Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in Clostridioides difficile

Johann Peltier, Audrey Hamiot, Julian R Garneau, Pierre Boudry, Anna Maikova, Eliane Hajnsdorf, Louis-Charles Fortier, Bruno Dupuy, Olga Soutourina

Type I toxin-antitoxin systems contribute to mobile genetic elements maintenance in *Clostridioides difficile*

Johann Peltier[1,2], Audrey Hamiot[1§], Julian R. Garneau[3], Pierre Boudry[1#], Anna Maikova[1,2,4§§], Eliane Hajnsdorf[5], Louis-Charles Fortier[3], Bruno Dupuy[1] and Olga Soutourina[1,2,6]*

1. Laboratoire Pathogenèse des Bactéries Anaérobies, CNRS-2001, Institut Pasteur, Université de Paris, F-75015 Paris, France.

2. Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, 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 of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, 143028, Russia

5. UMR8261, CNRS, Université de Paris, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, 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

§§ Present address: Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, 195251, Russia

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. In type I TA, synthesis of the toxin protein is prevented by the transcription of an antitoxin RNA during normal growth. The first type I TA modules were recently identified in the human enteropathogen *Clostridioides* (formerly *Clostridium*) *difficile*. Here, we report the characterization of five additional type I TA systems present within phiCD630-1 and phiCD630-2 prophage regions of *C. difficile* strain 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.

INTRODUCTION

Clostridioides difficile is a medically important human enteropathogen that became a key public health concern over the last two decades in industrialized countries [1, 2]. This strictly anaerobic spore-forming Gram-positive bacterium is a major cause of antibiotic-associated nosocomial diarrhoea in adults [3]. The main virulence factors of *C. difficile* are two toxins, TcdA and TcdB, produced by all toxigenic strains [4] and some isolates produce a binary toxin *Clostridium difficile* transferase (CDT). Additional factors, such as adhesins, pili, and flagella, involved in the interactions with the host during colonization have also been identified [5]. However, 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) [6]. 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 [7]. For example, the largely used laboratory strain 630 carries two homologous prophages, phiCD630-1 and phiCD630-2, while the NAP1/B1/027 epidemic strain R20291 carries one prophage (phi-027). The importance of prophages in the evolution and virulence of many pathogenic bacteria has clearly been demonstrated [8]. 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 [7, 8, 9]. 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 for *Escherichia coli* phages to range between 10^{-5} for phage P1 to $< 10^{-6}$ for phage lambda [10, 11]. To our knowledge, no experiments have been conducted to evaluate prophage loss rates in *C. difficile*.

TA modules are widespread in bacteria and archaea. These loci comprise two genes encoding a stable toxin and an unstable antitoxin [12]. Overexpression of the toxin has either bactericidal or bacteriostatic effects on the host cell while the antitoxin is able to neutralize the toxin action or production. 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 [12]. 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 [13, 14]. 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 [15]. Major functions suggested for TA modules include plasmid maintenance, abortive phage infection and persistence, however, their role in persister cell formation in the presence of antibiotics remains a subject of controversy [16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]. 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.

We recently reported the identification of the first type I TA systems associated with CRISPR arrays in *C. difficile* genomes [27]. The co-localization and co-regulation by the general stress response Sigma B factor and biofilm-related factors of TA and CRISPR components 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

104 homologous phiCD630-1 and phiCD630-2 prophages in *C. difficile* strain 630. In the present
105 work, we characterize additional type I TA modules highly conserved within *C. difficile*
106 prophages and provide experimental evidence of their contribution to prophage maintenance
107 and stability. Moreover, we demonstrate here that inducible toxicity caused by type I toxins
108 can be used as a counter selection marker in allele exchange genome editing procedures by
109 promoting the elimination of plasmid-bearing cells, largely improving their efficiency.

110 **RESULTS**

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

112 Multiple TA modules have been discovered in bacterial chromosomes including
113 prophage regions [12]. In *C. difficile*, we have recently identified several type I TA pairs
114 adjacent to CRISPR arrays, two of them being located inside the phiCD630-1 and phiCD630-
115 2 prophages of the strain 630 (*CD0956.2*-RCd10 and *CD2907.1*-RCd9, respectively) [27]. To
116 determine whether other type I TA modules might be present within phiCD630-1, we
117 performed a bioinformatics analysis on the phiCD630-1 sequence. Due to the small size of
118 the toxin-encoding genes, standard methods of open reading frame (ORF) detection and gene
119 annotation can hinder the identification of all toxin homologs. Moreover, prophages are
120 characterized by a very high gene density, which can impede such detection of small and
121 often overlapping coding regions. We therefore used the tBlastn program using the
122 previously identified type I toxin CD0956.2 as a query, as it does not depend on annotation
123 and ORF detection. We identified gene *CD0977.1* and two other novel putative genes,
124 unannotated on the genome, that we named *CD0904.1* and *CD0956.3*. These genes code for
125 small proteins of 47, 35 and 34 amino acids, respectively (Fig. 1A). Prophages phiCD630-1
126 and phiCD630-2 share a large region of homology with almost identical sequences, which
127 include a duplication of *CD0977.1* and *CD0956.3* (named *CD2889* and *CD2907.2* in

phiCD630-2, respectively) (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 [28] (Fig. S1A-C). The presence of a consensus RBS sequence (AGGAGG) 7-8 nucleotides upstream of the respective ATG start codons suggest that the corresponding proteins are produced (data not shown). 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) [29]. Analysis of our previous TSS mapping data [28] and sequence alignments (Fig. S1) 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 for both the toxin and antitoxin genes (Fig. S1 and data not shown) [30]. 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 P_{tet} promoter were constructed to overexpress *CD0904.1*, *CD0956.3* and *CD0977.1* toxin genes (pT) or toxin-antitoxin modules (pTA) in *C. difficile* 630Δ*erm*. Antisense RNAs are expressed from their own promoter in pTA. Growth of 630Δ*erm* 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 630Δ*erm*/pT strains was completely inhibited when ATc was present in the medium, while strains 630Δ*erm*/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 [28]. Most of these type I c-di-GMP-responsive riboswitches negatively control downstream genes by premature termination of transcription in the presence of c-di-GMP [28, 31]. 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 630Δ*erm*/pT, unlike 630Δ*erm*/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 [27], 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 630Δ*erm*/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 630Δ*erm*/p, 630Δ*erm*/pT (*CD0977.1*) and 630Δ*erm*/pTA (*CD0977.1*-RCd11) strains (Fig. 2A and 2B). 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 an 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. The toxin overexpression in the presence of ATc inducer resulted in a decreased amount of the major 150-nt RCd11 antitoxin expressed from chromosomal location (lanes "pT" compared in the absence and in the presence of ATc). Similarly, for the strain carrying the entire TA locus on pTA plasmid expressing the antitoxin from its own strong promoter, the toxin overexpression after ATc induction led to a decrease in the 150-nt RCd11 antitoxin level (lanes "pTA" compared under conditions "-ATc" and "+ATc"). To determine the impact of c-di-GMP on

177 the antitoxin transcripts, we elevated c-di-GMP intracellular levels in the 630Δ*erm* wild type
178 strain by expressing the gene *dccA*, coding for a diguanylate cyclase involved in c-di-GMP
179 production, from a plasmid (p*dccA*) (Fig. 2C), as previously reported [28]. A c-di-GMP-
180 regulated read-through transcript of about 400 nt, as well as a terminated transcript of about
181 140 nt were detected in this strain by Northern blotting using a riboswitch-specific probe. In
182 contrast, abundance of the 150-nt RCd11 antitoxin and toxin transcripts was not affected by
183 fluctuations of c-di-GMP levels. Elevated ci-di-GMP intracellular level could be associated
184 with biofilm growth conditions. As for some other type I TA transcripts in our previous study
185 [27], we detected by qRT-PCR analysis up to 20-fold increase in *CD0977.1* toxin gene
186 expression in biofilms as compared to planktonic culture, but no difference for short RCd11
187 form amount (data not shown).

188 We then mapped the transcriptional start (TSS) and termination sites for the genes of
189 the potential RCd11/RCd12-*CD0977.1/CD2889* TA modules by 5'/3'RACE analysis (Fig.
190 S3, Table S1). The results obtained agreed well with the transcript lengths deduced from TSS
191 mapping, RNA-seq and Northern blot. Taken together, these data suggest the presence of two
192 tandem TSS for RCd11, i.e. P_1 associated with c-di-GMP-dependent riboswitch, yielding a
193 premature terminated transcript of ~140 nt, primary read-through transcript of ~400 nt
194 (referred to as long transcript hereafter) and a processed transcript of ~300 nt, and P_2 located
195 downstream from the riboswitch, yielding a transcript of ~150 nt (referred to as short
196 transcript hereafter) (Fig. S1). All these transcripts except for riboswitch-associated
197 terminated transcript shared the same Rho-independent terminator (Fig. S3). According to the
198 position of the Northern blotting probes, the long 400-nt transcript could be detected with
199 both riboswitch- and RCd11-specific probes, while terminated 140-nt transcript could be
200 revealed only with riboswitch-specific probe and RCd11-specific probe hybridized to 300-nt
201 and 150-nt transcripts (Fig. 2A).

