What's a SNP between friends: the lineage of *Clostridioides difficile* R20291 can effect research outcomes.

Jorge Monteford[1]*, Terry W. Bilverstone[1,2]*, Patrick Ingle[1,2], Sheryl Philip[1,3], Sarah A. Kuehne[1,4], Nigel P. Minton[1,2].

* These authors contributed equally.

[1] Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Centre for Biomolecular Sciences, The University of Nottingham, Nottingham, NG7 2RD, UK.

[2] NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, NG7 2RD, UK.

Current Address:

[3]Biocatalysts Limited, Unit 1, Cefn Coed, Parc Nantgarw, Cardiff, CF15 7QQ, UK,

[4]Oral Microbiology Group, School of Dentistry and Institute of Microbiology and Infection, College of Medical and Dental Sciences, The University of Birmingham, Birmingham, B5 7EG, UK.

Corresponding author: nigel.minton@nottingham.ac.uk

Key Words

Clostridioides difficile R20291; motility; biofilm; toxin production; conjugation; genomic variation

Abstract:

Clostridioides difficile R20291 is the most studied PCR-Ribotype 027 isolate. The two predominant lineages of this hypervirulent strain, however, exhibit substantive phenotypic differences and possess genomes that differ by a small number of nucleotide changes. It is important that the source of R20291 is taken into account in research outcomes.

Clostridioides difficile (formerly *Clostridium difficile* [1]) is the leading cause of hospital-associated diarrhoea in the developed world. Its prevalence in recent years has been attributed to the emergence of hypervirulent strains, and in particular those belonging to BI/NAP1/PCR ribotype 027 (RT 027) which elaborate high titres of Toxin A/B, produce binary toxin and exhibit an increased propensity to form spores [2]. The first RT 027 strain to have its genome sequenced was strain R20291 [3] responsible for a major outbreak in 2006 at Stoke Mandeville Hospital, UK. Consequently, R20291 has become one of the most studied laboratory strains of *C. difficile.*

Full exploitation of clostridial genome sequence data has relied on the application of forward and reverse genetics tools [4], most notably ClosTron technology based on intron re-targeting [5]. Initial attempts to generate mutants in R20291, however, found that the effective transfer of the ClosTron plasmid was dependent on the R20291 stock used. Transfer was reproducibly possible using CRG0825, a stock of R20291 obtained by Nottingham's Clostridia Research Group (CRG) from the UK Anaerobe reference unit (ARU), Cardiff, UK. In comparison, transfer to a stock of R20291 (CRG2021) originating from the Brendon Wren laboratory at The London School of Hygiene and Tropical Medicine (LSHTM), was extremely ineffective. Consequently, the CRG0825 was taken forward in reverse genetic studies using the ClosTron and as the basis for the development of allelic-exchange (AE) technology based on *pyrE* alleles [6]. As a result, CRG0825 and its Δ*pyrE* derivative have been widely distributed to research laboratories wishing to study R20291.

The inefficient nature of CRG2021 as a conjugative recipient is not confined to the ClosTron plasmid but affects a range of different vectors which are transferred to CRG0825 at rates that are an order of magnitude higher (Fig.S1). To shed light on this phenomenon the genome sequences and the phenotypes of the two strains were compared. A third R20291 strain used by Novartis (CRG3661) was included for comparative purposes.

Genomic DNA from all three strains was subjected to Illumina paired-end sequencing and the reads assembled and aligned with the reference genome sequence (Accession number: FN545816). This analysis identified six single nucleotide polymorphisms (SNPs) across all three strains that deviate from the reference sequence, alongside thirteen insertions and eleven deletions (Table 1). In addition to the mutations that were conserved across all three strains, CRG0825 possessed three deletions and two SNPs that were not present in the reference or CRG2021 sequence, whilst the CRG3661 possessed three unique SNPs (Table 1). CRG2091 did not possess any unique mutations compared with the reference genome sequence.

Flagella likely play an important role in the conjugation process. We had previously noted that CRG0825 carried a single, polar flagella [7]. A separate study suggested that CRG2091 was peritrichously flagellated [8]. These differences were confirmed here using Transmission Electron Microscopy (TEM) and extended to establish that CRG3661 was also peritrichously flagellated (Fig. 1c-e). Further analysis demonstrated that CRG0825 exhibited an approximate 50% reduction in swimming motility relative to the other two strains (Fig. 1a). Moreover, consistent with its reduced motility, strain CRG0825 was also found to show a greater propensity to form biofilm, as measured by a biomass formation using crystal violet [9], than strains CRG2021 and CRG3661 (Fig. 1b).

Other studies have linked flagella-mediated motility with toxigenesis in *C. difficile* [10]. Therein, inactivation of early-stage flagella genes led to increased toxin production corresponding with enhanced *in vivo* virulence, whilst inactivation of late-stage flagella genes had the opposite effect [11, 12]. Accordingly, we assessed the levels of toxin production in the three strains using a commercial ELISA kit on 72h filter-sterilised

supernatants as described previously [13]. An approximately 3.5-fold increase in toxin production was observed for the CRG0825 compared with the CRG2021 strain which produced around 22% less combined Toxin A/B than CRG3661 (Fig. 2a).

Having established that genetic differences between the strains had affected the important virulence determinants of motility and toxin production, we tested to see whether the capacity to form endospores had been altered as spores represent a critical attribute of disease transmission. Under the conditions tested it was established that the final titre of spores obtained from 96h onwards was the same regardless of the strain (Fig. 2b). The first appearance of spores in cultures of CRG0825, however, was significantly delayed by some 24h compared to the other two strains (Fig. 2b).

Finally, the growth performance of each strain was compared. On complex medium, CRG0825 grew to a lower optical density (OD) during the exponential and stationary growth phases than the CRG3661 or CRG2021 strains, where the greatest disparity was observed between the CRG0825 and CRG3661 (Fig S2a). Intriguingly, the observed difference was reversed when the strains were cultured on minimal medium containing either glucose, fructose or mannitol (1% w/v) as the primary carbohydrate source (Fig. S2b-d).

