A helicase-containing module defines a family of pCD630-like plasmids in *Clostridium difficile*

Wiep Klaas Smits[1,2,*], J. Scott Weese[3], Adam P. Roberts[4,5],
Céline Harmanus[1], Bastian Hornung[1]

[1] Department of Medical Microbiology, Leiden University Medical Center, PO Box 9600, 2300RC, Leiden, The Netherlands

[2] Netherlands Centre for One Health

[3] Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada

[4] Department of Parasitology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

[5] Research Centre for Drugs and Diagnostics, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

Running title: pCD630 plasmids in *C. difficile*

* Corresponding author: Department of Medical Microbiology, Leiden University Medical Center, PO Box 9600, 2300RC, Leiden, The Netherlands
phone: +31-71-526 1229
fax: +31-71-526 6761
E-mail: w.k.smits@lumc.nl

Abstract

Clostridium difficile is a Gram-positive and sporulating enteropathogen that is a major cause of healthcare-associated infections. Even though a large number of genomes of this species have been sequenced, only a few plasmids have been described in the literature. Here, we use a combination of *in silico* analyses and laboratory experiments to show that plasmids are common in *C. difficile*. We focus on a group of plasmids that share similarity with the plasmid pCD630, from the reference strain 630. The family of pCD630-like plasmids is defined by the presence of a conserved putative helicase that is likely part of the plasmid replicon. This replicon is compatible with at least some other *C. difficile* replicons, as strains can carry pCD630-like plasmids in addition to other plasmids. We find two distinct sub-groups of pCD630-like plasmids that differ in size and accessory modules. This study is the first to describe a family of plasmids in *C. difficile*.

Keywords: Plasmid, replicon, helicase, replication

44 **Introduction**

45

46 *Clostridium difficile* (*Clostridioides difficile* [1]) is a Gram-positive, anaerobic and spore-
47 forming bacterium that can asymptomatically colonize the human gut [2]. It is ubiquitous
48 in the environment, and can also be found in the gastrointestinal tract of many animals.
49 The bacterium gained notoriety when it was identified as the causative agent of health
50 care associated diarrhea, and is increasingly implicated in community-associated
51 disease in many countries [2]. In hosts with a dysbiosis of the microbiome, such as
52 patients treated with broad-spectrum antimicrobials, conditions are favorable for *C.*
53 *difficile* germination and outgrowth [3]. *C. difficile* produces one or more toxins, that
54 cause symptoms ranging from diarrhea to potentially fatal toxic megacolon [2, 4].
55 Over the past two decades, genetic studies have of *C. difficile* have become
56 possible due to the generation of shuttle plasmids that can be transferred by
57 conjugation from *Escherichia coli* to *C. difficile* [5]. These plasmids mostly employ a
58 replicon derived from plasmid pCD6 for replication in *C. difficile* [6]. In 2006, the first
59 genome sequence of *C. difficile* became available, revealing the presence of another
60 plasmid, pCD630 [7].
61 Despite a great number of strains having been whole genome sequenced since
62 then, plasmid biology of *C. difficile* has been poorly explored. One reason is that
63 plasmid content is variable, and most studies on the evolution and/or transmission of *C.*
64 *difficile* focus on those genes conserved between all strains (the core genome) [8-11].
65 However, there is reason to assume that plasmids are common in *C. difficile*; for
66 instance, before the advent of the currently common typing schemes [12], plasmid

67 isolation had been proposed as an epidemiological tool [13]. The ratio of plasmid-
68 containing to plasmid-free strains in this study was found to be approximately 1:2,
69 suggesting that around 30% of strains of *C. difficile* may carry a plasmid. Furthermore,
70 hybridization-based analyses of total DNA from a collection of *C. difficile* strains suggest
71 the presence of DNA with significant similarity to pCD630 open reading frames (ORFs)
72 [14, 15].
73 Here, we define a family of plasmids that share a conserved helicase-containing
74 module and demonstrate that these plasmids are common in a diverse set of *C. difficile*
75 strains.
76
77 **Materials and methods**
78
79 *Strains and growth conditions*
80 Strains used in this study are listed in **Table 1**. For DNA isolation, strains were grown on
81 *Clostridium difficile* agar (CLO) plates (bioMérieux) and a single colony was inoculated
82 into pre-reduced brain-heart infusion broth (Oxoid) supplemented with 0.5 % w/v yeast
83 extract (Sigma-Aldrich) and *Clostridium difficile* selective supplement (Oxoid). Strains
84 were PCR ribotyped [16] in-house for this study. For some strains a PCR ribotype could
85 not be assigned (**Table 1**).
86
87 *Isolation of plasmid DNA from C. difficile*
88 Plasmids were isolated from 2 mL of *C. difficile* overnight culture using NucleoSpin
89 Plasmid Easypure columns (Macherey-Nagel); to increase yield, 10mg/mL lysozyme to

90 buffer A1 was added, as recommended by the manufacturer. Using PCR and
91 sequencing, we found that the DNA isolated using this kit is heavily contaminated with
92 chromosomal DNA. To isolate pure plasmid DNA, aliquots of the DNA were incubated
93 with PlasmidSafe ATP-dependent DNase (Epicentre) that digests linear, but not circular
94 double stranded DNA. After purification with a Nucleospin Gel and PCR Clean-up kit
95 (Macherey-Nagel), the absence of genomic DNA was confirmed by PCR using primers
96 targeting *gluD* (**Table 2**). Yields were generally very low, but the plasmid was readily
97 detectable by PCR.
98
99 *Reannotation of pCD630 and identification of a pCD630-like plasmid*
100 The pCD630 sequence was obtained from GenBank (AM180356.1). CDP01 and
101 CDP11 form a single open reading frame (ORF) and were treated as a single ORF in
102 our analyses. Protein sequences encoded by the ORFs of pCD630 were used as
103 BLAST queries against the non-redundant protein sequences database, limited to
104 taxid:1496 (*Clostridium difficile*). This identified the 8089 bp *Peptoclostridium difficile*
105 genome assembly 7032985, scaffold BN1096_Contig_85 (LK932541.1). To reconstitute
106 the plasmid from this contig, the DNA was circularized and a single copy of the 98 bp
107 direct repeat that was present at the terminus of the original contig was removed using
108 Geneious R10. The resulting 7991 bp sequence now encodes a full copy of a sequence
109 homologous to CDP07 of pCD630. Reannotation of plasmids was performed using
110 using an in-house pipeline. This pipeline incorporates the gene caller Prodigal (version
111 2.6.3) [17], RNAmmer (version 1.2) [18], Aragorn (version 1.2.38) [19], the CRISPR
112 recognition tool (version 1.2) [20], dbCAN (version 5.0) [21] and PRIAM (version March

