New tools and insights in physiology and chromosome dynamics of Clostridioides difficile

Oliveira Paiva, A.M.

Cover Page

Universiteit Leiden

The handle https://hdl.handle.net/1887/3158165 holds various files of this Leiden University dissertation.

Author: Oliveira Paiva, A.M.
Title: New tools and insights in physiology and chromosome dynamics of Clostridioides difficile
Issue Date: 2021-03-30

Identification of the unwinding region in the *Clostridioides difficile* chromosomal origin of replication

Ana M. Oliveira Paiva[1,2]

Erika van Eijk[1]

Annemieke H. Friggen[1,2]

Christoph Weigel[3]

Wiep Klaas Smits[1,2]

[1] Department of Medical Microbiology, Section Experimental Bacteriology, Leiden University Medical Center, Leiden, The Netherlands

[2] Center for Microbial Cell Biology, Leiden, The Netherlands

[3] Technische Universität Berlin, Institute of Biotechnology, Berlin, Germany

Abstract

Faithful DNA replication is crucial for the viability of cells across all kingdoms. Targeting DNA replication is a viable strategy for inhibition of bacterial pathogens. *Clostridioides difficile* is an important enteropathogen that causes potentially fatal intestinal inflammation. Knowledge about DNA replication in this organism is limited and no data is available on the very first steps of DNA replication. Here, we use a combination of *in silico* predictions and *in vitro* experiments to demonstrate that *C. difficile* employs a bipartite origin of replication that shows DnaA-dependent melting at *oriC2*, located in the *dnaA-dnaN* intergenic region. Analysis of putative origins of replication in different clostridia suggests that the main features of the origin architecture are conserved. This study is the first to characterize aspects of the origin region of *C. difficile* and contributes to our understanding of the initiation of DNA replication in clostridia.

Introduction

Clostridioides difficile (formerly *Clostridium difficile*) [1] is a gram-positive anaerobic bacterium. *C. difficile* infections (CDI) can occur in individuals with a disturbed microbiota and is one of the main causes of hospital-associated diarrhoea, but can also be found in the environment [2]. The incidence of CDI has increased worldwide since the beginning of the century [2,3]. Consequently, the interest in the physiology of the bacterium has increased as a way to understand its interaction with the host and the environment and to explore new pathways for intervention [4,5].

One such pathway is the replication of the chromosome. Overall, DNA replication is a highly conserved process across different kingdoms [6,7]. In all bacteria, DNA replication is a tightly regulated process that occurs with high fidelity and efficiency and is essential for cell survival. The process involves many different proteins that are required for the replication process itself, or to regulate and aid replisome assembly and activity [8-12]. Replication initiation and its regulation arguably are candidates for the search of novel therapeutic targets [4,13,14].

In most bacteria, replication of the chromosome starts with the assembly of the replisome at the origin of replication (*oriC*) and proceeds bidirectionally [10]. In the majority of bacteria, replication is initiated by the DnaA protein, an ATPase Associated with diverse cellular Activities (AAA+ protein) that binds specific sequences in the *oriC* region. The binding of DnaA induces DNA duplex unwinding, which subsequently drives the recruitment of other proteins, such as the replicative helicase, primase and DNA polymerase III proteins [10]. Termination of replication eventually leads to disassembly of the replication complexes [10].

In *C. difficile*, knowledge of DNA replication is limited. Though many proteins appear to be conserved between well-characterized species and *C. difficile*, only certain replication proteins have been experimentally characterized for *C. difficile* [15-17]. DNA polymerase C (PolC, CD1305) of *C. difficile* has been studied in the context of drug-discovery and appears to have a conserved primary structure similar to other low-[G+C] gram-positive organisms [15]. It is inhibited *in vitro* and *in vivo* by compounds that compete for binding with dGTP [18,19]. Helicase (CD3657), essential for DNA duplex unwinding, was found to interact in an ATP-dependent manner with a helicase loader (CD3654) and loading was proposed to occur through a ring-maker mechanism [17,20]. However, in contrast to helicase of the Firmicute *Bacillus subtilis*, *C. difficile* helicase activity is dependent on activation by the primase protein (CD1454), as has also been described for *Helicobacter pylori* [17,21]. *C. difficile* helicase stimulates primase activity at the trinucleotide 5'd(CTA), but not at the preferred trinucleotide 5'-d(CCC) [17].

DnaA of *C. difficile* has not been studied to date. Although no full-length structure has been determined for DnaA, individual domains of the DnaA protein from different organisms have been characterized [22-25]. DnaA proteins generally comprise four domains [25]. Domain I is involved in protein-protein interactions and is responsible for DnaA oligomerization [25-33]. Little is known about a specific function of domain II and this domain may even be absent [34]. It is thought to be a flexible linker that promotes the proper conformation of the other DnaA domains [27,35]. Domain III and Domain IV are responsible for the DNA binding. Domain III contains the AAA+ motif and is responsible for binding ATP, ADP and single-stranded DNA, as well as certain regulatory proteins [36-39]. Recent studies have also revealed the importance of this domain for binding phospholipids present in the bacterial membrane [40]. The C-terminal Domain IV contains a helix-turn-helix motif (HTH) and is responsible for the specific binding of DnaA to so-called DnaA boxes [34,41,42].

DnaA boxes are typically 9-mer non-palindromic DNA sequences, and the *E. coli* DnaA box consensus sequence is TTWTNCACA [43,44]. The boxes can differ in their affinity for DnaA, and even demonstrate different dependencies on the ATP co-factor [45,46]. Binding of domain IV to the DnaA boxes promotes higher-order oligomerization of DnaA, forming a filament that wraps around DNA [24,47,48]. It is thought that the interaction of the DnaA filament with the DNA helix introduces a bend in the DNA [24,46]. The resulting superhelical torsion facilitates the melting of the adjacent A+T-rich DNA Unwinding Element (DUE) [24,49,50]. Upon melting, the DUE provides the entry site for the replisomal proteins. Another conserved structural motif, a triplet repeat called DnaA-trio, is involved in the stabilization of the unwound region [51,52].

The *oriC* region has been characterized in several bacterial species. These analyses show that *oriC* regions are quite diverse in sequence, length and even chromosomal location, all of which contribute to species-specific replication initiation requirements [53,54]. In Firmicutes, including *C. difficile*, the genomic context of the origin regions appears to be conserved and encompasses the *rnpA-rpmH-dnaA-dnaN* genes [16,55].

The *oriC* region can be continuous (i.e. located at a single chromosomal locus) or bipartite [44]. Bipartite origins were initially identified in *B. subtilis* [56] but more recently also in *H. pylori* [57]. The separated subregions of the bipartite origin, *oriC1* and *oriC2*, are usually separated by the *dnaA* gene. Both *oriC1* and *oriC2* contain clusters of DnaA boxes, and one of the regions contains the DUE region. The DnaA protein binds to both subregions and places them in close proximity to each other, consequently looping out the *dnaA* gene [57,58]. In *H. pylori*, DnaA domain I and II are important for maintaining the interactions between both *oriC* regions [33].

In this study, we identified the putative *oriC* of *C. difficile* through *in silico* analysis and demonstrate DnaA-dependent unwinding of the *oriC2* region *in vitro*. Clear conservation of the origin of replication organization is observed throughout the clostridia. The present study contributes to our understanding of clostridial DNA replication initiation in general, and replication initiation of *C. difficile* specifically.

Materials and Methods

Sequence alignments and structure modelling

Multiple sequence alignment of amino acid sequences was performed with Protein BLAST (blastP suite, https://blast.ncbi.nlm.nih.gov/Blast.cgi) for individual alignment scores and the PRALINE program (http://www.ibi.vu.nl/programs/pralinewww/) [59] for multiple sequence alignment. Sequences were retrieved from the NCBI Reference Sequences. DnaA protein sequences from *C. difficile* 630Δ*erm* (CEJ96502.1), *C. acetobutylicum* DSM 1731 (AEI33799.1), *Bacillus subtilis* 168 (NP_387882.1), *Escherichia coli* K-12 (AMH32311.1), *Streptomyces coelicolor* A3(2) (TYP16779.1), *Mycobacterium tuberculosis* RGTB327 (AFE14996.1), *Helicobacter pylori J99* (Q9ZJ96.1) and *Aquifex aeolicus* (WP_010880157.1) were selected for alignment. The alignment was visualized in JalView version 2.11, with colouring by percentage identity.

Secondary structure prediction and homology modelling were performed using Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) [60] using the intensive default settings. Phyre2 modelling of *C. difficile* 630Δ*erm* DnaA (CEJ96502.1) was performed with 3 templates from *A. aeolicus* (PDB 2HCB, chain C), *B. subtilis* (PDB 4TPS, chain D) and *E. coli* (PDB 2E0G, chain A) and 21 residues were modelled *ab initio*. 95% of the residues were modelled with >90% confidence. Graphical representation was performed with the PyMOL Molecular Graphics System, Version 1.76.6. Schrödinger, LLC.

Prediction of the *C. difficile oriC*

To identify the *oriC* region of *C. difficile* the genome sequence of *C. difficile* 630Δ*erm* (GenBank accession no. LN614756.1) was analyzed through different software in a stepwise procedure [61].

The GenSkew Java Application (http://genskew.csb.univie.ac.at/) was used with default settings for the analysis of the normal and the cumulative skew of two selectable nucleotides of the genomic nucleotide sequence ($[G - C]/[G + C]$). Calculations were performed with a

window size of 4293 bp and a step size of 4293 bp. The inflexion values of the cumulative GC skew plot are indicative of the chromosomal origin (*oriC*) and terminus of replication (*ter*).

Prediction of superhelicity-dependent helically unstable DNA stretches (SIDDs) was performed in the vicinity of the inflexion point of the GC-skew plot, in 2.0 kb fragments comprising intergenic regions from nucleotide position 4291795 to 745 (*oriC1*) and 466 to 2465 (*oriC2*) of the *C. difficile* 630Δ*erm* chromosome. Prediction of the SIDDs in the different clostridia (Table 1) was performed in the vicinity of the inflexion points of the GC-plot retrieved from DoriC 10.0 database (http://tubic.tju.edu.cn/doric/public/index.php) [62], in 2.0 kb fragments comprising intergenic regions summarized in Table 1. The SIST program (https://bitbucket.org/benhamlab/sist_codes/src/master/) [63] was used to predicted free energies G(x) by running the melting transition algorithm only (SIDD) with default values (copolymeric energetics; default: σ =–0.06; T = 37°C; x= 0.01M) and with superhelical density σ = -0.04.

We performed the identification of the DnaA box clusters by search of the motif TTWTNCACA with one mismatch (Supplementary Information) in the leading strand on a 4432 bp sequence between the nucleotide position 4291488 to 2870 of the *C. difficile* 630Δ*erm* chromosome, using Pattern Locator (https://www.cmbl.uga.edu//downloads/programs/Pattern_Locator/patloc.c) [64]. Identification of the DnaA boxes in the different clostridia (Table 1) was performed with the same pattern motif in the leading strand of the intergenic regions summarized in Table 1.

Table 1 - Clostridia intergenic regions used for SIDD analysis.

Clostridia (GenBank accession no.)	oriC1*[1] DoriC ID*[2]	oriC2 DoriC ID*
C. difficile R20291 (NC_013316.1)	4189900 to 561 ORI93010593	780 to 2780 ORI93010592
C. botulinum A Hall (NC_009698.1)	3759361 to 800 ORI92010336	510 to 2510 ORI92010335
C. sordellii AM370 (NZ_CP014150	3549121 to 662 ORI97012279	561 to 2561 ORI97012278
C. acetobutylicum DSM 1731 (NC_015687.1)	3941422 to 961 ORI94010884	1040 to 3040 ORI94010883
C. perfringens str.13 (NC_003366.1)	3030241 to 810 ORI10010054	881 to 2881 ORI10010053
C. tetani E88 (NC_004557.1)	52001 to 54000 ORI10010089	50081 to 52081 ORI10010088

*[1] 2.0 kb fragments selected for SIDD analysis comprising the intergenic regions
*[2] DoriC 10.0 intergenic regions from http://tubic.tju.edu.cn/doric/public/index.php

DnaA-trio sequences and ribosomal binding sites where manually predicted based on Richardson et al. [51] and Vellanoweth and Rabinowitz [65], respectively.