The net result of our analysis was that the two R2091 strain CRG0825 and CRG2021 exhibit significant phenotypic differences. Aside from its greater efficiency as a recipient in conjugations with *E. coli* donor strains, CRG0825 was less motile and exhibited a greater propensity to form biofilm, as measured by a standard crystal violet assay. These differences may represent a consequence of its apparent possession of a single, polar flagella as opposed to the peritrichous flagella of CRG2021, as visualised under TEM. CRG0825 was also shown to produce higher levels of toxins, delayed sporulation and different growth characteristics to CRG2021 on rich and minimal media. The cause of these phenotypic differences are undoubtedly the SNPs and Indels present in CRG0825. A number of pivotal questions emerge from these observations.

What are the specific causes of the observed changes in phenotype? Many of the changes appear linked to flagella and motility, yet none of the five mutations in CRG0825 reside directly within, or flank any known flagella genes, and are most likely impinging on the regulation of these processes. Moreover, regulation of flagella, toxin production and virulence are known to be linked in *C. difficile* [10-12]. Three of the four non-synonymous SNPs present in CRG0825 do indeed affect apparent regulatory genes, namely *vncR,* TCS HK and the anti-sigma factor *rsbW.* Two of the three non-synonymous changes in CRG3661 are also in regulatory genes, *codY* and a *gntR* family regulator. However, to pin down exactly which SNP(s) or Indels, are responsible for the observed phenotypic differences between CRG0825 and CRG2021, for instance, would require a considerable effort in which all combinations of mutation would need to be generated in allele replacement experiments during which the generation of additional changes would need to be excluded.

How did these changes arise? Following their discovery, correspondence with Val Hall at the ARU revealed that at the time R20291 was sent to Nottingham, it was routine ARU practice to "*keep a small number of isolates that are used as internal lab controls on agar plates, sub-culturing weekly plate-to-plate and retrieving fresh cultures from the original vial periodically".* The sequence presented here is from Nottingham's -80° C, Master seed bank (red tube) prepared immediately on receipt of the strain in 2006. We can conclude, that during the repeated subculture of the R20291 stock at the ARU in 2006, the 5 described mutations arose. This practice of maintaining a stock plate no longer takes place at the ARU. The consequences of subculturing have previously been noted in the case of the *C. difficile* strain 630, where deliberate, repeated subculture led to the emergence of two very different cell lines (630Δ*erm* and 630E) carrying distinct SNPs, inversions and deletions and which exhibited differences in motility, spore formation and toxin production, as well as overall virulence in the hamster model of infection [14].

What are the lessons to be learnt? The take home message of this investigation is that stock cultures need to be appropriately maintained. At Nottingham, a traffic light

system is used to store bacterial cultures. Upon receipt of cultures, aliquots of cells (never single colonies) are used to inoculate an overnight which following addition of 10% glycerol is allocated to three 2 ml screwed capped tubes with a red, amber and green coloured cap insert and stored at -80° C. Red tubes remain untouched and are stored in a separate freezer. Green tubes represent the working stock which may be restocked from the amber tube where necessary.

The R20291 strain maintained at LSHTM has a genome sequence consistent with the sequence held at GenBank (Accession number: FN545816). The differences listed in Table 1 are common to all strains, and therefore likely represent errors in the original sequence. The strain CRG3661 began its journey at LSHTM and found its way to Novartis via the Trevor Lawley laboratory at the Sanger Institute, and thence to Nottingham. It is not clear when its three mutations arose. The Nottingham CRG0825 apparently arose as a result of repeated subculture at the ARU.

What is the way forward? It is clearly advisable that the genome sequences of any stock culture of any bacterial strain stored in a laboratory should be confirmed, regardless of source, prior to use. This principal should equally apply to any mutant derivative of a strain made by whatever means, to ensure that additional SNPs/ Indels have not arisen.

On the specific subject of studies dealing with R20291, it is important that experimentalists are aware of the differences between the strain lineages discussed here, and that the strain used is made clear in any meeting presentation or published article. CRG0825 is a widely distributed strain, owing to its superior conjugative efficiencies and its usage in the development of AE mutagenesis technologies [6]. The absence of polymorphisms specific to CRG2021, however, suggests that this strain is the closest ancestor of the original R20291 clinical isolate. Although the lack of a characterised uracil auxotroph, in addition to difficulties concerning conjugal transfer, formally reduced its attractiveness compared to CRG0825, recent advances that improve gene transfer frequencies [15, 16] and the advent of multiple CRISPR-Cas methodologies for use in *C. difficile* research [17-20], have improved the tractability of CRG2021 to genetic studies. For those researchers who wish

to use AE technologies based on *pyrE* [6], the requisite auxotrophic uracil mutant of CRG2021 is available from www.plasmidvectors.com.

References

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

2. Cartman S.T., Heap J.T., Kuehne S.A., Cockayne A., Minton N.P. The emergence of 'hypervirulence' in *Clostridium difficile*. *Int. J. Med. Microbiol*. 2010;**300(6):**387-95.

3. Stabler R.A., He M., Dawson L. et al. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol*. 2009;**10(9):**R102

4. Minton N.P., Ehsaan M., Humphreys C.M., Little G.T., Baker J., Henstra A.M., Liew F., Kelly M.L., Sheng L., Schwarz K., Zhang Y. A roadmap for gene system development in *Clostridium. Anaerobe*. 2016;**41:**104-112.

5. Heap J.T., Kuehne S.A., Ehsaan M., Cartman S.T., Cooksley C.M., Scott J.C., Minton N.P. *J. Microbiol. Methods*. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. 2010;**80(1):**49-55.

6. Ng Y.K., Ehsaan M., Philip S., Collery M.M., Janoir C., Collignon A., Cartman S.T., Minton N.P. Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: Allelic Exchange using *pyrE* alleles. *PLoS ONE.* 2013;**8(2):**e56051.

