



HAL
open science

What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of *Clostridium difficile* strain 630 and derivatives.

Mark M Collery, Sarah A Kuehne, Shonna M McBride, Michelle L Kelly, Marc Monot, Alan Cockayne, Bruno Dupuy, Nigel P Minton

► To cite this version:

Mark M Collery, Sarah A Kuehne, Shonna M McBride, Michelle L Kelly, Marc Monot, et al.. What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of *Clostridium difficile* strain 630 and derivatives.. *Virulence*, 2016, 10.1080/21505594.2016.1237333 . pasteur-01370882

HAL Id: pasteur-01370882

<https://pasteur.hal.science/pasteur-01370882>

Submitted on 23 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

1 **What's a SNP between friends: the influence of single nucleotide polymorphisms on**
2 **virulence and phenotypes of *Clostridium difficile* strain 630 and derivatives**

3

4 **Running Title:** What's a SNP between Friends?

5

6 Mark M Collery^{1¶}, Sarah A Kuehne^{1,2¶}, Shonna M McBride³, Michelle L Kelly¹, Marc Monot⁴,

7 Alan Cockayne¹, Bruno Dupuy⁴, and Nigel P Minton^{1,2*}

8

9

10 1 Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC),

11 School of Life Sciences, the University of Nottingham, Nottingham, UK

12 2 NIHR Biomedical Research Unit in Gastrointestinal and Liver Diseases at Nottingham

13 University Hospitals NHS Trust and The University of Nottingham

14 3 Department of Microbiology and Immunology, Emory Antibiotic Resistance Center,

15 Emory University, Atlanta, USA

16 4 Laboratoire Pathogénèse des Bactéries Anaérobies, Institut Pasteur, Paris, France

17

18 ***Corresponding to:** Nigel P Minton; Email: nigel.minton@nottingham.ac.uk

19

20 [¶]These authors contributed equally to the work.

21

22 **Keywords:** single nucleotide polymorphism; sporulation; motility; toxin expression;
23 mutation; Clostron; virulence

24 **Abstract**

25

26 *Clostridium difficile* is a major cause of antibiotic induced diarrhoea worldwide,
27 responsible for significant annual mortalities and represents a considerable economic
28 burden on healthcare systems. The two main *C. difficile* virulence factors are toxins A and
29 B. Isogenic toxin B mutants of two independently isolated erythromycin-sensitive
30 derivatives (630E and 630 Δ erm) of strain 630 were previously shown to exhibit
31 substantively different phenotypes. Compared to 630, strain 630E and its progeny grow
32 slower, achieve lower final cell densities, exhibit a reduced capacity for spore-formation,
33 produce lower levels of toxin and are less virulent in the hamster infection model. By the
34 same measures, strain 630 Δ erm and its derivatives more closely mirror the behaviour of
35 630. Genome sequencing revealed that 630 Δ erm had acquired seven unique Single
36 Nucleotide Polymorphisms (SNPs) compared to 630 and 630E, while 630E had nine SNPs
37 and a DNA inversion not found in the other two strains. The relatively large number of
38 mutations meant that the identification of those responsible for the altered properties of
39 630E was not possible, despite the restoration of three mutations to wildtype by allelic
40 exchange and comparative RNAseq analysis of all three strains. The latter analysis revealed
41 large differences in gene expression between the three strains, explaining in part why no
42 single SNP could restore the phenotypic differences. Our findings suggest that strain
43 630 Δ erm should be favoured over 630E as a surrogate for 630 in genetic-based studies.
44 They also underline the importance of effective strain curation and the need to genome re-
45 sequence master seed banks wherever possible.

46

47 Introduction

48

49 *Clostridium difficile* is a Gram-positive, anaerobic spore-forming bacterium capable of
50 causing a range of diseases from mild diarrhoea to potentially fatal toxic
51 pseudomembranous colitis. The toxigenic effects of *C. difficile* are caused by the activities
52 of two large, glucosylating toxins. The two toxins are 308kDa (toxin A) and 270kDa (toxin B)
53 in size¹⁻³ and are encoded by the chromosomally located genes *tcdA* and *tcdB*,
54 respectively. Both are cytopathic to cultured cells due to disruption of the cytoskeleton,
55 although TcdB is thought to be up to 1000-times more potent.¹ Historically, toxin A was
56 regarded as the main causative agent of the symptoms of *C. difficile* infection (CDI). Pivotal
57 data was provided by Lyerly et al.⁴ who were only able to detect disease when hamsters
58 were subject to intragastric challenge with purified TcdA alone and not with TcdB. The
59 latter could, however, cause disease symptoms if prior damage to the mucosa had been
60 inflicted by co-administration of sub-lethal concentrations of toxin A. Furthermore, co-
61 administration of both toxins led to more severe disease symptoms. To accommodate
62 these data, it was generally accepted that both toxins acted in concert to bring about
63 disease symptoms, with toxin A leading to the initial damage to the colon allowing the
64 subsequent access of the more potent toxin B.

65 During the 1990's *C. difficile* strains were isolated from symptomatic patients that only
66 produced toxin B (A-B+)^{5,6}. These findings suggested that toxin B, at least in certain strains,
67 is capable of causing disease without the help of toxin A. It has been reported since that
68 toxin B, in A-B+ strains, is modified and seems to be an evolutionary hybrid of *C. difficile*
69 toxin B and *Clostridium sordellii* lethal toxin.⁷

70 With the development of genetic systems, assumptions of the relative importance of
71 the two toxins could be tested through the creation, and *in vivo* assay, of isogenic mutants
72 in which production of either toxin had been ablated. Initial findings made by Lyras et al.⁸
73 appeared to turn the perceived view on its head, through the demonstration that a *tcdA*
74 mutant producing TcdB alone (A-B+) was capable of causing disease in the hamster model
75 while a *tcdB* mutant producing only TcdA (A+B-) did not. These data were, however,
76 almost immediately questioned by a second study conducted in the Minton laboratory⁹
77 showing that both *tcdA* and *tcdB* *C. difficile* mutants, and therefore TcdA and TcdB alone,
78 were independently capable of causing disease. Interestingly, a strain has recently¹⁰ been
79 isolated from a clinical case of CDI, that only produces TcdA (A+B-)

80 The possible reasons for the observed difference in outcomes of the two studies have
81 been discussed previously.¹¹ Both studies agree on the virulence potential of toxin B, but
82 uncertainties remain about the different outcomes concerning the effects of toxin A. In the
83 work presented here, we have hence focused on comparisons of the parental strains and
84 the strains only producing toxin A (A+B-). In essence, both sets of mutants were generated
85 by insertional inactivation of the toxin genes of the *C. difficile* strain 630¹² and, once
86 created, were tested in the hamster infection model. However, in order to implement the
87 available gene tools in strain 630 (at the time the only strain for which a genome sequence
88 was available), it was necessary to first isolate a variant that had become sensitive to
89 erythromycin, thereby allowing the use of an *ermB* gene as a selective, genetic marker.
90 Both studies used such an erythromycin-sensitive derivative of strain 630, but they were
91 independently isolated. In our study (Minton group),⁹ we used the strain 630 Δ *erm*,
92 isolated in the Mullany laboratory (UCL, London, UK) after 30 repeated subcultures of

93 strain 630 in non-selective media.¹³ In parallel, the Rood laboratory (Monash, Australia)
94 independently isolated the erythromycin sensitive strain JIR8094 (also referred to as 630E),
95 ¹⁴ through an undisclosed number of subcultures of strain 630 in non-selective media. Both
96 strains are reported to possess the same specific deletion of *ermB*.^{13,14}

97 We have previously hypothesised¹¹ that the different outcomes of the two studies^{8,9}
98 are a direct consequence of the use of the two, independently isolated erythromycin-
99 sensitive strains, 630 Δ *erm* and 630E. We suggested that during repeated subculture,
100 ancillary mutations arose which impacted on the virulence potential of one or other of the
101 two strains in the presence of different toxin gene alleles. In the current piece of work, we
102 have set out to test this hypothesis. We have undertaken side-by-side comparisons of
103 630 Δ *erm* and 630E, and the A+B- mutant derivatives, in a variety of assays to establish
104 phenotypic differences. In parallel, we have determined the genome sequences of the
105 various strains used in the two studies.^{8,9} Then, we have used our newly developed allelic
106 exchange methodologies¹⁵ to correct a number of SNPs in strain 630E back to wild-type
107 and assessed the consequences. Furthermore we have performed RNAseq experiments
108 comparing the transcriptome of 630, 630 Δ *erm* and 630E at three different time points. The
109 RNA data were related to the whole genome data to draw our final conclusions.

110

111 **Results**

112 **Generation of a ClosTron insertion in *tcdB* of 630E**

113 Although considered unlikely, the possibility existed that mutants made by the
114 insertion of a plasmid element carrying *ermB*¹⁴ might behave differently to an equivalent
115 mutant made by the insertion of a group II intron incorporating *ermB*.¹⁶ Our initial step

116 was, therefore, to create a *tcdB* mutant of strain 630E using ClosTron technology.
117 Accordingly, the ClosTron plasmid pMTL007C-E2::Cdi-tcdB-1511a that had previously been
118 used to generate strain 630 Δ *erm* A+B-⁹ was used to create an equivalent mutant in strain
119 630E as described.⁹ The resulting mutant, 630E A+B-CT, was verified by PCR, Sanger
120 sequencing and shown by Southern blot to carry a single group II intron insertion (**Fig.**
121 **S1A**). Parental strains, original A+B- mutants and the newly obtained mutant were tested
122 for production of toxin A in a Western blot (**Fig. S1B**). As expected all strains produced
123 toxin A.

124

125 **Phenotypic characterisation of strains**

126 In order to establish whether all strains were phenotypically identical a range of assays
127 were performed, comparing growth, motility and spore properties. An analysis of growth
128 rates using the procedure described in Materials and Methods showed that strain 630 Δ *erm*
129 and derivatives grew to the highest optical density, closely followed by strain 630 and
130 630E, and derivatives thereof (**Fig. 1A and 1B**). The data clearly demonstrated that strain
131 630 Δ *erm* and its 630 Δ *erm* A+B- derivative had relatively higher growth rates and achieved
132 higher optical densities ($p < 0.0001$, unpaired t-test at 24 h) than strain 630E and its
133 derivatives, with strain 630E A+B-CT growing the least (**Fig. 1B**). It was also apparent,
134 shown by plate motility assay (**Fig. 2**), that strains 630 and 630 Δ *erm* were motile, while
135 630E was not. Only 630 and 630 Δ *erm*, but not 630E, form pseudopod-like structures,
136 which are characteristic for swarming motility in bacteria.

137 Following the protocols of Burns et al,¹⁷ comparative differences in the numbers of
138 colony forming units (CFUs) obtained following heat shock were assessed between the

139 strains, as a crude estimate of spore formation.¹⁸ On this basis, strains 630 Δ *erm* and
140 630 Δ *erm* A+B- produced a greater numbers of spores than 630E and its derivatives, which
141 failed to produce any spores until 72 h. The total number of CFU/mL at this time point was
142 10³ times fewer than that obtained with 630 Δ *erm* or 630 Δ *erm* A+B- (**Fig. 3A**). Interestingly
143 parental strain 630 produced very few spores before 72 h, but spore counts increased from
144 72 h onwards and reached similar levels to strains 630 Δ *erm* and 630 Δ *erm* A+B- by the end
145 of the experiment. The reduction in spore formation may in part be due to the observed
146 reduction in OD as the 630E strains enter stationary phase, which might also explain their
147 predilection to flocculate. Indeed comparing percentage sporulation (relative to vegetative
148 cell count), confirmed the observation that 630 Δ *erm* and 630 Δ *erm* A+B- have a higher
149 sporulation frequency than both 630E (and derivatives) and strain 630.

150 The germination of the 630E strains was also comparatively reduced and did not reach
151 the same level as that of 630 Δ *erm* and its progeny. At the last time point (240 min) strains
152 630E and 630E A+B- reach the same level of CFU/ml as strain 630. The observed delay
153 could be due to the previously observed reduced cell growth of the 630E strains (**Fig. 3B**).

154

155 **Toxin Production**

156 Measurements of the amounts of toxin being produced by 630E and 630 Δ *erm* and
157 their derivatives were undertaken using both the *C. DIFFICILE TOX A/B II*TM ELISA assay kit
158 from TechLab, measuring toxin A and B, and kits from TGCbiomics, specifically measuring
159 either only toxin A or only toxin B. The results of the 72 h time point are shown in **Fig. 4**.

160 As shown in **Fig. 4A**, toxin production of 630, 630 Δ *erm* or 630 Δ *erm* A+B- clustered
161 together, as did toxin production of 630E, 630E A+B- or 630E A+B- CT with the latter three

162 showing no statistical differences between them ($P > 0.05$, one-way ANOVA with Fisher's
163 LSD test). There was, however, a statistically significant difference between the first three
164 strains (630, 630 Δ *erm* or 630 Δ *erm* A+B-) and the second set of three strains (630E, 630E
165 A+B- or 630E A+B- CT) ($P < 0.0001$). The *C. DIFFICILE TOX A/B II*[™] ELISA does not
166 differentiate between toxin A and B. In order to be able to quantify each toxin, the kits
167 from TGCbiomics were used (**Fig. 4B and 4C**). The toxin A ELISA showed significantly higher
168 production in strain 630 Δ *erm* compared to 630E ($p < 0.0016$) and also confirmed the
169 previous observations that strains with impaired *tcdB*, produce more toxin A^{8,9} (**Fig. 4B**).
170 No toxin B production was seen, as expected, in the *tcdB*-mutants. Strains 630 and
171 630 Δ *erm*, however, both produced significantly more toxin B than 630E ($p < 0.0001$) (**Fig.**
172 **4C**).

173

174 **Whole genome sequencing**

175 To establish whether strains 630E and 630 Δ *erm*, and derivatives, contained any
176 additional changes to the *ermB* gene deletion, relative to the parent strain 630, the
177 following strains were sequenced using Next Generation Sequencing platforms: 630E A+B-,
178 630E A+B- CT on Illumina HiSeq (GATC, Germany) and 630 Δ *erm* A-B on a Roche 454
179 (Deepseq, University of Nottingham, UK) and the data compared to the published genome
180 of the parental strain 630¹² and previously sequenced 630 Δ *erm* Δ *pyrE*.¹⁹ We used a
181 frequency of 70% as a cut-off for SNP calling and found multiple SNPs, InDels and other
182 minor changes, both common and unique to 630, 630E and 630 Δ *erm*. In total, two SNPs in
183 coding regions with non-synonymous changes were found that were common to all three
184 strains (in CD630_11900, encoding an acyl-CoA N-acyltransferase where SNP changes

185 phenylalanine133 to leucine and in CD630_13880, a pseudo gene where a frameshift is
186 introduced). In addition to these, we found in both 630 and 630 Δ erm strains three SNPs
187 (two in intergenic regions and one in a coding region of CD630_2667, encoding the BC
188 domain of a glucose PTS, changing valine228 to isoleucine). 630 Δ erm had seven unique
189 changes compared to 630 and 630E (including six non-synonymous SNPs in coding
190 regions), while 630E had eleven SNPs (with nine non-synonymous SNPs in coding regions)
191 not found in the other two strains. SNPs were confirmed by Sanger sequencing and
192 thereafter by RNAseq data (see below). Indeed, the SNPs found in the DNA-seq data were
193 validated by using the RNA-seq sequence reads mapped on the genome sequence with
194 Bowtie2²⁰ with each position being checked using Tablet²¹. A complete list of SNPs and
195 other small changes are indicated in **Table 1** and in supplementary **Table S1**. During the
196 preparation of this manuscript a new sequence of 630 was published by Riedel *et al.*²². We
197 incorporated their data into **Table 1** (and **Table S1**). Overall this new sequence shows very
198 few disparities to the original one. However two SNPs found in our data were attributed to
199 mistakes in the original sequence (in CD630_17670 and CD630_31561). Another paper was
200 recently published by van Eijk *et al.*,²³ resequencing 630 Δ erm. Overall there are very few
201 discrepancies between their data and our findings, confirming the quality of both data sets.
202 We have incorporated their findings into **Table S1**.