All output data was obtained as raw text files and further processed with Prism 8.3.1 (GraphPad, Inc, La Jolla, CA) and CorelDRAW X7 (Corel).

Strains and growth conditions

E. coli strains were grown aerobically at 37°C in lysogeny broth (LB, Affymetrix) supplemented with 15 µg/mL chloramphenicol or 50 µg/mL kanamycin when required. *E. coli* strain DH5α (Table 2) for DnaA containing plasmid and *E. coli* MC1061 strain (Table 2) was used to maintain the *oriC* containing plasmids. *E. coli* MS3898 strain, kindly provided by Alan Grossman (MIT, Cambridge, USA) (Table 2) was used for recombinant DnaA expression. *E. coli* transformation was performed using standard procedures [66]. The growth was followed by monitoring the optical density at 600 nm (OD$_{600}$).

Table 2 - *E. coli* strains used in this study.

Name	Relevant Genotype/Phenotype*	Origin
DH5α	F− endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK−mK+), λ−	Laboratory collection
MC1061	str. K-12 F− λ− Δ(ara-leu)7697 [araD139]B/r Δ(codB-lacI)3 galK16 galE15 e14− mcrA0 relA1 rpsL150(StrR) spoT1 mcrB1 hsdR2(r−m+)	Laboratory Collection
CYB1002	ΔdnaA zia::pKN500(miniR1) asnB32 relA1 spoT1 thi-1 ilv192 mad1 recA1 λimm434 F- pBB42 (lacI; TetR)	Grossman lab

Construction of the plasmids

For overexpression of DnaA, the *dnaA* nucleotide sequence (CEJ96502.1) from *C. difficile 630Δerm* (GenBank accession no. LN614756.1) was amplified by PCR from *C. difficile 630Δerm* genomic DNA using primers oEVE-7 and oEVE-21 (Table 3). The PCR product was subsequently digested with *NcoI* and *BglII*. The vector pAV13 [67] (Table 4), containing *B. subtilis dnaA* cloned in pQE60 (Qiagen) was kindly provided by Alan Grossman (MIT, Cambridge, USA) and was digested with the same enzymes and ligated to the digested fragment to yield vector pEVE40 (Table 4).

Table 3 - Oligonucleotides used in this study.

Name	Sequence (5'>3') *
oEVE-7	CAGTCCATGGATATAGTTTCTTTATGGGACAAAACC
oEVE-21	CGGCAGATCTTCCCTTCAAATCTGATATAATTTTGTCTATTTTAG
oAP30	AATT<u>GAATTC</u>TTTGTCCCATAAAGAAACTATATCC
oAP31	TGGG<u>CTGCAG</u>TTCAACCCTTTAGTCCTATTAAAGTCC
oAP32	AATT<u>GAATTC</u>TTTGCTAGGATTTTTTGATTAC
oAP33	TGGG<u>CTGCAG</u>TTGACAAAATTATATCAGATTTG
oAP40	TGGG<u>CTGCAG</u>TTGCTAGGATTTTTTGATTAC
oAP41	AATT<u>GAATTC</u>TTTCAACCCTTTAGTCCTATTAAAGTCC
oAP56	CAGCGAGTCAGTGAGCGAGGAAG
oAP57	GATTGATTTAATTCTCATGTTTGAC

* Restriction enzyme cleavage sites used underlined

To construct a plasmid carrying the complete predicted *oriC*, the predicted *oriC* region (nucleotide 4292150 to 1593 from *C. difficile* 630 GenBank accession no. LN614756.1) was amplified by PCR from *C. difficile* 630Δ*erm* genomic DNA using primers oAP40 and oAP41 (Table 3). The PCR product was subsequently digested with *EcoRI* and *PstI* and ligated into pori1ori2 (Table 4), kindly provided by Anna Zawilak-Pawlik (Hirszfeld Institute of Immunology and Experimental Therapy, PAS, Wroclaw, Poland), that was digested with the same enzymes, to yield vector pAP205 (Table 4).

Table 4 - Plasmids used in this study.

Name	Relevant features*	Source/Reference
pAV13	lacI^q, P$_{T5}$ expression vector; *km*	[67]
pEVE40	P$_{T5}$ - DnaA-6xHis; *km*	This study
pori1ori2	*H. pylori oriC1oriC2; amp*	[57]
pAP76	*C. difficile oriC2; amp*	This study
pAP83	*C. difficile oriC1*; amp	This study
pAP205	*C. difficile oriC1oriC2; amp*	This study

* *amp* – ampicillin resistance cassette, *km* – kanamycin resistance cassette

For the cloning of the predicted *oriC1* region (nucleotide 4292150 to 24 of *C. difficile* 630Δ*erm* genomic DNA) the primer set oAP30/oAP31 (Table 3) was used. The amplified fragment was

digested with *EcoRI* and *PstI* and inserted onto pori1ori2 (Table 4) digested with the same enzymes, yielding vector pAP83 (Table 4). For the cloning of the predicted *oriC2* region (nucleotide 1291 to 1593 of *C. difficile* 630Δ*erm* genomic DNA) the primer set oAP32/oAP33 (Table 3) was used. The amplified fragment was digested with *EcoRI* and *PstI* and inserted onto pori1ori2 (Table 4) digested with the same enzymes, yielding vector pAP76 (Table 4).

All DNA sequences introduced into the cloning vectors were verified by Sanger sequencing. For *oriC* containing vectors, primers oAP56 and oAP57 (Table 3) were used for sequencing.

Overproduction and purification of DnaA-6xHis

Overexpression of DnaA-6xHis was carried out in *E. coli* strain CYB1002 (Table 2), harbouring the expression plasmid pEVE40 (Table 4). Cells were grown in 800 mL LB and induced with 1mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6 for 3 hours. The cells were collected by centrifugation at 4°C and stored at -80°C. Cells were resuspended in Binding buffer (1X Phosphate buffer pH7.4, 10 mM Imidazol, 10% glycerol) lysed by French Press and collected in phenylmethylsulfonyl fluoride (PMSF) at 0.1 mM (end concentration). Separation of the soluble fraction was performed by centrifugation at 13000xg at 4°C for 20 min. Purification of the protein from the soluble fraction was done in Binding buffer on a 1 mL HisTrap Column (GE Healthcare) according to the manufacturer's instructions. Elution was performed with Binding buffer in stepwise increasing concentrations of imidazole (20, 60, 100, 300 and 500 mM). DnaA-6xHis was mainly eluted at a concentration of imidazole equal to or greater than 300mM.

Fractions containing the DnaA-6xHis protein were pooled together and applied to Amicon Ultra Centrifugal Filters with 30 kDa cutoff (Millipore). Buffer was exchanged to Buffer A (25 mM HEPES-KOH pH 7.5, 100 mM K-glutamate, 5 mM Mg-acetate, 10% glycerol). The concentrated DnaA protein was subjected to size exclusion chromatography on an Äkta pure instrument (GE Healthcare). 200 μL of concentrated DnaA-6xHis was applied to a Superdex 200 Increase 10/30 column (GE Healthcare) in buffer A at a flow rate of 0.5 ml min^{-1}. UV detection was done at 280 nm. The column was calibrated with a mixture of proteins of known molecular weights (Mw): thyroglobulin (669 kDa), apoferritin (443 kDa), β-amylase (200 kDa), albumin (66 kDa) and carbonic anhydrase (29 kDa). Eluted fractions containing DnaA-6xHis of the expected molecular weight (51 kDa) were quantified and visualized by Coomassie. Pure fractions were aliquoted and stored at -80°C for further experiments.

Immunoblotting and detection

For immunoblotting, proteins were separated on a 12% SDS-PAGE gel and transferred onto nitrocellulose membranes (Amersham), according to the manufacturer's instructions. The membranes were probed in PBST (PBS pH 7,4, 0,05% (v/v) Tween-20) with the mouse anti-his antibody (1:3000, Invitrogen) and the respective secondary antibody goat anti-mouse-HRP (1:3000, DAKO) were used. The membranes were visualized using the chemiluminescence detection kit Clarity ECL Western Blotting Substrates (Bio-Rad) in an Alliance Q9 Advanced machine (Uvitec).

P1 nuclease Assay

For the P1 nuclease assay, 100 ng pAP205 plasmid was incubated with increasing concentrations of DnaA-6xHis (0.14, 0.54, 1 and 6.3 µM), when required, in P1 buffer (25mM Hepes-KOH (pH 7.6), 12% (v/v) glycerol, 1mM $CaCl_2$, 0.2mM EDTA, 5mM ATP, 0.1 mg/ml BSA), at 30°C for 12 min. 0.75 unit of P1 nuclease (Sigma), resuspended in 0.01 M sodium acetate (pH 7.6) was added to the reaction and incubated at 30°C for 5 min. 220 µl of buffer PB (Qiagen) was added and the fragments purified with the miniElute PCR Purification Kit (Qiagen), according to manufacturer's instructions. Digestion with BglII, NotI or ScaI (NEB) of the purified fragments was performed according to the manufacturer's instructions for 1 hour at 37°C. Digested samples were resolved on 1% agarose gels in 0.5xTAE (40 mM Tris, 20 mM CH_3COOH, 1 mM EDTA PH 8.0) and stained with 0.01 mg/mL ethidium bromide solution afterwards. Visualization of the gels was performed on the Alliance Q9 Advanced machine (Uvitec). Images were processed in CorelDraw X7 software. For all experiments at least three independent replicates were performed with various concentrations of DnaA. To quantify the results, background-corrected band intensities were determined using ImageJ, values were normalized against the total signal in a lane in MS Excel, and plotted using GraphPad.

Results

C. difficile DnaA protein

C. difficile 630Δ*erm* encodes a homolog of the bacterial replication initiator protein DnaA (GenBank: CEJ96502.1; CD630DERM_00010). Alignment of the full-length *C. difficile* DnaA amino acid sequence with selected DnaA homologs from other organisms demonstrates a sequence identity of 35% to 67%, with an even higher similarity (57% to 83%, Fig. 1A). *C. difficile* DnaA displays a greater sequence identity between the low-[G+C] Firmicutes (> 60%). When compared with the extensively studied DnaA proteins from *E. coli* and *B. subtilis*, the

full-length protein has 43% and 62% identity, and a similarity of 63% and 78%, respectively (Fig. 1A).

To assess the structural properties of *C. difficile* DnaA, we predicted the secondary structure and generated a model of the protein using Phyre2 [60] (Fig. 1B). The predicted DnaA model is based on three DnaA structures from different organisms: *A. aeolicus* (residues 101 to 318 and 334 to 437)[24] for domain III and IV, and *B. subtilis* (residues 2 to 79) [29] and *E. coli* (residues 5 to 97) [27] for domain I and II.

Domain I of DnaA mediates interactions with a diverse set of regulators and is involved in DnaA oligomerization [25,33]. We observe limited homology of *C. difficile* DnaA domain I with the equivalent domain of the selected organisms (Fig. 1A), although the overall fold is conserved (Fig. 1B). Nevertheless, some residues (P45, F48) appear to be conserved in most of the selected organisms (Fig. 1A), though no functional role for these residues is known. Potentially, these residues might be involved in protein-protein interactions or DnaA oligomerization, as these functions have been mapped to domain I of DnaA [25-33].

Domain II is a flexible linker that is possibly involved in aiding the proper conformation of the DnaA domains, and thus requires a minimal length for DnaA function *in vivo* [35]. No clear sequence similarity is observed on domain II and modelling of the *C. difficile* DnaA protein suggests a putative disordered nature of this domain (Fig. 1).

Domain III is responsible for binding to the co-factors ATP and ADP, and in conjunction with domain IV essential for DNA binding [36,38,39]. Within domain III we readily identified the Walker A and Walker B motifs (WA and WB in Fig. 1A) of the AAA+ fold (residues 135-317), crucial for binding and hydrolyzing ATP. This domain is highly conserved among all the selected organisms (Fig. 1A) and comprises a structural centre of β-sheets (Fig. 1B, pink domain). Other features of the AAA+ ATPase fold are present and conserved between the organisms, such as the sensor I and sensor II motifs required for the nucleotide binding (I and II, Fig.1A). The arginine finger motif (the equivalent of R285 of *E.coli* DnaA in the VII box), important for the ATP dependent activation of DnaA [36], is conserved in *C. difficile* DnaA as well (R256 in motif box VII; Fig. 1A).