7. Baban S.T., Kuehne S.A., Barketi-Klai A., Cartman S.T., Kelly M.L., Hardie K.R., Kansau I., Collignon A., Minton N.P. The role of flagella in *Clostridium difficile* pathogenesis: comparison between a non-epidemic and an epidemic strain. *PLoS One*. 2013;**8(9):**e73026.

8. Martin M.J., Clare S., Goulding D., Faulds-Pain A., Barquist L., Browne H.P., Pettit L., Dougan G., Lawley T.D., Wren B.W. The *agr* locus regulates virulence and colonization genes in *Clostridium difficile* 027. *J. Bacteriol*. 2013;**195(16):**3672-81.

9. Đapa T., Leuzzi R., Ng Y.K., Baban S.T., Adamo R., Kuehne S.A., Scarselli M., Minton N.P., Serruto D., Unnikrishnan M. Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*. *J Bacteriol*. 2013;**195(3):**545-55.

10. Aubry A., Hussack G., Chen W., KuoLee R., Twine S.M., Fulton K.M., Foote S., Carrillo C.D., Tanha J., Logan S.M. Modulation of toxin production by the flagellar regulon in *Clostridium difficile*. *Infect. Immun.* 2012;**80(10):**3521-3532.

11. Dingle T.C., Mulvey G.L., Armstrong G.D. Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. *Infect Immun*. 2011;**79(10):**4061-4067.

12. Anjuwon-Foster B.R., Tamayo R. A genetic switch controls the production of flagella and toxins in *Clostridium difficile*. *PLoS Genet.* 2017;**13(3):**e1006701-e1006701.

13. Bilverstone T.W., Kinsmore N.L., Minton N.P., Kuehne S.A. Development of *Clostridium difficile* R20291ΔPaLoc model strains and in vitro methodologies reveals CdtR is required for the production of CDT to cytotoxic levels. *Anaerobe.* 2017;**44:**51-54.

14. Collery M.M., Kuehne S.A., McBride S.M., Kelly M.L., Monot M., Cockayne A., Dupuy B., Minton, N.P. What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of *Clostridium difficile* strain 630 and derivatives. *Virulence* 2017;**8(6):**767-781.

15. Kirk J.A., Fagan R.P. Heat shock increases conjugation efficiency in Clostridium difficile. *Anaerobe*, 2016;**42:**1-5.

16. Woods C., Humphreys C.M., Rodrigues R.M., Ingle P., Rowe P., Henstra A.M., Köpke M., Simpson S.D., Winzer K., Minton N.P. A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe*. 2019;**59:**184-191.

17. McAllister K.N., Bouillaut L., Kahn J.N., Self W.T., Sorg J.A. Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Sci. Rep.* 2017;**7(1)**:14672.

18. Hong W., Zhang J., Cui G., Wang L., Wang Y. Multiplexed CRISPR-Cpf1-Mediated Genome Editing in *Clostridium difficile* toward the Understanding of Pathogenesis of *C. difficile* Infection. *ACS Synth. Biol.* 2018;**7(6):**1588-1600.

19. Maikova A., Kreis V., Boutserin A., Severinov K., Soutourina O. Using an Endogenous CRISPR-Cas System for Genome Editing in the Human Pathogen *Clostridium difficile. Appl. Environ. Microbiol*. 2019;**85(20):**e01416-19.

20. Cañadas IC, Groothuis D, Zygouropoulou M, Rodrigues R, Minton NP. RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*. *ACS Synth. Biol.* 2019;**8(6):**1379-1390.

ACKNOWLEDGEMENT

This work was supported by the award of a PhD scholarship to JM by CONACYT (Consejo nacional de ciencia y tecnologia, Mexico), through the UK Biotechnology and Biological Sciences Research Council [grant number BB/J014508/1] and Medical Research Council [grant number G090085] and by the Swiss National Science Foundation [Sinergia grant CRSII5_180317]. The views expressed are those of the authors and not necessarily those of the funders.

Author contributions

Conceived the experiments: NPM. Performed the experiments: JM, TWB and PI. Undertook genome sequence determination and analysis: SP, TWB and NPM. Analysed the data: JM, TWB, PI, SAK and NPM. Wrote the paper: TWB and NPM. All authors read and commented on the final manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Table 1: Genomic mutations compared with the reference genome sequence for R02921.

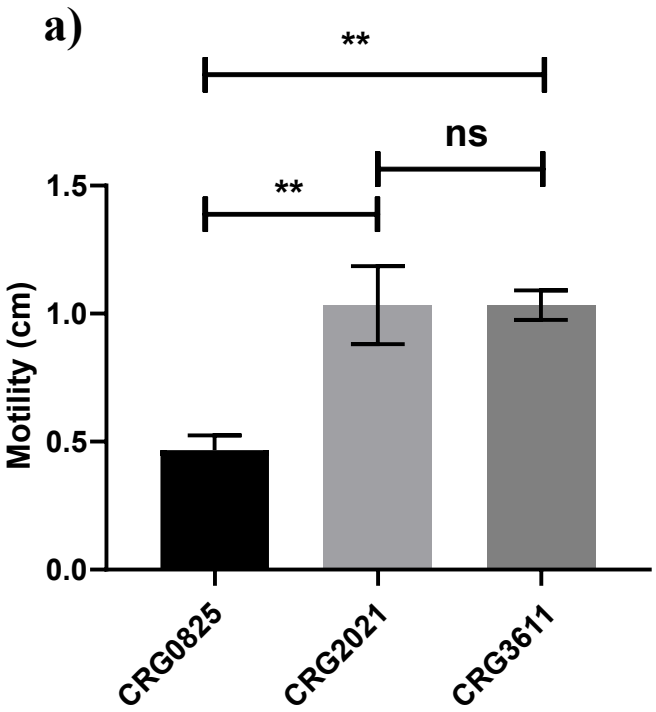
The region encompassing the mutation was aligned with multiple *C. difficile* genome sequences using NCBI Blastn. *Insertion here results in a frameshift mutation to a full-length pseudogene encoding an 87 AA protein. This gene without mutation encodes only 8 AAs. TCS-HC: Two-component system histidine kinase.

