Type I toxin-antitoxin systems contribute to mobile genetic elements maintenance in
*Clostridioides difficile* and can be used as a counter-selectable marker for chromosomal
manipulation

Johann Peltier¹,²,⁵, Audrey Hamiot¹,², Julian R. Garneau³, Pierre Boudry¹,²#, Anna Maikova¹,²,⁴,⁵,
Louis-Charles Fortier³, Bruno Dupuy¹,² and Olga Soutourina¹,²,⁵,*

1 Laboratoire Pathogenèse des Bactéries Anaérobies, Institut Pasteur, 75724 Paris Cedex 15, France
2 Université Paris Diderot, Sorbonne Paris Cité, 75724 Paris Cedex 15, France
3 Université de Sherbrooke, Faculty of medicine and health sciences, Department of microbiology
and infectious diseases, 3201 rue Jean Mignault, Sherbrooke, QC, J1E 4K8, Canada
4 Center for Data-Intensive Biomedicine and Biotechnology, Skolkovo Institute of Science and
Technology, Moscow, 143028, Russia
5 Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198,
Gif-sur-Yvette, France
6 Institut Universitaire de France (IUF)

* To whom correspondence should be addressed. Tel: +33 169826206; E-mail: olga.soutourina@i2bc.paris-saclay.fr
§ Present address: UMR UMET, INRA, CNRS, Univ. Lille 1, 59650 Villeneuve d’Ascq, France.
# Present address: Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud,
Université Paris-Saclay, 91198, Gif-sur-Yvette cedex, France

Keywords: toxin-antitoxin, small noncoding RNA, *cis*-antisense RNA, prophage stability

Running title: new type I TA systems in *C. difficile*
**ABSTRACT**

Toxin-antitoxin (TA) systems are widespread on mobile genetic elements as well as in bacterial chromosomes. According to the nature of the antitoxin and its mode of action for toxin inhibition, TA systems are subdivided into different types. The first type I TA modules were recently identified in the human enteropathogen *Clostridoides* (formerly *Clostridium*) *difficile*. In type I TA, synthesis of the toxin protein is prevented by the transcription of an antitoxin RNA during normal growth. Here, we report the characterization of five additional type I TA systems present within phiCD630-1 and phiCD630-2 prophage regions of *C. difficile* 630. Toxin genes encode 34 to 47 amino acid peptides and their ectopic expression in *C. difficile* induces growth arrest. Growth is restored when the antitoxin RNAs, transcribed from the opposite strand, are co-expressed together with the toxin genes. In addition, we show that type I TA modules located within the phiCD630-1 prophage contribute to its stability and mediate phiCD630-1 heritability. Type I TA systems were found to be widespread in genomes of *C. difficile* phages, further suggesting their functional importance. We have made use of a toxin gene from one of type I TA modules of *C. difficile* as a counter-selectable marker to generate an efficient mutagenesis tool for this bacterium. This tool enabled us to delete all identified toxin genes within the phiCD630-1 prophage, thus allowing investigation of the role of TA in prophage maintenance. Furthermore, we were able to delete the large 49 kb phiCD630-2 prophage region using this improved procedure.

**INTRODUCTION**

*Clostridoides difficile* is a medically important human enteropathogen that became a key public health concern over the last two decades in industrialized countries. This strictly anaerobic spore-forming Gram-positive bacterium is a major cause of antibiotic-associated nosocomial diarrhoea in adults. The main virulence factors of *C. difficile* are two toxins, TcdA and TcdB, produced by all toxigenic strains. A binary toxin CDT is also present in
some isolates, as well as additional factors that help colonization, like adhesins, pili, and flagella. Many questions remain unanswered regarding the success of this pathogen and its adaptation within the phage-rich gut environment.

*C. difficile* genome sequencing revealed the mosaic nature of its chromosome, which is composed of more than 10% of mobile genetic elements including integrated bacteriophages (prophages). Recent studies revealed a high prevalence of prophages in *C. difficile* genomes, each genome harbouring between one and up to five prophages, either integrated into the chromosome or maintained as stable extrachromosomal circular DNA elements. For example, the largely used laboratory strain 630 carries two homologous prophages, phiCD630-1 and phiCD630-2, while the epidemic ribotype 027 strain R20291 carries one prophage (phi-027). The importance of prophages in the evolution and virulence of many pathogenic bacteria has been clearly demonstrated. In *C. difficile*, all phages identified so far are temperate and can adopt a lysogenic lifecycle, and some of them have been shown to contribute to virulence-associated phenotypes. This includes modulation of toxin production and complex crosstalk between bacterial host and phage regulatory circuits. When integrated into *C. difficile* genomes, prophages are stably maintained and replicated along with the host chromosome. However, when they are excised, either spontaneously or following induction by antibiotics or the exposure to other stress conditions, prophages can sometimes be lost during cell division and segregation. The rate of spontaneous phage loss under natural conditions has been estimated to range between $10^{-5}$ for phage P1 to $<10^{-6}$ for phage lambda.

TA modules are widespread in bacteria and archaea. These loci comprise two genes encoding a stable toxin and an unstable antitoxin. Overexpression of the toxin has either bactericidal or bacteriostatic effects on the host cell while the antitoxin is able to neutralize the toxin action. For all identified TA modules, the toxin is always a protein. The RNA or protein...
nature and the mode of action of the antitoxin led to the classification of TA modules into six types \(^1\). In type I systems, the antitoxin is a small antisense RNA targeting toxin mRNA for degradation and/or inhibition of translation, while in type III systems, the antitoxin RNA binds directly to the toxin protein for neutralization \(^12, 13\). For other TA types, both the toxin and the antitoxin are proteins. In most studied type II TA systems, the proteinaceous antitoxin forms a complex with its cognate toxin leading to toxin inactivation \(^14\). Major functions of TA modules include plasmid maintenance, abortive phage infection and persister cell formation \(^15, 16, 17, 18, 19, 20\). TA loci are commonly found on mobile genetic elements, in particular plasmids in which they were initially discovered and extensively studied. However, the roles of chromosomally-encoded TA modules, including those within prophage genomes, remain largely unexplored. Likewise, the individual contribution from each of the chromosomal TA modules in bacteria remains to be investigated.

We recently reported the identification of the first type I TA systems associated with CRISPR arrays in \textit{C. difficile} genomes \(^21\). The co-localization and co-regulation by the general stress response Sigma B factor and biofilm-related factors suggested a possible genomic link between these cell dormancy and adaptive immunity systems. Interestingly, two of these functional type I TA pairs are located within the homologous phiCD630-1 and phiCD630-2 prophages in \textit{C. difficile} strain 630. In the present work, we describe the identification of additional type I TA modules highly conserved within \textit{C. difficile} prophages and provide experimental evidence of their contribution to prophage maintenance and stability.

The unique properties of type I TA systems offer interesting possibilities for biotechnological and therapeutic applications. We demonstrate in the present work that inducible toxicity caused by type I toxins largely improve the efficiency of allele exchange genome editing procedures by promoting the elimination of plasmid-bearing cells.
Altogether, our data provide important insights into the function and possible applications of type I TA systems in *C. difficile*.

**RESULTS**

**Identification of novel type I TA pairs in *C. difficile***

Multiple TA modules have been discovered in bacterial chromosomes including prophage regions. In *C. difficile*, we have recently identified several type I TA pairs adjacent to CRISPR arrays, two of them being located inside the phiCD630-1 and phiCD630-2 prophages of the strain 630 (*CD0956.2*-RCd10 and *CD2907.1*-RCd9, respectively). To determine whether other type I TA modules might be present within phiCD630-1, we performed a bioinformatics analysis on the phiCD630-1 sequence. Due to the small size of the toxin-encoding genes, standard methods of open reading frame (ORF) detection and gene annotation can hinder the identification of all toxin homologs. Moreover, prophages are characterized by a very high gene density, which can impede such detection of small and often overlapping coding regions. We therefore used the tBlastn program (protein query to search against translated genomes) using a previously identified type I toxin *CD0956.2* as a query, as it can overcome annotation and ORF detection defects. We identified gene *CD0977.1* and two other novel putative genes, unannotated on the genome, that we named *CD0904.1* and *CD0956.3*. These genes code for small proteins of 47, 35 and 34 amino acids, respectively (Fig. 1A). Prophages phiCD630-1 and phiCD630-2 share a large region of homology with almost identical sequences, which include a duplication of *CD0977.1* and *CD0956.3* (Fig. 1B). In contrast, *CD0904.1* is unique to phiCD630-1 and no other toxin gene homolog could be identified within phiCD630-2. Transcript reads were detected in regions of these putative genes by RNA-seq (Fig. S1A and S1B) and a consensus RBS sequence (AGGAGG) was present 7-8 nucleotides upstream of the respective ATG start codons,
suggesting that the corresponding proteins might be produced. In addition, all three putative proteins carried a hydrophobic N-terminal region and a positively charged tail, which are characteristic features of type I toxins (Fig. 1A). Analysis of our previous TSS mapping data and sequence alignments (Fig. S1C) suggested the presence of potential antisense RNAs of these toxin-encoding genes with the presence of TSS associated with Sigma A- and Sigma B-dependent promoter elements in both sense and antisense directions (Fig. S1). Antitoxins of CD0977.1, CD0904.1 and CD0956.3, located on phiCD630-1, were hereafter named RCd11, RCd13 and RCd14, respectively, and those of CD2889 and CD2907.2, found in phiCD630-2, were named RCd12 and RCd15.