203 It may be assumed, that during the repeated subculture of strain 630 undertaken in
204 the Mullany¹³ and Rood¹⁴ laboratories, sub-populations within the culture were isolated
205 carrying SNPs. However, it seems improbable that the two SNPs (**Table 1**), common to all
206 three strains, arose independently. Rather we hypothesise that these SNPs might be
207 sequencing mistakes. This theory gains weight through the new sequencing data by Riedel

208 *et al.*²². Two SNP changes which we identified originally between the published 630
209 sequence and our data were confirmed by Riedel et al to also be the sequence of their 630
210 seed stock. Unfortunately the genome announcement²² does not state the exact source of
211 their 630 strain. As mentioned above another three SNPs were only found in 630 and
212 630 Δ *erm*, two of these are in intergenic regions which showed no expression in our
213 RNAseq experiment, and the third is located in a PTS gene in 630 and 630 Δ *erm* (position
214 3080703, Val₂₂₈Ile). Rather than having occurred independently it is more likely that these
215 SNPs arose in the Mullany laboratory, subsequent to provision of chromosomal DNA to the
216 Sanger Centre for determination of the 630 genome sequence¹², and before the strain 630
217 was passaged to obtain 630 Δ *erm*. At the time, *C. difficile* strains in the Mullany laboratory
218 were routinely stored at 4°C as Robertson’s Cooked Meat stocks, as opposed to being
219 frozen at -80°C in 10% glycerol (A.R. Roberts, personal communication). On this basis, the
220 traditional microbiological practice of using Robertson’s Cooked Meat to curate strains
221 might not be ideal as strains are not entirely dormant and genome changes can occur over
222 time. The SNPs that were found to be unique to 630 Δ *erm* and 630E (n=8 and n=11,
223 respectively) can be assumed to have been accrued at some point after the two 630
224 populations diverged, that is when the strain was sent to the Rood laboratory. It is most
225 likely, although not certain, that the majority, if not all of the strain-specific SNPs arose
226 during the repeated subculture experiments undertaken to isolate the *ermB* deletion
227 strains 630E and 630 Δ *erm*.

228 Changes specific to 630 Δ *erm* include SNPs in three intergenic regions, which all have
229 been determined with a coverage of over 150 and 100 % frequency (see **Table 1**). The
230 other five changes comprise four non-synonymous SNPs and an insertion. The insertion

231 has previously been reported by Rosenbusch *et al.*²⁴ and was confirmed by van Eijk *et al.*²³
232 and is an 18 bp duplication in *spo0A*, the master regulator of sporulation. This insertion
233 might be responsible for the reduced sporulation frequency seen in strain 630 and also in
234 630E and derivatives, which do not carry this duplication (which does not have this
235 duplication) (**Fig. 3A**). The SNPs have been found in the following genes: CD630_08260,
236 encoding a ferric uptake regulator (*perR* homologue) (Thr₄₁Ala); CD630_19070, encoding
237 an alcohol dehydrogenase homologue (*eutG*) (Gly₂₅₂Glu); and CD630_35630, encoding a
238 transcriptional regulator of the GntR family (Ala₉₁Val).

239 In contrast, strain 630E contains a larger number of non-synonymous SNPs including
240 changes that result in nonsense mutations and in one case the inversion of a small
241 segment of DNA preceding a flagella operon. We found two changes in intergenic regions,
242 one with 100 % frequency and a coverage of 94 (position 3528736); the other at a much
243 lower frequency (41 %), but confirmed a 150 bp inversion by Sanger sequencing in the
244 promoter region of *flgB*, the first gene in a F3 flagella operon (early flagella genes). Non-
245 synonymous SNPs were found in CD630_07610, encoding a putative RNA helicase
246 (Asp₁₃₆Tyr); CD630_14040, encoding an oligopeptide transporter (Glu₅₃₆Gly);
247 CD630_20270, encoding a hydrolase (Gly₃₇₃Glu); CD630_29430, encoding a phage
248 replication protein (Asn₂₁₀Asp); and CD630_33790, encoding a conjugative transposon
249 protein (Glu₆₃Asp). Finally, there is another SNP at position 3034953, in gene CD630_26270
250 (Gly₆₈Cys), encoding a conserved hypothetical protein. Interestingly the new genome
251 sequence from Eijk *et al.*²³ suggested an “A” at position 3034953 in contrast to the earlier
252 annotation suggesting “C”. The new annotation is in line with our RNAseq data (**Table S1**)
253 and taken into account our sequencing data (**Table 1**) suggests that this is indeed a

254 mutation in 630E and was miss-annotated in the original sequence. In one instance the
255 nucleotide substitution resulted in the creation of a nonsense, stop codon, and as a
256 consequence a severe, premature truncation of the encoding protein. Thus, the stop codon
257 introduced into CD630_12740 encoding a topoisomerase I (*topA*) homologue (Gln₃₈₆*)
258 truncated the protein from 695 amino acids to 385 amino acids. Conversely, in the case of
259 the glucose PTS operon *ptsG-BC*, the conversion of the stop codon of *ptsG-B* gene
260 (CD630_26670) to a Glu codon (*₅₂₄Glu) resulted in its fusion to the coding region of the
261 immediately downstream *ptsG-C* gene.

262 **Virulence testing of 630Δ*erm*, 630E and mutants using an *in vivo* model**

263 In order to confirm previous data and to rule out differences in experimental set up in
264 different laboratories, the virulence of 630E and derivatives was assessed using the
265 hamster infection model in our laboratory (University of Nottingham) as previously
266 described.⁹

267 **Figure 5** shows the times from infection to endpoint (in days) for the hamsters
268 infected with 630E, 630E A+B- and 630E A+B- CT. For comparative purposes data for
269 infection with 630Δ*erm* and 630Δ*erm* A+B- from a previous study⁹ is also included. The
270 latter emphasises the fact that all eight hamsters infected with strain 630Δ*erm* were
271 colonised and succumbed to *C. difficile* disease (with an average time of 3.25 days from
272 infection to endpoint). This is in direct contrast to what is observed with 630E where of the
273 five animals successfully infected, only three were colonised till the respective endpoints
274 and of these, two succumbed to disease (at day two and six). Two animals lost colonisation
275 after days 15 and 18, respectively.

276 In our previous study⁹ seven, of the eight animals infected with 630 Δ erm A+B- (as also
277 shown in **Fig. 5**), succumbed to disease with an average time to death of colonised
278 hamsters being just under two days. One animal showed no signs of disease until the
279 experimental endpoint, but was found not to have been colonised. Here of the 11 animals
280 infected with the equivalent mutant of strain 630E (630E A+B-), only two animals
281 succumbed to CDI (on day two and nine). Four animals in this group were never colonised,
282 one lost colonisation after day three and the others were colonised till endpoint. Six
283 animals were infected with 630E A+B- CT, and of these two hamsters developed infection
284 (day three and five). Two of the surviving animals lost colonisation after day 15 and 18
285 respectively. (**Fig. 5** and **Table S2**).

286 The difference between the average time to death of all hamsters administered
287 630 Δ erm and 630 Δ erm A+B- was found not to be statistically significant (one-way ANOVA,
288 $p=0.5355$) (results from Kuehne *et al.*⁹). Similarly, the differences between the average
289 times to death of all animals administered 630E, 630E A+B- and 630E A+B- CT was not
290 statistically significant (one-way ANOVA, $p=0.8919$). In contrast, the difference between
291 the 630 Δ erm strain (and derivative) and the 630E strain (and derivatives) was statistically
292 significant (one-way ANOVA, $p>0.0001$).

293

294 **Correction of SNPs in strain 630E**

295 In view of the large number of SNPs present in strain 630E, it was impractical to
296 change them all back to the 630 parental sequence. We therefore selected just three
297 specific mutations present in 630E and converted them back to the sequence present in
298 the parental strain, 630.

299 Our principal target was to remove the stop codon from within the topoisomerase I
300 gene, CD630_12740, as this enzyme plays a central role in the regulation of DNA negative
301 supercoiling and its inactivation is likely to result in extensive pleiotropic effects. Indeed, in
302 some bacteria its inactivation is lethal.²⁵⁻²⁷ Moreover, bacterial genes related to
303 pathogenesis and virulence have been shown to be sensitive to *topA* mutation in *E. coli*,²⁸
304 *S. flexneri*,²⁹ *Yersinia enterocolitica*³⁰ and *Salmonella*.³¹ We therefore converted the “T”
305 nucleotide at position 1480649 in 630E back to an “A” nucleotide, thereby removing the
306 nonsense stop codon and allowing the production of full length native topoisomerase
307 enzyme.

308 As a second target we elected to correct the inversion of DNA upstream of the F3
309 flagella operon. As strain 630E is non-motile, and as the inverted region encompasses the
310 non-coding region immediately upstream of the *flgB* gene, it is likely to have disrupted the
311 promoter responsible for both *flgB* expression and the genes in the downstream operon.
312 The inversion is therefore likely to be the principal cause of the loss of motility in 630E.
313 Furthermore, factors affecting flagella expression can also influence toxin expression
314 levels.^{32, 33}

315 Finally, we sought to correct the fusion of the two PTS components *ptsG-B* and *ptsG-C*,
316 by resurrection of the stop codon of *ptsG-B* through the conversion of the “C” nucleotide
317 at position 3079815 back to an “A” nucleotide. As glucose is known to affect toxin
318 production, through catabolite repression,^{34, 35} it was reasoned that this particular SNP
319 could be affecting toxin expression, and therefore virulence.

320 The plasmids carrying the 630 wildtype alleles necessary for the correction of the
321 three targeted SNPs were assembled as described in Materials and Methods and then used

322 to effect the replacement of the 630E mutant alleles by allelic exchange.¹⁵ To verify that
323 the mutant clones obtained were correct, each targeted region was amplified by PCR using
324 appropriate oligonucleotide primers and the DNA fragments obtained subjected to Sanger
325 sequencing on both DNA strands. In every case, clones carrying the desired 'corrected'
326 sequence were obtained. The new strains were named after the genes or regions that
327 were corrected, namely 630E_*topA*, 630E_CD2667 and 630E_*flgB*, respectively.

328 To assess the effects of the changes on the characteristics of the mutant strains,
329 growth rate, sporulation and germination, motility, and *in vitro* cytotoxicity and toxin
330 production (ELISA) were measured. None of the three corrected mutants exhibited any
331 difference in growth rate compared to the parental strain 630E (data not shown). Similarly,
332 sporulation and germination remained unaffected (data not shown). Toxicity testing
333 revealed no difference to 630E using the *C. DIFFICILE TOX A/B II*[™] ELISA assay kit from
334 TechLab (**Fig. 4A**). To quantify toxin A and toxin B individually the ELISA kits from
335 TGCbiomics were used to assay 630E_*topA* and 630E_CD2667 (**Fig. 4B**). No differences
336 were measured for toxin A, but the strain 630E_*topA* showed significantly higher levels of
337 toxin B than the parental strain 630E.

338

339 **Transcriptomic comparison of 630, 630Δ*erm* and 630E**

340 RNA was extracted from strains 630, 630Δ*erm* and 630E at 6, 14 and 24 h and used in
341 an RNAseq experiment as described in Materials and Methods. The Principal Component
342 Analysis (PCA) (**Fig. 6**) showed that strain 630Δ*erm* and 630E are closely correlated on a
343 transcriptional level which is significantly separated from 630. While this result implies that
344 both strains are fundamentally different to the parental strain, it does not indicate that the

345 differences to 630 are the same for both strains. The analysis depicted by the Venn
346 diagram (**Fig. 7**) confirms the results of the PCA, showing that the majority of differentially
347 expressed genes are observed comparing 630 Δ *erm* and 630E to 630. From a total of 1337
348 differentially expressed genes (**Table S3** contains all the genes differentially expressed
349 along the growth and also comparisons between the strains), only 139 were common
350 between all three strains. A total of 345 were common between 630E and 630 Δ *erm*, 60
351 were common between 630 and 630 Δ *erm* and 58 were common between 630 and 630E.

352 Most of the 345 genes differentially expressed in both, 630 Δ *erm* and 630E, were either
353 up or down-regulated in the same way highlighting again how distinct the two strains are
354 from the parental strain 630 (**Table S3**). In TY medium used for the transcriptomic
355 experiments, known as a non-optimal for spore production, a total of 44 sporulation genes
356 were differentially expressed in both 630 Δ *erm* and 630E, and all of these were
357 downregulated at 14 and 24 h compared to 6 h. No further differentially expressed
358 sporulation genes appeared in 630 Δ *erm*, however, our analysis showed a further 22
359 sporulation genes, of which 21 were downregulated, in 630E. Amongst these was the
360 master regulator of sporulation *spo0A*. Nine genes classed as stress-related are
361 differentially expressed in all three strains (five upregulated), with a further three in 630
362 (all upregulated), seven in 630 Δ *erm* (four upregulated) and 16 in 630E (11 upregulated).
363 Nine genes related to secretion are down regulated in 630E and one gene related to type
364 IV pili is upregulated. In comparison only one secretion gene (putative pilus assembly
365 ATPase) is differentially expressed only in 630 Δ *erm* (down regulated) and none in 630.
366 Metabolism is also highly differentially regulated in the three strains. 90 genes were
367 uniquely, differentially expressed in 630E, 50 in 630 Δ *erm* and 30 in 630. In particular the

368 amino acid metabolism stands out for 630E with the majority of genes being
369 downregulated. (**Table S3**).

370 RNAseq data can be used to independently corroborate genome re-sequencing data.
371 Thus, it was apparent that those changes identified by CLC Bio as being present with a
372 frequency of 70% or less, except for CD630_20102, were not real accordingly to the
373 RNAseq analysis (**Table S1**). This increases the confidence in disregarding changes
374 identified by NGS with a low frequency. In most cases, the SNPs and Indels identified by
375 NGS were confirmed by the RNAseq analysis, with the following exceptions: for SNPs in
376 630E we found two disagreements notably in CD630_33790 and CD630_29430, which both
377 had 100 % frequency and a high coverage (around 200 reads) in the DNA sequence
378 analysis, but only low coverage in the RNAseq experiment. Due to the low coverage of
379 these regions during the RNAseq experiment, which is indicative of low or no expression
380 under the examined conditions, a sequencing error cannot be excluded. For the SNP in
381 CD630_12740 the RNA coverage corroborated the genomic data for 630E, but was in
382 disagreement with the genomic data for 630 and 630 Δ *erm*. For three SNPs in 630 Δ *erm*
383 similar scenarios were observed. CD630_19070 had very low RNA coverage, CD630_35650
384 showed ambiguous RNA data with low coverage for 630E. CD630_08260, the *perR*
385 homologue, had convincing DNA data, with frequencies of 98-100 % and coverage of at
386 least a 100 which was corroborated for 630 Δ *erm* by RNAseq coverage.

387 In terms of actual expression data (**Table S1**), CD630_07610 (the RNA helicase),
388 CD630_14040 (oligopeptide transporter), CD630_20270 (hydrolase) and CD630_29430
389 (phage replication protein) all showed differential expression in 630E compared to the
390 other two strains, with the first two showing reduced expression and the latter two an

391 increase. CD630_29430, however, also showed an increase in expression in 630 at the
392 later time point. Changes in CD630_26670 in 630 Δ *erm* and 630E both seem to lead to
393 severely reduced expression. CD630_12740 (*topA*) only showed differential expression at
394 24 h in 630 Δ *erm* and CD630_12140 (*spo0A*) expression was severely reduced in 630E.

395

396 **Discussion**

397 Previously, two studies ^{8,9} have attempted to use isogenic mutants defective in the
398 production of either toxin A or toxin B to determine the relative importance of these two
399 virulence factors in CDI using the hamster infection model. However, despite generating
400 essentially equivalent A+B-insertion mutants in ostensibly the same strain of *C. difficile*
401 (630), contradictory outcomes were obtained in terms of the importance of toxin A. Thus,
402 a *tcdB* mutant created in the one study ⁸ producing only TcdA did not cause disease in the
403 hamster, whereas the equivalent ClosTron mutant made in our laboratory (Minton group) ⁹
404 remained virulent. The work undertaken here has provided compelling evidence that the
405 reason for the observed conundrum resides in the use of two different erythromycin-
406 sensitive derivatives of strain 630.

407 Here we have shown that both erythromycin-sensitive derivatives, 630E ¹⁴ and
408 630 Δ *erm* ¹³ carry a significant number of SNPs compared to the published sequence.
409 Moreover, it is clear that whilst the phenotypic properties of 630 Δ *erm* and its mutant
410 derivatives closely resemble that of the parent strain 630, strain 630E and its progeny
411 exhibit substantive differences. Thus, whereas latter strains exhibit reduced growth rates,
412 are less proficient in spore formation and are non-motile, 630 Δ *erm* strains mirror the
413 behaviour of the 630 parental strain with respect to these phenotypes. Furthermore, 630E

414 strains produce reduced amounts of toxin and both struggle to colonise hamsters, and
415 once colonised, animals are less likely to succumb to disease. In short, 630E and its
416 derivatives (i.e., 630E A-B+ and 630E A+B- CT) are less virulent than 630 Δ *erm* and its
417 mutant counterparts (i.e., 630 Δ *erm* A+B-).

418 The altered properties of 630E and its derivatives are undoubtedly a consequence of
419 the observed SNPs. However, the substantive number of changes involved makes it
420 difficult to assign any particular SNP to a specific alteration in the observed phenotype,
421 particularly as a combination of mutagenic changes could be responsible. Whilst it is now
422 possible to make precise changes to the genome using allelic exchange methodologies¹⁵ it
423 is not practically feasible to make all of the sequential rational changes needed to
424 definitively identify the mutation(s) responsible for a particular phenotype. As such, we
425 only corrected three specific SNPs that we reasoned may be making a significant
426 contribution. The outcomes of these experiments only emphasised the difficulty of such an
427 undertaking, and served to highlight the dangers involved in making assumptions. Thus,
428 whilst it seemed reasonable to assume that the DNA inversion within the promoter region
429 of the flagella operon was likely to have caused the observed non-motile phenotype, this
430 surprisingly proved not to be the case. Re-inversion of the 150 bp region failed to restore
431 motility. Clearly other SNPs are at least partly responsible for the observed lack of motility.
432 Singling out any other SNP as the culprit would in the absence of experimental evidence be
433 counterproductive.

