

# SCIENTIFIC REPORTS



OPEN

## Generation of a fully erythromycin-sensitive strain of *Clostridioides difficile* using a novel CRISPR-Cas9 genome editing system

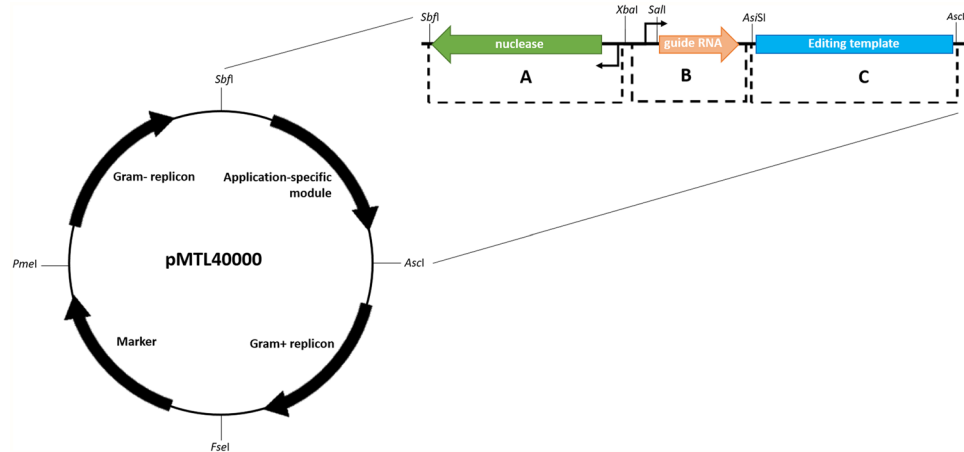
Patrick Ingle<sup>1</sup>, Daphne Groothuis<sup>1</sup>, Peter Rowe<sup>1,4</sup>, He Huang<sup>2</sup>, Alan Cockayne<sup>1</sup>, Sarah A. Kuehne<sup>1,3</sup>, Weihong Jiang<sup>2</sup>, Yang Gu<sup>2</sup>, Christopher M. Humphreys<sup>1</sup> & Nigel P. Minton<sup>1</sup>

Understanding the molecular pathogenesis of *Clostridioides difficile* has relied on the use of *ermB*-based mutagens in erythromycin-sensitive strains. However, the repeated subcultures required to isolate sensitive variants can lead to the acquisition of ancillary mutations that affect phenotype, including virulence. CRISPR-Cas9 allows the direct selection of mutants, reducing the number of subcultures and thereby minimising the likelihood of acquiring additional mutations. Accordingly, CRISPR-Cas9 was used to sequentially remove from the *C. difficile* 630 reference strain (NCTC 13307) two *ermB* genes and *pyrE*. The genomes of the strains generated (630 $\Delta$ *erm*\* and 630 $\Delta$ *erm*\* $\Delta$ *pyrE*, respectively) contained no ancillary mutations compared to the NCTC 13307 parental strain, making these strains the preferred option where erythromycin-sensitive 630 strains are required. Intriguingly, the *cas9* gene of the plasmid used contained a proximal frameshift mutation. Despite this, the frequency of mutant isolation was high (96% and 89% for *ermB* and *pyrE*, respectively) indicating that a functional Cas9 is still being produced. Re-initiation of translation from an internal AUG start codon would produce a foreshortened protein lacking a RuvCI nucleolytic domain, effectively a 'nickase'. The mutation allowed *cas9* to be cloned downstream of the strong P<sub>tht</sub> promoter. It may find application elsewhere where the use of strong, constitutive promoters is preferred.

*Clostridioides difficile* (formerly *Clostridium difficile*<sup>1</sup>) is a Gram-positive, anaerobic, endospore-forming bacterium, which is the leading cause of antibiotic-associated diarrhoea worldwide. *C. difficile* associated infection (CDI) is characterised by a range of disease symptoms ranging from mild, self-limiting diarrhoea to the potentially fatal pseudomembranous colitis (PMC) and places a significant burden on healthcare facilities. The principle virulence factors are the two large, glucosylating toxins A and B<sup>2</sup>. Efforts to identify and characterise those other factors important in CDI have centred on genomic studies and the application of mutagenic gene tools<sup>3</sup>.

The first *C. difficile* genome sequence determined was that of the erythromycin (Em) resistant (<sup>R</sup>), outbreak strain, 630<sup>3,4</sup>. Its genome contained two sequence identical *erm(B)* genes, termed *erm1(B)* and *erm2(B)*, which formed part of the mobilizable, non-conjugative transposon Tn5398<sup>5</sup>. To enable the use of Em<sup>R</sup> genes as selectable markers in available gene tools, sensitive (<sup>S</sup>) variants of strain 630 were required. These were independently isolated by two laboratories using serial passage in antibiotic-free media. Mullany and co-workers (UCL, London, UK) performed 30 such subcultures to isolate 630 $\Delta$ *erm*<sup>6</sup>, whilst the Rood laboratory (Monash, Australia) performed an undisclosed number of subcultures to generate 630E<sup>7</sup>. Both 630 $\Delta$ *erm* and 630E possess an identical 2.4 kb deletion in Tn5398 that entirely removed *erm2(B)* together with the gene encoding CD630\_2008, a predicted plasmid-partitioning protein.

<sup>1</sup>Clostridia Research Group, BBSRC/EPSC Synthetic Biology Research Centre (SBRC), School of Life Sciences, The University of Nottingham, Nottingham, NG7 2RD, UK. <sup>2</sup>Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, 200032, China. <sup>3</sup>Present address: School of Dentistry, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B5 7SA, UK. <sup>4</sup>Present address: Deep Branch Biotechnology, NG8 1AA, Nottingham, United Kingdom. Correspondence and requests for materials should be addressed to N.P.M. (email: [nigel.minton@nottingham.ac.uk](mailto:nigel.minton@nottingham.ac.uk))



**Figure 1.** Modular vector design of pMTL40000, showing three CRISPR-specific sub-modules within the application specific module of the pMTL80000 vector series. Module A consists of a nuclease and promoter (reverse orientation) and is flanked by *SbfI* and *XbaI*. Module B consists of the RNA component(s) of the nuclease in module A, along with a promoter. This is flanked by *XbaI* and *AsiSI*, and contains an internal *SalI* site between the promoter and guide RNA. Module C consists of the homology cassette derived from the left- and right- homology arms of the gene of interest, and is flanked by *AscI* and *AsiSI* restriction sites.

The availability of 630 $\Delta$ *erm* and 630E allowed the use of ClosTron and allelic exchange tools which capitalised on the use of *ermB* genes to generate insertional<sup>8</sup> or clean deletion mutants<sup>9</sup>. However, it became apparent that the phenotypic behaviour of 630 toxin mutants, for instance, was influenced by which Em<sup>S</sup> strain was employed as the progenitor<sup>2,10</sup>. It subsequently transpired that during their derivation, multiple single nucleotide polymorphisms (SNPs) and Indels (Insertion/deletions) had accumulated within their genomes compared to the parental 630 strain<sup>11</sup>. Furthermore, phenotypic comparisons of 630 $\Delta$ *erm* and 630E revealed the latter grew more slowly than the former and to a lower final cell density, produced lower toxin titres and was less virulent in the hamster model of *C. difficile* infection<sup>11</sup>.