We investigated the interaction between *CD0977.1* toxin mRNA and the short and long RCd11 RNAs to determine whether they form kissing complexes as in the case of the RCd9/CD2907.1 TA pair [27]. The results shown in Figure S4 reveal no difference in duplex formation with toxin mRNA for long and short antitoxin forms under native or full RNA duplex conditions suggesting that no kissing intermediate is formed during binding in native conditions *in vitro*. It should be noticed that only a fraction of *CD0977.1* can interact with both antitoxin forms even when they are in excess, indicating that it is tightly folded in these experimental conditions (Fig. S4).

It is in the nature of type I antitoxins to be short-lived in contrast to the stable toxin mRNA [14]. 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 630Δ*erm*/p carrying an empty vector, the half-life of the major short 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. 2D). Interestingly, depletion of the RNA chaperone protein Hfq, which generally increases the intracellular half-life of sRNAs and stabilizes the interactions between sRNAs and their target mRNAs, 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. S5). By contrast, the stable *CD0977.1* toxin mRNA was further stabilized to over 120 min half-life in the strains depleted for the ribonucleases RNase III, RNase J and RNase Y, that could be involved in toxin and antitoxin RNA decay. For antitoxin RCd11 RNA, we also observed a stabilization in strain depleted for RNase Y (Fig. S5) suggesting that this ribonuclease contributes 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. 2E). 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 then studied (Fig. 2E). 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_4 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) (Fig. S1), 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_4 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 and the riboswitch were present (pDIA6785) and the one in which only the toxin gene is expressed (pDIA6335) served as a positive and as a negative control for the assay, respectively. All plasmids were introduced into *C. difficile* 630Δ*erm* and the corresponding strains were grown

252 on BHI plates supplemented with 10 and 100 ng/ml ATc to induce *CD0977.1* toxin
253 expression. Growth of the strain carrying pDIA6816 was similar to that observed for the
254 control strain carrying pDIA6785 in the presence of 10 ng/ml ATc and was slightly defective
255 in the presence of 100 ng/ml ATc (Fig. 3A and Fig. S6A). In contrast, the strain carrying
256 pDIA6817 did not grow in the presence of 10 or 100 ng/ml ATc, similarly to the negative
257 control strain. Similar results were obtained when the strains were grown in an automatic
258 plate reader for 20 h in liquid medium in the presence of 5 ng/ml ATc (Fig. 3A).
259 Interestingly, induction of toxin expression on BHI plate with a lower dose of ATc (5 ng/ml)
260 led to a partial reversion of the growth defect of the strain carrying pDIA6817 unlike the
261 negative control strain (Fig. S6A). To further investigate the promoters functionality, a series
262 of promoter fragments fused to the *phoZ* reporter gene was created in the wild type strain,
263 and alkaline phosphatase (AP) activity was measured. After 4 h of growth in TY broth, the
264 P_2 promoter fragment exhibited a reporter activity 1.4-fold lower than that of the full length
265 promoter region, comprising P_1 and P_2, while the AP activity from the P_1 promoter
266 associated with the Cdi1_4 riboswitch fragment was 5.7-fold lower (Fig. 3B). This suggests
267 that the P_1 promoter activity is weaker than that of P_2, providing a rationale for the
268 incompetence of the long antitoxin form expressed from P_1 for toxin inhibition. Of note, the
269 promoter fragment comprising P_1 and the disrupted P_2 retained AP activity (Fig. S6B),
270 suggesting that the nucleotide substitutions in P_2 do not prevent expression of the antitoxin
271 transcripts. To determine whether promoters activity differs in various growth conditions, AP
272 activity was next measured after 10 h of growth in TY broth and under nutrient starvation
273 conditions. In these conditions, the full length promoter region (P_1 and P_2) exhibited a slight
274 decrease in activity while activity from the P_1 or P_2 promoter fusions was not significantly
275 different to that observed during the exponential growth phase (Fig. 3B), suggesting that the
276 expression of the antitoxin transcripts was not strongly modulated in these conditions. Taken

277 together, these results suggest that the short antitoxin transcript driven by promoter P_2 is
278 crucial for the efficient inactivation of the toxin, while the longer antitoxin transcript directed
279 by P_1 is dispensable.

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

281 Nucleotide sequences of short RCd11, lying within phiCD630-1 and short RCd12, lying
282 within phiCD630-2, are almost identical with only 3 mismatches located near the 3' end, in
283 the region overlapping with the toxin transcript (Fig. S7A and S8). The structure prediction
284 suggested that the 3' part folded similarly with two conserved hairpin structures in both
285 antitoxin short and long form predictions (Fig. S8). We therefore wondered whether RCd12
286 could cross-react with the transcript of the non-cognate toxin CD0977.1. To answer this
287 question, we generated constructs in which *CD0977.1* toxin gene under the control of the P_{tet}
288 promoter and different antitoxin genes with their own promoter were co-expressed from the
289 same plasmid but from distant locations (Fig. 3C). As anticipated, expression of RCd11 in
290 *trans* (pDIA6791) counteracted the toxicity associated with the expression of the cognate
291 toxin both on plate and in liquid culture (Fig. 3C and Fig. S6C). Replacement of RCd11 with
292 RCd12 (pDIA6792) led to the same result (Fig. 3C and Fig. S6C). By contrast, *in trans*
293 expression of the more divergent RCd10 (the antitoxin of CD0956.2 toxin from a previously
294 characterized TA module lying within phiCD630-1 [27]) (pDIA6793) (Fig. S7B) failed to revert
295 the growth defect induced by *CD0977.1* expression (Fig. 3C and Fig. S6C). These data
296 indicate that antitoxins act in a highly specific manner to repress their cognate toxins, not
297 only when they are expressed from the native convergent TA configuration, but also when
298 expressed in *trans*. However, the specificity of interaction is permissive for at least 3
299 mismatches allowing RCd12 expressed from phiCD630-2 to efficiently prevent CD0977.1
300 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 [18]. 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 pMTL84121-derived plasmids in which each TA module of phiCD630-1was cloned and expressed under the control of their respective native promoter in *C. difficile* 630Δ*erm*. *C. difficile* 630Δ*erm* harbouring the empty vector pMTL84121 was used as a control. After 7 passages in TY broth in the absence of antibiotic pressure, pMTL84121 was maintained by only 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 mutants deleted for toxin genes of TA modules in *C. difficile* 630Δ*erm*. For this purpose, we constructed a new Allele-Coupled Exchange (ACE) vector, derived from pMTL-SC7315, a *codA*-based "pseudosuicide" plasmid [32]. The *codA* cassette was here replaced with the *CD2517.1* toxin gene placed under the control of the P_{tet} inducible promoter (Fig. S9A). The functionality of RCd8-*CD2517.1* type I TA module in *C. difficile* was previously demonstrated [27]. 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), greatly facilitating the isolation of *C. difficile* deletion mutant generated by double cross-over allele exchange (Fig. S9C 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. S9B). Using this new tool, we first deleted the 49.3 kb phiCD630-2 locus to prevent any interfering cross-talk with phiCD630-1 (Fig. S10). A multiple deletion mutant of toxin-encoding genes *CD0904.1*, *CD0956.2*, *CD0956.3* and *CD0977.1* (phiCD630-1ΔT4) was then generated in the ΔphiCD630-2 background.

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* 630Δ*erm* ΔphiCD630-2 with primers flanking the *attL* and *attR* sites of phiCD630-1 (Fig. S11A). A PCR product with a size of 88 bp corresponding to a region with the excised prophage was detected (Fig. S11B) 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. S11A and B). We could also deduce the *attB, attP, attL* and *attR* sites from the sequencing of the PCR products (Fig. S11C). However, the frequency of phiCD630-1 excision, as measured by quantitative PCR (qPCR), was very low (~ 0.015 %) (Fig. S12). 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 [33], 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. S11D).

In an attempt to artificially increase the excision rate of phiCD630-1, we ectopically expressed the putative excisionase *CD0912* of phiCD630-1 from the inducible P_{tet} 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 [34]. Induction of *CD0912* expression with 10 ng/ml ATc in *C. difficile* 630Δ*erm* resulted in a high excision rate of phiCD630-1 (~ 70 %), indicating that CD0912 functions as an excisionase for phiCD630-1 (Fig. S12). 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. S12).

