Virulence and genomic diversity among clinical isolates of ST1 (BI/NAP1/027) *Clostridioides difficile*

Qiwen Dong[a,b], Huaiying Lin[b], Marie-Maude Allen[c], Julian R. Garneau[c] Jonathan K. Sia[d], Rita C. Smith[b], Fidel Haro[b], Tracy McMillen[e], Rosemary L. Pope[b,f], Carolyn Metcalfe[b], Victoria Burgo[b], Che Woodson[b], Nicholas Dylla[b], Claire Kohout[b], Anitha Sundararajan[b], Evan S Snitkin[g], Vincent B. Young[g,h], Louis-Charles Fortier[c], Mini Kamboj[e], Eric G. Pamer[a,b,f#]

[a]Department of Medicine, University of Chicago, Chicago, Illinois, USA

[b]Duchossois Family Institute, University of Chicago, Chicago, Illinois, USA

[c]Department of Microbiology and Infectious Diseases, Université de Sherbrooke, Sherbrooke, Quebec, Canada

[d]Immunology Program, Memorial Sloan Kettering Cancer Center, New York City, New York, USA

[e]Infection Control, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA

[f]Committee on Immunology, University of Chicago, Chicago, Illinois, USA

[g]Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA

[h]Department of Microbiology & Immunology, University of Michigan, Ann Arbor, MI, USA

Running Head: cdtR mutation attenuates ST1 *C. difficile* virulence

#Address correspondence to Eric G. Pamer, egpamer@uchicago.edu

Abstract

Clostridioides difficile (C. difficile), a leading cause of nosocomial infection, produces toxins that damage the colonic epithelium and results in colitis that varies from mild to fulminant. Variation in disease severity is poorly understood and has been attributed to host factors (age, immune competence and intestinal microbiome composition) and/or virulence differences between *C. difficile* strains, with some, such as the epidemic BI/NAP1/027 (MLST1) strain, being associated with greater virulence. We tested 23 MLST1(ST1) *C. difficile* clinical isolates for virulence in antibiotic-treated C57BL/6 mice. All isolates encoded a complete Tcd pathogenicity locus and achieved similar colonization densities in mice. Disease severity varied, however, with 5 isolates causing lethal infections, 16 isolates causing a range of moderate infections and 2 isolates resulting in no detectable disease. The avirulent ST1 isolates did not cause disease in highly susceptible Myd88$^{-/-}$ or germ-free mice. Genomic analysis of the avirulent isolates revealed a 69 base-pair deletion in the N-terminus of the *cdtR* gene, which encodes a response regulator for binary toxin (CDT) expression. Genetic deletion of the 69 base-pair *cdtR* sequence in the highly virulent ST1 R20291 *C. difficile* strain rendered it avirulent and reduced toxin gene transcription in cecal contents. Our study demonstrates that a natural deletion within *cdtR* attenuates virulence in the epidemic ST1 *C. difficile* strain without reducing colonization and persistence in the gut. Distinguishing strains on the basis of *cdtR* may enhance the specificity of diagnostic tests for *C. difficile* colitis.

Keywords

C. difficile, ST1(RT027), clinical isolates, virulence, prophage, toxins

Introduction

Clostridioides difficile is a Gram-positive, spore-forming anaerobic bacterium, and the leading cause of nosocomial infections in the United States.[1–3] Infections are acquired by oral ingestion of *C. difficile* spores, which are prevalent in the environment and can survive for extended periods of time on contaminated surfaces. Upon ingestion, *C. difficile* spores germinate, produce toxins and cause colitis and, in severe cases, can result in mortality. The major virulence factors of *C. difficile* are toxins A (*tcdA*) and B (*tcdB*), which are encoded in the Pathogenicity Locus (PaLoc).[4,5] These toxins glycosylate and thereby inactivate host GTPases, triggering the death of intestinal epithelial cells and lead to gut inflammation.[6]

C. difficile species is comprised of hundreds of strain types across more than 6 phylogenetic clades. PCR- and sequencing-based approaches, including PCR-ribotyping (RT) and multilocus sequencing typing (MLST/ST), have been used to characterize *C. difficile* strain types. More recently, whole-genome sequencing (WGS) has greatly contributed to our understanding of *C. difficile* diversity, evolution, and epidemiology.[7] Almost two decades ago, the BI/NAP1/027 strain, characterized as ST1 by MLST, emerged as a cause of severe nosocomial outbreaks and increased *C. difficile* infection (CDI) incidence in North America and Europe. Since then, the prevalence of ST1 has declined but it remains among the most frequently isolated strains in hospital and community-acquired CDI cases in the US.[2,8–11] The ST1 *C. difficile* strain encodes the additional CDT toxin (encoded by *cdtA* and *cdtB* and also referred to as binary toxin), which is an ADP-ribosyltransferase that modifies actin and disrupts cellular cytoskeleton organization.[12] The ST1 strains have

higher MICs to several antibiotics, most notably fluroquinolones, and produce higher amounts of toxins A and B compared to non-ST1 strains.[13,14] The relative virulence of the ST1 strain is controversial, however, with some studies demonstrating clinical disease severities similar to other strains.[15–17] Host factors can impact the severity of CDI, including underlying diseases, immune competence and microbiome composition.[18,19] Whether genetic variants of ST1 explain diverse disease manifestations is unknown.

To determine intra-strain type virulence diversity, we used an antibiotic-treated mouse model of *C. difficile* infection to test a panel of PaLoc- and CdtLoc-encoding ST1 *C. difficile* clinical isolates to quantify disease severity.[20] Clinical *C. difficile* isolates with identical PaLoc caused a range of disease severities, with two isolates causing no detectable disease in antibiotic-treated wild type, germ-free mice or Myd88-defecient mice. We identified a 69-bp deletion in the *cdtR* gene of these two avirulent isolates, which encodes a LytTR family response regulator that regulates CDT expression. The 69-bp deletion in the *cdtR* leads to reduced CDT toxin and PaLoc gene expression, resulting in loss of virulence and confirming previous studies implicating CdtR as regulator of CDT and Tcd toxins expression.[21] Our study is the first to describe virulence diversity within a single strain type and demonstrates the critical role of CdtR for ST1 *C. difficile* virulence.

Results

Clinical ST1 *C. difficile* isolates demonstrate variable severities in mice.

We focus on a group of 23 *C. difficile* isolates belonging to ribotype 027 epidemic strains (here referred to as ST1) isolated from patients with diarrhea during a molecular

surveillance program at Memorial Sloan Kettering Cancer Center 2013-2017.[22] Whole-genome Illumina sequencing of these isolates allows us to compare them to the public collections. We plotted a UMAP analysis of the presence or absence of unique coding sequences (annotated proteins or un-annotated protein clusters) across top 10 STs of *C. difficile* strains in Patric (date: Feb. 10 2021).[23] Different STs cluster individually and our ST1 isolates overlap with other ST1 *C. difficile* included in the analysis, confirming their strain type **(Figure 1A)**. These 23 isolates demonstrate high genome-wide similarity by average nucleotide identity (ANI) score above 99.8% and encode identical Pathogenicity Locus (PaLoc) sequences **(Figure S1A-S1B)**. To study if close-related *C. difficile* isolates may have variable virulence, mice treated with antibiotics (metronidazole, neomycin, vancomycin in drinking water with clindamycin intraperitoneal injection) were orally infected with each of these isolates at a dose of 200 spores and *C. difficile* pathogenicity was monitored throughout a 7-day-timecourse **(Figure 1B)**. Mice infected with different ST1 isolates displayed a spectrum of disease severity, including variable weight loss and mortality **(Figure 1C and Figure S1C)**. The widely used ST1 lab strain R20291 was included in parallel for virulence comparison. Within our ST1 collection, 5 isolates resulted in mortality in mice. A few isolates caused more severe weight loss than R20291, including ST1-49, ST1-11 and ST1-12, while most ST1 isolates caused moderate and non-lethal infections. Two isolates, ST1-75 and ST1-35 demonstrated no impact on mouse body weights. No apparent colonization deficiency was observed in any of these isolates **(Figure S1D)**. The variable pathogenicity induced by a group of ST1 isolates with identical PaLoc suggested additional regulatory mechanisms of *C. difficile* virulence. Therefore, we sought

to examine other genomic factors that are responsible for attenuated virulence of *C. difficile* isolates ST1-75 and ST1-35.

Two ST1 *C. difficile* isolates demonstrate avirulent phenotype

Among the *C. difficile* isolates that we examined using antibiotic-treated mice, two isolates, ST1-75 and ST1-35, caught our attention due to their strikingly attenuated phenotypes **(Figure 1B)**. Almost no weight loss was observed throughout the 7-day-timecourse, and low acute disease scores were displayed in mice infected with ST1-75 or ST1-35, in comparison to mice infected with R20291 **(Figure 2A-2B and 2D-2E)**. This avirulent phenotype was not due to colonization deficiency of ST1-75 or ST1-35, as the fecal CFU recovered from the mice infected with these two isolates were comparable to R20291-infected mice on both early and late days post-infection **(Figure 2C and 2F)**. Fecal levels of TcdA and TcdB were also measured, and similar levels were seen in the feces at day +1 post-infection from mice infected with ST1-75 and ST1-35, compared to R20291**(Figure 2G-2H)**.

To further investigate this avirulent phenotype, we inoculated ST1-75 into MyD88$^{-/-}$ mice, which lack the adaptor protein for Toll-like receptor signaling.[24] MyD88$^{-/-}$ mice fail to recruit neutrophils to the colonic tissue during early stages of *C. difficile* infection, and display markedly increased susceptibility to *C. difficile* induced colitis.[25] Here, MyD88$^{-/-}$ mice were treated with antibiotics and infected with either ST1-75 or R20291. Mice infected with R20291 quickly succumbed to infection 2 days after spore inoculation, whereas all MyD88$^{-/-}$ mice infected with ST1-75 survived the experiment with minimal

weight loss or disease scores **(Figure 3A-3B)**. Consistent with our results with wildtype mice, no deficiencies of colonization or toxin production were observed day+1 post infection of MyD88$^{-/-}$ mice **(Figure 3C-3D)**. These data suggest that the attenuation of the avirulent strain is independent of MyD88-mediated host innate immunity.

Germ-free mice are highly susceptible to *C. difficile* infection because they microbiome-mediated colonization resistance against *C. difficile*.[26,27] To investigate whether the gut microbiome renders ST1-75 avirulent, germ-free mice were infected with ST1-75 or R20291. Similarly, we observed no mortality or weight loss in the mice with ST1-75 infection, whereas mice infected with R20291 quickly lost weight and died (or >20% weight loss) **(Figure 3E)**. Milder diarrhea was observed in mice with ST1-75 compared to R20291 **(Figure 3F).** We observed no differences in colonization or fecal toxins between ST1-75 and R20291 up to 24 hours post infection **(Figure 3G-3H)**. Similar attenuation was also seen in ST1-35 infected germ-free mice **(Figure S2)**. In contrast, isolates that demonstrated relatively mild pathogenicity in antibiotic-treated mice, such as ST1-25 and ST1-67 **(Figure 1B)**, led to severe weight loss and diarrhea in germ-free mice **(Figure S2),** reaffirming the protective role of the gut microbiome during *C. difficile* infection. However, the attenuation of ST1-75 and ST1-35 in mice is independent of the gut microbiome.

Novel prophages identified in avirulent strains do not impact ST1 *C. difficile* virulence.

We next sought to determine the genetic factors that may abrogate *C. difficile* virulence in ST1-75 and ST1-35. Fully circularized genomes of 14 ST1 isolates were successfully obtained using Nanopore and Illumina hybrid assembly, and pangenomic analysis was

161 conducted on these 14 genomes and R20291 using Anvi'o pangenomics workflow.[28] A
162 group of gene clusters that is unique to ST1-75 and ST1-35 stood out, which are enriched
163 for phage-related genes **(Figure 4A)**. We then applied PHASTER, a tool for phage
164 identification in bacterial genomes, to discover two unique prophages in the genomes of
165 ST1-75 and ST1-35. One prophage resides on a 41-kb plasmid in ST1-75 and ST1-35 with
166 4-5 copies per cell and here is named as phiCD75-2 **(Figure S3A)**. Blasting phiCD75-2
167 found high similarities to reported *C. difficile* phages phiCD38-2 (99.8% identity) and
168 phiCDHS1 (94.7% identity).[29–32] In addition, a unique 54-kb segment was found as inserted
169 into the chromosomal DNA of ST1-75 and ST1-35 around position 1.29 Mbp and here is
170 named as phiCD75-3 **(Figure S3A)**. PhiCD75-3 does not show high similarity to any
171 described *C. difficile* phages to date.
172
173 Lysogenic bacteriophages have been identified in many *C. difficile* genomes, and play an
174 important role in shaping *C. difficile* evolution. However, their roles in *C. difficile* biology,
175 especially virulence, are not well-characterized.[33,34] To investigate the potential role of
176 these two prophages on *C. difficile* virulence, we induced lytic phage particles of phiCD75-
177 2 and phiCD75-3 from ST1-75 culture and infected R20291 to generate R20291 lysogens
178 harboring these prophages. We were able to generate R20291 derivatives carrying
179 phiCD75-2, phiCD75-3 or both prophages in their genomes **(Figure 4B)**. Whole-genome
180 sequencing of lysogenic R20291 strains confirmed that phiCD75-3 was inserted *in situ* as
181 in ST1-75 at 1.29 Mbp. Antibiotic-treated mice infected with R20291 lysogenic strains
182 **(Figure 4B)** followed the curve of virulent infection as 10% weight loss and 4-6 disease
183 scores were seen during the peak of symptomatic infection **(Figure 4C-4D).** The seemingly

faster recovery in the lysogens was not a reproducible finding. Similar levels of colonization and toxin production were also observed **(Figure 4E and S3B-S3D)**. Here, we discover two prophages in avirulent strains ST1-75 and ST1-35, that are not present in R20291 or other ST1 strains from our collection of. However, these two prophages do not appear to impact the virulence of R20291 in antibiotic-treated mice.

Mutations in the *cdtR* gene eliminate ST1 *C. difficile* virulence in mice.

