

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



**Characterising regulation of virulence factors affecting the
interaction between *Clostridioides difficile* and the intestinal
microbiome**

Mark Andrew Harrison

**Thesis submitted in accordance with the requirements for the
degree of**

**Doctor of Philosophy
of the
University of London**

January 2022

Department of Infection Biology

Faculty of Infectious and Tropical Diseases

**London School of Hygiene & Tropical Medicine, University of
London**

Funded by Medical Research Council

Research group affiliation(s): Dr Lisa Dawson

Professor Brendan Wren

Declaration

I, Mark Harrison, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



Mark Harrison

Countersigned by supervisors



Lisa Dawson



Brendan Wren

Abstract

Clostridioides difficile is the leading cause of antibiotic associated diarrhoea, causing significant morbidity and mortality worldwide. Treatment of *C. difficile* infection (CDI) with antibiotics can be effective, however, up to 30% of patients suffer from infection recurrence, often multiple episodes, either through infection relapse or reinfection. Even treatment with narrow spectrum antibiotics such as fidaxomicin, which is considered microbiome sparing, is still associated with a significant proportion of patients experiencing recurrence. Therefore, treatments targeting *C. difficile*'s specific virulence factors, including their expression and regulation, offer a promising route to manage CDI and reduce recurrence, however, a greater understanding of their underlying biology is required.

A major *C. difficile* virulence factor that provides it with a competitive advantage over certain species of the gut microbiota is the production of the antibacterial compound *p*-cresol. *p*-cresol is produced by the fermentation of tyrosine to give the precursor *p*-HPA which is converted to *p*-cresol by the actions of HpdBCA decarboxylase, encoded on the *hpdBCA* operon. Three different transcriptional reporter systems were used, and evaluated, to characterise the *hpdBCA* promoter and found that supplementation of growth media with *p*-HPA drives expression of *hpdBCA* leading to high level *p*-cresol production. This response was found across strains representing all five *C. difficile* clades showing that this is an extremely well conserved pathway. Additionally, exogenous *p*-HPA was found to be inhibitory to *C. difficile* growth, an effect enhanced by inactivation of HpdBCA, suggesting conversion of *p*-HPA to *p*-cresol provides *C. difficile* with a growth advantage. Similarly to *p*-cresol, *p*-HPA was also found to inhibit commensal species commonly found in the gut, with Gram-negative strains more susceptible than Gram-positive strains. Confocal microscopy revealed that HpdB is expressed ubiquitously by all cells within a *C. difficile* culture highlighting the importance of this response to *C. difficile*. These findings validate the hypothesis that targeting *p*-cresol production by inhibition of HpdBCA would be effective across all *C. difficile* infections regardless of causative strain.

An alternative approach to targeting *C. difficile* virulence factors directly is to target the regulators that control their expression and understand how they influence *C. difficile*'s lifestyle. The *sinR-CD2215* locus is formed of two transcriptional regulators and is involved in the control of *C. difficile* lifestyle choice with a major role in sporulation regulation. The work in this study describes key regulatory controls of the *sinR-CD2215* locus including autoregulation by SinR itself, together with identification of its binding site, and control of the operon's expression by major regulators Spo0A, CodY and CcpA.

This study has added significant insight into the regulatory controls of key virulence factors by *C. difficile* and highlights the importance of these controls to *C. difficile* infection.

Acknowledgements

Firstly, I would like to thank my supervisors Dr Lisa Dawson and Professor Brendan Wren for all of their support and the huge number and range of opportunities they have given me over the course of my PhD.

I would like to give a special mention to Catherine for her constant support, advice and generally making our little lab a nice place to come to every day. A big thanks to Ian and Jen for your help but most of all for all the random little chats. A big mention to Alex, it was your incredible support throughout my masters project that led me back to LSHTM all these years later. I'd also like to thank the rest of the Lab 280 group for all the time you've taken to help me and for being a great group to work in.

I would like to thank our collaborators who took time to contribute their expertise and knowledge to our projects, including Dr Rivka Isaacson and Janina Muench at KCL, Dr Isabelle Martin-Verstraete and Dr Bruno Dupuy at the Institut Pasteur, Dr Henrik Strahl at Newcastle University and finally the team at Proxima concepts.

I'd like to thank my family for their constant and amazing support throughout everything I do. Finally, and I've saved the best for last... I would like to thank my wonderful partner Amy, without whom I wouldn't have made it this far, your unwavering support and love means more than you know.

Contents

Declaration.....	2
Abstract.....	3
Acknowledgements	4
List of figures.....	8
List of tables	10
List of abbreviations.....	11
Chapter 1: Introduction	12
1.1 <i>C. difficile</i> infection	12
1.1.1 Epidemiology	12
1.1.2 Diagnosis.....	14
1.1.3 Infection outcome	15
1.2 CDI Treatment.....	16
1.2.1 Antibiotics.....	16
1.2.2 Faecal Microbial Transplant	17
1.2.3 Antibody treatments and vaccines.....	18
1.3 <i>C. difficile</i> virulence factors.....	19
1.3.1 Toxin production	19
1.3.2 Sporulation	21
1.3.3 Antimicrobial resistance	22
1.3.4 Gene reporter systems in <i>C. difficile</i> for the study of virulence factors.....	23
1.4 Regulation of virulence factors in <i>C. difficile</i>	25
1.4.1 The role of the <i>sin</i> locus in virulence factor regulation	26
1.5 The microbiome and dysbiosis in CDI.....	29
1.5.1 The role of <i>p</i> -cresol in CDI and potential disruption of the microbiome	31
1.6 Aims and objectives.....	34
Chapter 2: Materials and Methods.....	36
2.1 <i>C. difficile</i> growth	36
2.2 DNA manipulation	36
2.2.1 Polymerase chain reaction	36
2.2.2 Agarose Gel Electrophoresis.....	37
2.2.3 DNA Extraction	37
2.3 RNA extraction.....	38
2.3.1 cDNA Synthesis.....	39

2.4 Cloning techniques	39
2.4.1 Restriction-ligation cloning.....	39
2.4.2 Gibson assembly.....	40
2.4.3 Site directed mutagenesis	40
2.4.4 DNA sequencing	40
2.5 Cell Transformation	40
2.5.1 <i>E. coli</i>	40
2.5.2 <i>C. difficile</i>	41
2.6 qRT-PCR	41
2.7 Electromobility Shift Assays (EMSAs)	41
2.8 Reporter assays.....	42
2.8.1 SNAP-tag assay	42
2.8.2 <i>gusA</i> assay	43
2.8.3 <i>phoZ</i> assay	43
2.9 Phenotypic assays.....	43
2.9.1 Growth curves	43
2.9.2 High pressure liquid chromatography (HPLC)	44
2.9.3 Colony forming units	45
2.9.4 Phosphate release assay.....	46
2.9.5 Confocal microscopy	47
2.9.6 Preparation of samples for quantification of SinR and CD2215 by western blot.....	48
2.9.7 Western blotting.....	48
2.9.8 Motility assays	49
2.10 Statistical analysis.....	49
Chapter 3: <i>Clostridioides difficile</i> <i>para</i> -cresol production is induced by the precursor <i>para</i> -hydroxyphenylacetate	51
3.1. Research paper cover sheet	51
3.2 Additional results and discussion	73
Chapter 4: Production of <i>p</i> -cresol by decarboxylation of <i>p</i> -HPA by all five lineages of <i>Clostridioides difficile</i> provides a growth advantage.	75
4.1 Research paper cover sheet	75
Chapter 5: Characterisation of regulation and expression of the <i>C. difficile</i> <i>sinR-CD2215</i> locus	118
5.1 Introduction.....	118
5.2 Results.....	122
5.2.1 Transcriptional reporter construction	122