Strains ΔphiCD630-2 phiCD630-1::*erm* and ΔphiCD630-2 phiCD630-1-ΔT4::*erm* carrying pDIA6867 or a control vector were then grown in TY supplemented with 7.5 µg/ml Tm and 10 ng/ml ATc. After 16 and 24 hrs of incubation at 37°C, measurement of the OD_{600} revealed a dramatic growth defect of ΔphiCD630-2 phiCD630-1::*erm* expressing the excisionase compared to the strains carrying the control vector (Fig. 5B). Expression of the excisionase in ΔphiCD630-2 phiCD630-1-ΔT4::*erm* also resulted in a growth defect, although at a lesser extent. 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 contribute to 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 [35]. 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. S13). 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. S14). These findings imply that *C. difficile* phages could recombine with plasmids to exchange genetic material, as already proposed for *E. coli* phages [36, 37].

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 known TA systems, i.e. (i) the toxins are membrane-associated proteins having 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 to date in other bacteria.

RCd11-*CD0977.1* and RCd12-*CD2889* TA pairs are duplications respectively located within the homologous regions of phiCD630-1 and phiCD630-2 prophages. Two tandem TSS were identified for RCd11 (and RCd12), with the first one 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 [28, 38]. 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 [38]. 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. 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 terminated transcript were detected, and no read-through seemed to occur [39]. However, our RACE-PCR and Northern-blot analysis indicated the presence of a transcript downstream from these riboswitches (Fig. 2, Fig. S3 and Table S1). 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 (Fig. 2). *In vitro* interaction assays showed that both long and short RCd11 antitoxin RNAs could form a duplex with

423 *CD0977.1* mRNA with the same efficiency and only a fraction of toxin mRNA was included
424 in these complexes probably due to extensive RNA folding (Fig. S4 and S8). Despite these
425 data, we found that the shorter and more abundant RCd11 transcript alone was sufficient to
426 ensure complete CD0977.1 toxin inactivation under our conditions. In accordance, the
427 analysis of promoter activities showed that c-di-GMP-independent P_2 promoter driving the
428 short RCd11 transcription was much stronger than the P_1 promoter. In contrast, the longer
429 antitoxin transcript associated with the cdi1_4 riboswitch could only counteract the
430 CD0977.1 toxicity when the toxin gene was expressed at low levels. This suggests that this
431 antitoxin transcript might be involved in the tight regulation of CD0977.1 production and
432 might be crucial to prevent toxin translation under conditions where expression levels of the
433 toxin gene would be slightly higher than those of the short antitoxin transcript. The c-di-GMP
434 levels would then be critical in this regulation since elevated levels would result in a
435 decreased abundance of the short RCd11 transcript and consequently in growth inhibition.
436 From our previous studies of type I TA pairs, a larger link with biofilm-related control could
437 be suggested for these TA systems since biofilm conditions affected the expression of several
438 other TA transcripts independently from their association with c-di-GMP-responsive
439 riboswitch [27].

440 In this work, we detected a natural background excision of the phiCD630-1 prophage
441 and we identified the phage excisionase gene, *CD0912*. Expression of *CD0912* from a
442 plasmid promoted high levels of prophage excision from the host chromosome, mimicking
443 prophage induction under stressful conditions. While phiCD630-1 and phiCD630-2 of *C.*
444 *difficile* 630 share a large region of duplicated sequence, it is worth noting that *CD0912* is
445 located in the variable region and has no homolog in phiCD630-2. Interestingly, no obvious
446 putative excisionase-encoding gene could be identified in phiCD630-2 although natural
447 excision of this prophage could also be detected in the course of our experiments. Moreover,

448 expression of *CD0912* had no impact on the excision rate of phiCD630-2, suggesting that
449 phiCD630-2 might encode an atypical, yet to be identified excisionase. TA systems have
450 been suggested to play three important biological functions, i.e., stabilization of mobile
451 genetic elements (post-segregational killing), abortive phage infection and persister cell
452 formation [40]. Prophage maintenance is among the suggested function of TA including a
453 recent example of type II TA system that stabilizes prophage in *Shewanella oneidensis* [41] and
454 another type II TA system promoting the maintenance of an integrative conjugative element
455 in *Vibrio cholerae* [42]. This physiological function in prophage stabilization was also
456 suggested for type I TA modules but had never been experimentally demonstrated prior to
457 this study [43]. Prophage excision upon expression of *CD0912* made phiCD630-1 prone to be
458 lost by the host cells and we could thus show that type I TA systems are important to
459 maintain the episomal form of the phage into the host cell. In the *C. difficile* cells expressing
460 the excisionase gene, the frequency of phiCD630-1ΔT4 loss was higher than that of wildtype
461 phiCD630-1 and excision of phiCD630-1 was associated with a strong growth defect, which
462 can be attributed to the post-segregational killing mechanism. The unstable antitoxin is likely
463 degraded in daughter cells where the phage has been lost after cell division upon excision,
464 leading to the toxin production from its stable mRNA and to the growth inhibition of the new
465 cell. Expression of the excisionase gene in cells carrying the prophage devoid of the toxin
466 genes also resulted in a moderate growth defect, suggesting that a supplementary TA system
467 might be present in phiCD630-1 or that the excisionase has an additional function affecting
468 the cell growth. Several experimental conditions were tested in this work to induce the loss of
469 the phage from the cells. Surprisingly, four passages of the strain carrying phiCD630-1 with
470 the intact toxin genes grown in TY broth with constant expression of the excisionase gene
471 from a plasmid resulted in approximately 99% of loss of this prophage (data not shown). This
472 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 [9, 36, 37]. It is therefore tempting to speculate that this TA system 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* strain 630 to the phiCD38-2 prophage [9]. 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 *in silico* search for TA systems in *C. difficile* phages reveals that this 1.9-kb region in phiCD38-2 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 [44, 45], they are likely to be maintained and observed after their transfer by recombination events, even if they bring no selective advantage.

Thus, the large distribution of type I TA modules within *C. difficile* prophages argues in favour of their functional importance for prophage acquisition and transfer between *C. difficile* strains. The position of these TA modules in the extremities of the prophage is also consistent with their role in the entire prophage maintenance. In addition to prophage excision, other genetic events could lead to potential prophage loss such as recombination with other homologous phages. The presence of phage-like elements, cryptic phage rudiments and incomplete prophages in bacterial chromosomes including *C. difficile* genome attests on the frequency of such events. We could hypothesize that TA systems will contribute not only to episomal prophage stability, but also to the maintenance of integrated prophage.

Besides their biological functions, TA modules are also versatile tools for a multitude of purposes in basic research and biotechnology [46]. 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* [47, 48]. In this study, we engineered an inducible counter-selection marker based on the *C. difficile CD2517.1* toxin gene of the *CD2517.1*-RCd8 TA module. Artificial expression of *CD2517.1* from a plasmid in *C. difficile* leads to an immediate interruption of the bacterial growth [27]. Taking advantage of this feature, we generated novel vectors for allele exchange in *C. difficile* 630 (pMSR) and in *C. difficile* ribotype 027 strains and other ribotypes strains (pMSR0). 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 P_{tet} leakiness. In contrast, expression of the RCd8 antitoxin from the pMSR vector was not required since the *CD2517.1*-RCd8 TA module is

523 naturally present within the chromosome of *C. difficile* 630. Native expression of RCd8 was
524 therefore sufficient to prevent CD2517.1 production from the plasmid. Our vectors are
525 derived from those developed by Cartman *et al.*, which use the *codA* gene coding for cytosine
526 deaminase as a counter-selection marker for allelic exchange mutations [32]. However, *codA*-
527 based counter-selection was somewhat ineffective in our hands and false-positive counter-
528 selected colonies with the plasmid still integrated into the chromosome were repeatedly
529 found. This was reported by the authors as the consequence of loss-of-function mutations in
530 genes leading to the bypass of the counter-selection. Our system proved to be much more
531 efficient than all the others we have tested so far, and we did not observe any false-positive
532 clones so far. The false-positive rate could be estimated to less than 0.1% since all counter-
533 selected clones tested during about 50 mutant constructions attempts were thiamphenicol-
534 sensitive, indicative of the plasmid loss. We successfully used this system to construct
535 multiple mutants in various *C. difficile* strains, including the ΔT4 mutant, as well as deletion
536 of a large chromosomal region of 50 kb corresponding to the phiCD630-2 prophage, and
537 gene insertion into the bacterial chromosome (*ermB* gene). We therefore expect these new
538 vectors to become invaluable genetic tools that will foster research in *C. difficile*.

