P. & Fairweather, N. F. *Clostridium difficile* has two parallel and essential Sec secretion systems. *J. Biol. Chem.* **286**, 27483–27493 (2011).
26. de Kok, S. De *et al.* Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth. Biol.* **3**, 97–106 (2014).
27. Pfaffl, M. W. in *Real-time PCR* (ed. Dorak, T.) 63–82 (Taylor & Francis, 2006).
28. Collins, J., Auchtung, J. M., Schaefer, L., Eaton, K. A. & Britton, R. A. Humanized microbiota mice as a model of recurrent *Clostridium difficile* disease. *Microbiome* **3**, 35 (2015).
29. Auchtung, J. M., Robinson, C. D., Farrell, K. & Britton, R. A. in *Clostridium difficile: Methods and Protocols* (eds Roberts, A. P. & Mullany, P.) 235–258 (Springer, 2016).
30. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
31. Griffiths, D. *et al.* Multilocus sequence typing of *Clostridium difficile*. *J. Clin. Microbiol.* **48**, 770–778 (2010).
32. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
33. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
34. Roca, A. I., Abajian, A. C. & Vigerust, D. J. ProfileGrids solve the large alignment visualization problem: influenza hemagglutinin example. *F1000 Res.* **2**, 2 (2013).
35. Dingle, T. C., Mulvey, G. L. & Armstrong, G. D. Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. *Infect. Immun.* **79**, 4061–4067 (2011).
36. Popoff, M. R., Rubin, E. J., Gill, D. M. & Boquet, P. Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infect. Immun.* **56**, 2299–2306 (1988).
37. Smith, C. J., Markowitz, S. M. & Macrina, F. L. Transferable tetracycline resistance in *Clostridium difficile*. *Antimicrob. Agents Chemother.* **19**, 997–1003 (1981).



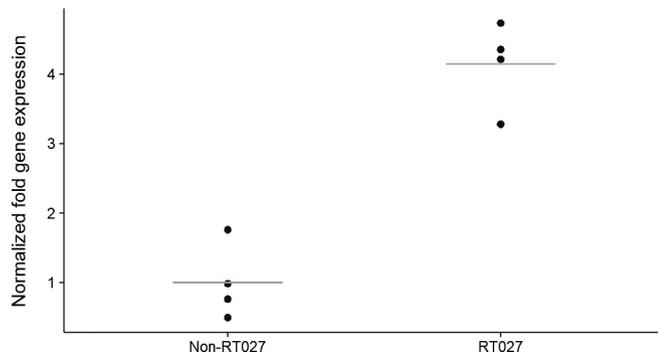
Extended Data Figure 1 | Phylogenetic organization of *C. difficile* MLST profiles. Maximum likelihood tree based upon concatenated multi locus sequence typing genes of the 399 current profiles available at <https://pubmlst.org/cdifficile/>³¹. Stars indicate position of strains used in this

study, with red stars indicating sequence types possessing either the TreR L172I amino acid substitution (ST1, ST41) or four-gene insertion (ST11). Tree constructed using MEGA7 (ref. 32).