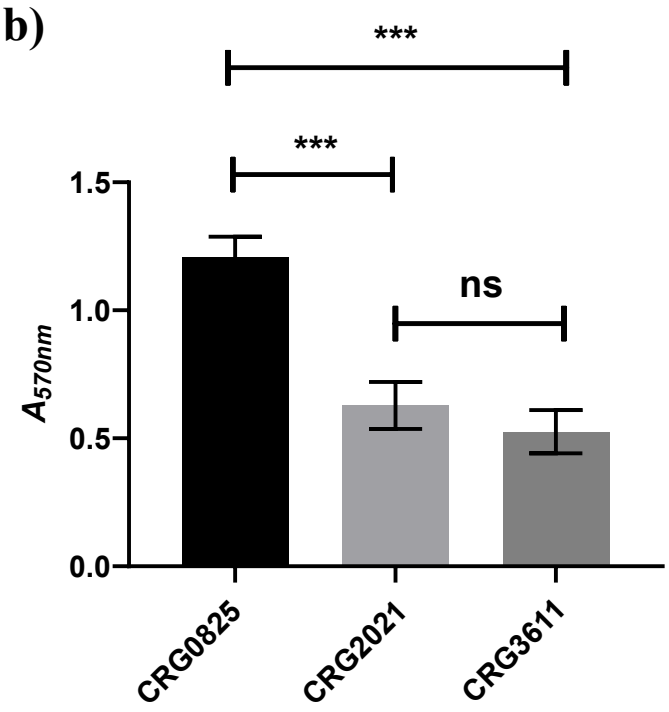
Figure Legends

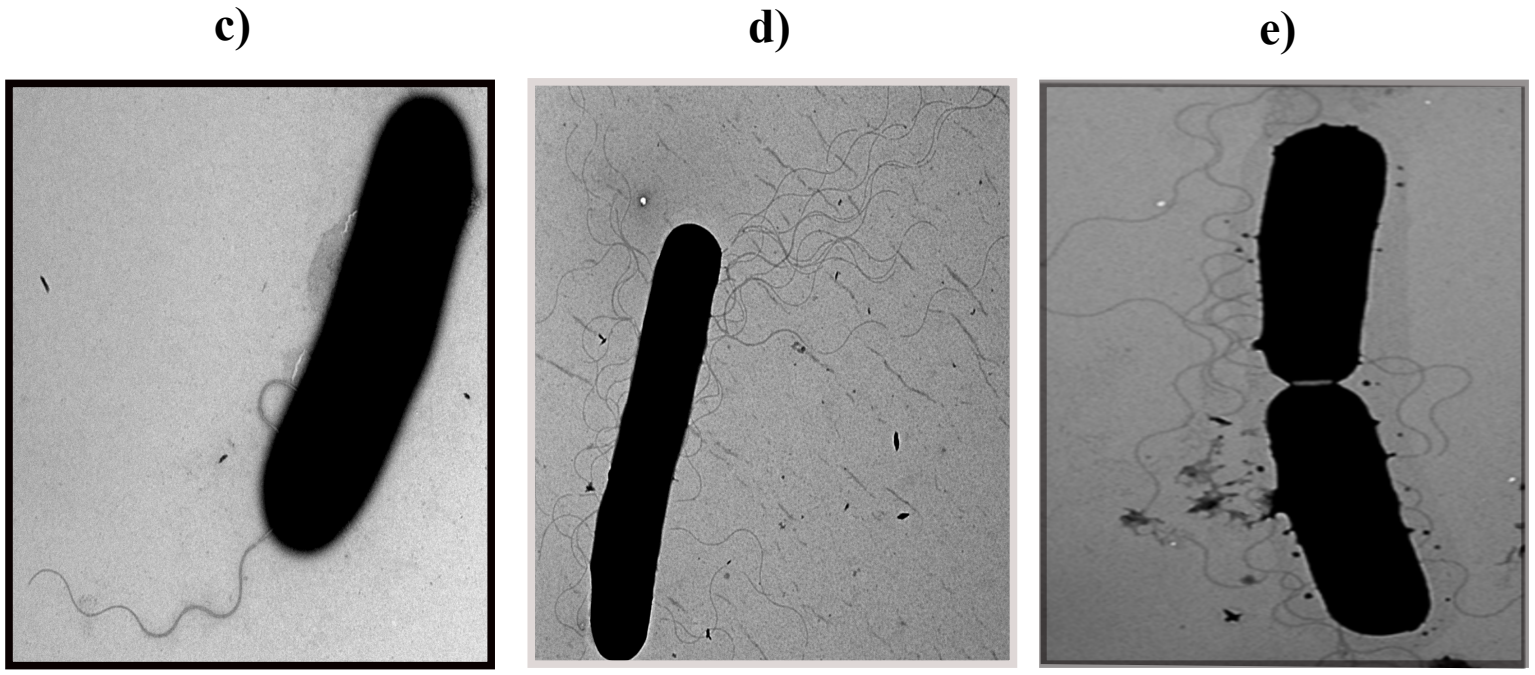
Fig. 1: Motility, biofilm formation and Transmission electron microscopy of R20291 derivatives. a) The three derivatives of *C. difficile* R20291 were assessed for their motility characteristics using a swarming motility assay. Motility is represented by the distance travelled from the initial inoculum to the outermost edge of the ensuing halo following 48h incubation. **b**) The three R20291 derivatives were assessed for their propensity to form biofilms by means of a biofilm assay. Biofilm production is represented by the enumeration of crystal violet dye extracted from 120h broth cultures. Data represent the mean ±SD of three independent experiments. Statistical significance according to One-way ANOVA followed by the Dunnet's multiple comparison test. P=**<0.01; P-***<0.001. Transmission electron microscopy analysis of **c)** CRG0825; **d)** CRG2091; **e)** CRG3661.