539 **MATERIALS AND METHODS**

540 **Plasmids, bacterial strains construction and growth conditions**

541 *C. difficile* and *Escherichia coli* strains and plasmids used in this study are presented in Table
542 S2. *C. difficile* strains were grown anaerobically (5 % H_2, 5 % CO_2, and 90 % N_2) in TY [49] or
543 Brain Heart Infusion (BHI, Difco) media in an anaerobic chamber (Jacomex). When
544 necessary, cefoxitin (Cfx; 25 µg/ml), cycloserine (Cs; 250 µg/ml) and thiamphenicol (Tm;
545 7.5 µg/ml) were added to *C. difficile* cultures. *E. coli* strains were grown in LB broth, and
546 when needed, ampicillin (100 µg/ml) or chloramphenicol (15 µg/ml) was added to the culture
547 medium. The non-antibiotic analogue anhydrotetracycline (ATc) was used for induction of

548 the P_{tet} promoter of pRPF185 vector derivatives in *C. difficile* [50]. Strains carrying pRPF185
549 derivatives were generally grown in TY medium in the presence of 250 ng/ml ATc and 7.5
550 µg/ml Tm for 7 h, unless stated otherwise. Growth curves were obtained using a GloMax
551 plate reader (Promega).

552 All primers used in this study are listed in Table S3. Details of vector construction are
553 described in the Supplementary materials.

554 The resulting derivative plasmids were transformed into the *E. coli* HB101 (RP4) and
555 subsequently mated with the appropriate *C. difficile* strains (Table S2). *C. difficile*
556 transconjugants were selected by sub-culturing on BHI agar containing Tm (15 µg/ml), Cfx
557 (25 µg/ml) and Cs (250 µg/ml).

558 **Mutagenesis approach and mutant construction**

559 To improve the efficiency of the allele exchange mutagenesis in *C. difficile*, we made use of
560 the inducible toxicity of the CD2517.1 type I toxin that we previously reported [27]. To
561 construct the pMSR vector, used for allele exchange in *C. difficile* 630Δ*erm*, the *codA* gene
562 was removed from the "pseudosuicide" vector pMTL-SC7315 [32] by inverse PCR and
563 replaced by a 1169 bp fragment comprising the entire P_{tet} promoter system and the
564 downstream *CD2517.1* toxin gene. This fragment was amplified from pDIA6319 plasmid [27]
565 and the purified PCR product was cloned into the linearized plasmid. In parallel, the pMSR0
566 vector, for allele exchange in *C. difficile* ribotype 027 strains and other ribotypes, was
567 constructed by removing the *codA* gene from the vector pMTL-SC7215 by inverse PCR and
568 replacing it with the *CD2517.1*-RCd8 TA region with *CD2517.1* under the control of the P_{tet}
569 promoter, as described above, and RCd8 under the control of its own promoter. For deletions,
570 allele exchange cassettes were designed to have between 800 and 1050 bp of homology to the
571 chromosomal sequence in both up- and downstream locations of the sequence to be altered.

The homology arms were amplified by PCR from *C. difficile* strain 630 genomic DNA (Table S3) and purified PCR products were directly cloned into the PmeI site of pMSR vector using NEBuilder HiFi DNA Assembly. To insert P_{thl}-*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 S3). The P_{thl}-*ermB* cassette was amplified from the Clostron mutant *cwp19* [33, 51]. 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. coli* strain NEB10β and all inserts were verified by sequencing. Plasmids were then transformed into *E. coli* HB101 (RP4) and transferred by conjugation into the appropriate *C. difficile* strains. The adopted protocol for allele exchange was similar to that used for the *codA*-mediated allele exchange [32], 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 [52].

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 [53]. Coomassie staining was performed for loading and fractionation controls. Western blotting was performed with anti-HA antibodies (1:2, 000) (Osenses) using standard methods.

Alkaline phosphatase activity assays

C. difficile strains containing the *phoZ* reporter fusions were grown and harvested in exponential growth phase, at the onset of stationary phase and under nutrient starvation conditions. Starvation conditions correspond to a 1 h incubation of exponentially grown cells in PBS buffer at 37 °C in anaerobic conditions. Alkaline phosphatase assays were performed as described previously [54].

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 [55]. 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 [28].

RNA band shift assay

Templates for the synthesis of RNA probes were obtained by PCR amplification using the Term and T7 oligonucleotides (Table S3). The three RNAs (*CD0977.1* toxin mRNA, the short and the long RCd11 antitoxin forms) were synthesized by T7 RNA polymerase and RNA concentrations were monitored by measuring the absorbance at 260 nm. Just before use, *CD0977.1* was also transcribed with (α-^{32}P) UTP yielding uniformly labelled RNA and traces were added to the unlabelled *CD0977.1*. This strategy was used because of the very low efficiency of 5' labelling of all these transcripts. *CD0977.1* transcript was incubated with increasing concentrations of RCd11 short or long RNAs under two different conditions referred as Native and Full RNA duplex conditions as in [27]. The complexes were immediately loaded on native polyacrylamide gels to control for hybridization efficiency. RNA levels were quantified by phosphoimagery.

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 S3), 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.

ACKNOWLEGEMENTS

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, research project № 19-54-15003) to O.S., and a Vernadski fellowship to A.M.

We would like to thank Shonna McBride for the gift of the pMC358 vector and 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 toxin CD0977.1 with the newly identified CD0904.1, CD0956.2 and CD0956.3. The hydrophobic and positively charged amino acids are indicated in red and blue, respectively. (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* 630Δ*erm* 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) A schematic of the *CD0977.1*-RCd11 TA pair genomic region and of the corresponding transcripts as identified by 5'/3'RACE and Northern blot. The Cdi1_4 riboswitch and the identified promoters are represented. The position of the different probes used in the Northern blot experiments is shown. (B) 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 the absence (- ATc) or in the presence (+ ATc) of 250 ng/ml of the inducer ATc. (C) Northern blot of total RNA from *C. difficile* carrying p (empty vector) or p*dccA* (expression of the diguanylate cyclase encoding gene *dccA*) in the presence of 250 ng/ml ATc. (D) Northern blot of total RNA from *C. difficile* 630Δ*erm* 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), the antitoxin (AT RCd11/RCd12) or the Cdi1_4/Cdi1_5 riboswitch (Riboswitch 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. The half-lives for toxin and antitoxin transcripts were estimated from three independent experiments. (E) 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 the 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* (C) was

assessed. (A) and (C) Growth of *C. difficile* 630Δ*erm* 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. Plotted values represent means ± standard deviations ($N = 3$). (B) Alkaline phosphatase activity of the RCd11 promoter::*phoZ* reporter fusions measured after 4 (exponential) and 10 h (stationary) of growth in TY broth or under nutrient starvation conditions. A schematic of the *CD0977.1*-RCd11 TA pair genomic region and of the locations and sizes of promoter fragments constructed for the *phoZ* reporter fusions is shown. Values represent means ± standard deviations ($N = 4$). ** $P \leq 0.01$, *** $P \leq 0.001$ and **** $P \leq 0.0001$ by a two-way ANOVA followed by a Dunnett's or Tukey's multiple comparison test.

Figure 4. Impact of TA 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* 630Δ*erm* was determined after seven passages (every 12 hours) in TY broth without thiamphenicol. Values represent means ± standard deviations ($N = 3$). ** $P \leq 0.05$ and *** $P \leq 0.001$ by an unpaired *t* test.

Figure 5. Impact of TA 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 h were plated on erythromycin-containing agar plates and cells that lost the prophage were selectively killed. (B) Strains ΔphiCD630-2 phiCD630-1::*erm* and ΔphiCD630-2 phiCD630-1-ΔT4::*erm* carrying a vector control or pDIA6867 (overproducing the excisionase CD0912) were inoculated at an initial optical

density at 600 nm (OD_{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 and bacterial growth was determined by measurement of the OD_{600nm} after 16 h and 24 h. ** $P \leq 0.01$ and **** $P \leq 0.0001$ by a two-way ANOVA followed by a Tukey's multiple comparison test. (C) Maintenance of prophages in strains ΔphiCD630-2 phiCD630-1::*erm* and ΔphiCD630-2 phiCD630-1-ΔT4::*erm* carrying pDIA6867 after 24 h of growth as in (B) was quantified by plating serial dilutions on agar plates supplemented or not with 2.5 µg/ml Erm.. Values represent means ± standard deviations ($N = 3$). *** $P \leq 0.001$ 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 *CD0904.1*-RCd13 (A) and *CD0956.3*-RCd14 (B) and *CD0977.1*-RCd11 (C) TA 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. Potential processing site is shown in (C) by vertical green arrow. 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 promoter region of RCd14 antitoxin was deduced from the alignment in (D). (D) Alignment between the sequences of *CD0904.1*-RCd13 (top line) and *CD0956.3*-RCd14 (bottom line) using EMBOSS Needle. Green boxes show the open reading frames of *CD0904.1* and *CD0956.3* and green arrows indicate the direction of the genes. Blue boxes show the -10 and the -35 boxes of the promoter regions and black boxes show the transcription start site of RCd13 and RCd14. In (A), (B) and (C), the red boxes show the Sigma A-dependent promoter -10 and -35 elements and the yellow boxes show the Sigma B-dependent promoter element.