434 Equally negative was the observed outcome of correcting the mutation in
435 CD630_12740 that results in a truncation of the encoded topoisomerase I enzyme. Given
436 this enzyme controls DNA supercoiling, and given that its mutation in certain bacteria is

437 either a lethal event ²⁵⁻²⁷ and/or is involved in the regulation of virulence factors, ²⁸⁻³⁰ it
438 seemed likely that its presence would result in pleiotropic effects that could have
439 contributed to the observed phenotypic changes. However, its correction, with the
440 exception of a measurable increase in toxin B levels, seemingly had no effects on the
441 behaviour of the strain, at least for those properties measured. The reasons are not clear.
442 In other bacteria, mutations of *topA* are only isolated if compensatory mutations arise
443 elsewhere in the genome. ²⁶ Whether any of the other SNPs present in 630E (eg., the RNA
444 helicase mutation) are negating the effects of the TopA truncation is currently unknown.

445 To understand the differences observed further, we analysed the transcriptome of
446 630, 630 Δ *erm* and 630E, comparing expression at 6 h to 14 h and to 24 h (**Table S3**). The
447 data corroborated the phenotypic analysis, showing vastly different transcriptomes for all
448 three strains. While 630E and 630 Δ *erm* cluster together in the PCA (**Fig. 6**), this only
449 highlights how different the two strains really are from the progenitor. The analysis clearly
450 shows that the three strains are very different from each other and also serves as an
451 explanation as to why the change of a single SNP could not restore any given phenotype.
452 Overall 630E seems the most divergent with many genes differentially expressed involved
453 in metabolism and regulation (**Table S3**). Additionally 32 genes grouped under the
454 descriptor 'cell factor', many of which play a role in energy metabolism, are differentially
455 expressed in 630E, with only six of these being upregulated. In contrast out of 19 genes in
456 630, 11 are upregulated and, out of 10 in 630 Δ *erm*, six were upregulated. The number of
457 genes downregulated in energy metabolism in 630E might relate to the growth differences
458 seen between the strains.

459 Interestingly 22 genes involved in sporulation are differentially expressed, 21 of these
460 downregulated, in 630E versus one in 630 and two in 630 Δ *erm*. This is consistent with the
461 observed delay in sporulation and reduced amount of spores produced by 630E. Secretion
462 also seems most affected in 630E, with 10 genes differentially expressed, compared to
463 none in 630 and one in 630 Δ *erm*. A general defect in secretion could affect the secretion
464 of certain virulence or adhesion factors. Furthermore, 33 cell wall genes are differentially
465 expressed in 630E (compared to 14 each in 630 Δ *erm* and 630). This may also contribute to
466 the observed colonisation deficiencies. Mobile elements are, however, more differentially
467 transcribed in 630 and 630 Δ *erm* (15 each) versus 630E (nine). As in many cases different
468 pathways were affected, we propose that this could at least in part explain the different
469 adaptability and virulence of the two strains. In both strains many regulators were
470 differentially affected providing a further basis for the observed phenotypic variation
471 between strains.

472

473 **Conclusion**

474 Our study has established that the parental strains (630E and 630 Δ *erm*) used in the
475 two previous studies, that explored the relative roles of toxin A and toxin B in disease,^{8,9}
476 are phenotypically and genetically distinct. Here we also reveal that the three strains (630,
477 630 Δ *erm* and 630E) have vastly different transcriptomes, which no doubt lead to the
478 different phenotypes observed. This immense diversity also underlines our finding that by
479 restoring just one SNP, the entire transcriptome cannot be changed. The presence of SNPs
480 in strain 630E significantly affects its transcriptome which in turn has a significant impact
481 on growth, sporulation and finally virulence of this strain in the hamster model of infection

482 under the conditions tested. Data (such as motility, toxicity and virulence) obtained with
483 strain 630 Δ *erm* reflects more accurately the behaviour of the parent strain 630. As such, it
484 may be concluded that 630 producing toxin A alone will cause disease in the hamster. As a
485 consequence, toxin A should remain a target for the rational development of effective
486 countermeasures against *C. difficile*.

487 This study has also highlighted a number of issues that need to be borne in mind in the
488 future. At a specific level, if researchers wish to undertake genetic-based studies with
489 strain 630, then the use of strain 630 Δ *erm* should be favoured over strain 630E. At a more
490 fundamental level, researchers need to effectively curate their strains to prevent the
491 inadvertent isolation of SNPs. Ideally, master seed banks need to be established as frozen
492 glycerol stocks. Moreover, the genome of the stored strain should be re-sequenced as part
493 of the storage process whenever a strain is received from external sources, regardless of
494 whether it has been re-sequenced in the sending laboratory.

495

496 **Materials and Methods**

497 **Bacterial strains and routine culture conditions**

498 Bacterial strains and plasmids used in this study are listed in **Table 2**. *E. coli* was
499 cultured aerobically at 37°C with shaking at 200 rpm in LB medium with chloramphenicol
500 supplementation (25 μ g/ml) where appropriate. *C. difficile* was cultured in TY (tryptose
501 yeast) medium supplemented with thiamphenicol (15 μ g/ml) where appropriate. When
502 needed, *C. difficile* strains were plated on BHIS agar (Brain Heart Infusion agar [Oxoid]
503 supplemented with 5 mg/ml yeast extract [Oxoid] and 0.1% [wt/vol] cysteine
504 [Calbiochem]) supplemented with d-cycloserine (250 μ g/ml), cefoxitin (8 μ g/ml) [Oxoid]

505 (BHIScc). Fluorocytosine selections were carried out on *C. difficile* minimal medium
506 (CDMM) as described previously.¹⁵ All *C. difficile* cultures were incubated at 37°C
507 anaerobically in an anaerobic MACS1000 workstation (Don Whitley, Yorkshire, UK).

508 **Mutant nomenclature**

509 For the sake of simplicity, *C. difficile* strains that carried a *tcdA* insertional mutant were
510 referred to as A-, those carrying a *tcdB* mutant as B-, and those strains carrying a mutation
511 in both genes as A-B-. To avoid any ambiguity, if the gene was not inactivated it was
512 referred to as A+ or B+, as appropriate. Thus, a *tcdA* mutant of 630 Δ *erm* constructed using
513 ClosTron technology as described,^{9, 16} was designated 630 Δ *erm* A-B+. The equivalent
514 mutant in 630E constructed through the insertion of a replication-deficient plasmid,
515 according to the method of O'Connor et al.,¹⁴ was designated 630E A-B+. When ClosTron
516 technology was used in 630E, this was clarified by adding a 'CT' suffix, 630E A-B+CT to the
517 strain designation.

518

519 **Whole genome sequencing and bioinformatics**

520 Genomic DNA from strains 630E A+B-, 630E A+B-CT and 630 Δ *erm* A+B- was prepared
521 by phenol:chloroform extraction. 630E A+B-, 630E A+B-CT, 630 Δ *erm*(Δ *pyrE*) were
522 sequenced on Illumina highseq (GATC, Germany) and 630 Δ *erm* A+B- on a Roche 454
523 (Deepseq, Nottingham, UK) and the data compared to the published genome of 630^{11,19}
524 using CLC genomic workbench. All raw sequencing data have been deposited in the
525 sequence read archive (SRA) under the study name PRJNA304508. The accession number is
526 SRP066836. The sequencing data for 630 Δ *erm* Δ *pyrE* had been obtained previously¹⁹ and
527 with no additional changes, other than the *pyrE* deletion, present compared to 630 Δ *erm*,

528 were used to analyse the parental strain 630 Δ *erm*. We used a frequency of 70% as a cut-off
529 for SNP calling. SNPs, InDels and inversions were confirmed by amplifying a few hundred
530 base pairs up- and downstream of the area of interest (primers are listed in **Table S4**) and
531 the amplicon was Sanger sequenced (Source BioScience, UK). This confirmation was done
532 on all strains including the parental strains (630, 630 Δ *erm*, 630E) and the derivatives
533 (630 Δ *erm* A+B-, 630E A+B-, 630E A+B-CT).

534

535 **Correction of SNPs and reversal of 150-bp region within the flagellar operon**

536 Using the method described by Cartman et al.¹⁵ we “corrected” the two SNPs and an
537 inversion in 630E to the 630 Δ *erm* genotype. A stretch of DNA corresponding to
538 approximately 500 bp either side of the area to be altered was synthesised by Biomatik
539 and cloned into plasmid pMTL-SC7315 λ 2.3. This vector was transformed by
540 electroporation into *E. coli* CA434 cells³⁶ and subsequently conjugated into 630E. Single
541 crossover colonies were identified as those growing faster on plates containing
542 thiamphenicol. Following overnight incubation on CDMM containing 5-fluorocytosine,
543 colonies were incubated on BHIScc plates with and without thiamphenicol. Those strains
544 that had lost the plasmid (both wildtype and double crossover) were unable to grow on
545 thiamphenicol. SNP corrections were confirmed by PCR (Primers see **Table S4**) and Sanger
546 sequencing (Source BioScience, UK).

547

548 **ClosTron mutagenesis**

549 A *tcdB* mutant was generated in the 630E background according to the published
550 method,¹⁶ using the same plasmid that was used to generate the equivalent mutant in

551 630 Δ erm.⁹ This newly created strain was referred to as 630E A+B- CT.

552

553 **Southern and Western blot**

554 The Southern and Western blot were performed as described in Kuehne et al., 2010.⁹

555

556 ***In vivo* testing of mutants**

557 *In vivo* testing was carried out in Syrian Golden hamsters (Charles River, Germany) as
558 previously described.⁹ Briefly, clindamycin was administered orally on day -five to render
559 the animals susceptible to infection. On day zero, 10,000 spores were administered orally.
560 Animals were assessed for signs of CDI (weight loss, wet tail, lethargy, lack of response to
561 stimulus) six times a day for the first five days, and once daily for the following 14 days. At
562 this point animals that failed to display signs of CDI were euthanised. Faecal pellets were
563 collected daily from day zero to endpoint, homogenised and plated on *C. difficile* fructose
564 agar (CDFA). *C. difficile* colonies were sub-cultured onto BHIS agar and the genotype was
565 established by PCR (primers in **Table S4**) followed by Sanger sequencing (results in **Table**
566 **S2**). At the experimental endpoint, part of the caecum of each animal was collected. This
567 was also used to plate on CDFA to verify colonisation.

568

569 ***In vitro* testing of mutants**

570 **Growth curves:** To assess the effects of SNPs and “corrected” SNPs/inversion on the
571 growth characteristics of all strains, we performed growth curve experiments over 24 h. A
572 180 μ L volume of TY medium was inoculated with 20 μ L of an overnight culture in 96-well
573 plates and incubated for 24 h in a GloMax-Multi Microplate Multimode Reader (Promega,

574 USA). Samples were shaken every h and OD₆₀₀ measurements were taken immediately
575 after.

576 **Motility assays:** 2xYTG (tryptone (1.6%), yeast (1%), NaCl (0.4%), Gelzan (0.24%)[Sigma-
577 Aldrich] and glucose (0.5%)) agar was utilised. 25 mL were poured into each petri dish and
578 let to solidify at room temperature for 15 min. The plates were then dried at 37°C for 30
579 min. The plates were placed into the anaerobic cabinet 24 h before use. 2 µL volumes of
580 overnight culture were 'dropped' onto each plate. Plates were incubated anaerobically for
581 48 h. Motility was assessed by eye and the plates photographed.

582 **Sporulation and germination assays:** Sporulation and germination assays were carried out
583 as previously described.^{17,18} Briefly, for the sporulation assay cultures were grown for five
584 days, with 2 X 500 µl samples taken at 0, 24, 48, 72, 96 and 120 h. One sample from each
585 time point was heated to 65°C for 30 min while the other sample was kept at room
586 temperature. After this time, samples were serially diluted from 10⁰ to 10⁻⁷ in PBS. 3 X 20µl
587 of each dilution was spotted onto BHIS plates containing 0.1% taurocholic acid and were
588 incubated for 24 h. The following day, colonies were counted and CFU/mL were calculated.
589 A 630Δ*erm spo0A::CT* mutant strain (containing a ClosTron insertion in the *spo0A* gene³⁷),
590 which is unable to form spores, was used as a negative control.

591 Germination was measured as a function of the ability of a germinated spore to
592 outgrow in the absence of taurocholate. Spore stocks were prepared as previously
593 described by Heeg *et al.*, 2012³⁸ and stored at -20°C. The optical density of spore
594 suspensions (OD₆₀₀) was adjusted to 1.0 and 450 µl was used per measurement. This
595 equated to approximately 2.5 X 10⁷ spores. Spore suspensions were heat treated at 60°C
596 for 25 min to kill any remaining vegetative cells and then centrifuged and resuspended in

597 BHIS with the germinant taurocholic acid (0.1%) in a total volume of 20 ml. Samples were
598 taken at 5, 10, 20, 30, 60, 90, 120, 180 and 240 min, briefly centrifuged, washed and
599 resuspended in phosphate-buffered saline (PBS), the samples were then diluted and plated
600 on plain BHIS agar. Plates were incubated for 24 h before the CFUs were enumerated.
601 Colonies that grew on these plates were considered to be germinated vegetative bacteria.

602

603 **Toxin A/B ELISA:** ELISA assays were performed using the *C. DIFFICILE TOX A/B II*[™] kit from
604 TechLab and the TGC-E002-1-separate detection of *C. difficile* toxins A and B kit from
605 TGCbiomics according to the manufacturers' instructions. Cultures of *C. difficile* were
606 grown in TY medium without glucose for 72 h, at which time 1 ml samples were taken,
607 centrifuged and the supernatant filter-sterilised and used for the ELISAs. A 1:2 dilution was
608 used for the toxin B ELISA kit from TGCbiomics. To quantify toxins a standard curve with
609 pure toxin (the native antigen company) was established for the TechLab ELISA and also
610 Toxin A ELISA from TGCbiomics.

611

612 **RNAseq:** Total RNA was extracted from 2 independent biological replicates of 630,
613 630 Δ *erm* and 630E strains at 3 time points (18 samples). Bacteria were grown in TY broth
614 medium after 6, 14 and 24 h as previously described.³⁵ The mRNA was treated with
615 MicroExpress kit (Ambion). For oriented RNA-seq library construction, the Truseq
616 stranded RNA seq Illumina kit was used according to manufacturer's instructions before
617 sequencing using the Illumina HiSeq 2500 machine. Sequencing reads were mapped using
618 Bowtie³⁹ to the reannotated 630 reference genome⁴⁰ complemented with the known
619 ncRNA.⁴¹ Statistical analyses were performed on each strand coverage count with DESeq2

620 ⁴² using the 6 h value as a reference for reporting the expression data of 14 and 24 h. A
621 gene was considered differentially expressed when the fold change was > 2 and the *P value*
622 was < 0.05.

623 The RNA-seq data discussed in this publication have been deposited in NCBI's Gene
624 Expression Omnibus database under the accession no. GSE72006

625 RNA-seq coverage visualization is available through the COV2HTML software:⁴³

630	14H - 6H	http://mmonot.eu/COV2HTML/visualisation.php?str_id=-20
	24H - 6H	http://mmonot.eu/COV2HTML/visualisation.php?str_id=-22
630 Δ erm	14H - 6H	http://mmonot.eu/COV2HTML/visualisation.php?str_id=-24
	24H - 6H	http://mmonot.eu/COV2HTML/visualisation.php?str_id=-26
630E	14H - 6H	http://mmonot.eu/COV2HTML/visualisation.php?str_id=-28
	24H - 6H	http://mmonot.eu/COV2HTML/visualisation.php?str_id=-30

626

627 **Graphs and statistical analyses**

628 All graphs were generated and statistical analyses were performed using GraphPad
629 PRISM 6.02. Statistical analysis comprised either 2-way ANOVA for multiple comparisons or
630 unpaired t test for pairwise comparisons. All experiments were carried out in triplicate
631 unless stated otherwise.

632

633 **Ethics statement**

634 This work was reviewed and approved locally by the Animal Welfare and Ethical
635 Review Body (formerly the Ethical Review Committee) at the University of Nottingham and
636 performed under a project licence (PPL 40/3590) granted under the Animal (Scientific

637 Procedures) Act, 1986, by the UK Home Office. The work was performed in accordance
638 with the NC3R^s ARRIVE guidelines.⁴⁴

639

640 **Acknowledgement**

641 We would like to thank Haitham Hussain for providing *C. difficile* 630 Δ erm, Dena Lyras
642 for providing *C. difficile* 630E and toxin mutants and Rory Cave for helping with PCRs and
643 Sanger sequencing. We would also like to acknowledge Sean Roberts who was involved in
644 creating the 630E A+B- CT strain. The opinions expressed in this paper are those of the
645 authors and do not necessarily represent those of Merck Sharp & Dohme Limited or the
646 Medical Research Council (UK).