Lysogenic R20291 strains with either or both prophages did not recapitulate the avirulent phenotype of ST1-75 or ST1-35. A closer look at the chromosomal genomes of ST1-75 and ST1-35 led us discover a common mutation in their *cdtR* gene, which was reported previously as a transcriptional regulator for binary toxin (CDT) genes, *cdtA* and *cdtB*.[35] A unique 69-bp deletion was found in the *cdtR* gene of ST1-75 and ST1-35, leading to an in-frame deletion of 23 amino acids **(Figure 5A)**. To investigate if there is a possible loss of function of CdtR resulted from the deletion, we accessed the transcriptional level of *cdtB* in mouse cecum following infection of ST1-75 or R20291. More than a 2-log reduction of *cdtB* transcripts was observed in ST1-75 group **(Figure 5B)**, suggesting an important role of these 69 base pairs for a fully functional *cdtR* gene. Next, we applied CRISPR-mediated genome editing approach to generate CdtR mutants using parental R20291 strain to study the contribution of CdtR to *C. difficile* virulence **(Figure S4A and 5A)**. In accordance with a previous report[21], knocking out *cdtR* either by deleting the whole gene (CdtRKO8.1 and CdtRKO10.3), or introducing a proximal premature stop codon (CdtRstop4.2 and CdtRstop8) led to a loss of pathogenicity in antibiotic-treated mice **(Figure S4B-S4C)**, confirming a critical role of CdtR for *C. difficile* virulence. Moreover, deleting the exact

same 69-bp region, as in ST1-75/35, in the *cdtR* of R20291 (CdtRmut6.1 and CdtRmut8.1) again eliminated the virulence of *C. difficile* **(Figure 5C-5D)**. Thus, loss of the 69-bp in *cdtR* explains the avirulence phenotype of ST1-75/35. On the other hand, colonization of these CdtR mutants, assessed by CFU at day+1 post infection, was comparable to that of R20291 **(Figure S4D and S4F)**, suggesting that CdtR is not required for colonization. Interestingly, while the fecal levels of TcdA and TcdB of CdtR mutants were comparable to R20291 in the early phase (day+1 post-infection), a significantly reduced level at a later time point (7-days post infection) was observed upon infection of CdtR mutants **(Figure S4E and 5E)**, supporting a role of CdtR in regulating PaLoc toxins production. Further, infecting germ-free mice with CdtRmut6.1 results in no weight loss or diarrhea, perfectly recapitulating ST1-75 and ST1-35 phenotypes in germ-free mice **(Figure 5F-5G and Figure S4G)**. Collectively, CRISPR-edited CdtR mutant strains mimic phenotypes of ST1-75 and ST1-35 in mice, demonstrating that *cdtR* gene is necessary for *in vivo* virulence. Furthermore, the 69-bp region in *cdtR*, which is deleted in ST1-75/35, is necessary for proper CdtR function through mechanisms yet to be determined.

Mutations in *cdtR* reduce PaLoc toxin transcription *in vivo*

CdtR mutants produce significantly reduced fecal toxins at a later time point (7-days post infection) **(Figure 5E and S4E)**, a phenotype that was confirmed in ST1-75 and ST1-35 **(Figure S5A-S5B)**. To examine whether the 69-bp deletion in *cdtR* impacts PaLoc toxin production, we harvested cecal contents from germ-free mice infected with ST1-75 or R20291. In contrast to fecal toxin levels, we observed a significantly reduced cecal toxin level in mice infected with ST1-75 compared to R20291 **(Figure 6A)**. This was not due to

a slightly lower CFU of cecal ST1-75 in germ-free mice **(Figure S5C-S5D)**. We further validated the reduced toxin production in cecal content by RT-qPCR and we observed a 50-fold reduction of the *tcdA* and *tcdB* transcripts in the cecum of mice infected with ST1-75 **(Figure 6B)**. Additionally, transcriptions of other PaLoc genes including *tcdE* and *tcdR* were also reduced in the cecum of mice infected with ST1-75 **(Figure 6B)**. *TcdE* is a putative holin that mediates toxin secretion.[36–38] Reduced *tcdE* likely further impacts the amount of toxins that may reach to the intestinal epithelium. TcdR is a positive regulator of the PaLoc[39,40] and is likely the common target of CdtR, which results in the observed downregulation of many PaLoc genes. Interestingly, *cdtR* transcripts were comparable between ST1-75 and R20291, suggesting that the 69-bp deletion does not impact the transcripts stability but leads to a nonfunctioning product. These results were further confirmed with CdtRmut6.1 strain, though to a lesser extent **(Figure S5E)**. Collectively, we demonstrated that a natural mutation found in *cdtR* of two ST1 clinical isolates results in reduced binary toxin production, reduced PaLoc toxin production (and likely secretion) in cecum of infected mice, and attenuated *C. difficile* virulence, independent of host innate immunity, colonization burden, microbiome constitution, or any noticeable impact of incidentally discovered prophages within these strains. This difference of toxin production in cecum at 24-hour post infection is however, not reflected in feces in parallel, but could be reflected at a later time point, likely due to cumulative differences over time.

cdtR* is versatile and more prevalent than *cdtA* and *cdtB

Our data support a regulatory role of CdtR outside CdtLoc, so we hypothesize that CdtR may evolve to impact virulence beyond regulating CDT binary toxins. To test this

253 possibility, we surveyed the presence of *cdtR*, *cdtA* and *cdtB* in two major *C. difficile*
254 clinical collections.[23,41] As expected, the majority of clade 2 strains, including the epidemic
255 ST1/RT027 strains, contains the CdtLoc with the presence of all three genes. Other
256 subgroups of *C. difficile* strains, including MLST5 and MLST11 were also reported to
257 encode CDT **(Figure 7A)**.[42,43] Unexpectedly, many strain types of *C. difficile* that were
258 reported as CDT-negative also encode *cdtR*, such as MLST2, MLST8 from clade 1**(Figure**
259 **7A)**. The higher prevalence of *cdtR* over *cdtA* and *cdtB* supports the possibility that CdtR
260 functions beyond regulating CDT. Additional work is needed to evaluate the functions of
261 CdtR in these CDT-negative strains. On the contrary, the presence of *cdtR* in CDT-positive
262 strains may lose its function, such as in MLST11 strains, where a premature stop codon
263 was found and results in a *cdtR* pseudogene[42], as well as here in the case of ST1-75/35. To
264 evaluate the prevalence of *cdtR* mutations that may lead to a loss of function, we aligned
265 all *cdtR* genes in MLST1 strains from the two described collections and found a few strains
266 having similar truncations at the proximal end and may have lost CdtR function, yet we
267 did not find the exact same deletion as in ST1-75/35 within almost 500 strains **(Figure**
268 **S6A)**.
269
270 The high similarity between ST1-75 and ST-35 genetically and phenotypically led us
271 reason whether they are clonal. We performed a core-genome SNP analysis across all ST1
272 isolates from our collection with R20291 as the reference. ST1-75 and ST1-35 shared all
273 SNPs when compared to the genome of R20291**(Figure S6B).** Additional clinical evidence
274 supports ST1-35 and ST1-75 being clonal strains isolated from two patients that shared the
275 same hospital room a few days apart **(Figure S6C)**. Our survey on CdtLoc genes suggest

that this locus is very versatile during evolution and *cdtR* is more prevalent than *cdtA* and *cdtB*, which may regulate virulence beyond CDT.

Discussion

Mouse models are valuable tools to study how *C. difficile* strain variations may result in variable disease severities, thanks to the advantages of their identical genetic, immune background and controlled microbiome compositions. Here we focused on a group of clinical *C. difficile* isolates belonging to the RT027/MLST1, with high genomic similarity, that all encode PaLoc and CdtLoc. We found that these similar *C. difficile* isolates caused variable disease severities in mice and that a very specific mutation in the *cdtR* gene rendered two clinical isolates, ST1-75 and ST1-35, avirulent. Avirulence was solely dependent on the *cdtR* mutation, as we obtained similar observations using MyD88$^{-/-}$ mice, and germ-free mice, which was also further validated with CRISPR-edited *cdtR* mutants. Lower transcripts of binary toxin gene *cdtB*, toxins A *tcdA*, toxin B *tcdB*, together with other PaLoc genes including regulator *tcdR* and putative holin *tcdE*, were observed in mouse cecum infected with the CdtR mutants. Our data support a critical role of CdtR in regulating *C. difficile* toxin production and secretion, which is essential to ST1 virulence. However, all the other ST1 isolates in this study encoded an intact CdtLoc with wildtype *cdtR*, whose variations in virulence are likely attributable to alternative mechanisms.

The presence of a binary toxin locus has been associated with epidemic strains and hypervirulence of *C. difficile*.[44,45] CDT belongs to the family of ADP-ribosylating toxins that consist of two components: CDTa (*cdtA*), the enzymatic active ADP-ribosyltransferase

299 which modifies cellular actin, and CDTb (*cdtB*), the binding component facilitates CDTa
300 translocation. However, despite knowing their enzymatic activities, experimental evidence
301 is very limited to support critical roles of CDT in *C. difficile* virulence.[46] CDTb was
302 reported to induce TLR2-dependent pathogenic inflammation, which suppresses a
303 protective eosinophilic response and enhances virulence of RT027 strains, however, *C.*
304 *difficile* lacking CDTb still causes acute disease in mice.[47] On the other hand, CdtR, as the
305 transcriptional regulator of *cdtA* and *cdtB*[6,35,48], has been previously linked to PaLoc toxin
306 production[21], suggesting a role as a major virulence regulator. Here, we demonstrated a
307 critical role of CdtR as a determinant of *C. difficile* virulence within ST1 strains. A natural
308 69-bp deletion in *cdtR* that was found in two clinical isolates can reverse the virulence of
309 a wildtype strain, by downregulating the expression of PaLoc genes and binary toxin genes.
310 Additionally, higher prevalence of *cdtR* over *cdtA* or *cdtB* was found while surveying
311 CdtLoc on clinical isolates from public databases. This suggests CdtR may have evolved
312 to function beyond regulating *cdtA* and *cdtB*. Systematically examining the target genes of
313 CdtR may give us insights on its additional functions, which may also help unveil the
314 mechanisms by which CdtR regulates the PaLoc genes.
315
316 ST1-75 and ST1-35 are avirulent in susceptible mouse models despite producing toxins,
317 albeit at reduced levels. This is intriguing because it is well appreciated that toxin
318 expression is necessary for *C. difficile* virulence.[46,49] However, our data indicate that toxin
319 production is not sufficient for causing CDI. The amount of toxin being produced and
320 released is likely impact the development of disease. The patients from whom we isolated
321 ST1-75 or ST1-35 had an overall mild clinical assessment, whose symptoms may be

attributable to causes other than *C. difficile* infection. Current CDI diagnoses largely depend on the detection of *TcdB* gene or toxin B positivity in feces and may lead to overdiagnosis of CDI. We, together with other reports, suggest the importance of quantifying toxins to evaluate CDI cases.[50–52] Incorporating adjunctive biomarkers, such as IL-1β, better distinguishes CDI from asymptomatic carriage and non-CDI diarrhea.[53] Here, CdtR regulates both toxin production and secretion, and is essential for *C. difficile* virulence in mice, suggesting it may serve as an adjunctive biomarker for CDI diagnosis.

Apart from characterizing CdtR, we also identified two prophages in ST1-75 and ST1-35. Prophages have been identified in many *C. difficile* genomes, and play important roles in shaping *C. difficile* evolution.[33] While prophages are highly prevalent in *C. difficile*, little is known about how prophages impact *C. difficile* biology. A couple of pioneering studies have shown that prophages can affect *C. difficile* gene expression, impacting toxin production.[29,54,55] In this study we identified two prophages in ST1-75/35 and named them phiCD75-2 and phiCD75-3. By making R20291 lysogenic strains harboring either or both prophages, we observed minimal impacts on *C. difficile* virulence by neither of the prophages in antibiotic-treated mice. PhiCD38-2 was shown to increase PaLoc gene expression and toxin production in some RT027 isolates, but not in all of them, suggesting that the genetic background influences the impact of a newly acquired prophage.[29] This may explain why phiCD75-2 (a phiCD38-2 derivative) did not increase toxin production ST1-75. Certain phages also impact phase variation of the cell surface protein, biofilm formation, and carry genes involved in quorum sensing, inferring their roles in bacterial fitness.[30,56,57] It would be very intriguing to investigate how phiCD75-2 and phiCD75-3

may impact *C. difficile* fitness, including gene expression, antibiotic resistance, and interspecies competition.

In summary, we demonstrate that ST1 *C. difficile* clinical isolates with identical PaLoc display variable virulence *in vivo*. Among them, two clonal clinical isolates, ST1-75 and ST1-35, were avirulent in mice, due to a 69-bp deletion mutation in their *cdtR* genes. These data suggest that specific *cdtR* genetic variants within the same strain type may predict disease occurrence and severity. Routine detection of these variants may enhance the specificity of NAATs for CDI diagnosis. Our data also corroborate recent clinical observations that toxin detection is unreliable as the sole criterion to distinguish between *C. difficile* infection and colonization.

Experimental model and subject details

C. difficile clinical isolate collection and classification

Toxigenic *C. difficile* -positive stool specimens were collected at Memorial Sloan Kettering Cancer Center between 2013-2017. *C. difficile* isolates were recovered by plating onto brain heart infusion (BHI) agar plates supplemented with yeast extract, L-cysteine (BHIS), and the antibiotics D-cycloserine and cefoxitin (BHI and yeast extract were from BD Biosciences, and the other components were from Sigma-Aldrich) in an anaerobic chamber (Coylabs). Individual colonies that were able to grow in the presence of these antibiotics and that had the characteristic phenotype of *C. difficile* were selected, isolated, and subjected to whole-genome sequencing and MLST classification.[58]

Mouse husbandry

Wild-type C57BL/6 mice, aged 6 to 8 weeks, were purchased from the Jackson Laboratories. MyD88$^{-/-}$ mice were maintained in augmentin (0.48 g/L and 0.07 mg/L of amoxicillin and clavulanate respectively) in the drinking water in specific-pathogen-free (SPF) facility at the University of Chicago. Germ-free C57Bl/6J mice were bred and maintained in plastic gnotobiotic isolators within the University of Chicago Gnotobiotic Core Facility and fed ad libitum autoclaved standard chow diet (LabDiets 5K67) before transferring to BSL2 room for infection. Mice housed in the BSL2 animal room are fed irradiated feed and provided with acidified water. All mouse experiments were performed in compliance with University of Chicago's institutional guidelines and were approved by its Institutional Animal Care and Use Committee.

Method details

C. difficile spore preparation and numeration

C. difficile sporulation and preparation was processed as described previously[59] with minor modifications. Briefly, single colonies of *C. difficile* isolates were inoculated in deoxygenated BHIS broth and incubated anaerobically for 40-50 days. *C. difficile* cells were harvested by centrifugation and five washes with ice-cold water. The cells were then suspended in 20% (w/v) HistoDenz (Sigma, St. Louis, MO) and layered onto a 50% (w/v) HistoDenz solution before centrifugating at 15,000 × g for 15 minutes to separate spores from vegetative cells. The purified spores pelleted at the bottom were then collected and washed for four times with ice-cold water to remove traces of HistoDenz, and finally resuspended in sterile water. Prepared spores were heated to 60°C for 20 min to kill vegetative cells, diluted and plated on both BHIS agar and BHIS agar containing 0.1%

(w/v) taurocholic acid (BHIS-TA) for numeration. Spore stocks for mouse infection were verified to have less than 1 vegetative cell per 200 spores (as the infection dose).

Virulence assessment of clinical isolates in mice

SPF mice were treated with antibiotic cocktail containing metronidazole, neomycin and vancomycin (MNV) in drinking water (0.25g/L for each antibiotic) for 3 days, 2 days after removing MNV, the mice were received one dose of clindamycin (200 µg/mouse) via intraperitoneal injection. Mice were then the next day infected with 200 *C. difficile* spores via oral gavage. Germ-free mice were infected with 200 *C. difficile* spores via oral gavage without antibiotic treatments.

Following infection, mice were monitored and scored for disease severity by four parameters[60]: weight loss (> 95% of initial weight = 0, 95%–90% initial weight = 1, 90%–80% initial weight = 2, < 80% = 3), surface body temperature (> 95% of initial temp= 0, 95%–90% initial temp = 1, 90%–85% initial temp = 2, < 85% = 3), diarrhea severity (formed pellets = 0, loose pellets = 1, liquid discharge = 2, no pellets/caked to fur = 3), morbidity (score of 1 for each symptoms with max score of 3; ruffled fur, hunched back, lethargy, ocular discharge).