A
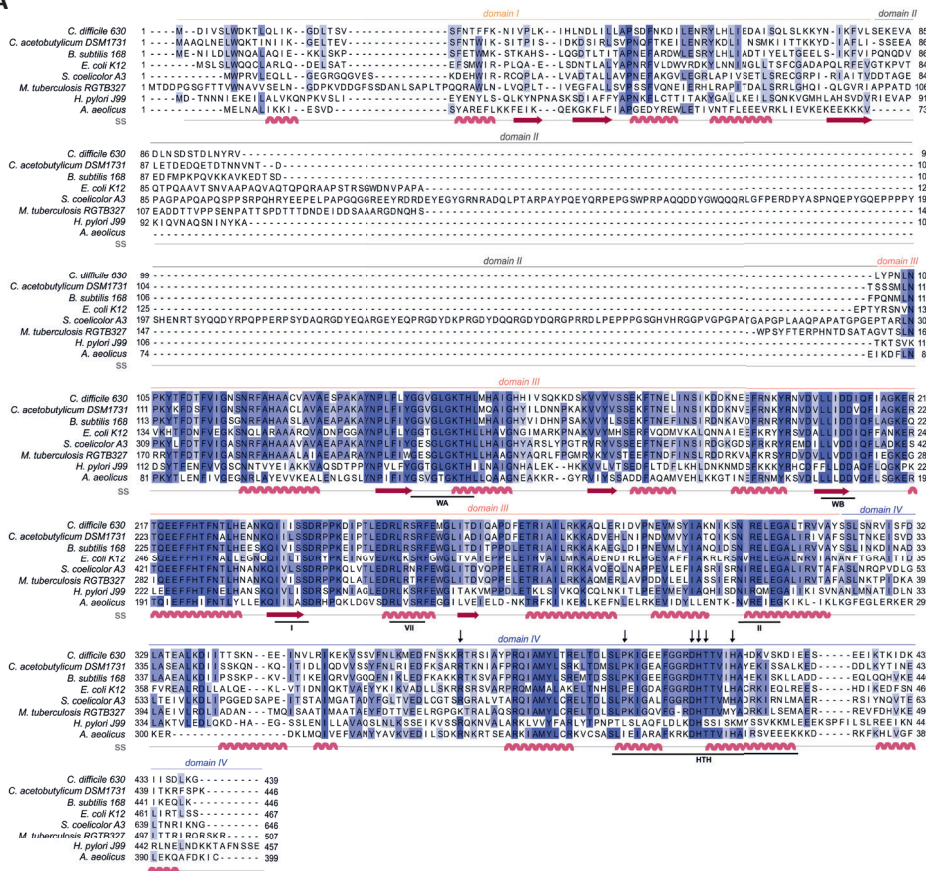
B
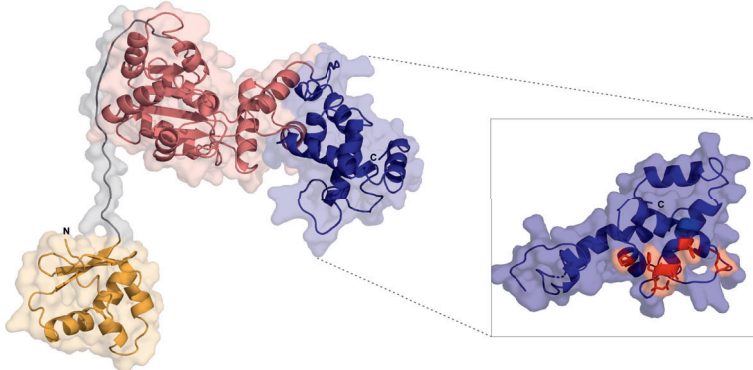
Fig. 1 - *C. difficile* DnaA DNA binding domain is conserved. A) Multiple sequence alignment (PRALINE) of *C. difficile* DnaA with homologous proteins retrieved from GenBank. The amino acid sequences from *C. difficile* 630Δerm (CEJ96502.1), *C. acetobutylicum* DSM 1731 (AEI33799.1), *B. subtilis* 168 (NP_387882.1), *E. coli* K-12 (AMH32311.1), *S. coelicolor* A3(2) (TYP16779.1), *M. tuberculosis* RGTB327 (AFE14996.1), *H. pylori* J99 (Q9ZJ96.1) and *Aquifex aeolicus* (WP_010880157.1) were used. Residues are colored according to sequence identity conservation using blue shading (dark blue more conserved), as

analysed in JalView. Secondary structure prediction (ss) is indicated, according to Phyre2 modelled structure. DnaA domains are represented, with the conserved AAA+ ATPase fold motifs Walker A, Walker B, VII box, sensor I and sensor II highlighted (WA, WB, I, VII and II motifs), as well as the domain IV helix-turn-helix (HTH). Residues involved in the base-specific recognition of the 9-mer DnaA box sequence are identified with an arrow. **B)** Structural model of *C. difficile* DnaA determined by Phyre2. Domains are coloured as in alignment. Both the N-terminus and the C-terminus are indicated in the figure. The DnaA domain IV is enhanced (inset) with the DnaA-box binding specific residues represented in red sticks.

The C-terminal domain IV of the DnaA protein (residues 317 to 439, Fig. 1A), contains the HTH motif required for the specific binding to DnaA-boxes [23,34]. Previous studies identified several residues involved in specific interactions with the DnaA boxes, that bind through hydrogen bonds and van der Waals contacts with thymines present in the DNA sequence [41,42,68]. The residues are conserved among all Firmicutes and *E. coli,* including the residues R371 (position R399 in *E. coli*), P395 (P423), D405 (D433), H406 (H434), T407 (T435), and H411 (H439), (Fig. 1B inset, red residues) [42]. Structural modelling of *C. difficile* DnaA predicts these residues to be exposed, providing an interface for DNA binding (Fig. 1B). Residues involved in base-specific recognition of the DnaA box sequence are conserved between the Firmicutes and *E. coli* (Fig. 1A), suggesting that *C. difficile* DnaA is likely to recognize the consensus DnaA box TTWTNCACA [43]. Notably, with the exception of a single arginine, these residues are not conserved between *C. difficile* and *Thermotoga maritima* DnaA (Fig. S1). As the latter recognizes an extended 12-bp motif [52,69], this provides additional support for the notion that *C. difficile* DnaA recognizes a classical 9-bp DnaA box. In addition, residues found to be involved in non-specific interactions with the phosphate backbone of the DNA (some of which contribute to sequence specificity) [42,68] appear less conserved between the selected organisms (Fig. 1A).

Expression and purification of DnaA-6xHis

To allow for *in vitro* characterization of DnaA activity, we recombinantly expressed the *C. difficile* DnaA with a C-terminal 6xHis-tag in *E. coli* cells. To prevent the co-purification of *C. difficile* DnaA with host DnaA protein, *E. coli* strain CYB1002 was used (a kind gift of A.D. Grossman). This strain is a derivative of *E. coli* MS3898, that lacks the *dnaA* gene and replicates in a DnaA-independent fashion [70]. Induction of the DnaA-6xHis protein was confirmed by Coomassie staining and immunoblotting with anti-his antibody at the expected molecular weight of 51 kDa (Fig. S2A, red arrow). Upon overexpression of DnaA-6xHis, smaller fragments were observed, which accumulated with a prolonged time of expression (Fig. S2A), most likely corresponding to proteolytic fragments of the DnaA-6xHis protein.

Purification of the recombinant DnaA-6xHis showed a clear band at the expected size when eluted at 300 mM imidazole concentration, but several lower molecular size bands were observed (Fig. S2B). Therefore, the eluted fractions where further purified with size exclusion chromatography (SEC). This yielded a single product at the expected molecular weight of DnaA-6xHis, and its identity was confirmed by western-blot with anti-his antibody (Fig. S2C, red arrow). A minor band of lower molecular weight (approximately 38 kDa, <1% of total protein) was observed (Fig. S2C, green asterisk), which may reflect some instability of the N-terminus of the DnaA-6xHis protein, as it appears to have retained the C-terminal 6xHis tag.

In silico prediction of the *oriC* region

To identify the *oriC* region and the elements that are part of it (DUE, DnaA-trio and DnaA boxes) we performed different prediction approaches in a stepwise procedure, as initially described [61].

We first analyzed the DNA asymmetry of the genome of *C. difficile* 630Δ*erm* (GenBank accession no. LN614756.1) [71], by plotting the normalized difference of the complementary nucleotides (GC-skew plot) [72]. *C. difficile* 630Δ*erm* has a circular genome of 4293049 bp and an average [G+C] content of 29.1%. We used the GenSkew Java Application (http://genskew.csb.univie.ac.at/) for determining the chromosomal asymmetry. Asymmetry changes in a GC-skew plot can be used to predict the origin of replication region and the terminus region of bacterial genomes. Based on this analysis, the origin is predicted at approximately position 1 of the chromosome. The terminus location is predicted at approximately 2.18 Mbp from the origin region (Fig. 2A). These results were confirmed when artificially reassigning the starting position of the chromosomal assembly (data not shown). The gene organization in the putative origin region is *rnpA-rpmH-dnaA-dnaN* (position 4291488 to 2870, Fig. 2B), identical to the origin of *B. subtilis* [16,73], and therefore encompasses the *dnaA* gene (CD630DERM_00010).

We next used the SIST program [63] to localize putative DUEs in the intergenic regions in the chromosomal region predicted to contain the *oriC*. Hereafter we refer to these regions as *oriC1* (in the intergenic region of *rpmH-dnaA*) and *oriC2* (in the intergenic region *dnaA-dnaN*), in line with nomenclature in other organisms [57,73] (Fig. 3B). SIST identifies helically unstable AT-rich DNA stretches (Stress-Induced Duplex Destabilization regions; SIDDs) [57,63]. In regions with lower free energy ($G_{(x)} < y$ kcal/mol), the double-stranded helix has a high probability to become single-stranded DNA. With increasing negative superhelicity ($\sigma = -0.06$, Fig. 2C, green line) regions of both *oriC1* and *oriC2* become single-stranded DNA ($G_{(x)} < 2$ kcal/mol). At low negative superhelicity ($\sigma = -0.04$, Fig. 3C, red line) short stretches of DNA of approximately 27
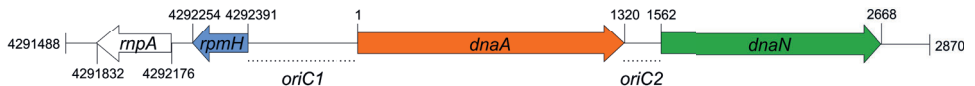
bp were identified with a significantly lower free energy. These regions with lower free energy at a negative superhelicity of −0.04 and −0.06 are potential DUE sites. The nucleotide sequence of the possible unwinding elements identified are represented in detail in Fig. 3 (grey boxes).
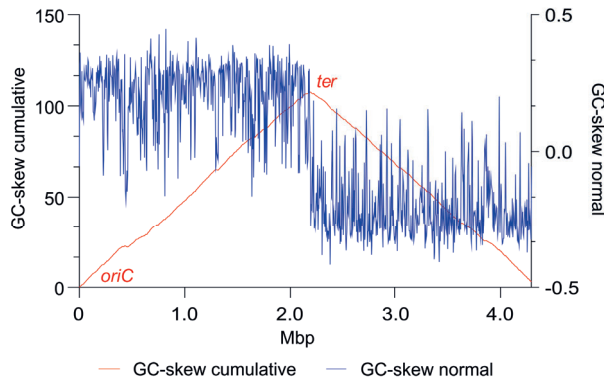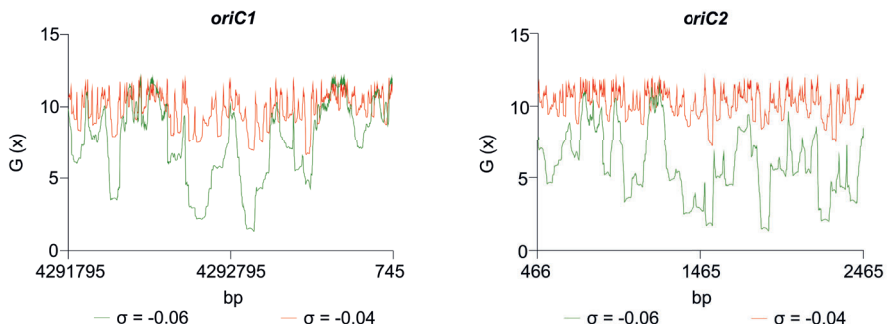

