

## Carriage of three plasmids in a single human clinical isolate of *Clostridioides difficile*

Anna M. Roseboom<sup>a</sup>, Quinten R. Ducarmon<sup>a</sup>, B.V.H. Hornung<sup>a</sup>, C. Harmanus<sup>a</sup>, M.J.T. Crobach<sup>a</sup>, Ed J. Kuijper<sup>a</sup>, Rolf H.A.M. Vossen<sup>b</sup>, Susan L. Kloet<sup>b</sup>, Wiep Klaas Smits<sup>a,c</sup>

<sup>a</sup>Department of Medical Microbiology and Leiden University Center of Infectious Diseases (LU-CID), Leiden University Medical Center, Leiden, The Netherlands

<sup>b</sup> Department of Human Genetics, Leiden Genome Technology Center, Leiden University Medical Center, Leiden, The Netherlands

<sup>c</sup> Centre for Microbial Cell Biology, Leiden, The Netherlands

\*Corresponding author: W.K. Smits, Department of Medical Microbiology, Postzone E4-P, Leiden University Medical Center, Albinusdreef 2, 2333ZA, Leiden, The Netherlands. Phone: +31-71-5261229. E-mail: [w.k.smits@lumc.nl](mailto:w.k.smits@lumc.nl)

AR: ORCID: 0000-0001-9576-4473; E-mail: [annarozad@hotmail.com](mailto:annarozad@hotmail.com)

QRD: ORCID 0000-0001-7077-2127; E-mail: [q.r.ducarmon@lumc.nl](mailto:q.r.ducarmon@lumc.nl)

BVHH: E-mail: [Bastian.Hornung@gmx.de](mailto:Bastian.Hornung@gmx.de)

CH: ORCID 0000-0002-5328-3273, E-mail: [c.harmanus@lumc.nl](mailto:c.harmanus@lumc.nl)

MJTC: ORCID 0000-0002-4743-0986, E-mail: [m.j.t.crobach@lumc.nl](mailto:m.j.t.crobach@lumc.nl)

EJK: ORCID 0000-0001-5726-2405; E-mail: [e.j.kuijper@lumc.nl](mailto:e.j.kuijper@lumc.nl)

RHAMV: ORCID ; E-mail: [r.h.a.m.vossen@lumc.nl](mailto:r.h.a.m.vossen@lumc.nl)

SLK: ORCID 0000-0001-9189-0640; E-mail: [s.l.kloet@lumc.nl](mailto:s.l.kloet@lumc.nl)

WKS: ORCID 0000-0002-7409-2847; E-mail: [w.k.smits@lumc.nl](mailto:w.k.smits@lumc.nl)

## Abstract

A subset of clinical isolates of *Clostridioides difficile* contains one or more plasmids and these plasmids can harbor virulence and antimicrobial resistance determinants. Despite their potential importance, *C. difficile* plasmids remain poorly characterized. Here, we describe a human clinical isolate that carries three plasmids from three different plasmid families that are therefore compatible. For two of these, we identify a region capable of sustaining plasmid replication in *C. difficile*. Together, our data advance our understanding of *C. difficile* plasmid biology.

## Keywords

Complete genome sequence, plasmid compatibility, *C. difficile* carriage

## Highlights

- The complete circular genome sequence is provided for a *C. difficile* isolate harboring three plasmids
- These three plasmids (pJMR5-1, pJMR5-4 and pJRM5-W) are therefore compatible in a single strain
- Sequence analysis suggest a modular nature of plasmid families to which the pJMR-plasmids belong
- A functional replicon was cloned from pJMR5-1 (pCD-ECE1 family) and pJMR5-W (pCD-WTSI1 family) and plasmids carrying this replicon are compatible with plasmid pCD630

## (1) Introduction

The Gram-positive anaerobic spore-forming bacterium *Clostridioides difficile* is responsible for healthcare-associated and community-acquired infectious diarrhea with potentially fatal consequences [1]. The symptoms of a *C. difficile* infection (CDI) are related to the expression of one or more toxins, but virulence of this bacterium is multifactorial [2].

In many Gram-positive pathogens, virulence factors are encoded on plasmids [3, 4]. For *C. difficile*, sporadic reports indicate that toxins and resistance determinants can be carried on extrachromosomal elements [5-7]. It is estimated that ~10-70% of *C. difficile* isolates carry one or more plasmids, but information on plasmid functions is sparse [8]. In particular, very few replicons have been identified and almost no experimental evidence on plasmid compatibility is available. Regions sufficient for replication in *C. difficile* have been cloned from the plasmids pCD6 and pCD-METRO from strains CD6 and IB136, respectively [5, 9]. An *in silico* analysis of publicly available sequence data suggests that specific families of plasmids may co-exist [8, 10], but these predictions have not been validated experimentally.

Here, we show carriage of plasmids from three different plasmid families in a single isolate of *C. difficile* derived from a human patient and report on the identification of a region sufficient for plasmid maintenance for two of these plasmids. Together, these data significantly advance our understanding of *C. difficile* plasmid biology.

## (2) Materials and methods

### (2.1) Bacterial strain and growth conditions

All bacterial strains are listed in **Table 1**. The isolates used in this study are derived from the collection of isolates at the Dutch national Reference Laboratory (NRL) for *C. difficile*, which is housed at the Leiden University Medical Center. Strains JMR1, JMR2, and JMR5 have been isolated from one or more patients at the same healthcare facility in the context of diagnostics. These strains were collected over a period of five months and were sent to the NRL for sentinel surveillance. Strain JMR7 was collected in the context of a multicenter study aimed at determining the value of *C. difficile* colonization screening at hospital admission in an endemic setting [11].