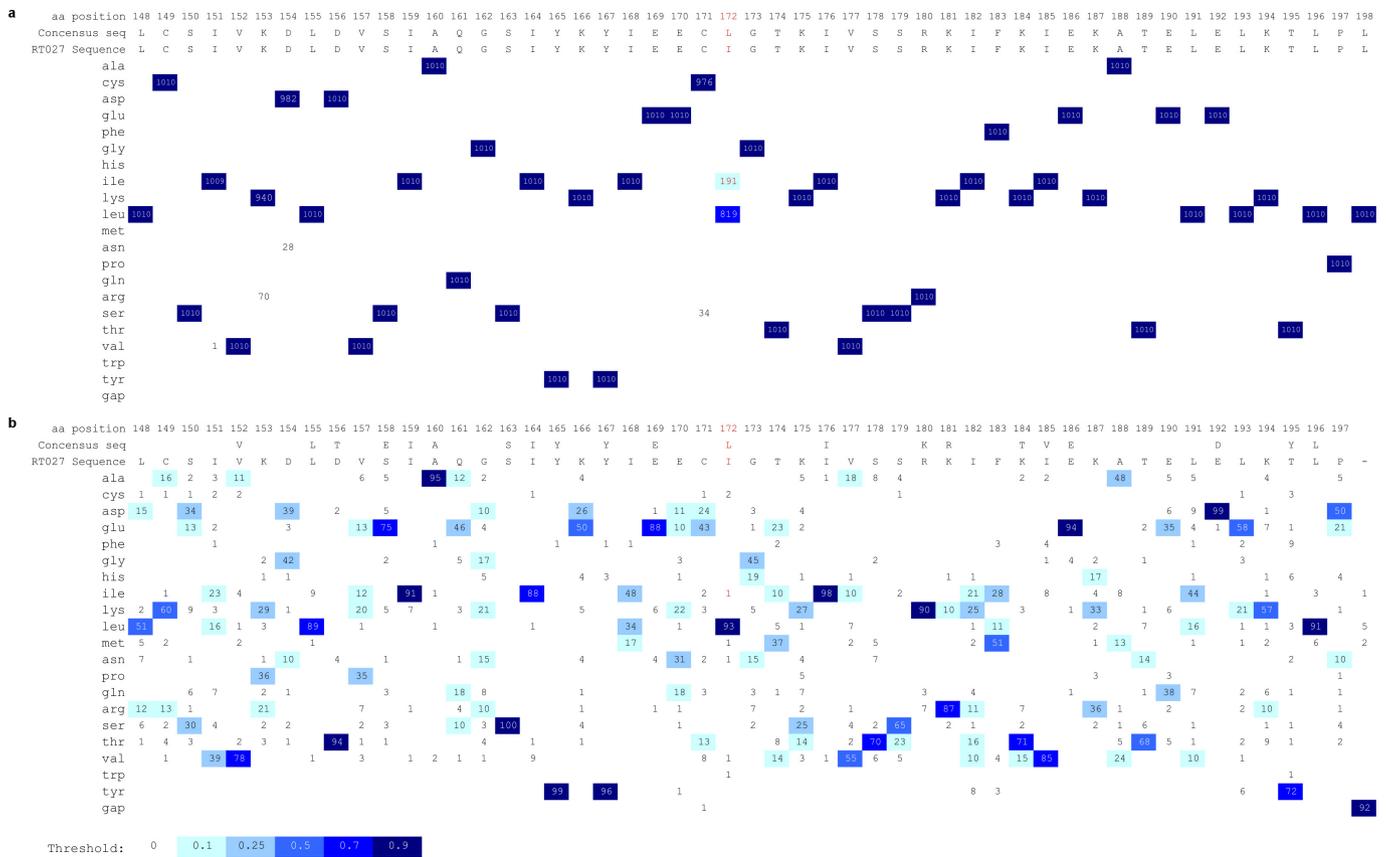


Extended Data Figure 2 | Growth of *C. difficile* strains. **a**, The majority of strains can grow on 50 mM trehalose. Dashed grey line and band indicate mean growth and s.d. in DMM without a carbon source. Solid lines indicate mean growth yield ($A_{600\text{ nm}}$) for groups: non-RT027/078 ($n = 10$), RT027 ($n = 8$), and RT078 ($n = 3$). **b**, Deletion of *treA* ablates the ability of both CD630 (RT12) and R20291 (RT027) to grow on trehalose.

This phenotype can be restored by supplying *treA* on an inducible plasmid ($n = 3$ for each strain/group). **c**, RT244 strains (DL3110 and DL3111) possessing the TreR L172I mutation are capable of growth on 10 mM trehalose ($n = 3$ for each strain/group). **d**, CD1015 Δ *ptsT* can metabolize 50 mM trehalose ($n = 4$ for each strain/group). For **a–d**, points represent biologically independent samples, solid bars are means.