Figure S2. Impact of toxin CD0977.1 expression on cell viability and morphology. Growth (A) and viability (B) of *C. difficile* 630Δ*erm* strain carrying the pRPF185-based plasmids (empty: p or with *CD0977.1* toxin gene under the control of the P_{tet} 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 ($N = 3$). * $P \leq 0.05$ 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

788 of cells with length above 2 standard deviations relative to the 630/p control strain mean
789 length.

790 **Figure S3. Sequence of type I *CD0977.1*-RCd11 TA locus in *C. difficile*.** The thick
791 horizontal arrows below the double-stranded sequences show the toxin and antitoxin
792 transcripts and the direction of transcription. The transcriptional start sites for sense and
793 antisense transcripts identified by 5'/3'RACE and TSS mapping are indicated by vertical
794 arrows with their genomic location. Line thickness corresponds to the proportion of observed
795 extremities. The genomic location of 5'- and 3'-ends of the transcripts are indicated above the
796 sequence. Potential processing site is shown by vertical green arrow. The inverted repeats at
797 the position of transcriptional terminators are indicated by thin black arrows. The positions of
798 Sigma A-dependent -10 and -35 promoter elements of antitoxin (AT) are shown in red boxes.
799 The positions of Sigma A-dependent -10 and -35 elements promoter, ribosome binding site,
800 translation initiation codon and stop codon of toxin (T) mRNA are shown in blue boxes. The
801 positions of Sigma B-dependent promoter elements are shown in green boxes for both TA
802 genes.

803 **Figure S4. Analysis of TA RNA duplex formation by RNA band shift assay.**
804 Radiolabeled *CD0977.1* transcript was incubated with increasing concentrations of RCd11
805 short or long RNAs under two different conditions referred as native and full RNA duplex
806 conditions. Native *CD0977.1*:RCd11 complexes were formed at 37 °C for 5 min in TMN
807 buffer, and full duplexes were obtained after a denaturation-annealing treatment in TE Buffer
808 (2 min 90°C, 30 min 37°C). The complexes were immediately loaded on native
809 polyacrylamide gels to control for hybridization efficiency. RNA levels were quantified by
810 phosphoimagery and complex proportion is indicated.

Figure S5. 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. The half-lives for toxin and antitoxin transcripts have been estimated from three independent experiments.

Figure S6. 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* (C) was assessed. (A) and (C) Growth of *C. difficile* 630Δ*erm* 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. (B) Alkaline phosphatase activity of the RCd11 promoter P_1 + disrupted P_2::*phoZ* reporter fusions and the promoterless *phoZ* measured after 4 h (stationary) of growth in TY broth. Values represent means ± standard deviations ($N = 3$).

Figure S7. Nucleotide alignment of antitoxins. (A) Nucleotide alignment of RCd11 and RCd12 using LALIGN. (B) Nucleotide alignment of RCd11 and RCd10 using LALIGN.

Figure S8. Secondary structure prediction of *CD0977.1* mRNAs and corresponding antitoxin RCd11 RNAs. The RNA secondary structure predictions were performed by Mfold software. The predictions for long and short forms of RCd11 antitoxins are shown. Ribosome binding site (SD), translation initiation codon and stop codon positions are highlighted. The positions of mismatches in RCd12 AT sequence are indicated.

Figure S9. 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* 630Δ*erm* 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 S10. Deletion of the phiCD630-2 prophage from *C. difficile* 630Δ*erm* using the newly developed allele exchange method. (A) Schematic representation of phiCD630-2 in *C. difficile* 630Δ*erm*. The location of primers used to screen for mutants is represented. (B) PCR products amplified using the indicated primers from the parental strain 630Δ*erm* (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 S11. 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 S12. Impact of excisionase (CD0912) overproduction on excision of phiCD630-1 and phiCD630-1-ΔT4. The frequency of prophage excision was estimated by quantitative PCR as described in Materials and Methods section. 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 ($N = 3$).

Figure S13. 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).

Figure S14. 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. Red dot with tag "small variant 1" indicates CD0904.1. Red dot with tag "small variant 2" indicates CD0956.3.

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. Sebaihia M*, et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet* **38**, 779-786 (2006).

7. Fortier LC. Bacteriophages Contribute to Shaping *Clostridioides (Clostridium) difficile* Species. *Front Microbiol* **9**, 2033 (2018).

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

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

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

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

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

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

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

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

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

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

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

19. 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).

20. Wen Y, Behiels E, Devreese B. Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. *Pathogens and disease* **70**, 240-249 (2014).

21. Yamaguchi Y, Inouye M. Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat Rev Microbiol* **9**, 779-790 (2011).

22. Song S, Wood TK. Toxin/Antitoxin System Paradigms: Toxins Bound to Antitoxins Are Not Likely Activated by Preferential Antitoxin Degradation. *Advanced biosystems* **4**, e1900290 (2020).

23. Trastoy R*, et al.* Mechanisms of Bacterial Tolerance and Persistence in the Gastrointestinal and Respiratory Environments. *Clinical microbiology reviews* **31**, (2018).

24. Goormaghtigh F*, et al.* Reassessing the Role of Type II Toxin-Antitoxin Systems in Formation of *Escherichia coli* Type II Persister Cells. *mBio* **9**, (2018).

25. Fraikin N, Goormaghtigh F, Van Melderen L. Type II Toxin-Antitoxin Systems: Evolution and Revolutions. *Journal of bacteriology* **202**, (2020).

26. Jurenas D, Van Melderen L. The Variety in the Common Theme of Translation Inhibition by Type II Toxin-Antitoxin Systems. *Frontiers in genetics* **11**, 262 (2020).

27. 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).

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

29. Fozo EM, Makarova KS, Shabalina SA, Yutin N, Koonin EV, Storz G. Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. *Nucleic acids research* **38**, 3743-3759 (2010).

30. Soutourina O. Type I Toxin-Antitoxin Systems in Clostridia. *Toxins* **11**, (2019).

31. Sudarsan N*, et al.* Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321**, 411-413 (2008).

32. 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).

33. 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).

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

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

36. 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).

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

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

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

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

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

42. Wozniak RA, Waldor MK. A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS genetics* **5**, e1000439 (2009).

43. Brantl S, Muller P. Toxin-Antitoxin Systems in *Bacillus subtilis*. *Toxins* **11**, (2019).

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

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

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

47. 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).

48. 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).

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

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

51. 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).

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

53. 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).

54. Edwards AN, Anjuwon-Foster BR, McBride SM. RstA Is a Major Regulator of *Clostridioides difficile* Toxin Production and Motility. *mBio* **10**, (2019).