Fig. 2: Toxin and sporulation profiles of R20291 derivatives. The three derivatives of R20291 were assessed for **a)** Their ability to produce and secrete toxin through a combined ELISA for TcdA and TcdB on sterile-filtered 72h supernatants **b)** Their ability to

form heat-resistant endospores (heat-resistant colony forming units HR-CFU/mL) across six time-points between 0 and 120h. Data represent the mean ±SD of three independent experiments. Statistical significance according to One-way ANOVA followed by the Dunnet's multiple comparison test (P=*<0.05; ****<0.0001).


a)
**
ns
**
1.5
1.0
0.5
0.0
Motility (cm)
CRG0825
CRG2021
CRG3611


b)
***
***
ns
1.5
1.0
0.5
0.0
A570nm
CRG0825
CRG2021
CRG3611


c)
d)
e)

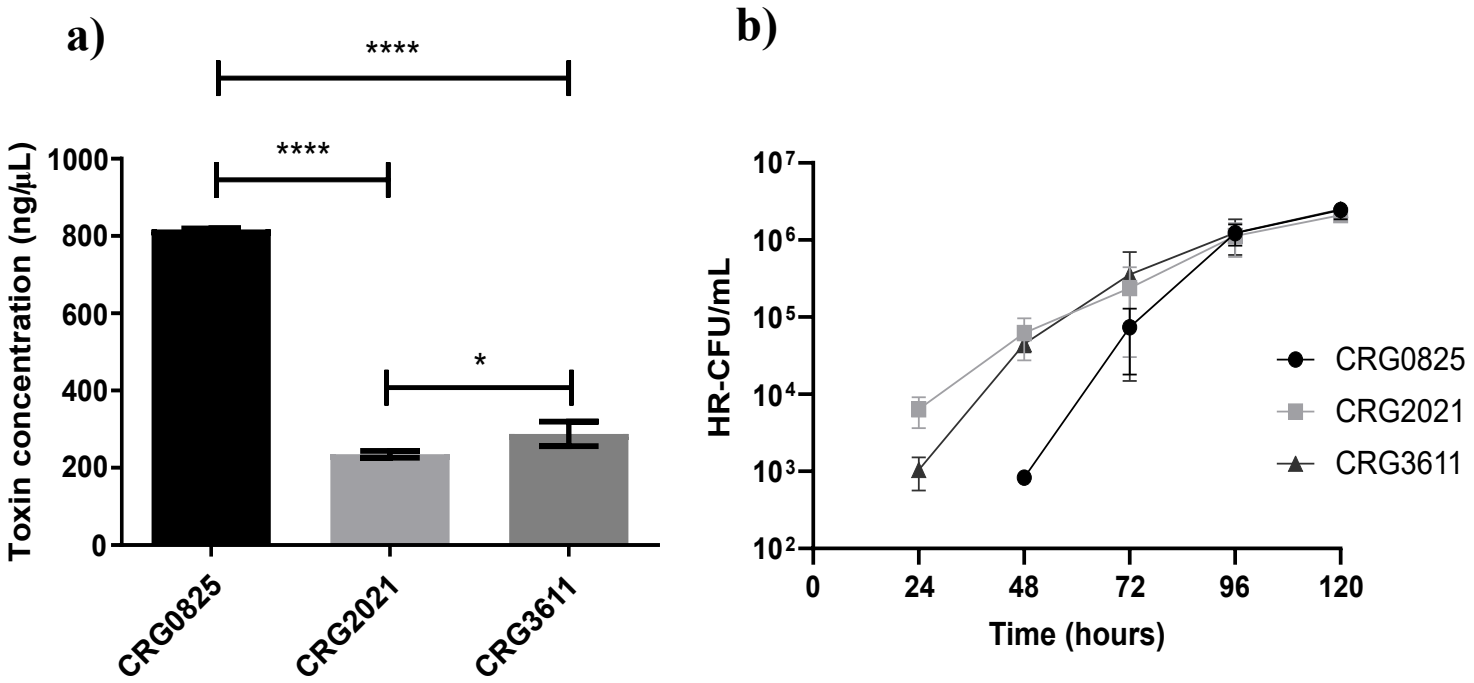

a)
****
****
*
Toxin concentration (ng/µL)
1000
800
600
400
200
0
CRG0825
CRG2021
CRG3611
b)
HR-CFU/mL
10^7
10^6
10^5
10^4
10^3
10^2
0
24
48
72
96
120
Time (hours)
CRG0825
CRG2021
CRG3611

Table 1: Genomic mutations of the three R20291 stocks compared with the reference genome sequence.