Fig. 2 - Prediction of the *C. difficile* origin of replication. A) GC skew analysis of the *C. difficile* 630Δerm (LN614756.1) genome sequence. Normal GC skew analysis ([G − C]/[G + C]) performed on leading strand (blue line) and respective cumulative GC skew plot (red line). Calculations were performed with a window size of 4293 bp and a step size of 4293 bp. The *origin (oriC) and terminus (ter) regions* are indicated. **B)** Representation of the predicted origin region and genomic context (from residues at position 4291488 to 2870 of the *C. difficile* 630 Δerm chromosome). The *rmpA, rpmH* (blue arrow), *dnaA* (orange arrow) and *dnaN* (green arrow) genes are indicated. Putative origins in intergenic regions are represented *oriC1* (*rpmH-dnaA*) and *oriC2* (*dnaA-dnaN*). **C)** SIDD analysis of 2.0 kb fragments comprising

oriC1 (nucleotide 4291795 to 745) and *oriC2* (nucleotide 466 to 2465). Predicted free energies G(x) for duplex destabilization at a superhelical density of σ = -0.06 (green) or σ = -0.04 (red).
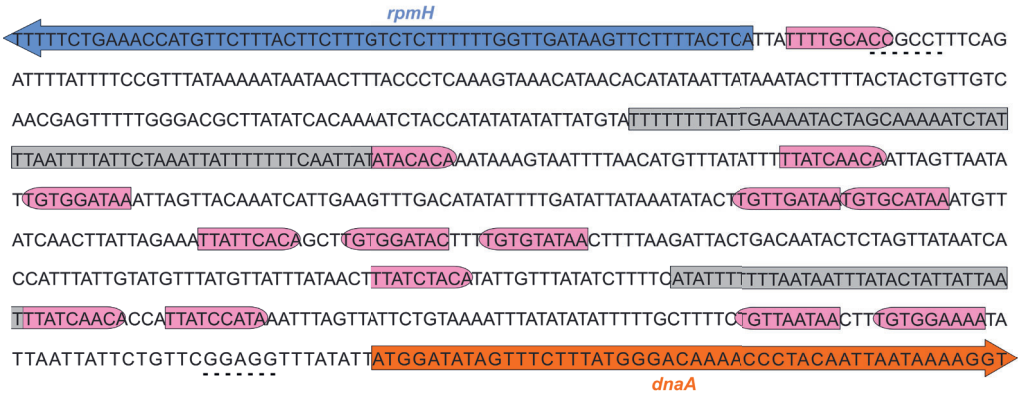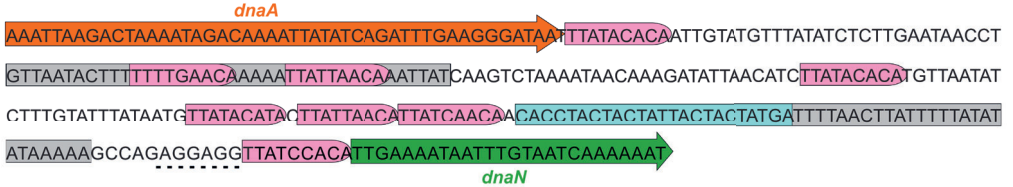
Fig. 3 - Identification of the *C. difficile* *oriC* region. Nucleotide sequence of the *oriC1* region (nucleotide 4292328 to 48 of the *C. difficile* 630Δerm LN614756.1 genome sequence) and *oriC2* region (nucleotide 1274 to 1587). Identification of the possible unwinding AT-rich regions previously identified in the SIDD analysis (grey boxes). The putative DnaA boxes found are represented (pink boxes) and orientation in the leading (right) and lagging strand (left) are shown. Possible DnaA-trio sequence is denoted (light blue boxes). Coding sequence of the genes *rpmH* (blue arrow), *dnaA* (orange arrow) and *dnaN* (green arrow) and respective putative ribosome binding sites (dashed line) are indicated. Pattern identification is described in Material and Methods.

We then performed the identification of DnaA box clusters through a search of the consensus DnaA box TTWTNCACA containing up to one mismatch, using Pattern Locator [64]. 22 putative DnaA boxes were identified in both the leading and lagging strand in the predicted *C. difficile* *oriC* regions (Fig. 3, pink boxes), 14 in the *oriC1* region and 8 in the *oriC2* region. Both the consensus DnaA box TTWTNCACA and variant boxes are found. A cluster of DnaA boxes was proposed to contain at least three boxes with an average distance lower than 100 bp in between [61]. At least one such cluster can be found in each origin region (Fig. 3).

Though these are not crucial to origin function, we also manually identified the putative ribosomal binding sites for the annotated genes (Fig. 3, dashed line) based on previously identified characteristics [65].

Finally, we manually predicted DnaA-trio sequences (3'-[G/A]A[T/A]$_{n>3}$-5' preceded by a GC-cluster) in the predicted *oriC* regions, as this motif is required for successful replication in both *E. coli* and *B. subtilis* [51] and can also be identified in *E. coli* [74], though a role in the binding of DnaA to ssDNA has yet to be experimentally demonstrated in this organism. We identified a clear DnaA-trio in the lagging strand upstream of a predicted DUE region in the *oriC2* region, with the nucleotide sequence 5'-CACCTACTACTATTACTACTATGA-3' (Fig. 3, light blue box), but no clear DnaA-trio was identified in the *oriC1* region.

From all the observations, we anticipate that a bipartite origin is located in the *dnaA* chromosomal region of *C. difficile* with unwinding occurring downstream of *dnaA*, at the *oriC2* region.

DnaA-dependent unwinding

To analyze DnaA-dependent unwinding of *oriC*, we used the purified *C. difficile* DnaA-6xHis protein and the predicted *oriC* sequence, to perform P1 nuclease assays as previously described [57,75]. Localized melting resulting from DnaA activity exposes ssDNA to the action of the ssDNA-specific P1 nuclease. After incubation of a vector containing the *oriC* fragment with DnaA protein and cleavage by the P1 nuclease, the vector is purified and digested with different endonucleases to map the location of the unwound region.

We constructed vectors, based on pori1ori2 [57], harbouring *C. difficile oriC1* (pAP76) or *oriC2* (pAP83) individually (Fig. S3A), as well as the complete *oriC* region (pAP205) (Fig. 4A). For a more accurate determination of the unwound region, the vectors were subjected to digestion by two different restriction enzymes (BglII and NotI), resulting in different restriction patterns. A limited spontaneous unwinding of the plasmid was observed in the *C. difficile oriC*-containing vectors (Fig. 4A and S3B). No DnaA-dependent change in restriction pattern was observed when using the single *oriC* regions (Fig. S3B), suggesting *oriC1* and *oriC2* individually lack the requirements for DnaA-dependent unwinding.

We did observe a DnaA-dependent change in digestion patterns for the *oriC1oriC2*-containing vector pAP205 (Fig. 4). Digestion of this vector with BglII in the absence of DnaA-6xHis and P1 nuclease resulted in a linear DNA fragment (4638 bp) due to the presence of a unique BglII restriction site (Fig. 4B, upper panel, first lane). The addition of P1 nuclease leads to the appearance of a faint band between 1650 and 3000 bp (Fig. 4B, upper panel, second lane), consistent with previous observations that the presence of a plasmid DUE can result in low-level spontaneous unwinding due to the inherent instability of these AT-rich regions [76]. Upon

the addition of the DnaA-6xHis protein, the observed band becomes more intense, suggesting a strong increase in unwinding (Fig. 4B, upper panel, red arrow).

Digestion of pAP205 with NotI in the absence of DnaA-6xHis and P1 nuclease results in fragments of 3804 and 842 bp, due to two NotI recognition sites in the vector (Fig 4B, lower panel, first lane). In the presence of just P1 nuclease, a similar low level of spontaneous unwinding is observed, resulting in the appearance of two additional faint bands, one between 1650 and 3000 bp and other between 1000 and 1650 bp (Fig. 4B, lower panel, second lane). The addition of DnaA-6xHis results in an increase in intensity of both these bands in a dose-dependent manner (Fig. 4A, lower panel, red arrows).
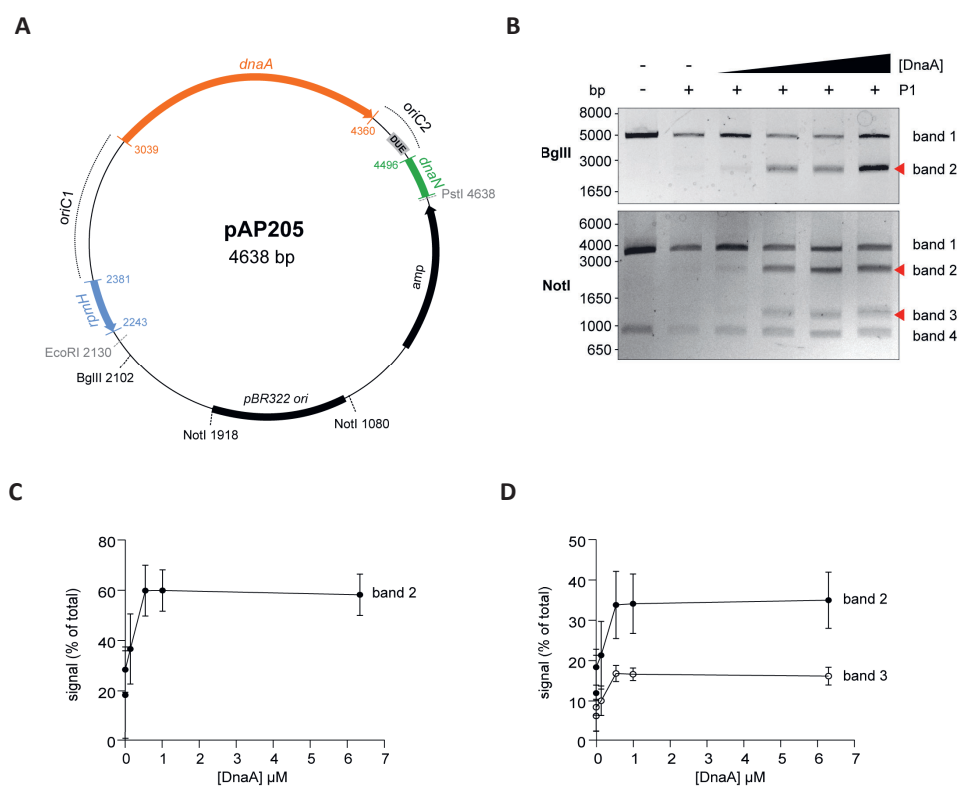
Fig. 4 - Identification of the unwinding region in *C. difficile* oriC. A) Representation of the *oriC1oriC2* containing vector pAP205 used in the P1 nuclease assay. The predicted *oriC1* and *oriC2* regions (dotted lines) and included genes are represented, *rpmH* (blue), *dnaA* (orange), and *dnaN* (green). The *bla* gene, the pBR322 plasmid origin of replication and the positions of used restriction sites are marked. The unwinding region (DUE) is denoted in a grey circle. **B)** P1 nuclease assay of the *oriC1oriC2*-containing vector pAP205. Digestion of the vector (lane 1) with different restriction enzymes *BglII* (upper panel) or *NotI* (lower panel). Treatment of the fragments with P1 nuclease only (lane 2) and incubated with increasing amounts of *C. difficile* DnaA protein (lanes 3-6). The DNA fragments were separated in a 1% agarose gel and analyzed after ethidium bromide staining. Fragments resulting from DnaA-dependent

unwinding are indicated with a red arrow (see Results for details). A typical result is shown. **C)**. Quantification of band 2 (black circles) of the P1/*BglII* digested vector. **D)**. Quantification of bands 2 (black circles) and 3 (open circles) of the P1/*NotI* digested vector. For panels C and D, error bars indicate the standard deviation of the mean of n=3 independent experiments.