5.2.2 Identification of a single promoter controlling <i>sinR-CD2215</i> expression.....	124
5.2.3 Expression from the <i>sinR-CD2215</i> promoter in <i>sinR-CD2215</i> , <i>CD2215</i> and <i>spo0A</i> mutants.....	127
5.2.4 Promoter expression in a single <i>sinR</i> deletion mutant.....	131
5.2.5 Regulators affecting expression from the <i>sinR-CD2215</i> promoter	132
5.2.6 Identification of the <i>sinR</i> binding site.....	136
5.2.7 SinR and CD2215 protein quantification	139
5.2.8 Effect of mutation of <i>sinR-CD2215</i> on motility and toxin production.....	141
5.3 Discussion	143
Chapter 6: Discussion.....	149
6.1 <i>P</i> -cresol production	149
6.1.1 Utilisation of exogenous <i>p</i> -HPA by <i>C. difficile</i> for <i>p</i> -cresol production	149
6.1.2 Comparison of transcriptional reporters.....	151
6.1.3 Regulation of <i>hpdBCA</i> expression.....	152
6.1.4 Inhibitory effects of <i>p</i> -HPA	154
6.1.5 Mechanism of <i>p</i> -HPA toxicity	156
6.1.6 Cellular response to <i>p</i> -HPA.....	157
6.1.7 Conservation of <i>hpdBCA</i> induction by <i>p</i> -HPA leads to high level <i>p</i> -cresol production across all five <i>C. difficile</i> clades	158
6.1.8 <i>p</i> -cresol production from tyrosine metabolism	159
6.2 Regulation of <i>sinR-CD2215</i> expression	160
6.2.1 Characterisation of the <i>sinR-CD2215</i> promoter and the roles of SinR and CD2215 in <i>sinR-CD2215</i> expression	160
6.2.2 Roles of Spo0A, CodY and CcpA in <i>sinR-CD2215</i> expression	163
6.3 Final conclusions.....	166
6.4 Future work	166
6.4.1 Further research into <i>p</i> -cresol production by <i>C. difficile</i>	166
6.4.2. Further characterisation of <i>sinR</i> and <i>CD2215</i> and factors involved in their expression	170
7. References	172
Appendices	189
Glossary.....	202

List of figures

Chapter 1 – *Clostridioides difficile* background

Figure 1.1. Diagram of the pathway of <i>p</i> -cresol production in <i>C. difficile</i>	32
--	----

Chapter 3 – *Clostridioides difficile* *para*-cresol production is induced by the precursor *para*-hydroxyphenylacetate (published manuscript)

Figure 1. Overview of the SNAP, <i>phoZ</i> , and <i>gusA</i> reporter constructs.....	55
Figure 2. Visualization of the SNAP-tag reporter activity under the control of the <i>hpdBCA</i> promoter and a constitutive <i>fdx</i> promoter.....	56
Figure 3. Visualization of the <i>phoZ</i> reporter activity under the control of the <i>hpdBCA</i> promoter and a constitutive <i>fdx</i> promoter.....	57
Figure 4. Visualization of the <i>gusA</i> reporter activity under the control of the <i>hpdBCA</i> promoter.....	58
Figure 5. Use of the <i>phoZ</i> reporter to determine a concentration-dependent response to <i>p</i> -HPA for induction of the <i>hpdBCA</i> operon.....	59
Figure 6. HPLC analysis to detect tyrosine, <i>p</i> -HPA, and <i>p</i> -cresol in spent culture media.....	59
Figure 7. Mutation of the potential promoter regions in the <i>hpdBCA</i> -SNAP-tag reporter constructs..	61
Figure 8. Mutation of the inverted repeat located within the <i>hpdBCA</i> upstream region leads to an inability to respond to <i>p</i> -HPA.....	62
Figure S1. Uncropped SNAP-tag gel.....	72
Figure 9. qRT-PCR of 630 Δ <i>erm hpdC</i> expression in response to exogenous <i>p</i> -HPA.....	73

Chapter 4 - Production of *p*-cresol by decarboxylation of *p*-HPA by all five lineages of *Clostridioides difficile* provides a growth advantage (submitted manuscript)

Figure 1. <i>hpdBCA</i> expression, <i>p</i> -HPA and <i>p</i> -cresol production by strains representing each of <i>C. difficile</i> clades 1-5.....	104
Figure 2. Confocal microscopy to confirm expression and determine localisation of HpdB.....	106
Figure 3. Growth analysis of 630 Δ <i>erm</i> and 630 Δ <i>erm hpdC::CT</i> in the presence of <i>p</i> -HPA.....	107
Figure 4. Effect of <i>p</i> -HPA and <i>p</i> -cresol on sporulation rate and total cell count.....	108
Figure 5. Analysis of growth of <i>C. difficile</i> and gut commensal species in the presence of <i>p</i> -HPA.....	109
Figure 6. Assessing the effect of <i>p</i> -HPA compared to pH on membrane integrity of <i>C. difficile</i> compared to <i>E. coli</i>	110
Figure 7. Mutation of <i>codY</i> reduces <i>p</i> -HPA conversion to <i>p</i> -cresol via a reduction in expression of the <i>hpdBCA</i> operon.....	111
Figure S1. Confirmation of HpdB-SNAP fusion.....	112
Figure S2. Growth of <i>C. difficile</i> in pH matched to the presence of <i>p</i> -HPA.....	112
Figure S3. Growth of <i>C. difficile</i> in acidic BHIS media.....	113
Figure S4. Growth of gut commensals in pH matched to the presence of <i>p</i> -HPA.....	114
Figure S5. <i>p</i> -cresol production from 630 Δ <i>erm</i> and 630 Δ <i>erm</i> Δ <i>codY</i>	115
Figure S6. Alignment of <i>hpdBCA</i> -like operons carried by cryptic clades C-I, C-II and C-III to <i>hpdBCA</i> operon from 630.....	115

Chapter 5 - Characterisation of regulation and expression of the *C. difficile* *sinR-CD2215* locus

Figure 5.1. Structure of the *sin* locus in *B. subtilis*.....119

Figure 5.2. Diagram showing design of four reporter constructs for characterisation of regulatory controls of the *sinR-CD2215* locus.....123

Figure 5.3. Visualisation of SNAP-tag reporter activity under control of putative promoters of CD2215.....124

Figure 5.4. Expression from P_{sinR} -Short and P_{sinR} -Short ΔP_1126

Figure 5.5. Quantification of SNAP-tag expression from reporters P_{sinR} -Short and P_{sinR} -Long at exponential phase growth.....128

Figure 5.6. *phoZ* reporter activity under the control of the *sinR-CD2215* promoter in 630 Δerm and 630 Δerm *spo0A::CT*.....130

Figure 5.7. Quantification of SNAP-tag expression from reporter P_{sinR} -Long in 630 Δerm (Paris) and 630 $\Delta erm \Delta sinR$131

Figure 5.8. Quantification of SNAP-tag expression from reporters P_{sinR} -Long and P_{sinR} -Long *codY* mut in 630 Δerm133

Figure 5.9. Quantification of expression from the *sinR-CD2215* promoter in 630 Δerm (Paris) and $\Delta ccpA$ using P_{sinR} -Long-SNAP and P_{sinR} -Long-*phoZ* reporters.....135

Figure 5.10. Electromobility shift assay to determine binding of SinR and CD2215 to the *sinR-CD2215* upstream region.....137

Figure 5.11. Quantification of SNAP-tag expression from reporters P_{sinR} -Long and P_{sinR} -Long ΔIR138

Figure 5.12. Western blot with anti-His for his-tagged SinR and CD2215.....140

Figure 5.13. Swimming motility of *sinR* and *CD2215* knockout strains.....142

Figure 5.14 . Summary of findings of regulatory factors involved in *sinR-CD2215* expression.....148

List of tables

Chapter 3 - *Clostridioides difficile* para-cresol production is induced by the precursor para-hydroxyphenylacetate (published manuscript)

Table 1. Strains and plasmids used in this study.....	65
Table 2. Oligonucleotides used in this study.....	66

Chapter 4 - Production of *p*-cresol by decarboxylation of *p*-HPA by all five lineages of *Clostridioides difficile* provides a growth advantage (submitted manuscript)

Table 1. Strains and plasmids used in this study.....	102
Table 2. Oligos used in this study.....	103
Supplementary table 1. Peptides identified by mass spectrometry of HpdB-SNAP-tag.....	116