To determine whether these novel potential TA pairs are functional, pRPF185-derivatives with anhydrotetracycline (ATc)-inducible Ptet promoter were constructed to overexpress CD0904.1, CD0956.3 and CD0977.1 toxin genes (pT) or toxin-antitoxin modules (pTA) in C. difficile 630Δerm. Antisense RNAs are expressed from their own promoter in pTA. Growth of 630Δerm carrying the different pT and pTA vectors on BHI plates was indistinguishable in the absence of ATc inducer (Fig. 1C). In contrast, growth of the 630Δerm/pT strains was completely inhibited when ATc was present in the medium, while strains 630Δerm/pTA showed a reversion of the growth defect. These results demonstrate that CD0904.1, CD0956.3 and CD0977.1 encode potent toxins and are associated with antisense RNAs that function as antitoxins.

**Detailed characterization of the CD0977.1-RCd11 TA pair**

Intriguingly, predicted riboswitches responding to the c-di-GMP signalling molecule, cdi1_4 and cdi1_5, precede RCd11 and RCd12 antisense RNAs. We therefore sought to further characterize the CD0977.1-RCd11 TA pair. In agreement with the data above, addition of ATc to liquid cultures in exponential growth phase led to an immediate growth arrest of
strain 630Δerm/pT, unlike 630Δerm/p (Fig. S2A). In addition, the growth arrest was accompanied by a drop of colony-forming units (CFUs) (Fig. S2B). Similarly to previous observations with other C. difficile type I TA modules \(^{21}\), the analysis of liquid cultures by light microscopy showed that toxin overexpression was accompanied by an increase in cell length in about 10% of the cells (Fig. S2C). Their length was above the mean length value of 630Δerm/p control strain with two standard deviations (10.5 µm). The co-expression of the entire TA module led to the partial reversion of this phenotype.

Using Northern blotting, we detected both toxin and antitoxin transcripts in the 630Δerm/p, 630Δerm/pT (CD0977.1) and 630Δerm/pTA (CD0977.1-RCd11) strains (Fig. 2A). In the absence of ATc inducer, a major transcript of about 300 nt was detected in all three strains with a CD0977.1-specific probe. When using a RCd11-specific probe, transcripts of about 150, 300 and 400 nt were observed. Under inducing conditions, a reverse correlation between the relative toxin and antitoxin transcript abundance was noticed. To determine the impact of c-di-GMP on the antitoxin transcripts, we elevated c-di-GMP intracellular levels in the 630Δerm wild type strain by expressing the diguanylate cyclase gene dccA from a plasmid (pdccA) (Fig. 2B), as previously reported \(^{22}\). A c-di-GMP-regulated read-through transcript of about 400 nt, as well as a terminated transcript of about 140 nt were detected in this strain by Northern blotting using a riboswitch-specific probe. In contrast, abundance of toxin transcripts was not affected by fluctuations of c-di-GMP levels.

We then mapped the transcriptional start (TSS) and termination sites for the genes of the potential RCd11/RCd12-CD0977.1/CD2889 TA modules by 5'/3'RACE analysis (Fig. S3, Table S3). The results obtained agreed generally well with the transcript lengths deduced from TSS mapping, RNA-seq and Northern blot. Taken together, these data suggest the presence of two tandem TSS for RCd11, i.e. \(P_1\) associated with c-di-GMP-dependent riboswitch and \(P_2\) located downstream from the riboswitch (Fig. S1).
It is in the nature of type I antitoxins to be short-lived in contrast to the stable toxin mRNA. To determine the half-lives of toxin and antitoxin RNAs of the CD0977.1-RCd11 module, C. difficile strains were grown in TY medium until late-exponential phase and rifampicin was added to block transcription. Samples were taken at different time points after rifampicin addition for total RNA extraction and Northern blot analysis with toxin and antitoxin-specific probes. In a control strain 630Δerm/p carrying an empty vector, the half-life of major 150-nt antitoxin transcript for RCd11 was estimated to be about 8 min while the half-life of CD0977.1 toxin mRNA was estimated to about 89 min (Fig. 2C). Interestingly, depletion of the RNA chaperone protein Hfq resulted in a moderate destabilization of CD0977.1 toxin mRNA and antitoxin RCd11 RNA with the half-life of 64 min and 4 min, respectively (Fig. S4). By contrast, the stable CD0977.1 toxin mRNA was further stabilized to over 120 min half-life in the strains depleted for RNase III, RNase J and RNase Y. For antitoxin RCd11 RNA, we also observed a stabilization in strain depleted for RNase Y (Fig. S4) suggesting that this ribonuclease could contribute to antitoxin RNA degradation.

To confirm the protein nature of CD0977.1 and assess its subcellular localization, we constructed a derivative of CD0977.1 with an HA tag fused to the C-terminus of CD0977.1 expressed from a plasmid under the control of the inducible P_{tet} promoter (pT-HA). C. difficile strain carrying pT-HA was grown to mid-exponential phase, induced with ATc for 90 min, and whole cell extracts were prepared. Induction of CD0977.1 expression immediately stopped the growth, as revealed by OD_{600} measurements (data not shown), suggesting that the HA-tag does not interfere with toxin activity. HA-tagged CD0977.1 was detectable by Western blotting with anti-HA antibodies (Fig. 2D). No band was observed in a whole cell extract of a control strain producing untagged CD0977.1 protein. The distribution of HA-tagged CD0977.1 within supernatant, cell wall, membrane and cytosolic compartments was
studied (Fig. 2D). HA-tagged CD0977.1 was only detected in the membrane fraction, indicating the association of CD0977.1 with the cell membrane of *C. difficile*.

**The antitoxin transcript controlled by cdi1_5 riboswitch is dispensable for efficient toxin inactivation**

To get further insights into the function of abundant short (transcribed from *P_2*) and less abundant long RCd11 antitoxin transcripts (transcribed from *P_1*), we generated new plasmid constructs that allowed the inducible expression of the *CD0977.1* toxin gene under the control of the *P_{tet}* promoter and the expression of different forms of the RCd11 antisense RNA (Fig. 3A). The first construct, yielding pDIA6816, lacked the cdi1_5 riboswitch and its associated promoter (*P_1*) but retained the *P_2* promoter of the antitoxin. On the opposite, the second construct, yielding pDIA6817, retained the *P_1* promoter and the associated riboswitch but had a disrupted *P_2* promoter. The construct in which both promoters of RCd11 were present (pDIA6785) served as a positive control for the assay, and the RCd11 promoterless construct (pDIA6335) was included as a negative control. All plasmids were introduced into *C. difficile 630Δerm* and the corresponding strains were grown on BHI plates supplemented with 10 and 100 ng/ml ATc to induce *CD0977.1* toxin expression. Growth of the strain carrying pDIA6816 was similar to that observed for the control strain carrying pDIA6785 in the presence of 10 ng/ml ATc and was slightly defective in the presence of 100 ng/ml ATc (Fig. 3A and Fig. S5A). In contrast, the strain carrying pDIA6817 did not grow in the presence of 10 or 100 ng/ml ATc, similarly to the negative control strain. Similar results were obtained when the strains were grown in an automatic plate reader for 20 h in liquid medium in the presence of 5 ng/ml ATc (Fig 3A). Interestingly, induction of toxin expression on BHI plate with a lower dose of ATc (5 ng/ml) led to a partial reversion of the growth defect of the strain carrying pDIA6816, unlike the negative control strain (Fig. S5A). These results suggest
that the short antitoxin transcript driven by promoter $P_2$ is crucial for the efficient inactivation of the toxin, while the longer antitoxin transcript directed by $P_1$ is dispensable.