We quantified the intensity of the bands from three independent P1 nuclease assays in order to determine the reproducibility of the assay (Fig. 4C, 4D and Fig. S4). For the BglII-digested vector, we observed a DnaA-dependent increase of 20 to 60% of the total signal for the band between 1650 and 3000bp (Fig. 4C, band 2). For the NotI-digested vector, the signals of the second and third band increase from approximately 10% of the total signal to approximately 35% (1650-3000bp, band 2) and 20% (1000-1650bp, band 3) of total signal in the lane (Fig. 4D). The observed increase was highly consistent and appeared to saturate around 0.54-1 uM of DnaA (Fig. 4C and 4D). The quantification also revealed a concomitant decrease in the signal for the upper bands in the gels of the BglII and NotI digests (Fig. S4, band 1).

The DnaA-dependent appearance of the ~2000 bp band in the BglII digest, and the ~1200 and ~2200bp bands in the NotI digest localize the DnaA-dependent unwinding of the *C. difficile oriC* in the *oriC2* region (Fig. 4A, grey rectangle, DUE). Moreover, these results suggest that *C. difficile* has a bipartite origin of replication, as successful DnaA-dependent unwinding of *C. difficile* in the *oriC2* region requires both *oriC* regions (*oriC1* and *oriC2*).

**Conservation of the origin organisation in related Clostridia**

Our results suggest that the origin organization of *C. difficile* resembles that of a more distantly related Firmicute, *B. subtilis*. To extend our observations, we evaluated the genomic organization of the *oriC* region in different organisms phylogenetically related to *C. difficile.* We followed a similar approach as described above for *C. difficile* 630Δ*erm,* taking advantage of the DoriC 10.0 database [62]. Importantly, our results with respect to the *C. difficile* origin of replication described above were largely congruent with the DoriC 10.0 database despite being based on different methods (a notable exception is the prediction for *C. difficile* strain 630; data not shown). We retrieved the predicted *oriC* regions from the DoriC 10.0 database and performed an in-depth analysis of these regions for the closely related *C. difficile* strain R20291 (NC_013316.1), as well as the more distantly related *C. botulinum* A Hall (NC_009698.1), *C. sordellii* AM370 (NZ_CP014150), *C. acetobutylicum* DSM 1731 (NC_015687.1), *C. perfringens* str.13 (NC_003366.1) and *C. tetani* E88 (NC_004557.1) (Table 1).

Similar to *C. difficile* 630Δ*erm*, the genomic context of the origin contains the *rpmH-dnaA-dnaN* region for most of the clostridia selected and mirrors that of *B. subtilis* (Fig. 5). The only

exception is *C. tetani* E88 where the uncharacterized CLOTE0041 gene lies upstream of the *dnaA-dnaN* cluster (Fig. 5).

We also identified the possible DnaA boxes for the selected clostridia (Fig. 5, pink semi-circle). Across the analyzed clostridia, *oriC1* region presented more variability in the number of putative DnaA boxes, from 9 to 19, whereas *oriC2* contained 5 to 9 DnaA boxes, with *C. tetani* E88 with the lowest number of possible DnaA boxes, both at the *oriC1* (9 boxes) and *oriC2* (5 boxes) regions (Fig. 5, pink semi-circle). In all the organisms we observe at least 1 DnaA cluster in each origin region, as also observed for *C. difficile* 630Δ*erm*.

Prediction of DUEs using the SIST program [63] identified several helically unstable regions that are candidate sites for unwinding (Fig. 5, dashed lines, and Fig. S5). Notably, in all cases, one such region in *oriC2* (Fig. 5, grey circle) is preceded immediately by the manually identified DnaA-trio (Fig. 5, light blue circle). Based on our experimental data for *C. difficile* 630Δ*erm*, we suggest that in all analyzed clostridia, DnaA-dependent unwinding occurs at a conserved DUE downstream of the DnaA-trio in the *oriC2* region (Fig. 5).
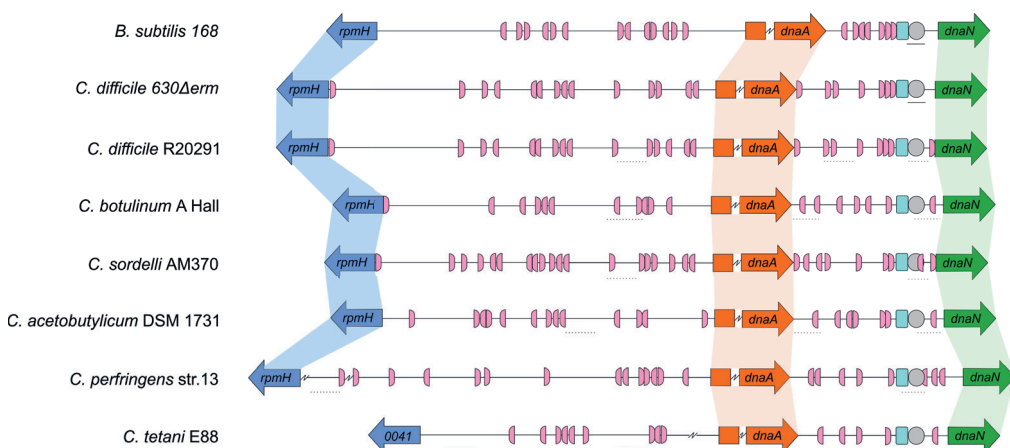
Fig. 5 - Comparison of the clostridial *oriC* regions. Representation of the origin region and genomic context of B. subtilis, *C. difficile* 630Δ*erm* chromosome and the predicted regions for *C. difficile* R20291, *C. botulinum* A Hall, *C. sordellii* AM370, *C. acetobutylicum* DSM 1731, *C. perfringens* str.13, *C. tetani* E88 (see Table 1). The *rpmH* (blue arrow), *dnaA* (orange arrow) and *dnaN* (green arrow) genes are indicated. Predicted DnaA-boxes are indicated by pink boxes and orientation on the leading (right) and lagging strand (left) are shown. Identification of the experimentally identified unwinding AT-rich regions (lines) and the SIDD-predicted helical instability are shown (dashed lines). The putative DUE is denoted (grey circle). Possible DnaA-trio sequences are shown in light blue boxes. See Material and Methods for detailed information. Alignment of the represented chromosomal regions is based on the location of the DnaA-trio.

Discussion

Chromosomal replication is an essential process for the survival of the cell. In most bacteria, DnaA protein is the initiator protein for replication and through a cascade of events leads to the successful loading of the replication complex onto the origin of replication [10,77].

Initial characterization of bacterial replication has been assessed in the model organisms *E. coli* and *B. subtilis* [11]. Despite the similarities (location in an intergenic region, presence of a DUE, several DnaA boxes in both orientations) the structure of the replication origins and the regulation mechanisms are variable among bacteria [44]. In contrast to *E. coli*, *B. subtilis* origin region is bipartite, with two intergenic regions upstream and downstream the *dnaA* gene. In *C. difficile* the genomic organization in the predicted cluster *rnpA-rpmH-dnaA-dnaN,* and the presence of AT-rich sequences in the intergenic regions is consistent with a bipartite origin, as in *B. subtilis (*Fig. 3).

The origin region contains several DnaA-boxes with different properties that are recognized by the DnaA protein. The specific binding of DnaA to the DnaA-boxes is mediated mainly through domain IV of the DnaA protein. From DNA bound structures of DnaA it was possible to identify several residues involved in the contact with the DnaA boxes, some of which confer specificity [41,42,68]. Analysis of the of *C. difficile* DnaA homology in domain IV did not show any difference in the residues involved on the DnaA-box specificity (Fig.1, vertical arrows), suggesting the same consensus motif conservation as the DnaA-box TTWTNCACA for *E.coli* [43]. The conserved DnaA-box motif allowed us to identify several DnaA boxes along the intergenic regions of the *oriC*. Like in the bipartite origin of *B. subtilis,* we identified at least one cluster of DnaA-boxes in the *C. difficile oriC1* and *oriC2* regions (Fig. 3 and 5). In the case of *B. subtilis*, it has been shown that different DnaA boxes fulfil different roles in replication initiation: two out of three DnaA boxes immediately upstream of the DnaA-trio are part of the basal unwinding system (i.e. required for DnaA-dependent strand separation), whereas other DnaA affect coordination and regulation of DNA replication [52]. For *C. difficile*, we also find three DnaA boxes immediately upstream of the DnaA trio (Fig. 3 and Fig. S6), but the role of these boxes has not been experimentally verified to date.

The P1 nuclease assays place a region in which DnaA-dependent unwinding occurs in the *oriC2* region of *C. difficile*, supported by the presence of the several features on the *oriC2*, such as the identified DUE and DnaA-trio, both required for unwinding [49,51]. The presence of both *oriC* regions (*oriC1* and *oriC2*) is required for melting *in vitro*, as observed for other bipartite origins [44]. In contrast to the bipartite origin identified in *H. pylori* [57], we did not observe unwinding of the *oriC2* region alone. Though this may be a specific aspect of *C. difficile oriC2*, we cannot

exclude that differences in the experimental setup (e.g. DnaA protein purification) could affect these observations. Nevertheless, our data are consistent with DnaA binding the DnaA-box clusters in both *oriC* regions, leading to potential DnaA oligomerization, loop formation, and unwinding at the AT-rich DUE site.

When analyzing the origin region between different clostridia, features similar to those of *C. difficile* are observed, such as conservation of DnaA-box clusters within both *oriC* regions in the vicinity of the *dnaA* gene. Similar to *C. difficile* and *B. subtilis*, a putative DUE element, preceded by the DnaA-trio, was also located within the *oriC2* region (Fig. 4 and 6). Thus, the overall origin organization and mechanism of DNA replication initiation is likely to be conserved within the Firmicutes [16]. As spacing of the DnaA-boxes are determinants for the species-specific effective replication [23,53], these similarities do no exclude the possibilities that subtle differences in replication initiation exist, and further studies are required. For instance, our work does not address which DnaA boxes in either *oriC1* or *oriC2* are important for unwinding, and whether the requirement is due to DnaA-dependent changes in structure of origin DNA (as has been shown for *B. subtilis*) [52], or as a cis-acting regulatory element like DARS/DatA [8,74]. Further experiments could provide insights into the DnaA-box conservation and affinities and establish which DnaA boxes are crucial for origin firing and/or transcriptional regulation

Several proteins can interact with the *oriC* region or DnaA, including YabA, Rok, DnaD/DnaB, Soj and HU [11,16]. In doing so they shape the origin conformation and/or stabilize the DnaA filament or the unwound region, consequently affecting replication initiation.

YabA or Rok affect *B. subtilis* replication initiation [12,78,79], but no homologs of these proteins have been identified in *C. difficile* [4]. Similarly, no homologs are identified of other well-characterized DnaA-interacting proteins from gram-negative bacteria [4], such as Hda, DiaA/HobA [25] or HdaB [80]; it is unknown how *C. difficile* regulates DnaA activity.

In *B. subtilis*, DnaD, DnaB and DnaI helicase loader proteins associate sequentially with the origin region resulting in the recruitment of the DnaC helicase protein [11,81-83]. In *B. subtilis*, DnaD binds to DnaA and it is postulated that this affects the stability of the DnaA filament and consequently the unwinding of the *oriC* [31,32,84]. *B. subtilis* DnaB protein also affects the DNA topology and has been shown to be important for recruiting *oriC* to the membrane [85,86]. *C. difficile* lacks a homologue for the DnaB protein, although the closest homolog of the DnaD protein (CD3653) [4] may perform similar functions in the origin remodelling [17]. Direct interaction of DnaA-DnaD through the DnaA domain I was structurally determined and the residues present at the interface were solved [31]. Despite high variability of this domain

between organisms, half of the identified contacts for the DnaA-DnaD interaction are conserved within *C. difficile*, the S22 (S23 in *B. subtilis* DnaA), T25 (T26), F48 (F49), D51 (D52) and L68 (L69) (Fig.1) [31,32]. This might suggest a similar interaction surface for CD3653 on *C. difficile* DnaA. Characterization of the putative interaction between CD3653 and DnaA, and the resulting effect on DnaA oligomerization and origin melting awaits purification and functional characterization of CD3653.