Position	Gene	Locus Tag	Type	Reference ID	Mutated ID	AA substitution
[1] Common to all three strains						
132924	*met-tRNA*		Insertion	-	A	
132939	*met-tRNA*		SNP	G	T	
132955	*met-tRNA*		SNP	C	A	
132958-59	*met-tRNA*		SNP	TT	CG	
143463	Intergenic		Insertion	-	A	
206399	Intergenic		Insertion	-	A	
581481	Intergenic		Insertion	-	A	
581488	Intergenic		Insertion	-	A	
581495	Intergenic		Insertion	-	A	
1564432	Intergenic		Deletion	A	-	
1568676	Ruberythrin	CDR20291_1323	SNP	C	A	Gln138Lys
1578167	Intergenic		Deletion	T	-	
1578203	Intergenic		Insertion	-	A	
1592813	Intergenic		SNP	A	T	
1864417	Pseudogene	CDR20291_1576	Insertion	-	T	Ser7Frameshift*
1899596	Intergenic		Deletion	A	-	
2235738	Membrane protein	CDR20291_1913	Deletion	T	-	Val83Frameshift
2262060	Intergenic		Insertion	-	A	
2264190	Intergenic		Deletion	T	-	
2298111	Intergenic		Insertion	-	T	
2361948	Intergenic		SNP	C	A	
2361957	Intergenic		Insertion	-	A	
2367942	Intergenic		Insertion	-	T	
2578157	Intergenic		Deletion	T	-	
2674744	Intergenic		Deletion	T	-	
2680787	Intergenic		Insertion	-	T	
2772179	Pseudogene**	CDR20291_2368	Deletion	T	-	
3077986	Intergenic		Deletion	A	-	
3162098	Intergenic		Deletion	T	-	
3361915	Intergenic		Deletion	A	-	
[2] Specific to CRG0825						
9694	*rsbW*	CDR20291_3551	SNP	G	T	Gly82Val
358260	*rbsK*	CDR20291_0302	Deletion	A	-	Met57Stop
2077305	Intergenic	(CDR20291_1777 to CDR20291_1778)	Deletion	C	-	
2120669	*vncR*	CDR20291_1806	SNP	A	G	Asp202Gly
2881467	TCS-HK***	CDR20291_2456	Deletion	T	-	Leu434Stop
[3] Specific to CRG3661						
1340128	*codY*	CDR20291_1115	SNP	T	A	Try146Asn
3292465	*gntR* regulator	CDR20291_2781	SNP	T	C	Ile58Met
3472928	Intergenic	(CDR20291_2929 to CDR20291_2928)	SNP	G	A	

The region encompassing the mutation was aligned with multiple *C. difficile* genome sequences using NCBI Blastn. *Insertion here results in a frameshift mutation to a full-length pseudogene encoding an 87 AA protein. This gene without mutation encodes only 6 AAs. **Putative competence membrane protein ***TCS-HC: Two-component system histidine kinase. The gene resides immediately downstream of an adjacent gene (CDR20291_2457) encoding a putative response regulator.

What's a SNP between friends: the lineage of *Clostridioides difficile* R20291 can effect research outcomes.

Jorge Monteford[1]*, Terry W. Bilverstone[1,2]*, Patrick Ingle[1,2], Sheryl Philip[1,3], Sarah A. Kuehne[1,4], Nigel P. Minton[1,2].

* These authors contributed equally.

[1] Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Centre for Biomolecular Sciences, The University of Nottingham, Nottingham, NG7 2RD, UK.

[2] NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, NG7 2RD, UK.

Corresponding author: nigel.minton@nottingham.ac.uk

Current Address:

[3]Biocatalysts Limited, Unit 1, Cefn Coed, Parc Nantgarw, Cardiff, CF15 7QQ, UK,

[4]Oral Microbiology Group, School of Dentistry and Institute of Microbiology and Infection, College of Medical and Dental Sciences, The University of Birmingham, Birmingham, B5 7EG, UK.

Supplementary material

Table S1: *C. difficile* strains used in this study

Strain	Description
CRG0825	R20291 sent to Nottingham by Val Hall in 2006 from the Anaerobe Reference Unit (ARU), Cardiff, UK.
CRG2021	R20291 sent to Nottingham by Lisa Dawson in 2010 from the laboratory of Brendan Wren at the London School of Hygiene and Tropical Medicine (LSHTM), UK. Originally obtained from the ARU, Cardiff, UK.
CRG3661	R20291 sent to Nottingham by Meera Unnikrishnan in 2013 from Novartis, Sienna, Italy. Sent to Novartis by Trevor Lawley at the Sanger Institute, Cambridge, UK. Originally obtained from the Brendan Wren laboratory at LSHTM, London UK.
CRG1375	R20291 (CRG0825) *spo0A* ClosTron mutant [1].

Experimental

Bacterial Strains and Growth Conditions

Strains were routinely cultured anaerobically at 37 °C in an anaerobic MACS1000 workstation (Don Whitely, Yorkshire, UK) in BHIS (Brain Heart Infusion supplemented with yeast extract [5 mg.ml−1] and L-cysteine [0.1% w/v]) medium supplemented with d-

cycloserine (250 µg.ml−1), cefoxitin (8 µg.ml−1) and thiamphenicol (15 µg.ml−1) or Em (10 µg.ml−1) where appropriate.

Comparative conjugations

Conjugations of shuttle vectors into *C. difficile* R20291 were performed as described in [2]. Briefly, *E. coli* CA434 donor strains harbouring either pMTL82151, pMTL83151 or pMTL84151 were grown overnight in LB supplemented with chloramphenicol and kanamycin. From which, aliquots (1ml) were pelleted, washed in PBS and resuspended in 200 µl of *C. difficile* R20291 cultures grown anaerobically overnight in BHIS broth. The resulting conjugal mixtures were spotted onto BHIS plates lacking antibiotics and incubated anaerobically for 24h. Subsequent growth was resuspended in PBS (500µl) and spread onto BHIS plates supplemented with d-cycloserine and cefoxitin, both with and without thiamphenicol. After 72h, Thiamphenicol (Tm) resistant (^{R}) CFU.ml^{-1} and total *C. difficile* CFU.ml^{-1} values were determined from the subsequent growth of *C. difficile* R20291 strains in the presence (Tm^{R}) or absence (total) of thiamphenicol. Conjugation efficiency was calculated as the Tm^{R} CFU.ml^{-1} divided by total *C. difficile* CFU.ml^{-1}.

24h growth curve

The growth characteristics of *C. difficile* R02921 was assessed by manual growth curve. Therein, colonies of *C. difficile* R20291 were subcultured into fresh BHIS broth in an Anaerobic Work Station (Don Whitley, UK), at 37°C with an anaerobic atmosphere comprising 80% N_2, 10% H_2 and 10% CO_2. The resultant cultures were then diluted 1/100 in fresh BHIS and grown to an optical density value (OD_{600nm}) of 0.2-0.5. This generated replicates of each strain with similar starting OD values for downstream growth assessment. Once target OD values had been reached, the cultures were diluted 1/100 in fresh medium incubated for 24h. 1ml of sample was taken for each replicate at hourly intervals which was then assessed for its optical density using a Novaspec II spectrophotometer (Geminibv, Netherlands).