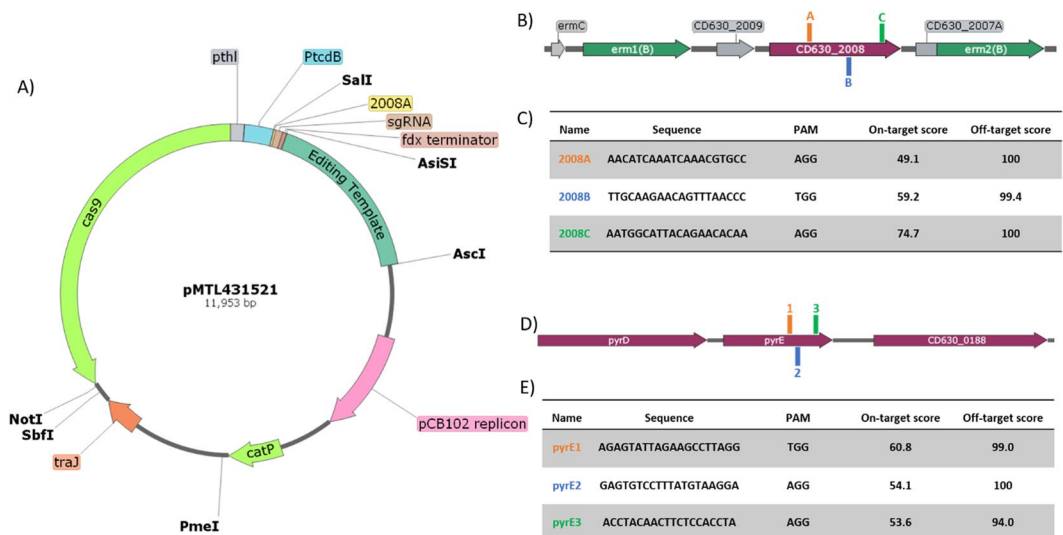
It is likely that the ancillary mutants that arose in 630 $\Delta$ *erm* and 630E were a consequence of the repeated subcultures needed to isolate a spontaneous Em<sup>S</sup> strain. This could be avoided by direct deletion of *ermB* from the 630 genome. Whilst this can be relatively easily achieved through the use of allelic exchange in combination with counter selection markers<sup>12</sup>, the desired mutants may not be directly selected. Rather, a sequential process is used in which single crossover integrants are first selected on media supplemented with an appropriate antibiotic. Following their purification, these cells are then plated on selective media containing the necessary counter selective agent to identify the desired double crossover mutant<sup>9</sup>. One consequence of the necessary repeated passage through single cell, clonal selection steps is the increased probability of isolating SNPs and Indels. A more rapid route to mutant generation would be preferable.

The recent exemplification of CRISPR-Cas9 mutagenesis in clostridia<sup>13–15</sup>, including *C. difficile*<sup>16,17</sup>, offers the facility to directly select double crossover mutants in the absence of traditional counter selection markers. The associated reduction in the number of steps needed to isolate mutants should minimise the risk of isolating variants in which secondary mutations have arisen. To test this assumption we sought to implement our previously developed CRISPR-Cas9 system<sup>14</sup> in *C. difficile* and recreate a strain 630 variant sensitive to Em<sup>R</sup>. Our design took the opportunity to delete both copies of *ermB*, thereby reducing the risk of reversion to Em<sup>R</sup>. During the course of this work, we isolated a CRISPR-Cas9 variant (trCas9) which, in the configuration used, facilitated its deployment in mutant generation.

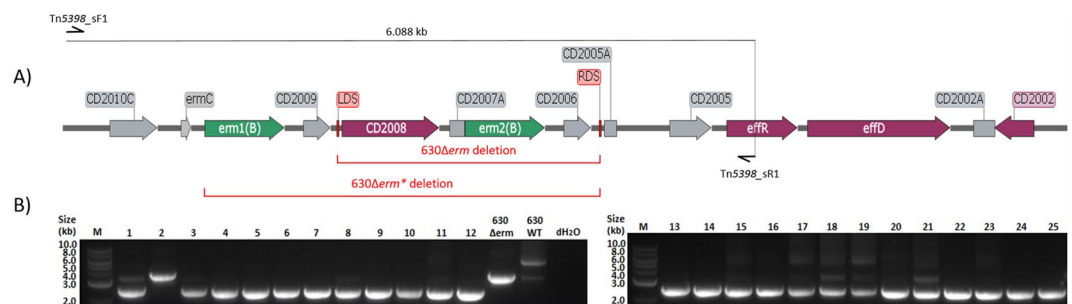
## Results

**Modularisation of a CRISPR-Cas9 vector.** We have previously constructed a CRISPR-Cas9 vector for use in a number of industrially relevant clostridia, including *Clostridium ljungdahlii*<sup>14</sup>. As *C. difficile* is also an acetogen<sup>18</sup>, we based our vector on this system. To build in utility we first modularised the vector to conform to the standardised, pMTL80000 modular vector system in which the application module (between the *SbfI* and *AscI* restriction enzyme sites) became the components required for CRISPR-Cas9 genome editing. The CRISPR-specific components were divided into the following sub-modules; (A) nuclease, (B) guide RNA, and (C) editing template. Submodules A and B are flanked with unique 6–8 nucleotide restriction sites; *SbfI* and *XbaI* for submodule A, and *XbaI*, and *AsiSI* for submodule B. The latter also contains a *SalI* site between the guide RNA promoter and the seed region of the guide RNA, enabling high throughput assembly methods to rapidly exchange guide targeting loci. Due to submodule C containing relatively long (>1500 bp) sequences, the 8 nucleotide *AsiSI* and *AscI* restriction sites were selected to reduce the likelihood of unwanted occurrence of recognition sequences within the homology arms that comprise the editing template.

Initially the vector pMTLcas-ptA previously shown to generate a *pyrE* knockout in *C. ljungdahlii*<sup>14</sup> was adapted to the modular format<sup>19</sup>. Early iterations of the vector were designed to incorporate an alternative restriction site, *MauBI*, between submodules B and C, however, this was later substituted to *AsiSI* as commercially available *MauBI* proved relatively inefficient in DNA cleavage (METHODS). The final pMTL40000 vector design can



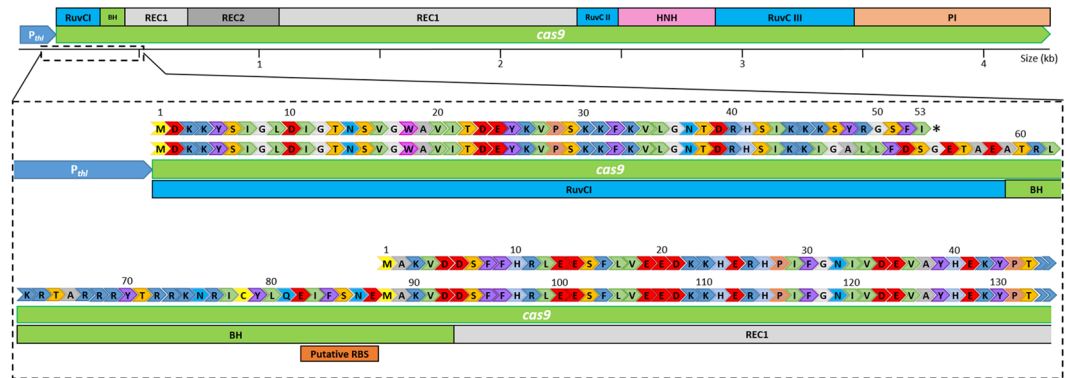
**Figure 2.** CRISPR-Cas9 genome editing vector and retargeting crRNA sequences. **(A)** Diagrammatic representation of the CRISPR-Cas9 vector developed in this study. The requisite CRISPR-Cas9 components, namely *S. pyogenes* *cas9*, under the control of the *C. acetobutylicum* thiolase promoter ( $P_{thl}$ ); an sgRNA component, consisting of a gRNA handle and crRNA (2008A) under the control of *C. difficile* toxin B promoter ( $P_{tcdB}$ ), and an editing template, containing upstream and downstream chromosomal regions flanking the deletion target site, cloned between *AsiI* and *AscI* restriction sites. The plasmid backbone consists of the *E. coli*-clostridia shuttle vector pMTL83151. **(B)** Positions of 20-nt crRNA retargeting sequences within the region targeted for deletion from Tn5398 labelled A–C, corresponding to the 2008A–C sequences listed in **(C)**, along with on- and off-target scores provided by the Benchling CRISPR guide design tool. The protospacer-associated motif (PAM) for each crRNA sequence is also stated. Similarly, locations **(D)**; not to scale) of the three crRNA retargeting sequences **(E)** present in the three CRISPR-Cas9 vectors utilised for truncation of 234 bp at the 3'-end of *pyrE*.