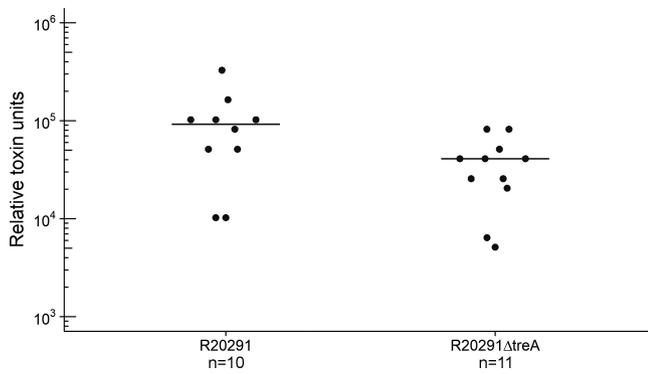


Extended Data Figure 3 | RT027 strains express *treA* at a significantly higher level than non-RT027 strains in the presence of 25 mM trehalose. Each data point ($n = 4$ ribotypes per group) represents gene expression from a different, biologically independent, strain and is an average from two to five independent experiments. $P = 0.029$, Mann-Whitney-Wilcoxon test (two-sided). Bar indicates mean expression.

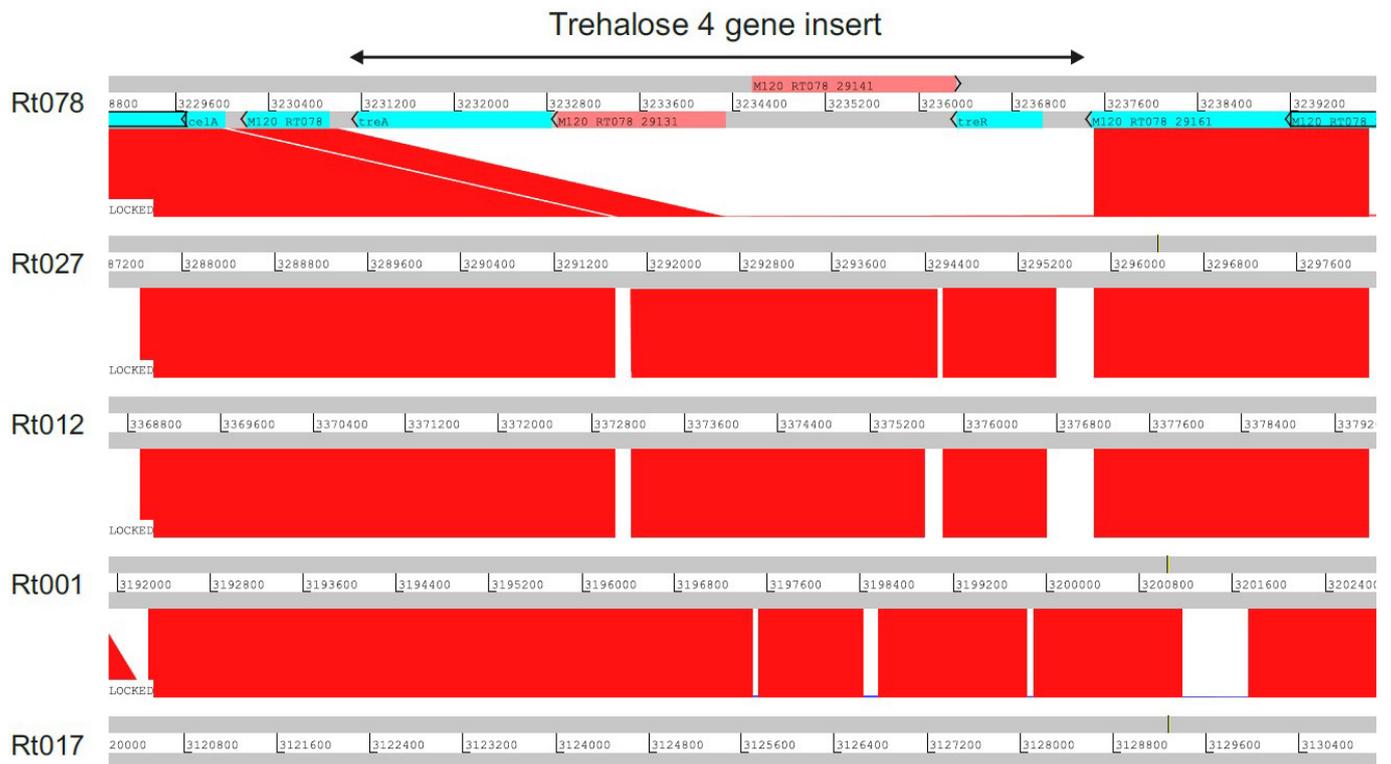


Extended Data Figure 4 | RT027 strains have an L172I mutation at a highly conserved site. a, The *treR* genes from available *C. difficile* whole-genome sequencing files on the NCBI database (accessed 11 May 2017) were identified by tblastn and translated to protein sequences. Sequence fragments shorter than 240 amino acids were discarded and the remaining 1,010 sequences aligned with Clustal Omega³³. All 191 sequences containing the L172I SNP also contained the *thyA* gene, a marker for the RT027 lineage; *thyA* was not found in any other genomes. Numbers indicate the number of sequences with a corresponding amino acid in that position. Multiple sequence alignment visualization generated with