Isolates were routinely cultured on TSS plates (Tryptic Soy Agar with 5% sheep blood, bioMérieux, The Netherlands) or CLO plates (selective *C. difficile* medium containing cefoxitin, amphotericin B and cycloserine, bioMérieux, The Netherlands). Capillary electrophoresis PCR ribotyping was performed at the Dutch National Reference Laboratory for *C. difficile*, according to standard procedures [12] and further characterized using a multiplex PCR targeting the 16S rRNA gene, *gluD*, and the genes encoding the large clostridial toxins and binary toxin [13].

Laboratory strains of *C. difficile* were cultured anaerobically at 37°C in liquid BHIY (1.5 % w/v Brain Heart Infusion [Oxoid], 0.5% w/v yeast extract [Sigma]) or on BHIY agar plates (BHIY, 1.5% agar w/v), supplemented with *C. difficile* selective Supplement (CDSS [Oxoid]) and 25 µg/mL thiamphenicol when appropriate, in a Don Whitley VA-1000 workstation (10% CO<sub>2</sub>, 10% H<sub>2</sub> and 80% N<sub>2</sub> atmosphere). *E. coli* strains were cultivated aerobically at 37°C, 200 rpm in Luria-Bertani (LB) broth or on LB agar plates, supplemented with 50 µg/mL ampicillin, 25 µg/mL chloramphenicol and/or 50 µg/mL kanamycin when required. Stocks were made in 15% w/v glycerol and stored at –80°C.

## **(2.2) DNA isolation and whole genome sequencing of *C. difficile***

Initial analysis of the draft genome of strain JMR5, including the tentative identification and annotation of extrachromosomal elements, was performed based on short-read Illumina sequence data (ERS2564723 from PRJEB25045) as described [10, 11]. The complete genome was generated based on long read sequencing on the Pacific Biosciences (PacBio) Sequel platform. To prepare high molecular weight total DNA, cells from 5 mL of overnight culture were pelleted and processed using the Qiagen Genomic-tip 100/g, according to the manufacturer's instructions. PacBio sequencing libraries were generated according to the manufacturer's multiplexed microbial library preparation protocol, part number 101-696-100, version 7, July 2020 release using the SMRTbell Express Template Prep Kit v2.0 with the following modifications: genomic DNA was sheared using Speed 34 on the Megaruptor 3 (Diagenode) and an additional size selection step of 6-50kb fragments on the Blue Pippin (Sage Science) was included for the final SMRT bell library. The libraries were sequenced on a Sequel II platform (Pacific Biosciences) using the Sequel II Binding Kit v2.0, Sequencing Primer v4, Sequencing Kit v2.0 and a 30hr movie time.

## **(2.3) Data analysis and visualization**

Raw PacBio sequence data was assembled using Flye (v2.9)[14] and the start position was fixed using Circlator and the “fixstart” parameter [15]. Assembly quality was subsequently inspected using QUAST [16] and assembly completeness using BUSCO (v5.3.2, dataset clostridia\_odb10 creation date 2020-03-06)[17]. Prokka (v1.14.6) was used for rapid genome annotation and to obtain protein sequences, with the “–kingdom Bacteria”, “–genus Clostridioides” and “–species difficile” parameters [18]. Multi-locus sequence typing (MLST) was performed using mlst (v2.19.0) with the PubMLST *C. difficile* database updated to October 21<sup>st</sup> 2021 [19]. Alignments of plasmid families were visualized using clinker [20]. Amrfinderplus (v3.1.23)[21] with the database version of

December 21<sup>st</sup> 2021 was run to identify acquired antimicrobial resistance (AMR) genes, genes with point mutations conferring resistance to antimicrobials (specifically *gyrA*, *gyrB*, *murG*, *rpoB*, *rpoC* and *23S*) and virulence factors (including the *C. difficile* toxin genes) with the following options: “-n” “--organism Clostridioides\_difficile”, “—plus”. The MobileElementFinder webserver was used to detect potential mobile elements with default settings.[22].

#### **(2.4) Data and code availability**

Raw sequence data generated for this study is available at the European Nucleotide Archive under BioProject PRJEB53950. All bioinformatic tools used for the analyses are freely available through the references provided. Plasmid sequences have also been deposited independently as ON887052 (pJMR5-1), ON887053 (pJMR5-4) and ON887054 (pJMR5-W) in GenBank. The complete genome sequence of JMR5 is GCA\_944989955.1 (assembly)/ERS12289077 (sample). Previously generated whole genome sequences [11] are available under BioProject number PRJEB25045. Sequences for pCD-ECE1 (LR594544.1), pCD-ECE4 (LR594545.1), pCD630 (AM180356.2) and pCD-WTSI1 (MG019959.1) were retrieved from GenBank.

#### **(2.5) Confirmation of extracellular nature of the plasmids**

In order to confirm the extrachromosomal nature of the plasmids, a PlasmidSafe DNase (PSD; Epicentre) experiment was performed as described previously [5]. In short, total genomic DNA was isolated and incubated in the presence or absence of PSD for 24h. Chromosomal DNA is fragmented and susceptible for degradation by PSD in this procedure, whereas small circular double stranded DNA (plasmids) is not. After incubation, the remaining DNA was purified and amplified with primers specific for loci on the chromosome (oWKS-1070/oWKS-1071), pJMR5-1 (oAR-1/oAR-2), pJMR5-4 (oAR-3/oAR-4) and pJMR5-W (oAR-5/oAR-6) using MyTaq DNA polymerase (Meridan) according to the instructions of the manufacturer. The sequences of all oligonucleotides are listed in **Table 2**.