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

Fig. 1

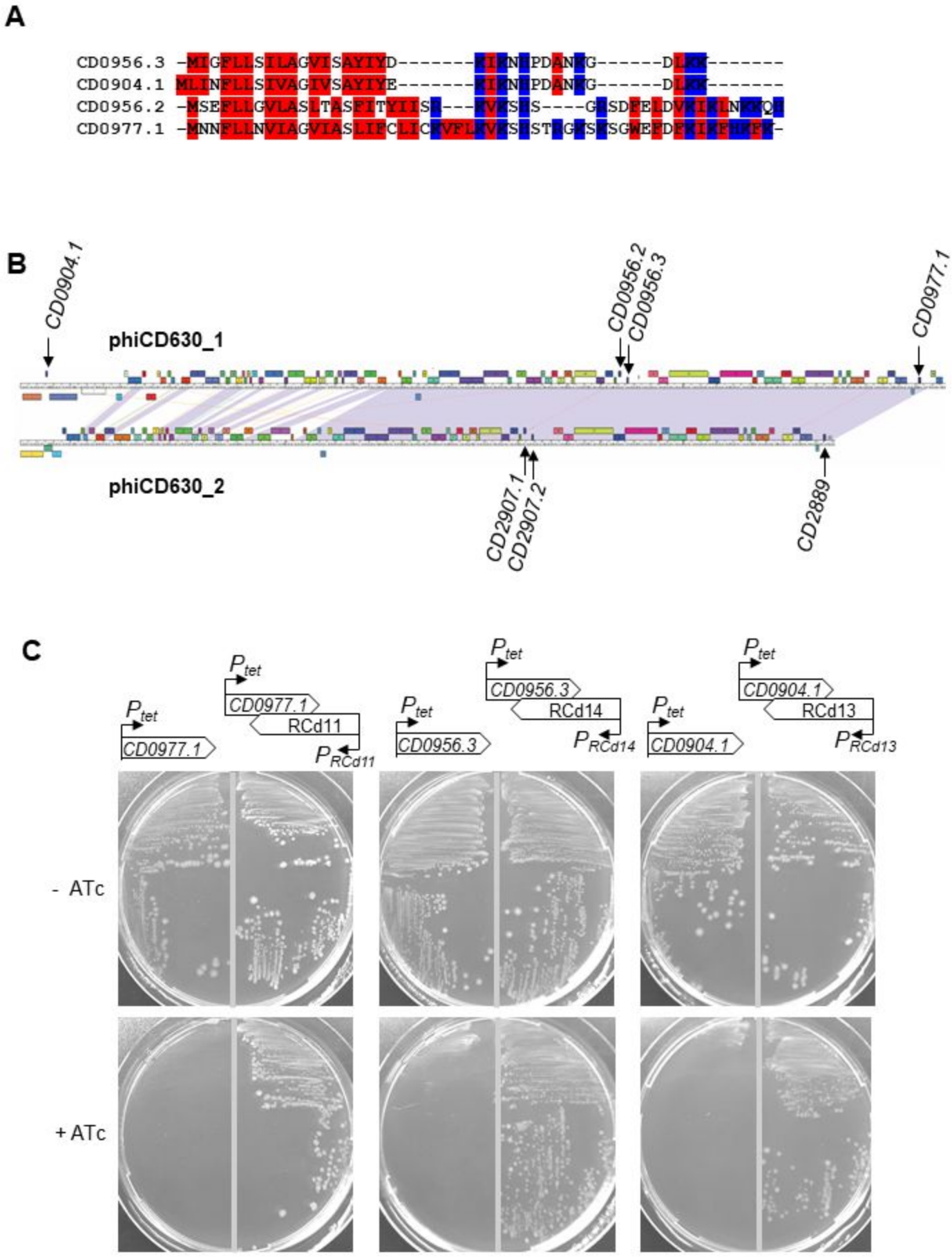

A
CD0956.3
CD0904.1
CD0956.2
CD0977.1
B
CD0904.1
phiCD630_1
CD0956.2
CD0956.3
CD0977.1
phiCD630_2
CD2907.1
CD2907.2
CD2889
C
Ptet
CD0977.1
RCd11
PRCd11
CD0956.3
RCd14
PRCd14
CD0904.1
RCd13
PRCd13
- ATc
+ ATc

Fig. 2

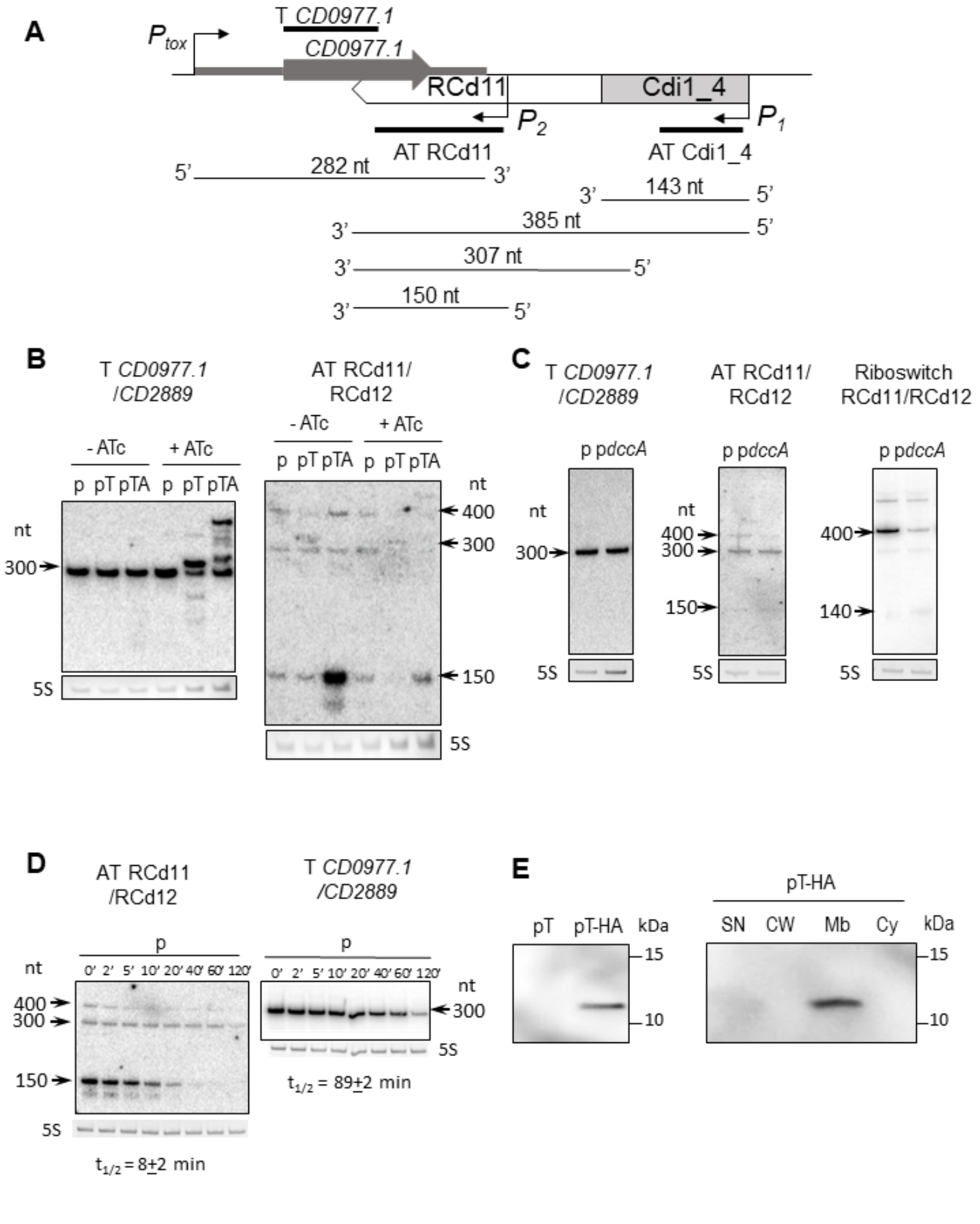

A
P tox
T CD0977.1
CD0977.1
RCd11
Cdi1_4
P2
P1
AT RCd11
AT Cdi1_4
5'
282 nt
3'
3'
143 nt
5'
3'
385 nt
5'
3'
307 nt
5'
3'
150 nt
5'
B
T CD0977.1 /CD2889
AT RCd11/ RCd12
- ATc
+ ATc
p pT pTA p pT pTA
nt
300
400
150
5S
C
T CD0977.1 /CD2889
AT RCd11/ RCd12
Riboswitch RCd11/RCd12
p pdccA
nt
300
400
150
140
5S
D
AT RCd11 /RCd12
T CD0977.1 /CD2889
p
0' 2' 5' 10' 20' 40' 60' 120'
nt
400
300
150
5S
t1/2 = 8±2 min
t1/2 = 89±2 min
E
pT pT-HA kDa
SN CW Mb Cy kDa
pT-HA
15
10

