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

Introduction

Clostridium difficile (*Clostridioides difficile* [1]) is a Gram-positive, anaerobic and spore-forming bacterium that can asymptomatically colonize the human gut [2]. It is ubiquitous in the environment, and can also be found in the gastrointestinal tract of many animals. The bacterium gained notoriety when it was identified as the causative agent of health care associated diarrhea, and is increasingly implicated in community-associated disease in many countries [2]. In hosts with a dysbiosis of the microbiome, such as patients treated with broad-spectrum antimicrobials, conditions are favorable for *C. difficile* germination and outgrowth [3]. *C. difficile* produces one or more toxins, that cause symptoms ranging from diarrhea to potentially fatal toxic megacolon [2, 4].

Over the past two decades, genetic studies have of *C. difficile* have become possible due to the generation of shuttle plasmids that can be transferred by conjugation from *Escherichia coli* to *C. difficile* [5]. These plasmids mostly employ a replicon derived from plasmid pCD6 for replication in *C. difficile* [6]. In 2006, the first genome sequence of *C. difficile* became available, revealing the presence of another plasmid, pCD630 [7].

Despite a great number of strains having been whole genome sequenced since then, plasmid biology of *C. difficile* has been poorly explored. One reason is that plasmid content is variable, and most studies on the evolution and/or transmission of *C. difficile* focus on those genes conserved between all strains (the core genome) [8-11]. However, there is reason to assume that plasmids are common in *C. difficile*; for instance, before the advent of the currently common typing schemes [12], plasmid

isolation had been proposed as an epidemiological tool [13]. The ratio of plasmid-containing to plasmid-free strains in this study was found to be approximately 1:2, suggesting that around 30% of strains of *C. difficile* may carry a plasmid. Furthermore, hybridization-based analyses of total DNA from a collection of *C. difficile* strains suggest the presence of DNA with significant similarity to pCD630 open reading frames (ORFs) [14, 15].

Here, we define a family of plasmids that share a conserved helicase-containing module and demonstrate that these plasmids are common in a diverse set of *C. difficile* strains.

Materials and methods

Strains and growth conditions

Strains used in this study are listed in **Table 1**. For DNA isolation, strains were grown on *Clostridium difficile* agar (CLO) plates (bioMérieux) and a single colony was inoculated into pre-reduced brain-heart infusion broth (Oxoid) supplemented with 0.5 % w/v yeast extract (Sigma-Aldrich) and *Clostridium difficile* selective supplement (Oxoid). Strains were PCR ribotyped [16] in-house for this study. For some strains a PCR ribotype could not be assigned (**Table 1**).

Isolation of plasmid DNA from C. difficile

Plasmids were isolated from 2 mL of *C. difficile* overnight culture using NucleoSpin Plasmid Easypure columns (Macherey-Nagel); to increase yield, 10mg/mL lysozyme to

buffer A1 was added, as recommended by the manufacturer. Using PCR and sequencing, we found that the DNA isolated using this kit is heavily contaminated with chromosomal DNA. To isolate pure plasmid DNA, aliquots of the DNA were incubated with PlasmidSafe ATP-dependent DNase (Epicentre) that digests linear, but not circular double stranded DNA. After purification with a Nucleospin Gel and PCR Clean-up kit (Macherey-Nagel), the absence of genomic DNA was confirmed by PCR using primers targeting *gluD* (**Table 2**). Yields were generally very low, but the plasmid was readily detectable by PCR.