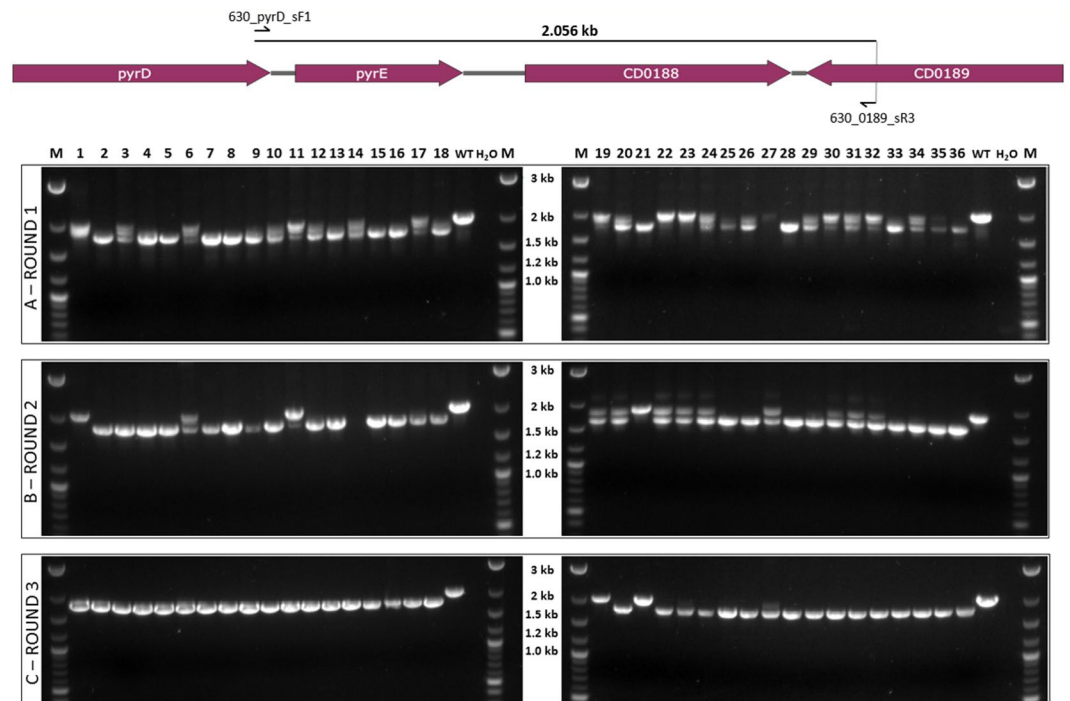
**Figure 3.** Colony PCR screening of putative 630 $\Delta$ *erm*\* strains. **(A)** Diagrammatic representation of the 9.6 kb *C. difficile* 630 Tn5398 mobilizable transposon sequence. The *erm* (**B**) genes are shown in green, other named ORFs shown in purple, and gene remnants and pseudogenes are shown in grey. Red brackets denote the genomic region deleted in 630 $\Delta$ *erm*, flanked by the left- (LDS) and right-deletion sites (RDS), and targeted for deletion in the generation of 630 $\Delta$ *erm*\*. Schematic binding sites of screening PCR primers Tn5398\_sF1 and Tn5398\_sR1 (not to scale) and the size of the expected wild-type product are also shown. **(B)** Colony PCR screening of 25 putative 630 $\Delta$ *erm*\* mutants using Tn5398\_sF1 and Tn5398\_sR1 primers. 'M' denotes DNA marker, 'dH<sub>2</sub>O' denotes negative control lane.

be seen in Fig. 1. For pMTL431511-CLAU-pyrE, the pCB102 Gram-positive replicon was used, along with the CatP selective marker, and ColE1 + *tra* Gram-negative replicon module. The CRISPR-specific application module comprises **(A)**  $P_{thl}$  controlling the expression of *cas9* RNA **(B)** the  $P_{araE}$  promoter controlling the expression of sgRNA containing the *pyrE* targeting seed region, and **(C)** the homology cassette consists of 1 kb regions immediately up- and down-stream of the *pyrE* gene.

**Implementation of the modular CRISPR vector in *C. difficile*.** Following modularisation of the CRISPR vector the plasmid was adapted for use in *C. difficile*. The nuclease selected for submodule A remained the *cas9* gene from *Streptococcus pyogenes* under the control of the strong thiolase promoter from *Clostridium acetobutylicum* as it is known to function effectively in *C. difficile*<sup>8,19</sup>. For submodule B, the sgRNA cassette was



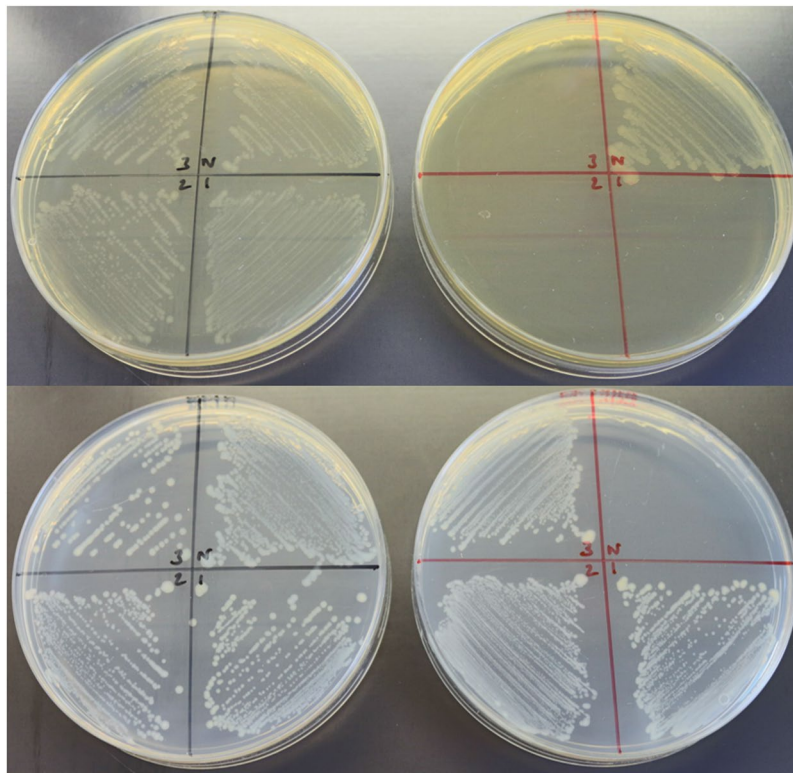
**Figure 4.** Diagrammatic representation of trCas9 sequence and annotated features. Top: Domain organisation within wild-type *S. pyogenes* Cas9. Bottom: Domain organisation and DNA sequence of our trCas9 and upstream thiolase promoter ( $P_{th}$ ) from pMTL431511-CLAU-pyrE. The frameshift-generating adenine base insertion at position 130 within the cas9 RuvCI domain, which results in a truncated polypeptide of 53 amino acids, is indicated in capital letters. Putative ribosome binding site and downstream trCas9 ORF lacking 87 amino acids from the N-terminal end of Cas9 also indicated.



**Figure 5.** Colony PCR screening of putative  $630\Delta erm^* \Delta pyrE$  strains. Top: Diagrammatic representation of the *C. difficile* 630 chromosomal region containing *pyrE*, including schematic binding sites of screening primers 630\_pyrD\_sF1 and 630\_0189\_sR3 (not to scale). Bottom: Agarose gel electrophoresis following colony PCR screens for the desired 234 bp truncation of *pyrE* following initial streaking of thiamphenicol resistant transconjugant colonies to purity (A) and two subsequent colony PCR screens following an additional one (B), or two (C), rounds of passaging on BHIS agar supplemented with thiamphenicol. 'M' denotes DNA marker, 'dH<sub>2</sub>O' denotes negative control lane.

placed under the control of the native *C. difficile* Toxin B promoter  $P_{tcdB}$ , and an editing template for homologous recombination comprising approximately 1 kb up- and down-stream genomic regions flanking the desired 3.6 kb deletion region within Tn5398 (Fig. 2A) made up submodule C. To thoroughly test our CRISPR-Cas9 genome editing plasmid, and since guide RNA prediction methods are in their infancy, we generated three CRISPR-Cas9 vectors targeting the CD630\_2008 gene in Tn5398, with each containing a different 20 nucleotide crRNA sequence (Fig. 2B,C) from those identified using the Benchling CRISPR guide design tool<sup>20</sup>. These vectors, pMTL431521-CDF-2008A, pMTL431521-CDF-2008B and pMTL431521-CDF-2008C, were conjugated into our *C. difficile* 630 strain (CRG856) and the resulting thiamphenicol-resistant transconjugants which appeared