**RCd12 counteracts toxic activity of non-cognate CD0977.1 toxin**

Nucleotide sequences of short RCd11, lying within phiCD630-1 and short RCd12, lying within phiCD630-2, are almost identical with only 3 mismatches located near the 3’ end, in the region overlapping with the toxin transcript (Fig. S6A and S7). We therefore wondered whether RCd12 could cross-react with the transcript of the non-cognate toxin CD0977.1. To answer this question, we generated new constructs in which CD0977.1 toxin gene under the control of the $P_{tet}$ promoter and different antitoxins with their own promoter were co-expressed from the same plasmid but from distant locations (Fig. 3B). As anticipated, expression of RCd11 in trans counteracted the toxicity associated with the expression of the cognate toxin both on plate and in liquid culture (Fig. 3B and Fig. S5B). Replacement of RCd11 with RCd12 led to the same result. By contrast, expression of the more divergent RCd10 (the antitoxin of CD0956.2 toxin from a previously characterized TA module lying within phiCD630-1$^{21}$) in trans failed to revert the growth defect induced by CD0977.1 expression (Fig. S5B and 3B). These data indicate that antitoxins act in a highly specific manner to repress their cognate toxins, not only when they are expressed from the native convergent TA configuration, but also when expressed in trans. However, the specificity of interaction is permissive for at least 3 mismatches allowing RCd12 expressed from phiCD630-2 to efficiently prevent CD0977.1 toxin production from phiCD630-1.

**TA modules confer plasmid stabilization**

TA systems have been initially discovered on plasmids where they confer maintenance of the genetic element $^{17}$. Plasmid loss results in a rapid decrease in the levels of the unstable antitoxin, which allows the stable toxin to inhibit cell growth. To test whether the TA
modules located on phiCD630-1 could contribute to plasmid maintenance, we assessed the stability of plasmids in which the toxin genes and their cognate antitoxin were co-expressed under the control of their respective native promoter in C. difficile 630Δerm. Bacteria harbouring the empty vector pMTL84121 were used as a control. After 7 passages in TY broth in the absence of antibiotic pressure, pMTL84121 was maintained by 1.0% (+/-0.4%) of the bacterial population (Fig. 4). In contrast, plasmids expressing TA pairs were still present in 22.3 (+/- 5.4%) to 63.6 % (+/- 14.3%) of total cells. These results indicate that the four TA pairs can confer plasmid maintenance.

**Deletion of phiCD630-1 toxin genes in C. difficile 630Δerm**

In order to determine whether the TA modules contribute to phiCD630-1 stability, we undertook the construction of deletion mutants of toxin genes in C. difficile 630Δerm. For this purpose, we constructed a new Allele-Coupled Exchange (ACE) vector, derived from pMTL-SC7315, a codA-based “pseudosuicide” plasmid that replicates at a rate lower than that of the host chromosome 23. The codA cassette was here replaced with the CD2517.1 toxin gene placed under the control of the Ptet inducible promoter (Fig. S8A). We demonstrated the functionality of RCd8-CD2517.1 type I TA module in C. difficile in a previous study 21. In our new vector, designated pMSR, the inducible toxic expression of CD2517.1 is used as a counter-selection marker to screen for plasmid excision and loss (see Materials and Methods) and this greatly facilitates the isolation of C. difficile deletion mutant generated by double cross-over allele exchange (Fig. S8C and D). We also constructed a second vector, pMSR0, for allele exchange in C. difficile ribotype 027 strains and other ribotypes (see Materials and Methods and Fig S8B). By using this new tool, we first deleted the 49.3 kb phiCD630-2 locus to prevent any interfering cross-talk with phiCD630-1 (Fig. S9). A multiple deletion mutant of toxin-encoding genes CD0904.1, CD0956.2, CD0956.3 and CD0977.1 (ΔT4) was then generated in the ΔphiCD630-2 background (Table S1).
TA systems are involved in maintenance of phiCD630-1 in the host cells

Because the loss of an integrated phage from cells first requires its excision from the host genome, we sought to determine whether spontaneous excision of phiCD630-1 from chromosomal DNA occurred. To do so, we performed a PCR on genomic DNA from C. difficile 630Δerm ΔphiCD630-2 with primers flanking the attL and attR sites of phiCD630-1 (Fig. S10A). A PCR product with a size of 88 bp corresponding to the excised prophage was detected (Fig. S10B) and DNA sequencing of this amplicon confirmed the complete removal of phiCD630-1 from the host chromosome. A second PCR-based assay showed that the excised prophage (PCR product of 117 bp) was present as an extrachromosomal circular form in the host cell (Fig. S10A and B). We could also deduce the attB, attP, attL and attR sites from the sequencing of the PCR products (Fig. S10C). However, the frequency of phiCD630-1 excision, measured by quantitative PCR (qPCR), was very low (~ 0.015 %) (Fig. S11). To screen for the presence/absence of phiCD630-1 in the host cells, we introduced, using our new ACE vector, the ermB gene placed under the control of the strong thl promoter of Clostridium acetobutylicum as previously described 24, into an innocuous location (between CD0946.1 and CD0947) of phiCD630-1 and phiCD630-1-ΔT4 (Fig. 5A). Starting from the overnight cultures, the ΔphiCD630-2 phiCD630-1::erm and ΔphiCD630-2 phiCD630-1-ΔT4::erm strains were subcultured four times in fresh medium and cells were screened for erythromycin resistance by plating onto non-selective and erythromycin-containing agar plates. Nearly 100% of cells from both strains were found to still be resistant to erythromycin in these conditions, indicating that they had retained the prophage (Fig S10D).

In an attempt to artificially increase the excision rate of phiCD630-1, we ectopically expressed the putative excisionase CD0912 from the inducible Ptet promoter, yielding pDIA6867. CD0912, identified in a bioinformatics search, is a 109 amino acid protein with a predicted DNA-binding domain similar to the HTH-17 superfamily and the excisionase (Xis)
family. Induction of CD0912 expression with 10 ng/ml ATc in C. difficile 630Δerm resulted in a high excision rate of phiCD630-1 (~ 70 %), indicating that CD0912 functions as an excisionase for phiCD630-1 (Fig. S11). Expression of CD0912 in strains ΔphiCD630-2 phiCD630-1::erm and ΔphiCD630-2 phiCD630-1ΔT4::erm caused excision at a similar rate, suggesting that TA systems do not affect phiCD630-1 excision (Fig. S11).

Strains ΔphiCD630-2 phiCD630-1::erm and ΔphiCD630-2 phiCD630-1-ΔT4::erm carrying pDIA6867 were then inoculated at an initial OD₆₀₀ of 0.005 in TY supplemented with 7.5 μg/ml Tm and 10 ng/ml ATc. After 24 hrs of incubation at 37°C, measurement of the OD₆₀₀ revealed a dramatic growth defect of ΔphiCD630-2 phiCD630-1::erm compared to the ΔphiCD630-2 phiCD630-1-ΔT4::erm (Fig. 5B). In addition, plating of the cells bearing pDIA6867 on non-selective and erythromycin-containing agar plates revealed that phiCD630-1::erm was still present in more than 90% of the total population while phiCD630-1-ΔT4::erm remained in less than 10 % of the cells (Fig. 5C). These results thus show that TA modules are important for phiCD630-1 maintenance after its excision and highlight the impact of the toxin expression on the cell growth upon the loss of prophage. Together, these data demonstrate that TA modules mediate phiCD630-1 heritability.

Type I TA are prevalent in C. difficile phage genomes

Since we identified additional toxin variants of type I TA systems after careful inspection of phiCD630-1 full genome, we decided to re-scan for possible ORFs in every available phage genomes of C. difficile using the permissive algorithm of the NCBI ORFfinder software. ORFs with minimal length of 60 nucleotides as well as nested ORFs were detected. A blastP search against the corresponding proteins allowed the identification of toxin homologs in all C. difficile prophage genomes (functional phages) (Fig. 6). Moreover, toxin sequence alignments revealed the high conservation of the hydrophobic N-terminal region, as well as
the lysine-rich, positively charged region at the C-terminus. Hence, these data suggest the functionality of the toxins and reinforce their proposed role for phage maintenance and preservation.

Despite these conserved regions, alignment of toxins also revealed small variations among sequences. We therefore sought to explore the possible relationship between phage phylogeny and the observed toxin variants. A whole genome comparison of all phages included in this study was performed to create phylogenetic groups (phiCD119-like viruses, phiCD38-2-like viruses and phiMMP04-like viruses), as previously described. A clear link between phage groups and specific toxin variants could be established, suggesting an independent acquisition of TA systems in different groups of phages (Fig. 6 and Fig. S12).