Appendices

Table 1. Bacterial strains used in this study.....	189-193
Table 2. Plasmids used in this study.....	194-195
Table 3. Primers used in this study.....	196-198
Table 4. Components of growth media used in this study.....	199
Table 5. Reagents used in this study.....	200-201

List of abbreviations

BHI – Brain heart infusion
BHIS – Brain heart infusion supplemented yeast extract and cysteine
CC – Cycloserine cefoxitin
CDI - *Clostridioides difficile* infection
CDT – *C. difficile* toxin (binary toxin)
CFE – Cell free extract
CFU – Colony forming unit
DEPC - Diethyl pyrocarbonate
DTT – Dithiothreitol
EDTA - Ethylenediaminetetraacetic acid
EIA - Enzyme immunoassay
EMSA – Electromobility shift assay
FMT – Faecal microbial transplant
GDH - Glutamate dehydrogenase
GFP – Green fluorescent protein
HTH – Helix-turn-helix
MIC – Minimum inhibitory concentration
MOPS - 3-(*N*-morpholino) propane sulfonic acid
ml – millilitre
µg - microgram
NAAT - Nucleic acid amplification test
NHS – National Health Service
NMR – Nuclear magnetic resonance
OD_{590 nm} – Optical density 590 nm
***p*-cresol** – *para*-cresol
***p*-HPA** – *para*-hydroxyphenylacetic acid
PBS – Phosphate buffered saline
PCR - Polymerase chain reaction
PHE - Public Health England
PMC - Pseudomembranous colitis
PTS - Phosphotransferase system
qRT-PCR - Quantitative reverse transcriptase polymerase chain reaction
RACE - Rapid amplification of cDNA ends
RT - Ribotype
SDM - Site directed mutagenesis
SDS-PAGE - sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SOC - Super optimal broth with catabolite repression
TAE - Tris-acetic acid-EDTA
TBE - Tris-boric acid-EDTA
TBS - Tris buffered saline
Tm – Thiamphenicol
v/v – volume/volume
w/v – weight/volume

Chapter 1: Introduction

Clostridioides difficile is a Gram-positive, spore forming bacterium that is a leading cause of antibiotic associated diarrhoea worldwide. It was previously known as *Clostridium difficile* but has recently been reclassified as *Clostridioides difficile* as a result of the *Clostridium* genus being limited to *Clostridium butyricum* and its related species (1). *C. difficile* causes significant morbidity and mortality in the UK and globally. *C. difficile* infection (CDI) can range from asymptomatic colonisation in infants to pseudomembranous colitis (PMC) and death in adults (2, 3). Critically, 20-30% of patients suffer from infection recurrence, often multiple recurrences, and this is a major driving factor behind the costs of CDI (4). The challenges presented by CDI are numerous and high in human and financial terms. Infection in the UK has fallen in recent years from a dramatic high; in 2007/08 55,498 cases were reported compared to 13,177 cases in 2019/20 (5). However, CDI prevalence has stabilised at approximately 12,000 cases per year and remains a significant drain on the NHS (5) and healthcare systems worldwide with the economic cost of CDI in the USA estimated at \$5.4 billion in the year 2014 (6).

1.1 *C. difficile* infection

1.1.1 Epidemiology

C. difficile was first identified as being a cause of PMC and antibiotic associated diarrhoea in 1978 (7, 8) however, it did not receive much attention until the early 2000's when a severe outbreak was reported in the Quebec province of Canada (9) which was quickly followed by further reports from around the world (10, 11). These outbreaks were associated with a strain belonging to ribotype (RT) 027, which was found to be more resistant to antibiotics with a key feature of these strains being the acquisition of high level fluoroquinolone resistance (12, 13), with later analysis revealing resistance had arisen in two separate epidemic RT027 lineages (14). A second highly virulent ribotype, RT078, emerged shortly after RT027 and was found to have similar rates of disease severity and mortality (15, 16). Recent evidence from animal models of infection suggests that these epidemic strains are more virulent than non-epidemic strains (17). Interestingly, the hyper-virulent ribotype RT027 strain,

associated with large numbers of infection in the late 2000's, prevalence fell from 55% of CDI cases in 2007-08 to 21% in 2009-10 (18). It has been suggested that the drop in prevalence of RT027 cases is linked to a reduction in use of fluoroquinolones as these antimicrobials are a major risk factor for CDI (18).

Soon after the emergence of the epidemic strains reports were made of infections arising in community settings in people who were previously considered to be at low risk of infection (19, 20). It is now recognised that the epidemiology of CDI is changing with a large proportion of cases being community acquired, with estimates suggesting that this accounts for approximately 40% of cases (21-24) with data from Public Health England (PHE) suggesting that up to 64.3% of cases are now community acquired in England (5). RT078 strains are of particular concern as they are more likely to cause infections in the community and in younger patients (16). A major caveat to community acquired cases is that the vast majority of patients who develop CDI have had exposure to healthcare settings with 82% associated with outpatient care and 94% associated with any form of healthcare exposure (25, 26). In addition, it has been shown that *C. difficile* can be found extensively throughout the environment of healthcare facilities (27) so whilst infections are spread outside of healthcare settings it appears CDI and healthcare are inextricably linked.

The two main risk factors associated with CDI are receipt of broad-spectrum antibiotics and advanced age of 65 years or over (28, 29). Antibiotics lead to dysbiosis of the gut microbiome and as a result *C. difficile* can establish and maintain infection where previously the healthy microbiome provided colonisation resistance. Antibiotic exposure is associated with both healthcare and community acquired CDI infection, although risk varies by antibiotic class, for example tetracyclines are not associated with CDI whereas fluoroquinolones and clindamycin are considered high risk (30-32). Those receiving a longer course of therapy are at increased risk than those receiving shorter courses (33). Advanced age is a major risk factor for CDI, particularly in the over 65s (34, 35). Lessa *et al.* found a risk ratio for this age group of 8.65 compared to those under 65 (25). In addition, the elderly have a

higher risk of infection with the hypervirulent RT027 strains (35). Changes in immunocompetence and the microbiome with age may partially be the reason behind increased susceptibility. A number of alterations to the microbiome have been identified with increasing age such as decreases in *Bifidobacterium* (36) and increases in *Bacteroides* (37).

A number of other risk factors have been identified and associated with infection including length of stay in hospital (29, 38), receipt of proton pump inhibitors (26, 39, 40), and comorbidities such as inflammatory bowel disease (41) and renal failure (42). A recently identified risk factor for CDI is diarrhoeal episodes, with the finding that these episodes can lead to long term colonisation with *C. difficile* and when tested in a mouse model of infection it was found that mice who were treated with increasing doses of laxatives were more likely to experience blooms of *C. difficile* growth (43). Most recently, data from a mouse model of infection has shown that diet can affect the severity of infection with a high fat, high protein diet proving to exacerbate CDI whilst a high carbohydrate diet protects against disease (44). Therefore, whilst age and antibiotic receipt are major risk factors there are a growing number of other risk factors being identified which may help to explain how people previously thought to be at low risk are now increasingly developing active CDI.

1.1.2 Diagnosis

Diagnosis of CDI should follow a multi-step algorithm and is complicated by the fact that a significant proportion of patients carry *C. difficile*, including toxigenic strains, asymptomatically (45). Indeed, there is evidence that asymptomatic carriage is protective of active and symptomatic infection (46). Recently, guidelines state that testing should be carried out on patients suspected of CDI as indicated by unexplained and new diarrhoea, which is defined as ≥ 3 un-formed stools within 24 hours (47). There is currently no evidence for asymptomatic patients to be screened, or for carriers of *C. difficile* to be isolated (47). It is vital to take in to account the clinical presentation of patients to avoid unnecessary testing (48). Recent recommendations on diagnosis differ depending on whether clinicians agree not to submit stool samples for testing with the knowledge that a patient is receiving laxatives (47). Current testing methods include nucleic acid amplification test (NAAT), glutamate

dehydrogenase (GDH) test, toxin assays and toxigenic cultures. All of these assays are imperfect, thus requiring the need for a multi-step algorithm for diagnosis in order to mitigate the effects of potentially over or under diagnosing. The current policy in the UK is to use a multistep algorithm for diagnosis, for example GDH testing may be used to screen patients and any positive results undergo confirmatory testing by use of a toxin immunoassay (49). If the system to avoid testing patients receiving laxatives is in place then testing may be done by either NAAT alone or a stool toxin test resulting in a much simpler testing strategy (47).