647

648 **References**

- 649 1. Govind R, Dupuy B. Secretion of *Clostridium difficile* toxins A and B requires the
650 holin-like protein TcdE. PLoS Pathogens 2012; 8:e1002727.
- 651 2. Jank T, Giesemann T, Aktories K. Rho-glucosylating *Clostridium difficile* toxins A and
652 B: new insights into structure and function. Glycobiology 2007; 17:15R-22R.
- 653 3. Sun X, Savidge T, Feng H. The enterotoxicity of *Clostridium difficile* toxins. Toxins
654 (Basel) 2010; 2:1848-80.
- 655 4. Lyerly DM, Krivan HC, Wilkins TD. *Clostridium difficile*: its disease and toxins. Clinical
656 Microbiology Reviews 1988; 1:1-18.
- 657 5. Lyerly DM, Barroso LA, Wilkins TD, Depitre C, Corthier G. Characterization of a toxin
658 A-negative, toxin B-positive strain of *Clostridium difficile*. Infection Immunity 1992;
659 60:4633-9.
- 660 6. Kim J, Pai H, Seo MR, Kang JO. Clinical and microbiologic characteristics of tcdA-
661 negative variant *Clostridium difficile* infections. BMC Infectious Diseases 2012; 12:109.
- 662 7. Chaves-Olarte E, Low P, Freer E, Norlin T, Weidmann M, von Eichel-Streiber C,
663 Thelestam M. A novel cytotoxin from *Clostridium difficile* serogroup F is a functional hybrid
664 between two other large clostridial cytotoxins. Journal of Biological Chemistry 1999;
665 274:11046-52.
- 666 8. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, Poon R,
667 Adams V, Vedantam G, Johnson S, et al. Toxin B is essential for virulence of *Clostridium*
668 *difficile*. Nature 2009; 458:1176-9.
- 669 9. Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP. The role of
670 toxin A and toxin B in *Clostridium difficile* infection. Nature 2010; 467:711-3.

- 671 10. Monot M, Eckert C, Lemire A, Hamiot A, Dubois T, Tessier C, Dumoulard B, Hamel B,
672 Petit A, Lalande V, et al. *Clostridium difficile*: New Insights into the Evolution of the
673 Pathogenicity Locus. *Science Reports* 2015; 5:15023.
- 674 11. Kuehne SA, Cartman ST, Minton NP. Both, toxin A and toxin B, are important in
675 *Clostridium difficile* infection. *Gut Microbes* 2011; 2:252-5.
- 676 12. Sebahia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson
677 NR, Roberts AP, Cerdeno-Tarraga AM, Wang H, et al. The multidrug-resistant human
678 pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genetics* 2006;
679 38:779-86.
- 680 13. Hussain HA, Roberts AP, Mullany P. Generation of an erythromycin-sensitive
681 derivative of *Clostridium difficile* strain 630 (630Deltaerm) and demonstration that the
682 conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites.
683 *Journal of Medical Microbiology* 2005; 54:137-41.
- 684 14. O'Connor JR, Lyras D, Farrow KA, Adams V, Powell DR, Hinds J, Cheung JK, Rood JI.
685 Construction and analysis of chromosomal *Clostridium difficile* mutants. *Molecular*
686 *Microbiology* 2006; 61:1335-51.
- 687 15. Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP. Precise manipulation of the
688 *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype
689 and toxin production. *Applied and Environmental Microbiology* 2012; 78:4683-90.
- 690 16. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, Minton NP. The
691 Clostron: Mutagenesis in *Clostridium* refined and streamlined. *Journal of Microbiological*
692 *Methods* 2010; 80:49-55.
- 693 17. Burns DA, Heap JT, Minton NP. SleC is essential for germination of *Clostridium*
694 *difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate.
695 *Journal of Bacteriology* 2010; 192:657-64.
- 696 18. Burns DA, Heeg D, Cartman ST, Minton NP. Reconsidering the sporulation
697 characteristics of hypervirulent *Clostridium difficile* BI/NAP1/027. *PLoS One* 2011;
698 6:e24894.
- 699 19. Ng YK, Ehsaan M, Philip S, Collery MM, Janoir C, Collignon A, Cartman ST, Minton
700 NP. Expanding the repertoire of gene tools for precise manipulation of the *Clostridium*
701 *difficile* genome: allelic exchange using *pyrE* alleles. *PLoS One* 2013; 8:e56051.
- 702 20. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature*
703 *Methods* 2012; 9:357-9.
- 704 21. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D.
705 Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform*
706 2013; 14:193-202.
- 707 22. Riedel T, Bunk B, Thurmer A, Sproer C, Brzuszkiewicz E, Abt B, Gronow S, Liesegang
708 H, Daniel R, Overmann J. Genome Resequencing of the Virulent and Multidrug-Resistant
709 Reference Strain *Clostridium difficile* 630. *Genome Announcements* 2015; 3.
- 710 23. van Eijk E, Anvar SY, Browne HP, Leung WY, Frank J, Schmitz AM, Roberts AP, Smits
711 WK. Complete genome sequence of the *Clostridium difficile* laboratory strain 630Deltaerm
712 reveals differences from strain 630, including translocation of the mobile element CTn5.
713 *BMC Genomics* 2015; 16:31.
- 714 24. Rosenbusch KE, Bakker D, Kuijper EJ, Smits WK. *C. difficile* 630Deltaerm Spo0A
715 regulates sporulation, but does not contribute to toxin production, by direct high-affinity
716 binding to target DNA. *PLoS One* 2012; 7:e48608.

- 717 25. Ahmed W, Menon S, Godbole AA, Karthik PV, Nagaraja V. Conditional silencing of
718 topoisomerase I gene of *Mycobacterium tuberculosis* validates its essentiality for cell
719 survival. FEMS Microbiology Letters 2014; 353:116-23.
- 720 26. Stupina VA, Wang JC. Viability of *Escherichia coli topA* mutants lacking DNA
721 topoisomerase I. The Journal of Biological Chemistry 2005; 280:355-60.
- 722 27. Suerbaum S, Brauer-Steppkes T, Labigne A, Cameron B, Drlica K. Topoisomerase I of
723 *Helicobacter pylori*: juxtaposition with a flagellin gene (*flaB*) and functional requirement of
724 a fourth zinc finger motif. Gene 1998; 210:151-61.
- 725 28. Tessier MC, Graveline R, Crost C, Desabrais JA, Martin C, Drolet M, Harel J. Effects
726 of DNA supercoiling and topoisomerases on the expression of genes coding for F165(1), a
727 P-like fimbriae. FEMS Microbiology Letters 2007; 277:28-36.
- 728 29. Ni Bhriain N, Dorman CJ. Isolation and characterization of a *topA* mutant of *Shigella*
729 *flexneri*. Molecular Microbiology 1993; 7:351-8.
- 730 30. Darwin AJ, Miller VL. Identification of *Yersinia enterocolitica* genes affecting survival
731 in an animal host using signature-tagged transposon mutagenesis. Molecular Microbiology
732 1999; 32:51-62.
- 733 31. Galan JE, Curtiss R, 3rd. Expression of *Salmonella typhimurium* genes required for
734 invasion is regulated by changes in DNA supercoiling. Infection and Immunity 1990;
735 58:1879-85.
- 736 32. Aubry A, Hussack G, Chen W, KuoLee R, Twine SM, Fulton KM, Foote S, Carrillo CD,
737 Tanha J, Logan SM. Modulation of toxin production by the flagellar regulon in *Clostridium*
738 *difficile*. Infection and Immunity 2012.
- 739 33. Baban ST, Kuehne SA, Barketi-Klai A, Cartman ST, Kelly ML, Hardie KR, Kansau I,
740 Collignon A, Minton NP. The Role of Flagella in *Clostridium difficile* Pathogenesis:
741 Comparison between a Non-Epidemic and an Epidemic Strain. PLoS One 2013; 8:e73026.
- 742 34. Antunes A, Martin-Verstraete I, Dupuy B. CcpA-mediated repression of *Clostridium*
743 *difficile* toxin gene expression. Molecular Microbiology 2011; 79:882-99.
- 744 35. Antunes A, Camiade E, Monot M, Courtois E, Barbut F, Sernova NV, Rodionov DA,
745 Martin-Verstraete I, Dupuy B. Global transcriptional control by glucose and carbon
746 regulator CcpA in *Clostridium difficile*. Nucleic Acids Research 2012; 40:10701-18.
- 747 36. Williams DR, Young DI, Young M. Conjugative plasmid transfer from *Escherichia coli*
748 to *Clostridium acetobutylicum*. Journal of General Microbiology 1990; 136:819-26.
- 749 37. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. The ClosTron: a
750 universal gene knock-out system for the genus *Clostridium*. Journal of Microbiological
751 Methods 2007; 70:452-64.
- 752 38. Heeg D, Burns DA, Cartman ST, Minton NP. Spores of *Clostridium difficile* clinical
753 isolates display a diverse germination response to bile salts. PLoS One 2012; 7:e32381.
- 754 39. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient
755 alignment of short DNA sequences to the human genome. Genome Biology 2009; 10:R25.
- 756 40. Monot M, Boursaux-Eude C, Thibonnier M, Vallenet D, Moszer I, Medigue C,
757 Martin-Verstraete I, Dupuy B. Reannotation of the genome sequence of *Clostridium difficile*
758 strain 630. Journal of Medical Microbiology 2011; 60:1193-9.
- 759 41. Soutourina OA, Monot M, Boudry P, Saujet L, Pichon C, Sismeiro O, Semenova E,
760 Severinov K, Le Bouguenec C, Coppee JY, et al. Genome-wide identification of regulatory
761 RNAs in the human pathogen *Clostridium difficile*. PLoS Genetics 2013; 9:e1003493.

762 42. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion
763 for RNA-seq data with DESeq2. *Genome Biology* 2014; 15:550.

764 43. Monot M, Orgeur M, Camiade E, Brehier C, Dupuy B. COV2HTML: a visualization
765 and analysis tool of bacterial next generation sequencing (NGS) data for postgenomics life
766 scientists. *OmicS* 2014; 18:184-95.

767 44. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience
768 research reporting: The ARRIVE guidelines for reporting animal research. *Journal of*
769 *Pharmacology and Pharmacotherapy* 2010; 1:94-9.

770

771

772

773 **Figure legends**

774

775 **Figure 1.** Growth curves of strain 630 and its derivatives. A. 630 (630), 630 Δ *erm* and 630E were
776 grown in TY-broth for 24 h in a 96 well plate reader. The optical density at 600 nm was measured
777 every 30 min. B. This graph shows the same growth as A. and in addition the growth of derivatives
778 630 Δ *erm* A+B-, 630E A+B- and 630E A+B- CT.

779

780 **Figure 2.** Motility assays. The assay was carried out by inoculating overnight cultures onto motility
781 agar plates and incubating anaerobically for 48 h. Strains 630, 630 Δ *erm* and 630E were compared
782 for their ability to swarm.

783

784 **Figure 3.** Sporulation and Germination. A. Sporulation over 120 h comparing heat treated CFUs of
785 strains 630, 630 Δ *erm*, 630 Δ *erm* A+B-, 630E, 630E A+B- and 630E A+B- CT with a non-sporulating
786 control (*spo0A*). B. The extent of germination of the indicated strains was measured over 250 min
787 as the ability of germinated spores to form colonies on plates lacking taurocholate.

788

789 **Figure 4.** Toxin ELISAs. A. The *C. DIFFICILE TOX A/B II*[™] ELISA assay kit from TechLab was used to
790 measure combined toxin A and B in strains 630, 630 Δ *erm*, 630 Δ *erm* A+B-, 630E, 630E A+B-, 630E
791 A+B- CT, 630E_*topA*, 630E_*flgB* and 630E_CD2667 grown in TY for 72 h. B. and C. Toxin ELISAs
792 TGcbiomics, measuring the toxins separately were used to quantify toxin A (B) and toxin B (C)
793 produced by strains 630, 630 Δ *erm*, 630 Δ *erm* A+B-, 630E, 630E A+B- and 630E A+B- CT grown in TY
794 for 72 h. Statistics were performed using one-way ANOVA with Fisher's LSD test.

795

796 **Figure 5.** Infection to endpoint in the Hamster infection model. Groups of Golden Syrian Hamsters
797 were challenged with *C. difficile* 630E (5 hamsters), 630E A+B- (11 hamsters) and 630E A+B- CT (6
798 hamsters). The graph represents the time from inoculation to endpoint. The maximal duration of
799 the experiment was set to 20 days. Animals represented in open symbols, have not been colonized
800 despite challenge or lost colonization before day 20. Details can be seen in **Table 1**. The dotted line
801 separates this experiment from data obtained by Kuehne et al. ⁹, which are represented here as a
802 comparator. In that study 8 hamsters were infected with *C. difficile* 630 Δ *erm* and another 8
803 hamsters with 630 Δ *erm* A+B-.

804

805 **Figure 6.** Principal Component Analysis (PCA). The Principal Component Analysis (PCA) visualizes
806 the variance of the data in a single graph. The axis represent the two largest variances of the data;
807 PC1 accounts for 42% and PC2 accounts for 21%, that means that 63% of the total variance of the
808 dataset is explained in this graph. The third component accounts for less than 10% and further
809 components have a value that falls rapidly. The PCA represents the RNAseq data (at three different
810 time points, six, 14 and 24 h) in duplicate for 630 (blue), 630 Δ *erm* (green) and 630E (orange). The
811 different time points are represented as dots in the different shades of the respective colour as
812 indicated in the colour legend.

813

814 **Figure 7.** Venn Diagram representing differentially expressed genes in the three different strains.
815 The diagram summarises the output from the RNAseq data, comparing strains 630, 630 Δ *erm* and
816 630E. It depicts all differentially expressed genes and shows how many genes are differentially
817 expressed in all strains, in two of the strains or are unique to just one strain.

818

Table 1. Single Nucleotide Polymorphisms (SNPs) and other changes found after re-sequencing.

Gene	Description	Position	630 ¹²	630 ²²	630	C	F	630 Δ erm	C	F	630E	C	F	AA
630E														
CD630_07610	Putative ATP-dependent RNA helicase	933139	G	-	-			-			T	199	100	Asp136Tyr
CD630_14040	Putative oligopeptide transporter	1626977	A	-	-			-			G	187	100	Glu536Gly
CD630_20270	N-carbamoyl-L-amino acid hydrolase	2339506	G	-	-			-			A	73	100	Gly373Glu
CD630_26670	PTSG-BC	3079815	A	-	-			-			C	165	100	*524Glu
CD630_26270	Hypothetical protein	3034953	C	A	A			A			-	156	100	Gly68Cys
CD630_33790	conjugative transposon protein	3951559	C	-	-			-			A	307	100	Glu63Asp
CD630_12740	<i>topA</i>	1480649	C	-	-			-			T	146	100	Gln386*
CD630_29430	Putative phage replication protein	3422569	T	-	-			-			C	195	100	Asn210Asp
IG	Intergenic region	309208	-	-	-			-			INV	129	41	
IG	Intergenic region	3528736	G	-	-			-			T	94	100	
630Δerm														
CD630_19070	<i>eutG</i>	2209236	G	-	-			A	127	97				Gly252Glu
CD630_35650	GntR family transcriptional regulator	4166495	G	-	-			A	182	100	-			Ala91Val
IG	Intergenic region	2937176	C	-	-			A	173	100	-			
IG	Intergenic region	3005866	T	-	-			G	156	100	-			
IG	Intergenic region	3591103	G	-	-			A	211	100	-			
CD630_12140	<i>spo0A</i>	1413057	-	-	-			AGAATGTA GGAAATAT AG	112	40	-			Insertion
CD630_08260	Ferric uptake regulator	1000995	A	-	-	100	100	G	153	100	-	170	98	Thr41Ala
630														
CD630_32450	<i>prdR</i>	3797112	C	-	T	117	100	-			-			Glu261Lys

CD630_02050	Transcription antiterminator, PTS operon regulator	268934	G	-	T	123	97	-			-			Gly165Cys
630 and 630Δerm														
CD630_2667	PTSG-BC	3080703	C	-	T	75	100	T	141	100	-			Val228Ile
IG	Intergenic region	2203033	A	-	T	105	100	T	174	99				
IG	Intergenic region	4007463			C	10	100	C	62	100				
630 and 630Δerm and 630E														
CD630_11900	acyl-CoA N-acyltransferase	1391850	T	-	C	118	100	C	133	100	C	143	100	Phe133Leu
CD630_13880	pseudo	1607453	INS ¹	-	T	119	94	T	175	94	T	124	88	Thr16fs
Mistake in original sequence														
CD630_17670	<i>gapB</i>	2044514	C	G	G			G			G			
CD630_31561	pseudo	3686535	INS ¹	A	A			A			A			

The table shows the Single Nucleotide Polymorphism (SNP) changes in 630 Δ erm and 630E compared to the reference 630¹² and also the new annotation by Riedel *et al* (The column 'Gene' represents the gene (or intergenic region (IG)) in which the change occurs, the column 'position' indicates the exact nucleotide position of the change.).²² It also contains SNP frequency (F) and genomic coverage (C) as well as the resultant amino acid change (AA). No change from the original 630 annotation¹² is represented by a dash (-).

Table 2. Strains and plasmids used in this study.