Interestingly, an extended search outside C. difficile phages revealed the presence of other toxin homologs inside plasmids of C. difficile and Paeniclostridium sordellii, a closely related species (Fig. S13). These findings imply that C. difficile phages could recombine with plasmids to exchange genetic material, as already proposed for E. coli phages.

DISCUSSION

In this study, we identified and characterized novel functional type I TA modules in C. difficile 630 prophages. Although these modules share characteristic features of TA systems, i.e. (i) toxic proteins are membrane-associated and have a positively charged tail, (ii) the toxin mRNA is much more stable than the antitoxin RNA, (iii) artificial expression of the toxin genes inhibits bacterial growth unless their cognate antitoxin RNA is co-expressed; they do not present sequence homology with other TA modules identified in other bacteria.

RCd11-CD0977.1 and Rcd12-CD2889 TA pairs are duplications respectively located within the homologous phiCD630-1 and phiCD630-2 prophages. Two alternative RCd11 (and Rcd12) antitoxin transcripts were identified and the longer, less abundant form is associated with the cdi1_4 (and cdi1_5) c-di-GMP-responsive riboswitch. C-di-GMP is a
second messenger in bacterial systems and a key signal in the control of critical lifestyle choices, such as the transition between planktonic and biofilm growth. C-di-GMP has been found to regulate important functions in *C. difficile*, including motility, production of type IV pili, cell aggregation and biofilm formation, through control of gene expression by C-di-GMP-dependent riboswitches. Sixteen predicted C-di-GMP sensing riboswitches are encoded in the *C. difficile* 630 genome and the regulatory function of five of them has been investigated so far. Interestingly, cdI1_4 and cdI1_5 riboswitches were recently reported to be insensitive to an elevation of the C-di-GMP levels, and only transcript reads corresponding to the riboswitch sequences themselves (the short form) were detected, and no read-through seemed to occur. In contrast, our RACE-PCR and Northern-blot analysis indicated the presence of a transcript downstream from these riboswitches. Moreover, our data suggested that cdI1_4 and cdI1_5 are functional riboswitches responding to C-di-GMP since the abundance of the downstream transcripts was significantly reduced in the presence of high levels of C-di-GMP. Despite these data, we found that the shorter and more abundant RCd11 transcript alone was sufficient to ensure complete CD0977.1 toxin inactivation under our conditions. In contrast, the longer antitoxin transcript associated with the cdI1_4 riboswitch could only counteract the CD0977.1 toxicity when the toxin gene was expressed at low levels. This suggests that this antitoxin transcript might be involved in the tight regulation of CD0977.1 production and might be crucial to prevent toxin translation under conditions where expression levels of the toxin gene would be slightly higher than those of the short antitoxin transcript. The C-di-GMP levels would then be critical in this regulation since elevated levels would result in a decreased abundance of the short RCd11 transcript and consequently in growth inhibition.

In this work, we detected a natural background excision of the phiCD630-1 prophage and we identified the phage excisionase gene, *CD0912*. Expression of *CD0912* from a
plasmid promoted high levels of prophage excision from the host chromosome. While phiCD630-1 and phiCD630-2 of *C. difficile* 630 share a large region of duplicated sequence, it is worth noting that *CD0912* is located in the variable region and has no homolog in phiCD630-2. Interestingly, no obvious putative excisionase-encoding gene could be identified in phiCD630-2 although natural excision of this prophage could also be detected in the course of our experiments. Moreover, expression of *CD0912* had no impact on the excision rate of phiCD630-2. This suggests that phiCD630-2 might encode an atypical, yet to be identified excisionase. TA systems have been shown to play three important biological functions, i.e., stabilization of mobile genetic elements (post-segregational killing), abortive phage infection and persister cell formation. Prophage maintenance is among the suggested function of TA including a recent example of type II TA system that stabilizes prophage in *Shewanella oneidensis* and another type II TA system promoting the maintenance of an integrative conjugative element in *Vibrio cholerae*. This physiological function in prophage stabilization was also suggested for type I TA modules but had never been experimentally demonstrated prior to this study. Prophage excision upon expression of *CD0912* made phiCD630-1 prone to be lost by the host cells and we could thus show that type I TA systems are important to maintain the episomal form of the phage into the host cell. Indeed, the frequency of phiCD630-1ΔT4 loss was higher than that of wildtype phiCD630-1. Moreover, maintenance of phiCD630-1 in the cells expressing the excisionase gene was associated with a strong growth defect, which can be attributed to the post-segregational killing mechanism. We can speculate that the unstable antitoxin is degraded in daughter cells where the phage has been lost and the stable toxic protein prevents the growth of the new cell. Expression of the excisionase gene in cells carrying the prophage devoid of the toxin genes also resulted in a growth defect, although to a lesser extent. This suggests that a supplementary TA system might be present in phiCD630-1 or that the excisionase has an additional function affecting...
the cell growth and that remains to be uncovered. Several experimental conditions were tested in this work to induce the loss of the phage from the cells. Surprisingly, four passages of the strain carrying phiCD630-1 with the intact toxin genes in rich medium with constant expression of the excisionase gene resulted in approximately 99% of loss of this prophage. This is likely due to a progressive enrichment of the cell population surviving the loss of the phage since the growth rate of this population is higher than that of the population bearing the phage. In any case, these data suggest that the identification and the overexpression of the phage excisionase-encoding genes could provide an easy and efficient way to cure *C. difficile* strains from their prophages.

Toxins of type I TA systems are relatively small proteins, and this is probably one of the reasons why they have remained uncharacterized and unexplored in *C. difficile* and in other organisms. In this study, we have come to realize that standard methods of annotation are unable to detect all toxin homologs present in prophage genomes. Novel toxin homologs, previously unannotated, could thus be detected inside plasmids of *C. difficile* and *P. sordellii*. It has been proposed that phages could recombine with plasmids during infection of the same or different bacterial species to exchange genetic material. It is therefore tempting to speculate that this TA system could originate from *P. sordellii*, or at least that it has the ability to disseminate, through horizontal gene transfer involving conjugation and recombination, from one species to another. Intriguingly, it was previously noticed that a 1.9-kb region could have been transferred from the plasmid of a *C. difficile* 630 strain to the phiCD38-2 prophage. It was suggested that this recombination event had led to the acquisition of parA, a gene assumed to help the newly created chimeric phage to autonomously replicate and segregate as a circular plasmid. Our analysis of TA systems in phiCD38-2 shows that this 1.9-kb region also carried a TA (gp33) that presumably contributes to the phage maintenance and stability. It is interesting to observe that TA
encoding regions can relocate from one mobile genetic element to another in this fashion, and that genes in proximity to the TA being transferred (i.e. parA gene) have more chances to become fixed in the newly integrated DNA. In the latter case, the region transferred seems to provide two complementary and beneficial features to the phage, i.e. the capacity to segregate successfully to the daughter cell, and the death of the cells upon curing if the phage has not been sequestered in both dividing cells. However, since TA systems behave as selfish elements that promote their propagation within bacterial genomes at the expense of their host 35, 36, they are likely to be maintained and observed after their transfer by recombination events, even if they bring no selective advantage.

Besides their biological functions, TA modules are also versatile tools for a multitude of purposes in basic research and biotechnology 37. For example, the MazF toxin-encoding gene from E. coli is used as a counter-selection marker for chromosomal manipulation in Bacillus subtilis and C. acetobutylicum 38, 39. In this study, we engineered an inducible counter-selection marker based on the C. difficile CD2517.1 toxin gene. When CD2517.1, cloned under the control of an inducible promoter on a plasmid, is expressed in C. difficile upon addition of the inducer, bacterial growth is immediately interrupted 21. Taking advantage of this feature, we generated a novel vector for allele exchange in C. difficile 630 and another vector for use in C. difficile ribotype 027 strains and other ribotypes strains. It should be noted that expression of the RCd8 antitoxin from the pMSR0 vector was required to counteract the basal expression of CD2517.1 toxin gene due to the Ptet leakiness. In contrast, expression of the RCd8 antitoxin from the pMSR vector was not required since the CD2517.1-RCd8 TA module is naturally present within the chromosome of C. difficile 630. Native expression of RCd8 was therefore sufficient to prevent CD2517.1 production from the plasmid. The system developed by Cartman et al., which uses the codA gene coding for cytosine deaminase as a counter-selection marker for allelic exchange mutations was
appealing at first since no genomic manipulations were needed to apply the counter-selection. However, codA-based counter-selection was somewhat ineffective in our hands and false-positive counter-selected colonies with the plasmid still integrated into the chromosome were repeatedly found. This was reported by the authors as the consequence of loss-of-function mutations in either codA or upp genes leading to the bypass of the counter-selection. Our newly developed system proved to be much more efficient than all the others we have tested so far, and we did not observe any false-positive clones so far. We successfully used this system to construct multiple mutants in various C. difficile strains, including the ΔT4 mutant lacking the four toxin genes reported in this study, deletion of a large chromosomal region of 50 kb corresponding to the phiCD630-2 prophage, and gene insertion into the bacterial chromosome (ermB gene). We therefore expect these new vectors to become invaluable genetic tools that will foster research in C. difficile.