Quantification of fecal colony forming units

Fecal pellets or cecal content from *C. difficile* infected mice were harvested and resuspended in deoxygenated phosphate-buffed saline (PBS), diluted and plated on BHI agar supplemented with yeast extract, taurocholic acid, L-cysteine, D-cycloserine and cefoxitin (CC-BHIS-TA) at 37°C anaerobically for overnight.[61]

Cell-based assay to quantify fecal and cecal toxin

412 The presence of *C. difficile* toxins was determined using a cell-based cytotoxicity assay as
413 previously described with minor modifications.[61] Briefly, Chinese hamster ovary cells
414 (CHO/dhFr-, ATCC#CRL-9096) were incubated in a 96-well plate overnight at 37°C. Ten-
415 fold dilutions of supernatant from resuspended fecal or cecal content were added to
416 CHO/dhFr- cells, incubated overnight at 37°C. Cell rounding and death was scored the
417 next day. The presence of *C. difficile* toxins was confirmed by neutralization by antitoxin
418 antisera (Techlab, Blacksburg, VA). The data are expressed as the log10 reciprocal value
419 of the last dilution where cell rounding was observed.

420 **DNA extraction, RNA extraction and reverse transcription**

421 Fecal DNA was extracted using DNeasy PowerSoil Pro Kit (Qiagen), and RNA was
422 isolated from cecal contents using RNeasy PowerMicrobiome Kit (Qiagen) according to
423 the manufacturer's instructions, respectively. Complementary DNA was generated using
424 the QuantiTect reverse transcriptase kit (Qiagen) according to the manufacturer's
425 instructions.

426 **Quantitative polymerase chain reaction (qPCR)**

427 Quantitative PCR was performed on genomic DNA or complementary DNA using primers
428 (listed in Table 1) with PowerTrack SYBR Green Master Mix (Thermo Fisher). Reactions
429 were run on a QuantStudio 6 pro (Thermo Fisher). Relative abundance was normalized by
430 ΔΔCt.

431 **Generation of *C. difficile cdtR* mutants using CRISPR**

432 CRISPR editing on *C. difficile* strains R20291 was performed as described in.[62] The
433 primers were listed in Table 1[63,65,67,69,71]. Briefly, donor regions for homology were
434 generated by separately amplifying regions ~500 bp upstream and ~500 bp downstream of

435 the target of interest. The resulting regions were cloned into pCE677 between NotI and
436 XhoI sites by Gibson Assembly. Geneious Prime (v11) was used to design sgRNAs
437 targeting each deleted target. sgRNA fragments were then amplified by PCR from pCE677,
438 using an upstream primer that introduces the altered guide and inserted at the MscI and
439 MluI sites of the pCE677-derivative with the appropriate homology region. Regions of
440 plasmids constructed using PCR were verified by Sanger sequencing. Plasmids were then
441 passaged through NEBturbo *E. coli* strain before transformation into *Bacillus subtilis* strain
442 BS49. The CRISPR-Cas9 deletion plasmids which harbor the oriT (Tn916) origin of
443 transfer, were then introduced into *C. difficile* strains by conjugation.[64] *C. difficile* colonies
444 were then screened for proper mutations in the genomes by PCR and correct clones were
445 further validated by whole-genome sequencing.

446 **Whole-genome sequencing and assembly**

447 DNA was extracted using the QIAamp PowerFecal Pro DNA kit (Qiagen). Libraries were
448 prepared using 100 ng of genomic DNA using the QIAseq FX DNA library kit (Qiagen).
449 Briefly, DNA was fragmented enzymatically into smaller fragments and desired insert size
450 was achieved by adjusting fragmentation conditions. Fragmented DNA was end repaired
451 and 'A's' were added to the 3'ends to stage inserts for ligation. During ligation step,
452 Illumina compatible Unique Dual Index (UDI) adapters were added to the inserts and
453 prepared library was PCR amplified. Amplified libraries were cleaned up, and QC was
454 performed using Tapestation 4200 (Agilent Technologies). Libraries were sequenced on
455 an Illumina NextSeq 500 or MiSeq platform to generate 2x150 or 2x250 bp reads
456 respectively. Illumina reads were assembled into contigs using SPAdes[66] and genes were
457 called and annotated using Prokka (v1.14.6).[68]

Samples for Nanopore and Illumina hybrid assemblies were extracted using the NEB Monarch Genomic DNA Purification Kit. DNA was QC'ed using genomic Tapestation 4200. Nanopore libraries were prepared using the Ligation Sequencing Kit (SQK-LSK109), the Native Barcoding Expansions 1-12 (EXP-NBD104) and 13-24 (EXP-NBD114), and the NebNext Companion Module for Oxford Nanopore Technologies (E7180S). The shearing steps and first ethanol wash were eliminated to ensure high concentrations of long fragments. Using R9.4.1 flow cells, libraries were run on a MinION for 72 hours at -180V. The Nanopore and Illumina hybrid assemblies were completed using Unicycler (v0.4.8)[70] either with the untrimmed or trimmed Illumina reads. The assemblies with less number of circularized contigs were used for genome analysis.

Multiple sequence alignment for Pathogenicity loci

Illumina whole-genome of twenty-five *C. difficile* strains including two reference genome: R20291 (accession: FN545816.1) and CD630 (accession: NC_009089.1), along with 23 in-house strains ST1-10, ST1-11, ST1-12, ST1-19, ST1-20, ST1-23, ST1-25, ST1-26, ST1-27, ST1-35, ST1-49, ST1-5, ST1-53, ST1-57, ST1-58, ST1-6, ST1-62, ST1-63, ST1-65, ST1-67, ST1-68, ST1-69, and ST1-75 were included in the pathogenicity locus (PaLoc) analysis. The PaLoc region of R20291 (NCBI accessions NC_013316, 706,660 – 725,022 bp) were extracted as the query to BLAST[72] against a local database of all the above genomes. Hits with at least 85% query coverage and 85% percent identity were extracted and multiple sequence alignment were performed using Geneious Prime 2022.0.1 with default settings to compare their nucleotide differences.

Binary toxin genes prevalence analysis

C. difficile isolates (N=827) from BioProject PRJEB4556 were downloaded from NCBI, and assembled into contigs using SPAdes.[66] A collection of 2143 *C. difficile* genomes from Patric (date: Feb. 10 2021)[23] were also downloaded. MLST was determined on those contigs by mlst.[73] ST type with less than 3 isolates were removed. Binary toxin *cdtA*, *cdtB* and *cdtR* from R20291 (NCBI accessions NC_013316) were used as query to BLAST[72] against the assembled contigs, and hits with at least 85% identity and 85% coverage of the query are considered a valid match.

UMAP (Uniform Manifold Approximation and Projection) analysis

A subset of isolate contigs of 199 ST1, 50 ST2, 50 ST3, 49 ST6, 50 ST8, 50 ST11, 42 ST14, 50 ST15, 50 ST17, 50 ST37 and 50 ST42, totaling 690 isolates were selected from the above Patric collection. They were all sequenced by short read technology, and they are the top 10 abundant ST groups except ST1 in the Patric collection. Genes were called and annotated from their contigs using Prokka (v1.14.6).[68] By combining the 23 isolates from this study, we constructed a matrix of 731 isolates by 8025 annotated genes and hypothetical protein clusters. Specifically, hypothetical protein clusters were formed by clustering hypothetical proteins at 50% identity using cd-hit.[74,75] Any protein sequences that were at least 50% similar fall into an artificially cluster. UMAP analysis was performed on the basis of the presence/absence of the genes/hypothetical protein clusters by setting the n_neighbors parameter to 675.

Core-genomes SNPs analysis on ST1 isolates

SNP analysis was done on the 23 ST1 isolates from our collection against R20291 using snippy (v4.6.0).[76] Then the core SNP was extracted, recombination removed using gubbins (v2.4.1)[77], and a phylogenetic tree was built using fastTree (v2.1.10).[78,79]

MLST1 *cdtR* SNPs analysis

cdtR hits without starting position at the beginning of *C. difficile* contigs were chosen to further examine their nucleotide differences in *cdtR* gene to R20291 and ST1-75. Five such isolates were found either from Patric collection or BioProject PRJEB4556[23,41], and multiple sequences alignment were performed in Geneious Prime 2022.0.1 with default settings.

FastANI

Genomic similarity between clinical isolates were calculated using FastANI (v 1.32)[80] and presented as ANI score.

Pangenomic analysis of ST1 isolates using Anvi'o

Fifteen circularized genomes of ST1 isolates generated by the Nanopore and Illumina hybrid assemblies were used for pangenomic analysis. Default settings were used based on the Anvi'o workflow for microbial pangenomcis with adjustments for minbit as 0.5 and mcl-inflation as 10.[28,81–83] Annotations were performed with NCBI Clusters of Orthologous Genes (COG).[84] Accessory genomes were grouped by gene clusters that are not present in all 15 isolate genomes.

Prophage identification using PHASTER

Three complete prophages were identified in the genome sequence of strain ST1-75 using PHASTER.[85] Based on Blastn analyses, the phiCD75-1 prophage corresponds to phi027, a prophage highly conserved among R027 isolates.[86] The phiCD75-2 prophage is homologous at 99.82% to the well-described phiCD38-2 phage [29], whereas the phiCD75-3 prophage seems to be a new phage with no close homolog in public databases. The detection of ORF and gene annotation on phage genomes were performed with PROKKA

[68], using an E-value threshold of 10E-3 for function assignment. The most recent PHASTER database (last update Dec 22, 2020) was implemented into PROKKA to improve function prediction and overall annotation quality of phage proteins.[85] The genomes were reorganized so that they started with the terminase gene. Genomic maps were generated using Benchling and finalized with Inkscape v1.2.

Prophage induction and phage amplification

To confirm the functionality of the phiCD75-2 and phiCD75-3 prophages, induction was performed in TY (2% yeast extract, 3% tryptose, pH 7.4) using two different strategies described previously. The first method was a treatment with 2.5 µg/mL mitomycin C (Novus Biologicals), and the second one was UV irradiation for 10 sec at a wavelength of 365 nm.[87] The induced cultures were clarified by centrifugation, then filtered on a 0.22 µm filter and the presence of infectious phage particles was confirmed by plating on bacterial lawns of the R20291 epidemic strain using a soft agar overlay method.[87] Six isolated phage plaques obtained with each induction lysate were picked, diffused in 500 µL phage buffer (50mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM $MgSO_4$) and the identity of the induced phages was determined by PCR using primers specific for phiCD75-2 (LCF0312 5'-AGCGGTATCGGCTTGGTTGTAGAT-3' and LCF0313 5'-TGCTAGTTTCCTGTCAAGGTCGCT-3') and for phiCD75-3 (LCF1242 5'-CGACCCACCTAAAGGTATTCA-3' and LCF1243 5'-GTTCTTTAGTCCAGTTCCCATTTC-3'). Mitomycin C treatment led to induction of the phiCD75-3 prophage whereas UV treatment led to induction of the phiCD75-2 prophage. Each phage was then plaque-purified 3 times to obtain pure phage cultures. Amplification to high titers (> 10^8 pfu/mL) was done in TY broth using strain R20291 as the host and

methods described elsewhere.[29] Phage genomic DNA was extracted from each lysate and a restriction profile was established using XbaI and HindIII to further confirm the identity of the phages, as done before.[29]

Creation of new lysogens

New lysogens of strain R20291 carrying either phiCD75-2, phiCD75-3 or both phages were created using a method described before.[29] Briefly, soft agar overlays containing high titers (> 10^8 pfu/mL) of phage phiCD75-2, phiCD75-3, or both phages in equivalent amounts were prepared and bacterial dilutions of the wildtype R20291 strain were spread on top of the lawn. Phage-resistant colonies that grew after overnight incubation were picked, re-streaked 3 times on TY agar, and the presence of the respective prophage was detected by PCR and by confirming the phage-resistant phenotype upon re-infection with the corresponding phage(s).

Quantification and statistical analysis

Results represent means ± SD. Statistical significance was determined by the unpaired t test and one-way ANOVA test. Statistical analyses were performed using Prism GraphPad software v9.3.1 (∗ $p < 0.05$; ∗∗ $p < 0.01$; ∗∗∗ $p < 0.001$; ∗∗∗∗ $p < 0.0001$

Data availability

Whole-genome sequence data were uploaded to National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject accessions PRJNA885086 and PRJNA595724.

Acknowledgments

We thank Dr. Craig D Ellermeier for his generously providing us the CRISPR editing system of *C. difficile*, Dr. Olga Soutourina for providing R20291 strain. We thank Melanie Spedale and the animal facilities of the University of Chicago for their help on mouse experiments. We thank Dr. Nhu Nguyen, Dr. Zhenrun Zhang and the rest of the Pamer lab members for helpful discussions. This work was supported by the National Institutes of Health R01 AI095706. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Q.D. and E.G.P conceived the project. Q.D., H.L. and N.D. analyzed the data. Q.D., J.K.S., R.C.S., M.M.A., J.R.G., F.H., R.L.P., V.B., C.M., C.W., A.S. and C. K. performed experiments. M.K. and T.M. isolated *C. difficile* isolates. V.B.Y. and E.S.S. sequenced clinical isolates. Q.D., H.L., L.C.F. and E.G.P interpreted the results and wrote the manuscript.

Declaration of interests

None.

References

1. Lessa, F.C., Mu, Y., Bamberg, W.M., Beldavs, Z.G., Dumyati, G.K., Dunn, J.R., Farley, M.M., Holzbauer, S.M., Meek, J.I., Phipps, E.C., et al. (2015). Burden of Clostridium difficile Infection in the United States. New England Journal of Medicine *372*, 825–834. 10.1056/NEJMoa1408913.

2. Guh, A.Y., Mu, Y., Winston, L.G., Johnston, H., Olson, D., Farley, M.M., Wilson, L.E., Holzbauer, S.M., Phipps, E.C., Dumyati, G.K., et al. (2020). Trends in U.S. Burden of Clostridioides difficile Infection and Outcomes. New England Journal of Medicine *382*, 1320–1330. 10.1056/NEJMoa1910215.

3. Abt, M.C., McKenney, P.T., and Pamer, E.G. (2016). Clostridium difficile colitis: pathogenesis and host defence. Nat. Rev. Microbiol. *14*, 609–620. 10.1038/nrmicro.2016.108.

4. Rupnik, M., Wilcox, M.H., and Gerding, D.N. (2009). Clostridium difficile infection: new developments in epidemiology and pathogenesis. Nat Rev Microbiol *7*, 526–536. 10.1038/nrmicro2164.

5. Cohen, S.H., Tang, Y.J., and Silva, J. (2000). Analysis of the pathogenicity locus in Clostridium difficile strains. J Infect Dis *181*, 659–663. 10.1086/315248.

6. Kordus, S.L., Thomas, A.K., and Lacy, D.B. (2022). Clostridioides difficile toxins: mechanisms of action and antitoxin therapeutics. Nat Rev Microbiol *20*, 285–298. 10.1038/s41579-021-00660-2.

7. Knight, D.R., Elliott, B., Chang, B.J., Perkins, T.T., and Riley, T.V. (2015). Diversity and Evolution in the Genome of Clostridium difficile. Clin Microbiol Rev *28*, 721–741. 10.1128/CMR.00127-14.

8. Rao, K., Micic, D., Natarajan, M., Winters, S., Kiel, M.J., Walk, S.T., Santhosh, K., Mogle, J.A., Galecki, A.T., LeBar, W., et al. (2015). Clostridium difficile ribotype 027: relationship to age, detectability of toxins A or B in stool with rapid testing, severe infection, and mortality. Clin Infect Dis *61*, 233–241. 10.1093/cid/civ254.