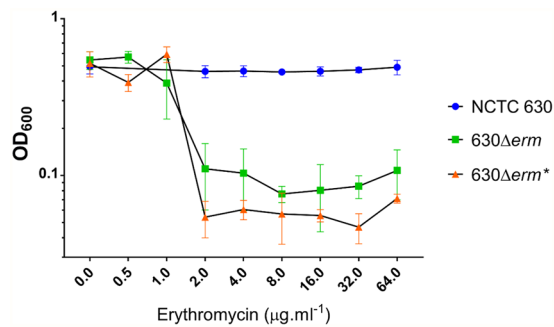


**Figure 6.** Phenotypic characterisation of NCTC 630 and  $630\Delta erm^* \Delta pyrE$  strains. Growth of *C. difficile* strains NCTC 630 (N) and three independently generated  $630\Delta erm^* \Delta pyrE$  mutants (#1–3) after 48 hours on BHIS (top left), BHIS supplemented with  $10 \mu\text{g}\cdot\text{ml}^{-1}$  Em (top right), *C. difficile* minimal media (CDMM) supplemented with uracil (bottom left) and CDMM supplemented with uracil and 5-fluoroorotic acid (bottom right).

after 48–72 hours were screened via colony PCR (Fig. 3). Of the 25 colonies screened, all but one generated a 2.44 kb-sized product signifying the deletion of both *erm(B)* genes from Tn5398 with an editing efficiency of 96%. Eleven of these 25 PCR screens also showed bands at 6.088 kb and/or 3.682 kb indicating the presence of wild-type 630 or  $630\Delta erm$  Tn5398 sequence, respectively, whilst thirteen appeared as pure mutants. From these thirteen pure mutants, three independently generated strains, one from each CRISPR-Cas9 vector, were carried forward for plasmid loss, with the subsequent thiamphenicol-sensitive strains designated as  $630\Delta erm^*$ .

**The *cas9* gene of pMTL431521 contains a frame shift.** In parallel to the generation of the *erm(B)* deletion variants, the entire nucleotide sequence of the plasmids used were determined using appropriate primers and Sanger sequencing. Unexpectedly, a single adenine base insertion within a poly-A region starting at nucleotide position 130 was discovered which resulted in a frameshift mutation. This same single base insertion was also found (Fig. S1) in the parental plasmid, pMTLcas-pta<sup>14</sup>. Closer inspection of the region of sequence affected identified an AUG start codon some 135 nucleotides 3' to the frameshift that was preceded by a sequencing bearing some resemblance to a RBS which would allow a truncated Cas9 (trCas9) protein lacking 87 amino acids from the N-terminal end of Cas9 to be produced (Fig. 4). The deleted Cas9 domain encompasses a RuvCI nucleolytic domain. The trCas9 produced would, therefore, most likely represent a 'nickase' variant of Cas9 which produces single-strand nicks instead of DSBs. Cas9 nickase variants have been previously shown to be highly effective for genome editing in clostridia<sup>13</sup>.

**Generation of  $630\Delta erm^* \Delta pyrE$ .** Having successfully demonstrated that our CRISPR-Cas9 system was effective, despite the premature termination of the Cas9 polypeptide, we sought to further validate this system via the truncation of *pyrE* in each of our independent  $630\Delta erm^*$  strains. An editing template encompassing approximately 1 kb up- and down-stream of the 234 bp to be deleted from the 3'-end of the *pyrE* gene was constructed via SOE PCR and cloned into pMTL431521-CDF-2008A. HiFi assembly was again used to insert the three chosen crRNA sequences (Fig. 2D,E) from the 16 identified within the 234 bp target region, generating vectors pMTL431521-CDF-pyrE1, pMTL431521-CDF-pyrE2 and pMTL431521-CDF-pyrE3. These vectors were transferred into each of the independent  $630\Delta erm^*$  strains and 12 resulting thiamphenicol-resistant colonies from each conjugation were colony PCR screened for the desired deletion (Fig. 4). Whilst 32/36 colonies (89%) contained the 1.822 kb band indicative of this deletion, over half of these also contained a 2.056 kb band consistent with the wild-type *pyrE* sequence. Previous CRISPR-Cas9 studies have reported that the isolation of pure mutants following such mixed phenotypes from PCR screens can be improved via additional streaking of the mixed



**Figure 7.** Determination of the minimum inhibitory concentration to Em of *C. difficile* 630 strains. Growth after 24 hours of *C. difficile* 630 strains in BHIS broth supplemented with various amounts of Em was determined via optical density measurements at 600 nm (OD<sub>600</sub>). Symbols represent mean values from three independent experiments and error bars indicate the standard deviation from the mean.

strains on selective media<sup>14</sup>. Therefore, we streaked all 36 colonies onto the same BHIS media supplemented with thiamphenicol twice more, performing colony PCR screens after each round of streaking (Fig. 5). After the first round of screening, 13/36 colonies were pure mutant with a further 19/36 displaying a mixed wild-type and mutant phenotype. The number of pure mutants obtained increased to 21 after one additional streak, and to 32 after a third round of screening, confirming the observations by Huang *et al.*<sup>14</sup>. Three independently generated pure *pyrE*-truncated mutants were sub-cultured to lose the plasmid and named CRG 630Δ*erm*\*Δ*pyrE* #1–3.

**Whole genome sequencing of CRG 630Δ*erm*\* strains.** We next performed Illumina whole genome sequencing of our independent CRG 630Δ*erm*\* mutants, along with the parental CRG 630 strain, to assess whether any genomic variants had accumulated in these strains during CRISPR-Cas9 mutagenesis. Illumina reads from each strain were aligned to the 630 reference sequence AM180355.1 (Table S1). Whilst we were able to show that no additional mutations relative to the parental strain were present, we detected 17 variants present in each of our CRG 630 strains (parent and mutants) compared to the published reference genome. These variants consisted of 13 single nucleotide polymorphisms, three point insertions and a 21-nucleotide deletion from the CD630\_26850 ORF resulting in an in-frame deletion of seven amino acids from the mature protein. Searches for conserved protein domains within the amino acid sequence of CD630\_26850 identified C-terminal SpoIIE and serine phosphatase domains. This 21-nt internal deletion within CD630\_26850 was confirmed via PCR and Sanger sequencing with flanking primers in our CRG 630 strain and its progeny, but was not observed in our 630Δ*erm* strain.

To investigate the effects of this 21-nt deletion, or indeed any of the other variants, on the ability of our CRG 630 strain and its progeny to form heat-resistant spores over five days we performed a rate of sporulation assay (Fig. S2). Development of heat-resistant CFU for CRG 630 and its progeny was lower, yet comparable, with that of 630Δ*erm* and this difference was not statistically significant. Furthermore, there was no significant difference between our CRG 630 strain and the 630Δ*erm*\* derivatives indicating the deletion of the 3.6 kb genomic region containing both *erm(B)* genes had no effect upon the proficiency of these strains to develop heat-resistant spores.

**Regeneration of 630Δ*erm*\* and 630Δ*erm*\*Δ*pyrE*.** From these whole genome sequencing results it was clear that our aim of generating a new Em<sup>s</sup> *C. difficile* 630 strain free from genomic variants had not been achieved due to an accumulation of ancillary mutants, and in particular the 21-bp deletion. Hence, we sought to repeat our CRISPR-Cas9 mutagenesis in a newly acquired reference strain of *C. difficile* 630 from the NCTC culture collection (NCTC 13307; ‘NCTC 630’). Accordingly, vectors pMTL431521-CDF-2008A, pMTL431521-CDF-2008B and pMTL431521-CDF-2008C were conjugated into NCTC 630 and, as before, three independent 630Δ*erm*\* strains lacking both *erm(B)* genes were obtained, one from each conjugation. Similarly, for the regeneration of 630Δ*erm*\*Δ*pyrE*, vectors pMTL431521-CDF-pyrE1, pMTL431521-CDF-pyrE2 and pMTL431521-CDF-pyrE3 were conjugated into the three 630Δ*erm*\* replicates and the desired truncation of *pyrE* observed in each case (Fig. S3).