Motility assay

The motility of *C. difficile* R20291 derivatives was assessed by swimming motility assays as previously described [3]. Therein, single colonies of R20291 were isolated using a toothpick and stabbed onto the centre of semi-solid BHIS plate containing 0.3% (w/v) agar. Following 48h incubation as described above, the diameter of the ensuing halo was measured. Motility is represented as the distance between the centremost and outermost points of detected colonisation (cm).

Crystal Violet, Biofilm Assay

The assay was undertaken essentially as described by Dapa and co-workers [4]. A starter culture of *C. difficile* was prepared by inoculating fresh BHIs broth containing 0.1M glucose with an overnight culture of the desired strain in a 1:100 dilution. A 1ml aliquot of this culture added to each well of a 24-well Tissue culture plate (Costar,USA) and incubated anaerobically for 120h. Plates were pre-reduced in the anaerobic cabinet for 48h prior to use. To avoid liquid evaporation, each plate was wrapped in parafilm. Following incubation, wells were washed with PBS and the plate allowed to dry for 10m. The biofilm was stained with 1ml of filter-sterilised 0.2% (w/v) crystal violet solution per well and incubated for 30m at 37°C under anaerobic conditions. The staining solution was removed and the wells were washed twice with PBS. The plate was removed from the anaerobic cabinet and 1ml methanol was added to the wells for the removal of the dye from the biofilm and it was incubated for 30m at room temperature. The methanol extracted dye was diluted 1:1, 1:10 and 1:100 and the absorbance A570 was measured with Ultrospec 500 pro spectrophotometer.

Sporulation assay

Cultures were generated for each strain with similar staring OD values as described for the 24h growth curve. For the sporulation assay, cultures were incubated for a 120h period. Samples were taken at 24h intervals which were heated at 65°C for 30m in order to eradicate vegetative cells and diluted $1x10^{-1}$-$1x10^{-8}$ before plating onto BHIS supplemented with 0.1% taurocholate germinant. Spores were then enumerated for each R20291 derivative alongside a ClosTron insertional mutant for the master regulator of sporulation *spo0A* [1].

Detection of combined TcdA and TcdB

Combined TcdA and TcdB was detected as previously described [5]. Cultures of each strain were collected after 72h, the OD measured, and normalised to the lowest OD value. Normalised samples were centrifuged and the supernatant passed through a 0.22µm filter and frozen at -20°C until required (<1 week). Thawed samples were diluted $1X10^{1}$ – $1X10^{8}$ in sterile PBS and processed for ELISA quantification of total TcdA and TcdB using a C. DIFFICILE Tox A/B II detections kit (TechLab, USA) according to the manufacturer's instructions. Absorbance values were converted into toxin concentration by determining the R2 value of the assay's standard curve by running a range of defined combined TcdA and TcdB toxin standards from 0-125ng/ml (The Native Antigen Company).

Genome Sequencing

Chromosomal DNA of each strain was prepared and subjected to Illumina paired-end sequencing by DeepSeq (University of Nottingham) using the MiSeq v3 600 platform. Paired reads were trimmed, before mapping the trimmed reads to the reference genome sequence for R20291 (Accession number: FN545816) using the quality-based variant detection workflow from CLC Genomics Workbench (Qiagen, Germantown, USA). The software was then used to identify single nucleotide variations (SNVs), insertions and deletions compared with the reference genome sequence. Sequencing reads were

deposited to the NCBI Sequencing Reads Archive with the Bioproject accession PRJNA689976 and the following individual accession numbers: CRG2021 (SRR13366486); CRG0825 (SRR13366485); CRG03661 (SRR13366484).