9. See, I., Mu, Y., Cohen, J., Beldavs, Z.G., Winston, L.G., Dumyati, G., Holzbauer, S., Dunn, J., Farley, M.M., Lyons, C., et al. (2014). NAP1 Strain Type Predicts Outcomes From Clostridium difficile Infection. Clinical Infectious Diseases *58*, 1394–1400. 10.1093/cid/ciu125.

10. Giancola, S.E., Williams, R.J., and Gentry, C.A. (2018). Prevalence of the Clostridium difficile BI/NAP1/027 strain across the United States Veterans Health Administration. Clin Microbiol Infect *24*, 877–881. 10.1016/j.cmi.2017.11.011.

11. C. difficile Infections | A.R. & Patient Safety Portal https://arpsp.cdc.gov/profile/eip/cdi.

12. Gerding, D.N., Johnson, S., Rupnik, M., and Aktories, K. (2014). Clostridium difficile binary toxin CDT. Gut Microbes *5*, 15–27. 10.4161/gmic.26854.

13. Gargis, A.S., Karlsson, M., Paulick, A.L., Anderson, K.F., Adamczyk, M., Vlachos, N., Kent, A.G., McAllister, G., McKay, S.L., Halpin, A.L., et al. (2022). Reference Susceptibility Testing and Genomic Surveillance of Clostridioides difficile, United States, 2012–17. Clinical Infectious Diseases, ciac817. 10.1093/cid/ciac817.

14. Warny, M., Pepin, J., Fang, A., Killgore, G., Thompson, A., Brazier, J., Frost, E., and McDonald, L.C. (2005). Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe. Lancet *366*, 1079–1084. 10.1016/S0140-6736(05)67420-X.

15. Cloud, J., Noddin, L., Pressman, A., Hu, M., and Kelly, C. (2009). Clostridium difficile strain NAP-1 is not associated with severe disease in a nonepidemic setting. Clin Gastroenterol Hepatol *7*, 868-873.e2. 10.1016/j.cgh.2009.05.018.

16. Walk, S.T., Micic, D., Jain, R., Lo, E.S., Trivedi, I., Liu, E.W., Almassalha, L.M., Ewing, S.A., Ring, C., Galecki, A.T., et al. (2012). Clostridium difficile ribotype does not predict severe infection. Clin Infect Dis *55*, 1661–1668. 10.1093/cid/cis786.

17. Bauer, K.A., Johnston, J.E.W., Wenzler, E., Goff, D.A., Cook, C.H., Balada-Llasat, J.-M., Pancholi, P., and Mangino, J.E. (2017). Impact of the NAP-1 strain on disease severity, mortality, and recurrence of healthcare-associated Clostridium difficile infection. Anaerobe *48*, 1–6. 10.1016/j.anaerobe.2017.06.009.

18. Revolinski, S.L., and Munoz-Price, L.S. (2019). Clostridium difficile in Immunocompromised Hosts: A Review of Epidemiology, Risk Factors, Treatment, and Prevention. Clin Infect Dis *68*, 2144–2153. 10.1093/cid/ciy845.

19. Seekatz, A.M., and Young, V.B. (2014). Clostridium difficile and the microbiota. J Clin Invest *124*, 4182–4189. 10.1172/JCI72336.

20. Lewis, B.B., Carter, R.A., Ling, L., Leiner, I., Taur, Y., Kamboj, M., Dubberke, E.R., Xavier, J., and Pamer, E.G. (2017). Pathogenicity Locus, Core Genome, and Accessory Gene Contributions to Clostridium difficile Virulence. mBio *8*, e00885-17. 10.1128/mBio.00885-17.

21. Lyon, S.A., Hutton, M.L., Rood, J.I., Cheung, J.K., and Lyras, D. (2016). CdtR Regulates TcdA and TcdB Production in Clostridium difficile. PLoS Pathog *12*, e1005758. 10.1371/journal.ppat.1005758.

22. Miles-Jay, A., Young, V.B., Pamer, E.G., Savidge, T.C., Kamboj, M., Garey, K.W., and Snitkin, E.S. (2021). A multisite genomic epidemiology study of Clostridioides difficile infections in the USA supports differential roles of healthcare versus community spread for two common strains. Microb Genom *7*, 000590. 10.1099/mgen.0.000590.

23. Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., Gabbard, J.L., et al. (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. *45*, D535–D542. 10.1093/nar/gkw1017.

24. Lebeis, S.L., Bommarius, B., Parkos, C.A., Sherman, M.A., and Kalman, D. (2007). TLR signaling mediated by MyD88 is required for a protective innate immune

response by neutrophils to Citrobacter rodentium. J Immunol *179*, 566–577. 10.4049/jimmunol.179.1.566.

25. Jarchum, I., Liu, M., Shi, C., Equinda, M., and Pamer, E.G. (2012). Critical role for MyD88-mediated neutrophil recruitment during Clostridium difficile colitis. Infect. Immun. *80*, 2989–2996. 10.1128/IAI.00448-12.

26. Buffie, C.G., Bucci, V., Stein, R.R., McKenney, P.T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., et al. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature *517*, 205–208. 10.1038/nature13828.

27. Reeves, A.E., Koenigsknecht, M.J., Bergin, I.L., and Young, V.B. (2012). Suppression of Clostridium difficile in the gastrointestinal tracts of germfree mice inoculated with a murine isolate from the family Lachnospiraceae. Infect Immun *80*, 3786–3794. 10.1128/IAI.00647-12.

28. Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N., Yousef, M., Fogarty, E.C., et al. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol *6*, 3–6. 10.1038/s41564-020-00834-3.

29. Sekulovic, O., Meessen-Pinard, M., and Fortier, L.-C. (2011). Prophage-stimulated toxin production in Clostridium difficile NAP1/027 lysogens. J Bacteriol *193*, 2726–2734. 10.1128/JB.00787-10.

30. Sekulovic, O., Ospina Bedoya, M., Fivian-Hughes, A.S., Fairweather, N.F., and Fortier, L.-C. (2015). The Clostridium difficile cell wall protein CwpV confers phase-variable phage resistance. Mol Microbiol *98*, 329–342. 10.1111/mmi.13121.

31. Nale, J.Y., Spencer, J., Hargreaves, K.R., Buckley, A.M., Trzepiński, P., Douce, G.R., and Clokie, M.R.J. (2016). Bacteriophage Combinations Significantly Reduce Clostridium difficile Growth In Vitro and Proliferation In Vivo. Antimicrobial Agents and Chemotherapy *60*, 968–981. 10.1128/AAC.01774-15.

32. Shan, J., Ramachandran, A., Thanki, A.M., Vukusic, F.B.I., Barylski, J., and Clokie, M.R.J. (2018). Bacteriophages are more virulent to bacteria with human cells than they are in bacterial culture; insights from HT-29 cells. Sci Rep *8*, 5091. 10.1038/s41598-018-23418-y.

33. Fortier, L.-C. (2018). Bacteriophages Contribute to Shaping Clostridioides (Clostridium) difficile Species. Front Microbiol *9*, 2033. 10.3389/fmicb.2018.02033.

34. Heuler, J., Fortier, L.-C., and Sun, X. (2021). Clostridioides difficile phage biology and application. FEMS Microbiol Rev *45*, fuab012. 10.1093/femsre/fuab012.

35. Carter, G.P., Lyras, D., Allen, D.L., Mackin, K.E., Howarth, P.M., O'Connor, J.R., and Rood, J.I. (2007). Binary toxin production in Clostridium difficile is regulated by

CdtR, a LytTR family response regulator. J Bacteriol *189*, 7290–7301. 10.1128/JB.00731-07.

36. Govind, R., and Dupuy, B. (2012). Secretion of Clostridium difficile toxins A and B requires the holin-like protein TcdE. PLoS Pathog *8*, e1002727. 10.1371/journal.ppat.1002727.

37. Tan, K.S., Wee, B.Y., and Song, K.P. (2001). Evidence for holin function of tcdE gene in the pathogenicity of Clostridium difficile. J Med Microbiol *50*, 613–619. 10.1099/0022-1317-50-7-613.

38. Govind, R., Fitzwater, L., and Nichols, R. (2015). Observations on the Role of TcdE Isoforms in Clostridium difficile Toxin Secretion. J Bacteriol *197*, 2600–2609. 10.1128/JB.00224-15.

39. Dupuy, B., Raffestin, S., Matamouros, S., Mani, N., Popoff, M.R., and Sonenshein, A.L. (2006). Regulation of toxin and bacteriocin gene expression in Clostridium by interchangeable RNA polymerase sigma factors. Mol Microbiol *60*, 1044–1057. 10.1111/j.1365-2958.2006.05159.x.

40. Mani, N., Lyras, D., Barroso, L., Howarth, P., Wilkins, T., Rood, J.I., Sonenshein, A.L., and Dupuy, B. (2002). Environmental response and autoregulation of Clostridium difficile TxeR, a sigma factor for toxin gene expression. J Bacteriol *184*, 5971–5978. 10.1128/JB.184.21.5971-5978.2002.

41. Eyre, D.W., Cule, M.L., Wilson, D.J., Griffiths, D., Vaughan, A., O'Connor, L., Ip, C.L.C., Golubchik, T., Batty, E.M., Finney, J.M., et al. (2013). Diverse sources of C. difficile infection identified on whole-genome sequencing. N Engl J Med *369*, 1195–1205. 10.1056/NEJMoa1216064.

42. Bilverstone, T.W., Minton, N.P., and Kuehne, S.A. (2019). Phosphorylation and functionality of CdtR in Clostridium difficile. Anaerobe *58*, 103–109. 10.1016/j.anaerobe.2019.102074.

43. Chen, R., Feng, Y., Wang, X., Yang, J., Zhang, X., Lü, X., and Zong, Z. (2017). Whole genome sequences of three Clade 3 Clostridium difficile strains carrying binary toxin genes in China. Sci Rep *7*, 43555. 10.1038/srep43555.

44. Barbut, F., Decré, D., Lalande, V., Burghoffer, B., Noussair, L., Gigandon, A., Espinasse, F., Raskine, L., Robert, J., Mangeol, A., et al. (2005). Clinical features of Clostridium difficile-associated diarrhoea due to binary toxin (actin-specific ADP-ribosyltransferase)-producing strains. J Med Microbiol *54*, 181–185. 10.1099/jmm.0.45804-0.

45. Young, M.K., Leslie, J.L., Madden, G.R., Lyerly, D.M., Carman, R.J., Lyerly, M.W., Stewart, D.B., Abhyankar, M.M., and Petri, W.A. (2022). Binary Toxin Expression by Clostridioides difficile Is Associated With Worse Disease. Open Forum Infect Dis *9*, ofac001. 10.1093/ofid/ofac001.

745 46. Carter, G.P., Chakravorty, A., Pham Nguyen, T.A., Mileto, S., Schreiber, F., Li, L.,
746 Howarth, P., Clare, S., Cunningham, B., Sambol, S.P., et al. (2015). Defining the
747 Roles of TcdA and TcdB in Localized Gastrointestinal Disease, Systemic Organ
748 Damage, and the Host Response during Clostridium difficile Infections. mBio *6*,
749 e00551. 10.1128/mBio.00551-15.

750 47. Cowardin, C.A., Buonomo, E.L., Saleh, M.M., Wilson, M.G., Burgess, S.L., Kuehne,
751 S.A., Schwan, C., Eichhoff, A.M., Koch-Nolte, F., Lyras, D., et al. (2016). The
752 binary toxin CDT enhances Clostridium difficile virulence by suppressing protective
753 colonic eosinophilia. Nat Microbiol *1*, 16108. 10.1038/nmicrobiol.2016.108.

754 48. Metcalf, D.S., and Weese, J.S. (2011). Binary toxin locus analysis in Clostridium
755 difficile. J Med Microbiol *60*, 1137–1145. 10.1099/jmm.0.028498-0.

756 49. Kuehne, S.A., Cartman, S.T., Heap, J.T., Kelly, M.L., Cockayne, A., and Minton,
757 N.P. (2010). The role of toxin A and toxin B in Clostridium difficile infection. Nature
758 *467*, 711–713. 10.1038/nature09397.

759 50. Alonso, C.D., Pollock, N.R., Garey, K.W., Gonzales-Luna, A.J., Williams, D.N.,
760 Daugherty, K., Cuddemi, C., Villafuerte-Gálvez, J., White, N.C., Chen, X., et al.
761 (2022). Higher In Vivo Fecal Concentrations of Clostridioides difficile Toxins A and
762 B in Patients With North American Pulsed-Field Gel Electrophoresis Type
763 1/Ribotype 027 Strain Infection. Clin Infect Dis *75*, 2019–2022. 10.1093/cid/ciac406.

764 51. Alonso, C.D., Kelly, C.P., Garey, K.W., Gonzales-Luna, A.J., Williams, D.,
765 Daugherty, K., Cuddemi, C., Villafuerte-Gálvez, J., White, N.C., Chen, X., et al.
766 (2022). Ultrasensitive and Quantitative Toxin Measurement Correlates With Baseline
767 Severity, Severe Outcomes, and Recurrence Among Hospitalized Patients With
768 Clostridioides difficile Infection. Clin Infect Dis *74*, 2142–2149.
769 10.1093/cid/ciab826.

770 52. Guh, A.Y., Hatfield, K.M., Winston, L.G., Martin, B., Johnston, H., Brousseau, G.,
771 Farley, M.M., Wilson, L., Perlmutter, R., Phipps, E.C., et al. (2019). Toxin Enzyme
772 Immunoassays Detect Clostridioides difficile Infection With Greater Severity and
773 Higher Recurrence Rates. Clinical Infectious Diseases *69*, 1667–1674.
774 10.1093/cid/ciz009.

775 53. Gálvez, J.A.V., Pollock, N.R., Alonso, C.D., Chen, X., Xu, H., Wang, L., White, N.,
776 Banz, A., Miller, M., Daugherty, K., et al. (2022). Stool interleukin-1β differentiates
777 Clostridioides difficile infection from asymptomatic carriage and non-C.difficile
778 infection diarrhea. Clin Infect Dis, ciac624. 10.1093/cid/ciac624.

779 54. Govind, R., Vediyappan, G., Rolfe, R.D., Dupuy, B., and Fralick, J.A. (2009).
780 Bacteriophage-Mediated Toxin Gene Regulation in Clostridium difficile. J Virol *83*,
781 12037–12045. 10.1128/JVI.01256-09.

782 55. Goh, S., Chang, B.J., and Riley, T.V. (2005). Effect of phage infection on toxin
783 production by Clostridium difficile. J Med Microbiol *54*, 129–135.
784 10.1099/jmm.0.45821-0.

785 56. Hargreaves, K.R., Kropinski, A.M., and Clokie, M.R.J. (2014). What does the
786 talking?: quorum sensing signalling genes discovered in a bacteriophage genome.
787 PLoS One *9*, e85131. 10.1371/journal.pone.0085131.

788 57. Nale, J.Y., Chutia, M., Carr, P., Hickenbotham, P.T., and Clokie, M.R.J. (2016). "Get
789 in Early"; Biofilm and Wax Moth (Galleria mellonella) Models Reveal New Insights
790 into the Therapeutic Potential of Clostridium difficile Bacteriophages. Front
791 Microbiol *7*, 1383. 10.3389/fmicb.2016.01383.

792 58. Griffiths, D., Fawley, W., Kachrimanidou, M., Bowden, R., Crook, D.W., Fung, R.,
793 Golubchik, T., Harding, R.M., Jeffery, K.J.M., Jolley, K.A., et al. (2010). Multilocus
794 Sequence Typing of Clostridium difficile. Journal of Clinical Microbiology *48*, 770–
795 778. 10.1128/JCM.01796-09.