113 2015) [22]. The plasmid derived from LK932541 was submitted to GenBank as pCD-
114 ISS1 (GenBank: pending accessioning).
115
116 *Identification of pCD630-like plasmids in short read archives*
117 In order to identify other pCD630-like plasmids in sequence databases, paired end
118 Illumina sequences from study PRJEB2101 (ERR017367-ERR017371, ERR022513,
119 ERR125908-ERR125911) were downloaded from the short read archive of the
120 European Nucleotide Archive (ENA). Short reads were assembled and visualized in
121 PLACNETw [23] to determine likely replicons. The contigs corresponding to pCD630-
122 like plasmids were downloaded and imported into Geneious R10 software (Biomatters
123 Ltd) for circularization and removal of terminal repeats; afterwards all plasmids which
124 could be circularized were compared with BLASTN (version 2.40) [24] to pCD630 and
125 the sequences were restructured to start at the base corresponding to base 2903 in
126 pCD630. Afterwards the plasmids were annotated using the in-house pipeline as
127 described above, and submitted to GenBank as pCD-WTSI1 (GenBank: MG019959),
128 pCD-WTSI2 (GenBank: MG019960), pCD-WTSI3 (GenBank: MG019961), pCD-WTSI4
129 (GenBank: MG019962). Alignments of the pCD630-like plasmids were performed in
130 Geneious R10 (Biomatters Ltd) and the alignment figure was prepared using Adobe
131 Illustrator CC (Adobe Systems Inc).
132
133 *Polymerase chain reaction*
134 Oligonucleotides used in this work are listed in **Table 2**. To confirm the presence of
135 pCD630 in derivatives of *C. difficile* strain 630, PCR was performed with oWKS-1629

136 and oWKS-1630 (targeting CDP04); oWKS-1631 and oWKS-1632 (targeting CDP07);
137 oWKS-1633 and oWKS-1634 (targeting CDP10). As a control for chromosomal DNA, a
138 PCR was performed targeting the *gluD* gene that is used as a target for *C. difficile*
139 identification, using primers oWKS-1070 and oWKS-1071. To screen a collection of *C.*
140 *difficile* strains for the presence of pCD630-like plasmids, a PCR was performed with
141 primers oWKS-1651 and oWKS-1652 that targets a region of CDP07 conserved among
142 the 6 full length plasmids identified in this work. Fragments were separated on 0.5x
143 TAE (20 mM Tris, 10mM acetic acid, 0.5mM EDTA) agarose, stained with ethidium
144 bromide and imaged on a Gel Doc XR system (BioRad). Images were captured using
145 QuantityOne (BioRad) and prepared for publication using Adobe Photoshop CC (Adobe
146 Systems Inc) and CorelDRAW X8 (Corel Corporation).
147

148 **Results and discussion**

149

150 *pCD630 is present in C. difficile strain 630 and some of its derivatives*
151 The most commonly used laboratory strains of *C. difficile* are derived from the reference
152 strain 630 (PCR ribotype 012 [7]) by serial passaging and screening for loss of
153 resistance to the antimicrobial erythromycin [25]). It was generally assumed that during
154 this passaging the plasmid pCD630 that is present in strain 630 was lost. Indeed, our *de*
155 *novo* assembly of the 630Δ*erm* genome sequence using single molecule real-time
156 (SMRT) sequencing did not report the presence of the plasmid [26]. Recently, however,
157 one study showed the presence of reads mapping to pCD630 in a genome
158 resequencing project of another isolate of 630Δ*erm* [27]. This prompted us to revisit our

159 whole genome sequencing data (ENA:PRJEB7326). If the plasmid was maintained in
160 630Δ*erm*, we expected to be able to find reads mapping back to the pCD630 reference
161 sequence (GenBank: AM180356.1) in this dataset. Indeed, when we performed a
162 reference assembly of the short reads (ENA: ERR609091) against the pCD630, we
163 found approximately 0.8% of the reads mapping to the plasmid. The original *de novo*
164 assembly overlooked the plasmid due to a low number of plasmid-mapping reads as the
165 result of a size fractionation step (the plasmid is <8kb, and SMRT sequencing was
166 performed on high MW DNA). Notably, both a *de novo* assembly of the plasmid based
167 on a small number of SMRT reads, as well as the reference assembly using a large
168 number of Illumina reads shows a 100% congruence with the published reference
169 sequence for pCD630 (data not shown). This indicates that, despite the lack of selective
170 pressure and repeated culturing under laboratory conditions, the plasmid has remained
171 unchanged.

172 We confirmed the presence of pCD630 and the extrachromosomal nature of the
173 plasmid. To do so, we performed a miniprep on a *C. difficile* liquid culture and treated
174 the resulting DNA with PlasmidSafe DNase, that selectively removes linear double
175 stranded (sheared) but not circular DNA. A PCR using primers against three ORFs of
176 pCD630 (*cdp04, cdp07 and cdp10*) and one chromosomal locus (*gluD*) showed that the
177 DNase treated samples were negative for the *gluD* PCR, but positive for all three
178 plasmid loci (**Figure 1A**).

179 The results above suggest that pCD630 is stably maintained
180 extrachromosomally. Next, we wanted to verify the presence of the plasmid in multiple
181 derivatives of strain 630, to see if plasmid-loss could be documented. We previously

182 analyzed 630Δ*erm* from our laboratory as well as from the laboratory where it was
183 generated to determine the chromosomal location of the mobile element CTn*5*, in
184 comparison with the ancestral strain 630 and the independently derived 630E strain
185 [26]. We found that pCD630 was readily detected on total genomic DNA from all these
186 strains, with the exception of the 630E isolate in our collection (**Figure 1B**). 630E and
187 630Δ*erm* demonstrate notable phenotypic differences [25, 28] and we wondered
188 whether these might be in part due to loss of the pCD630 plasmid. We performed a
189 reference assembly using the whole genome sequencing data available from the study
190 by Collery *et al* (ENA: PRJNA304508), that compares 630Δ*erm* and 630E [28]. The
191 assembly showed that both these strains contain pCD630 and indicate that the loss of
192 plasmid is not a general feature of 630E strains. We conclude that the observed
193 phenotypic differences are not likely due to loss of the plasmid. It was reported that the
194 isolate of *C. difficile* 630 stored at in the collection of the DSMZ (www.dsmz.de) lacks
195 the pCD630 plasmid [27, 29]. We requested both 630 (DSMZ 26845) and 630Δ*erm*
196 (DSMZ 27543) and checked for the presence of the plasmid. Our results confirm the
197 absence of pCD630 from DSMZ 26485 (**Figure 1B**), in line with the analysis of
198 Dannheim *et al* [27].

199 In other organisms, the presence of certain replicons can negatively affect the
200 maintenance of other replicons (plasmid incompatibility); this has not been documented
201 for *C. difficile* to date. If pCD630 would be incompatible with other replicons (such as the
202 pCB102 and pCD6) [5, 30], this could result in loss of the pCD630 plasmid in genetically
203 modified *C. difficile*. We therefore tested whether pCD630 was lost in strains
204 chromosomally modified using Clostron mutagenesis [30, 31], Allele Coupled Exchange

205 [32, 33] or carrying a replicative plasmid [34, 35]. We found that all of these carried
206 pCD630, suggesting that pCD630 is compatible with pCB102 and pCD6-based
207 replicons (**Figure 1C**). Similar results were obtained with multiple mutants (data not
208 shown).

209 Together, our data clearly shows that pCD630 persists in the absence of
210 selection, but also that pCD630 can be lost. Thus, care should be taken to verify
211 plasmid content when comparing presumed isogenic laboratory strains even when they
212 are derived from the same isolate.