## (2.6) Molecular biology procedures

All vectors were constructed with restriction/ligation cloning using a linearized pAP24-based vector [23] as the backbone and a DNA fragment of interest as insert. Plasmids were isolated from 1 mL overnight culture using NucleoSpin Plasmid Easypure columns (Macherey-Nagel) from the Plasmid Easypure Kit (Nucleospin).

Putative replicon fragments were obtained through amplification PCR using high fidelity enzyme Q5 DNA polymerase (NEB) and primers including a KpnI or NcoI restriction site (**Table 2**). After digestion with KpnI/NcoI, the amplicons were ligated into similarly digested pAP24-backbone [23], as described previously [5], and transformed to *E.coli* DH5 $\alpha$  with selection on chloramphenicol-containing LB agar plates. All constructs were verified by Sanger sequencing using primers listed in

### **Table 2.**

For conjugation, *E.coli* CA434 competent cells [9] were transformed using purified plasmid DNA and transformants were selected on LB agar supplemented with chloramphenicol and kanamycin. Conjugative transfer of the plasmids from CA434 to *C. difficile* 630 $\Delta$ erm [24] was performed according to standard procedures [9]. The transconjugants, selected on thiamphenicol containing media, were screened with PCR using MyTaq DNA polymerase (Meridan) directly on colony material or on purified DNA. DNA extractions were performed using the DNeasy Blood and Tissue (Qiagen) kit after enzymatic cell lysis as prescribed by the manufacturer. To confirm identity of the transconjugants, the presence of *gluD* (on the chromosome) and *traJ* (on the plasmid backbone) was verified in a PCR using primers oWKS-1070/oWKS-1071 and oWKS-1387/oWKS-1388, respectively (**Table 2**). The presence of the pCD630 plasmid in the DNA samples of the TCs was confirmed by performing a PCR using primers oWKS-1633 and oWKS-1634 (**Table 2**).

### **(3) Results and Discussion**

#### **(3.1) Characterization of the plasmid-containing isolates**

Three of the four clinical isolates (JMR1, JMR2 and JMR5) have been isolated from a single patient in a single healthcare facility as part of the ongoing activities of the Dutch National Reference Laboratory for *C. difficile*. Medical records show that the patient initially experienced mild symptoms, but eventually succumbed to the infection after several more CDI episodes. The fourth isolate (JMR7) was collected as part of a study looking at *C. difficile* colonization at hospital admission [11]. As the patient did not give full consent during this study [11], only limited metadata is available for this isolate. We note, however, that JMR5 and JMR7 were collected in the same time period and place, from patients of the same age and gender. Therefore, it is likely that JMR1, JMR2, JMR5, and JMR7 are all longitudinal samples from the same patient suffering from recurrent CDI over a period of at least 4 months. If the samples are indeed derived from the same patient, we would expect all *C. difficile* isolates to be similar in terms of PCR ribotype and toxin profile.

Therefore, we performed capillary PCR ribotyping and a multiplex PCR targeting the toxin genes. This data showed that all JMR isolates belong to PCR ribotype 081; this PCR ribotype is reported to be part of the phylogenetic clade 1 of non-epidemic isolates and is not among the most commonly found PCR ribotypes in Europe [25]. In agreement with the ribotyping result, the multiplex PCR shows that all isolates are toxin A and toxin B positive, but negative for binary toxin.

Additionally, we extracted the relevant core genome MLST information (cgMLST) from the previously generated Illumina sequences (GenBank PRJEB25045)[11]. This showed that all isolates belong to the same cluster type, with 0 allele differences. Together, these data support the notion that they are derived from a single persistently infected individual.

#### **(3.2) Confirmation of the extrachromosomal nature of pJMR5-1, pJMR5-2 and pJMR5-3**

Isolate JMR7 was subjected to short-read whole genome sequencing, as part of the aforementioned study. During the analyses, carriage of extrachromosomal elements was predicted, according to in-

house established methods [10]. The predictions indicate possible extrachromosomal elements belonging to the pCD-ECE1, pCD-ECE4 and pCD-WTSI1 families, of 6.5, 15.4 and 21.9 kb size, respectively.

We extracted the contigs corresponding to the predicted extrachromosomal elements, and circularized them after removal of the terminal repeats. To confirm the presence and the circular dsDNA nature of the elements *in vivo*, we performed a PlasmidSafe DNase (PSD) analysis using primers specifically directed at regions of each putative plasmid [26]. The analysis showed that a product of the expected size was obtained in PCRs targeting each of the putative plasmids, even for PSD-treated samples. In contrast, a PCR product for the chromosomal locus *gluD* was only obtained in the samples that were not treated with PSD (**Figure 1**). We note a decrease in the strength of the signal on gel for the pCD-WTSI family plasmid upon treatment with PSD; as this is the largest of the predicted plasmid, we attribute this to partial fragmentation of the dsDNA during the DNA isolation.

These data indicate that all JMR strains contain circular dsDNA that is extrachromosomal, consistent with the presence of a plasmid.