ProfileGrid³⁴. **b**, The TreR protein sequence from RT027 strain R20291 was blasted against non-*C. difficile* sequences in the NCBI database and the top 99 matches (along with R20291_{treR}) aligned with Clustal Omega. The leucine at position 172 was found to be conserved in 93 of 99 non-*C. difficile* sequences. To confirm the importance of this residue, TreR was blasted against all non-clostridial sequences in the NCBI database and the top 500 hits saved. After removal of duplicate species, 191 sequences were aligned with Clustal Omega. The leucine at residue 172 was conserved in 83% of sequences (data not shown).



Extended Data Figure 5 | A *treA* knockout strain decreased toxin production 48 h after infection. Mice were gavaged with 10^4 spores of either R20291 or R20291 Δ *treA* and provided with 5 mM trehalose in drinking water. Points represent toxin levels from individual mice (R20291 $n = 10$, R20291 Δ *treA* $n = 11$) euthanized 48 h after infection. Bars are means. Mice gavaged with R20291 Δ *treA* had significantly lower toxin levels ($P = 0.0268$; Wilcoxon-Mann-Whitney test (two-sided), median 40,960, IQR 23,040–46,080 versus 92,160, IQR 51,200–102,400).



Extended Data Figure 6 | The four-gene trehalose insertion is only present in the RT078 lineage. Artemis comparison tool displaying pairwise comparisons between *C. difficile* RT078 genome (M120) sequence and genome sequences from other *C. difficile* ribotypes (ribotypes indicated on the left). Numbers between grey bars indicate the genomic

region where the trehalose four-gene insert is located (3231169–3237057). Regions of sequence homology are displayed in red. The trehalose four-gene insert of RT078 (indicated by the arrow on the top) was observed in RT078, but was absent in other ribotypes.

Extended Data Table 1 | Compounds conferring at least 1.5-fold growth advantage in Biolog Phenotypic Microarray plates PM1 or PM2

	Compound	CD630	CD2015	
PM1:	N-acetyl-D-glucosamine	+	+	
	L-proline	-	-	
	D-trehalose	+	+	
	D-mannose	+	+	
	D-sorbitol	-	+	
	D-mannitol	+	+	
	D-fructose	-	+	
	α -D-glucose	+	+	
	α -Keto-Butyric acid	+	-	
	L-serine	+	-	
	L-threonine	-	-	
	glycyl-L-proline	-	-	
	PM2:	N-acetyl-neuraminic acid	+	+
		D-arabitol	-	+
arbutin		+	+	
D-melezilose		+	+	
salicin		+	+	
D-tagatose		+	+	
D-glucosamine		+	+	
β -hydroxy-butyric acid		+	+	
α -keto valeric acid		+	+	
hydroxy-L-proline		-	+	
L-leucine	+	+		
L-methionine	-	+		

Results of individual experiments where growth was (+) or was not (-) increased by at least 1.5-fold over DMM control.

Extended Data Table 2 | Spontaneous *C. difficile* mutants able to utilize 10 mM trehalose

Strain	Ribotype	Nonsense mutation*	Missense mutation	Insertions/ deletions	Number of independent isolates
3014	001	18	-	-	1
2012	002	63	-	-	1
2012	002	89	-	-	2
2012	002	15	-	-	1
2012	002	22	-	-	1
2012	002	-	S20I	-	1
2012	002	64	-	-	1
2012	002	24	-	-	1
2012	002	20	-	-	1
1014	014	-	S41I, T118K	1	1
1014	014	-	T118K	-	1
2048	053	70	-	-	1

*Numbers refer to the positions in the consensus TreR amino acid sequence that become a premature stop codon.