**Phenotypic characterisation of NCTC 630Δ*erm*\* and 630Δ*erm*\*Δ*pyrE* strains.** Having regenerated 630Δ*erm*\* and 630Δ*erm*\*Δ*pyrE* mutants in the NCTC 630 strain, we next confirmed the expected phenotypes of these strains by assessing growth after plating on appropriate media (Fig. 6). All three independent 630Δ*erm*\* mutants were unable to grow on BHIS media supplemented with 10 µg.ml<sup>-1</sup> Em. Furthermore, only the three 630Δ*erm*\*Δ*pyrE* strains were able to grow on CDMM supplemented with uracil and 5-FOA, as in the parental NCTC 630 strain the product of the *pyrE* gene converts 5-FOA into the toxic compound 5-FUMP. To determine whether the deletion of both *ermB* genes from *C. difficile* 630 is sufficient to prevent the reversion to Em<sup>R</sup> exhibited by 630Δ*erm*, we conducted serial passages of the parental 630 strain, 630Δ*erm*\* mutants and 630Δ*erm* in antibiotic-free, liquid media. After thirty such passages, neither 630Δ*erm* nor 630Δ*erm*\* exhibited a reversion to Em<sup>R</sup> which would allow these strains to grow on BHIS media supplemented with 10 µg.ml<sup>-1</sup> Em. We then determined the minimum inhibitory concentration of these strains to Em using the broth microdilution method

(Fig. 7). The MIC of both 630 $\Delta erm$  and 630 $\Delta erm^*$  strains was determined to be 2  $\mu\text{g}\cdot\text{ml}^{-1}$ , whilst the NCTC 630 strain was resistant up to 64  $\mu\text{g}\cdot\text{ml}^{-1}$ , the highest Em concentration tested.

**Whole genome sequencing of NCTC 630 $\Delta erm^*$  and 630 $\Delta erm^* \Delta pyrE$  strains.** We then performed Illumina sequencing on our new NCTC 630 strain as well as each triplicate of the 630 $\Delta erm^*$  and 630 $\Delta erm^* \Delta pyrE$  progeny. All genomic variants identified in these strains relative to the CP010905.2 reference sequence are outlined in Table S2 in Supplementary Information. The NCTC 630 strain was found to contain three genomic variants relative to the CP010905.2 reference sequence, consisting of a single adenosine base insertion in a non-coding region downstream of *argS*, and two non-synonymous SNPs in the *rpoB* and *perR* genes. Unsurprisingly, these variants were present in all of the subsequent mutants. Of the three triplicate 630 $\Delta erm^* \Delta pyrE$  strains, one (replicate #3) had accumulated no SNPs nor indels after the two rounds of CRISPR-Cas9 mutagenesis. Interestingly, in both rounds of mutagenesis this strain required the fewest passages to confirm the loss of the CRISPR plasmid. Meanwhile, replicates #1 and #2, which required up to four additional passages to confirm loss of the CRISPR, were found to have accumulated one and two SNPs, respectively. 630 $\Delta erm^* \Delta pyrE$  replicate #1 contained an additional non-synonymous SNP in *perR*, which encodes a peroxide-responsive repressor, whilst 630 $\Delta erm^* \Delta pyrE$  replicate #2 contained two additional non-synonymous SNPs; one in CDIF630\_00864 encoding a putative lipoprotein and another introducing a premature stop codon in *eftA3*, which encodes an electron transfer flavoprotein alpha subunit. Each of these variants occurred during the first round of CRISPR-Cas9 mutagenesis to generate the 630 $\Delta erm^*$  strains. Thus, no variants were accumulated in any of our triplicates during the second round of CRISPR-Cas9 genome editing to obtain the 630 $\Delta erm^* \Delta pyrE$  mutants.

We then aligned our NCTC 630 sequence data against other 630 and 630 $\Delta erm$  reference sequences available from the NCBI database to see how this NCTC sequence compares to other 630 strains (Table S3). The NCTC 630 strain sequenced in this study is most similar to that sequenced by Riedel *et al.*<sup>21</sup>, (CP010905.2) who obtained their strain from the DSM culture collection (DSM 27543). As this DSM strain was acquired from the NCTC, the three variants we observe in our NCTC 630 strain most likely arose in the intervening time between transfer from the NCTC and sequencing of the DSM strain. Meanwhile, we observed 16 conflicts (13 SNPs, 3 point insertions) when comparing our NCTC 630 sequence with that of the 630 sequence from Sebahia *et al.*<sup>3</sup>, only five of which are also present in both 630 $\Delta erm$  sequences. This suggests that the first 630 strain sequenced (AM180355.1) had diverged from a progenitor strain from which 630 $\Delta erm$  was generated, as we would expect all variants called in AM180355.1 to be present in the 630 $\Delta erm$  sequences.

## Discussion

In this study, we used CRISPR-Cas9 genome editing to generate 630 $\Delta erm^*$ , a fully Em<sup>S</sup> variant of a *C. difficile* reference strain 630 (NCTC 13307), by removing both *ermB* genes present on the mobilizable transposon Tn5398. The CRISPR-Cas9 system developed in this study was then further exemplified via the generation of 630 $\Delta erm^* \Delta pyrE$ , a truncated *pyrE* deletion mutant compatible with the pre-existing allele-coupled exchange methodology. Crucially, replicates of each mutant were obtained lacking any SNPs or indels which can arise during mutagenesis. This genome editing, performed in triplicate for each deletion target using three separate plasmids with unique crRNA retargeting sequences, was observed to be highly efficient in each instance. Observed editing efficiencies in this study were 96% for 630 $\Delta erm^*$  and 89% for 630 $\Delta erm^* \Delta pyrE$ .

The efficiency of mutant generation observed here represents a higher frequency than obtained using the aTet-inducible system of McAllister and coworkers<sup>16</sup> who described mutant generation in *pyrE* and *selD* at frequencies of 50 and 20%, respectively. A similar system based on a lactose inducible promoter was used to 'knock-out' *spoA* (100%) and 'knock-in' a fluorescent *PpFbFPm* reporter gene<sup>17</sup>. Mutant colonies of the latter could not initially be detected and required that the primary transconjugants were subcultured overnight in liquid media and re-plated to single colonies before the presence of mutants could be demonstrated. In the experiment described, 80% were mutants. Interestingly, in both cases<sup>16,17</sup> mutants were obtained in the absence of inducer indicating that neither promoter system (aTet or lactose) was sufficiently repressed under non-inducing conditions to prevent Cas9-mediated genome cleavage.

Another recent study regenerated 630 $\Delta erm$  by deleting both *ermB* genes, as was done in this study, utilising the Cpf1 endonuclease instead of Cas9<sup>22</sup>. However, this CRISPR-Cpf1 mediated *ermB* deletion was performed in a lab strain of 630 and no whole genome sequencing of the subsequent mutants was reported. Hence, our 630 $\Delta erm^*$  strain is preferable for use in future studies as we have confirmed a lack of ancillary mutations relative to the NCTC 630 reference strain.