1.1.3 Infection outcome

Infection outcome typically varies according to the age of the patient colonised or infected, although increasingly caveats are being recognised to this dogma. Infants are commonly colonised asymptotically with reported rates of carriage varying heavily, with one study reporting rates as high as 71% (3, 50, 51). Whilst neonates and children are often asymptomatic carriers, they may also develop disease and symptoms including diarrhoea and PMC (reviewed by Enoch *et al.* (52)). Furthermore, there is evidence suggesting *C. difficile* leads to microbiome changes in colonised neonates (53). Asymptomatic carriage may also occur in adults (54, 55), however this is considered to be less common than in the young. Asymptomatic carriage is a key reservoir for infection spread in healthcare settings, with Sheth *et al.* reporting that 20% of patients with *C. difficile* colonisation acquired the bacterium from asymptotically colonised patients within the hospital (56). Furthermore, it has been shown that toxigenic strains are also carried asymptotically, further demonstrating the importance of these patients in the transmission of *C. difficile* within healthcare facilities (57-59).

Symptomatic CDI ranges from mild diarrhoea to much more serious complications such as toxic megacolon (2). Overall mortality rate estimates have varied with the latest figures from PHE for 2017/18 showing a case fatality rate of 15.2% (60). However, this a significant decrease from the case fatality rate 10 years ago which was 42% (60). There is similar evidence in a drop in mortality rate in the USA across a 10 year period from 2005 to 2014 (61). A key issue associated with CDI is the rate of

infection recurrence which has been estimated to be as high as 35% (62), with subsequent episodes of infection recurrence increasing in probability (63). Infection recurrence can occur via either relapse, i.e., the same strain causing symptoms repeatedly, or via reinfection, where a patient becomes infected with a new strain of *C. difficile*. Results vary between studies reporting if recurrence was due to relapse or reinfection, with one study finding near identical rates between the two (64) and others finding recurrence due to relapse in more than 60% of cases (62, 65). Recently, it has been shown to be possible to identify when infection recurrence is due to relapse or reinfection via whole genome sequencing (66) but this is not routinely done in hospitals due to the cost and complexity.

1.2 CDI Treatment

1.2.1 Antibiotics

The most recent treatment guidelines from PHE, published in 2013, recommend antibiotic treatment as first line therapy for moderate or severe disease with metronidazole and vancomycin respectively (67). In the USA, management guidelines were updated in 2017 and recommend a number of different antibiotic treatment strategies depending on severity of disease and infection recurrence (47). These treatment strategies largely recommend the use of vancomycin or fidaxomicin. Vancomycin and metronidazole are both effective against *C. difficile* but both are associated with significant levels of infection recurrence, this is partially as a result of the additional dysbiosis they cause to the microbiome (68, 69).

Fidaxomicin is a narrow spectrum antibiotic with strong activity against Clostridia, however other Gram-positives as well as Gram-negative organisms are resistant (70). Therefore, fidaxomicin offers the advantage of affecting fewer species of the gut microbiome and is considered microbiome sparing (68, 71). Fidaxomicin has been shown to significantly reduce the likelihood of infection recurrence (72), for example Cornely *et al.* showed a recurrence rate of 35.5% with vancomycin compared with 19% with fidaxomicin (73). Fidaxomicin is considerably more expensive than vancomycin, however it

has been shown to be cost effective, which is in large part due to reduced infection recurrence resulting in patients spending fewer days in hospital (74).

Recently, a novel antimicrobial for CDI treatment; ridinilazole, has completed Phase II trials and has been shown to be effective and safe (75). Importantly, similarly to fidaxomicin, ridinilazole has a narrow spectrum of activity with high levels of activity against Clostridia whilst sparing other organisms indigenous to the gut including both Gram-negative and Gram-positive organisms (76, 77). In a Phase II trial, Vickers *et al.* found a sustained clinical response (i.e., no recurrence within 30 days) in 66.7% of patients treated by ridinilazole versus 42.4% of those treated by vancomycin (75). Currently, there is limited data comparing fidaxomicin and ridinilazole however a small study has found that ridinilazole is more targeted and sparing of the microbiome than fidaxomicin (78) which bodes well for improving reductions in infection recurrence.

1.2.2 Faecal Microbial Transplant

Where antibiotic therapy has failed to treat recurrent infection, particularly multiple episodes of recurrence, the use of faecal microbial transplant (FMT) has been shown to be highly effective. This strategy aims to restore colonisation resistance to *C. difficile* by restoration of the gut microbiota, which may be vital to CDI clearance as the microbiota has been associated with *C. difficile* clearance independent of the adaptive immune response in an animal model (79). A systematic review by Moayyedi *et al.* found that FMT is superior to vancomycin for the treatment of recurrent *C. difficile* associated diarrhoea (80). Furthermore, there have been calls to consider FMT as a first line treatment for CDI following the finding of it reducing three month mortality (81). Whilst expensive, FMT has been demonstrated to be potentially cost effective, particularly in cases of recurrent CDI (82-84) with one group estimating savings of up to AU\$4000 per patient (85). In addition, despite the potentially unappealing nature of FMT it is generally well accepted by patients (86).

Despite efficacy against recurrent CDI there are concerns around FMT including regulatory considerations and patient safety. There is little long-term safety data following FMT use and this issue has become especially acute since the death of a patient as a result of the transfer of a multi-drug resistant *Escherichia coli*. There have also been other reports of death linked to FMT (87). As a result of these cases the FDA have published advice regarding the screening procedures to be undertaken on donor stool samples and the need to standardise and more closely regulate FMT donations has been acknowledged (88). As a result, the race continues to identify a defined cocktail of commensal bacteria that could replace FMT with several groups and companies involved.

1.2.3 Antibody treatments and vaccines

A major advance in the treatment of recurrent CDI is the development of Bezlotoxumab, a monoclonal antibody targeted at Toxin B. Bezlotoxumab has been shown to be highly effective in reducing infection recurrence in high risk patients (89, 90). Interestingly, Wilcox *et al.* showed that Bezlotoxumab treatment was effective in reducing infection recurrence, however, this efficacy did not improve upon concurrent administration of a second monoclonal antibody, Actoxumab, targeted at Toxin A (89). Outside of Bezlotoxumab's clinical trials a small study in Finland demonstrated that amongst severe cases of CDI Bezlotoxumab prevented recurrence in 63% of cases (91). Bezlotoxumab has a high initial cost at approximately \$4500 / 1,000 mg vial. In a recent review by Alonso *et al.* the authors found evidence both for and against Bezlotoxumab in terms of cost effectiveness (92).

Currently, there is no vaccine available for *C. difficile* and there is little prospect of one being successfully developed and entering the market within the next five years. Past attempts at generating a vaccine focused on *C. difficile* toxins A and B as these have been shown to generate a specific antibody response (93). Inactivated toxins have been tested in both animal models as well as in humans and have been shown to be both safe and immunogenic (94-96). Despite these promising results, Sanofi Pasteur were forced to terminate a large Phase III clinical trial of their Cdiffense vaccine early after finding that their vaccine was ineffective at preventing CDI (97). Both Pfizer and Valneva

are still developing toxin-based vaccines having shown safety and immunogenicity in Phase II trials (98, 99) with Pfizer's candidate undergoing Phase III trials and Valneva's planned to follow suit.

A major problem with development of vaccines that are toxin targeted are that whilst clinical disease may not manifest, the vaccines will not prevent colonisation of patients, therefore these patients will then become a reservoir for transmission of the infection (100). As a result, there is now more focus on identification of targets that will prevent colonisation. A number of possible candidates have been identified such as CD0873, a *C. difficile* adhesin part of a tyrosine import system (101) and exosporium proteins CdeC and CdeM (102).