Extended Data Table 3 | Strains, primers and plasmids

Strains	Ribotype	MLST (Clade)	Note	Reference
CD630	12	54 (1)	Erythromycin sensitive	35
CD630 Δ <i>treA</i>	12	54 (1)		This Study
CD630 pRFP185-P _{atC} - <i>ptsT</i>	12	54 (1)		This Study
CD630 Δ <i>treA</i> pRFP185-P _{atC} - <i>treA</i>	12	54 (1)		This Study
R20291	27	1 (2)		24
R20291 Δ <i>pyrE</i>	27	1 (2)		24
R20291 Δ <i>treA</i>	27	1 (2)		This Study
R20291 Δ <i>treA</i> pRFP185-P _{atC} - <i>treA</i>	27	1 (2)		This Study
CD1015	78	11 (5)		†
CD1015 Δ <i>pyrE</i>	78	11 (5)		This Study
CD1015 Δ <i>ptsT</i>	78	11 (5)		This Study
CD1015 Δ <i>ptsT</i> pRFP185-P _{atC} - <i>ptsT</i>	78	11 (5)		This Study
VPI10463	3	12 (1)	High toxin producer	
CD196	27	1 (2)	Ancestral RT027 strain	36
CD1007	053-163	63 (1)		†
CD1014	014-20	2 (1)		10
CD2012	2	8 (1)		†
CD2018	unique UM isolate			†
CD2046	unique UM isolate			†
CD2048	053-163	63 (1)		10
CD37	9	3 (1)	Non-toxicogenic strain	37
CD4004	2	8 (1)		†
CD4011	1	3 (1)		†
CD2015	27	1 (2)		10
CD3017	27	1 (2)		10
CD4010	27	1 (2)		10
CD4012	27	1 (2)		†
CD4015	27	1 (2)		10
CD2001	78	11 (5)		†
CD2058	78	11 (5)		†
DL3110	244	41 (2)	contains L172I in <i>treR</i>	11
DL3111	244	41 (2)	contains L172I in <i>treR</i>	11

Primers	Note	Sequence 5' - 3'
IBS1.2		atatcaagcttttgaaccacgctgacgtggaatagaagattatgtgcccagatagggtg
EBS1	Insertional deletion of <i>treA</i> in CD630	cagattgtacaaatggtgataacagataagtcattatataactaccttcttctgt
EBS2		cgcaagtttctaatttcggtttctatc.gatagaggaaggtgtct
CR064	CD630 <i>treA</i> insertion check	gcaacaatgatggtataggtgataaaatgg
CR065		ggaacagaaccatcaggttagca
JCb092	Upstream homology arm R20291 <i>treA</i>	agaccttaaggaggatagggt
JCb093		aggagaaacgtacataggatcaacca
JCb094	Downstream homology arm R20291 <i>treA</i>	cctgaaacttattgaataaataaactacac
JCb095		gtttgatactgatgaggcccta
JCb096		gccctccatcagatcaaacggggatcctctagatcgac
JCb097	Bridging oligos for ligase cycling	ctcctaagtacgtttctcctcctgaaactattgaaata
JCb098		cgaaatc.gagctc.ggtacc.cagaccttaaggaggatag
JCb135	Upstream homology arm CD1015 <i>pyrE</i>	atacctgcagagggaactttttatctcag
JCb136		acaacatctcagcaattatctttg
JCb137	Spacer from pMTL-YN2	gcggccgctgatccatgatgacc
JCb138		actagccatctgccattcagg
JCb139	Downstream homology arm CD1015 <i>pyrE</i>	gcggccgctgatccatgatgacc
JCb140		atatgcccgcatacaataataaaatataaataat
JCb141	Bridging oligos for ligase cycling	ataatgtcgaagatgtgtgcccctgctcatatg
JCb142		gaatggcgaatggcgtgtaataaaactaattatt
JCb153	Upstream homology arm CD1015 <i>ptsT</i>	atagaattc.aaggacc.caggaaattgacc
JCb154		ataggatccatctactctcttctctttttataaag
JCb155	Downstream homology arm CD1015 <i>ptsT</i>	ataggatcccaaatgacaataataataaattccctgg
JCb156		atacctatc.cctgctggtgctatggttaca
JCb167	CD1015 Δ <i>ptsT</i> check	cggaatttctttatctcatttgg
JCb168		cccaattgttgagcactt
JCb225	Conformation of <i>pyrE</i> knockout/repair	atgggaatggcggaataac
JCb226		gcttggagcagctacaacaga
JCb211	CD1015 <i>pyrE</i> repair	atagccggcccttactcctaattcctgaactctc
JP048	qPCR for <i>C. difficile</i> 16S DNA	ttgagcgattactcctgtaaga
JP049		ccatcctgactgctcacct
CR045	qPCR for <i>C. difficile</i> <i>treA</i> DNA	tacgctgatgctcctgctat
CR046		cgctcctttataatctgctttc