213

214 *A pCD630-like plasmid is present in a strain with reduced metronidazole susceptibility*
215 We wondered whether there are more pCD630-like plasmids. As a first step, we set out
216 to identify coding sequences with homology to pCD630 ORFs in GenBank. Using
217 default settings, we identified a single 8089 bp contig that encodes proteins with
218 homology to CDP01, CDP04-6 and CDP08-11 (*Peptoclostridium difficile* genome
219 assembly 7032985, scaffold BN1096_Contig_85; GenBank: LK932541) (**Figure 2**).

220 This sequence stems from a study that compares three non-toxigenic PCR
221 ribotype 010 strains of *C. difficile*, with differing susceptibility to metronidazole [36].
222 Strain 7032985 was classified as intermediate resistant to metronidazole. If we assume
223 that the contig represents a pCD630-like plasmid, we expect DNA from this strain to
224 remain positive in a PCR that targets the plasmid upon treatment with PlasmidSafe
225 DNase. We found that the PCR targeting *cdp07*, but not chromosomal locus *gluD,*
226 results in a clear signal when using a template treated with PlasmidSafe DNase (**Figure**

227 **1D**). Having confirmed that the contig is extrachromosomal in nature, we will refer to the
228 putative plasmid as pCD-ISS1 hereafter (**Table 3**).
229 To further analyze pCD-ISS1, we circularized the LK932451 contig to yield a
230 putative plasmid of 7991bp, performed an automated annotation (GenBank: pending
231 accessioning) and compared the annotated pCD-ISS1 sequence to that of pCD630
232 (**Figure 2**).Overall, the two plasmids are highly similar. Of note, the ORF that
233 corresponds to the DEAD/DEAH helicase like protein (CDP07 in pCD630) was not
234 annotated in the LK932541 contig due to its linear nature, but is evident in the pCD-
235 ISS1 sequence. Similarly, we found that CDP1 (gene remnant) and CDP11 (doubtful
236 CDS) of pCD630 are in fact a single 201bp ORF, as annotated in the LK932541 contig.
237 A revised annotation of pCD630 has been submitted to ENA (AM180356) to reflect this.
238 Though the pCD-ISS1 and pCD630 plasmids are co-linear, there is a single region that
239 is divergent. The region of pCD630 encompassing the ORFs encoding CDP02 and
240 CDP03 is absent from pCD-ISS1; the latter contains an ORF encoding a RNA
241 polymerase sigma factor protein (Interpro:IPR013324) in this region. The pCD-ISS1
242 annotation does not identify an ORF encoding a homolog of CDP05 of pCD630. This is
243 the result of a 2bp deletion; it suggests that CDP05 (previously annotated as a doubtful
244 CDS) may not be a true coding sequence. Both pCD630 and pCD-ISS1 encode phage-
245 related functions. Most notably, CDP04 and its homolog encode a phage capsid protein
246 with similarity to the HK97-like major capsid proteins of tailed phages of the
247 Caudovirales order. Caudovirales are common *C. difficile* phages [37-40]. However,
248 beside the phage capsid, pCD630 and pCD-ISS11 lack genes encoding other proteins
249 required for virion formation, such as the large terminase subunit and the portal protein.

250 Therefore, it is highly unlikely that phage particles can be produced from these
251 plasmids. In line with this, we find that the genes encoding the phage proteins are
252 poorly, if at all, expressed (unpublished observations). It appears therefore that (part of)
253 a viral genome has been incorporated into the plasmid, or that the viral genome has
254 been transformed into a plasmid.

255 Together, these data suggest the existence of plasmids closely related to
256 pCD630 in at least two different PCR ribotypes (010 and 012).

257

258 *pCD630-like plasmids can be identified in short reads from whole genome sequence*
259 *projects*

260 Above, we showed the existence of at least one pCD630-like plasmid. We wondered if
261 we could extend the family by interrogating the wealth of raw, non-annotated, sequence
262 data in the public domain. We downloaded a selection of sequence reads from ENA,
263 corresponding to 10 different strains (see Materials and Methods). To identify
264 extrachromosomal replicons, we used graph-based tool for reconstruction of plasmids
265 [23]. We validated this tool on our short read sequence data from our 630Δ*erm*
266 sequence (ERR609091)[26] and found that is readily identifies the pCD630 plasmid
267 (data not shown).

268 Surprisingly, we found only two strains with a single replicon (i.e. only the
269 chromosome). The other 8 analyzed datasets suggested the presence of at least one
270 other replicon. Strikingly, 6 contained a replicon that shared similarity to pCD630. Of
271 these, 4 could be circularized due to the presence of direct repeats at the ends of the
272 contig and therefore likely represent complete plasmid sequences, as was the case for

273 pCD-ISS1 (**Table 3**). These plasmids - hereafter referred to as pCD-WTSI1, pCD-
274 WTSI2, pCD-WTSI3 and pCD-WTSI4 – are all significantly larger than pCD630 and
275 pCD-ISS1 (**Figure 2**). The smaller pCD630-like contigs without flanking repeats (that
276 may represent either complete, or incomplete plasmids) were not further studied.

277 To gain further insight in the group of large pCD630-like plasmids, we performed
278 an automated annotation of plasmids pCD-WTSI1 (GenBank: MG019959), pCD-WTSI2
279 (GenBank: MG019960), pCD-WTSI3 (GenBank: MG019961) and pCD-WTSI4
280 (GenBank: MG019962). The homology with the small pCD630-like plasmids is confined
281 to the region encoding CDP6-CDP10 of pCD630. Within this region, it is noteworthy that
282 the ORF encoding the Arc-type ribbon-helix-helix protein (Pfam: PF12651) CDP09 of
283 pCD630 appears to be replaced with another putative DNA binding protein, a helix-turn-
284 helix XRE protein (InterPro:IPR010982) in the pCD-WTSI group of plasmids. Further,
285 we noted that the CDP06, that encodes a truncated homolog of CDP07, appears to be
286 fused with CDP07 to form a hybrid protein nearly identical in size to CDP07. This
287 suggests that the CDP06-07 arrangement may be the result of an (incomplete) gene
288 duplication event. The proteins are putative superfamily 2 helicase fused to an N-
289 terminal CHC2 zinc finger domain, with homology to the corresponding TOPRIM
290 domain of DnaG-like primases. They also contains a third domain of unknown function
291 C-terminal of the helicase domain.