MATERIALS AND METHODS

Plasmids, bacterial strains construction and growth conditions

C. difficile and Escherichia coli strains and plasmids used in this study are presented in Table S1. C. difficile strains were grown anaerobically (5 % H₂, 5 % CO₂, and 90 % N₂) in TY or Brain Heart Infusion (BHI, Difco) media in an anaerobic chamber (Jacomex). When necessary, cefoxitin (Cfx; 25 μg/ml), cycloserine (Cs; 250 μg/ml) and thiamphenicol (Tm; 7.5 μg/ml) were added to C. difficile cultures. E. coli strains were grown in LB broth, and when needed, ampicillin (100 μg/ml) or chloramphenicol (15 μg/ml) was added to the culture medium. The non-antibiotic analogue anhydrotetracycline (ATc) was used for induction of the P_{tet} promoter of pRPF185 vector derivatives in C. difficile. Strains carrying pRPF185 derivatives were generally grown in TY medium in the presence of 250 ng/ml ATc and 7.5 μg/ml Tm for 7 h, unless stated otherwise. Growth curves were obtained using a GloMax plate reader (Promega).
All primers used in this study are listed in Table S2. For expression of TA genes under the control of their native promoters, the CD0977.1-RCd11, CD0956.2-RCd10, CD0904.1-RCd13, CD0956.3-RCd14 TA modules and their associated promoters were amplified by PCR and cloned into BamHI and XhoI sites of pMTL84121. For inducible expression of C. difficile genes, we used the pDIA6103 derivative of pRPF185 vector expression system lacking the gusA gene. To assess the activity of the putative toxin CD0977.1 and the functionality of its cognate antitoxin, the CD0977.1 gene (-66 to +222 relative to the translational start site) and the RCd11-CD0977.1 region (-220 to +608 relative to the translational start site) were amplified by PCR and cloned into StuI and BamHI sites of pDIA6103 vector under the control of the ATc-inducible P_{tet} promoter; yielding pDIA6335 and pDIA6337. All other constructs described below were performed with NEBuilder HiFi DNA Assembly (NEB). For this purpose, the pDIA6103 vector was linearized by inverse PCR at the MCS for cloning of genes under the control of the ATc-inducible P_{tet} promoter. To assess the toxicity of CD0904.1 and CD0956.3 and the functionality of their cognate antitoxin, the CD0904.1 (-43 to +109 relative to the translational start site) and CD0956.3 genes (-35 to +108 relative to the translational start site) as well as the RCd13-CD0904.1 (-43 to +230 relative to the translational start site) and the RCd14-CD0956.3 regions (-35 to +248 relative to the translational start site) were amplified by PCR and cloned into the linearized pDIA6103 vector, yielding pDIA6866, pDIA7030, pDIA6934 and pDIA7031, respectively. For inducible toxin expression and co-expression of its cognate antitoxin, the RCd11-CD0977.1 TA region with the two RCd11 promoters (-66 to +608 relative to the translational start site of CD0977.1) or the RCd11-CD0977.1 TA region with the P_{2} promoter only (-66 to +306 relative to the translational start site of CD0977.1) were amplified by PCR and cloned into the linearized pDIA6103 vector, yielding pDIA6785 and pDIA6816, respectively. For inducible toxin expression and co-expression of its cognate antitoxin with
the promoter $P_1$ only, the identified transcriptional start site (TSS) G of $P_2$ promoter was replaced with a T and the -10 box TATAAT was replaced with TTTTTT, using an inverse PCR approach, to disrupt the $P_2$ promoter; yielding pDIA6817. To investigate the action of cognate (RCd11) and non-cognate (RCd10 and RCd12) antitoxins on toxin CD0977.1 when co-expressed in trans (from a site distant from the vector MCS), we used an inverse PCR approach to construct different plasmids on the basis of pDIA6335. RCd11 (-158 to +386 relative to the first TSS), RCd12 (-158 to +386 relative to the TSS) and RCd10 (-167 to +282 relative to the TSS) with their respective promoter were amplified by PCR and cloned into pDIA6335 linearized by inverse PCR at a site distant from the vector MCS, yielding pDIA6791, pDIA6792 and pDIA6793, respectively. For subcellular localization of toxin CD0977.1, we used an inverse PCR approach to construct CD0977.1-HA-expressing plasmids (pDIA6622) on the basis of pDIA6335 with primers designed to introduce the hemagglutinin HA-tag sequence at the C-terminal part of toxin coding regions, directly upstream of the stop codon (Table S2). For expression of the CD0912 excisionase, the CD0912 gene (-37 to +333 relative to the translational start site) was amplified by PCR and cloned into the linearized pDIA6103 vector, yielding pDIA6867.

The resulting pRPF185 derivative plasmids were transformed into the E. coli HB101 (RP4) and subsequently mated with the appropriate C. difficile strains (Table S1). C. difficile transconjugants were selected by sub-culturing on BHI agar containing Tm (15 µg/ml), Cfx (25 µg/ml) and Cs (250 µg/ml).

**Mutagenesis approach and mutant construction**

To improve the efficiency of the allele exchange mutagenesis in C. difficile, we made use of the inducible toxicity of the CD2517.1 type I toxin that we previously reported. To construct the pMSR vector, used for allele exchange in C. difficile 630Δerm, the codA gene
was removed from the “pseudosuicide” vector pMTL-SC7315 by inverse PCR and replaced by a 1169 bp fragment comprising the entire \( P_{\text{tet}} \) promoter system and the downstream \( CD2517.1 \) toxin gene. This fragment was amplified from pDIA6319 plasmid and the purified PCR product was cloned into the linearized plasmid. In parallel, the pMSR0 vector, for allele exchange in \( C. \text{ difficile} \) ribotype 027 strains and other ribotypes, was constructed by removing the \( \text{codA} \) gene from the vector pMTL-SC7215 by inverse PCR and replacing it with the \( CD2517.1\)-RCd8 TA region with \( CD2517.1 \) under the control of the \( P_{\text{tet}} \) promoter, as described above, and RCd8 under the control of its own promoter. For deletions, allele exchange cassettes were designed to have between 800 and 1050 bp of homology to the chromosomal sequence in both up- and downstream locations of the sequence to be altered. The homology arms were amplified by PCR from \( C. \text{ difficile} \) strain 630 genomic DNA (Table S1) and purified PCR products were directly cloned into the PmeI site of pMSR vector using NEBuilder HiFi DNA Assembly. To insert \( P_{\text{ahi}}\text{-ermB} \) into the phiCD630-1 prophage, within the intergenic region between \( CD0946.1 \) and \( CD0947 \) genes, homology arms (~900 bp up- and downstream of the insertion site) were amplified by PCR from strain 630 genomic DNA (Table S1). The \( P_{\text{ahi}}\text{-ermB} \) cassette was amplified from the Clostron mutant \( cwp19 \). Purified PCR products were all assembled and cloned together into the PmeI site of pMSR vector using NEBuilder HiFi DNA Assembly. All pMSR-derived plasmids were initially transformed into \( E. \text{ coli} \) strain NEB10β and all inserts were verified by sequencing. Plasmids were then transformed into \( E. \text{ coli} \) HB101 (RP4) and transferred by conjugation into the appropriate \( C. \text{ difficile} \) strains. The adopted protocol for allele exchange was similar to that used for the \( \text{codA} \)-mediated allele exchange except that counter-selection was based on the inducible expression of the \( CD2517.1 \) toxin gene. Transconjugants were selected on BHI supplemented with Cs, Cfx and Tm, and then restreaked onto fresh BHI plates containing Tm. After 24 h, faster-growing single-crossover
integrants formed visibly larger colonies. One such large colony was restreaked once or twice
on BHI Tm plate to ensure purity of the single crossover integrant. Purified colonies were
then restreaked onto BHI plates containing 100 ng/ml ATc inducer to select for cells in which
the plasmid had been excised and lost. In the presence of ATc, cells in which the plasmid is
still present produce CD2517.1 at toxic levels and do not form colonies. Growing colonies
were then tested by PCR for the presence of the expected deletion.