796 59. Sorg, J.A., and Sonenshein, A.L. (2008). Bile salts and glycine as cogerminants for
797 Clostridium difficile spores. J Bacteriol *190*, 2505–2512. 10.1128/JB.01765-07.

798 60. Abt, M.C., Lewis, B.B., Caballero, S., Xiong, H., Carter, R.A., Sušac, B., Ling, L.,
799 Leiner, I., and Pamer, E.G. (2015). Innate Immune Defenses Mediated by Two ILC
800 Subsets Are Critical for Protection against Acute Clostridium difficile Infection. Cell
801 Host Microbe *18*, 27–37. 10.1016/j.chom.2015.06.011.

802 61. Keith, J.W., Dong, Q., Sorbara, M.T., Becattini, S., Sia, J.K., Gjonbalaj, M., Seok, R.,
803 Leiner, I.M., Littmann, E.R., and Pamer, E.G. (2020). Impact of Antibiotic-Resistant
804 Bacteria on Immune Activation and Clostridioides difficile Infection in the Mouse
805 Intestine. Infect Immun *88*, e00362-19. 10.1128/IAI.00362-19.

806 62. Kaus, G.M., Snyder, L.F., Müh, U., Flores, M.J., Popham, D.L., and Ellermeier, C.D.
807 (2020). Lysozyme Resistance in Clostridioides difficile Is Dependent on Two
808 Peptidoglycan Deacetylases. J Bacteriol *202*, e00421-20. 10.1128/JB.00421-20.

809 63. Edwards, A.N., Krall, E.G., and McBride, S.M. (2020). Strain-Dependent RstA
810 Regulation of Clostridioides difficile Toxin Production and Sporulation. J Bacteriol
811 *202*, e00586-19. 10.1128/JB.00586-19.

812 64. McAllister, K.N., Bouillaut, L., Kahn, J.N., Self, W.T., and Sorg, J.A. (2017). Using
813 CRISPR-Cas9-mediated genome editing to generate C. difficile mutants defective in
814 selenoproteins synthesis. Sci Rep *7*, 14672. 10.1038/s41598-017-15236-5.

815 65. Angione, S.L., Sarma, A.A., Novikov, A., Seward, L., Fieber, J.H., Mermel, L.A.,
816 and Tripathi, A. (2014). A novel subtyping assay for detection of Clostridium
817 difficile virulence genes. J Mol Diagn *16*, 244–252. 10.1016/j.jmoldx.2013.11.006.

66. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics *70*, e102. 10.1002/cpbi.102.

67. Babakhani, F., Bouillaut, L., Sears, P., Sims, C., Gomez, A., and Sonenshein, A.L. (2013). Fidaxomicin inhibits toxin production in Clostridium difficile. Journal of Antimicrobial Chemotherapy *68*, 515–522. 10.1093/jac/dks450.

68. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069. 10.1093/bioinformatics/btu153.

69. Wroblewski, D., Hannett, G.E., Bopp, D.J., Dumyati, G.K., Halse, T.A., Dumas, N.B., and Musser, K.A. (2009). Rapid molecular characterization of Clostridium difficile and assessment of populations of C. difficile in stool specimens. J Clin Microbiol *47*, 2142–2148. 10.1128/JCM.02498-08.

70. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol *13*, e1005595. 10.1371/journal.pcbi.1005595.

71. Metcalf, D., Sharif, S., and Weese, J.S. (2010). Evaluation of candidate reference genes in Clostridium difficile for gene expression normalization. Anaerobe *16*, 439–443. 10.1016/j.anaerobe.2010.06.007.

72. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421. 10.1186/1471-2105-10-421.

73. Seemann, T. (2023). mlst.

74. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659. 10.1093/bioinformatics/btl158.

75. L, F., B, N., Z, Z., S, W., and W, L. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics (Oxford, England) *28*. 10.1093/bioinformatics/bts565.

76. Seemann, T. (2023). Snippy.

77. Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J., and Harris, S.R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res *43*, e15. 10.1093/nar/gku1196.

78. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol *26*, 1641–1650. 10.1093/molbev/msp077.

79. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490. 10.1371/journal.pone.0009490.

80. Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun *9*, 5114. 10.1038/s41467-018-07641-9.

81. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat Methods *12*, 59–60. 10.1038/nmeth.3176.

82. van Dongen, S., and Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. Methods Mol Biol *804*, 281–295. 10.1007/978-1-61779-361-5_15.

83. Benedict, M.N., Henriksen, J.R., Metcalf, W.W., Whitaker, R.J., and Price, N.D. (2014). ITEP: An integrated toolkit for exploration of microbial pan-genomes. BMC Genomics *15*, 8. 10.1186/1471-2164-15-8.

84. Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res *43*, D261-269. 10.1093/nar/gku1223.

85. Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res *44*, W16-21. 10.1093/nar/gkw387.

86. Sekulovic, O., and Fortier, L.-C. (2015). Global transcriptional response of Clostridium difficile carrying the CD38 prophage. Appl Environ Microbiol *81*, 1364–1374. 10.1128/AEM.03656-14.

87. Sekulović, O., and Fortier, L.-C. (2016). Characterization of Functional Prophages in Clostridium difficile. In Clostridium difficile: Methods and Protocols Methods in Molecular Biology., A. P. Roberts and P. Mullany, eds. (Springer), pp. 143–165. 10.1007/978-1-4939-6361-4_11.

Table 1 Oligonucleotides used in this study

Name	Sequence 5'-3'	Reference	Purpose
tcdA_qFor	GTATGGATAGGTGGAGAAGTCA	Babakhani 2012 J Antimicrobial Chemotherapy	Forward primer for tcdA transcription analysis
tcdA_qRev	CTCTTCCTCTAGTAGCTGTAATGC	Babakhani 2012 J Antimicrobial Chemotherapy	Reverse primer for tcdA transcription analysis
tcdB_qFor	AGCAGTTGAATATAGTGGTTTAGTTAGAGTTG	Wroblewski 2009 J Clin Microbiol.	Forward primer for tcdB transcription analysis
tcdB_qRev	CATGCTTTTTTAGTTTCTGGATTGAA	Wroblewski 2009 J Clin Microbiol.	Reverse primer for tcdB transcription analysis
tcdE_qFor	ATAAACCTAGGAGGCGTTATGAATATGA	Edwards 2020 J Bacteriology	Forward primer for tcdE transcription analysis
tcdE_qRev	TTATTGCACTTAAACATCCTAATAATGTATCAAA	Edwards 2020 J Bacteriology	Reverse primer for tcdE transcription analysis
tcdR_qFor	AGCAAGAAATAACTCAGTAGATGATT	Edwards 2020 J Bacteriology	Forward primer for tcdR transcription analysis
tcdR_qRev	TTATTAAATCTGTTTCTCCCTCTTCA	Edwards 2020 J Bacteriology	Reverse primer for tcdR transcription analysis
cdtB_qFor	GCAGTTAAGTGGGAAGATAG	Angione 2014 J Mol Diagn	Forward primer for cdtB transcription analysis
cdtB_qRev	TCCATACCTACTCCAACAAT	Angione 2014 J Mol Diagn	Reverse primer for cdtB transcription analysis
cdtR-2_qFor	TTGAAACAAGCGCTATTCCACA	This study	Forward primer for cdtR transcription analysis
cdtR-2_qRev	TGTACACGAATAAAGCATGCATC	This study	Reverse primer for cdtR transcription analysis
rpsJ_qFor	GATCACAAGTTTCAGGACCTG	Metcalf 2010 Anaerobe	Forward primer for rspJ transcription analysis
rpsJ_qRev	GTCTTAGGTGTTGGATTAGC	Metcalf 2010 Anaerobe	Reverse primer for rspJ transcription analysis
adk_qFor	GTGTATGTGATGTATGCCAAG	Metcalf 2010 Anaerobe	Forward primer for adk transcription analysis
adk_qRev	CCTAAGGCTGCGACAATATC	Metcalf 2010 Anaerobe	Reverse primer for adk transcription analysis
R_cdtR_up_For	AAACAGCTATGACCGCGGCCGCCTAAACACACA TTATCATCTCTCTG	This study	Forward primer amplifying upstream region of cdtR gene for cdtRKO, cdtRstop and cdtRmut
R_cdtRKO_up_Rev	AACTTTCAGTTTAGCGGTCTGGGCGCCTAAATAC CCTCCTATAAAAAATTCAAAAG	This study	Reverse primer amplifying upstream region of cdtR gene for cdtRKO
R_cdtRKO_down_For	GGCGCCCAGACCGCTAAACTGAAAGTTTAAATA GAAAAAAGAGATGTCTCAAGATAAG	This study	Forward primer amplifying downstream region of cdtR gene for cdtRKO
R_cdtRKO_down_Rev	TTATTTTTATGCTAGCTCGAGTAAGTCTTGTGCA TAAATGTTATTAGG	This study	Reverse primer amplifying downstream region of cdtR gene for cdtRKO
R_cdtRstop_up_Rev	AACTTTCAGTTTAGCGGTCTGGGCGCCTTATCAA AAATTAATATATCCACTAAATACCC	This study	Reverse primer amplifying upstream region of cdtR gene for cdtRstop
R_cdtRstop_down_For	GGCGCCCAGACCGCTAAACTGAAAGTTTTTTGA TAACGATGTTATAAGATTATATTATAA	This study	Forward primer amplifying downstream region of cdtR gene for cdtRstop
R_cdtRstop/mut_down_Rev	TTATTTTTATGCTAGCTCGAGATCTGATAAAGAC CTTAAACTTTTATAG	This study	Reverse primer amplifying downstream region of cdtR gene for cdtRstop and cdtRmut
R_cdtRmut_up_Rev	TTATAATATAATCTTATAACATCGTTATCAAAAA TTAATATATCCACTAAATACCC	This study	Reverse primer amplifying upstream region of cdtR gene for cdtRmut
R_cdtRmut_down_For	GGGTATTTAGTGGATATATTAATTTTTGATAACG ATGTTATAAGATTATATTATAA	This study	Forward primer amplifying downstream region of cdtR gene for cdtRmut
R_cdtR_sgRNA1c_For	aattaaactgtaaatggccaAATAATATTTTAATAAAA GAGTTTTAGAGCTAGAAATAGC	This study	Forward primer gRNA Cloning into pCE677 digested with MscI and MluI
CDEP3876	aaccatctaaaaatagttgcagagcttACGCGTC	Kaus 2020 J Bacteriol.	Universal Reverse primer for gRNA Cloning into pCE677 digested with MscI and MluI

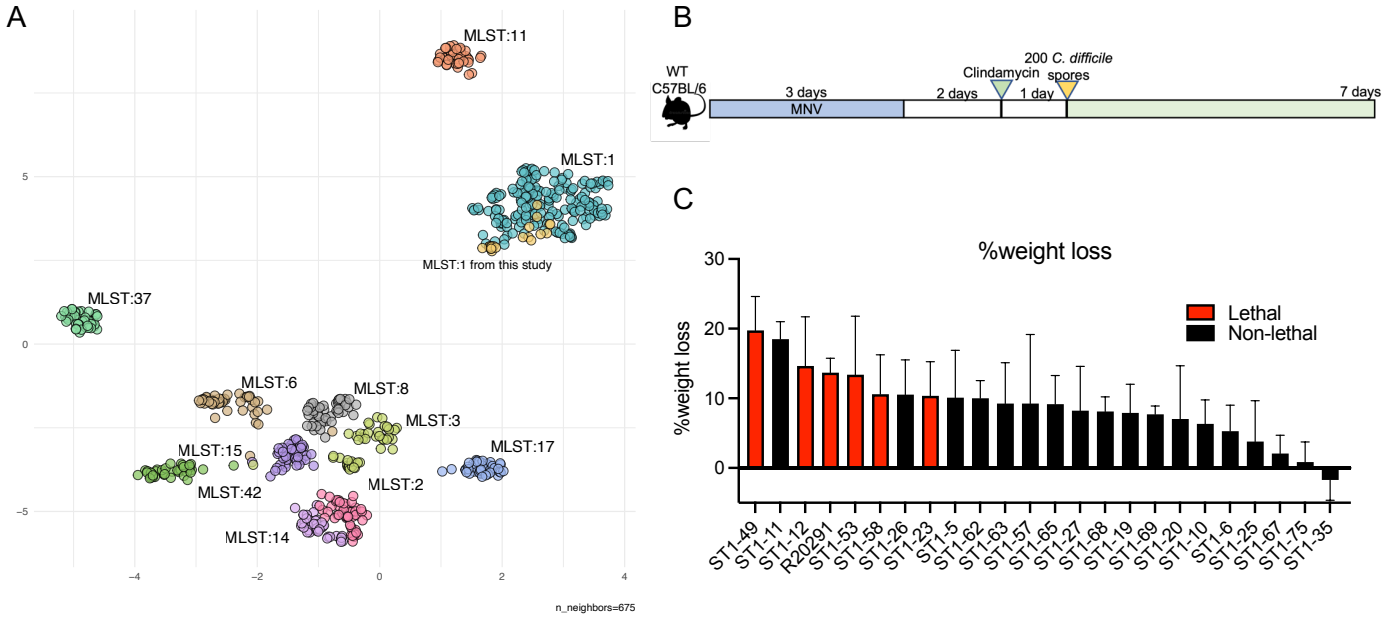


Figure 1. Clinical ST1 *C. difficile* isolates demonstrated variable virulence in mice treated with antibiotics. (A) Plot of the a UMAP analysis of the presence or absence of unique coding sequences (annotated proteins or un-annotated protein clusters) across top 10 STs of *C. difficile* strains in Patric. (B) Mouse experiment schematic: wildtype C57BL/6 mice were treated with metronidazole, vancomycin, and neomycin (MNV, 0.25 g/L for each) in drinking water for 3 days and followed by one intraperitoneal injection of clindamycin (200 µg/mouse) 2 days after antibiotic recess. Then, mice were inoculated with 200 *C. difficile* spores via oral gavage. Daily body weight and acute disease scores were monitored for 7 days post infection. (C) %Max weight loss to baseline were calculated using the lowest weights within 7 days post infection divided by day 0 weights. N (Number of mice per strain-infected group) = 5-8 except for ST1-62 and ST1-68, which have 2 mice per group.