292 The pCD-WTSI plasmids all contain a highly similar accessory module of ~8kb.
293 Within this module, notable functions include an integrase (Pfam: PF00589), a
294 recombinase (Pfam: PF00239), a Cro-C1-type HTH protein (Pfam: PF01381), a
295 penicillinase repressor (Pfam: PF03965), and an RNA polymerase sigma factor (Pfam:

296 PF08281 & Pfam: PF04542). The combination is suggestive of integration of mobile
297 genetic element(s) into the plasmid backbone.
298 In the short read archive, we only identified large pCD630-like plasmids so far.
299 Though we cannot exclude the existence of more small pCD630-like plasmids, we
300 consider it likely that the pCD-WTSI plasmids represent a more widely distributed form
301 of the pCD630-like plasmid family.
302
303 *pCD630-like plasmids have a modular organization*
304 Above, we have identified 6 plasmids sharing significant homology in a region that
305 encompasses an ORF encoding a putative helicase. Moreover, we have shown that the
306 large and small pCD630-like plasmids are remarkably similar, but that certain genes
307 appear to have been exchanged. Thus, the organization of these plasmids, like those of
308 mobile elements in *C. difficile* [41, 42] and plasmids in other organisms [43], is modular.
309 None of the pCD630-like plasmids encodes a previously characterized replication
310 protein; yet, it is clear that the plasmid is efficiently maintained in the absence of
311 obvious selection (**Figure 1**). Based on the finding that all 6 plasmids contain homologs
312 of the pCD630 CDP6-10, we propose that this region (or part of it) forms the replicon of
313 the plasmids. The DEAD-DEAH family helicase CDP07 and its homologs, that also
314 contain a CHC2 zinc finger domain (InterPro: IPR002694) that aligns with the
315 corresponding domain in DnaG-like DNA primases, appear to be the most likely
316 candidate to be the replication proteins for this family of plasmids. As noted above, in
317 the large pCD630-like plasmids the helicase is a CDP06-7 hybrid protein; this may
318 underlie the signals corresponding to these ORFs in microarray and comparative

319 genome hybridization studies, but also suggests that CDP06 itself is probably
320 dispensable for plasmid maintenance. CDP09 is likely also not crucial for the function of
321 the replicon, as it is replaced by another protein in the group of large pCD630-like
322 proteins. It is conceivable that CDP09 and the HTH XTRE proteins serve a regulatory
323 function for instance in controlling the copy number of the plasmids. The small pCD630-
324 like plasmids have an estimated copy number of 4-5, based on average read coverage
325 for chromosomal loci and the plasmid contigs. For the large plasmids, this is 9-10.
326 Consistent with a regulatory rather than an essential function, we noted that in a
327 previous microarray identification more strains appear to contain homologs of CDP6-10
328 than any of the other pCD630 genes, and that several strains harboring CDP6-8 and
329 CDP10 do not contain CDP09 [14]. The same study also found strains that carry
330 homologs of CDP02-03, but not any of the other genes of pCD630. Combined with our
331 observation that this particular region is replaced with a single ORF in pCD-ISS1,
332 suggest that CDP02-03 have been horizontally acquired. In line with this notion, we
333 found that CDP02 has homology to HNH endonucleases (PFAM01844.17), and genes
334 encoding these homing endonucleases are considered as selfish genetic elements [44].
335

336 *pCD630-like plasmids are common in diverse ribotypes*

337 The identification of 6 plasmids carrying a conserved putative replication module,
338 allowed us to determine the most conserved regions within this module. We designed
339 primers against one such region, to be able to identify pCD630-like plasmids by PCR.
340 We tested these primers in a PCR reaction on chromosomal DNA from strains 630Δ*erm*
341 (WKS1241), yielding a positive signal (**Figure 3**). Next, we tested a collection of 43

342 strains of diverse PCR ribotypes to see if pCD630-like plasmids could be identified. We
343 found DNA from 11 isolates gave a signal similar or greater than our positive control,
344 630Δ*erm* (32.6%); this includes strains of PCR ribotypes 012, 015, 017, and 081
345 (**Figure 3**). Interestingly, strain 630 and derivatives are PCR ribotype 012 as well [7].
346 Those samples that were weakly positive on total DNA, appear negative on
347 PlasmidSafe DNase treated DNA and are therefore likely false positives. Alternatively,
348 these could represent isolates in which the plasmid is integrated into the chromosome.
349 Isolating and characterizing these plasmids is part of our ongoing work. We noted that
350 strain EK29, that presumably contains a pCD630-like plasmid [15], appears negative in
351 this PCR. We interpret this to mean that the PCR likely fails to detect certain pCD630-
352 like plasmids, suggesting that the actual number of strain containing pCD630-like
353 plasmids may be even higher. Our data suggests that pCD630-like plasmids are
354 common, and not limited to PCR ribotype 010 (strain 7032985) and 012 (strains 630
355 and derivatives).

356 The high prevalence of pCD630-like plasmids in these strains raises some
357 interesting questions. There is little to no information on the function of these plasmids
358 in *C. difficile* cells. The plasmids from the pCD630-family lack characterized
359 determinants for antimicrobial resistance and are therefore unlikely to play a major role
360 in drug resistance. Instead, they appear to harbor phage remnants or (partial) mobile
361 genetic elements. It is documented that (pro)phages can modulate the expression of the
362 major toxins [45, 46], affect the expression of cell wall proteins [47] and are up-
363 regulated during infection [48]; a role in virulence of *C. difficile* is therefore certainly
364 conceivable.

365 This study has only looked at plasmids of the pCD630 family and found that it
366 occurs among diverse *C. difficile* strains. Based on our limited survey, we found
367 plasmids in 5 different PCR ribotypes, and in strains of different toxinotypes (including
368 both toxigenic and non-toxigenic strains). It will be of interest to see if the pCD630-
369 family of plasmids is the most common, or that other plasmids are equally widely
370 distributed. A broad survey of available genome sequences will likely reveal other
371 families of plasmids and some of these may be limited to specific strains or clades of *C.*
372 *difficile*.

373 The distribution of pCD630-like plasmids suggests that this family was acquired
374 early during the evolution of *C. difficile*, or that the plasmid is capable of horizontal
375 transfer. The pCD630-like plasmids do not encode any characterized conjugation
376 proteins (**Figure 2**); however, they might be transferable dependent on other mobile
377 elements or conjugative plasmids. Of note in this respect is that the mobile element
378 ICE*Bs1* (which is related to Tn*916*, a conjugative transposon common in *C. difficile*) can
379 mobilize plasmids [49], the pathogenicity locus of *C. difficile* can get transferred by a so
380 far unidentified mechanism likely to rely on integrated conjugative elements [50] and in
381 archaea vesicle-mediated plasmid transfer has been documented [51].

382 We found that pCD630-like plasmids are compatible with different replicons
383 (**Figure 1C**). To our knowledge, no plasmid incompatibility has been described for *C.*
384 *difficile* and sequence analysis did not reveal clear candidate genes for an
385 incompatibility system in the plasmids analyzed. Considering the high plasmid
386 prevalence (**Figure 3**), and the fact that existing genetic tools for *C. difficile* depend on
387 the conjugative transfer of shuttle plasmids with a pCB102 or pCD6 replicon [5], one