Fig. 3

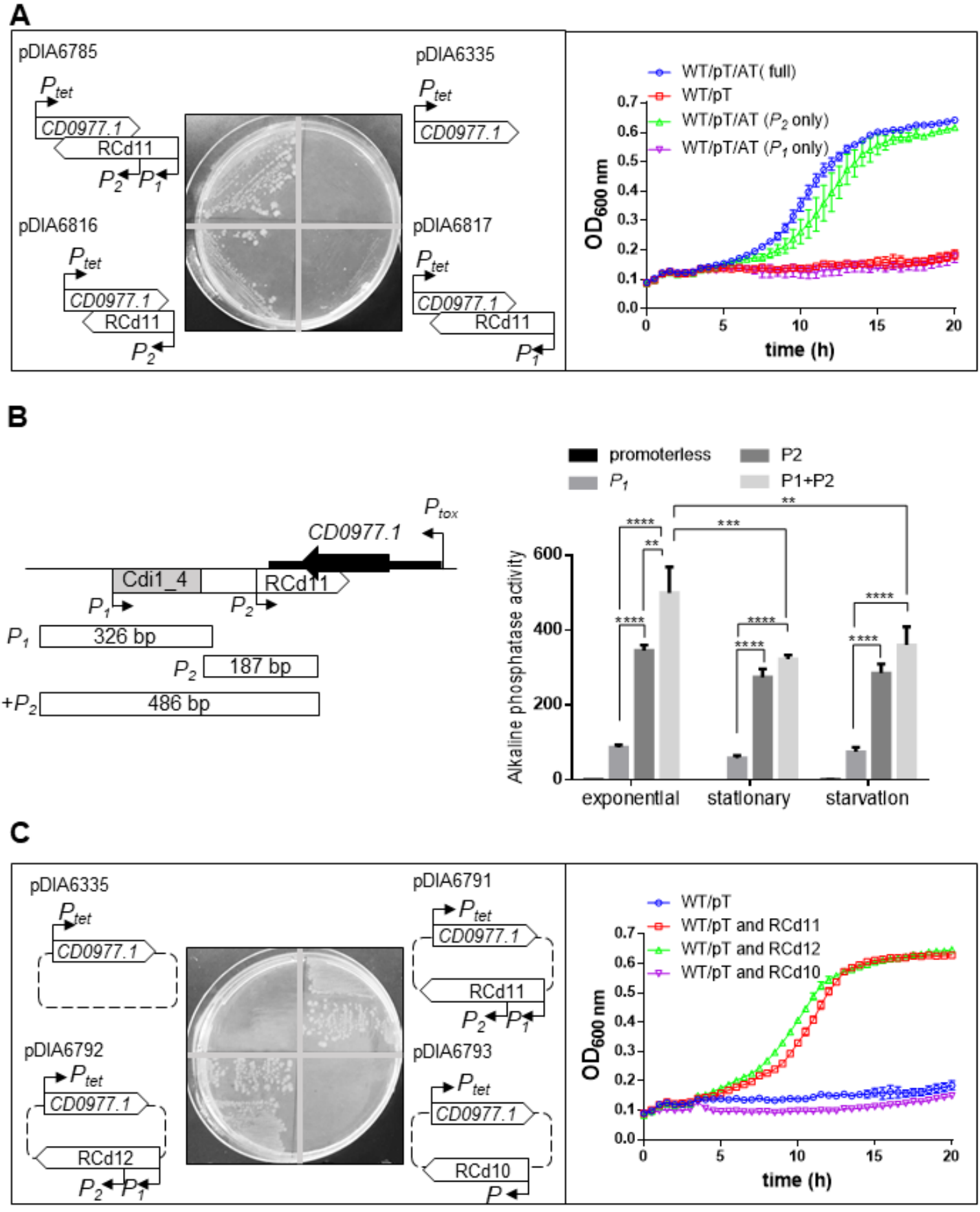

A
pDIA6785
P_{tet}
CD0977.1
RCd11
P_2 P_1
pDIA6335
P_{tet}
CD0977.1
pDIA6816
P_{tet}
CD0977.1
RCd11
P_2
pDIA6817
P_{tet}
CD0977.1
RCd11
P_1
WT/pT/AT(full)
WT/pT
WT/pT/AT (P_2 only)
WT/pT/AT (P_1 only)
$OD_{600\ nm}$
time (h)
B
P_{tox}
CD0977.1
Cdi1_4
RCd11
P_1
P_2
P_1 326 bp
P_2 187 bp
P_1+P_2 486 bp
promoterless
P2
P_1
P1+P2
Alkaline phosphatase activity
exponential
stationary
starvation
C
pDIA6335
P_{tet}
CD0977.1
pDIA6791
P_{tet}
CD0977.1
RCd11
P_2 P_1
pDIA6792
P_{tet}
CD0977.1
RCd12
P_2 P_1
pDIA6793
P_{tet}
CD0977.1
RCd10
P
WT/pT
WT/pT and RCd11
WT/pT and RCd12
WT/pT and RCd10
$OD_{600\ nm}$
time (h)

Fig. 4

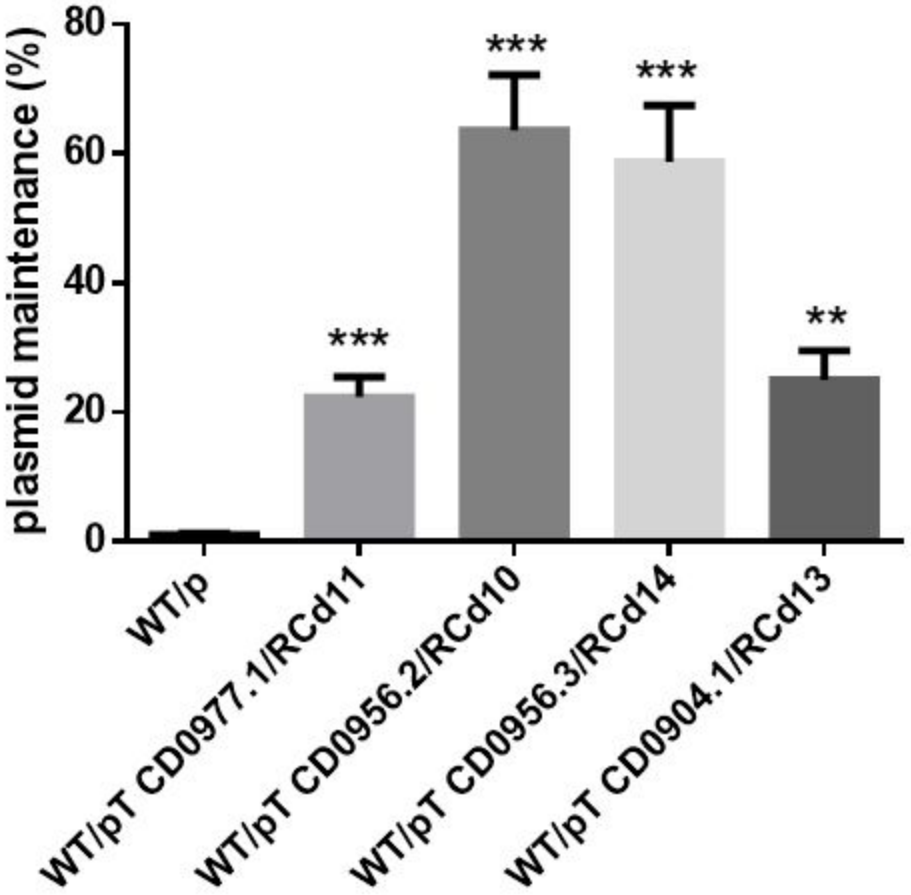

plasmid maintenance (%)
80
60
40
20
0
***
***
***
**
WT/p
WT/pT CD0977.1/RCd11
WT/pT CD0956.2/RCd10
WT/pT CD0956.3/RCd14
WT/pT CD0904.1/RCd13