Name	Description	Source
Bacterial strains		
<i>E. coli</i> TOP 10	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 ^R) <i>endA1</i> <i>nupG</i>	Invitrogen
<i>E. coli</i> CA434	Conjugation donor	Williams et al. ³⁶
<i>C. difficile</i> 630	Wild-type	Brendan Wren
<i>C. difficile</i> 630Δ <i>erm</i>	Erythromycin sensitive strain of <i>C. difficile</i> 630	Hussain et al. ¹³
<i>C. difficile</i> 630Δ <i>erm</i> (Δ <i>pyrE</i>)	<i>C. difficile</i> 630Δ <i>erm</i> containing a deletion in the <i>pyrE</i> gene	Ng et al. ¹⁹
<i>C. difficile</i> 630E	Erythromycin sensitive strain of <i>C. difficile</i> 630	Lyras et al. ⁸
<i>C. difficile</i> 630Δ <i>erm</i> A+B-	<i>C. difficile</i> 630Δ <i>erm</i> <i>tcdB</i> -1511a::intron <i>ermB</i>	Kuehne et al. ⁹
<i>C. difficile</i> 630E A+B-	<i>C. difficile</i> 630E	Lyras et al. ⁸
<i>C. difficile</i> 630E A+B- CT	<i>C. difficile</i> 630E <i>tcdB</i> -1511a:: intron <i>ermB</i>	This study
<i>C. difficile</i> 630E_ <i>topA</i>	<i>C. difficile</i> 630E	This study
<i>C. difficile</i> 630E_ <i>flgB</i>	<i>C. difficile</i> 630E	This study
<i>C. difficile</i> 630E_CD2667	<i>C. difficile</i> 630E	This study
Plasmids		
pMTL007C-E2: <i>tcdB</i> - 1511a	Clostron plasmid containing retargeted region to <i>tcdB</i> at IS 1511 (antisense oriented) for <i>C.</i> <i>difficile</i> 630Δ <i>erm</i> or 630E	Kuehne et al. ⁹
pMTL-	pMTL-SC7315λ2.3 containing 1,000 bp	This study, based on

SC7315λ2.3:: <i>topA</i>	homology arms to change nucleotide 1480649 from T to C in 630E	Cartman et al. ¹⁵
pMTL- SC7315λ2.3:: <i>flgB</i>	pMTL-SC7315λ2.3 containing 1,156 bp homology arms to reverse the inversion upstream of <i>flgB</i> in 630E	This study, based on Cartman et al. ¹⁵
pMTL- SC7315λ2.3:: <i>CD2667</i>	pMTL-SC7315λ2.3 containing 1,000 bp homology arms to change nucleotide 3079815 from C to A in 630E	This study, based on Cartman et al. ¹⁵

Supporting Information

Figure S1. Southern and Western blot of 630E A+B- CT. A. Southern blot showing the single Clostron insertion to create 630E A+B- CT (mutant 1, 2 and 3). The Clostron plasmid pMTL007C-E2:*tcdB*-1511a was used as a positive control and the parental strain 630E as a negative control. The probe used binds to the intron insertion. B. Western blot using an anti-TcdA antibody to detect production of toxin A in the supernatant of 630, 630 Δ *erm*, 630 Δ *erm* A+B-, 630E, 630E A+B- and 630E A+B- CT after 96 h.

Table S1. Genomic and transcriptomic differences between 630, 630 Δ *erm* and 630E.

SNPs, insertions (INS) and inversion (INV) are shown in this table as well as the corresponding RNAseq coverage. Changes between 630, 630 Δ *erm* and 630E are represented compared to the published genome of 630, the new 630 genome by Riedel *et al.*²² and also the newly published genome of 630 Δ *erm*²³.

See attached excel file: S1_Table

Table S2. Colonisation of hamsters with *C. difficile* strains.

Colonisation of hamsters with 630E, 630E A+B- and 630E A+B- CT was confirmed by plating faecal and caecal samples on CDFA and by extraction genomic DNA and PCR of the toxin A and B genes.

Strain given	Colonisation	630 PCR	<i>tcdA</i>	<i>tcdB</i>	Succumbed to CDI
630E		Caecum	+	+	
630E	Colonised to Day 18	Faecal	+	+	
630E	Colonised to Day 15	Faecal	+	+	
630E		Caecum	+	+	CDI
630E		Caecum	+	+	CDI
630EA+B-		Caecum	+	-	CDI
630EA+B-	Not colonised	Faecal	-	-	
630EA+B-	Not colonised	Faecal	-	-	
630EA+B-	Not colonised	Faecal	-	-	
630EA+B-	Not colonised	Faecal	-	-	

630EA+B-		Caecum	+	-	
630EA+B-		Caecum	+	-	CDI
630EA+B-		Caecum	+	-	
630EA+B-	Colonised to Day 3	Faecal	+	-	
630EA+B-		Caecum	+	-	
630EA+B-		Caecum	+	-	
630E A+B- CT		Caecum	+	CT	CDI
630E A+B- CT		Caecum	+	CT	CDI
630E A+B- CT	Colonised to Day 15	Faecal	+	CT	
630E A+B- CT	Colonised to Day 18	Faecal	+	CT	
630E A+B- CT		Caecum	+	CT	
630E A+B- CT		Caecum	+	CT	

Table S3. Transcriptomic differences between 630, 630Δerm and 630E.

Differentially expressed (DE) genes, comparing RNAseq data from 6 h to 14 h and 24 h in 630, 630Δerm and 630E. The genes are classed into cell factor, cell growth, cell wall, fermentation, membrane transport, amino acid metabolism, carbon metabolism, cofactor metabolism, lipid metabolism, nucleic acid metabolism, mobile elements, motility, operons, regulation, anaerobic respiration, secretion, sporulation, stress, translation, unknown and virulence factors. Downregulated genes are coloured in green and upregulated genes in red.

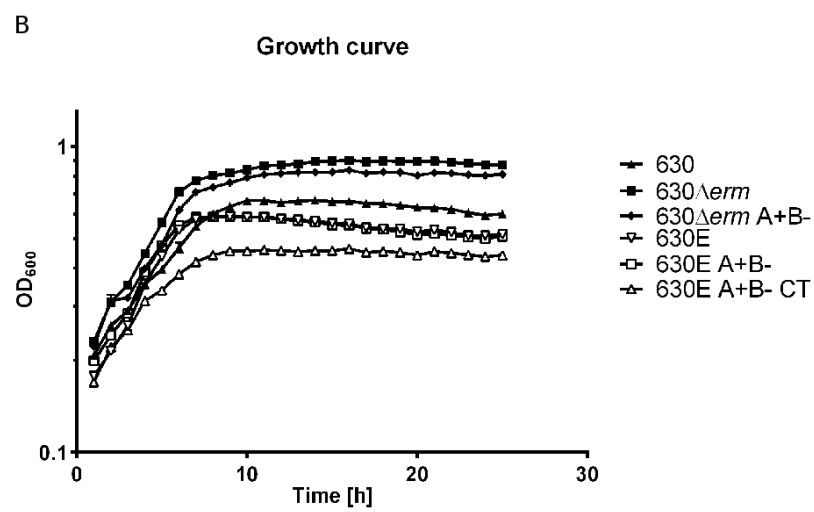
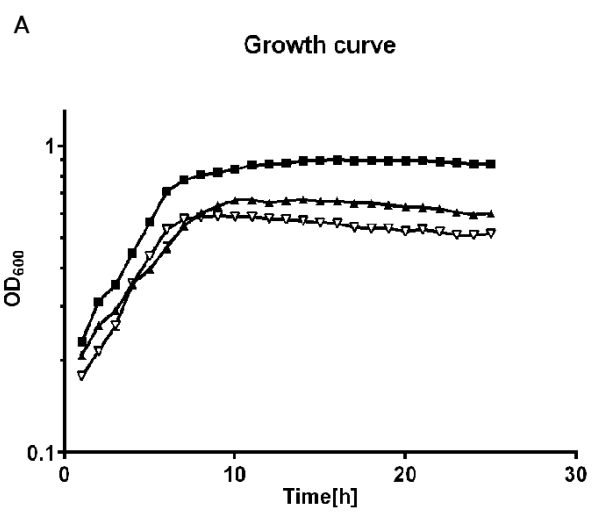
See attached excel file: S3_Table

Table S4. Primers used in this study.

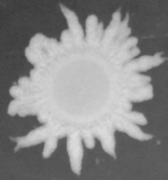
Primer name	Sequence (5'-3')	Explanation or SNP target (where applicable)
topAMC1	GATGCACAACAGGCAAGAAGAGTGC	To screen for chromosomal change
topAMC2b	CCCGATTGTAAAACAACACTAGACCAATTATG	To screen for chromosomal change
flgBMC1b	CTATCAAATACAGATGGAAGTTGTGGTG	To screen for chromosomal change
flgBMC2c	CGAGCATATGATTCTAACGTAGATACATTGAATG	To screen for chromosomal change

CD2667MC1	GTCACCTTATGAGTGAAGTTTGTAAATAAATGTGG	To screen for chromosomal change
CD2667MC2b	GCAAGAGCTGCTGCTGGAAGAC	To screen for chromosomal change
JRP3441 ⁸	GTTACCAGGAATACAACCAGAC	To test for interrupted <i>tcdB</i>
JRP2839 ⁸	CGGCCAGCCTCGCAGAGCAG	To test for interrupted <i>tcdB</i>
JRP3442 ⁸	GCACTTGCTTGATCAAAGCTCC	<i>toxA</i> specific PCR, positive for all strains (630E, 630E A+B-, 630E A+B-CT)
JRP2342 ⁸	CCGGAATTCGCTCTATTGGACTAGACCGTTG	<i>toxA</i> specific PCR, positive for all strains (630E, 630E A+B-, 630E A+B-CT)
Cdi-tcdB-F1 ⁹	TGATAGTATAATGGCTGAAGCTAATGCAGATAATGG	To test for <i>tcdB</i> ClosTron insertion
Cdi-tcdB-R1 ⁹	CTTGCATCGTCAAATGACC ATAAGCTAGCC	To test for <i>tcdB</i> ClosTron insertion
268934_F	TGTCAAGTGAATTAGAAAAGAAACCA	268934
268934_R	AAGTGAGCCGTGTTTTGAAAA	268934
309208_F	GCCAGTTGCCAAAAAGAGTC	309208
309208_R	GGCATAGCATCATTTAGTGTTTC	309208
MCSNP11	CGGGAAAAACAGCTGCTTTTAGTATCCC	933139
MCSNP12	CCTCTTGCTTGTAATCTCCTACCAATTC	933139
10000995_F	GATGAAGAAGTTGTTTGGCAAT	10000995
10000995_R	CCTACTTGGCTACACCTTTTACA	10000995
1391850_F	TCTGTCATTTGGAAAGGATGAA	1391850
1391850_R	TCTGTACTTGCTTTTGATATACTTGGA	1391850
Spo0A_F1	GGCATAGCTAAGGATGGAATTG	1413057
Spo0A_R1	GGAGTAGAGGAAAAGTTGACACAA	1413057
1480649_F	GCTTCAACAAGAAGGAGCAAA	1480649
1480649_R	TGCTGGTGGTTGTGTAATAATG	1480649
1607453_F	TTGAAGGTGTAACCTCAGTTGTAGG	1607453
1607453_R	TCCAAATAAAAGTCTATGAAAATGAA	1607453
1626977_F	TGGTGGTAGCAAAAACGAAA	1626977
1626977_R	TGCCATTGAATTTGTTGCAG	1626977
2044514_F	CATACTAAATGAGGGGTAAAATAAAGA	2044514
2044514_R	TTTTTCTGCCTTTTCTCTTTGTG	2044514
MCSNP1	GTTCTGTAATACCTTTTTCTTTAGCTATTTTAATTGC	2203033
MCSNP2	GTTACCGATATTATAGGCAAAACTGCCC	2203033
2209236_F	CCATTAGTGAGTGATGATTTACTTCC	2209236
2209236_R	GCAAGTTTTGCTATTTCTCTTTCTT	2209236
2320410_F	GTCACTGGTAGGAATTAATCTAACG	2320410
2320410_R	TGCTTTCACAAATGCTTTTCG	2320410
MCSNP13	GAAAGATGAATTTCTATCCATCTTCATCAAATGTGG	2339506
MCSNP14	GGGGCTGTTGACCTTGGACCC	2339506

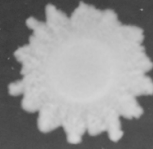
MCSNP3	CCTTTGATGTCTAGTTAATTTCTTCACTTATTTTAAGC	2937176
MCSNP4	GGAAAACCAGCAAAAGCTTGTATTATGATTCCC	2937176
3005866_F	AATATAATCCCAACCTTCCAAA	3005866
3005866_R	TTTGTTGAAGATTAGTGGTGATTG	3005866
MCSNP5	GTATATTTTTCTCTAGCTTTATCTCCATCAGGG	3034953
MCSNP6	GGAAAGGATAAACCCAGGTATAGTGCC	3034953
3079815_F	CCCGCTTTTACTTCATCTCC	3079815
3079815_R	GCATCAGAGATTTTGATTGCTTT	3079815
MCSNP7	CCTGCTACAAATTTTTCTTTCTGGC	3080703
MCSNP8	CTGCTTATCTTTATAAAAAGTTTATAAAAATTGAATTACCTC	3080703
MCSNP15	CTTCATATTGAGTGAAAGTCTGATTGAAGTTAGC	3422569
MCSNP16	CTCAACCGTGTGCCGTTTTCCCG	3422569
3526888_F	CTCTTTCCTGCATTCCCAAG	3526888
3526888_R	TTGTTGAGCAGATATAAAATCCCA	3526888
MCSNP17	CAATCTATTCAAAGATAAACTATAGTACTTCTTCTAC	3528736
MCSNP18	CCTACTCCTTTAGGTGTGAGATGG	3528736
3591103_F	GGCACTAGCTGCTCCTAATAAA	3591103
3591103_R	CCATATACCCCTATCCCTCCTT	3591103
3686535_F	TCTTCCAAGCTTTACCTGTTTG	3686535
3686535_R	GCTCTGTCCAGTTAATTG	3686535
3797112_F	TGCTCCTGTAATGCACCTG	3797112
3797112_R	CTGTAAAATACAAGTCACTCATTCCAA	3797112
MCSNP19	CCGTTCCAGACTGTTCAATGCTCC	3951559
MCSNP20	CCTAAGTAGTAGTTACTGGCAACAGCAC	3951559
MCSNP9	GCACCCTTAATAACTTGACCAGTAAAAAGG	4007463
MCSNP10	CGCCCGAAGCCGATTATCTAACC	4007463
4166495_F	GCATCAAGTAAGTATTTATGCTCTTCA	4166495
4166495_R	TGAACTTGGATAATTACAAGCCATT	4166495