388 can wonder whether some strains are refractory to genetic manipulation due to the
389 presence of plasmids from an incompatible plasmid group.
390
391 **Conclusions**
392 In this study we showed that plasmid pCD630 from strain 630 is the paradigm of a
393 family of plasmids that is defined by a module that encodes a conserved helicase. Most
394 of the family members belong to a group that is larger than pCD630, and that differ in
395 their accessory module. Plasmids from the pCD630-family are present in diverse *C.*
396 *difficile* strains. Our data warrant further investigation of the role of pCD630-like
397 plasmids – and plasmids in general - in *C. difficile* biology.
398
399
400 **Acknowledgements**
401 The following people are acknowledged for helpful discussions: Y. Anvar, W. Meijer, M.
402 Krupovic and the Experimental Bacteriology laboratory of E.J. Kuijper at the LUMC. We
403 thank the Wellcome Trust Sanger Institute for sharing unpublished raw data that was
404 used in this work via the European Nucleotide Archive. The European Nucleotide
405 Archive support desk is acknowledged for updating the AM180356 record to reflect the
406 findings of this study. This work was supported by the Netherlands Organisation for
407 Scientific Research [VIDI 016.141.310] and Departmental Funds to W.K.S.
408

409 **References**

410 [1] P.A. Lawson, D.M. Citron, K.L. Tyrrell, S.M. Finegold. Reclassification of Clostridium
411 difficile as Clostridioides difficile (Hall and O'Toole 1935) Prevot 1938. Anaerobe 40 (2016) 95-9.
412 [2] W.K. Smits, D. Lyras, D.B. Lacy, M.H. Wilcox, E.J. Kuijper. Clostridium difficile infection.
413 Nat Rev Dis Primers 2 (2016) 16020.
414 [3] C.M. Theriot, A.A. Bowman, V.B. Young. Antibiotic-Induced Alterations of the Gut
415 Microbiota Alter Secondary Bile Acid Production and Allow for Clostridium difficile Spore
416 Germination and Outgrowth in the Large Intestine. mSphere 1 (2016).
417 [4] K. Aktories, C. Schwan, T. Jank. Clostridium difficile Toxin Biology. Annu Rev Microbiol
418 (2017).
419 [5] N. Minton, G. Carter, M. Herbert, T. O'Keeffe, D. Purdy, M. Elmore, et al. The
420 development of Clostridium difficile genetic systems. Anaerobe 10 (2004) 75-84.
421 [6] D. Purdy, T.A. O'Keeffe, M. Elmore, M. Herbert, A. McLeod, M. Bokori-Brown, et al.
422 Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile
423 through circumvention of the restriction barrier. Mol Microbiol 46 (2002) 439-52.
424 [7] M. Sebaihia, B.W. Wren, P. Mullany, N.F. Fairweather, N. Minton, R. Stabler, et al. The
425 multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome.
426 Nat Genet 38 (2006) 779-86.
427 [8] M. He, F. Miyajima, P. Roberts, L. Ellison, D.J. Pickard, M.J. Martin, et al. Emergence
428 and global spread of epidemic healthcare-associated Clostridium difficile. Nat Genet 45 (2013)
429 109-13.
430 [9] N. Kumar, F. Miyajima, M. He, P. Roberts, A. Swale, L. Ellison, et al. Genome-Based
431 Infection Tracking Reveals Dynamics of Clostridium difficile Transmission and Disease
432 Recurrence. Clin Infect Dis 62 (2016) 746-52.
433 [10] D.W. Eyre, M.L. Cule, D.J. Wilson, D. Griffiths, A. Vaughan, L. O'Connor, et al. Diverse
434 sources of C. difficile infection identified on whole-genome sequencing. N Engl J Med 369
435 (2013) 1195-205.
436 [11] X. Didelot, D.W. Eyre, M. Cule, C.L. Ip, M.A. Ansari, D. Griffiths, et al. Microevolutionary
437 analysis of Clostridium difficile genomes to investigate transmission. Genome Biol 13 (2012)
438 R118.
439 [12] C.W. Knetsch, T.D. Lawley, M.P. Hensgens, J. Corver, M.W. Wilcox, E.J. Kuijper.
440 Current application and future perspectives of molecular typing methods to study Clostridium
441 difficile infections. Euro Surveill 18 (2013) 20381.
442 [13] C. Clabots, S. Lee, D. Gerding, M. Mulligan, R. Kwok, D. Schaberg, et al. Clostridium
443 difficile plasmid isolation as an epidemiologic tool. Eur J Clin Microbiol Infect Dis 7 (1988) 312-5.
444 [14] T. Janvilisri, J. Scaria, A.D. Thompson, A. Nicholson, B.M. Limbago, L.G. Arroyo, et al.
445 Microarray identification of Clostridium difficile core components and divergent regions
446 associated with host origin. J Bacteriol 191 (2009) 3881-91.
447 [15] G.L. Marsden, I.J. Davis, V.J. Wright, M. Sebaihia, E.J. Kuijper, N.P. Minton. Array
448 comparative hybridisation reveals a high degree of similarity between UK and European clinical
449 isolates of hypervirulent Clostridium difficile. BMC Genomics 11 (2010) 389.
450 [16] W.N. Fawley, C.W. Knetsch, D.R. MacCannell, C. Harmanus, T. Du, M.R. Mulvey, et al.
451 Development and validation of an internationally-standardized, high-resolution capillary gel-
452 based electrophoresis PCR-ribotyping protocol for Clostridium difficile. PLoS One 10 (2015)
453 e0118150.
454 [17] D. Hyatt, G.L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, L.J. Hauser. Prodigal:
455 prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11
456 (2010) 119.