### **(3.3) Complete genome sequence of *C. difficile* JMR5**

Next, we determined the complete genome sequence of JMR5, as a representative of the group of JMR strains, using single molecule real time sequencing. Assembly of the reads of the Pacific Bioscience Sequel platform yielded a single circular chromosome of 4.3 Mb and 3 smaller replicons (**Supplemental Figure 1**). The size and sequence of the latter matched the predicted circular elements discussed above, and are hereafter referred to as pJMR5-1 (for plasmid from **JMR5** from the pCD-ECE1 family), pJMR5-4 (for plasmid from **JMR5** from the pCD-ECE4 family) and pJMR5-W (for plasmid from **JMR5** from the pCD-WTSI1 family). Initially, we recovered only pJMR-4 and pJMR5-W in the *de novo* assembly, possibly as a result of size selection during the library preparation (see Materials and Methods). However, manual inspection of the data revealed a significant number of

reads matching pJMR5-1 and these reads fully matched the predicted sequence; indeed, altering the default settings of Flye [14] allowed automatic recovery of the pJMR5-1 plasmid, though in this case the plasmid contained a duplication (a well-known artefact of long-read sequencing of small plasmids) [15]. A split library preparation (<10kb and 10+ kb) may improve automatic recovery plasmids of <10kb in a de novo assembly using Flye.

The genome of JMR5 is 4.321.867 bp and has an average [G+C]-content of 29% , in line with other *C. difficile* genomes. An automated annotation identifies 3937 ORFs, 35 rRNA genes ,90 tRNA genes and a single tmRNA. Several AMR genes were identified by AMRfinderplus [21] on the JMR5 chromosome, namely bla-CDD1, erm(B), the vanGCd operon and a T82I point mutation in *gyrA*, associated with fluoroquinolone resistance. One AMR gene, tet(M), was present on pJMR5-W. No mobile elements were identified by MobileElementFinder [22].

We used PubMLST [19] to predict the sequence type of JMR5. The isolate was identified as belonging to ST9, consistent with previous studies that place the RT081 reference strain from the Leeds-Leiden collection in the same sequence type [27, 28].

### **(3.4) Comparison of the identified plasmids with reference plasmids**

The tentative assignment of the three plasmids to the pCD-ECE1, pCD-ECE4 and pCD-WTSI1 family of plasmids is based on regions of similarity with the reference sequences for these plasmids as described [10]. To gain more insight in the relatedness of the plasmids, we generated and visualized nucleotide alignments using clinker [20]. From these analyses, multiple observations were made.

First, we note that the ORF7-ORF8 region of pCD-ECE1 appears not to be conserved in pJMR5-1 (**Figure 2A**). This was unexpected, as ORF8 of pCD-ECE1 was hypothesized to encode a replication-associated protein based on limited homology to viral REP proteins [10]. If this region of the plasmid is involved in replication, the corresponding region in pJMR5-1 should also encode a replication protein. Indeed, analysis of the predicted amino acid sequence of the pJMR5-1 protein

using the PHYRE2 Protein Fold Recognition Server [29] indicates that it likely is a rolling circle replication initiator protein (PDB 4CIJ; confidence 98.8%; identity; 13%).

Second, the pJMR5-4 plasmid appears to contain a 11-gene insertion in comparison to the pCD-ECE4 reference sequence (**Figure 2B**), consistent with a previously reported large size-diversity within this family [10]. Automated annotation of gene function (see Materials and methods) suggests that this region may include a methylation sensitive restriction modification system as ORF12 of pJMR5-4 encodes a putative type II endonuclease, and ORF11 encodes a DNA (cytosine-5)-methyltransferase.

Third, pJMR5-W is substantially larger than other members of the pCD630/pCD-WTSI1 family (**Figure 2C**)[26]. Plasmids pCD630 and pCD-WTSI1 share a conserved region that encodes a helicase protein, believed to be important for replication, but differ in their accessory genes [26]. pJMR5-W also contains the conserved region but it appears that pJMR5-W contains two insertions compared to pCD-WTSI1; i) a single hypothetical gene that is also found at this site for other pCD-WTSI family plasmids [26] and ii) an 11-gene cluster suggestive of a mobile element (e.g. including a gene encoding an Tn916-like excisionase [ORF23] and an integrase [ORF24]). Interestingly, PHYRE2 [29] also predicts multiple regulator proteins (ORF14, ORF15, ORF19, ORF21), a putative rolling circle replication initiator protein (PDB 4CIJ, confidence 100, 28% identity), and a putative tetracycline resistance protein (ORF20)(PDB 3J25, confidence 100, 97% identity) in the island. Tn916-like elements are common vehicles for tetracycline resistance genes [30].

### **(3.5) A functional replicon from pJMR5-1 and pJMR5-W**

Based on the sequence analysis, we expected ORF8 of pJMR5-1 to be required for replication of the plasmid in *C. difficile*. To assess which region is sufficient for replication, we replaced the pCD6 replicon from an established *E.coli*-*C. difficile* shuttle vector [23] with different, but overlapping, regions of the pJMR5-1 plasmid. The obtained constructs were conjugated into the *C. difficile* laboratory strain 630Δerm [24, 31]; when transconjugants were obtained this was taken as a sign of

a functional replicon, whereas repeated failure to obtain transconjugants suggested that a critical determinant was missing. A region encompassing ORF7-ORF10 sustained plasmid replication, as did a region encompassing ORF8-ORF10 as transconjugants were readily obtained when conjugating plasmids pJMR25 and pJMR51 (**Table 3**). In contrast, a region encompassing only ORF8, or ORF7-ORF8 did not, as we failed to obtain transconjugants with plasmid pJMR20 and pJMR52 (**Table 3**). We conclude that the ORF8-ORF10 is sufficient to allow the plasmid to replicate (**Figure 2A**). Furthermore, preliminary data suggests that ORF10, encoding an Arc-type ribbon-helix-helix protein is essential for this function (data not shown).