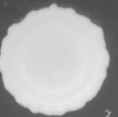
A
630



B
630 Δ erm

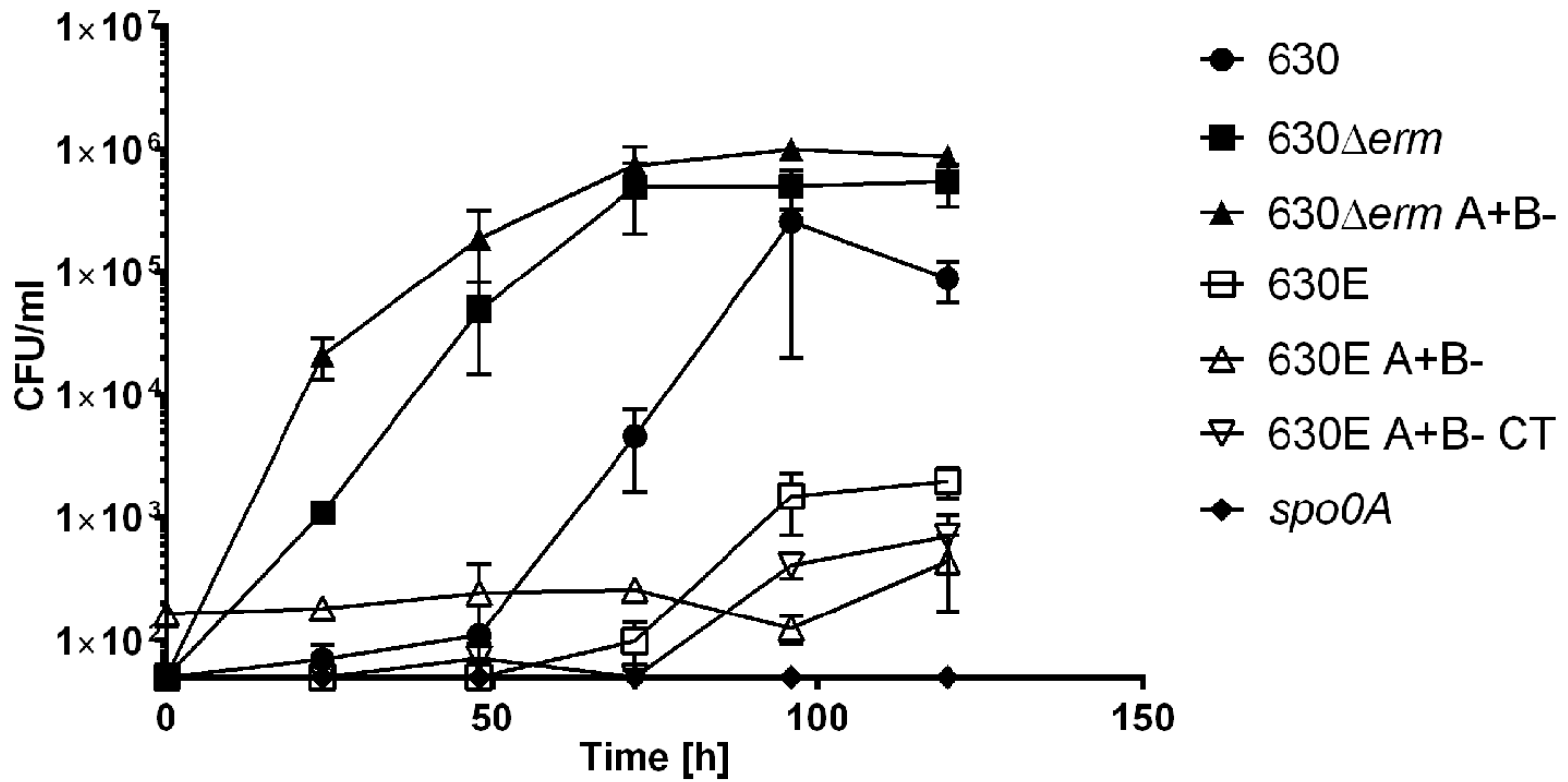


C
630E



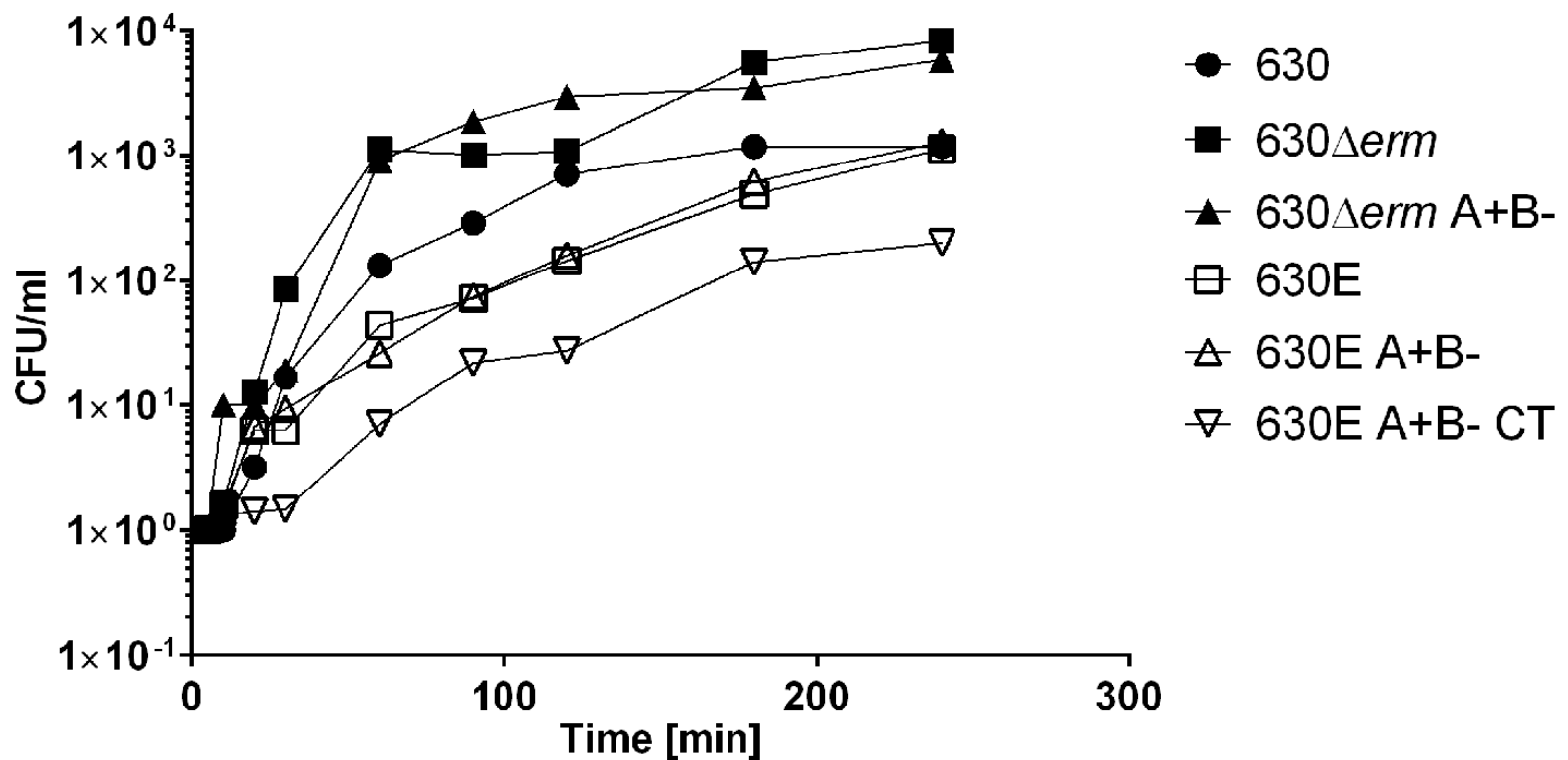
A

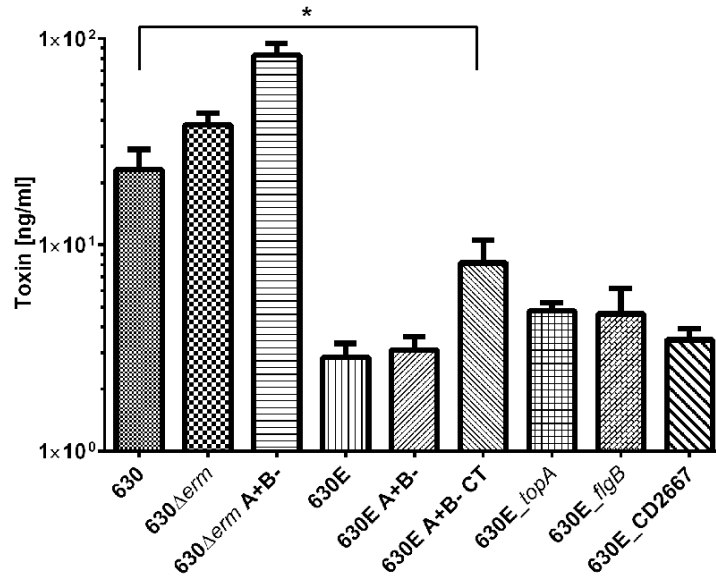
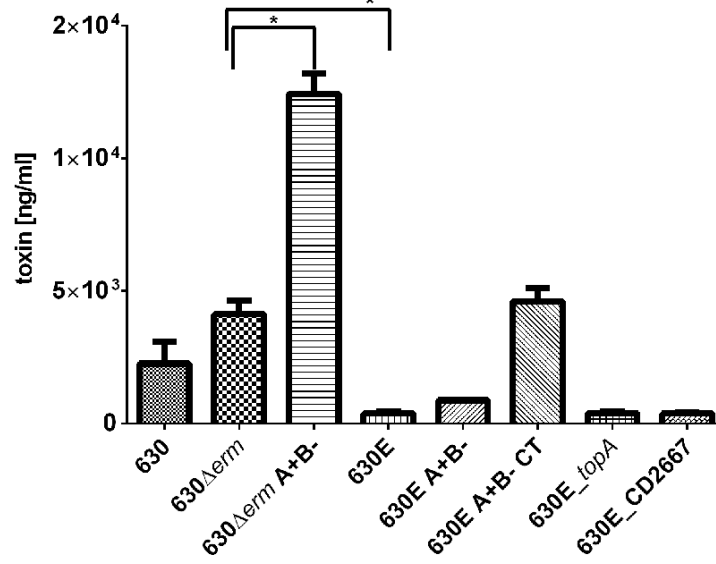
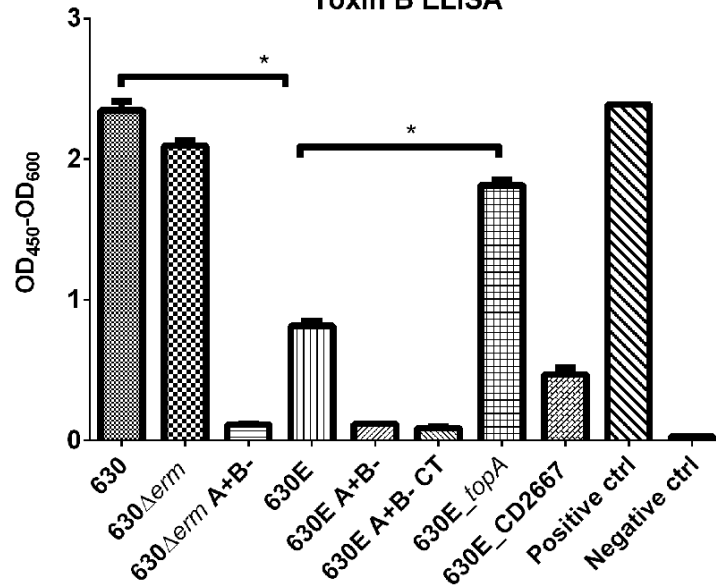
Sporulation



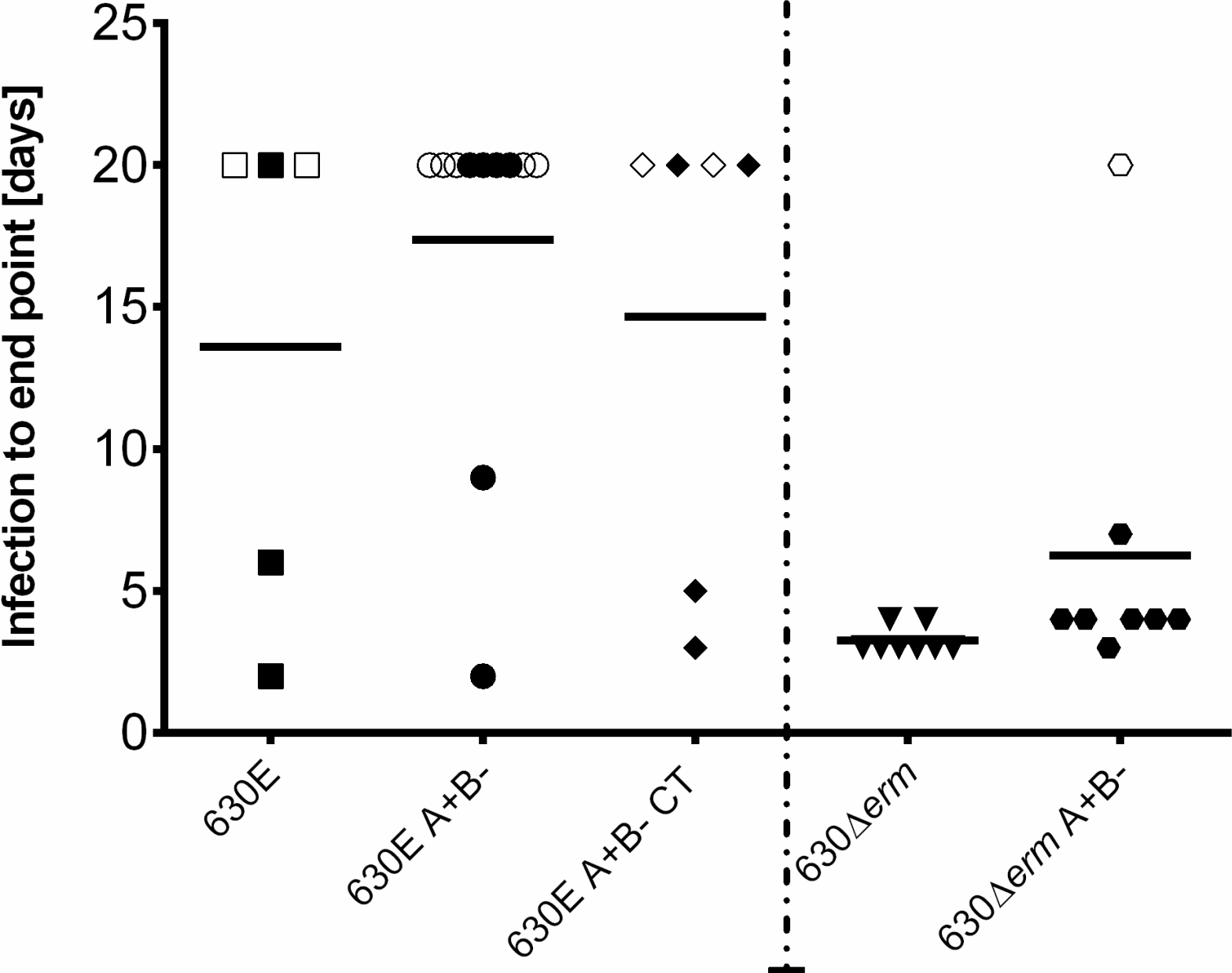
B

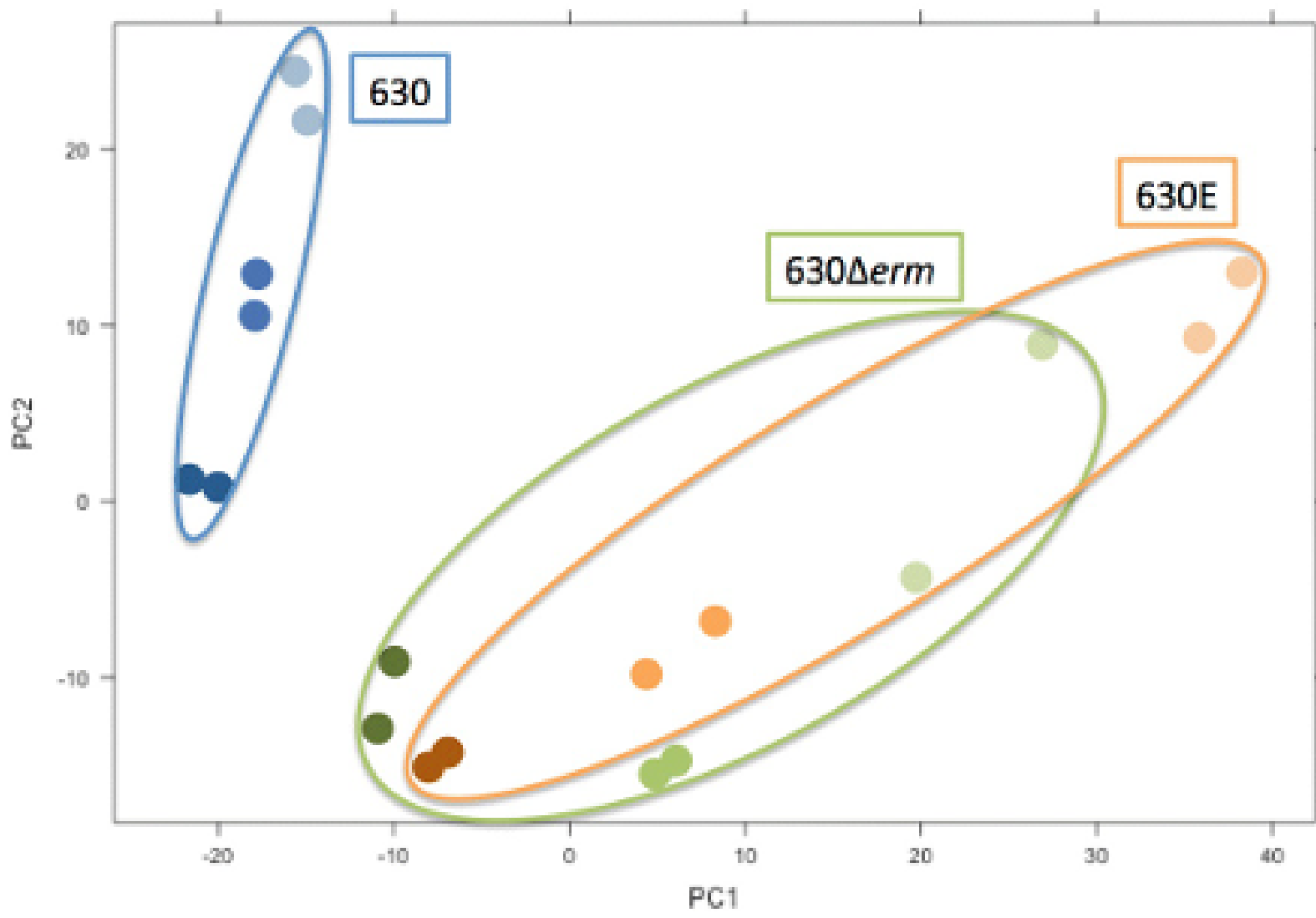
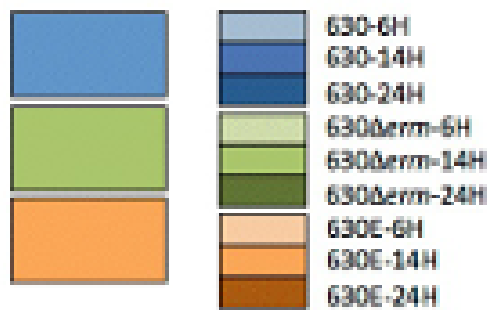
Germination