1.3 *C. difficile* virulence factors

1.3.1 Toxin production

C. difficile produces the large clostridial toxins (LCTs); Toxin A (TcdA) and Toxin B (TcdB) which are encoded on the pathogenicity locus (PaLoc) (103). The PaLoc is formed of five genes; *tcdA* and *tcdB*, as well as *tcdC*, a negative regulator of toxin production (104), *tcdR*, a RNA polymerase sigma factor that positively regulates expression (105) and *tcdE*, a holin-like protein involved in toxin secretion (106). The LCTs cause damage to the membrane of the colon leading to PMC typical of infection, with the importance of these toxins shown by Kuehne *et al.* who found that deletion of both toxin genes completely attenuated disease in a hamster model of infection (107). The mechanism of action of the LCTs is via glucosylation of GTPases including Rho, Rac and Cdc42 (108, 109) resulting in wide ranging effects including cytoskeleton damage and alteration in cell signalling pathways (110). It has been shown that TcdB is significantly more catalytically active than TcdA (111, 112) and may be more important to symptomatic disease as mutation of *tcdB* leads to attenuated disease (113). Furthermore, RT017 strains which lack TcdA cause significant morbidity and mortality most commonly in Asia (114-117). These findings in conjunction with the evidence that the efficacy of bezlotoxumab is not increased by adding treatment directed at TcdA suggests that TcdB is likely to be more important to disease (89).

The *tcdC* gene within the PaLoc is a negative regulator of toxin production and in hyper virulent strains, such as those of RT027, a single base pair deletion in *tcdC* has been identified (104, 118, 119). This

deletion leads to a truncated TcdC (104) and was thought to lead to de-repression of toxin production resulting in the high level of toxin production seen in hyper virulent strains (10). However, doubt has been cast on this by Cartman *et al.* who showed a lack of association between the deletion and toxin production (120) as well as Bakker *et al.* who showed that TcdC does not affect toxin production (121).

Production of the LCTs is an energy intensive process and as a result there are a number of regulators involved in the control of toxin production to ensure it is only carried out when it provides the maximum benefit for the bacterium. A major global regulator in *C. difficile* is CodY, which responds to the presence of branched chain amino acids (BCAAs) and GTP (122, 123). CodY has an important role in the regulation of toxin production by repressing it when the cell is in relatively nutrient rich conditions (124). Similarly, carbon catabolite protein A (CcpA) responds to the presence of carbohydrates and regulates genes involved in sugar uptake and fermentation, and like CodY, represses production of both TcdA and TcdB (125, 126). Other regulators involved in sporulation, Spo0A (127, 128) and RstA (129) have also been found to negatively regulate toxin production, likely to conserve energy for the cell during sporulation.

Another toxin produced by *C. difficile* whose role in disease is yet to be fully elucidated is binary toxin (CDT). The interest in this toxin stems from reports of patients with symptomatic CDI but who are infected with strains which are both Toxin A and Toxin B negative but CDT positive (130, 131). Whilst these strains are relatively rare, the number of strains positive for CDT has increased since the major outbreaks of the early 2000's (132). Furthermore, strains positive for CDT are associated with higher levels of infection recurrence and disease severity, including higher mortality rates (133, 134). The effects of CDT *in vivo* are relatively unclear however evidence suggests that CDT may be proinflammatory and as well as having a role in adherence and colonisation (135, 136). An important finding was that CDT may act synergistically with Toxin A, as strains with both Toxin A and binary toxin caused significantly faster death in a hamster model of infection than with one of these toxins alone (137).

1.3.2 Sporulation

The ability to form spores is vital to *C. difficile*, as they are essential for persistence and transmission (138). Spores are found throughout healthcare facilities (27), they are intrinsically resistant to antibiotics (139) and are resistant to commonly used hospital disinfectants (140), ethanol based hand sanitisers (141) and heat (142).

Sporulation in *C. difficile* is governed by the master sporulation regulator Spo0A (127, 143, 144). Spo0A regulates a wide variety of genes and crucially it controls expression of *sigH* and early sporulation genes; *spolIAA*, *spolIE*, and *spolIGA* which are responsible for controlling the activation of SigF and SigE. In addition to its role in sporulation, Spo0A has been found to have roles in the regulation of a number of other important virulence factors in *C. difficile*, including biofilm formation (145, 146), motility (127) and possibly toxin production (128). Crucial to the function of Spo0A is its phosphorylation, which is required for its activation, with the leading hypothesis for the mechanism behind phosphorylation in *Clostridium* species being a two-component system whereby sensor kinases respond to specific stimuli, either internal or environmental, and directly phosphorylate Spo0A (147, 148). Five genes encoding orphan histidine kinases, with homology to *kinA-E* genes which phosphorylate Spo0A in *Bacillus subtilis*, were suggested to play a similar role in *C. difficile* (149). CD1579 was shown to phosphorylate Spo0A, and CD2492 was shown to positively contribute to sporulation (149). Conversely, one such homolog, CD1492, has been found to be a negative regulator of sporulation (150). To date, no data has been published on the role of the other two orphan histidine kinases, CD1352 and CD1949, in sporulation. Interestingly, a recent finding by Oliveira *et al.* suggests the presence of additional factors involved in sporulation initiation, as they identified an orphan methyltransferase which when inactivated led to a decrease in sporulation (151). This recent finding underscores how the current knowledge of controls and triggers of sporulation in *C. difficile* is lacking. This is in spite of it being an important therapeutic target as shown by Srikhanta *et al.* who demonstrated that cephamycins, a group of β -lactam antibiotics, inhibited sporulation and in a mouse model of infection prevented infection recurrence (152).

1.3.3 Antimicrobial resistance

Multidrug resistance is a major feature of *C. difficile*, with strains resistant to antibiotics such as cephalosporins, clindamycin, fluoroquinolones and erythromycin found very commonly (153-158). *C. difficile* has been recognised as having a highly mobile genome which is thought to contribute to its ability to gain multidrug resistance with relative ease (159). As discussed above, the development of fluoroquinolone resistance is considered vital to the spread of the hypervirulent strains (12, 14), and resistance has since been shown at a high level to later generation fluoroquinolones such as moxifloxacin and gatifloxacin (160-162).

Current treatment recommendations make use of metronidazole, vancomycin or fidaxomicin for both primary and recurrent infection. Decreased susceptibility to metronidazole is not common although there have been outbreaks of particular concern, such as that in Israel in 2013 where 38 cases of decreased metronidazole susceptibility were recorded (163). The mechanism of resistance to metronidazole is not fully understood, however, acquisition of a plasmid was found to be associated with decreased susceptibility and when this plasmid was introduced into a previously susceptible strain, a 25-fold increase in MIC was observed (164). Fortunately, vancomycin resistance is extremely rare with very few cases reported worldwide (163, 165) and encouragingly, a recent longitudinal study in Europe found that *C. difficile* susceptibility to metronidazole and vancomycin is increasing, with the authors suggesting this could be owing to reduced use of these antimicrobials or greater strain diversity (166). As fidaxomicin is a much newer drug, it would be expected that resistance would take time to arise and thus far, promisingly, there have been only a few cases reported worldwide (166, 167).

1.3.4 Gene reporter systems in *C. difficile* for the study of virulence factors

The advantages of gene reporter systems for infection studies and investigation into virulence factors, such as toxin production and sporulation, when compared to qRT-PCR are numerous. Reporter systems are relatively low cost and are far more suitable for large scale screening and identification of elements such as -10 sites. Although a drawback to using gene reporter systems is the need to carry out cloning techniques to insert the reporters into either the genome or a plasmid so that they are under the control of the gene promoter of interest. Unfortunately, *C. difficile* has a number of features that make it incompatible with widely used reporters such as Green Fluorescent Protein. Many of these reporters have a requirement for oxygen to fully function (168), however, *C. difficile* is a strict anaerobe rendering it unsuited with these reporters. Another problem with *C. difficile* is its high level of background of green autofluorescence (169), therefore reporters that function at this wavelength are also unsuitable. Despite these difficulties, a number of reporters have recently been developed for use in *C. difficile* *in vitro* and *in vivo*, including SNAP-tag and CLIP-tag, *gusA*, *phoZ*, mCherry and light, oxygen or voltage domain (LOV) systems (170-174).