Our developed CRISPR-Cas9 system was notably characterised by the presence in the *cas9* coding sequence of a nucleotide insertion that resulted in the premature termination of the encoded product. Despite this, our system proved highly effective, suggesting that a functional Cas9 nuclease was still being produced. The most likely explanation for this phenomenon was that re-initiation of translation occurs at a downstream, internal AUG start codon that is preceded by a sequence reminiscent of a RBS sequence. The protein (trCas9) would, therefore, be foreshortened by some 87 amino acids from the N-terminus of the native Cas9 nuclease. As the deleted region encompasses a RuvCI nucleolytic domain<sup>23</sup> the variant Cas9 produced is most likely a nickase. Whilst nickase has reduced editing efficiencies compared to native Cas9, because their expression is less deleterious to the cell, larger numbers of primary transformants/transconjugant cells are obtained which, upon subsequent passages, result in mutants. As a consequence, the final total number of mutants obtained in other clostridia has been shown to be greater in comparison to the use of native Cas9<sup>13,15</sup>. Subsequent attempts to recreate plasmid pMTL431521 were repeated unsuccessfully, again resulting in the isolation of *cas9* variants in which various rearrangements and deletions had occurred that obviated the isolation of a fully functional *cas9* (data not shown). The most likely

Bacterial Strain	Description	Source
<i>E. coli</i> TOP10	F- <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) Φ80 <i>lacZ</i> Δ <i>M15</i> Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ( <i>ara</i> <i>leu</i> ) 7697 <i>galU</i> <i>galK</i> <i>rpsL</i> (Str <sup>R</sup> ) <i>endA1</i> <i>nupG</i>	Invitrogen
<i>E. coli</i> CA434	Conjugation donor	Purdy <i>et al.</i> <sup>28</sup>
<i>C. difficile</i> 630 (CRG856)	Ribotype-012 wild-type (Zurich, Switzerland)	P Mullany, UCL
<i>C. difficile</i> 630 (NCTC 13307)	Ribotype-012 wild-type, reference strain (Zurich, Switzerland)	National Collection of Type Cultures (NCTC), Public Health England
<i>C. difficile</i> 630Δ <i>erm</i>	Em <sup>S</sup> derivative of <i>C. difficile</i> 630 lacking <i>erm2(B)</i>	Hussein <i>et al.</i> <sup>6</sup>
<i>C. difficile</i> 630Δ <i>erm</i> *	Em <sup>S</sup> <i>C. difficile</i> 630 strain lacking <i>erm1(B)</i> and <i>erm2(B)</i>	This study
<i>C. difficile</i> 630Δ <i>erm</i> *Δ <i>pyrE</i>	<i>C. difficile</i> 630Δ <i>erm</i> * lacking 234 base-pairs from 3'-end of the <i>pyrE</i> gene	This study

**Table 1.** List of bacterial strains utilised in this study.

explanation for the repeated isolation of this variant is that the high expression of *cas9* when placed downstream of the strong  $P_{thi}$  promoter is detrimental to the *E. coli* host resulting in the selection of variants which no longer express a Cas9 nuclease with DSB activity.

The emergence of inadvertent mutations that reduce *cas9* expression underlines the importance of limiting the constitutive expression of these genes, particularly in the *E. coli* donor strain. Recently, a CRISPR vector for use in *C. sporogenes* was described<sup>24</sup> in which *cas9* was placed downstream of the promoter of the *C. sporogenes* ferredoxin gene,  $P_{fdx}$ . Similar to  $P_{thi}$ , this is a strong, constitutive promoter<sup>25</sup>. Closer inspection of the strategy employed reveals that the 5' end of the fragment carrying *cas9* was not cloned at an appropriate position downstream of the RBS of the  $P_{fdx}$  promoter of the vector used, pMTL83151. Rather the region of DNA encoding the 5'-end of *lacZ*α region of pMTL83151 remained. This is predicted to result in the production of a polypeptide comprising some sixteen amino acids from LacZα (MTMITNSSVPGDPLE) followed by a further 6 amino acids (WIRNTQ) derived from the out-of-frame fusion to the 5'-end of *cas9*, after which a TAG stop codon is encountered (Fig. S4). The *cas9* gene is not preceded by an appropriate RBS sequence. It is, therefore, highly likely that this vector exhibits Cas9 activity for the same reason as described here, through re-initiation of translation at a downstream, internal AUG start codon to produce trCas9.

*C. difficile* 630 is resistant to a wide range of antibiotics including Em, tetracycline and β-lactams, limiting the choice of markers which can be used in genetics studies. The previous generation of Em<sup>S</sup> derivatives of 630, 630Δ*erm* and 630E, permitted the use of the *ermB* marker in ClosTron mutagenesis for the identification of successful intron insertion into the target gene. Whilst the development of CRISPR-Cas9 genome editing allows mutagenesis to be performed in wild-type strains, the availability of an Em<sup>S</sup> 630 strain will still find use in studies which require multiple antibiotic resistance markers or where an *ermB* marker is to be used without risk of reversion to an Em<sup>R</sup> phenotype previously observed with 630Δ*erm* at low frequencies<sup>6</sup>. Here we present 630Δ*erm*\*, an Em<sup>S</sup> 630 strain which is sequence identical to the NCTC 13307 parental strain, except for the intended deletion within Tn5398.

Our initial attempts at generating a SNP-free Em<sup>S</sup> strain of 630 were hampered by the presence of mutations in our laboratory stock of 630 (CRG 630) prior to the onset of this work, most notably the 21 bp deletion. Our stock of 630 was acquired from the Brendan Wren laboratory (LSHTM) in 2010, who had previously acquired it from the Mullany laboratory. A fresh stock of 630 was reacquired from LSHTM in 2018, from Lisa Dawson, and the region encompassing the deletion PCR amplified from the culture. No evidence for the presence of the deletion was evident from Sanger sequencing of the amplified DNA fragment. On receipt in 2010, the strain would have been used to prepare glycerol stocks which have been stored from that date at -80 °C until used here. It would appear that the deletion variant arose during the sub-culture undertaken to create that stock.

Since its isolation from a patient with severe pseudomembranous colitis in 1982 by Wust *et al.*<sup>4</sup>, *C. difficile* 630 was supplied to Hachler, who in turn supplied this strain to Mullany, who then supplied it to Wren, from whom we obtained our CRG 630 strain. Mullany also deposited strains 630 and the Em<sup>S</sup> derivative 630Δ*erm* with the NCTC culture collection (NCTC 13307) who then supplied 630 to the DSMZ collection (DSM 27543). Given the variants accumulated in our CRG 630 strain and the differences between published 630 genomes, we reiterate our previous recommendation<sup>11</sup> that bacterial strains within laboratory collections are regularly re-sequenced. Furthermore, with the ever-decreasing cost of whole genome sequencing it is essential that this is performed after the generation of mutant strains to ensure no ancillary genomic variants have been acquired along the way. A similar recommendation was made in a recent review of *C. difficile* 630 reference genome sequences which advocated the resequencing of all laboratory 630 strains to allow for the tracing of 630 evolution, and drew parallels between the dissemination and propagation of 630 strains in laboratories and culture collections across the world to the *E. coli* long-term evolution experiment<sup>26</sup>.

Traditional approaches to mutant generation by allelic exchange in clostridia, has been reliant on counter selection markers in a two-step selection process reliant on the initial isolation of single cross-over integrants followed by the isolation of the desired double cross-over mutants. As a consequence, repeated passages through single colony isolations are required, a process that increases the chances of isolating ancillary genomic mutations. The use of CRISPR-Cas9 allows the direct selection of clonal populations which carry the desired mutant allele, reducing the number of passages and thereby the chances of isolating inadvertent mutants. Both approaches, however, can fall foul of the need to isolate mutants that have lost the plasmid used to create them. In many instances, this can require several passages through single colony re-streaks. As demonstrated here, the more passages that are undertaken, the greater the likelihood of selecting a derivative containing SNPs or Indels. In this respect, the number of passages used to cure the CRISPR-Cas9 vector employed to create the mutant should be wherever possible minimised.