Reannotation of pCD630 and identification of a pCD630-like plasmid

The pCD630 sequence was obtained from GenBank (AM180356.1). CDP01 and CDP11 form a single open reading frame (ORF) and were treated as a single ORF in our analyses. Protein sequences encoded by the ORFs of pCD630 were used as BLAST queries against the non-redundant protein sequences database, limited to taxid:1496 (*Clostridium difficile*). This identified the 8089 bp *Peptoclostridium difficile* genome assembly 7032985, scaffold BN1096_Contig_85 (LK932541.1). To reconstitute the plasmid from this contig, the DNA was circularized and a single copy of the 98 bp direct repeat that was present at the terminus of the original contig was removed using Geneious R10. The resulting 7991 bp sequence now encodes a full copy of a sequence homologous to CDP07 of pCD630. Reannotation of plasmids was performed using using an in-house pipeline. This pipeline incorporates the gene caller Prodigal (version 2.6.3) [17], RNAmmer (version 1.2) [18], Aragorn (version 1.2.38) [19], the CRISPR recognition tool (version 1.2) [20], dbCAN (version 5.0) [21] and PRIAM (version March

2015) [22]. The plasmid derived from LK932541 was submitted to GenBank as pCD-ISS1 (GenBank: MG266000).

Identification of pCD630-like plasmids in short read archives

In order to identify other pCD630-like plasmids in sequence databases, paired end Illumina sequences from study PRJEB2101 (ERR017367-ERR017371, ERR022513, ERR125908-ERR125911) were downloaded from the short read archive of the European Nucleotide Archive (ENA). Short reads were assembled and visualized in PLACNETw [23] to determine likely replicons. The contigs corresponding to pCD630-like plasmids were downloaded and imported into Geneious R10 software (Biomatters Ltd) for circularization and removal of terminal repeats; afterwards all plasmids which could be circularized were compared with BLASTN (version 2.40) [24] to pCD630 and the sequences were restructured to start at the base corresponding to base 2903 in pCD630. Afterwards the plasmids were annotated using the in-house pipeline as described above, and submitted to GenBank as pCD-WTSI1 (GenBank: MG019959), pCD-WTSI2 (GenBank: MG019960), pCD-WTSI3 (GenBank: MG019961), pCD-WTSI4 (GenBank: MG019962). Alignments of the pCD630-like plasmids were performed in Geneious R10 (Biomatters Ltd) and the alignment figure was prepared using Adobe Illustrator CC (Adobe Systems Inc).

Polymerase chain reaction

Oligonucleotides used in this work are listed in **Table 2**. To confirm the presence of pCD630 in derivatives of *C. difficile* strain 630, PCR was performed with oWKS-1629

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

Results and discussion

pCD630 is present in C. difficile strain 630 and some of its derivatives

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

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

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

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

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

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

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

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

A pCD630-like plasmid is present in a strain with reduced metronidazole susceptibility

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

This sequence stems from a study that compares three non-toxigenic PCR ribotype 010 strains of *C. difficile*, with differing susceptibility to metronidazole [36]. Strain 7032985 was classified as intermediate resistant to metronidazole. If we assume that the contig represents a pCD630-like plasmid, we expect DNA from this strain to remain positive in a PCR that targets the plasmid upon treatment with PlasmidSafe DNase. We found that the PCR targeting *cdp07*, but not chromosomal locus *gluD,* 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:
231 MG266000) 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.

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

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

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

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

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

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

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

The pCD-WTSI plasmids all contain a highly similar accessory module of ~8kb. Within this module, notable functions include an integrase (Pfam: PF00589), a recombinase (Pfam: PF00239), a Cro-C1-type HTH protein (Pfam: PF01381), a 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

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

pCD630-like plasmids are common in diverse ribotypes

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

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

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

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

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

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

can wonder whether some strains are refractory to genetic manipulation due to the presence of plasmids from an incompatible plasmid group.

Conclusions

In this study we showed that plasmid pCD630 from strain 630 is the paradigm of a family of plasmids that is defined by a module that encodes a conserved helicase. Most of the family members belong to a group that is larger than pCD630, and that differ in their accessory module. Plasmids from the pCD630-family are present in diverse *C. difficile* strains. Our data warrant further investigation of the role of pCD630-like plasmids – and plasmids in general - in *C. difficile* biology.