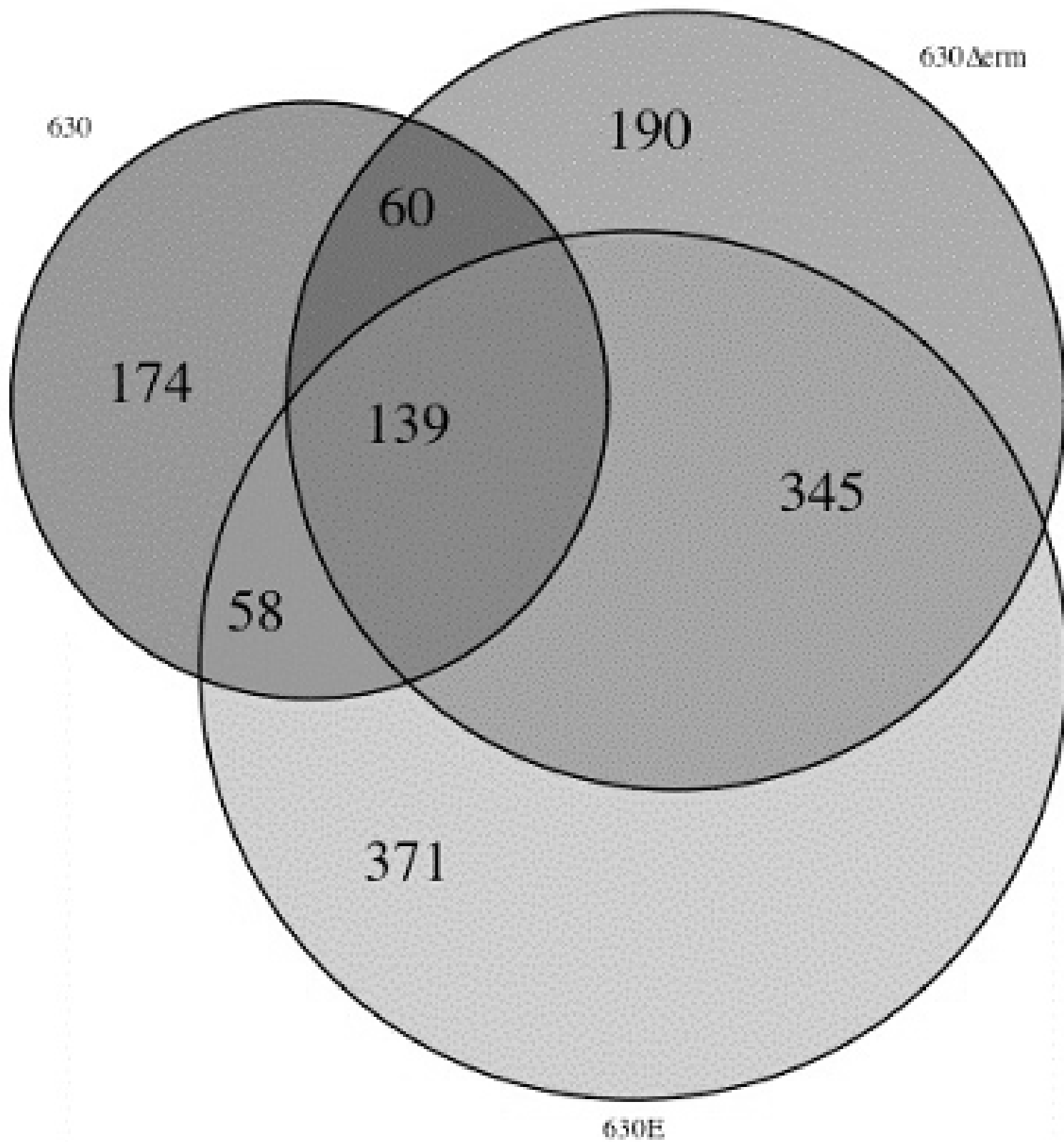


A**72 h Toxin A and B ELISA****B****Toxin A ELISA****C****Toxin B ELISA**

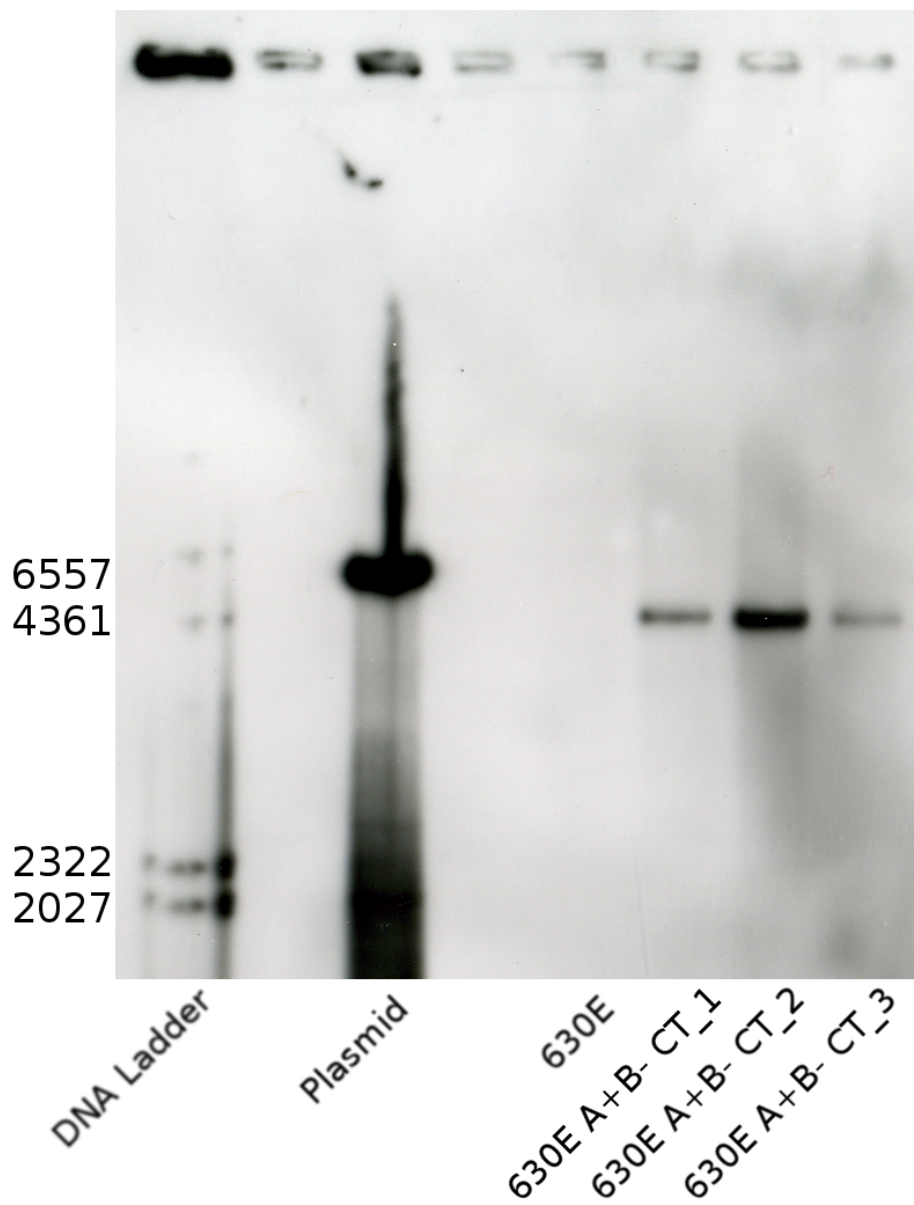
Hamster Infection Model



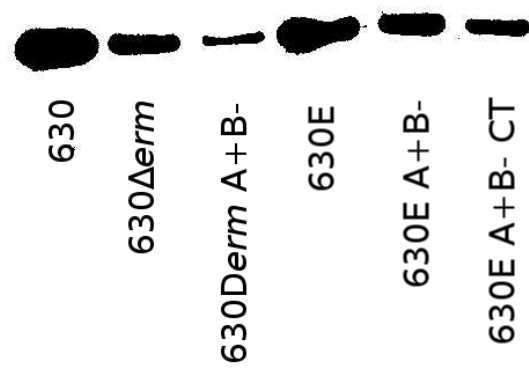




A



B



SNP between strains	Name
630E	
CD630_07610	
CD630_14040	
CD630_20270	
CD630_26670	PTSG-BC
CD630_26270	
CD630_33790	
CD630_12740	topA
CD630_29430	
IG	
IG	
CD630_20102	(pseudo)
630Δerm	
CD630_30360	
CD630_19070	eutG
CD630_35650	
IG	
IG	
IG	
CD630_12140	spo0A
CD630_08260	
CD630	
CD630_32450	prdR
CD630_02050	
CD630 and 630Δerm	
CD630_26670	PTSG-BC
IG	
IG	
CD630, 630Δerm and 630E	
CD630_22630	prsA
CD630_30890	
Identical for all strains	
CD630_11900	
CD630_13880	(pseudo)
Mistake in original sequence	

CD630_17670	gapB
CD630_31561	(pseudo)

Product

Putative ATP-dependent RNA helicase
Putative oligopeptide transporter
N-carbamoyl-L-amino acid hydrolase
PTS system, glucose-specific IIBC component
Conserved hypothetical protein
Putative conjugative transposon protein
DNA topoisomerase I
Putative phage replication protein
Upstream of flgB
-
Fragment of transcriptional regulator

Transporter, Major Facilitator Superfamily (MFS)
Ethanolamine iron-dependent Alcohol dehydrogenase
Transcriptional regulator, GntR family
Upstream of Sodium:glutamate/aspartate symporter fam
Upstream of Lipoprotein signal peptidase
Upstream Transcriptional regulator, GntR family
Stage 0 sporulation protein A
Ferric uptake regulator (perR)

Transcriptional regulator, sigma-54-dependent
Transcription antiterminator, PTS operon regulator

PTS system, glucose-specific IIBC component
Upstream of phage regulator

Peptidyl-prolyl cis-trans isomerase, PpiC-type
PTS system, glucose-like IIBC component

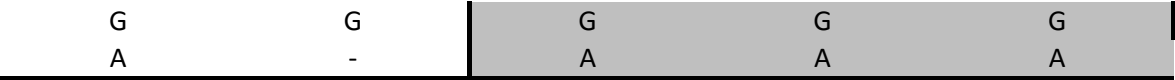
Putative acyl-CoA N-acyltransferase
Fragment of transcriptional regulator

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)
Fragment of conserved hypothetical protein

Metabolic Pathway (KEEG, Microscope)	Position	630 reference
RNA degradation	933,139	G
-	1,626,977	A
Purine metabolism	2,339,506	G
Phosphotransferase system (PTS)	3,079,815	A
-	3,034,953	C
-	3,951,559	C
-	1,480,649	C
-	3,422,569	T
-	309,208	-
-	3,528,736	G
-	2,320,410	A
-	3,526,888	T
Ethanolamine utilization	2,209,236	G
Sporulation	4,166,495	G
-	2,937,176	C
-	3,005,866	T
-	3,591,103	G
Sporulation	1,413,057	INS (18)*
superoxide radicals degradation	1,000,995	A
Proline degradation	3,797,112	C
Phosphotransferase system (PTS)	268,934	G
Phosphotransferase system (PTS)	3,080,703	C
-	2,203,033	A
-	4,007,463	-
-	2,623,495	G
Phosphotransferase system (PTS)	3,590,230	T
-	1,391,850	T
-	1,607,453	INS (1)

Glycolysis / Gluconeogenesis	2,044,514	C
-	3,686,535	INS (1)

Complete genome		DNA-Seq		
630 GenomeA	630 GenomeA	630wt	630Δerm (M)	630E (R)
<i>Riedel et al</i>	<i>Van Eijk et al</i>			
-	-	-	-	T
-	-	-	-	G
-	-	-	-	A
-	-	-	-	C
A	A	A	A	-
-	-	-	-	A
-	-	-	-	T
-	-	-	-	C
-	-	-	-	INV (150 bases)
-	-	-	-	T
-	-	-	-	C
-	-	-	A	-
-	A	-	A	-
-	-	-	-	A
-	A	-	A	-
-	G	-	G	-
-	A	-	A	-
-	AATGTAGGAAATAT	-	AATGTAGGAAATAT	-
-	G	-	G	-
-	-	T	-	-
-	-	T	-	-
-	T	T	T	C
-	T	T	T	-
-	-	C	C	-
-	-	-	-	T
-	-	G	-	-
-	C	C	C	C
-	T	T	T	T



-

same base as reference



DNA-Seq more than 70%

DNA-Seq less than 70%

RNA-Seq Mapping (coverage)			RNA-Seq* 630wt	
630wt	630Δerm (M)	630E (R)	14h / 6h	24h / 6h
G (42) / T (7)	G (27) / T (2)	T (3) / G (1)	1.00	1.00
A (58) G (46)	A (40) G (23)	G (3)	1.00	1.00
G (31) / A (2)	G (27) / A (6)	A (15) / G (5)	1.00	1.00
A (88) / C (49)	A (80)	C (1)	1.00	1.00
A (50) / C (25)	A (60) / C (40)	C (46) / A (6)	1.00	1.00
C (3)		C (1)	4.43	1.00
T (205) / C (136)	T (93) / C (24)	T (49) / C (1)	1.00	1.00
C (9) / T (2)	T (1)	C (2)	1.00	3.48
			1.00	1.00
			1.00	1.00
T (12)			1.00	1.00
G (3)	G (2)		1.00	5.91
A (47) / G (38)	G (17) / A (7)	G (5) / A (1)	1.00	1.00
			8.34	3.00
			1.00	1.00
			4.19	1.73
			1.00	1.00
G (250) / A (21)	G (>1500) / A (300)	G (268) / A (250)	7.12	1.00
C (3) / T (0)	T (3) / C (1)		1.00	1.00
T (1)			0.11	0.02
T (7)	T (5)		1.00	1.00
			1.00	1.00
			1.00	1.00
G (2)		G (1)	1.00	1.00
T (140) / G (10)	T (45) / G (7)	T (24)	1.00	1.00
C (34)	C (11)	C (5)	1.00	1.00
			1.00	1.00

G (43)	G (8)	G (32)	0.26	1.00
			1.00	1.00

RNA-Seq mapping = DNA-Seq	Difference ex
RNA-Seq mapping \neq DNA-Seq	INV 150 bp invers
<i>*RNA-Seq FC \geq 3, IG: gene expression after the mutation</i>	INS (18)* Identical in 6:

RNA-Seq* 630Δerm (M)		RNA-Seq* 630E (R)		Comments
14h / 6h	24h / 6h	14h / 6h	24h / 6h	
1.00	1.00	0.12	1.00	OK
1.00	1.00	1.00	0.27	OK
1.00	1.00	1.00	3.13	OK
0.05	0.03	0.32	0.01	OK
1.00	1.00	1.00	1.00	OK
4.5	1.00	5.28	1.00	RNA-Seq ≠ DNA-Seq
1.00	4.47	1.00	1.00	RNA-Seq ≠ DNA-Seq
1.00	1.00	1.00	5.10	OK with CD630wt
1.00	1.00	1.00	1.00	No expression
1.00	1.00	1.00	1.00	CD630_20101 ermC
1.00	1.00	1.00	1.00	Low expression
1.00	1.00	1.00	1.00	RNA-Seq ≠ DNA-Seq
1.00	1.00	1.00	1.00	RNA-Seq ≠ DNA-Seq
7.35	5.68	5.03	3.39	No expression
1.00	1.00	1.00	1.00	No expression
1.00	3.00	1.00	1.00	Gene after : CD630_30900 TreR
1.00	1.00	0.17	0.23	No expression
10.45	1.00	5.42	1.00	RNA-Seq ≠ DNA-Seq
1.00	1.00	1.00	1.00	RNA-Seq ≠ DNA-Seq
0.04	0.08	0.01	0.01	Low expression
0.05	0.03	0.32	0.01	OK
1.00	1.00	1.00	1.00	No expression
1.00	1.00	1.00	1.00	No expression
1.00	1.00	1.00	1.00	RNA-Seq ≠ DNA-Seq
1.00	13.07	1.00	7.38	RNA-Seq ≠ DNA-Seq
1.00	1.00	1.00	1.00	OK
1.00	1.00	1.00	1.00	No expression

1.00	1.00	1.00	3.54	OK
1.00	1.00	1.00	1.00	No expression

pression in agreement with SNP
ion promoter region of flgB
30 GenomeA and 630E (R) strains