457 [18] K. Lagesen, P. Hallin, E.A. Rodland, H.H. Staerfeldt, T. Rognes, D.W. Ussery.
458 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35
459 (2007) 3100-8.
460 [19] D. Laslett, B. Canback. ARAGORN, a program to detect tRNA genes and tmRNA genes
461 in nucleotide sequences. Nucleic Acids Res 32 (2004) 11-6.
462 [20] C. Bland, T.L. Ramsey, F. Sabree, M. Lowe, K. Brown, N.C. Kyrpides, et al. CRISPR
463 recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced
464 palindromic repeats. BMC Bioinformatics 8 (2007) 209.
465 [21] Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, Y. Xu. dbCAN: a web resource for automated
466 carbohydrate-active enzyme annotation. Nucleic Acids Res 40 (2012) W445-51.
467 [22] C. Claudel-Renard, C. Chevalet, T. Faraut, D. Kahn. Enzyme-specific profiles for
468 genome annotation: PRIAM. Nucleic Acids Res 31 (2003) 6633-9.
469 [23] L. Vielva, M. de Toro, V.F. Laza, F. de la Cruz. PLACNETw: a web-based tool for
470 plasmid reconstruction from bacterial genomes. Bioinformatics btx462 (2017) 1-3.
471 [24] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman. Basic local alignment search
472 tool. J Mol Biol 215 (1990) 403-10.
473 [25] W.K. Smits. SNP-ing out the differences: Investigating differences between Clostridium
474 difficile lab strains. Virulence (2016) 1-5.
475 [26] E. van Eijk, S.Y. Anvar, H.P. Browne, W.Y. Leung, J. Frank, A.M. Schmitz, et al.
476 Complete genome sequence of the Clostridium difficile laboratory strain 630Deltaerm reveals
477 differences from strain 630, including translocation of the mobile element CTn5. BMC Genomics
478 16 (2015) 31.
479 [27] H. Dannheim, T. Riedel, M. Neumann-Schaal, B. Bunk, I. Schober, C. Sproer, et al.
480 Manual curation and reannotation of the genomes of Clostridium difficile 630Deltaerm and C.
481 difficile 630. J Med Microbiol 66 (2017) 286-93.
482 [28] M.M. Collery, S.A. Kuehne, S.M. McBride, M.L. Kelly, M. Monot, A. Cockayne, et al.
483 What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence
484 and phenotypes of Clostridium difficile strain 630 and derivatives. Virulence (2016) 1-15.
485 [29] T. Riedel, D. Wetzel, J.D. Hofmann, S. Plorin, H. Dannheim, M. Berges, et al. High
486 metabolic versatility of different toxigenic and non-toxigenic Clostridioides difficile isolates. Int J
487 Med Microbiol 307 (2017) 311-20.
488 [30] J.T. Heap, S.T. Cartman, S.A. Kuehne, C. Cooksley, N.P. Minton. ClosTron-targeted
489 mutagenesis. Methods Mol Biol 646 (2010) 165-82.
490 [31] D. Bakker, A.M. Buckley, A. de Jong, V.J. van Winden, J.P. Verhoeks, O.P. Kuipers, et
491 al. The HtrA-like protease CD3284 modulates virulence of Clostridium difficile. Infect Immun 82
492 (2014) 4222-32.
493 [32] S.T. Cartman, M.L. Kelly, D. Heeg, J.T. Heap, N.P. Minton. Precise manipulation of the
494 Clostridium difficile chromosome reveals a lack of association between the tcdC genotype and
495 toxin production. Appl Environ Microbiol 78 (2012) 4683-90.
496 [33] P.T. van Leeuwen, J.M. van der Peet, F.J. Bikker, M.A. Hoogenkamp, A.M. Oliveira
497 Paiva, S. Kostidis, et al. Interspecies Interactions between Clostridium difficile and Candida
498 albicans. mSphere 1 (2016).
499 [34] R.P. Fagan, N.F. Fairweather. Clostridium difficile has two parallel and essential Sec
500 secretion systems. J Biol Chem 286 (2011) 27483-93.
501 [35] A.M. Oliveira Paiva, A.H. Friggen, S. Hossein-Javaheri, W.K. Smits. The Signal
502 Sequence of the Abundant Extracellular Metalloprotease PPEP-1 Can Be Used to Secrete
503 Synthetic Reporter Proteins in Clostridium difficile. ACS Synth Biol 5 (2016) 1376-82.
504 [36] I. Moura, M. Monot, C. Tani, P. Spigaglia, F. Barbanti, N. Norais, et al. Multidisciplinary
505 analysis of a nontoxigenic Clostridium difficile strain with stable resistance to metronidazole.