In 2019, analysis of the ORFs on the pCD-ECE4 plasmid failed to predict a replicon region in plasmids from this family [10]. Screening the ORFs of pJMR5-4 using PHYRE2 indicated that ORF5 likely encodes a protein of which the core domain shows homology to the archaeo-eukaryotic primase (AEP) domain of an archaeal primase protein (PDB 1V33; confidence 96.8, 32% identity)[32]. Primase-polymerases (PrimPols) are archaeo-eukaryotic primases (AEP) with both primase activity and DNA polymerase activity [33]. The AEP supergroup PrimPol-PV1 is widespread in diverse bacterial and archaeal plasmids and likely plays a crucial role in plasmid replication [34]. However, a region encompassing this protein did not sustain plasmid replication in *C. difficile* (**Table 3**). A larger region, additionally encompassing ORF1-ORF5 and an area of high repeat-density, similarly failed to sustain replication (**Table 3**). We conclude that this region does not contain a functional replicon. It is likely that one or more of the hypothetical genes in the conserved part of the pCD-ECE4 family of plasmids encodes the replication function.

The plasmid pJMR5-W contains the helicase-containing module that defines the pCD630/pCD-WTSI family of plasmids [26]. We cloned two different fragments of the highly conserved region into a vector as potential replicons; pJMR57 carries the complete conserved region as putative replicon, and pJMR28 includes only ORF1, encoding the putative helicase protein. We considered that the pJMR5-W replicon might not be compatible with the replicon of pCD630 due to high similarity (**Figure 3B**). Plasmid pCD630 is present in our recipient strain 630Δerm [24] so two

additional recipient strains were included; DSMZ 28645 (a  $630\Delta erm$  strain from the Leibniz Institute DSMZ collection)[35] and JIR8094 (also known as 630E)[36] both of which do not contain the pCD630 plasmid [26]. Conjugation of the three strains with pJMR28 (containing only the ORF9 fragment) repeatedly did not result in any growth on the selective plates, while conjugation with pJMR57 (containing the complete conserved region) did yield viable transconjugants carrying the vector. Therefore, we conclude that the conserved region characterizing all pCD630/pCD-WTSI plasmids comprises a functional replicon, whereas a fragment encoding only the helicase is not sufficient for plasmid maintenance. The fact that plasmid pJMR57 was successfully introduced into a recipient strain carrying pCD630 indicates that pCD630 and pCD-WTSI1 are compatible plasmids, at least for the duration of the experiment. This was confirmed by a PCR specifically targeting the pCD630 sequence on purified DNA from the obtained transconjugants (data not shown). Our experimental data is consistent with the *in silico* prediction that pCD-WTSI family plasmids appear compatible with most other plasmids, including others from the pCD-WTSI1 family (i.e. the replication function does not confer incompatibility)[8]. Of note, our results also show that the putative rolling circle replication protein of pJMR5-W (ORF22) – that is not present in pCD-WTSI1 or pCD630 – is not strictly required for replication, despite its structural similarity to a plasmid replication protein. However, we have not yet established whether a region of the plasmid that incorporates this protein might also be able to sustain replication so we cannot rule out such a function at this time.

#### **(4) Conclusions**

In this study we demonstrated carriage of three different plasmids in a clinical isolate obtained from a symptomatic patient suffering from rCDI. Our results suggest that for *C. difficile* i) non-conserved regions of putative plasmid families can be involved in plasmid replication (as exemplified by pJMR5-1), ii) plasmid replication proteins may exist that defy current *in silico* predictions of protein structure and function (pJMR5-4) and iii) the conserved module of a wide-spread plasmid family can sustain plasmid replication (pJMR5-W). Importantly, epidemiological evidence suggests that the

pJMR5-1, pJMR5-4 and pJMR5-W plasmids are retained during a persistent infection over a period of months.

## (5) References

- [1] W.K. Smits, D. Lyras, D.B. Lacy, M.H. Wilcox, E.J. Kuijper. *Clostridium difficile* infection. *Nat Rev Dis Primers* 2 (2016) 16020.
- [2] G. Vedantam, A. Clark, M. Chu, R. McQuade, M. Mallozzi, V.K. Viswanathan. *Clostridium difficile* infection: toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. *Gut Microbes* 3 (2012) 121-34.
- [3] V. Adams, J. Li, J.A. Wisniewski, F.A. Uzal, R.J. Moore, B.A. McClane, et al. Virulence Plasmids of Spore-Forming Bacteria. *Microbiol Spectr* 2 (2014).
- [4] S. Schwarz, J. Shen, S. Wendlandt, A.T. Fessler, Y. Wang, K. Kadlec, et al. Plasmid-Mediated Antimicrobial Resistance in Staphylococci and Other Firmicutes. *Microbiol Spectr* 2 (2014).
- [5] I.M. Boekhoud, B.V.H. Hornung, E. Sevilla, C. Harmanus, I. Bos-Sanders, E.M. Terveer, et al. Plasmid-mediated metronidazole resistance in *Clostridioides difficile*. *Nat Commun* 11 (2020) 598.
- [6] M. Pu, J.M. Cho, S.A. Cunningham, G.K. Behera, S. Becker, T. Amjad, et al. Plasmid Acquisition Alters Vancomycin Susceptibility in *Clostridioides difficile*. *Gastroenterology* 160 (2021) 941-5 e8.
- [7] G. Ramirez-Vargas, C. Rodriguez. Putative Conjugative Plasmids with *tcdB* and *cdtAB* Genes in *Clostridioides difficile*. *Emerg Infect Dis* 26 (2020) 2287-90.
- [8] W.K. Smits, A.M. Roseboom, J. Corver. Plasmids of *Clostridioides difficile*. *Curr Opin Microbiol* 65 (2022) 87-94.
- [9] 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. *Molecular microbiology* 46 (2002) 439-52.
- [10] B.V.H. Hornung, E.J. Kuijper, W.K. Smits. An in silico survey of *Clostridioides difficile* extrachromosomal elements. *Microb Genom* 5 (2019) e000926.
- [11] M.J.T. Crobach, B.V.H. Hornung, C. Verduin, M.C. Vos, J. Hopman, N. Kumar, et al. *Clostridioides difficile* colonization at admission to the hospital and onwards transmission: a multi-centre case-control study. Submitted (2022).
- [12] 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.
- [13] E.C.f.D.P.a. Control. Laboratory procedures for diagnosis and typing of human *Clostridium difficile* infection. Stockholm; 2018.
- [14] M. Kolmogorov, J. Yuan, Y. Lin, P.A. Pevzner. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37 (2019) 540-6.
- [15] M. Hunt, N.D. Silva, T.D. Otto, J. Parkhill, J.A. Keane, S.R. Harris. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16 (2015) 294.
- [16] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29 (2013) 1072-5.
- [17] M. Manni, M.R. Berkeley, M. Seppey, E.M. Zdobnov. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr Protoc* 1 (2021) e323.
- [18] T. Seemann. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30 (2014) 2068-9.
- [19] K.A. Jolley, J.E. Bray, M.C.J. Maiden. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3 (2018) 124.
- [20] C.L.M. Gilchrist, Y.H. Chooi. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* (2021).
- [21] M. Feldgarden, V. Brover, N. Gonzalez-Escalona, J.G. Frye, J. Haendiges, D.H. Haft, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* 11 (2021) 12728.