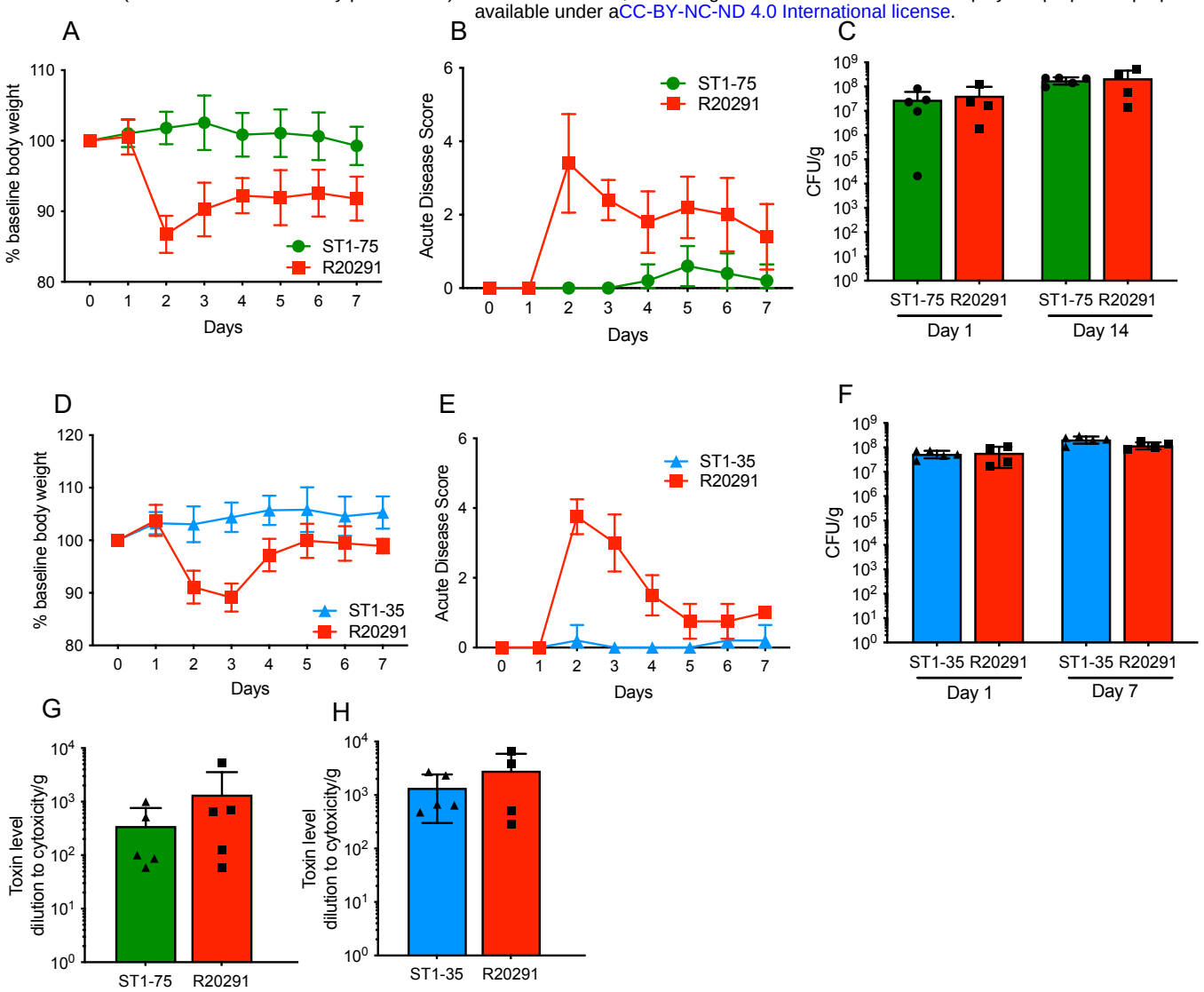


Figure 2. Two isolated clinical strains of *C. difficile* have no virulence in mice treated with antibiotics. Wildtype C57BL/6 mice (n=3-5 per group) were treated with metronidazole, vancomycin, and neomycin (MNV, 0.25 g/L for each) in drinking water for 3 days and followed by one intraperitoneal injection of clindamycin (200 μg /mouse) 2 days after antibiotic recess. Then, mice were inoculated with 200 *C. difficile* spores via oral gavage. Daily body weight and acute disease scores were monitored for 7 days post infection. (A, D) %Weight loss to baseline of mice infected with indicated strains. (B, E) Acute disease scores comprising weight loss, body temperature drop, diarrhea, morbidity of mice infected with indicated strains. (C, F) Fecal colony-forming units measured by plating on selective agar on indicated days.(G-H) Fecal toxins measured by CHO cell rounding assay 1 day post infection.

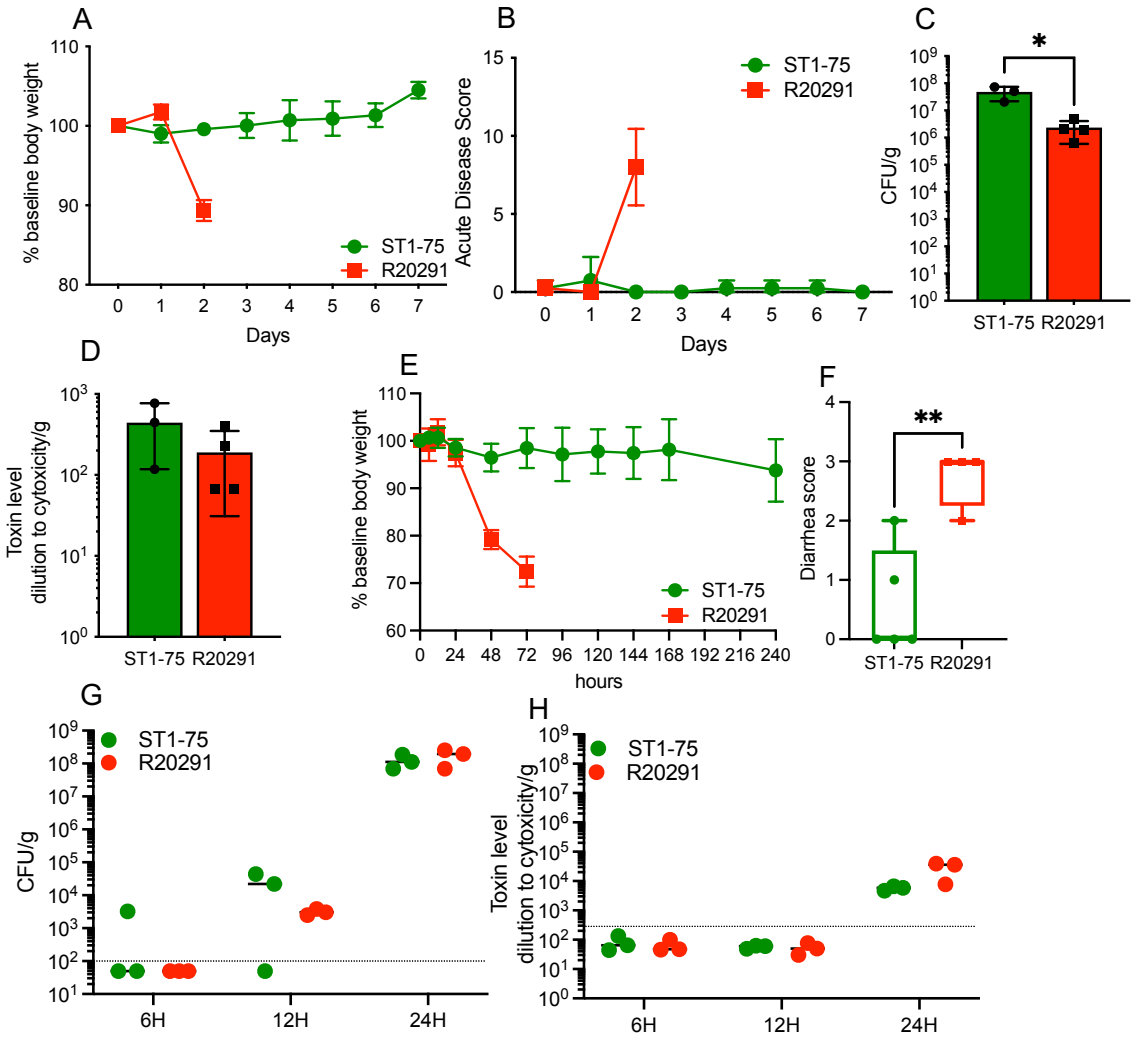


Figure 3. Avirulent *C. difficile* strain demonstrates no virulence in innate immune deficient mice and germ-free mice. (A-D) MyD88$^{-/-}$mice (n=4 per group) were treated with MNV and clindamycin before orally administered with 200 spores of *C. difficile* strains. Daily body weight and acute disease scores were monitored for 7 days post infection. (A) %Weight loss to baseline of mice infected with indicated strains. (B) Acute disease scores comprising weight loss, body temperature drop, diarrhea, morbidity of mice infected with indicated strains. (C) Fecal colony-forming units measured by plating on selective agar 1 day post infection. (D) Fecal toxins measured by CHO cell rounding assay 1 day post infection. (E-H) Germ-free mice (n=3 to 5) orally administered with 200 spores of indicated *C. difficile* strains. Daily body weight and acute disease scores were monitored for 10 days post infection. (E) %Weight loss to baseline of mice infected with indicated strains. (F) Diarrhea scores of mice infected with indicated strains 2 days post infection. (G) Fecal colony-forming units measured by plating on selective agar at 6, 12, and 24 hours post infection. (H) Fecal toxins measured by CHO cell rounding assay at 6, 12, and 24 hours post infection. Statistical significance was calculated by Unpaired t-test, * $p < 0.05$, ** $p < 0.01$.

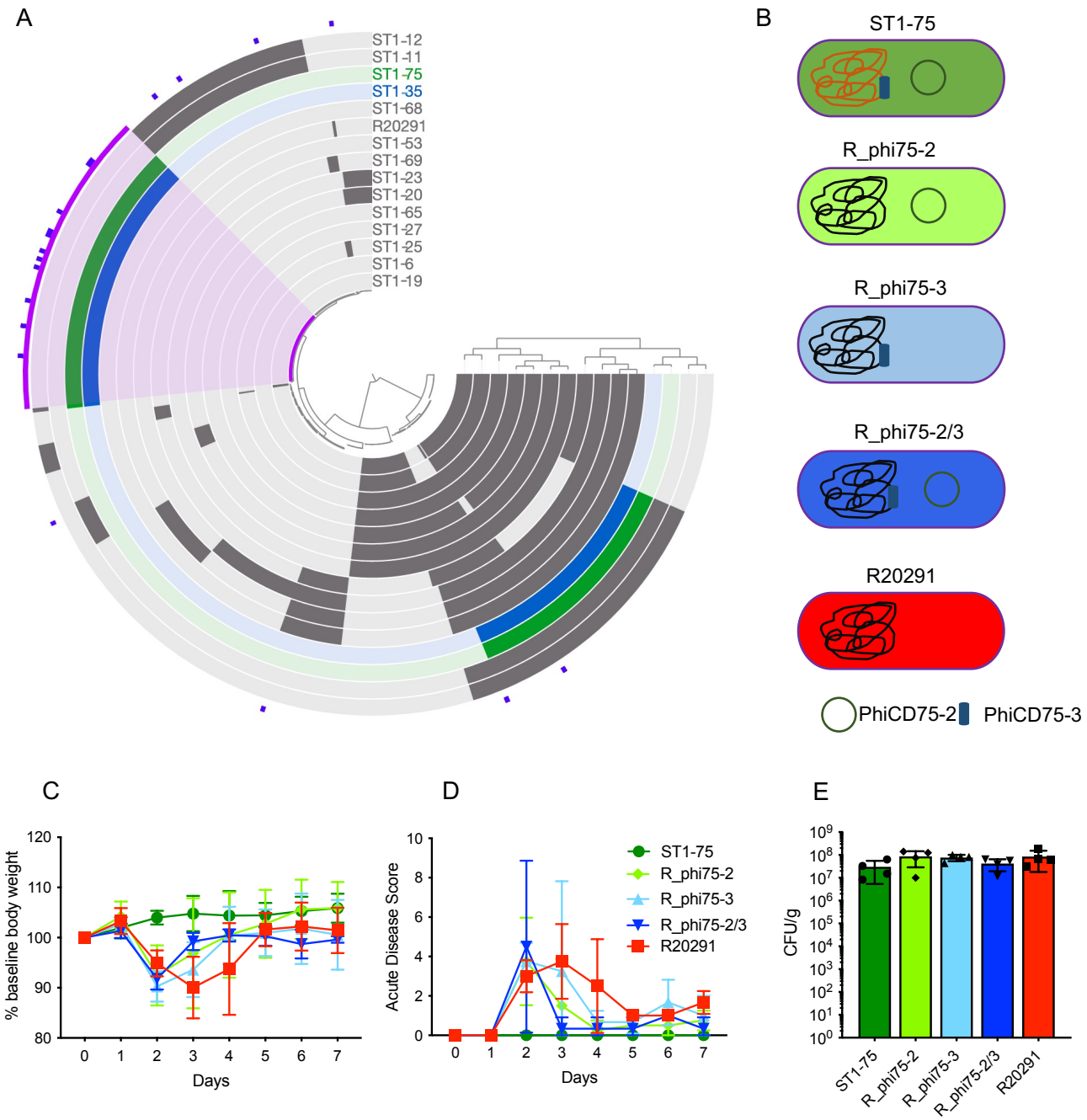


Figure 4. Prophages identified in avirulent *C. difficile* mildly impact virulence in mice treated with antibiotics. (A) Anvi'o plot displaying accessory genomes of ST1 isolates. Highlighted gene clusters in purple are unique to ST1-35 and ST1-75. Blue dashes (outmost layer) indicate phage-related genes by NCBI COG. (B) Schematic of mutant strains made using R20291 *C. difficile* strain. (C-E) Wildtype C57BL/6 mice (n=4 per group) were treated with MNV and clindamycin as preciously described. Then, mice were inoculated with 200 *C. difficile* spores via oral gavage. Daily body weight and acute disease scores were monitored for 7 days post infection. (C) %Weight loss to baseline of mice infected with indicated strains. (D) Acute disease scores comprising weight loss, body temperature drop, diarrhea, morbidity of mice infected with indicated strains. (E) Fecal colony-forming units measured by plating on selective agar 1 day post infection.