Acknowledgements

The following people are acknowledged for helpful discussions: Y. Anvar, W. Meijer, M. Krupovic and the Experimental Bacteriology laboratory of E.J. Kuijper at the LUMC. We thank the Wellcome Trust Sanger Institute for sharing unpublished raw data that was used in this work via the European Nucleotide Archive. The European Nucleotide Archive support desk is acknowledged for updating the AM180356 record to reflect the findings of this study. This work was supported by the Netherlands Organisation for Scientific Research [VIDI 016.141.310] and Departmental Funds to W.K.S.

References

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

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

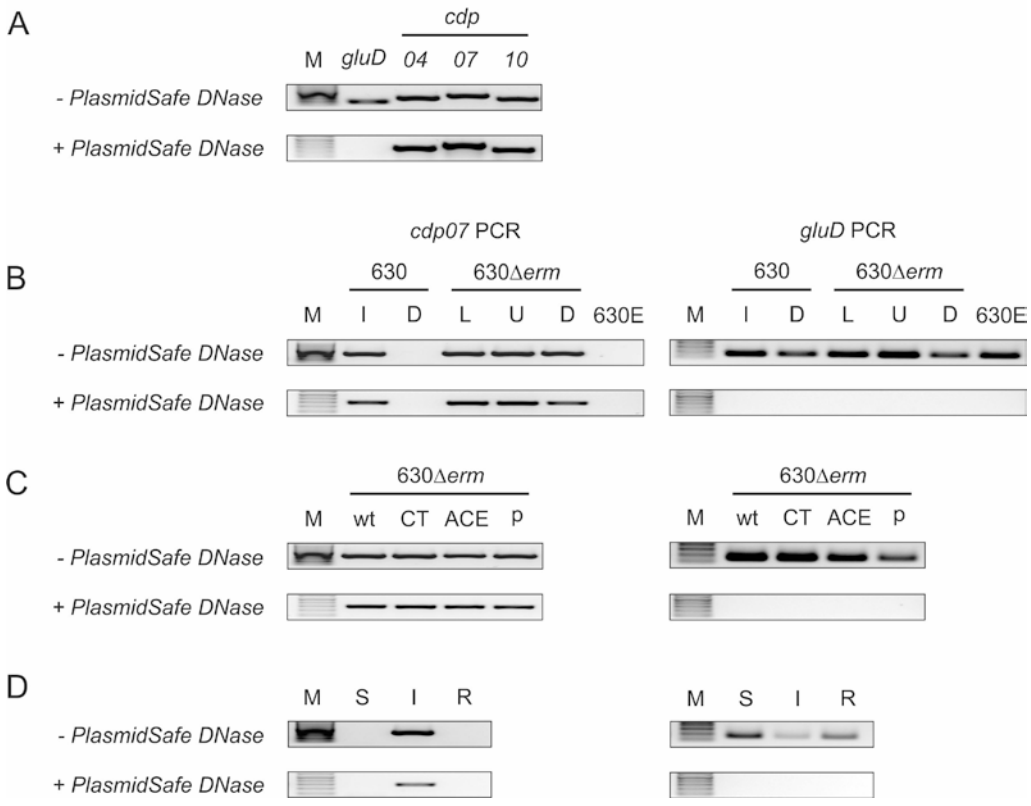
[37] K.R. Hargreaves, M.R. Clokie. A Taxonomic Review of Clostridium difficile Phages and Proposal of a Novel Genus, "Phimmp04likevirus". Viruses 7 (2015) 2534-41.
[38] K.R. Hargreaves, M.R. Clokie. Clostridium difficile phages: still difficult? Front Microbiol 5 (2014) 184.
[39] O. Sekulovic, L.C. Fortier.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