The Soj protein, also involved in chromosome segregation, has been shown to interact with DnaA via domain III, regulating DnaA-filament formation [87] and the *C. difficile* encodes at least one uncharacterized Soj homolog, but a role in DNA replication has not been experimentally demonstrated.

Bacterial histone-like proteins (such as HU and HBsu) can modulate DNA topology and might therefore influence *oriC* unwinding and replication initiation. However, the importance of HU for replication initiation has only been demonstrated for *E.coli* [58,88]. Several studies have shown HU independent origin unwinding even in gram-negative bacteria [57,89-91], suggesting that HU-dependence of origin unwinding may be limited to a narrow phylogenetic group. *C. difficile* encodes a homologue of HU, HupA [92] but whether this protein plays a role in DNA replication initiation remains to be established.

Finally, Spo0A, the master regulator of sporulation, binds to several Spo0A-boxes present in this the *oriC* region in *B. subtilis* [93]. Some of the Spo0A-boxes partially overlap with DnaA-boxes and binding of Spo0A can prevent the DnaA-mediated unwinding, thus playing a significant role in the coordination of between cell replication and sporulation [93]. In *C. difficile,* Spo0A-binding has previously been investigated [94], but a role in DNA replication has not been assessed.

For all the regulators with a *C. difficile* homolog discussed above (i.e. CD3653, Soj, HupA and Spo0A), further studies can be envisioned employing the P1 nuclease assays described here to assess the effects on DnaA-mediated unwinding of the origin. Our experiments show, however, they are not strictly required for origin unwinding (Fig. 4).

In summary, through a combination of different *in silico* predictions and *in vitro* studies, we have shown the DnaA-dependent unwinding in the *dnaA-dnaN* intergenic region in the bipartite *C. difficile* origin of replication. We have analysed the putative origin of replication in different clostridia and a conserved organization is observed throughout the Firmicutes, although different mechanisms and regulation could be behind the initiation of replication. The present study is the first to characterize the origin region of *C. difficile* and form the start

to further unravel the mechanism behind the DnaA-dependent regulation of *C. difficile* initiation of replication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

AMOP and WKS designed experiments. AMOP and CW performed the *in silico* analyses. AMOP, EVE and AF performed experiments. AMOP and WKS analysed data and wrote the manuscript. All authors read and approved the final version for submission.

Funding

Acknowledgements

Supplemental Information

Pattern search:
NTATCCACA
TNATCCACA
TTANCCACA
TTATCNACA
TTATCCNCA
TTATCCANA
TTATCCACN
TTNTCCACA
TGTGGATAN
TGTGGATNA
TGTGGNTAA
TGTNGATAA
TGNGGATAA
TNTGGATAA
NGTGGATAA
TGTGGANAA
TTWTNCACA
TGTGNAWAA
NTWTNCACA
TGTGNAWAN
TNWTNCACA
TGTGNAWNA
TTWNNCACA
TGTGNNWAA
TTWTNNACA
TGTNNAWAA
TTWTNCNCA
TGNGNAWAA
TTWTNCANA
TNTGNAWAA
TTWTNCACN
NGTGNAWAA

Supplementary Figures
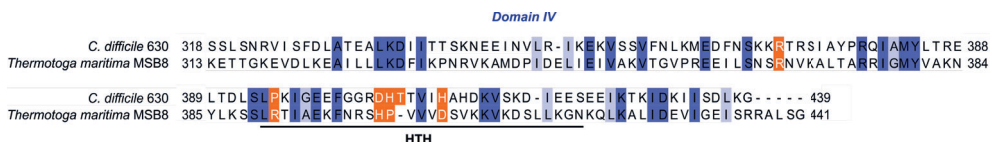
Fig. S1 - **Alignment of domain IV of the *C. difficile* and *Thermotoga maritima* DnaA protein.** Residues are coloured according to sequence identity conservation using blue shading (dark blue more conserved), as analysed in JalView, as for Figure 1. Residues involved in specific contacts with the 9-mer DnaA box sequence are indicated in orange. It is clear that the majority of these residues are not conserved between the two species, except *C. difficile* R370/*T. maritima* R366.
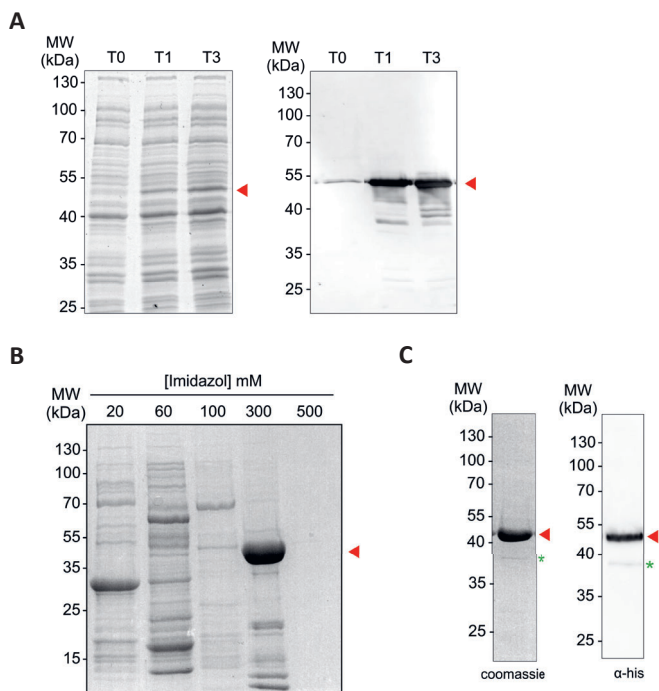

Fig. S2 -. Expression and purification of *C. difficile* DnaA-6xHis protein. A) *E. coli* expressing DnaA-6xHis cells were induced with 1 mM IPTG. Optical density-normalized samples before induction (T0), after 1 hour of induction (T1) and 3 hours of induction (T3) were resolved by 12% SDS-PAGE and immunoblotted with anti-his antibody. Induced DnaA is observed with the approximate molecular weight of 51 kDa (red arrow). Possible breakdown product is observed (blue arrow). **B)**. Samples of DnaA-6xHis HisTrap purification from the elution fraction 2 at binding buffer with different imidazole concentrations (20, 60, 100, 300 and 500 mM) were separated by 12% SDS– PAGE and stained with Coomassie brilliant blue. DnaA-6xHis is observed with an approximate molecular weight of 51 kDa (red arrow) and eluted in Binding buffer supplemented with >300 mM imidazole. **C)** Confirmation of size-exclusion fraction containing the *C. difficile* DnaA-6xHis and further used for analysis after protein purification resolved by

12% SDS-PAGE (Coomassie staining) and immunoblotted with anti-his antibody. DnaA-6xHis is observed with the approximate molecular weight of ~51 kDa (red arrow). Possible minor breakdown products are observed (green asterisk).

A

B
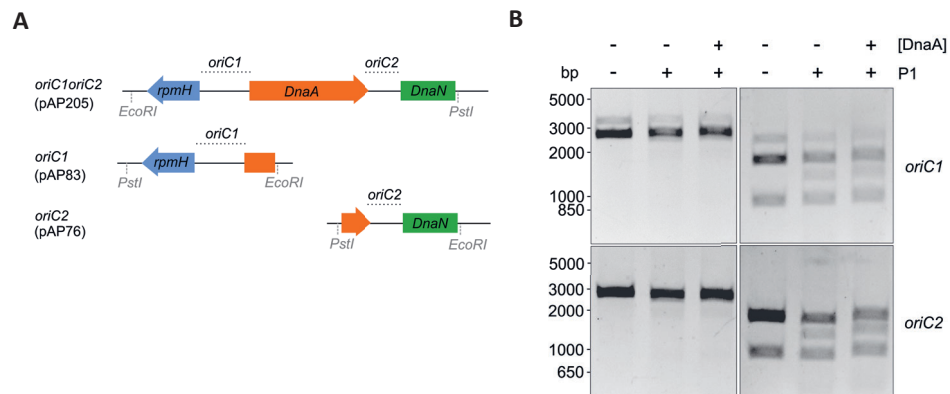
Fig. S3- P1 nuclease assay of the individual *C. difficile oriC* regions. A) Representation of the *oriC* regions present in the used vectors for P1 nuclease assay, *oriC1oriC2* (pAP205), *oriC1-* (pAP83) and *oriC2* (pAP76)-containing vectors. The predicted *oriC* regions (dotted lines) and included genes are represented, *rpmH* (blue), *dnaA* (orange), and *dnaN* (green). **B)** P1 nuclease assay of pAP83 (*oriC1*, upper panel) and pAP76 (*oriC2*, lower panel). Digestion of the vector with the restriction enzymes *BglII* (left panel) or *NotI* (right panel). Digestion of the vectors with the restriction enzymes (lanes 1-3). Treatment of the fragments with P1 nuclease only (lane 2) and incubated with 0.14 µM of *C. difficile* DnaA-6xHis protein (lane 3). Higher DnaA-6xHis were tested with the same profile (data not shown). The DNA fragments were separated in a 1% agarose gel and analyzed with ethidium bromide staining. Spontaneous unwinding is observed and no DnaA-dependent unwinding is detected.