SNAP-Tag is a mutant version of the human DNA repair enzyme O⁶-alkylguanine-DNA alkyltransferase which can bind a number of fluorescent substrates. These commercially available substrates fluoresce at wavelengths ranging from 437 to 670 nm. This is advantageous as it means it is possible to avoid the wavelengths that would be affected by *C. difficile*'s high background of green autofluorescence (470 nm), and, in addition, near-infrared substrates have been described for use *in vivo*. The SNAP-tag has been used for both transcriptional fusions and translational fusions, where it is fused to a protein of interest which can then be visualised by microscopy to investigate protein localisation (170). Disadvantages of the SNAP-tag include a high background signal when quantified by absorbance (175) as well as the relatively high costs of the substrates. An alternative to SNAP-tag is CLIP-tag, which is a further mutated version of the human DNA repair enzyme O⁶-alkylguanine-DNA alkyltransferase, resulting in CLIP-tag binding O²-labeled benzylcytosine substrates where the SNAP-tag binds O⁶-

labeled benzylguanine substrates, however, fewer CLIP-tag substrates are currently available than with the SNAP-tag (176, 177).

gusA and *phoZ* are both colorimetric reporters, using genes from other bacteria to act as transcriptional reporters. *gusA* is a glucuronidase gene whose activity is possible to quantify by absorbance measurement of the breakdown of *p*-nitrophenyl- β -D-glucuronide. It has been used as a reporter in both *C. perfringens* and *C. difficile* (171, 178). The use of *gusA* fusions have been compared to the SNAP-tag in a recent publication by Mordaka *et al.* who found that *gusA* was both more sensitive and had a lower background signal than the SNAP-tag (175), however this was in *Clostridium acetobutylicum* not *C. difficile*. *phoZ* is an alkaline phosphatase from *Enterococcus faecalis* which has been demonstrated to be a viable system for use as a transcriptional reporter in *C. difficile* and like *gusA* is quantified by absorbance measurement (172). These two systems cannot be used as translational reporters as they cannot be visualised by microscopy however due to the visible colour change either on a plate or in solution they can be used for genetic screens where the SNAP-tag cannot be so easily used. In addition, these two reporters, in particular *phoZ*, are significantly cheaper and less complex to use than the SNAP-tag.

LOV domains are fluorescent proteins which have been demonstrated to be viable translational reporters in *C. difficile* (174). However, LOV based reporters have a significant drawback in that their fluorescence in the green spectrum overlaps with *C. difficile*'s natural green autofluorescence (174). This means that these reporters are not suited for the detection of small changes in fluorescence and so are unlikely to be successfully applied to the study of low abundance proteins in *C. difficile* (174). Alternatively, the mCherry reporter system avoids the problem of *C. difficile*'s green autofluorescence as the encoded protein fluoresces in the red spectrum. mCherry has been successfully used for protein localisation and gene expression studies, however, a major disadvantage of the mCherry system is

that it requires oxygen to develop fluorescence and therefore it is not compatible with *C. difficile* samples that have not undergone fixation anaerobically (173, 179).

1.4 Regulation of virulence factors in *C. difficile*

The regulatory control of virulence factors in *C. difficile* is vital to the survival, lifecycle, and pathogenesis of the bacterium. In addition to Spo0A, *C. difficile* has a number of important regulators including CcpA and CodY which are involved in controlling the expression of virulence factors such as toxin production and sporulation.

CodY is a global nutritional regulator which acts to repress alternative metabolic pathways when nutrients needed for growth are readily available, therefore CodY primarily functions during the exponential phase of growth. Inactivation of CodY was found to lead to four-fold up-regulated expression of 146 genes and the four-fold down-regulation of 19 genes, with up to 350 CodY binding sites identified in the *C. difficile* genome (180). CodY is found in several Gram-positive bacteria including *B. subtilis* (181, 182) and *Staphylococcus aureus* (183). It is bound by BCAAs (122) and GTP (123) which act as co-factors and allow CodY to bind to DNA with high affinity. As the cell progresses towards and enters stationary phase, these nutrients become less available and CodY is less able to bind DNA and act as a repressor leading to de-repression of genes involved in alternative metabolic pathways. As well as having a vital role in expression of genes relating to nutrient acquisition and usage, CodY has also been found to have a role in the regulation of both toxin synthesis (184) and sporulation in *C. difficile* (185). CodY represses production of Toxins A and B via its regulation of the toxin specific sigma factor TcdR (184). CodY also represses sporulation with a number of mechanisms proposed that achieve this including repression of Spo0A and SigE (185).

Another important regulator in *C. difficile* is CcpA which, similarly to CodY, responds to nutrient availability. CcpA regulates gene expression in *C. difficile* in response to glucose, approximately 18% of all *C. difficile* genes' expression are regulated by the availability of glucose, of these 18%

approximately 50% are regulated by CcpA (126). A key feature of CcpA in *C. difficile* is that it represses expression of Toxins A and B in response to phosphotransferase system (PTS) sugar availability, via the direct binding of the *tcdB* promoter region and the 5' end of the coding sequence for *tcdA* (125). In addition, CcpA has been found to be a regulator of sporulation through its repression of genes involved in the early phases of sporulation such as *spo0A* and *sigF* (126). This demonstrates the importance of the link between nutrient availability and the expression of key virulence factors.

In addition to the relatively well characterised regulators described above, recently a number of other regulators have been reported to play a role in virulence factor expression. RstA has been shown to repress toxin production and motility whilst promoting sporulation (129, 186). Additionally, the recently identified regulator leucine-responsive regulatory protein (Lrp) has been shown to repress both toxin production and sporulation in R20291 (187).

Transcriptional regulators are of significant importance to the lifecycle and pathogenesis of *C. difficile* by providing a link between nutrient availability and the control of the expression of major virulence factors. A greater understanding of features affecting virulence factor regulation in *C. difficile* may allow for the identification of treatment strategies which could be used to remove or generate particular nutrients and/or metabolites from *C. difficile*'s environment and therefore influence *C. difficile*'s expression of virulence factors. Two potential strategies for this could be: i) identification and inhibition of transporter systems responsible for nutrient uptake or ii) identification of bacterial species which could compete for or produce specific nutrients and/or metabolites.

1.4.1 The role of the *sin* locus in virulence factor regulation

A major feature in *C. difficile*'s lifestyle is the choice to undergo sporulation. Spores are a major virulence factor as they are the agents responsible for *C. difficile* transmission (138). Spo0A is the master regulator of sporulation and when inactivated, both the 630 and R20291 strains are unable to

produce spores (138). As discussed above, the function of Spo0A is dependent on its phosphorylation state (149), which has been demonstrated in *B. subtilis* to result in dimerisation of Spo0A which is thought to be its active form (188).

The *sin* (sporulation inhibitor) locus has been well described in *B. subtilis*; an organism often used as a model for *C. difficile*. In *B. subtilis*, the *sin* locus is formed of two genes: *sinR_{BS}* and *sinI_{BS}*. SinR_{BS} is a repressor of sporulation whilst also regulating expression of genes related to biofilm formation and cannibalism (189, 190). The role of SinI_{BS} is to bind to SinR_{BS} to form a heterodimer and prevent the formation of the active SinR_{BS} tetramer (191, 192) thus lifting repression of genes involved in sporulation. Another factor involved in genes affected by the actions of SinR_{BS} is SlrR, which has been found to bind to SinR_{BS} and form a complex which represses genes involved in both motility and cell separation (193). Therefore, this locus has been shown to be key in whether a bacterium undergoes sporulation as well as a host of other functions in lifestyle choice.