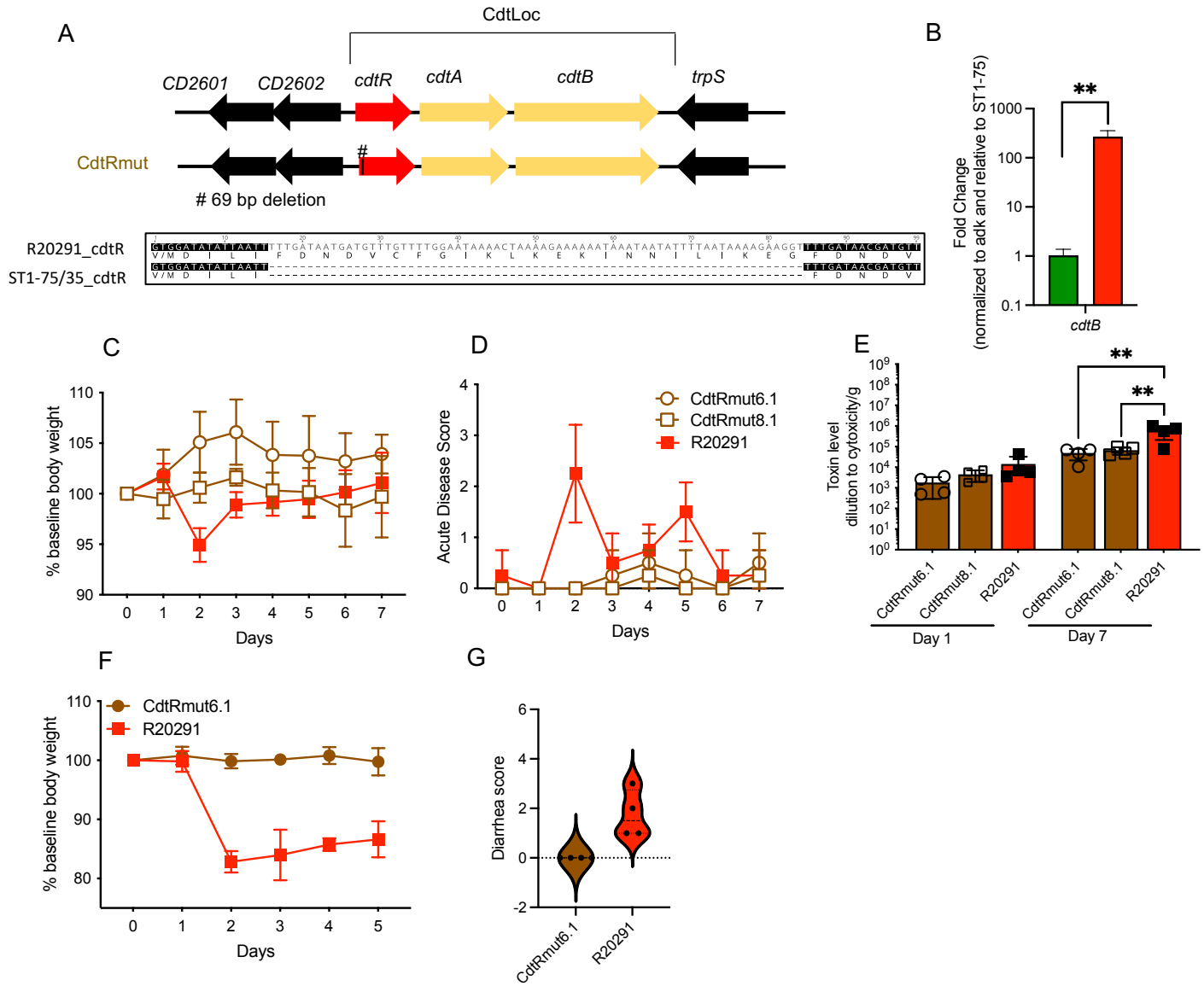


Figure 5. Binary toxin regulator *cdtR* contributes to *C. difficile* virulence in mice. (A) Deletion identified in ST1-35/75 and schematic of *cdtR* mutants generated using R20291 *C. difficile* strain. (B) Germ-free mice (n=3 per group) orally administered with 200 spores of indicated *C. difficile* strains. Binary toxin gene *cdtB* transcripts were measured by RT-qPCR on cecal contents harvested at 24 hours post infection. Transcripts were normalized to the *adk* and fold change is relative to ST1-75 condition. (C-E) Wildtype C57BL/6 mice (n=3 to 5 per group) were treated with MNV and clindamycin as preciously described. Then, mice were inoculated with 200 *C. difficile* spores via oral gavage. Daily body weight and acute disease scores were monitored for 7 days post infection. (C) %Weight loss to baseline of mice infected with indicated strains. (D) Acute disease scores comprising weight loss, body temperature drop, diarrhea, morbidity of mice infected with indicated strains. (E) Fecal toxins measured by CHO cell rounding assay on indicated days. (F-G) Germ-free mice (n=4) orally administered with 200 spores of indicated *C. difficile* strains. Daily body weight and were monitored for 5 days post infection. (F) %Weight loss to baseline of mice infected with indicated strains. (G) Diarrhea scores of mice infected with indicated strains 3 days post infection. Statistical significance was calculated by One-way ANOVA, * p <0.05, ** < p< 0.01.

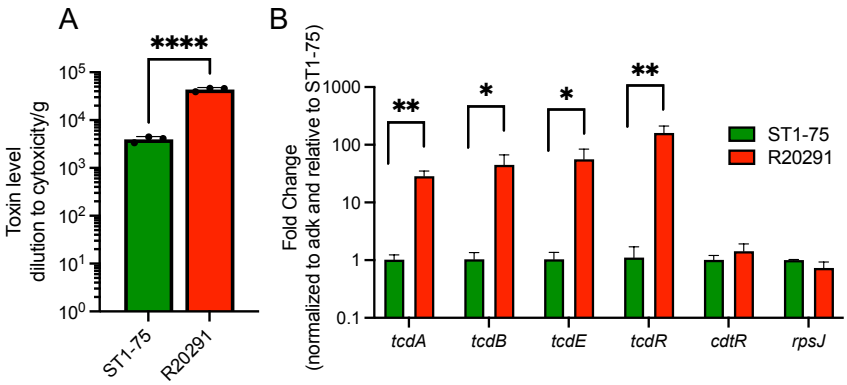


Figure 6. CdtR regulates PaLoc toxins transcription *in vivo*. Germ-free mice (n=4) were orally administered with 200 spores of indicated *C. difficile* strains and cecal contents were harvested at 24 hours post infection. (A) Cecal toxins measured by CHO cell rounding assay. (B) Indicated gene transcripts were measured by RT-qPCR. Transcripts were all normalized to the *adk* and fold change is relative to ST1-75 for each of the genes. Statistical significance was calculated by Unpaired t-test, * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$.

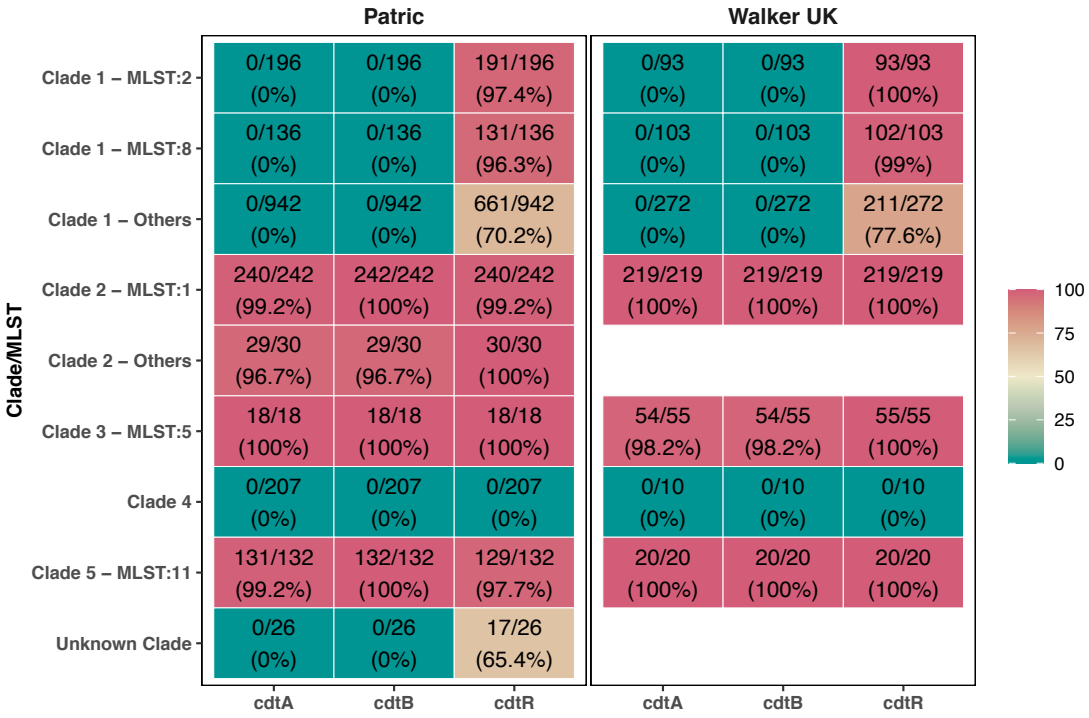
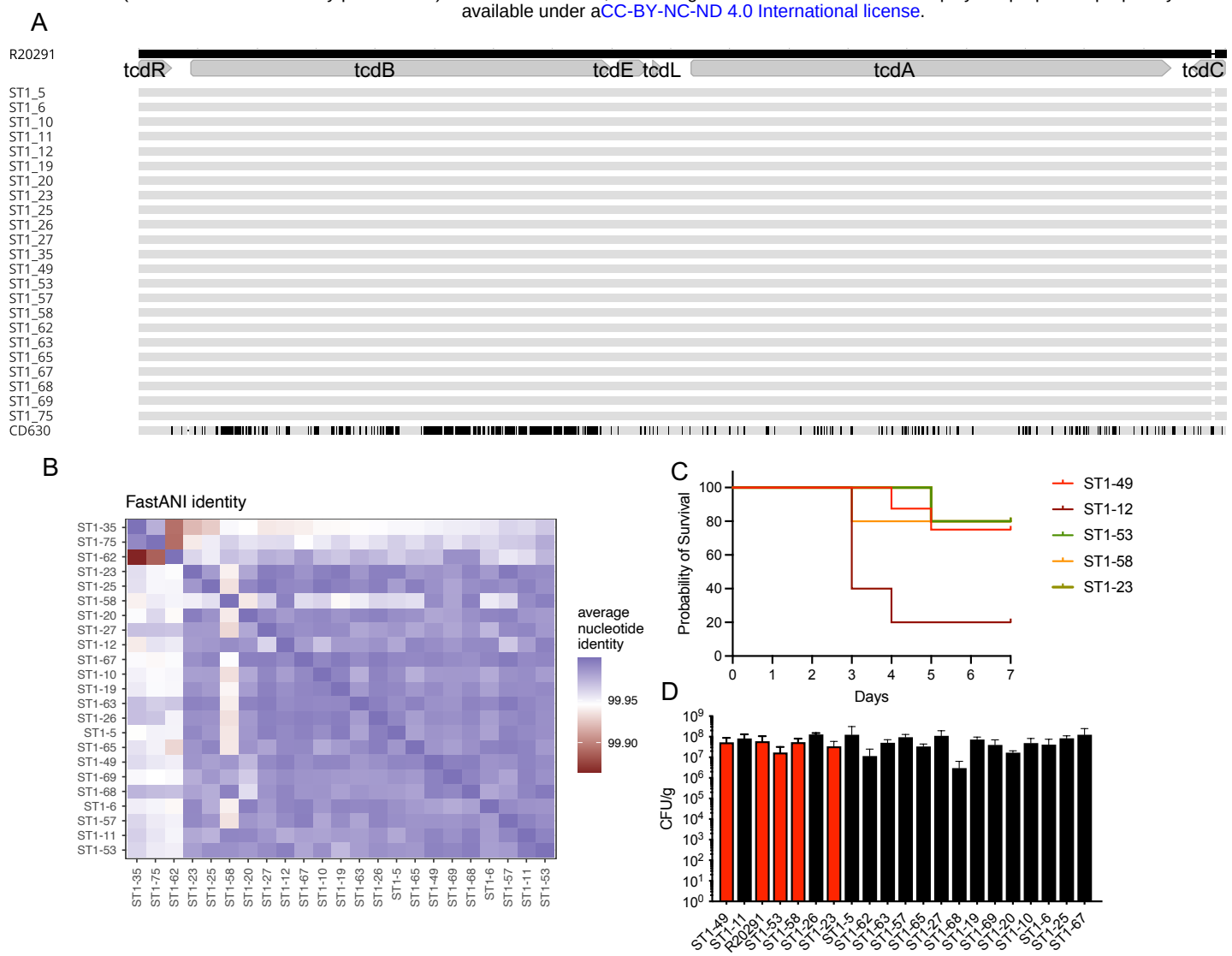
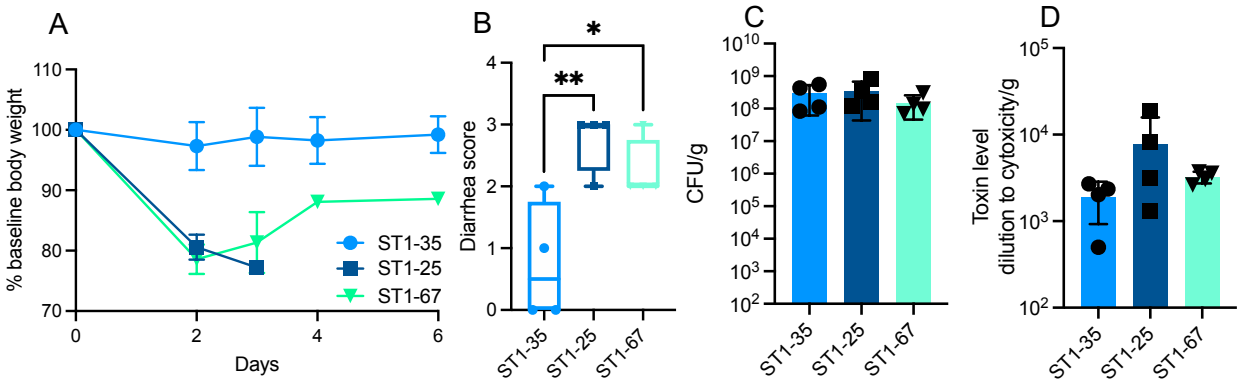


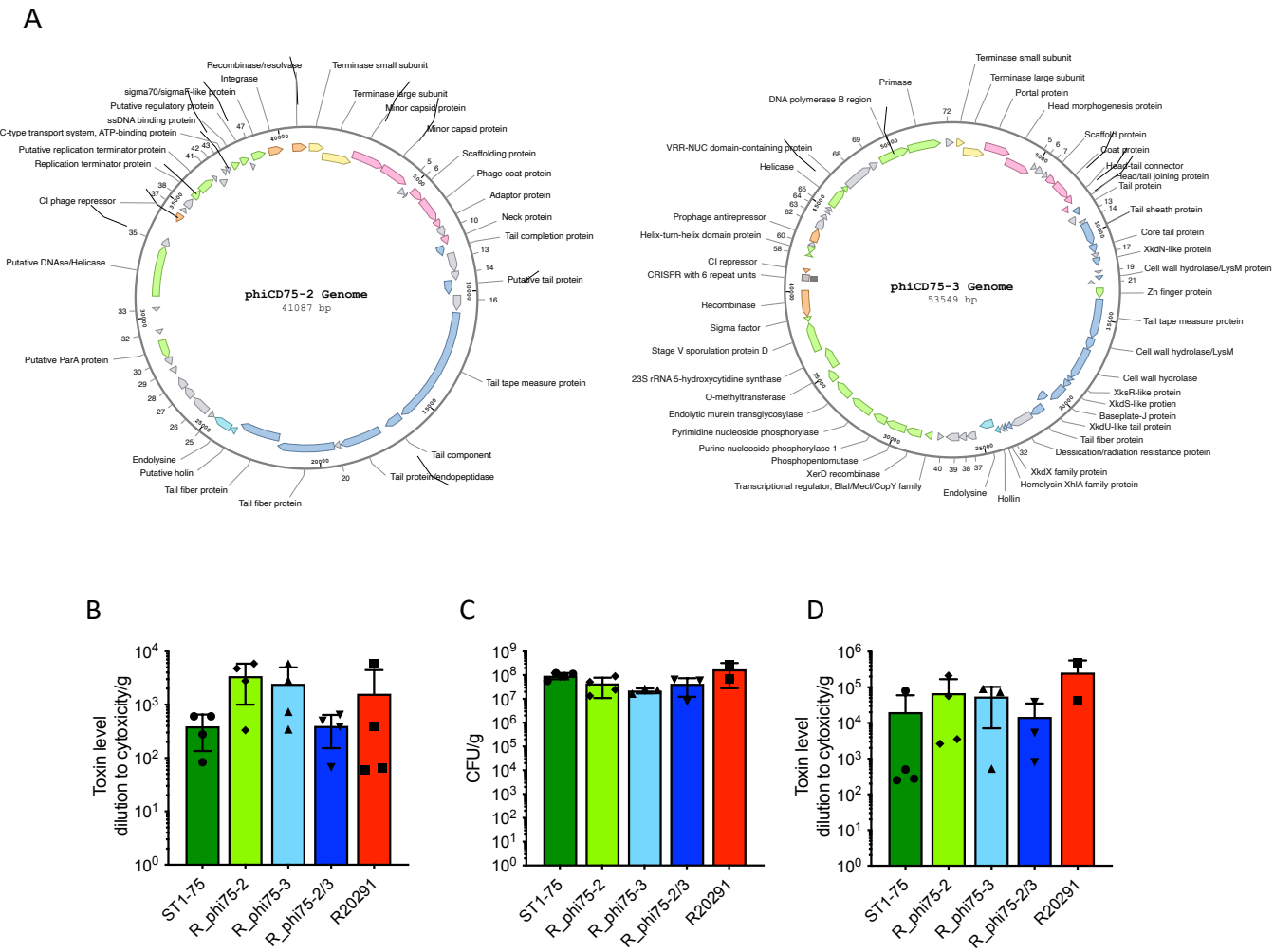
Figure 7. Binary toxin regulator *cdtR* is prevalent in clinical *C. difficile* isolates.
Binary toxin *cdtA, cdtB* and *cdtR* from R20291 were used as query to BLAST against the assembled contigs. Hits with at least 85% identity and 85% coverage of the query were considered a valid match. Numbers of match in total and percentages are presented.