- [22] M.H.K. Johansson, V. Bortolaia, S. Tansirichaiya, F.M. Aarestrup, A.P. Roberts, T.N. Petersen. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *J Antimicrob Chemother* 76 (2021) 101-9.
- [23] 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.
- [24] 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.
- [25] M.P. Bauer, D.W. Notermans, B.H. van Benthem, J.S. Brazier, M.H. Wilcox, M. Rupnik, et al. *Clostridium difficile* infection in Europe: a hospital-based survey. *Lancet (London, England)* 377 (2011) 63-73.
- [26] W.K. Smits, J.S. Weese, A.P. Roberts, C. Harmanus, B. Hornung. A helicase-containing module defines a family of pCD630-like plasmids in *Clostridium difficile*. *Anaerobe* 49 (2018) 78-84.
- [27] C.W. Knetsch, E.M. Terveer, C. Lauber, A.E. Gorbelenya, C. Harmanus, E.J. Kuijper, et al. Comparative analysis of an expanded *Clostridium difficile* reference strain collection reveals genetic diversity and evolution through six lineages. *Infect Genet Evol* 12 (2012) 1577-85.
- [28] A. Baktash, J. Corver, C. Harmanus, W.K. Smits, W. Fawley, M.H. Wilcox, et al. Comparison of Whole-Genome Sequence-Based Methods and PCR Ribotyping for Subtyping of *Clostridioides difficile*. *J Clin Microbiol* 60 (2022) e0173721.
- [29] L.A. Kelley, M.J. Sternberg. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4 (2009) 363-71.
- [30] A.P. Roberts, P. Mullany. Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol Rev* 35 (2011) 856-71.
- [31] H.A. Hussain. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630 erm) and demonstration that the conjugative transposon Tn916 E enters the genome of this strain at multiple sites. *Journal of Medical Microbiology* 54 (2005) 137-41.
- [32] D. Czernecki, P. Legrand, M. Tekpinar, S. Rosario, P.A. Kaminski, M. Delarue. How cyanophage S-2L rejects adenine and incorporates 2-aminoadenine to saturate hydrogen bonding in its DNA. *Nat Commun* 12 (2021) 2420.
- [33] L. Blanco, P.A. Calvo, A. Diaz-Talavera, G. Carvalho, N. Calero, A. Martinez-Carron, et al. Mechanism of DNA primer synthesis by human PrimPol. *Enzymes* 45 (2019) 289-310.
- [34] D. Kazlauskas, G. Sezonov, N. Charpin, C. Venclovas, P. Forterre, M. Krupovic. Novel Families of Archaeo-Eukaryotic Primases Associated with Mobile Genetic Elements of Bacteria and Archaea. *J Mol Biol* 430 (2018) 737-50.
- [35] T. Riedel, B. Bunk, A. Thurmer, C. Sproer, E. Brzuszkiewicz, B. Abt, et al. Genome Resequencing of the Virulent and Multidrug-Resistant Reference Strain *Clostridium difficile* 630. *Genome Announc* 3 (2015).
- [36] 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. *Molecular microbiology* 61 (2006) 1335-51.