Plasmids	Reference
pMTL-YN2	24
pMTL-YN2C	24
pMTL-YN4	24
pRFP185	25
Plasmid Maps	
pJC-R20291 <i>treA</i> KO	https://benchling.com/s/seq-DANCiRRu7fWnQrJCO9iu
pJC-CD1015 <i>pyrE</i> KO	https://benchling.com/s/seq-Km3QbgAKU66DkNtaOA4R
pRFP185-P _{atC} - <i>ptsT</i>	https://benchling.com/s/seq-hKTbVE2pFE8MpwVEqFce
pJC-CD1015 <i>pyrE</i> Repair	https://benchling.com/s/seq-d25nmFmvSPtR1iQB8vka
pRFP185-P _{atC} - <i>treA</i>	https://benchling.com/s/seq-TTthHzlU2fsfRZ94oD9V

†Clinical isolates obtained from the Michigan Department of Community Health. Collected from Michigan hospitals between December 2007 and May 2008. (References 10, 11, 24, 25, 35–37 are cited in the table.)

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

▶ Experimental design

1. Sample size

Describe how sample size was determined.

Sample size for animal experiments were determined using power analysis based upon prior experimental data. Sample sizes for bacterial growth and gene expression assays were based upon prior empirical work in the lab, however, no explicit power analysis was carried out.

2. Data exclusions

Describe any data exclusions.

No data were excluded from analysis

3. Replication

Describe whether the experimental findings were reliably reproduced.

All experiments were repeated on multiple days and used multiple batches of media to control for batch variation. All attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

No specific randomization of animals was carried out to avoid splitting mouse cages, however, all groups were checked to ensure no significant difference in the age, weight or sex of mice between groups prior to starting experiments.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators were not blinded to group allocation during data collection or analysis except for pathology scoring of mouse tissue samples.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

- | | |
|--------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The <u>exact sample size</u> (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement indicating how many times each experiment was replicated |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Clearly defined error bars |

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

R version 3.3.2 was used for all data analysis. Clustal Omega v 1.2.2 was used for protein alignments.
 JProfileGrid v 2.0.5 was used for protein alignment visualization.
 BLASTX was used to compare RT078 genome with the *C. difficile* protein database. This protein database, containing gene identifier information for each ribotype together with corresponding protein sequences, was created by using the Choose Search Set functionality of BLASTX. Phylogenetic trees were reconstructed using MEGA 7.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No restrictions.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

African Green Monkey kidney (Vero) cells.

b. Describe the method of cell line authentication used.

C. difficile Toxin assay used as verification.

c. Report whether the cell lines were tested for mycoplasma contamination.

No test for mycoplasma was performed

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No commonly misidentified cell lines were used.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

C57 b6 mice with a humanized microbiota were used in this study, described in detail in "Humanized microbiota mice as a model of recurrent *Clostridium difficile* disease" Collins et al 2015. Mice of both sexes were used, aged 6-8 weeks.

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Ileostomy effluent samples were collected anonymously from members of the Ostomy Association of Greater Lansing according to a protocol approved by the Institutional Review Board of Michigan State University (Protocol 10-736SM). For faecal samples, live subjects who were self-described as healthy and had not consumed antibiotics within the previous two months were recruited to provide faecal samples for human faecal bioreactor experiments. Informed consent was obtained prior to collection of samples and no identifying information was obtained along with the sample. Subjects who had ileostomies placed due to a previous, undisclosed illnesses were recruited to provide ileostomy effluents. Informed consent was obtained prior to collection of samples and no identifying information was obtained along with the sample. Ileostomy donors were recruited through the Ostomy Association of Greater Lansing, and were most likely residents of Lansing, Michigan USA and its surrounding counties