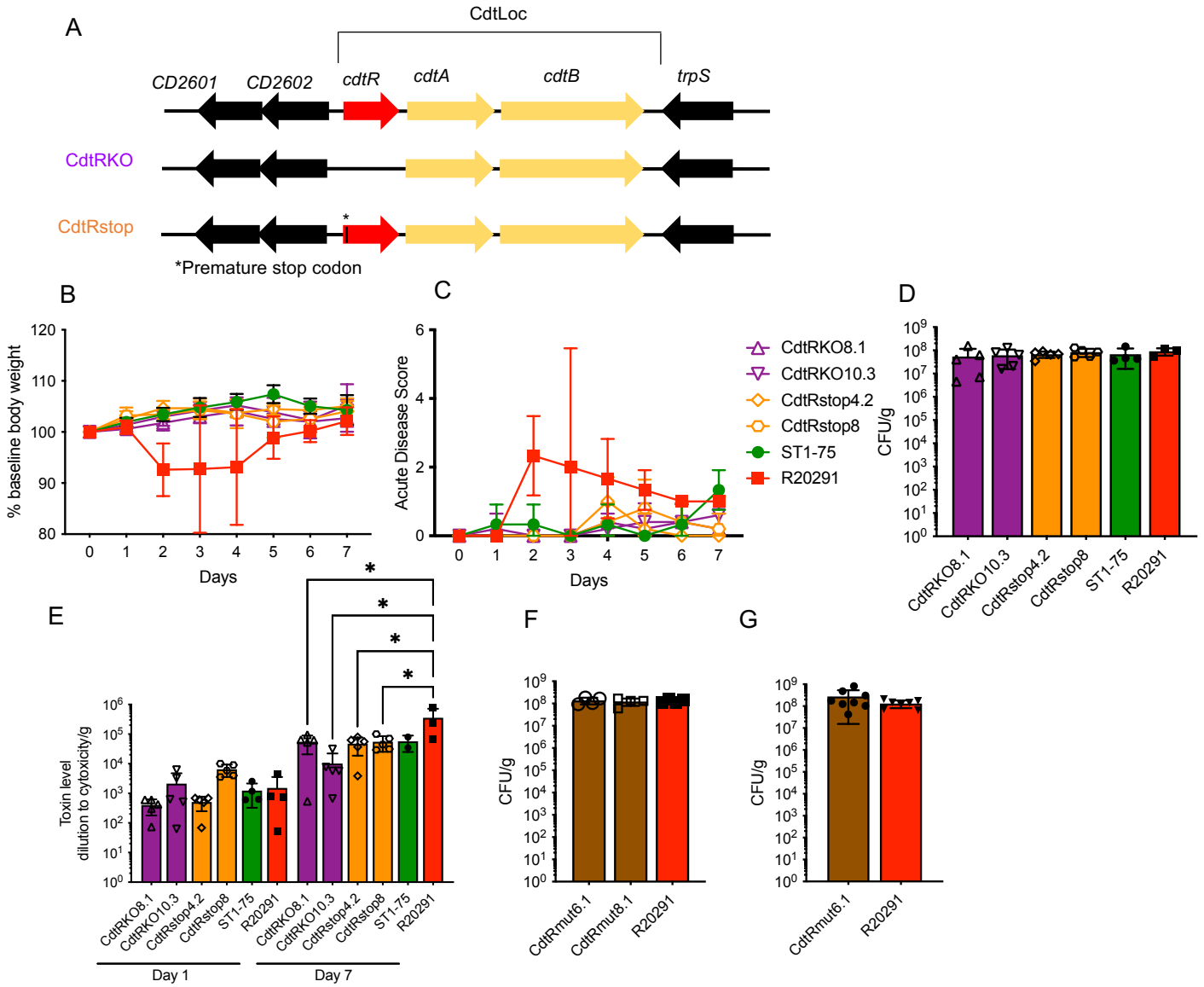
Supplemental Figure 1. ST1 *C. difficile* strains are closely related. (A) Pathogenicity loci were extracted from whole-genome sequence and multiple sequence alignment was performed for all strains. Each dash indicates one single-nucleotide polymorphism. (B) Average nucleotide identity was calculated with pairs of ST1 isolates. (C) Survival curve of indicated strain over a 7-day time course post infection. (D) Fecal colony-forming units measured by plating on selective agar on 1 day post infection from Figure 1C.



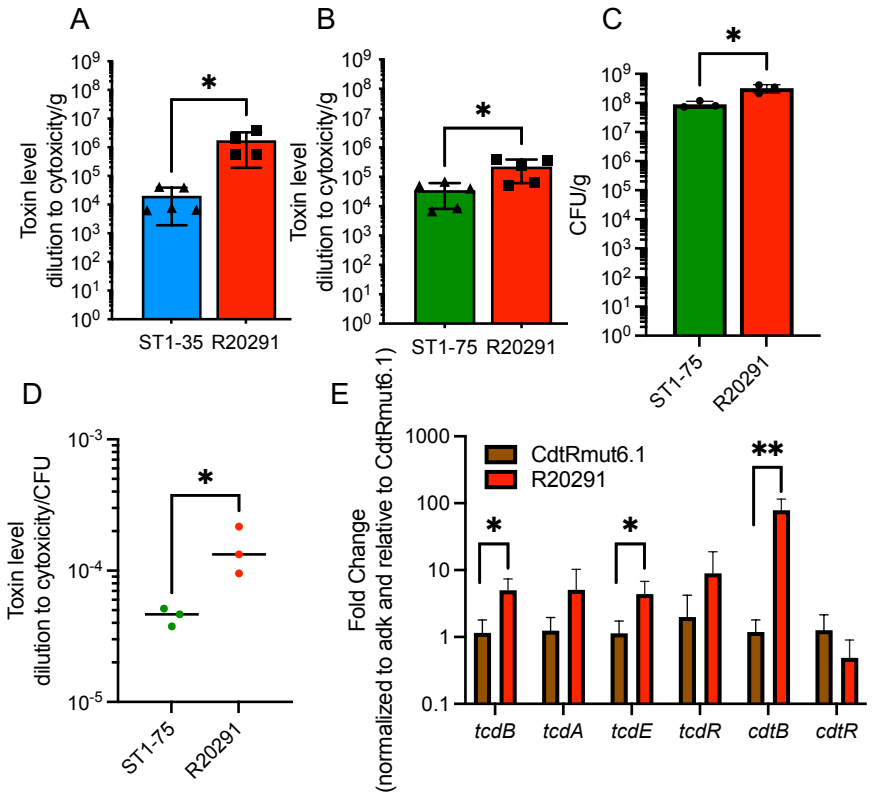
Supplemental Figure 2. Avirulent *C. difficile* strain demonstrates no virulence in germ-free mice. Germ-free mice (n=4 per group) orally administered with 200 spores of indicated *C. difficile* strains. Daily body weight and acute disease scores were monitored for 6 days post infection. (A) %Weight loss to baseline of mice infected with indicated strains. (B) Diarrhea scores of mice infected with indicated strains on 3 days post infection. (C) Fecal colony-forming units measured by plating on selective agar 1 day post infection. (D) Fecal toxins measured by CHO cell rounding assay 1 day post infection. Statistical significance was calculated by Unpaired t-test, * $p < 0.05$, ** $p < 0.01$.



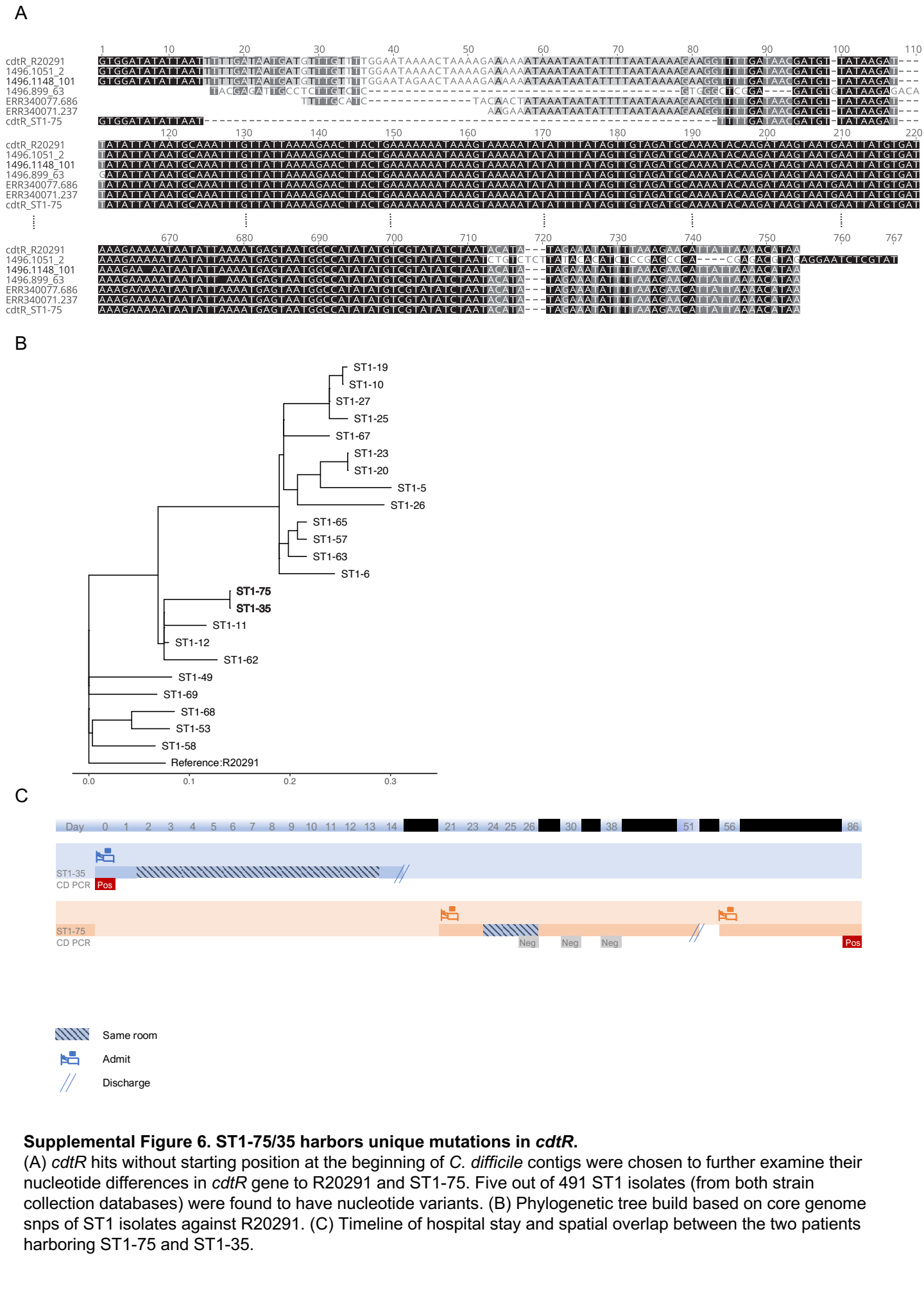
Supplemental Figure 3. Unique prophages identified in two avirulent *C. difficile* strains. (A) Schematic of prophages identified in avirulent strains. Each arrow represents a coding sequence whose product is indicated, when available. Numbers refer to the CDS when no function could be assigned. Gene products are colored by functional groups; yellow = packaging; red = head morphogenesis proteins; blue = tail morphogenesis proteins; cyan = lysis module; green = gene regulation and DNA replication; orange = lysogeny; grey = other or unknown function. Maps were generated with Benchling and finalized in Inkscape v1.2. (B-D) Wildtype C57BL/6 mice (n=4 per group) were treated with MNV and clindamycin as preciously described. Then, mice were inoculated with 200 *C. difficile* spores via oral gavage. (B) Fecal toxins measured by CHO cell rounding assay 1 day post infection. (C) Fecal colony-forming units measured by plating on selective agar 7 days post infection. (D) Fecal toxins measured by CHO cell rounding assay 7 days post infection.



Supplemental Figure 4. Binary toxin regulator *cdtR* does not impact *C. difficile* colonization in mice. (A) Schematic of *cdtR* mutants generated using R20291 *C. difficile* strain. (B-F) Wildtype C57BL/6 mice (n=3 to 5 per group) were treated with MNV and clindamycin as preciously described. Then, mice were inoculated with 200 *C. difficile* spores via oral gavage. Daily body weight and acute disease scores were monitored for 7 days post infection. (B) %Weight loss to baseline of mice infected with indicated strains. (C) Acute disease scores comprising weight loss, body temperature drop, diarrhea, morbidity of mice infected with indicated strains. (D, F) Fecal colony-forming units measured by plating on selective agar 1 day post infection (E) Fecal toxins measured by CHO cell rounding assay on indicated days. (G) Germ-free mice (n=4) orally administered with 200 spores of indicated *C. difficile* strains. Fecal colony-forming units measured by plating on selective agar 1 day post infection. Statistical significance was calculated by One-way ANOVA, * p <0.05.



Supplemental Figure 5. CdtR regulates PaLoc toxins production. (A-B) Wildtype C57BL/6 mice (n=3-5 per group) were treated with MNV and clindamycin as preciously described. Then, mice were inoculated with 200 *C. difficile* spores via oral gavage. (A) Fecal toxins measured by CHO cell rounding assay 7 days post infection for ST1-35. (B) Fecal toxins measured by CHO cell rounding assay 14 days post infection for ST1-75. (C-E) Germ-free mice (n=3-4) orally administered with 200 spores of indicated *C. difficile* strains and cecal contents were harvested at 24 hours post infection. (C) Cecal CFU measured by plating on selective agar. (D) Cecal toxin as in Figure 5A was normalized to CFU. (E) PaLoc and CdtLoc transcripts were measured by RT-qPCR. Transcripts were all normalized to the *adk* and fold change is relative to CdtRmut_6.1 for each of the genes. Statistical significance was calculated by Unpaired t-test, * $p < 0.05$, ** $p < 0.01$.



Supplemental Figure 6. ST1-75/35 harbors unique mutations in *cdtR*.
(A) *cdtR* hits without starting position at the beginning of *C. difficile* contigs were chosen to further examine their nucleotide differences in *cdtR* gene to R20291 and ST1-75. Five out of 491 ST1 isolates (from both strain collection databases) were found to have nucleotide variants. (B) Phylogenetic tree build based on core genome snps of ST1 isolates against R20291. (C) Timeline of hospital stay and spatial overlap between the two patients harboring ST1-75 and ST1-35.