Light microscopy

For light microscopy, bacterial cells were observed at 100x magnification on an Axioskop
Zeiss Light Microscope. Cell length was estimated for more than 100 cells for each strain
using ImageJ software.

Subcellular localization of HA-tagged toxins by cell fractionation and Western blotting

*C. difficile* cultures were inoculated from overnight grown cells in 10 ml of TY medium at an
optical density at 600 nm (OD$_{600}$) of 0.05. Cultures were allowed to grow for 3 hours before
the addition of 250 ng/ml ATc and incubation for 90 min. Then, cells were centrifuged and
proteins were extracted. Cell lysis, fractionation and protein analysis were performed as
previously described. Coomassie staining was performed for loading and fractionation
controls. Western blotting was performed with anti-HA antibodies (1:2,000) (Osenses) using
standard methods.

RNA extraction, Northern blot and 5'/3’RACE

Total RNA was isolated from *C. difficile* strains after 4, 6 or 10 hrs of growth in TY medium,
or 7.5 hrs in TY medium containing 7.5 µg/ml of Tm and 250 ng/ml of ATc for strains
carrying pRPF185 derivatives, as previously described. Starvation conditions
corresponded to a 1 h incubation of exponentially grown cells (6 h of growth) in PBS buffer
at 37°C. Northern blot analysis and 5'/3’RACE experiments were performed as previously
described.22

**Measurement of RNA decay by rifampicin assay**

For determination of toxin and antitoxin RNA half-lives the *C. difficile* strains were grown in
TY medium supplemented with 250 ng/ml ATc and 7.5 µg/ml Tm for 7.5 h at 37°C.
Samples were taken at different time points after the addition of 200 µg/mL rifampicin (0, 2,
5, 10, 20, 40, 60 and 120 min) and subjected to RNA extraction and Northern blotting.

**Plasmid stability assays**

Overnight cultures of *C. difficile* cells containing the pMTL84121 empty vector or the
pMTL84121 derivatives were grown in TY broth with Tm and used to inoculate (at 1%) 5 ml
of fresh TY broth without antibiotic. Every 10 to 14 hrs, 1% of the cultures were reinoculated
into fresh TY broth without antibiotic. After seven passages, CFUs were estimated on TY
plates supplemented or not with Tm to differentiate between the total number of cells and the
plasmid-containing cells.

**Quantification of the frequency of prophage excision**

The frequency of prophage excision in different *C. difficile* strains was estimated by
quantitative PCR on genomic DNA extracted using the NucleoSpin Microbial DNA kit
(Macherey-Nagel). The total chromosome copy number was quantified based on the
reference gene *dnaF* (*CD1305*) encoding DNA polymerase III. The number of chromosomes
devoid of phiCD630-1 was quantified by PCR amplification using primers flanking
phiCD630-1 (Table S2), which only results in PCR products when the prophage is excised.

**PhiCD630-1 stability assays**
Overnight cultures of *C. difficile* strain 630 ΔphiCD630-2 phiCD630-1::erm and ΔphiCD630-2 phiCD630-1-ΔT4::erm were used to inoculate 10 ml of TY broth at an initial OD$_{600}$ of 0.05. Every 10 to 14 h, cultures were subcultured at an initial OD$_{600}$ of 0.05. After four passages, the cultures were serially diluted and plated on BHI plates to estimate the total CFUs and on BHI plates supplemented with 2.5 µg/ml erythromycin to determine the number of CFUs in which phiCD630-1 was still present. For cells expressing the excisionase gene, overnight cultures of *C. difficile* strain 630 ΔphiCD630-2 phiCD630-1::erm and ΔphiCD630-2 phiCD630-1-ΔT4::erm carrying pDIA6867 were used to inoculate fresh TY broth with Tm and 10 ng/ml ATc at an initial OD$_{600}$ of 0.005. After 24 hrs of incubation at 37°C, cultures were serially diluted and plated on BHI plates to estimate the total CFUs and on BHI plates supplemented with 2.5 µg/ml erythromycin to determine the number of CFUs in which phiCD630-1 was still present. The OD$_{600}$ of the cultures was also measured to monitor cell growth.

ACKNOWLEDGEMENTS

This work was supported by Agence Nationale de la Recherche (“CloSTARn”, ANR-13-JSV3-0005-01 to O.S.), the Institut Universitaire de France (to O.S.), the University Paris-Saclay, the Institute for Integrative Biology of the Cell, the Pasteur Institute, the DIM-1HEALTH regional Ile-de-France program (LSP grant no. 164466), the CNRS-RFBR PRC 2019 (grant no. 288426 № 19-54-15003) to O.S., and a Vernadski fellowship to A.M.

We would like to thank Marc Monot for helpful discussions.

AUTHOR CONTRIBUTIONS

J.P. and O.S. conceived and coordinated the study, which was initiated by P.B. J.P. and O.S. performed the majority of the experiments. A.H. constructed vectors and deletion mutants. J.R.G. performed the *in-silico* analyses, A.M. performed growth curves and light microscopy.
L-C.F. and B.D. provided scientific insight into the design of the experiments. J.P. and O.S. wrote the paper and all authors reviewed and approved the final version of the manuscript.

COMPETING INTERESTS

The authors declare that they have no conflict of interest.

FIGURE LEGENDS

Figure 1. Identification and functionality of novel toxin genes within phiCD630-1. (A) Protein alignment of toxins CD0977.1 with the newly identified CD0904.1, CD0956.2 and CD0956.3. The hydrophobic amino acids are indicated in red. (B) Maps and alignment of the phiCD630-1 and phiCD630-2 genomes. The location of toxin genes in both prophages is indicated. (C) Growth of C. difficile 630Δerm strains harbouring the pRPF185-based plasmids on BHI agar plates supplemented with Tm and with (+ATc) or without 10 ng/ml of ATc inducer (-ATc) after 24 hrs of incubation at 37°C. Schematic representations of the constructs are shown.

Figure 2. Detection of CD0977.1 and RCd11 transcripts and CD0977.1-HA protein. (A) Northern blot of total RNA from C. difficile carrying p (empty vector), pT (expression of CD0977.1) or pTA (expression of CD0977.1 and its antitoxin) in absence (- ATc) or in presence (+ ATc) of 250 ng/ml of the inducer ATc. (B) Northern blot of total RNA from C. difficile carrying p (empty vector) or pdccA (expression of the diguanylate cyclase encoding gene dccA) in presence of 250 ng/ml ATc. (C) Northern blot of total RNA from C. difficile 630Δerm carrying an empty vector (wt/p) collected at the indicated time after addition of rifampicin. All Northern blots were probed with a radiolabelled oligonucleotide specific to the toxin (T CD0977.1/CD2889) or the antitoxin (AT RCd11/RCd12) transcript and 5S RNA at the bottom serves as loading control. The arrows show the detected transcripts with their estimated size. The relative intensity of the bands was quantified using the ImageJ software. (D) Detection and subcellular localization of the CD0977.1-HA protein. Immunoblotting with anti-HA detected a major polypeptide of ~12 kDa in whole cell extracts of C. difficile carrying pT-HA (CD0977.1-HA) grown in presence of 250 ng/ml of ATc but not in extracts...
of C. difficile carrying pT (non-tagged CD0977.1) (left panel). The culture of C. difficile carrying pT-HA was fractionated into supernatant (SN), cell wall (CW), membrane (Mb) and cytosolic (Cy) compartments and immunoblotted with anti-HA antibodies. Proteins were separated on 12% Bis-Tris polyacrylamide gels in MES buffer.

Figure 3. Impact of toxin-antitoxin co-expression on growth. The effect on the toxicity of CD0977.1 of long and short antitoxin transcripts expressed in cis (A) and in trans (B) was assessed. Growth of C. difficile 630Δerm strains harbouring the pRPF185-based plasmids on BHI agar plates supplemented with Tm and 10 ng/mL of ATc inducer after 24 hrs of incubation at 37°C and in TY broth at 37°C in the presence of 5 ng/mL ATc. Schematic representations of the constructs are shown.

Figure 4. Impact of toxin-antitoxin modules on plasmid loss in the absence of selection pressure. The stability of pMTL84121 (p, empty vector) and pMTL84121-derived vectors expressing the different TA modules in C. difficile 630Δerm was determined after seven passages (every 12 hours) in TY broth without thiamphenicol. Values represent means ± standard deviations (N = 3). ** P ≤ 0.05 and *** P ≤ 0.001 by an unpaired t test.