Supplementary Figures

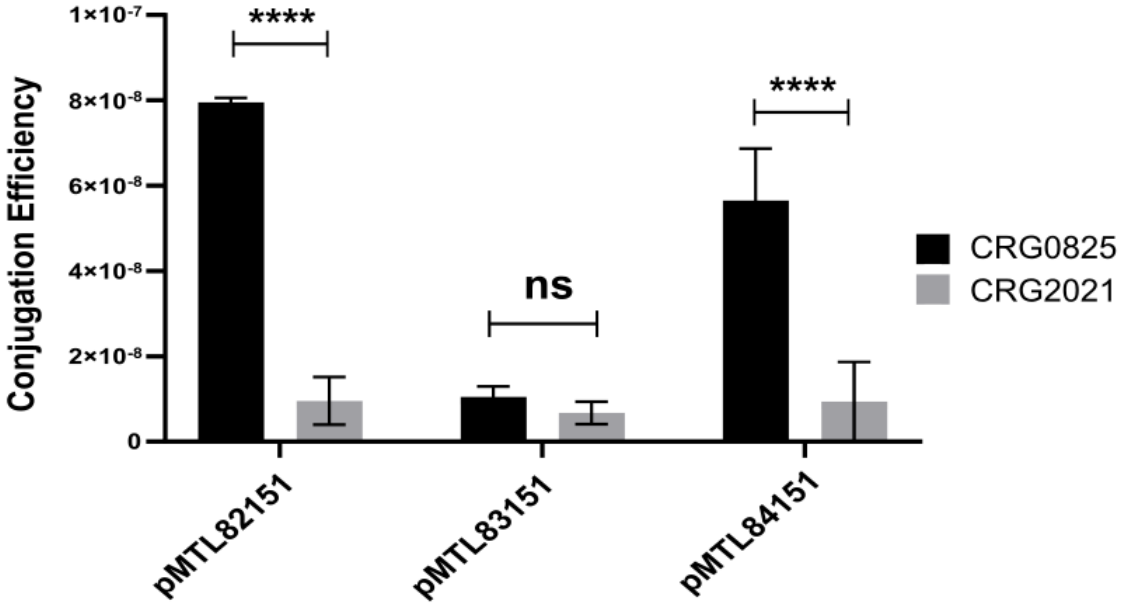


Fig S1: Comparative conjugation efficiencies of plasmid transfer from *E. coli* CA434 into *C. difficile* R20291 stocks CRG0825 and CRG2021. Conjugations from *E. coli* CA434 strains harbouring the indicated shuttle vectors, differing only in the Gram-positive replicon present, into *C. difficile* R20291 CRG2021 (grey bars) and CRG0825 (black bars) were performed as indicated in Materials and Methods. Conjugation efficiency was calculated as thiamphenicol resistant CFU.ml^{-1} divided by the total recipient *C. difficile* R20291 CFU.ml^{-1}. Data represent the mean ±SD of three independent experiments. Statistical significance was determined using multiple unpaired t-tests. P=****<0.0001; ns= not significant.

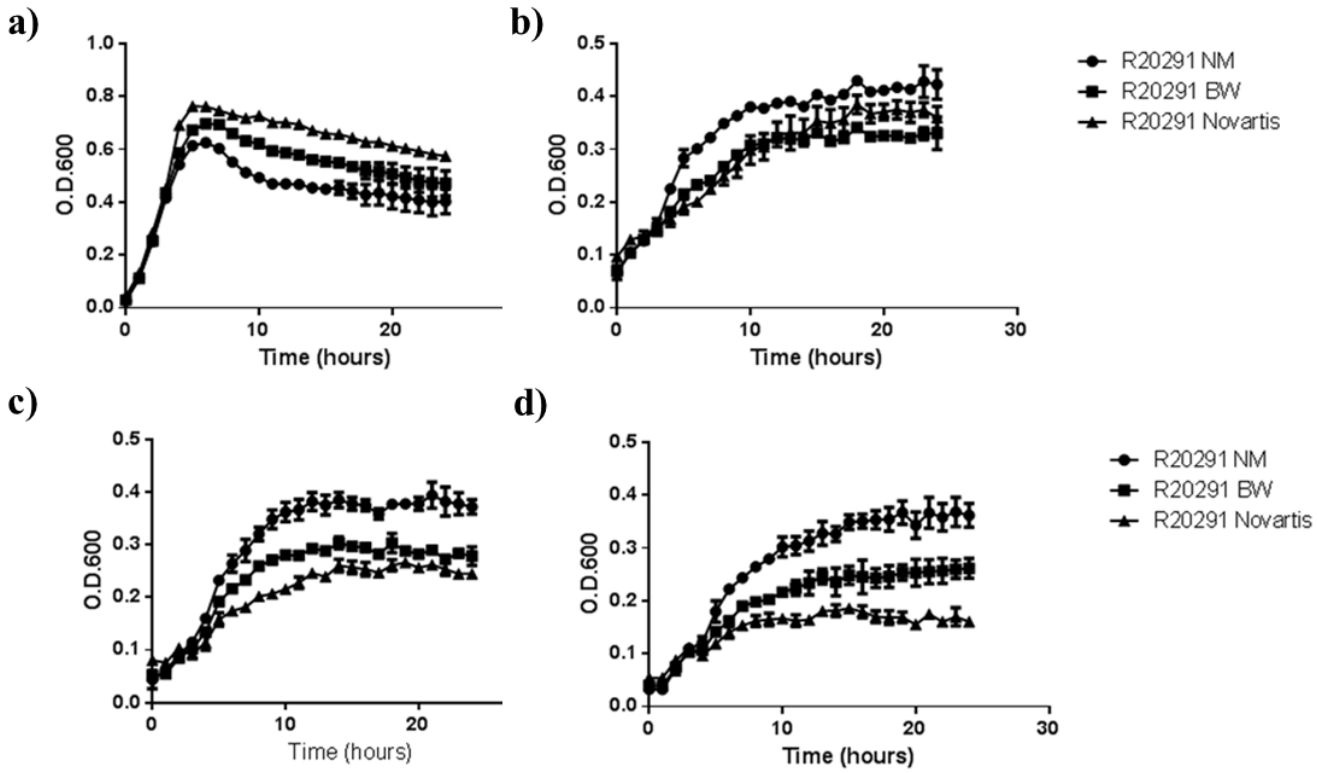


Figure S2: Comparison of growth characteristics between the three derivatives of R20291. R02921 strains were grown for 24h in a) BHIS broth; b) CDMM 1% (w/v) glucose; c) CDMM 1% fructose; d) CDMM 1% mannitol. Data points indicate the mean ±SD of three independent experiments.

References

1. Heap J.T., Kuehne S.A., Ehsaan M., Cartman S.T., Cooksley C.M., Scott J.C., Minton N.P. *J Microbiol Methods*. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. 2010;**80(1):**49-55.

2. Woods C., Humphreys C.M., Rodrigues R.M., Ingle P., Rowe P., Henstra A.M., Köpke M., Simpson S.D., Winzer K., Minton N.P. A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe*. 2019;**59:**184-191.
3. Baban S.T., Kuehne S.A., Barketi-Klai A., Cartman S.T., Kelly M.L., Hardie K.R., Kansau I., Collignon A., Minton N.P. The role of flagella in *Clostridium difficile* pathogenesis: comparison between a non-epidemic and an epidemic strain. *PLoS One*. 2013;**8(9):**e73026.
4. Đapa T., Leuzzi R., Ng Y.K., Baban S.T., Adamo R., Kuehne S.A., Scarselli M., Minton N.P., Serruto D., Unnikrishnan M. Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*. *J Bacteriol*. 2013;**195(3):**545-55.
5. Bilverstone T.W., Kinsmore N.L., Minton N.P., Kuehne S.A. Development of *Clostridium difficile* R20291ΔPaLoc model strains and in vitro methodologies reveals CdtR is required for the production of CDT to cytotoxic levels. *Anaerobe.* 2017;**44:**51-54.