C. difficile has two genes which show high similarity to *B. subtilis*' *sinR_{BS}*: CD2214 (*sinR*) and CD2215, but *C. difficile* has no ortholog to *sinI_{BS}* (194). Both of these genes resemble *sinR_{BS}* with both containing Helix-Turn-Helix (HTH) domains as well as multimerisation domains. The presence of the HTH domains suggests these proteins function, at least in part, through DNA binding. Girinathan *et al.* showed that similarly to *B. subtilis*' SinR_{BS} SinR and CD2215 are pleiotropic regulators and they have roles in the control of sporulation, toxin production and motility in *C. difficile* (194). However, in contrast to *B. subtilis* where three promoters control expression of *sinR_{BS}* and *sinI_{BS}* Girinathan *et al.* provided evidence that *sinR* and CD2215 are expressed as an operon (194). The authors showed that a double mutant of both *sinR* and CD2215 in the R20291 strain was asporogenic, whilst a single mutant of CD2215 had a hyper-sporulation phenotype suggesting that SinR enhances sporulation and CD2215 represses sporulation (194). This was followed up by showing that CD2215 binds, via its multimerisation domain, directly to SinR to inhibit it and therefore prevent SinR's genetic regulation

activities (195). This shows that CD2215 functions similarly to *B. subtilis*' SinI_{BS} in that it directly binds SinR and in so doing inhibits SinR's effects on gene expression. However, this finding does not explain the purpose of CD2215's HTH domain and what effects it has on gene expression. A further interaction found between *sinR-CD2215* and sporulation was that Dhungel and Govind determined that Spo0A acts to repress expression from the *sinR-CD2215* promoter via two Spo0A binding sites found upstream of *sinR* (196).

There is some debate over the interaction between SinR and CD2215 and the regulator CodY. CodY is a global regulator that usually functions as a repressor but can also be an enhancer (197). Work by Nawrocki *et al.* showed contrasting roles for CodY when looking at the 630 Δ *erm* strain in comparison to the R20291 strain (185). They found that mutation of CodY led to increased expression from the *sinR-CD2215* operon in 630 Δ *erm*, however, mutation of CodY in R20291 led to a decrease in expression of *sinR* and *CD2215* (185). The upstream regions of *sinR-CD2215* are identical in the two strains so the authors concluded that other differences in CodY mediated regulation may explain the differences seen in the two strains. Nawrocki *et al.* also found, using a transcriptional reporter, that mutation of a single base pair in the putative CodY binding site upstream of *sinR* did not lead to altered expression from the *sinR* upstream region suggesting this site is not bound by CodY, although the authors did state that a single base pair mutation may not have been sufficient to prevent CodY binding (185). Interestingly, Girinathan *et al.* provided evidence that CodY binds directly to the upstream region of *sinR-CD2215* leading to downregulation of *sinR-CD2215* expression in R20291, but, unfortunately this experiment was not carried out in 630 Δ *erm* or UK1 (194) making comparisons between studies difficult. Therefore, whilst there is consensus that CodY has an effect on the locus, there is yet to be agreement on whether this effect is positive or negative and either direct or indirect.

Further research into this operon, including its activities and the mechanisms that control its expression, may provide opportunities to exploit it in order to direct *C. difficile* lifestyle choice towards

pathways that are disadvantageous for infection persistence and therefore reduce the likelihood of infection recurrence and transmission.

1.5 The microbiome and dysbiosis in CDI

As discussed, the health of the gut microbiome is strongly associated with CDI risk, but it is now recognised that dysbiosis can play a role in a variety of diseases including those caused by a pathogen, such as in CDI, but also in non-communicable diseases such as inflammatory bowel disease (reviewed by Kostic *et al.* (198)). The gut microbiome is one of the better characterised anatomical sites however much is still unknown. With the advent of high throughput sequencing technologies and techniques such as 16S sequencing, knowledge of the microbiome has advanced quickly, yet, despite this there are large gaps in the knowledge especially regarding what is considered a healthy human microbiome. This issue is now beginning to be addressed with the publication of large scale studies, such as those from the MetaHIT consortium and the human microbiome project. Data from the MetaHIT consortium showed that up to 1150 species have been identified as constituents of the microbiome and that each individual harbours at least 160 individual species (199).

A number of publications are now exploring how the microbiome is affected by antibiotics and how even a single dose can be sufficient to induce long lasting alterations (200). Buffie *et al.* showed in mice treated with clindamycin, and subsequently infected with *C. difficile*, that infected animals did not recover populations of *Akkermansia*, *Enterobacteriaceae* and Mollicute phlotypes (200). A number of studies have found similar effects in humans, with changes occurring in the short term but also in the long term. For example, Zaura *et al.* found that up to a year after treatment with a single dose of ciprofloxacin microbiome diversity remained decreased (201).

Not only does dysbiosis leave patients susceptible to CDI, but changes to the microbiome also occur throughout the course of CDI. A greater understanding of the changes during CDI may lead to

improved treatment through precise manipulation of the microbiome to restore colonisation resistance. Despite varying methods used to assess the composition of the microbiome and the level of taxonomic rank determined, a number of consistent findings are beginning to emerge within CDI. Bacteria within the Bacteroidetes phylum, including the family *Lachnospiraceae*, have been found to be depleted whereas families within the Firmicute phylum have consistently found to be increased, such as *Enterococcaceae* (202-207). In addition, there is recent evidence that the microbiome can predict both treatment failure and infection recurrence. Khanna *et al.* showed that patients who responded to treatment had increases in *Ruminococcaceae*, *Rikenellaceae*, *Clostridiaceae*, *Bacteroides*, *Faecalibacterium* and *Rothia* compared to patients who did not respond to treatment (208). In the same study, patients who suffered infection recurrence were associated with increases in *Veillonella*, *Enterobacteriaceae*, *Streptococci*, *Parabacteroides* and *Lachnospiraceae* (208). In support of this data, *Veillonella* was positively associated with *C. difficile* colonisation in a study by Crobach *et al.*, who also found that *Eubacterium hallii* and *Fusicatenibacter* may indicate resistance to colonisation (209). In contrast to the findings by Khanna *et al.* results from Lee *et al.* showed in patients with ulcerative colitis that for some taxa the opposite results were true, as decreased *Veillonella*, *Lachnospiraceae* and *Enterobacteriaceae* were associated with reduced *C. difficile* recurrence (210). It is clear that much more research is needed to elucidate predictive changes especially in populations with existing health conditions such as ulcerative colitis.

Further work is now beginning to characterise the mechanics behind the interactions between *C. difficile* and the microbiome, for example, *B. thetaotaiaomicron* has been shown to produce succinate which *C. difficile* can use for fermentation to butyrate (211). It has been demonstrated that precise manipulation of the microbiome can restore colonisation resistance. For example, Buffie *et al.* showed that adoptive transfer of *Clostridium scindens* to mice results in improved survival following challenge with *C. difficile* as a result of *C. scindens* synthesising secondary bile acids which inhibit *C. difficile* growth (212). This finding was followed up by Reed *et al.* who showed transfer of the *bai* locus,

responsible for secondary bile acid synthesis in *C. scindens*, to four other commensal *Clostridium* species could inhibit *C. difficile* (213). As a result of findings such as this, a number of probiotics have been investigated as potential prophylactics or CDI treatment options. There has been a particular focus on administration of *Lactobacillus* species, and evidence is accumulating that this may be an effective strategy however there is no consensus on the dosage and strains that make the most effective probiotic (214-216) and as such there is currently no recommendation for their use clinically (47).