A

B
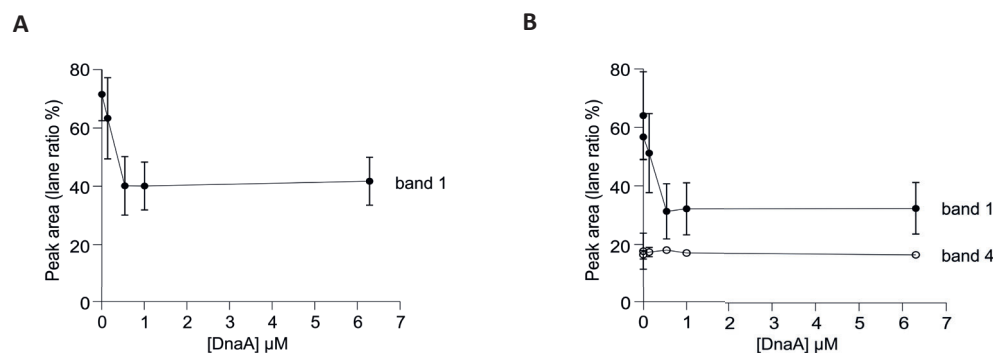
Fig. S4 - Quantification of the P1-independent bands. Data presented here are complementary to that of Figure 4C and 4D in the main body of the manuscript. Quantification was performed using ImageJ, and signals were normalized to the total signal in a lane. **A)** Results for P1/*BglII* digested vector. Shown is the signal (black circles) for the upper band of the gel (Figure 4B, upper panel). **B)** Results for the P1/*NotI* digested vector. Shown is the quantification of the signal of the upper (black circles) and lower (open circles) bands of the gel (Figure 4B, lower panel). Error bars indicate the standard deviation of the mean of n=3 independent experiments.
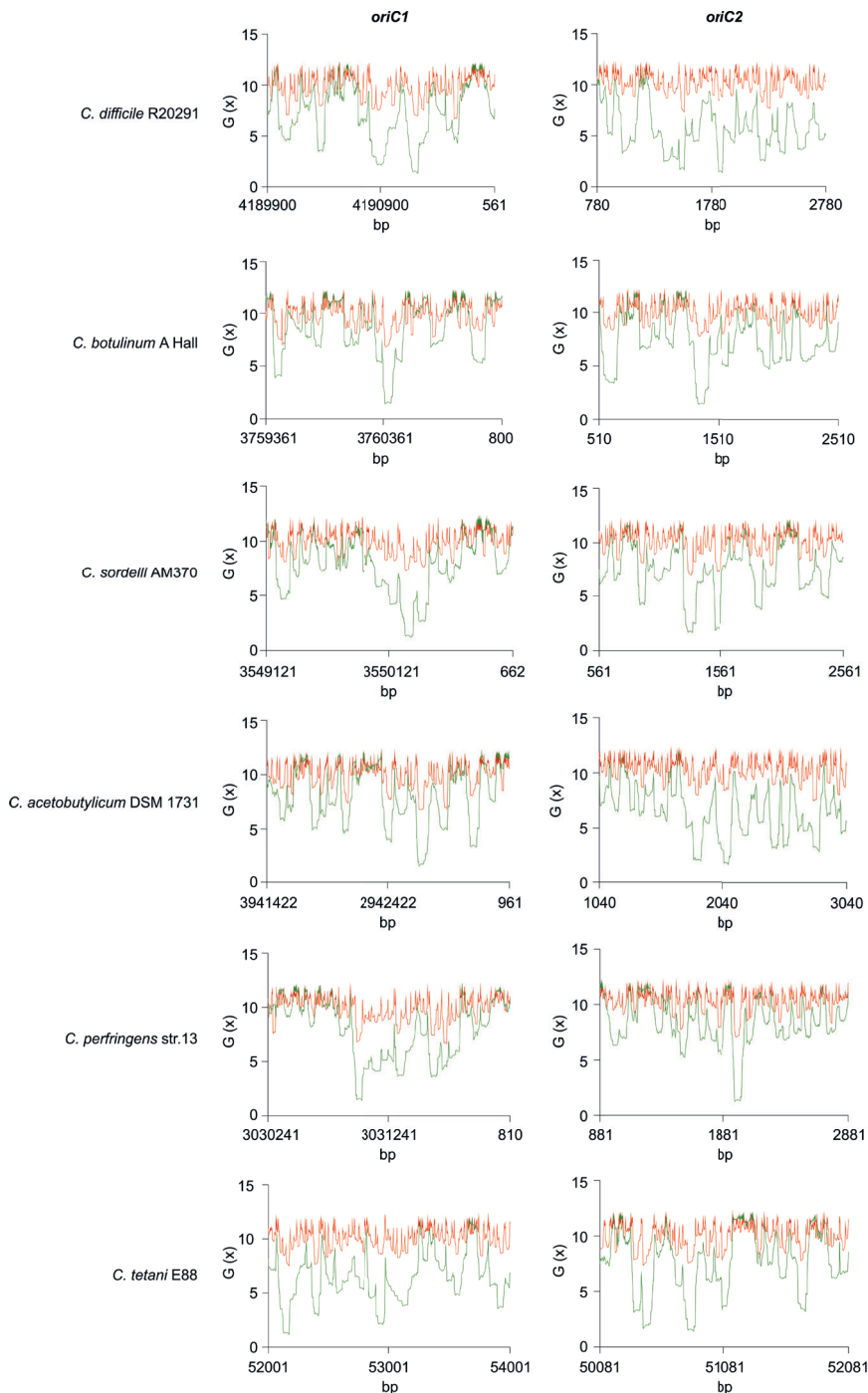
Fig. S5 - SIDD analysis different clostridia. Analyis of 2.0 kb fragments comprising *oriC1* and *oriC2* in *C. difficile* R20291, *C. botulinum A* Hall, *C. sordellii* AM370, *C. acetobutylicum* DSM 1731, *C. perfringens*

str.13, *C. tetani* E88 (see Table 1 in the main body of the manuscript). Nucleotide positioning is indicated. Predicted free energies G(x) for duplex destabilization at a superhelical density of σ = -0.06 (green) or σ = -0.04 (red).
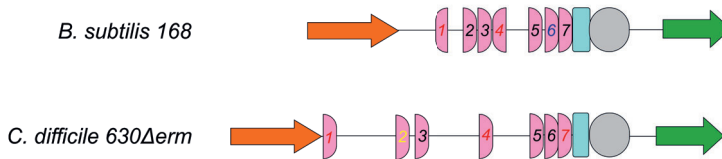
oriC2

Fig. S6 - **Comparison of the *B. subtilis* and *C. difficile oriC2*.** Representation of the *oriC2* region (the intergenic region between *dnaA* and *dnaN*) of *B. subtilis* and *C. difficile* chromosome. The *dnaA* and *dnaN* genes are represented by orange and green arrows, respectively. The DUE is represented by a grey circle. DnaA-trio sequences are shown in light blue boxes. DnaA-boxes are indicated by pink boxes and orientation on the leading (right) and lagging strand (left) are shown. DnaA boxes are numbered according to the *B. subtilis* nomenclature (Richardson, 2019), with numbers in blue (no mismatch from the TTATCCACA sequence, red (1 mismatch), black (2 mismatches) or yellow (3 mismatches). See Material and Methods for detailed information. Alignment of the represented chromosomal regions is based on the location of the DnaA-trio.

References

1 Lawson, P. A., Citron, D. M., Tyrrell, K. L. & Finegold, S. M. Reclassification of *Clostridium difficile* as Clostridioides difficile (Hall and O'Toole 1935) Prevot 1938. *Anaerobe* **40**, 95-99 (2016).

2 Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. *Clostridium difficile* infection. *Nature Reviews Disease Primers* **2**, 16020 (2016).

3 Warriner, K., Xu, C., Habash, M., Sultan, S. & Weese, S. J. Dissemination of *Clostridium difficile* in food and the environment: Significant sources of *C. difficile* community-acquired infection? *J Appl Microbiol* **122**, 542-553 (2017).

4 van Eijk, E., Wittekoek, B., Kuijper, E. J. & Smits, W. K. DNA replication proteins as potential targets for antimicrobials in drug-resistant bacterial pathogens. *J Antimicrob Chemother* **72**, 1275-1284 (2017).

5 Crobach, M. J. T. *et al.* Understanding *Clostridium difficile* Colonization. *Clinical microbiology reviews* **31** (2018).

6 O'Donnell, M., Langston, L. & Stillman, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harbor perspectives in biology* **5** (2013).

7 Bleichert, F., Botchan, M. R. & Berger, J. M. Mechanisms for initiating cellular DNA replication. *Science* **355** (2017).

8 Katayama, T., Ozaki, S., Keyamura, K. & Fujimitsu, K. Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. *Nat Rev Microbiol* **8**, 163-170 (2010).

9 Murray, H. & Koh, A. Multiple regulatory systems coordinate DNA replication with cell growth in *Bacillus subtilis*. *PLoS Genet* **10**, e1004731 (2014).

10 Chodavarapu, S. & Kaguni, J. M. Replication Initiation in Bacteria. *Enzymes* **39**, 1-30 (2016).

11 Jameson, K. H. & Wilkinson, A. J. Control of Initiation of DNA Replication in *Bacillus subtilis* and *Escherichia coli*. *Genes (Basel)* **8** (2017).

12 Schenk, K. *et al.* Rapid turnover of DnaA at replication origin regions contributes to initiation control of DNA replication. *PLoS Genet* **13**, e1006561 (2017).

13 Fossum, S. *et al.* A robust screen for novel antibiotics: specific knockout of the initiator of bacterial DNA replication. *FEMS microbiology letters* **281**, 210-214 (2008).

14 Grimwade, J. E. & Leonard, A. C. Targeting the Bacterial Orisome in the Search for New Antibiotics. *Front Microbiol* **8**, 2352 (2017).

15 Torti, A. *et al. Clostridium difficile* DNA polymerase IIIC: basis for activity of antibacterial compounds. *Current Enzyme Inhibition* **7** (2011).

16 Briggs, G. S., Smits, W. K. & Soultanas, P. Chromosomal replication initiation machinery of low-G+C-content Firmicutes. *J Bacteriol* **194**, 5162-5170 (2012).

17 van Eijk, E. *et al.* Primase is required for helicase activity and helicase alters the specificity of primase in the enteropathogen *Clostridium difficile. Open Biol* **6** (2016).

18 van Eijk, E. *et al.* Genome Location Dictates the Transcriptional Response to PolC Inhibition in *Clostridium difficile*. *Antimicrob Agents Chemother* **63** (2019).

19 Xu, W. C., Silverman, M. H., Yu, X. Y., Wright, G. & Brown, N. Discovery and development of DNA polymerase IIIC inhibitors to treat Gram-positive infections. *Bioorganic & medicinal chemistry* **27**, 3209-3217 (2019).

20 Davey, M. J. & O'Donnell, M. Replicative helicase loaders: ring breakers and ring makers. *Current Biology* **13**, R594-R596 (2003).

21 Bazin, A., Cherrier, M. V., Gutsche, I., Timmins, J. & Terradot, L. Structure and primase-mediated activation of a bacterial dodecameric replicative helicase. *Nucleic acids research* (2015).

22 Majka, J., Messer, W., Schrempf, H. & Zakrzewska-Czerwinska, J. Purification and characterization of the *Streptomyces lividans* initiator protein DnaA. Vol. 179 (1997).

23 Zawilak, A., Durrant, M. C., Jakimowicz, P., Backert, S. & Zakrzewska-Czerwinska, J. DNA binding specificity of the replication initiator protein, DnaA from *Helicobacter pylori*. *Journal of molecular biology* **334**, 933-947 (2003).
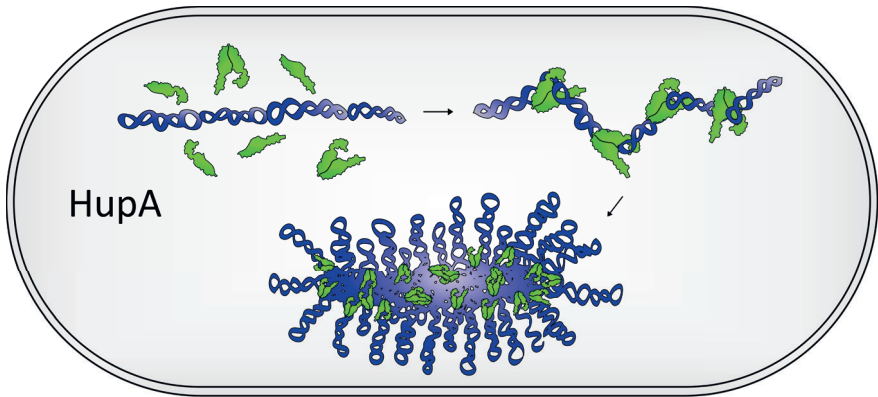
24 Erzberger, J. P., Mott, M. L. & Berger, J. M. Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat Struct Mol Biol* **13**, 676-683 (2006).

25 Zawilak-Pawlik, A., Nowaczyk, M. & Zakrzewska-Czerwinska, J. The Role of the N-Terminal Domains of Bacterial Initiator DnaA in the Assembly and Regulation of the Bacterial Replication Initiation Complex. *Genes (Basel)* **8** (2017).

26 Weigel, C. *et al.* The N-terminus promotes oligomerization of the *Escherichia coli* initiator protein DnaA. *Molecular microbiology* **34**, 53-66 (1999).

27 Abe, Y. *et al.* Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on oriC. *J Biol Chem* **282**, 17816-17827 (2007).

28 Natrajan, G., Noirot-Gros, M. F., Zawilak-Pawlik, A., Kapp, U. & Terradot, L. The structure of a DnaA/HobA complex from *Helicobacter pylori* provides insight into regulation of DNA replication in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 21115-21120 (2009).

29 Jameson, K. H. *et al.* Structure and interactions of the *Bacillus subtilis* sporulation inhibitor of DNA replication, SirA, with domain I of DnaA. *Molecular microbiology* **93**, 975-991 (2014).

30 Kim, J. S. *et al.* Dynamic assembly of Hda and the sliding clamp in the regulation of replication licensing. *Nucleic acids research* **45**, 3888-3905 (2017).

31 Martin, E. *et al.* DNA replication initiation in *Bacillus subtilis*: structural and functional characterization of the essential DnaA-DnaD interaction. *Nucleic acids research* (2018).

32 Matthews, L. A. & Simmons, L. A. Cryptic protein interactions regulate DNA replication initiation. *Molecular microbiology* **111**, 118-130 (2019).

33 Nowaczyk-Cieszewska, M. *et al.* The role of *Helicobacter pylori* DnaA domain I in orisome assembly on a bipartite origin of chromosome replication. *Molecular microbiology* (2019).