506 Antimicrob Agents Chemother 58 (2014) 4957-60.

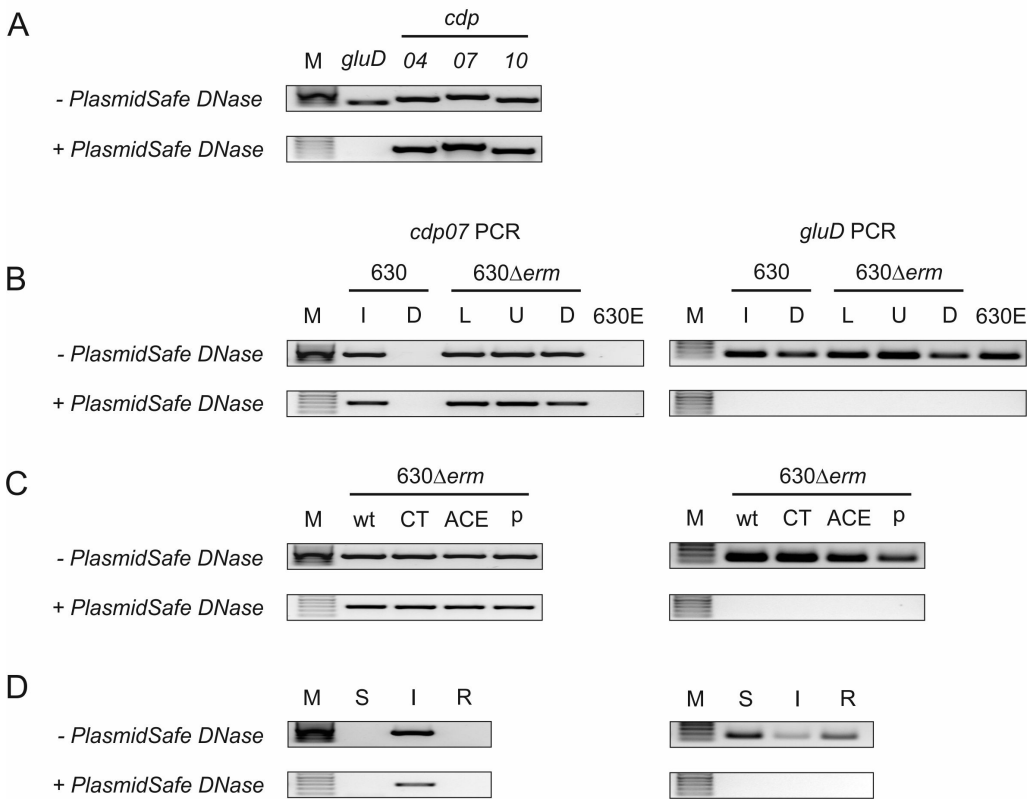
507 [37] K.R. Hargreaves, M.R. Clokie. A Taxonomic Review of Clostridium difficile Phages and
508 Proposal of a Novel Genus, "Phimmp04likevirus". Viruses 7 (2015) 2534-41.
509 [38] K.R. Hargreaves, M.R. Clokie. Clostridium difficile phages: still difficult? Front Microbiol 5
510 (2014) 184.
511 [39] O. Sekulovic, L.C. Fortier. Characterization of Functional Prophages in Clostridium
512 difficile. Methods Mol Biol 1476 (2016) 143-65.
513 [40] O. Sekulovic, J.R. Garneau, A. Neron, L.C. Fortier. Characterization of temperate
514 phages infecting Clostridium difficile isolates of human and animal origins. Appl Environ
515 Microbiol 80 (2014) 2555-63.
516 [41] M.S. Brouwer, A.P. Roberts, P. Mullany, E. Allan. In silico analysis of sequenced strains
517 of Clostridium difficile reveals a related set of conjugative transposons carrying a variety of
518 accessory genes. Mob Genet Elements 2 (2012) 8-12.
519 [42] M.S. Brouwer, P.J. Warburton, A.P. Roberts, P. Mullany, E. Allan. Genetic organisation,
520 mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced
521 strains of Clostridium difficile. PLoS One 6 (2011) e23014.
522 [43] R. Fernandez-Lopez, M.P. Garcillan-Barcia, C. Revilla, M. Lazaro, L. Vielva, F. de la
523 Cruz. Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid
524 evolution. FEMS Microbiol Rev 30 (2006) 942-66.
525 [44] D.R. Edgell. Selfish DNA: homing endonucleases find a home. Curr Biol 19 (2009)
526 R115-7.
527 [45] R. Govind, G. Vediyappan, R.D. Rolfe, B. Dupuy, J.A. Fralick. Bacteriophage-mediated
528 toxin gene regulation in Clostridium difficile. J Virol 83 (2009) 12037-45.
529 [46] O. Sekulovic, M. Meessen-Pinard, L.C. Fortier. Prophage-stimulated toxin production in
530 Clostridium difficile NAP1/027 lysogens. J Bacteriol 193 (2011) 2726-34.
531 [47] O. Sekulovic, L.C. Fortier. Global transcriptional response of Clostridium difficile carrying
532 the CD38 prophage. Appl Environ Microbiol 81 (2015) 1364-74.
533 [48] M. Meessen-Pinard, O. Sekulovic, L.C. Fortier. Evidence of in vivo prophage induction
534 during Clostridium difficile infection. Appl Environ Microbiol 78 (2012) 7662-70.
535 [49] C.A. Lee, J. Thomas, A.D. Grossman. The Bacillus subtilis conjugative transposon
536 ICEBs1 mobilizes plasmids lacking dedicated mobilization functions. J Bacteriol 194 (2012)
537 3165-72.
538 [50] M.S. Brouwer, A.P. Roberts, H. Hussain, R.J. Williams, E. Allan, P. Mullany. Horizontal
539 gene transfer converts non-toxigenic Clostridium difficile strains into toxin producers. Nat
540 Commun 4 (2013) 2601.
541 [51] S. Erdmann, B. Tschitschko, L. Zhong, M.J. Raftery, R. Cavicchioli. A plasmid from an
542 Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-
543 free cells. Nat Microbiol (2017).
544 [52] J.R. O'Connor, D. Lyras, K.A. Farrow, V. Adams, D.R. Powell, J. Hinds, et al.
545 Construction and analysis of chromosomal Clostridium difficile mutants. Mol Microbiol 61 (2006)
546 1335-51.
547 [53] H.A. Hussain, A.P. Roberts, P. Mullany. Generation of an erythromycin-sensitive
548 derivative of Clostridium difficile strain 630 (630Deltaerm) and demonstration that the
549 conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. J Med
550 Microbiol 54 (2005) 137-41.
551 [54] L.J. Deakin, S. Clare, R.P. Fagan, L.F. Dawson, D.J. Pickard, M.R. West, et al. The
552 Clostridium difficile spo0A gene is a persistence and transmission factor. Infect Immun 80
553 (2012) 2704-11.
554 [55] K.E. Rosenbusch, D. Bakker, E.J. Kuijper, W.K. Smits. C. difficile 630Deltaerm Spo0A
555 regulates sporulation, but does not contribute to toxin production, by direct high-affinity binding
556 to target DNA. PLoS One 7 (2012) e48608.