Plasmid	Description	Source
pMTL83151	<i>E. coli</i> -clostridia shuttle vector with pCB102 replicon	Heap <i>et al.</i> <sup>19</sup>
pMTLcas-pta	CRISPR-Cas9 genome editing vector harbouring <i>cas9</i> gene from <i>Streptococcus pyogenes</i>	Huang <i>et al.</i> <sup>14</sup>
pMTL431511-CLAU-pyrE	Modular CRISPR-Cas9 vector targeting <i>C. autoethanogenum pyrE</i>	This study
pMTL431521-CDF-2008A	pMTL431511-CLAU-pyrE with TcdB promoter driving sgRNA expression, 2008A crRNA sequence and Tn5398 editing template	This study
pMTL431521-CDF-2008B	pMTL431521-CDF-2008A with 2008B crRNA sequence	This study
pMTL431521-CDF-2008C	pMTL431521-CDF-2008A with 2008C crRNA sequence	This study
pMTL431521-CDF-pyrE1	pMTL431521-CDF-2008A containing <i>pyrE</i> editing template and <i>pyrE1</i> crRNA sequence	This study
pMTL431521-CDF-pyrE2	pMTL431521-CDF-pyrE1 with <i>pyrE2</i> crRNA sequence	This study
pMTL431521-CDF-pyrE3	pMTL431521-CDF-pyrE1 with <i>pyrE3</i> crRNA sequence	This study

**Table 2.** List of plasmids utilised in this study.

## Methods

**Bacterial strains and plasmids.** All bacterial strains and plasmids utilised in this study are listed in Tables 1 and 2 respectively. *E. coli* was cultured in LB medium, supplemented where appropriate with chloramphenicol (25 µg.ml<sup>-1</sup>), at 37 °C with horizontal shaking at 200 rpm. *C. difficile* was cultured anaerobically at 37 °C in an anaerobic MACS1000 workstation (Don Whately, Yorkshire, UK) in BHIS (Brain Heart Infusion supplemented with yeast extract [5 mg.ml<sup>-1</sup>] and L-cysteine [0.1% w/v]) medium supplemented with d-cycloserine (250 µg.ml<sup>-1</sup>), cefoxitin (8 µg.ml<sup>-1</sup>) and thiamphenicol (15 µg.ml<sup>-1</sup>) or Em (10 µg.ml<sup>-1</sup>) where appropriate. Confirmation of *pyrE* truncation was performed by plating strain onto *C. difficile* minimal medium described previously<sup>27</sup>, supplemented with uracil (20 µg.ml<sup>-1</sup>) and 5-fluoroorotic acid (4 mg.ml<sup>-1</sup>) where appropriate.

**Modularisation of the CRISPR-Cas9 vector.** First, *cas9* and *P<sub>thi</sub>* from pMTLcas-pta were amplified using Pthl-F-XbaI and Cas9R-NotI, with the resulting PCR product digested with *Xba*I and *Not*I and cloned into a pMTL83151 backbone linearised with the same enzymes. Following Sanger sequence confirmation with CatP-R1, *pyrE* targeting sgRNA cassette was generated and inserted described below. Using pMTLcas-pta as a template, the *P<sub>araE</sub>* promoter was amplified with sgRNA-F-XbaI and sgRNA-pyrE-sg3-SOE-R, with the latter primer containing the 20 nt *pyrE* targeting seed region as an overhang. Simultaneously, the sgRNA handle and its associated terminator were amplified using sgRNA-pyrE-sg3-SOE-F and sgRNA-R-MauBI-AscI primers, with the former containing the 20 nt *pyrE* targeting seed region as an overhang. Using the flanking sgRNA-F-XbaI and sgRNA-F-MauBI-AscI primers, these two PCR products were spliced via SOEing PCR using the overlapping seed region. The resulting spliced PCR product was digested with *Xba*I and *Asc*I, and cloned in the Cas9 containing pMTL83151 vector described above (also digested with *Xba*I and *Asc*I). Again, CatP-R1 was used for Sanger sequence confirmation of this intermediate vector, prior to insertion of the *pyrE* homology arm cassette, resulting in the final pMTL43151-CLAU-pyrE vector.

For homology cassette generation, primer pairs CLAU-pyrE-LHA-F-MauBI + CLAU-pyrE-LHA-R, and CLAU-pyrE-RHA-F + CLAU-pyrE-RHA-R-AscI were used to generate the *pyrE* LHA and RHA, respectively. Following this, the flanking primers CLAU-pyrE-LHA-F-MauBI and CLAU-pyrE-RHA-R-AscI were used to splice the homology arms, with the subsequent PCR product digested with *Mau*BI and *Asc*I. The digested homology arm cassette was cloned in a *Mau*BI and *Asc*I linearised backbone of the intermediate Cas9 and sgRNA containing pMTL83151 vector described above, to generate the final pMTL43151-CLAU-pyrE vector. Following the identification of issues regarding *Mau*BI digestion efficiency, the relevant primer pairs were adjusted in order to incorporate the *Asi*SI recognition site in place of the *Mau*BI site, as detailed in Supplementary Table S4. Final vector sequence confirmation was achieved using CatP-R1 and sgRNA-F-XbaI primers.

Having modularised the pMTLcas-pta into the pMTL40000 vector series as outlined in the results section, we next sought to modify this system for use in *C. difficile* via replacement of the *P<sub>araE</sub>* promoter controlling sgRNA expression. Accordingly, the promoter region of *tcdB* was PCR amplified from *C. difficile* 630 genomic DNA using primer pair P<sub>tcdB</sub>\_XbaI and P<sub>tcdB</sub>\_SalI and cloned between *Xba*I and *Sal*I restriction sites within pMTL431511-CLAU-pyrE.

For the deletion of a 3.6 kb fragment containing the *erm1(B)* and *erm2(B)* genes from Tn5398 an editing template for homologous recombination was generated. PCR primer pairs 630erm\_asiSI\_LF1 & 630erm\_LR1, and 630erm\_RF1 & 630erm\_ascI\_RR1, were used to amplify 0.977 kb and 1.025 kb homology arms, respectively, from *C. difficile* 630 genomic DNA. These homology arms were fused together using splicing by overlap extension PCR (SOE-PCR) with primers 630erm\_asiSI\_LF1 and 630erm\_ascI\_RR1 and the resulting editing template cloned into pMTL431511-CLAU-pyrE between *Asi*SI and *Asc*I restriction sites. To retarget the sgRNA submodule of the resulting vector, we identified 20-nucleotide guide retargeting sequences upstream of 5'-NGG-3 PAMs within the 3.6 kb chromosomal deletion fragment using the CRISPR guide design tool on Benchling ([www.benchling.com](http://www.benchling.com))<sup>20</sup>. Three guide sequences were selected with the highest on-target and off-target scores, across both DNA strands and incorporated into oligonucleotides with flanking 25-nucleotide regions of overlapping homology to the 3'-end of *P<sub>tcdB</sub>* and 5'-end of the sgRNA handle. These oligonucleotides were separately incorporated into *Sal*I-linearized vector using HiFi DNA assembly (NEB), the assembly reaction products were transformed into *E. coli* DH5α and the Sanger sequencing confirmed vectors named pMTL431521-CDF-2008A, pMTL431521-CDF-2008B and pMTL431521-CDF-2008C.

Similarly, for generating truncated *pyrE* mutants in *C. difficile* 630 lacking 234 bp from the 3' end of the gene, 1.338 and 1.026 kb homology arms were PCR amplified using primer pairs 630pyrE\_asiSI\_LF1 & 630pyrE\_LR1, and 630pyrE\_RF1 & 630pyrE\_ascI\_RR1, respectively, and fused together using SOE-PCR with primers 630pyrE\_asiSI\_LF1 and 630pyrE\_ascI\_RR1. The resulting editing template was cloned between *AsiSI* and *AscI* sites of pMTL431521-CDF-2008A. Next, this vector was linearized with *Sall* as previously to allow replacement of guide targeting sequences using HiFi assembly with oligonucleotides pyrE1\_HiFi, pyrE2\_HiFi and pyrE3\_HiFi generating the vectors pMTL431521-CDF-pyrE1, pMTL431521-CDF-pyrE2 and pMTL431521-CDF-pyrE3.

**CRISPR-Cas9 mutagenesis.** CRISPR-Cas9 vectors were transferred to *C. difficile* 630 strains via conjugation from *E. coli* CA434 donors as described previously<sup>28</sup>. Thiamphenicol resistant transconjugant colonies appeared within 72 hours and were streaked to purity. Resulting single colonies were used as the template in colony PCR screens for the desired deletions using primer pairs annealing to chromosomal regions flanking the editing template sequences, Tn5398\_sF1 & Tn5398\_sR1 for 630 $\Delta$ *erm*\* and CD630\_pyrD\_sF1 & CD630\_0189\_sR3 for 630 $\Delta$ *erm*\* $\Delta$ *pyrE*. Colony PCR products were separated using agarose gel electrophoresis.