Figure 5. Impact of toxin /antitoxin modules on prophage maintenance. (A) Schematic representation of the method used to quantify prophage maintenance. A cassette containing an erythromycin resistance gene (ermB) under control of the strong thl promoter of C. acetobutylicum was introduced into an innocuous location of phiCD630-1, within the intergenic region between CD0946.1 and CD0947 genes, encoding a hypothetical protein and a putative scaffold protein, respectively. Cultures grown for 24 hrs were plated on erythromycin-containing agar plates and cells that lost the prophage were selectively killed. (B) and (C): Strains ΔphiCD630-2 phiCD630-1::erm and ΔphiCD630-2 phiCD630-1-ΔT4::erm carrying a pDIA6867 (overproducing the excisionase CD0912) were inoculated at an initial optical density at 600 nm (OD\textsubscript{600nm}) of 0.005 in TY medium supplemented with 7.5 μg/ml Tm and 10 ng/ml ATc. Cultures were incubated at 37°C for 24 hrs and bacterial growth was determined by measurement of the OD\textsubscript{600nm} (B) and the maintenance of prophages was quantified by plating serial dilutions on agar plates supplemented or not with 2.5
μg/ml Erm (C). Values represent means ± standard deviations (N = 3). *** P ≤ 0.001 by an unpaired t test.

Figure 6. Relationship between phage phylogeny and toxin variants. Putative toxin protein sequences detected in all available phage genomes were aligned using MUSCLE (v3.8) algorithm (EMBL-EBI). All phage genomes were re-scanned for potential ORFs using the NCBI ORFfinder software and detected ORFs were translated into their corresponding protein sequence. Protein sequences were combined to create a local BLASTp database and confirmed functional toxins CD0956.2 (large variant) and CD0904.1 (short variant) were used as queries to search for putative toxins in the database. Significant hits (min 45% identity, min 28% coverage) were retrieved and the corresponding proteins were aligned using MUSCLE (v3.8) algorithm (EMBL-EBI). The protein sequence consensus is shown. A phylogenetic tree was built using the Poisson distance method and neighbour joining implemented in Seaview (v4.4.2). Residues were coloured according to high hydrophobicity (red) and low hydrophobicity (blue).

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. RNA-seq and TSS mapping profiles for the TA loci in C. difficile strain 630. The TAP-/TAP+ profile comparison for 5’-end RNA-seq data is aligned with RNA-seq data for CD0977.1-RCd11 (A) and CD0904.1-RCd13 TA (B) genomic regions. The TSS are indicated by red broken arrows in accordance with the positions of 5’-transcript ends shown by vertical green lines on the sequence read graphs corresponding to TSS. TSS corresponds to positions with significantly greater numbers of reads in TAP+ samples. 5’-end sequencing data show 51-bp reads matching to the 5’-transcript ends, while RNA-seq data show reads covering the whole transcript. Coding sequences are indicated by blue arrows and the regulatory RNAs are indicated by grey arrows. The promoter regions of the antitoxins are shown. The positions of TSS “+1” are represented. (C) Alignment between the sequences of CD0904.1-RCd13 (top line) and CD0956.3-RCd14 (bottom line) using
Figure S2. Impact of toxin CD0977.1 expression on cell viability and morphology. Growth (A) and viability (B) of C. difficile 630Δerm strain carrying the pRPF185-based plasmids (empty: p or with CD0977.1 toxin gene under the control of the Ptet promoter: pT) in TY broth in presence of 200 ng/ml ATc. The time point of ATc addition is indicated by a vertical arrow. Values represent means ± standard deviations (N = 3). * P ≤ 0.05 by a Student’s t test. (C) Selected images from light microscopy observations of 630/p, 630/pT and 630/pTA strains grown in TY broth for 1 h at 37°C after the addition of 250 ng/mL ATc. Cell length was estimated using the ImageJ software for at least 115 cells per strain. The mean values with standard deviations are indicated for each strain, as well as the proportion of cells with length above 2 standard deviations relative to the 630/p control strain mean length.

Figure S3. Sequence of type I CD0977.1-RCd11 TA locus in C. difficile. The thick horizontal arrows below the double-stranded sequences show the toxin and antitoxin transcripts and the direction of transcription. The transcriptional start sites for sense and antisense transcripts identified by 5'/3'RACE and TSS mapping are indicated by vertical arrows with their genomic location. Line thickness corresponds to the proportion of observed extremities. The genomic location of 5’- and 3’-ends of the transcripts are indicated above the sequence. The inverted repeats at the position of transcriptional terminators are indicated by thin black arrows. The positions of Sigma A-dependent -10 and -35 promoter elements of antitoxin (AT) are shown in red boxes. The positions of Sigma A-dependent -10 and -35 elements promoter, ribosome binding site, translation initiation codon and stop codon of toxin (T) mRNA are underlined in blue. The positions of Sigma B-dependent promoter elements are shown in green boxes for both TA genes.
Figure S4. Northern blots showing the stability of CD0977.1/CD2889 toxin (A) and RCd11/RCd12 antitoxin (B) transcripts in strains depleted for RNase III, RNase Y, RNase J and Hfq. To determine half-lives, samples were taken at the indicated time points after the addition of 200 µg/mL rifampicin. RNAs were extracted from strains CDIP369 (630/p), CDIP230 expressing an antisense RNA for the rncS gene encoding RNase III (AS rncS), CDIP53 strain expressing an antisense RNA for the hfq gene (AS hfq), CDIP55 strain expressing an antisense RNA for the rnJ gene encoding RNase J (AS rnJ) and CDIP57 strain expressing an antisense RNA for the rny gene encoding RNase Y (AS rny). All Northern blots were probed with a radiolabelled oligonucleotide specific to the toxin (T CD0977.1/CD2889) or the antitoxin (AT RCd11/RCd12) transcript and 5S RNA at the bottom serves as loading control. The relative intensities of the bands were quantified using ImageJ software.

Figure S5. Impact of toxin-antitoxin co-expression on growth. The effect on the toxicity of CD0977.1 of long and short antitoxin transcripts expressed in cis (A) and in trans (B) was assessed. Growth of C. difficile 630Δerm strains harbouring the pRPF185-based plasmids on BHI agar plates supplemented with Tm and the indicated concentration of ATc inducer after 24 hrs of incubation at 37°C. Schematic representations of the constructs are shown.

Figure S6. Nucleotide alignment of antitoxins. (A) Nucleotide alignment of RCd11 and RCd12 using LALIGN. (B) Nucleotide alignment of RCd11 and RCd10 using LALIGN.

Figure S7. Secondary structure prediction of CD0977.1 mRNAs and corresponding antitoxin.

RCd11 RNA. The RNA secondary structure prediction was performed by Mfold software. Ribosome binding site (SD), translation initiation codon and stop codon positions are highlighted. The positions of mismatches in RCd12 AT sequence are indicated.

Figure S8. New vector for efficient gene editing in C. difficile. Features of the pMSR (A) and pMSR0 (B) vectors used for allele exchange in C. difficile 630Δerm and C. difficile 027 ribotype strains, respectively. The toxin gene CD2517.1 is under the control of the ATc inducible promoter P_{tet} and the RCd8 antitoxin present in pMSR0 is under control of its own promoter. Schematic overview
of the allele exchange protocol (C) and of the inducible counterselection method used to isolate double cross-over clones (D). Isolated single cross-over integrants were restreaked on ATc-containing agar plates to induce synthesis of toxin CD2517.1. Cells that kept the pMSR plasmid (either integrated or excised) produced CD2517.1 and were selectively killed.

Figure S9. Deletion of the phiCD630-2 prophage from C. difficile 630Δerm using the newly developed allele exchange method. (A) Schematic representation of phiCD630-2 in C. difficile 630Δerm. The location of primers used to screen for mutants is represented. (B) PCR products amplified using the indicated primers from the parental strain 630Δerm (WT) and the ΔphiCD630-2 strain (Δ). A product of 1,348 bp could be amplified with primers JP528-JP527 if phiCD630-2 had been deleted, whereas a product of 1,339 bp could be amplified with primers JP570-JP527 if phiCD630-2 was still present.

Figure S10. Maintenance and site-specific excision of phiCD630-1 from genomic DNA of C. difficile 630. (A) Schematic representation of phiCD630-1 DNA excision from genomic DNA of C. difficile 630, and circularization. The location of primers used to demonstrate prophage excision is represented. (B) PCR products amplified using the indicated primers from C. difficile 630 genomic DNA. (C) DNA sequences within attP, attB, attL and attR as determined by Sanger sequencing. The central identical sequences where recombination occurs are shown in bold. Short segments of sequence surrounding the central identity region are shown in blue (bacterial sequences) and in red (phage sequences). (D) The maintenance of prophage in strains ΔphiCD630-2 phiCD630-1::erm and ΔphiCD630-2 phiCD630-1-ΔT4::erm was determined after four passages in TY broth by plating serial dilutions on agar plates supplemented or not with 2.5 μg/ml Erm. Values represent means ± standard deviations (N = 3).