—
┌
└
—
—
┌
┌
┌
—
—
—
—
—

Class	Gene ID	Name	Product	ES0 14h / GH	ES0 24h / GH
Cell Factor	CD610_0001		Putative peptidase, M20 family	0.23	
Cell Factor	CD610_06490		Putative peptidase, S9 family	1	5.88
Cell Factor	CD610_06600		Putative peptidase, S9 family	1	6.92
Cell Factor	CD610_07100		Putative iron-sulfur protein	1	3.32
Cell Factor	CD610_07110		Putative radical SAM superfamily protein	1	0.33
Cell Factor	CD610_07130		Putative radical SAM superfamily protein	1	0.28
Cell Factor	CD610_07130		Putative biotin/ipoate-protein ligase	1	0.31
Cell Factor	CD610_12770		Putative acyltransferase	1	3.08
Cell Factor	CD610_15600		Conserved hypothetical protein	1	4.12
Cell Factor	CD610_15610		Putative dehydrogenase	1	4.13
Cell Factor	CD610_16230		Conserved hypothetical protein	1	1.71
Cell Factor	CD610_16786.gpp		Pyruvate decarboxylase	1	4.96
Cell Factor	CD610_17630		Putative hydrolase, NUDIX family	1	5.13
Cell Factor	CD610_17640		Putative ATP7P-binding protein	1	1.85
Cell Factor	CD610_21552.nob		Biotin synthase	1	3.97
Cell Factor	CD610_21552.nob		Thiamine biosynthesis protein ThiH	1	6.04
Cell Factor	CD610_21731.cyt7		Carbonic anhydrase	1	0.29
Cell Factor	CD610_21850		Putative NAD(P)-binding oxidoreductase	1	0.25
Cell Factor	CD610_30390		Putative pyridine nucleotide-deulphohide oxidoreductase, class I	1	3.86
Cell Wall	CD610_02410		Phosphate threonine kinase	1	0.03
Cell Wall	CD610_02420		Phosphotransferase N-acetylglucosamine catalysis	1	0.04
Cell Wall	CD610_02430		Methyltransferase	1	0.04
Cell Wall	CD610_02440		Putative GDP-glycerol-3-phosphoethanolphosphate glycerophosphotransferase	1	0.06
Cell Wall	CD610_03060		Conserved hypothetical protein	1	4.41
Cell Wall	CD610_07860		Putative membrane protein	1	4.13
Cell Wall	CD610_08330		Putative membrane protein	1	1.7
Cell Wall	CD610_10290		Putative cell wall anchored protein	1	8.05
Cell Wall	CD610_10335.nmb		UDP-N-acetylglucosamine 2-epimerase (UDP-G2NAc-2-epimerase)	1	4.23
Cell Wall	CD610_10365.caps15		Putative N-acetylglucosamoyl-L-alanine amidase, autolysin	1	0.33
Cell Wall	CD610_15850		Putative N-acetylglucosamine kinase NAGK-like	1	4.51
Cell Wall	CD610_16450		Putative membrane protein	1	0.25
Cell Wall	CD610_16460		Putative peptidase, M19 family	1	0.23
Cell Wall	CD610_24580		Putative membrane protein	1	0.21
Fermentation	CD610_10544.dct2		Butyryl-CoA dehydrogenase	1	0.27
Fermentation	CD610_10550.effB		Electron transfer flavoprotein subunit beta	1	0.25
Fermentation	CD610_10550.effB		Electron transfer flavoprotein subunit alpha	1	0.28
Fermentation	CD610_10570.cit2		3-hydroxybutyryl-CoA dehydrogenase (Citronase)	1	0.31
Fermentation	CD610_20622.nme		Aldehyde-acyl dehydrogenase	1	0.26
Membrane Transport	CD610_03610		ABC-type transport system, multidrug-family ATP-binding protein	0.31	0.25
Membrane Transport	CD610_03620		Putative radical SAM family protein	0.33	0.26
Membrane Transport	CD610_04930		PTS system, mannose-specific IIC component	1	0.32
Membrane Transport	CD610_06410		ABC-type transport system, tauribiotic/multidrug-family ATP-binding protein	1	0.28
Membrane Transport	CD610_06450.effB		ABC-type transport system, lactibiotic family IIC component	1	0.32
Membrane Transport	CD610_08540.pqc		ABC-type transport system, oligopeptide-family permease	1	0.32
Membrane Transport	CD610_10550.caps15		ABC-type transport system, oligopeptide-family extracellular solute-binding protein	1	0.27
Membrane Transport	CD610_16290		ABC-type transport system, mannose-family permease	1	0.77
Membrane Transport	CD610_16280		ABC-type transport system, multidrug-family ATP-binding protein	1	4.87
Membrane Transport	CD610_16840.cmt		ABC-type transport system, akanaevorax-family extracellular solute-binding protein	1	4.33
Membrane Transport	CD610_15230		Putative transporter	1	3.3
Membrane Transport	CD610_17600		Putative permease	1	0.16
Membrane Transport	CD610_21020		Putative Hcp1/Hcp1 antipporter	1	0.04
Membrane Transport	CD610_21070		Xanthine/uracil/thiamine/acarboxate permease family protein	1	3.62
Membrane Transport	CD610_24880		PTS system, fructose-like IIC component	1	0.25
Membrane Transport	CD610_24890		Transcription antiterminator, PTS operon regulator	1	4.99
Membrane Transport	CD610_28700.kglT1		3-keto-3-deoxygluconate permease I	1	0.28
Membrane Transport	CD610_28830.a		PTS system, cellobiose-specific IIC component	1	0.3
Membrane Transport	CD610_28840.a		PTS system, cellobiose-specific IIC component	1	0.29
Membrane Transport	CD610_30320		Putative alpha-glucosylase	0.3	1
Membrane Transport	CD610_30330		PTS system, mannose-specific IIC component	0.29	0.3
Membrane Transport	CD610_30340		PTS system, mannose-specific IIC component	0.3	0.29
Membrane Transport	CD610_30350		PTS system, mannose-specific IIC component	1	0.31
Membrane Transport	CD610_30360		Transcription antiterminator, PTS operon regulator	0.31	1
Membrane Transport	CD610_30760.pgf		Tageline 4-phosphatase	1	0.24
Membrane Transport	CD610_34410		Putative pyridoxal phosphate-dependent transferase (PLP-dependent transferase)	1	3.83
Membrane Transport	CD610_34420		Putative peptidase, M24 family	1	3.97
Membrane Transport	CD610_34430		PTS system, lactulose/cellobiose-family IIC component	1	0.22
Membrane Transport	CD610_34450		PTS system, lactulose/cellobiose-family IIC component	1	3.67
Membrane Transport	CD610_35250		ABC-type transport system, iron-family permease	1	0.08
Membrane Transport	CD610_35260		ABC-type transport system, iron-family permease	1	0.06
Membrane Transport	CD610_35270		ABC-type transport system, iron-family ATP-binding protein	1	0.07
Metabolism Amino Acid	CD610_02650		Putative aminocarbonylase	1	0.01
Metabolism Amino Acid	CD610_14810		Putative tyrosine carbonylase	1	0.31
Metabolism Amino Acid	CD610_15470.kat		ATP phosphoribosyltransferase regulatory subunit	1	0.23
Metabolism Amino Acid	CD610_15480.kug		ATP phosphoribosyltransferase (ATP-PRTase) (ATP-PR1)	1	0.09
Metabolism Amino Acid	CD610_15490.kat		Histidinol phosphate aminotransferase (imidazole acetyl-phosphate transaminase)	1	0.17
Metabolism Amino Acid	CD610_15500.nh		Imidazole acetyl-phosphate dehydrogenase	1	0.18
Metabolism Amino Acid	CD610_15510.nh		Imidazole glycerol phosphate synthase subunit Hnf	1	0.15
Metabolism Amino Acid	CD610_15520.nh		1-5-phosphoribosyl-5-S-phosphoribosylaminomethyladenosine/imidazole 4-carbamoyl ion	1	0.12
Metabolism Amino Acid	CD610_15530.nh		Imidazole glycerol phosphate synthase subunit Hnf	1	0.12
Metabolism Amino Acid	CD610_15540		Histidine biosynthesis bifunctional protein HisH [Includes: Phosphoribosyl-AMP cyclitolase Ph	1	0.2
Metabolism Amino Acid	CD610_15550.nhm2		Homoserine dehydrogenase	1	0.73
Metabolism Amino Acid	CD610_17640		Putative phenylalanyl-RNA synthetase beta chain	1	0.28
Metabolism Amino Acid	CD610_20530.a		Diaminopimellic decarboxylase	1	0.27
Metabolism Amino Acid	CD610_20540.c		Asparaginase	1	0.25
Metabolism Amino Acid	CD610_21180.cnc		Threonine synthase	1	3.35
Metabolism Amino Acid	CD610_22520		Glutamate 2,3-aminomutase	1	3.08
Metabolism Amino Acid	CD610_29020		Putative pyridoxal phosphate-dependent transferase	1	0.17
Metabolism Amino Acid	CD610_30990		Putative aminohydrolase, M202 family	0.04	0.31
Metabolism Amino Acid	CD610_31000		Putative conjugative transposon carrier, DccC family	0.33	0.03
Metabolism Amino Acid	CD610_21870		Alpha-hydroxy acid oxidoreductase, FAD-dependent	1	0.17
Metabolism Carbon	CD610_24100.ppkK		Pyruvate phosphate dikinase (ATP-pyruvate, phosphate phosphotransferase)	0.31	1
Metabolism Carbon	CD610_24100.ppkR		Tartrate semialdehyde reductase	1	0.24
Metabolism Carbon	CD610_36140		Conserved hypothetical protein	1	0.21
Metabolism Cofactor	CD610_21690		Putative iron-sulfur binding protein	0.32	0.21
Metabolism Lipids	CD610_19070.a		Ethanolamine iron-dependent alcohol dehydrogenase	1	0.16
Metabolism Lipids	CD610_19080.a		Ethanolamine carbonylase structural protein, BMC Family	1	3.56
Metabolism Lipids	CD610_19090.pqc		Ethanolamine utilization protein, OTPase family	1	3.1
Metabolism Lipids	CD610_19120.a		Ethanolamine reductase factor for ammonia lyase eutC	1	0.88
Metabolism Lipids	CD610_19130.a		Ethanolamine ammonia lyase large subunit	1	12.31
Metabolism Lipids	CD610_19140.a		Ethanolamine ammonia lyase small subunit	1	28.68
Metabolism Lipids	CD610_19150.a		Ethanolamine carbonylase structural protein, BMC Family	1	17.38
Metabolism Lipids	CD610_19160.a		Ethanolamine carbonylase structural protein, BMC Family	1	21.53
Metabolism Lipids	CD610_19170.a		Ethanolamine acetylaldehyde reductase	1	0.16
Metabolism Lipids	CD610_19180.a		Ethanolamine carbonylase structural protein, BMC Family	1	14.47
Metabolism Lipids	CD610_19190.a		Ethanolamine acid cobaltation adenosyltransferase	1	21.74
Metabolism Lipids	CD610_19200.a		Putative phosphotransacetylase	1	22.6
Metabolism Lipids	CD610_19210		Putative ethanolamine utilization protein	1	15.57
Metabolism Lipids	CD610_19220.a		Ethanolamine carbonylase structural protein, BMC Family	1	13.82
Metabolism Lipids	CD610_19230		Ethanolamine carbonylase structural protein, BMC Family	1	12.09
Metabolism Lipids	CD610_19240.a		Ethanolamine carbonylase structural protein, BMC Family	1	0.72
Metabolism Lipids	CD610_19250.a		Ethanolamine utilization protein, Cupin barrels family	1	11.59
Metabolism Nucleic Acid	CD610_06200		Putative replicative helicase DnaB-like	1	7.06
Metabolism Nucleic Acid	CD610_12720		Putative magnesium chelatase	1	0.33
Metabolism Nucleic Acid	CD610_17760		Putative replication associated recombination protein A Rank	1	3.08
Metabolism Nucleic Acid	CD610_20730.nha		Xanthine dehydrogenase, molybdenum binding subunit	1	0.31
Metabolism Nucleic Acid	CD610_24750		Putative membrane protein	1	3.09
Metabolism Nucleic Acid	CD610_25210		Putative hydrolase	1	0.31
Metabolism Nucleic Acid	CD610_25240.d		Nucleoside-methylethyladenyltransferase	1	0.25
Metabolism Nucleic Acid	CD610_38420		Putative selenocysteine synthase	1	0.24
Metabolism Nucleic Acid	CD610_38430		Oligonucleotase	1	0.31
Mobile Element	CD610_03730		Putative membrane protein Tn916-like, Ctn1-Ctn17	3.14	1
Mobile Element	CD610_06200		Putative conjugative transposon protein Tn1549-like, Ctn2-Ctn21	1	0.28
Mobile Element	CD610_06210		Putative conjugative transposon protein Tn1579, Ctn3-Ctn8	1	0.22
Mobile Element	CD610_06250		Putative phage protein	1	8.03
Mobile Element	CD610_06300		Putative phage protein	1	3.1
Mobile Element	CD610_09380		Putative phage protein	1	3.16
Mobile Element	CD610_13630		Putative phage Xdk-like protein	1	3.16
Mobile Element	CD610_15640		Putative phage Xdk-like protein	1	3.88
Mobile Element	CD610_13650		Putative phage Xdk-like protein	1	4.87
Mobile Element	CD610_29030		Putative phage protein	1	3.16
Mobile Element	CD610_29120		Conserved hypothetical protein	1	3.76
Mobile Element	CD610_29130		Putative phage protein	1	3.1
Mobile Element	CD610_29140		Putative phage protein	1	3.39
Mobile Element	CD610_29190		Putative phage protein	1	4.13
Mobile Element	CD610_31800		Putative cell wall hydrolase Tn916-like, Ctn1-Ctn17	1	1.48
Mobility	CD610_02280.n		Flagellar motor switch protein FlhK	1	0.24
Mobility	CD610_02310.n		Flagellar assembly factor FlhW	1	0.03
Mobility	CD610_17940		Conserved hypothetical protein	1	3.11
OPRONS (Automatic)	CD610_17600		Putative iron-sulfur protein	1	0.27
OPRONS (Automatic)	CD610_16690		Putative endonuclease ribonuclease Tn1549-like, Ctn2-Ctn20	1	0.33
OPRONS (Automatic)	CD610_23970		Putative exonuclease	1	3.11
OPRONS (Automatic)	CD610_35030.a		BirA bifunctional protein [includes: biotin operon repressor and both--[acyl]-CoA-carboxylase]	1	3.31
OPRONS (Automatic)	CD610_35750		Putative sodium/calcium symporter	1	0.16
Regulators	CD610_02920		Transcriptional regulator, HTH-type	1	3.99
Regulators	CD610_06170		Putative membrane protein	1	0.18
Regulators	CD610_06270		Putative transcriptional regulator, activator Mor	1	5.63
Regulators	CD610_12460		Two-component response regulator	1	3.35
Regulators	CD610_12700		Two-component sensor histidine kinase	1	3.35
Regulators	CD610_17250		Transcriptional regulator, ArcA family	1	3.49
Regulators	CD610_18410		Two-component sensor histidine kinase	1	3.24
Regulators	CD610_23850		Putative oligonucleotide cyclase signaling protein	1	4.63
Regulators	CD610_23810.a		ABC-type transport system, phosphate-specific ATP-binding	1	4.23
Respiration Anaerobic	CD610_34760.a		ATP synthase translocator protein AtpZ	1	4.24
Spore/Spores	CD610_12130.a		Stage IV sporulation protein B, peptidase S55 family	1	3.68
Stress	CD610_03440		Putative beta-lactamase-like protein	1	0.21
Stress	CD610_11740		Putative beta-lactamase-inhibitor protein II	1	3.54
Stress	CD610_10250.a		Two-component sensor histidine kinase	1	3.61
UNKNOWN	CD610_11560		Conserved hypothetical protein	1	0.02
UNKNOWN	CD610_15830		Conserved hypothetical protein	1	4.18
UNKNOWN	CD610_15831		Conserved hypothetical protein	1	0.21
UNKNOWN	CD610_17170		Conserved hypothetical protein, DUF917 family	1	0.18
UNKNOWN	CD610_17170		Conserved hypothetical protein	1	0.2
UNKNOWN	CD610_20620		Conserved hypothetical protein	1	3.12
UNKNOWN	CD610_21570		Conserved hypothetical protein	1	5.84
UNKNOWN	CD610_21590		Conserved hypothetical protein with BMC-like jelly roll fold	1	0.28
UNKNOWN	CD610_23860		Conserved hypothetical protein	1	3.59
UNKNOWN	CD610_24300		Conserved hypothetical protein	1	0.28
UNKNOWN	CD610_28790		Conserved hypothetical protein	1	0.29
UNKNOWN	CD610_30220		Conserved hypothetical protein	1	3.28
UNKNOWN	CD610_30430		Conserved hypothetical protein	1	0.04
UNKNOWN	CD610_31230		Conserved hypothetical protein	1	3.04
UNKNOWN	CD610_31240		Conserved hypothetical protein	1	0.65
UNKNOWN	CD610_33512		Conserved hypothetical protein	1	3.3
Virulence Factors	CD610_06640.cdc		Negative regulator of toxin gene expression	1	0.16