**Illumina Sequencing and bioinformatics.** Genomic DNA was isolated from *C. difficile* overnight cultures using phenol-chloroform extraction and stored in Tris-HCl buffer (10 mM, pH 7.8). *C. difficile* genomic DNA was sequenced using the Illumina MiSeq platform (DeepSeq, Nottingham, UK) using 500 bp V2 SBS chemistry. Analysis of the generated reads and identification of single nucleotide polymorphisms, insertions and deletions was performed using the Basic Variant Detection tool within CLC Genomics Workbench (version 11.0.1) by mapping the trimmed Illumina paired-end reads to the reference sequences AM180355.1<sup>3</sup>, CP010905.2<sup>21</sup>, LN614756<sup>29</sup> or CP016318<sup>30</sup>. Variants were detected with minimum coverage, count and frequency of 10, 2 and 70%, respectively, with base quality filter settings of neighbourhood radius = 5, minimum central quality = 20 and minimum neighbourhood quality = 15, applied.

**Development of heat-resistant CFU over 5 days.** Sporulation assays observing the development of heat resistant CFU over 5 days were performed as described previously<sup>31,32</sup>. Briefly, sporulation cultures, including a 630 $\Delta$ *erm* $\Delta$ *spo0A* sporulation negative control<sup>12</sup>, were grown in triplicate in BHIS broth for 5 days, with two 500  $\mu$ l samples taken after 0, 24, 48, 72, 96 and 120 hours. From each time-point, one sample was heat-treated (65 °C, 30 minutes) whilst the other was incubated on the bench. Following incubations, samples were serially diluted from 10<sup>9</sup> to 10<sup>-7</sup> in PBS and three 20  $\mu$ l aliquots of each dilution were spotted onto BHIS agar supplemented with 0.1% taurocholic acid. After 24 hours, colonies were counted and CFU.ml<sup>-1</sup> values determined.

**MIC determinations.** Determinations of the minimum inhibitory concentration of *C. difficile* strains to Em was performed using the broth microdilution method described previously<sup>33</sup>. Overnight *C. difficile* cultures were sub-cultured into pre-reduced BHIS broth supplemented with Em concentrations ranging from 0 to 64  $\mu$ g.ml<sup>-1</sup>. After 24 hours incubation, OD<sub>600</sub> measurements were performed using a GloMax-Multi Microplate Multimode Reader (Promega, USA) and the MICs determined.

## References

- Lawson, P. A., Citron, D. M., Tyrrell, K. L. & Finegold, S. M. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe* **40**, 95–99 (2016).
- Kuehne, S. A. *et al.* The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* **467**, 711 (2010).
- Sebaihia, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* **38**, 779–786 (2006).
- Wust, J., Sullivan, N. M., Hardegger, U. & Wilkins, T. D. Investigation of an outbreak of antibiotic-associated colitis by various typing methods. Investigation of an Outbreak of Antibiotic-Associated Colitis by Various Typing Methods. *J. Clin. Microbiol.* **16**, 1096–1101 (1982).
- Farrow, K. A., Lyras, D. & Rood, J. I. Genomic analysis of the erythromycin resistance element Tn5398 from *Clostridium difficile*. *Microbiology* **147**, 2717–2728 (2001).
- Hussain, H. A., Roberts, A. P. & Mullany, P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630 $\Delta$ *erm*) and demonstration that the conjugative transposon Tn916 $\Delta$ E enters the genome of this strain at multiple sites. *J. Med. Microbiol.* **54**, 137–141 (2005).
- O'Connor, J. R. *et al.* Construction and analysis of chromosomal *Clostridium difficile* mutants. *Mol. Microbiol.* **61**, 1335–1351 (2006).
- Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P. & Minton, N. P. The ClosTron: A universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods* **70**, 452–464 (2007).
- Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T. & Minton, N. P. Precise Manipulation of the *Clostridium difficile* Chromosome Reveals a Lack of Association between the tcdC Genotype and Toxin Production. **78**, 4683–4690 (2012).
- Lyras, D. *et al.* Toxin B is essential for virulence of *Clostridium difficile*. *Nature* **458**, 1176–9 (2009).
- Collery, M. M. *et al.* What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of *Clostridium difficile* strain 630 and derivatives. *Virulence* **8**, 767–781 (2017).
- Ng, Y. K. *et al.* Expanding the Repertoire of Gene Tools for Precise Manipulation of the *Clostridium difficile* Genome: Allelic Exchange Using *pyrE* Alleles. *PLoS One* **8** (2013).
- Xu, T. *et al.* Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. *Appl. Environ. Microbiol.* **81**, 4423–4431 (2015).
- Huang, H. *et al.* CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. *ACS Synth. Biol.* **5**, 1355–1361 (2016).
- Li, Q. *et al.* CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnol. J.* **11**, 961–972 (2016).
- McAllister, K. N., Bouillaut, L., Kahn, J. N., Self, W. T. & Sorg, J. A. Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Sci. Rep.* **7**, 1–12 (2017).
- Wang, S. *et al.* Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clin. Microbiol. Infect.* **24**, 1095–1099 (2018).

18. Koepke, M., Straub, M. & Duerre, P. *Clostridium difficile* is an autotrophic bacterial pathogen. *PLoS One* **8**, 1–7 (2013).
19. Heap, J. T., Pennington, O. J., Cartman, S. T. & Minton, N. P. A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* **78**, 79–85 (2009).
20. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
21. Riedel, T. *et al.* Genome resequencing of the virulent and multidrug-resistant reference Strain *Clostridium difficile* 630. *Genome Announc.* **3**, e00276–15 (2015).
22. Hong, W., Zhang, J., Cui, G., Wang, L. & Wang, Y. Multiplexed CRISPR-Cpf1-mediated genome editing in *Clostridium difficile* toward the understanding of pathogenesis of *C. difficile* infection. *ACS Synth. Biol.* **7**, 1588–1600 (2018).
23. Nishimasu, H. *et al.* Crystal Structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
24. Guo, C., Allen, B. M., Hiam, K. J., Dodd, D. & Treuren, W. V. Depletion of microbiome-derived molecules in the host using *Clostridium* genetics. *bioRxiv* (2018).
25. Zhang, Y., Grosse-honebrink, A. & Minton, N. P. A universal mariner transposon system for forward genetic studies in the genus *Clostridium*. *PLoS One* **10**, e0122411 (2015).
26. Roberts, A. P. & Smits, W. K. The evolving epidemic of *Clostridium difficile* 630. *Anaerobe* 10–12, <https://doi.org/10.1016/j.anaerobe.2018.04.015> (2018).
27. Cartman, S. T. & Minton, N. P. A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. *Appl. Environ. Microbiol.* **76**, 1103–1109 (2010).
28. Purdy, D. *et al.* Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol.* **46**, 439–452 (2002).
29. van Eijk, E. *et al.* Complete genome sequence of the *Clostridium difficile* laboratory strain 630 $\delta$ erm reveals differences from strain 630, including translocation of the mobile element CTn5. *BMC Genomics* **16** (2015).
30. Dannheim, H. *et al.* Manual curation and reannotation of the genomes of *Clostridium difficile* 630 $\Delta$ erm and *C. difficile* 630. *J. Med. Microbiol.* **66**, 286–293 (2017).
31. Burns, D. A., Heap, J. T. & Minton, N. P. SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. *J. Bacteriol.* **192**, 657–664 (2010).
32. Burns, D. A., Heeg, D., Cartman, S. T. & Minton, N. P. Reconsidering the sporulation characteristics of hypervirulent *Clostridium difficile* BI/NAP1/027. *PLoS One* **6**, 1–7 (2011).
33. Hasselmann, C. Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution. *Clin. Microbiol. Infect.* **9**, 1–7 (2003).

## Acknowledgements

This work was funded by the Swiss National Science Foundation (Sinergia Grant CRSII3\_147603), the Marie Skłodowska-Curie ITN, CLOSPORE [Grant No. 642068] and the Biotechnology and Biological Sciences Research Council [Grant Numbers BB/L502030/1, BB/K00283X/1].

## Author Contributions

P.I., D.G., P.R., C.M.H., W.J. and Y.G. and N.P.M. planned the study. P.I., D.G., H.H. and P.R. performed experimental work and data analysis. P.I. and N.P.M. drafted the manuscript, all authors contributed to the final version.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-44458-y>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019