557 [56] T. Riedel, B. Bunk, A. Thurmer, C. Sproer, E. Brzuszkiewicz, B. Abt, et al. Genome
558 Resequencing of the Virulent and Multidrug-Resistant Reference Strain Clostridium difficile 630.
559 Genome Announc 3 (2015).
560

561

562 **Figure Legends**

563 **Figure 1. 630 and derivatives can contain pCD630. A.** *C. difficile* 630Δ*erm* [26]
564 contains the pCD630 plasmid. **B.** Some, but not all, 630-derived strains contain
565 pCD630. I=ISS D=DSMZ L=LUMC U=UCL [26]. **C.** Genetically modified 630Δ*erm*
566 strains still contain pCD630. wt = wild type, CT = Clostron mutant [30, 31], ACE = allelic
567 coupled exchange mutant [32, 33], p = containing a replicative plasmid [34, 35]. **D**.
568 Strain 7032985 (intermediate metronidazole susceptible; I) contains a pCD630-like
569 plasmid but strains 7032994 (metronidazole susceptible; S) and 7032989
570 (metronidazole resistant; R) do not. For oligonucleotides used, see Materials and
571 Methods. M = marker.

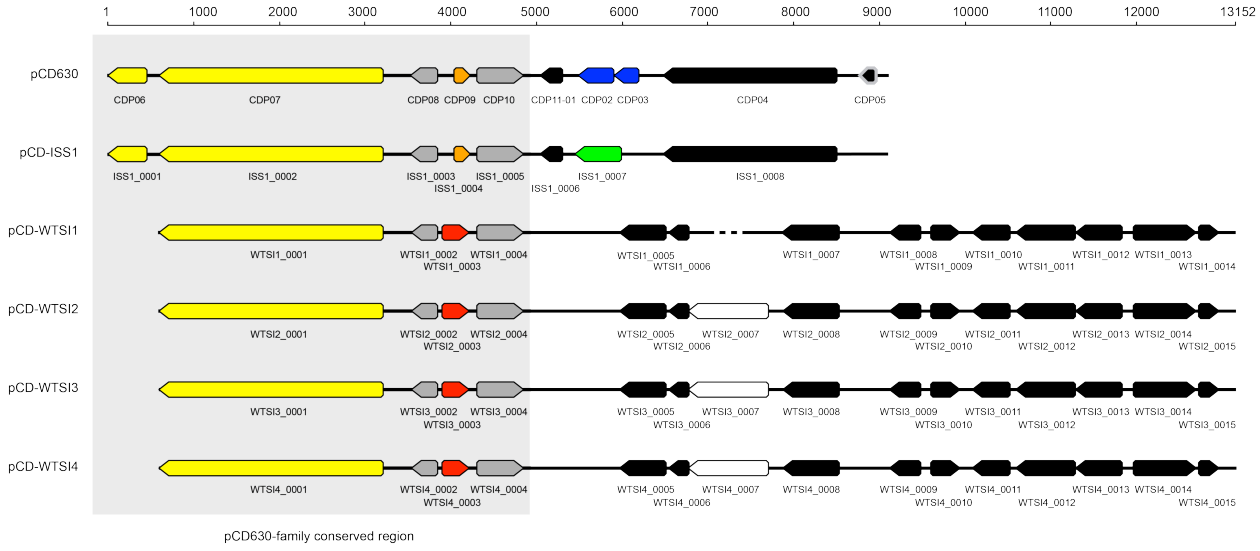


572

573

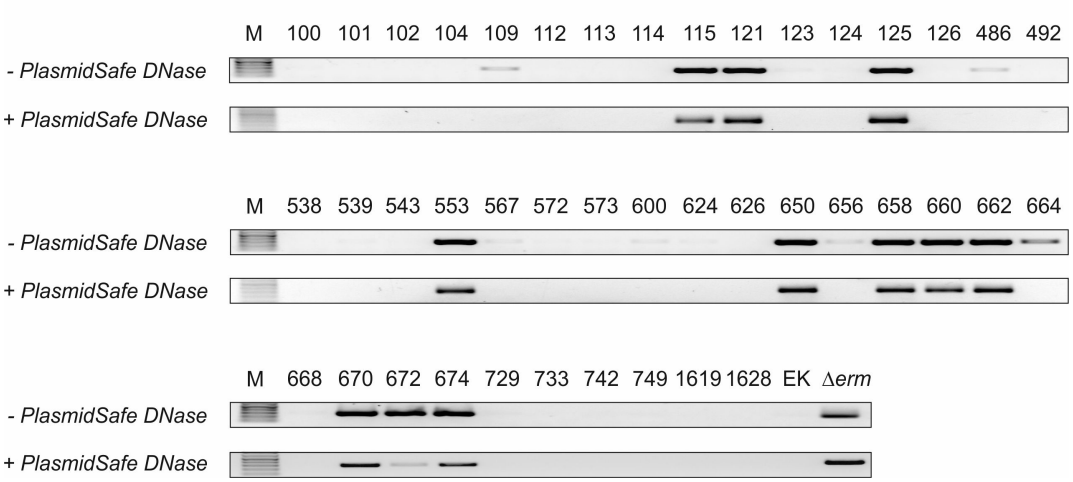
574

575 **Figure 2. Schematic representation of an alignment of pCD630-like plasmids.** Full-
576 length plasmids identified in this study were aligned. pCD-ISS1 is based on
577 GenBank:LK932541. pCD-WTSI-1 to pCD-WTSI4 are based on short read sequences
578 from ENA:PRJEB2101. The most striking differences are indicated with differently
579 colored ORFs. The conserved module encompassing the gene encoding a helicase is
580 boxed, the accessory module is indicated with black ORFs. The gray outline of CDP05
581 indicates it is annotated in AM180356.1 but is not predicted in our analysis.
582



583
584
585
586
587
588
589
590
591

592 **Figure 3. pCD630-like plasmids are present in diverse *C. difficile* strains.** A PCR
593 was performed against a conserved target region in the putative helicase protein using
594 oWKS-1651 and oWKS-1652. The presence of a pCD630-like plasmid results in a
595 positive signal in this PCR. M = marker, EK = EK29 [15], Δ*erm* = 630Δ*erm* [26].



596

597

598 **Tables.**
599
600 **Table 1. *S*trains used in this study.**
601

Strain	Labstock	PCR ribotype	References
630 (ISS)	WKS1705	012	P. Mastrantonio lab collection; [7]
630E	WKS1706	012	R. Britton lab collection; [52]
630Δ*erm* (LUMC)	WKS1241	012	[26, 53]
630Δ*erm* (UCL)	WKS1707	012	[53]
spo0A::CT	WKS1242	012	[54, 55]
pAP24	AP34	012	[35]
Δ*hpd*	AP58	012	[33]
DSMZ27543	WKS1941	012	[7, 27, 56]
DSMZ28645	WKS1943	012	[27, 53]
EK29	WKS1914	078	[15]
7032994	WKS1935	010	P. Spigaglia lab collection; [36]
7032985	WKS1937	010	P. Spigaglia lab collection; [36]
7032989	WKS1939	010	P. Spigaglia lab collection; [36]
100	WKS1950	002	J.S. Weese lab collection
101	WKS1951	258	J.S. Weese lab collection; [14]
102	WKS1952	002	J.S. Weese lab collection
104	WKS1953	137	J.S. Weese lab collection; [14]
109	WKS1954	085	J.S. Weese lab collection; [14]
112	WKS1955	009	J.S. Weese lab collection; [14]
113	WKS1956	002	J.S. Weese lab collection; [14]
114	WKS1957	015	J.S. Weese lab collection; [14]
115	WKS1958	015	J.S. Weese lab collection; [14]
121	WKS1959	001	J.S. Weese lab collection; [14]
123	WKS1960	039	J.S. Weese lab collection; [14]
124	WKS1961	002	J.S. Weese lab collection; [14]
125	WKS1962	081	J.S. Weese lab collection; [14]
126	WKS1963	010	J.S. Weese lab collection
486	WKS1964	Unknown	J.S. Weese lab collection
492	WKS1965	719	J.S. Weese lab collection
538	WKS1967	009	J.S. Weese lab collection; [14]
539	WKS1968	017	J.S. Weese lab collection
543	WKS1969	009	J.S. Weese lab collection

553	WKS1970	015	J.S. Weese lab collection
567	WKS1971	020	J.S. Weese lab collection
572	WKS1972	027	J.S. Weese lab collection
573	WKS1973	001	J.S. Weese lab collection
600	WKS1974	137	J.S. Weese lab collection
624	WKS1975	001	J.S. Weese lab collection
626	WKS1976	046	J.S. Weese lab collection
650	WKS1977	012	J.S. Weese lab collection
656	WKS1978	017	J.S. Weese lab collection; [14]
658	WKS1979	012	J.S. Weese lab collection; [14]
660	WKS1980	012	J.S. Weese lab collection; [14]
662	WKS1981	012	J.S. Weese lab collection; [14]
664	WKS1982	017	J.S. Weese lab collection; [14]
668	WKS1983	012	J.S. Weese lab collection; [14]
670	WKS1984	012	J.S. Weese lab collection; [14]
672	WKS1985	012	J.S. Weese lab collection; [14]
674	WKS1986	012	J.S. Weese lab collection; [14]
729	WKS1987	078	J.S. Weese lab collection
733	WKS1988	288	J.S. Weese lab collection
742	WKS1989	Unknown	J.S. Weese lab collection
749	WKS1990	001	J.S. Weese lab collection
1619	WKS1991	Unknown	J.S. Weese lab collection
1628	WKS1992	Unknown	J.S. Weese lab collection

604 **Table 2. Oligonucleotides used in this study.**

Name	Sequence (5' > 3')	Target
oWKS-1070	GTCTTGGATGGTTGATGAGTAC	Forward primer on *gluD*
oWKS-1071	TTCCTAATTTAGCAGCAGCTTC	Reverse primer on *gluD*
oWKS-1629	CTCGAGCGAATGCAAGAG	Forward primer on *cdp04*
oWKS-1630	CCAGTCACCTATGTGCATACC	Reverse primer on *cdp04*
oWKS-1631	ACCTACACAGATGCGTTCAG	Forward primer on *cdp07*
oWKS-1632	AAAGCACCTCATAGCCTTCC	Reverse primer on *cdp07*
oWKS-1633	AAAGTAGTTACGGGCGACAC	Forward primer on *cdp10*
oWKS-1634	TCACAGAAGGCTGCAAACTC	Reverse primer on *cdp10*
oWKS-1651	TAGTCTACCTCTGCACTTATTAG	Forward primer on pCD630-like helicase genes (including *cdp07*)
oWKS-1652	CATTAAAAGAGCTGGATATAAAAGC	Reverse primer on pCD630-like helicase genes (including *cdp07*)

605

606

607 **Table 3. Full length pCD630-like plasmids.**

Name	Source	Accessions	Size	Reference
pCD630	Strain 630	GenBank: AM180356	7881 bp	[7]; this study
pCD-ISS1	Strain 7032985	GenBank: LK932541 (contig) GenBank: pending accessioning (plasmid)	7991 bp	[36]; this study
pCD-WTSI1	Not specified	ENA: ERR017368 (Illumina reads) GenBank: MG019959	11777 bp	This study
pCD-WTSI2	Not specified	ENA:ERR022513 (Illumina reads) GenBank: MG019960	12526 bp	This study
pCD-WTSI3	Not specified	ENA: ERR125910 (Illumina reads) GenBank: MG019961	12525 bp	This study
pCD-WTSI4	Not specified	ENA: ERR125911 (Illumina reads) GenBank: MG019962	12488 bp	This study

608