**Table 1. Strains.**

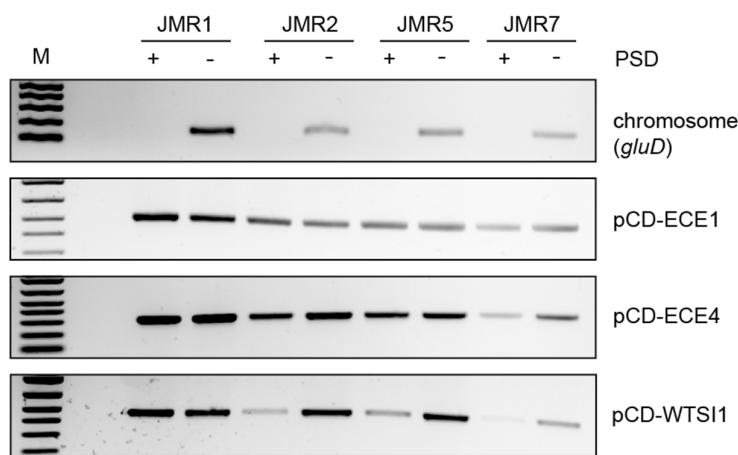
Strain	Description	Reference
<b><i>Clostridioides difficile</i></b>		
JMR1	Human clinical isolate; RT081. Contains three plasmids.	This study
JMR2	Human clinical isolate; RT081. Contains three plasmids.	This study
JMR5	Human clinical isolate; RT081. Contains pJMR5-1, pJMR5-4 and pJMR5-W plasmids.	This study
JMR7	Human clinical isolate; RT081. Contains three plasmids.	This study
630 $\Delta$ erm	Laboratory strain; RT012.	[24, 31]
JIR8094	Laboratory strain; RT012.	[36]
DSMZ 28645	Laboratory strain; RT012.	[35]
<b><i>Escherichia coli</i></b>		
AP24	DH5 $\alpha$ x pAP24 (pRPF185-derived vector containing sluc <sup>opt</sup> and pCD6 replicon, <i>catP</i> )	[23]
DH5 $\alpha$	<i>fhuA2 lac(del)U169 phoA glnV44</i> $\Phi$ 80' <i>lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	Laboratory stock
JMR20	DH5 $\alpha$ x pJMR20 (pAP24-derived vector containing pJMR5-1 ORF8 and flanking intergenic regions)	This study
JMR25	DH5 $\alpha$ x pJMR25 (pAP24-derived vector containing pJMR5-1 ORF7-ORF10 and flanking intergenic regions)	This study
JMR28	DH5 $\alpha$ x pJMR28 (pAP24-derived vector containing pJMR5-W ORF1 and flanking intergenic regions)	This study
JMR33	DH5 $\alpha$ x pJMR33 (pAP24-derived vector containing pJMR5-4 ORF5 and flanking intergenic regions)	This study
JMR46	DH5 $\alpha$ x pJMR46 (pAP24-derived vector containing pJMR5-4 ORF1-5 and flanking intergenic regions)	This study
JMR51	DH5 $\alpha$ x pJMR51 (pAP24-derived vector containing pJMR5-1 ORF8-10 and flanking intergenic regions)	This study
JMR52	DH5 $\alpha$ x pJMR52 (pAP24-derived vector containing pJMR5-1 ORF7-8 and flanking intergenic regions)	This study
JMR57	DH5 $\alpha$ x pJMR57 (pAP24-derived vector containing pJMR5-W ORF1-4 and flanking intergenic regions)	This study