Figure S11. Impact of excisionase (CD0912) overproduction on excision of phiCD630-1 and phiCD630-1-ΔT4. Excision rate of phiCD630-1 was higher in C. difficile carrying pDIA6867 (inducible expression of CD0912) in presence of 10 ng/ml of the inducer ATc when compared to the ΔphiCD630-2 strain without plasmid and was not impacted by the deletion of toxin genes. Values represent means ± standard deviations (N = 3).
Fig. S12. Heatmap and phylogenetic tree showing *C. difficile* relatedness for 27 sequenced *C. difficile* phage genomes available in the NCBI database. The Gegenees software (v2.2.1) was used to produce the heatmap of genome similarity. Similarity scores are based on a fragmented all-against-all pairwise alignment using BLASTn and the accurate alignment option (fragment size, 200; step size, 100). The colours reflect the similarity, ranging from low (red) to high (green). Phages were assigned to a genus if they clustered closely to another phage previously described as a member of that genus. The phylogenetic tree is based on the sequence similarity scores from the same whole-genome comparison and was constructed using the neighbour joining method with the SplitsTree4 software (v 4.13.1).

Fig. S13. Identification of toxin homologs outside *C. difficile* phages. Toxin homologs found outside *C. difficile* phages were aligned. The protein sequence consensus is shown at the bottom. Phylogenic analysis of toxins is also represented.

REFERENCES

1. Banawas SS. *Clostridium difficile* Infections: A Global Overview of Drug Sensitivity and Resistance Mechanisms. *Biomed Res Int* 2018, 8414257 (2018).
2. Rupnik M, Wilcox MH, Gerding DN. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nature reviews Microbiology* 7, 526-536 (2009).
3. Carroll KC, Bartlett JG. Biology of *Clostridium difficile*: implications for epidemiology and diagnosis. *Annual review of microbiology* 65, 501-521 (2011).
4. Vedantam G, Clark A, Chu M, McQuade R, Mallozzi M, Viswanathan V. *Clostridium difficile* infection: toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. *Gut Microbes* 3, 121-134 (2012).
5. Janoir C. Virulence factors of *Clostridium difficile* and their role during infection. *Anaerobe* 37, 13-24 (2016).
6. Fortier LC. Bacteriophages Contribute to Shaping *Clostridioides (Clostridium) difficile* Species. *Front Microbiol* 9, 2033 (2018).

7. Fortier LC, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4, 354-365 (2013).

8. Sekulovic O, Meessen-Pinard M, Fortier LC. Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 lysogens. *Journal of bacteriology* 193, 2726-2734 (2011).

9. Rosner JL. Formation, induction, and curing of bacteriophage P1 lysogens. *Virology* 48, 679-689 (1972).

10. Echols H. Constitutive integrative recombination by bacteriophage lambda. *Virology* 64, 557-559 (1975).

11. Page R, Peti W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nature chemical biology* 12, 208-214 (2016).

12. Brantl S. Bacterial type I toxin-antitoxin systems. *RNA biology* 9, 1488-1490 (2012).

13. Brantl S, Jahn N. sRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiol Rev* 39, 413-427 (2015).

14. Coray DS, Wheeler NE, Heinemann JA, Gardner PP. Why so narrow: Distribution of antisense regulated, type I toxin-antitoxin systems compared with type II and type III systems. *RNA biology* 14, 275-280 (2017).

15. Gerdes K, Christensen SK, Lobner-Olesen A. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* 3, 371-382 (2005).

16. Gerdes K, Maisonneuve E. Bacterial persistence and toxin-antitoxin loci. *Annual review of microbiology* 66, 103-123 (2012).

17. Hayes F. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* 301, 1496-1499 (2003).

18. Wang X, Wood TK. Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Applied and environmental microbiology* 77, 5577-5583 (2011).

19. Wen Y, Behiels E, Devreese B. Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. *Pathogens and disease* 70, 240-249 (2014).
20. Yamaguchi Y, Inouye M. Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat Rev Microbiol* **9**, 779-790 (2011).

21. Maikova A, *et al.* Discovery of new type I toxin-antitoxin systems adjacent to CRISPR arrays in *Clostridium difficile*. *Nucleic acids research* **46**, 4733-4751 (2018).

22. Soutourina OA, *et al.* Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. *PLoS genetics* **9**, e1003493 (2013).

23. Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. *Applied and environmental microbiology* **78**, 4683-4690 (2012).

24. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *J Microbiol Methods* **70**, 452-464 (2007).

25. Marchler-Bauer A, *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research* **39**, D225-229 (2011).

26. Rashid SR, Clokie MR, Millard AD. Draft Genome Sequences of Three Novel *Clostridium* Isolates from Northern Iraq. *Genome announcements* **4**, (2016).

27. Pogue-Geile KL, DasSarma S, King SR, Jaskunas SR. Recombination between bacteriophage lambda and plasmid pBR322 in *Escherichia coli*. *Journal of bacteriology* **142**, 992-1003 (1980).

28. Stone JC, Miller RC, Jr. Plasmid-phage recombination in T7 infected *Escherichia coli*. *Virology* **137**, 305-313 (1984).

29. Purcell EB, Tamayo R. Cyclic diguanylate signaling in Gram-positive bacteria. *FEMS microbiology reviews* **40**, 753-773 (2016).

30. McKee RW, Harvest CK, Tamayo R. Cyclic Diguanylate Regulates Virulence Factor Genes via Multiple Riboswitches in *Clostridium difficile*. *mSphere* **3**, (2018).

31. Harms A, Brodersen DE, Mitarai N, Gerdes K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol Cell* **70**, 768-784 (2018).

32. Yao J, *et al.* Type II toxin/antitoxin system ParESO /CopASO stabilizes prophage CP4So in *Shewanella oneidensis*. *Environmental microbiology* **20**, 1224-1239 (2018).

33. Wozniak RA, Waldor MK. A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS genetics* **5**, e1000439 (2009).
34. Brantl S, Muller P. Toxin-Antitoxin Systems in *Bacillus subtilis*. *Toxins* **11**, (2019).

35. Van Melderen L. Toxin-antitoxin systems: why so many, what for? *Curr Opin Microbiol* **13**, 781-785 (2010).

36. Magnuson RD. Hypothetical functions of toxin-antitoxin systems. *Journal of bacteriology* **189**, 6089-6092 (2007).

37. Unterholzner SJ, Poppenberger B, Rozhon W. Toxin-antitoxin systems: Biology, identification, and application. *Mobile genetic elements* **3**, e26219 (2013).

38. Zhang XZ, Yan X, Cui ZL, Hong Q, Li SP. *mazF*, a novel counter-selectable marker for unmarked chromosomal manipulation in *Bacillus subtilis*. *Nucleic acids research* **34**, e71 (2006).

39. Al-Hinai MA, Fast AG, Papoutsakis ET. Novel system for efficient isolation of *Clostridium* double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration. *Applied and environmental microbiology* **78**, 8112-8121 (2012).

40. Dupuy B, Sonenshein AL. Regulated transcription of *Clostridium difficile* toxin genes. *Molecular microbiology* **27**, 107-120 (1998).

41. Fagan RP, Fairweather NF. *Clostridium difficile* has two parallel and essential Sec secretion systems. *The Journal of biological chemistry*, (2011).

42. Heap JT, Pennington OJ, Cartman ST, Minton NP. A modular system for Clostridium shuttle plasmids. *J Microbiol Methods* **78**, 79-85 (2009).

43. Wydau-Dematteis S, *et al*. Cwp19 Is a Novel Lytic Transglycosylase Involved in Stationary-Phase Autolysis Resulting in Toxin Release in *Clostridium difficile*. *mBio* **9**, (2018).

44. Collins TJ. ImageJ for microscopy. *BioTechniques* **43**, 25-30 (2007).

45. Peltier J, *et al*. Cyclic diGMP regulates production of sortase substrates of *Clostridium difficile* and their surface exposure through ZmpI protease-mediated cleavage. *The Journal of biological chemistry* **290**, 24453-24469 (2015).

46. Peltier J, Soutourina O. Identification of c-di-GMP-Responsive Riboswitches. *Methods in molecular biology* **1657**, 377-402 (2017).
Fig. 4

![Bar graph showing plasmid maintenance (%) for different genotypes.](image-url)