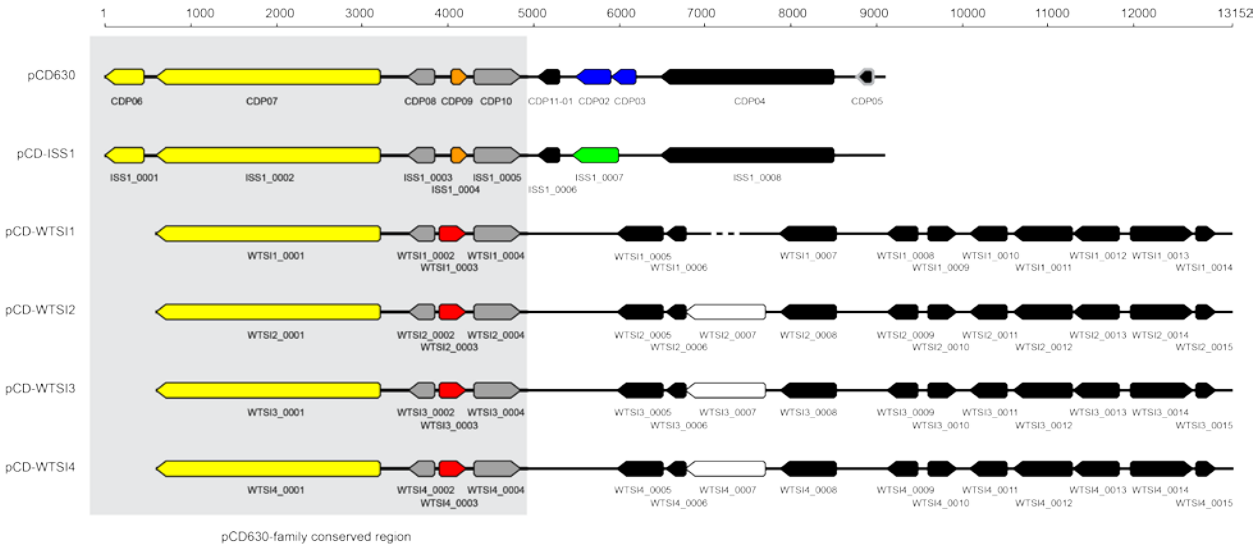
561

Figure Legends

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

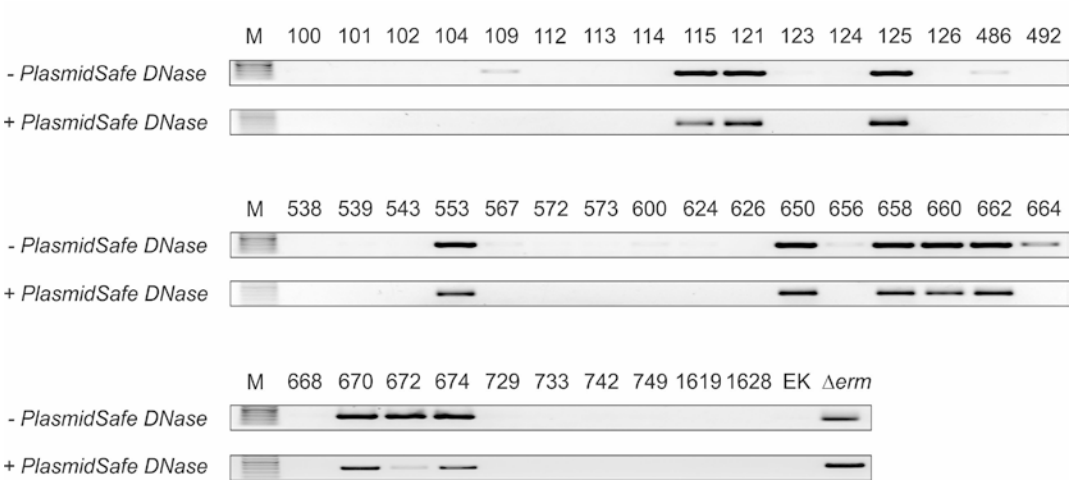


575 **Figure 2. Schematic representation of an alignment of pCD630-like plasmids.** Full-
576 length plasmids identified in this study were aligned. pCD-ISS1 is based on
577 GenBank:LK932541. pCD-WTSI-1 to pCD-WTSI4 are based on short read sequences
578 from ENA:PRJEB2101. The most striking differences 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. *Strains 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

602

603

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: MG266000 (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