**Table 2. Oligonucleotides used in this study.** Relevant restriction sites are underlined.

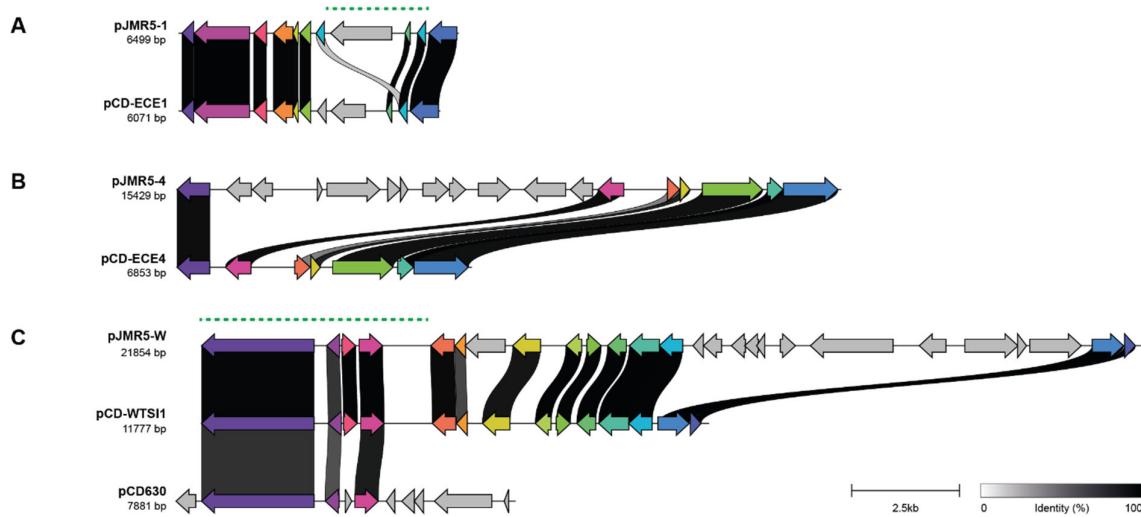
Name	Sequence 5' > 3'	Description	Reference
oAR-1	TTGATTATTATAGTGGAACTAATGGC	pJMR5-1 Forward	This study
oAR-2	ACACCATAACAGTATAAACAGC	pJMR5-1 Reverse	This study
oAR-3	AGGTATGGAGCAATCAGAAAAAG	pJMR5-4 Forward	This study
oAR-4	<u>GT</u> TTTTATAATCTCTGGTTGCATGTG	pJMR5-4 Reverse	This study
oAR-5	GGAACATGATTATAGCTAATAAT	pJMR5-W Forward	This study
oAR-6	CTGTCACAGGGTTGGAAAG	pJMR5-W Reverse	This study
oAR-12	<u>GTGGTAC</u> CTAACATCATTAAAGTAACATCA	Cloning of pJMR5-1 replicon Forward	This study
oAR-13	<u>CTCCC</u> ATGGAATTCTTAAATAGTTGACATTAG	Cloning of pJMR5-1 replicon Reverse	This study
oAR-14	<u>GTGGTAC</u> AAAAGCACCTCCTATAAAC	Cloning of pJMR5-1 replicon Forward	This study
oAR-15	<u>CTCCC</u> CATGGAGGAAGTATTATTTGAACTTTTAG	Cloning of pJMR5-1 replicon Reverse	This study
oAR-16	<u>GTGGTAC</u> AGCTACTTCAACCTTTCATG	Cloning of pJMR5-W replicon Forward	This study
oAR-17	<u>CTCCC</u> CATGGTATTAAAACCGTCAATTATACTTTTATTAAATG	Cloning of pJMR5-W replicon Reverse	This study
oAR-18	<u>GTGGTAC</u> AAAGTGATTTTATATATTATATGAATA	Cloning of pJMR5-4 replicon Forward	This study
oAR-19	<u>CTCCC</u> CATGGCTAATTATTAACTAATCTACTTACT	Cloning of pJMR5-4 replicon Reverse	This study
oAR-20	<u>GTGGTAC</u> CTATGTGAAGAAGAATGATTATATG	Cloning of pJMR5-4 replicon Forward	This study
oAR-21	AAAACGGCTTAAATTCAATAC	Sequencing replicons	This study
oAR-22	CCATACTCACTTTGCC	Sequencing replicons	This study
oAR-23	ACATGCTACAGAAGGTAAGAACGT	Sequencing replicons	This study
oAR-24	GGTATCTTATCATATGCACC	Sequencing replicons	This study
oAR-25	CTTAGGTAGTCAAAAGC	Sequencing replicons	This study
oAR-26	CCATTGAATTATGAATGAATATG	Sequencing replicons	This study
oAR-27	AGGTCTATTGTATCAGC	Sequencing replicons	This study
oAR-28	GCCTAGTATATCATCGCAG	Sequencing replicons	This study
oAR-29	GTATAAATAGTCACATAACTCTATAG	Sequencing replicons	This study
oAR-31	<u>CTCCC</u> CATGGTGTATGATTCTTGCTGTTTTAC	Cloning of pJMR5-W replicon	This study
oAR-36	ATGCTATGTTTGATGAGTATATGCT	Sequencing replicons	This study
oAR-37	CTGTACGTTGTTAATTAGTTAAGTGA	Sequencing replicons	This study
oWKS-1070	GTCTGGATGGTTGATGAGTAC	Targeting chromosome ( <i>gluD</i> Forward)	[26]
oWKS-1071	TTCCTAATTAGCAGCAGCTTC	Targeting chromosome ( <i>gluD</i> Reverse)	[26]
oWKS-1387	CAGATGAGGGCAAGCGGATG	Verification of transconjugants ( <i>traJ</i> Forward)	[5]
oWKS-1388	CGTCGGTGAGGCCAGAGTTTC	Verification of transconjugants ( <i>traJ</i> Reverse)	[5]
oWKS-1633	AAAGTAGTTACGGGCGACAC	Targeting pCD630	[26]
oWKS-1634	TCACAGAAGGCTGCAAACTC	Targeting pCD630	[26]

**Table 3. Results of conjugations**

Origin of putative replicon	Plasmid	Recipient strain	pCD630 in recipient (yes/no)	Transconjugants obtained (yes/no)
pJMR5-1	pJMR20	630Δerm	Yes	No
	pJMR25	630Δerm	Yes	Yes
	pJMR51	630Δerm	Yes	Yes
	pJMR52	630Δerm	Yes	No
pJMR5-4	pJMR33	630Δerm	Yes	Yes
	pJMR46	630Δerm	Yes	No
pJMR5-W	pJMR28	630Δerm	Yes	No
		JIR8094	No	No
		DSMZ 28645	No	No
	pJMR57	630Δerm	Yes	Yes
		JIR8094	No	Yes
		DSMZ 28645	No	Yes



**Figure 1. JMR strains contain three plasmids.** Total genomic DNA was incubated in the absence (-) or presence (+) of PlasmidSafe DNase (PSD) and subsequently amplified with primers specific for the indicated DNA (pCD-ECE1 family: oAR1/2, pCD-ECE4 family: oAR3/4, pCD-WTSI1 family: oAR4/5; see **Table 2** and Materials and Methods).



**Figure 2. Comparison of pJMR5 plasmids with reference plasmids.** Within panels, colored arrows indicate similar genes; links are drawn between similar genes on neighboring clusters and are shaded based on sequence identity (0% white, 100% black, identity threshold for visualization 0.30). Dashed green lines indicate regions capable of sustaining replication of a shuttle vector in *C. difficile*. **A.** Comparison of pJMR5-1 (ON887052) with pCD-ECE1 (LR594544.1). **B.** Comparison of pJMR5-4 (ON887053) with pCD-ECE4 (LR594545.1). **C.** Comparison of pJMR5-W (ON887054) with pCD630 (AM180356.2) and pCD-WTS11 (MG019959.1). Image is based on a visualization with clinker [20].

## Acknowledgements

The authors would like to acknowledge B. Nibbering for isolating the total genomic DNA for PacBio sequencing, S. Nooij for assistance with PacBio-data analysis and I. Sidorov for the local implementation of clinker at the LUMC.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Declarations of interest:** none

**Author contributions:** AR: investigation, formal analysis; QRD: formal analysis, writing – original draft, visualization; BVHH: conceptualization, methodology, formal analysis; CH: investigation; MJTC: formal analysis; EJK: supervision; RHAMV: investigation, formal analysis; SLK: project administration, supervision, resources; WKS: conceptualization, formal analysis, visualization, project administration, resources, writing – original draft. All authors contributed to review & editing.