34 Erzberger, J. P., Pirruccello, M. M. & Berger, J. M. The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *EMBO J* **21**, 4763-4773 (2002).

35 Nozaki, S. & Ogawa, T. Determination of the minimum domain II size of Escherichia coli DnaA protein essential for cell viability. *Microbiology* **154**, 3379-3384 (2008).

36 Kawakami, H., Keyamura, K. & Katayama, T. Formation of an ATP-DnaA-specific initiation complex requires DnaA Arginine 285, a conserved motif in the AAA+ protein family. *J Biol Chem* **280**, 27420-27430 (2005).

37 Cho, E., Ogasawara, N. & Ishikawa, S. The functional analysis of YabA, which interacts with DnaA and regulates initiation of chromosome replication in *Bacillus subtils*. *Genes Genet Syst* **83**, 111-125 (2008).

38 Ozaki, S. *et al.* A common mechanism for the ATP-DnaA-dependent formation of open complexes at the replication origin. *J Biol Chem* **283**, 8351-8362 (2008).

39 Ozaki, S. & Katayama, T. Highly organized DnaA-oriC complexes recruit the single-stranded DNA for replication initiation. *Nucleic acids research* **40**, 1648-1665 (2012).

40 Saxena, R., Fingland, N., Patil, D., Sharma, A. K. & Crooke, E. Crosstalk between DnaA protein, the initiator of *Escherichia coli* chromosomal replication, and acidic phospholipids present in bacterial membranes. *Int J Mol Sci* **14**, 8517-8537 (2013).

41 Blaesing, F., Weigel, C., Welzeck, M. & Messer, W. Analysis of the DNA-binding domain of *Escherichia coli* DnaA protein. *Molecular microbiology* **36**, 557-569 (2000).

42 Fujikawa, N. *et al.* Structural basis of replication origin recognition by the DnaA protein. *Nucleic acids research* **31**, 2077-2086 (2003).

43 Schaper, S. & Messer, W. Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J Biol Chem* **270**, 17622-17626 (1995).

44 Wolanski, M., Donczew, R., Zawilak-Pawlik, A. & Zakrzewska-Czerwinska, J. oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front Microbiol* **5**, 735 (2014).

45 Speck, C., Weigel, C. & Messer, W. ATP- and ADP-dnaA protein, a molecular switch in gene regulation. *EMBO J* **18**, 6169-6176 (1999).

46   Patel, M. J., Bhatia, L., Yilmaz, G., Biswas-Fiss, E. E. & Biswas, S. B. Multiple conformational states of DnaA protein regulate its interaction with DnaA boxes in the initiation of DNA replication. *Biochim Biophys Acta* (2017).

47   Ozaki, S., Noguchi, Y., Hayashi, Y., Miyazaki, E. & Katayama, T. Differentiation of the DnaA-oriC subcomplex for DNA unwinding in a replication initiation complex. *J Biol Chem* **287**, 37458-37471 (2012).

48   Scholefield, G. & Murray, H. YabA and DnaD inhibit helix assembly of the DNA replication initiation protein DnaA. *Molecular microbiology* **90**, 147-159 (2013).

49   Kowalski, D. & Eddy, M. J. The DNA unwinding element: a novel, cis-acting component that facilitates opening of the *Escherichia coli* replication origin. *EMBO J* **8**, 4335-4344 (1989).

50   Zorman, S., Seitz, H., Sclavi, B. & Strick, T. R. Topological characterization of the DnaA-oriC complex using single-molecule nanomanipuation. *Nucleic acids research* **40**, 7375-7383 (2012).

51   Richardson, T. T., Harran, O. & Murray, H. The bacterial DnaA-trio replication origin element specifies single-stranded DNA initiator binding. *Nature* **534**, 412-416 (2016).

52   Richardson, T. T. *et al.* Identification of a basal system for unwinding a bacterial chromosome origin. *EMBO J* **38**, e101649 (2019).

53   Zawilak-Pawlik, A. *et al.* Architecture of bacterial replication initiation complexes: orisomes from four unrelated bacteria. *Biochem J* **389**, 471-481 (2005).

54   Ekundayo, B. & Bleichert, F. Origins of DNA replication. *PLoS Genet* **15**, e1008320 (2019).

55   Ogasawara, N. & Yoshikawa, H. Genes and their organization in the replication origin region of the bacterial chromosome. *Molecular microbiology* **6**, 629-634 (1992).

56   Moriya, S., Fukuoka, T., Ogasawara, N. & Yoshikawa, H. Regulation of initiation of the chromosomal replication by DnaA-boxes in the origin region of the *Bacillus subtilis* chromosome. *EMBO J* **7**, 2911-2917 (1988).

57   Donczew, R., Weigel, C., Lurz, R., Zakrzewska-Czerwinska, J. & Zawilak-Pawlik, A. *Helicobacter pylori* oriC--the first bipartite origin of chromosome replication in Gram-negative bacteria. *Nucleic acids research* **40**, 9647-9660 (2012).

58   Krause, M., Ruckert, B., Lurz, R. & Messer, W. Complexes at the replication origin of *Bacillus subtilis* with homologous and heterologous DnaA protein. *Journal of molecular biology* **274**, 365-380 (1997).

59   Bawono, P. & Heringa, J. PRALINE: a versatile multiple sequence alignment toolkit. *Methods in molecular biology* **1079**, 245-262 (2014).

60   Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858 (2015).

61   Mackiewicz, P., Zakrzewska-Czerwinska, J., Zawilak, A., Dudek, M. R. & Cebrat, S. Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic acids research* **32**, 3781-3791 (2004).

62   Luo, H. & Gao, F. DoriC 10.0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic acids research* **47**, D74-D77 (2019).

63   Zhabinskaya, D., Madden, S. & Benham, C. J. SIST: stress-induced structural transitions in superhelical DNA. *Bioinformatics* **31**, 421-422 (2015).

64   Mrazek, J. & Xie, S. Pattern locator: a new tool for finding local sequence patterns in genomic DNA sequences. *Bioinformatics* **22**, 3099-3100 (2006).

65   Vellanoweth, R. L. & Rabinowitz, J. C. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo. *Molecular microbiology* **6**, 1105-1114 (1992).

66   Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning : a laboratory manual*. (Cold Spring Harbor Laboratory, 1989).

67   Smits, W. K., Merrikh, H., Bonilla, C. Y. & Grossman, A. D. Primosomal proteins DnaD and DnaB are recruited to chromosomal regions bound by DnaA in *Bacillus subtilis*. *J Bacteriol* **193**, 640-648 (2011).

68 Tsodikov, O. V. & Biswas, T. Structural and thermodynamic signatures of DNA recognition by *Mycobacterium tuberculosis* DnaA. *Journal of molecular biology* **410**, 461-476 (2011).

69 Ozaki, S., Fujimitsu, K., Kurumizaka, H. & Katayama, T. The DnaA homolog of the hyperthermophilic eubacterium *Thermotoga maritima* forms an open complex with a minimal 149-bp origin region in an ATP-dependent manner. *Genes to cells : devoted to molecular & cellular mechanisms* **11**, 425-438 (2006).

70 Sutton, M. D. & Kaguni, J. M. Novel alleles of the *Escherichia coli* dnaA gene. *Journal of molecular biology* **271**, 693-703 (1997).

71 van Eijk, E. *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**, 31 (2015).

72 Necsulea, A. & Lobry, J. R. A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol Biol Evol* **24**, 2169-2179 (2007).

73 Ogasawara, N., Moriya, S., von Meyenburg, K., Hansen, F. G. & Yoshikawa, H. Conservation of genes and their organization in the chromosomal replication origin region of *Bacillus subtilis* and *Escherichia coli*. *EMBO J* **4**, 3345-3350 (1985).

74 Katayama, T., Kasho, K. & Kawakami, H. The DnaA Cycle in *Escherichia coli*: Activation, Function and Inactivation of the Initiator Protein. *Front Microbiol* **8**, 2496 (2017).

75 Sekimizu, K., Bramhill, D. & Kornberg, A. Sequential early stages in the in vitro initiation of replication at the origin of the *Escherichia coli* chromosome. *J Biol Chem* **263**, 7124-7130 (1988).

76 Jaworski, P. *et al.* Unique and Universal Features of Epsilonproteobacterial Origins of Chromosome Replication and DnaA-DnaA Box Interactions. *Front Microbiol* **7**, 1555 (2016).

77 Duderstadt, K. E., Chuang, K. & Berger, J. M. DNA stretching by bacterial initiators promotes replication origin opening. *Nature* **478**, 209-213 (2011).

78 Goranov, A. I., Breier, A. M., Merrikh, H. & Grossman, A. D. YabA of *Bacillus subtilis* controls DnaA-mediated replication initiation but not the transcriptional response to replication stress. *Molecular microbiology* **74**, 454-466 (2009).

79 Seid, C. A., Smith, J. L. & Grossman, A. D. Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in *Bacillus subtilis*. *Molecular microbiology* **103**, 798-817 (2017).

80 Frandi, A. & Collier, J. HdaB: a novel and conserved DnaA-related protein that targets the RIDA process to stimulate replication initiation. *Nucleic acids research* **48**, 2412-2423 (2020).

81 Marsin, S., McGovern, S., Ehrlich, S. D., Bruand, C. & Polard, P. Early steps of *Bacillus subtilis* primosome assembly. *J Biol Chem* **276**, 45818-45825 (2001).

82 Velten, M. *et al.* A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Molecular cell* **11**, 1009-1020 (2003).

83 Smits, W. K., Goranov, A. I. & Grossman, A. D. Ordered association of helicase loader proteins with the *Bacillus subtilis* origin of replication in vivo. *Molecular microbiology* **75**, 452-461 (2010).

84 Ishigo-Oka, D., Ogasawara, N. & Moriya, S. DnaD protein of *Bacillus subtilis* interacts with DnaA, the initiator protein of replication. *J Bacteriol* **183**, 2148-2150 (2001).

85 Zhang, W. *et al.* The *Bacillus subtilis* DnaD and DnaB proteins exhibit different DNA remodelling activities. *Journal of molecular biology* **351**, 66-75 (2005).

86 Rokop, M. E., Auchtung, J. M. & Grossman, A. D. Control of DNA replication initiation by recruitment of an essential initiation protein to the membrane of *Bacillus subtilis*. *Mol Microbiol* **52**, 1757-1767 (2004).

87 Scholefield, G., Errington, J. & Murray, H. Soj/ParA stalls DNA replication by inhibiting helix formation of the initiator protein DnaA. *EMBO J* **31**, 1542-1555 (2012).

88 Chodavarapu, S., Felczak, M. M., Yaniv, J. R. & Kaguni, J. M. *Escherichia coli* DnaA interacts with HU in initiation at the E. coli replication origin. *Molecular microbiology* **67**, 781-792 (2008).

89 Jaworski, P. *et al.* Structure and Function of the Campylobacter jejuni Chromosome Replication Origin. *Front Microbiol* **9**, 1533 (2018).

90   Plachetka, M. *et al. Streptomycete* origin of chromosomal replication with two putative unwinding elements. *Microbiology* **165**, 1365-1375 (2019).

91   Makowski, L., Donczew, R., Weigel, C., Zawilak-Pawlik, A. & Zakrzewska-Czerwinska, J. Initiation of Chromosomal Replication in Predatory Bacterium *Bdellovibrio bacteriovorus*. *Front Microbiol* **7**, 1898 (2016).

92   Oliveira Paiva, A. M. *et al.* The Bacterial Chromatin Protein HupA Can Remodel DNA and Associates with the Nucleoid in *Clostridium difficile*. *Journal of molecular biology* **431**, 653-672 (2019).

93   Boonstra, M. *et al.* Spo0A regulates chromosome copy number during sporulation by directly binding to the origin of replication in *Bacillus subtilis*. *Molecular microbiology* **87**, 925-938 (2013).

94   Rosenbusch, K. E., Bakker, D., Kuijper, E. J. & Smits, W. K. *C. difficile* 630D*erm* Spo0A Regulates Sporulation, but Does Not Contribute to Toxin Production, by Direct High-Affinity Binding to Target DNA. *PloS one* (2012).

HupA

Clostridioides difficile