Fig. 5

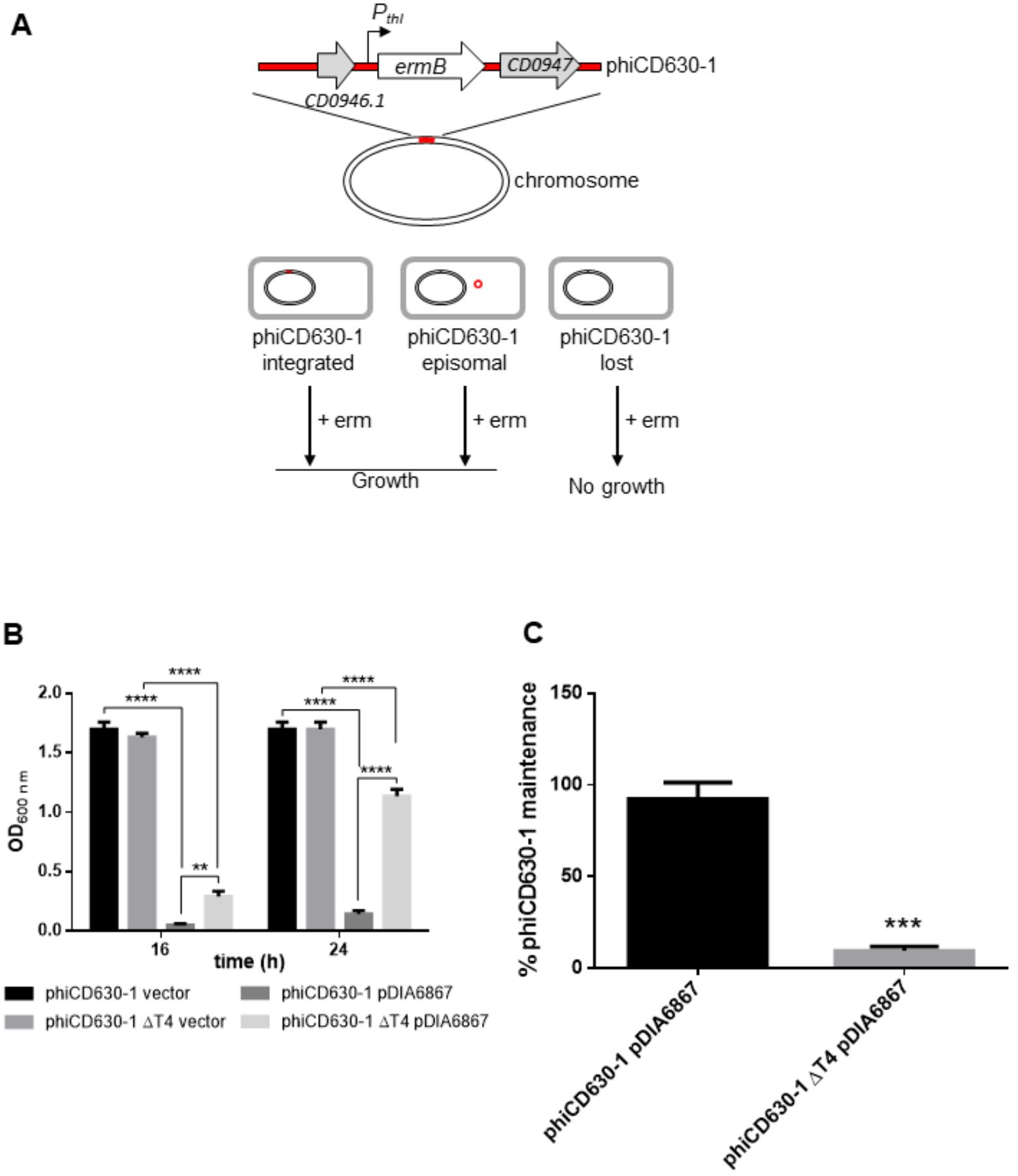
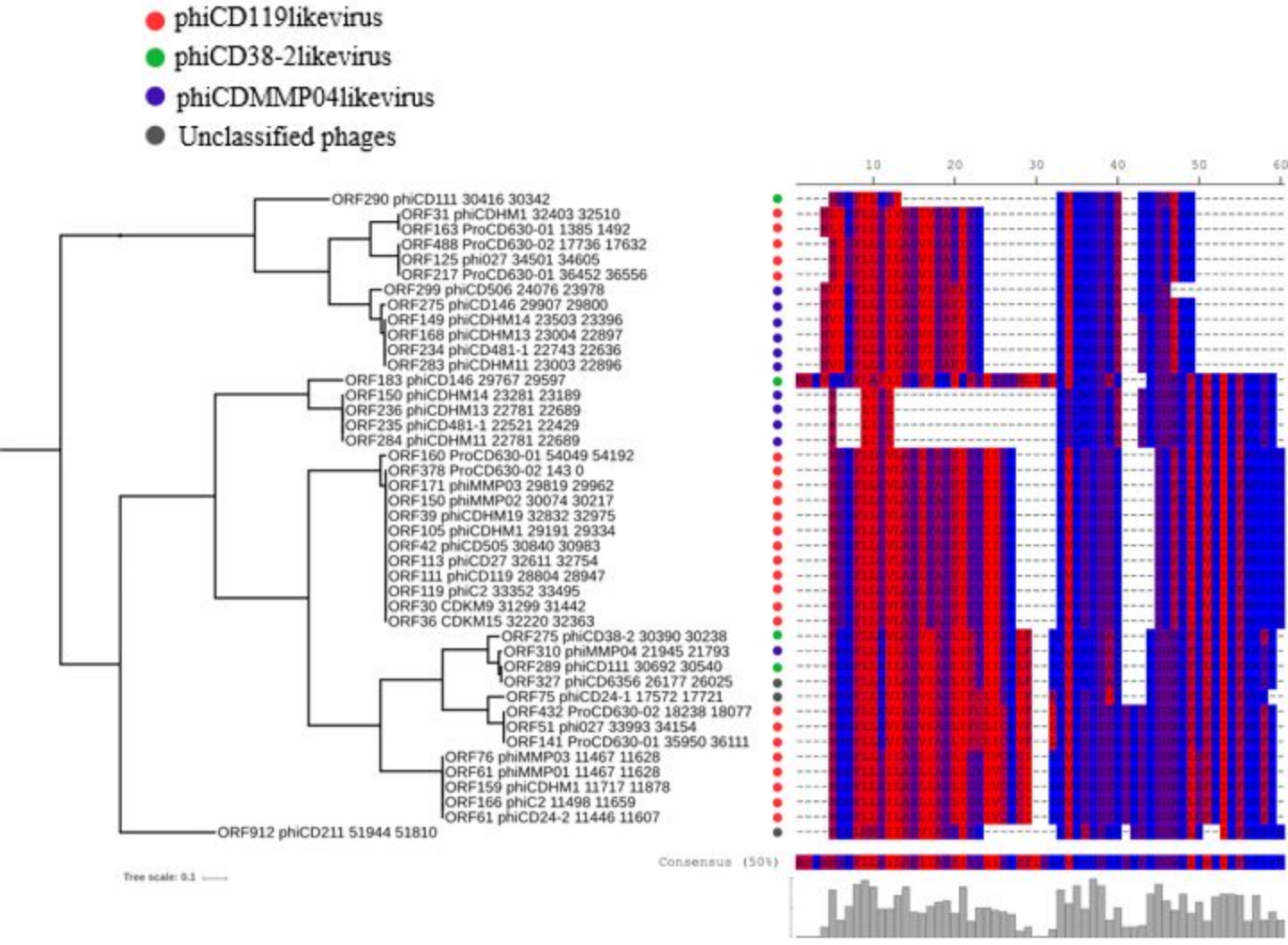

A
P_{thl}
ermB
CD0947
phiCD630-1
CD0946.1
chromosome
phiCD630-1 integrated
phiCD630-1 episomal
phiCD630-1 lost
+ erm
+ erm
+ erm
Growth
No growth
B
$OD_{600\ nm}$
time (h)
16
24
phiCD630-1 vector
phiCD630-1 ΔT4 vector
phiCD630-1 pDIA6867
phiCD630-1 ΔT4 pDIA6867
C
% phiCD630-1 maintenance
phiCD630-1 pDIA6867
phiCD630-1 ΔT4 pDIA6867

Fig. 6


phiCD119likevirus
phiCD38-2likevirus
phiCDMMP04likevirus
Unclassified phages
ORF290 phiCD111 30416 30342
ORF31 phiCDHM1 32403 32510
ORF163 ProCD630-01 1385 1492
ORF488 ProCD630-02 17736 17632
ORF125 phi027 34501 34605
ORF217 ProCD630-01 36452 36556
ORF299 phiCD506 24076 23978
ORF275 phiCD146 29907 29800
ORF149 phiCDHM14 23503 23396
ORF168 phiCDHM13 23004 22897
ORF234 phiCD481-1 22743 22636
ORF283 phiCDHM11 23003 22896
ORF183 phiCD146 29767 29597
ORF150 phiCDHM14 23281 23189
ORF236 phiCDHM13 22781 22689
ORF235 phiCD481-1 22521 22429
ORF284 phiCDHM11 22781 22689
ORF160 ProCD630-01 54049 54192
ORF378 ProCD630-02 143 0
ORF171 phiMMP03 29819 29962
ORF150 phiMMP02 30074 30217
ORF39 phiCDHM19 32832 32975
ORF105 phiCDHM1 29191 29334
ORF42 phiCD505 30840 30983
ORF113 phiCD27 32611 32754
ORF111 phiCD119 28804 28947
ORF119 phiC2 33352 33495
ORF30 CDKM9 31299 31442
ORF36 CDKM15 32220 32363
ORF275 phiCD38-2 30390 30238
ORF310 phiMMP04 21945 21793
ORF289 phiCD111 30692 30540
ORF327 phiCD6356 26177 26025
ORF75 phiCD24-1 17572 17721
ORF432 ProCD630-02 18238 18077
ORF51 phi027 33993 34154
ORF141 ProCD630-01 35950 36111
ORF76 phiMMP03 11467 11628
ORF61 phiMMP01 11467 11628
ORF159 phiCDHM1 11717 11878
ORF166 phiC2 11498 11659
ORF61 phiCD24-2 11446 11607
ORF912 phiCD211 51944 51810
Tree scale: 0.1
10
20
30
40
50
60
Consensus (50%)