1.5.1 The role of *p*-cresol in CDI and potential disruption of the microbiome

p-cresol is a phenolic antibacterial compound produced by *C. difficile*, through the fermentation of tyrosine via the intermediate *para*-hydroxyphenylacetate (*p*-HPA) (217) (Fig. 1.1). Recent evidence, has shown that the mutation of a tyrosine transporter leads to a decrease in *p*-cresol production, providing further evidence for the conversion of tyrosine with *p*-cresol as an end product in *C. difficile* (218). The pathway by which tyrosine is converted to *p*-HPA is currently unknown, however it has been shown that *p*-HPA is converted to *p*-cresol by the actions of *p*-HPA decarboxylase (HpdBCA), which is encoded by the *hpdBCA* operon (219). Each gene of the *hpdBCA* operon encodes a subunit of *p*-HPA decarboxylase and mutation of any of these genes renders *C. difficile* unable to produce *p*-cresol (220). To date, very little is known about factors that control or trigger *p*-cresol production. It has been shown that *p*-cresol production is below the limit of detection, when tested by mass spectrometry and nuclear magnetic resonance (NMR), following *C. difficile* culture in rich media (Brain Heart Infusion) whereas it is detected after *C. difficile* growth in less rich media (Yeast Peptone) (221). Others have reported that production of *p*-cresol requires the presence of *p*-HPA in the *C. difficile* growth media (219, 222).

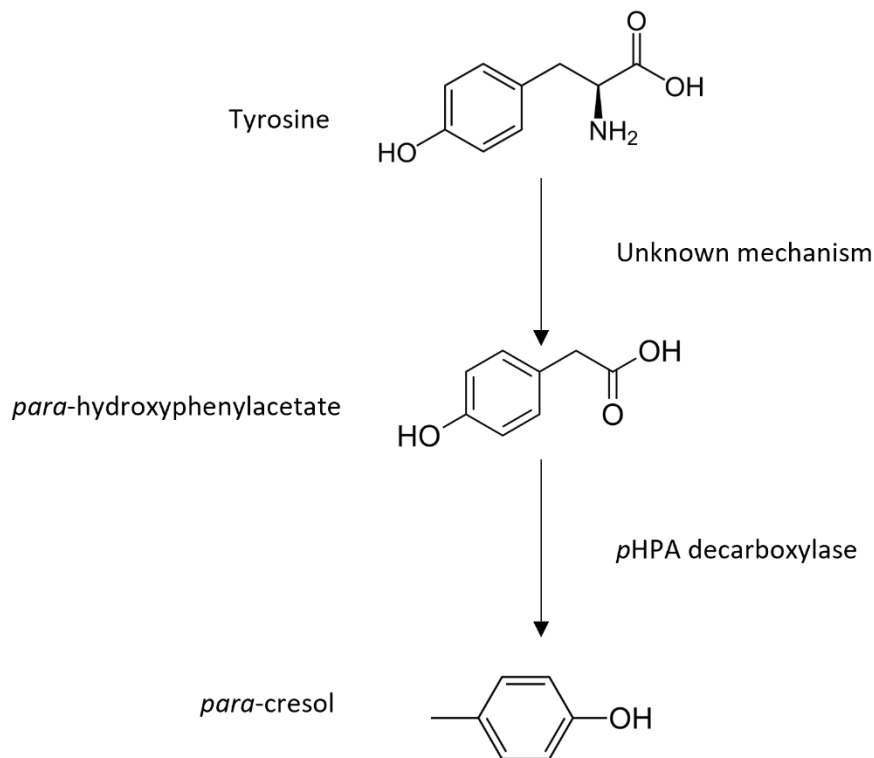


Figure 1.1. Diagram of the pathway of *p*-cresol production in *C. difficile*

It has been shown the hypervirulent R20291 strain (RT027, clinical isolate Stoke-Mandeville, 2006) both produces more *p*-cresol and is more tolerant to *p*-cresol than the 630 strain (ribotype 012, clinical isolate Zurich, Switzerland 1982) (220). This would suggest that increased *p*-cresol production may play a contributory role in virulence. *p*-cresol tolerance is not linked to production as it has been demonstrated that species that do not produce *p*-cresol can be more tolerant to its effects than *C. difficile*, such as *C. perfringens* and *Enterococcus faecium* (221). In recent work by Passmore *et al.*, it was demonstrated that *p*-cresol allows *C. difficile* to compete more effectively against other bacteria, with Gram-negative bacteria being particularly susceptible to *p*-cresol (223). Furthermore, in a mouse model of CDI relapse, mice infected with a mutant of *hpdC*, unable to produce *p*-cresol, showed significantly greater microbiome diversity seven days after infection compared to mice infected with the wild type 630 Δ *erm* (223). In addition to an overall effect on diversity, changes were found in the composition of the microbiome between the wild type and *hpdC* mutant consistent with findings that

Gram-negative species are more susceptible to *p*-cresol (223). In the wild type infected mice there was an increase in Firmicutes such as Erysipelotrichales, whereas in the *p*-cresol mutant infected mice there was an increase in Proteobacteria such as Enterobacteriales and Gammaproteobacteria (223). These findings are in line with the current literature as *p*-cresol tolerance is somewhat consistent with changes found in the microbiome. For example, bacteria that are highly tolerant, such as *Enterococci*, are found at significantly increased levels during CDI (202, 204-207). Interestingly, subjects who carry *C. difficile* asymptotically are also more likely to be more densely colonised with *Enterococci* (55).

Recently, three other intestinal species have been shown to have the ability to produce *p*-cresol in concentrations comparable to *C. difficile*; *Blautia hydrogenotrophica*, *Romboutsia lituseburensis* and *Olsenella uli* (224). Each of these three species have been shown to have an ortholog of the *hpdbCA* operon suggesting this operon is vital to production of high levels of *p*-cresol (224). Direct comparisons are difficult due to major differences in media used by Saito *et al.* and Passmore *et al.* however the amounts of *p*-cresol produced show large differences. Saito *et al.* show production by *C. difficile* of approximately 400-500 μM (224) whereas Passmore *et al.* shows production of up to approximately 25 mM when the growth media has been supplemented with the *p*-cresol precursor *p*-HPA (223). Therefore, whilst these other species have demonstrated the capacity to produce *p*-cresol whether they are capable of levels close to *C. difficile* remains to be determined.

These previous studies support the notion that *p*-cresol may play an important role in the pathogenesis of *C. difficile* by altering the gut microbiome or by perhaps maintaining dysbiosis. This would suggest that inhibition of *p*-cresol production may be a viable strategy to reduce the effect *C. difficile* has on the gut microbiome leading to a decrease in the time taken for colonisation resistance to recover, thereby reducing the chance of infection relapse.

1.6 Aims and objectives

The overall aim of this project was to characterise factors involved in the regulation of expression of genes associated with virulence factors. The main objectives of my thesis are on two different virulence systems in *C. difficile*: a) Regulation of *p*-cresol production and b) characterising the regulation of the pleiotropic regulators SinR and CD2215. I sought to identify and characterise the environmental factors involved in *p*-cresol production and the response mechanisms to these environmental factors. Additionally, my aim was to characterise the genetic regulation and expression of pleiotropic regulators SinR and CD2215, to unravel their roles in virulence.

a) Regulation of *p*-cresol production

I. Identify the triggers of *p*-cresol production

- Characterise the *hpdBCA* operon by identification of the promoter site(s) and the regulatory factors which control expression level.
- Use transcriptional reporter constructs to identify environmental conditions which lead to enhancement or repression of *hpdBCA* operon expression.
- Compare and contrast the efficacy of different transcriptional reporters to assess gene regulation in *C. difficile* (sensitivity, expense, and ease of method).

II. Characterise the effect of tyrosine metabolism on *C. difficile* and representatives of the beneficial gut microbiota

- Explore the effect of the tyrosine metabolite *p*-HPA on growth and viability of *C. difficile* and other species of the gut microbiota.
- Explore potential mechanisms of *p*-HPA toxicity.
- Assess the conservation of *p*-cresol production across *C. difficile* lineages.

b) Characterise regulation of the pleiotropic regulators SinR and CD2215

I. Identify regulatory controls of SinR and CD2215

- Utilise reporter constructs to investigate regulatory controls of *sinR* and *CD2215* expression.

- Unravel the mechanism of action of these regulatory factors on *sinR* and *CD2215* expression.

Chapter 2: Materials and Methods

All materials used in this study are listed in the appendices as follows: bacterial strains (Table 1), plasmids (Table 2), primers (Table 3), growth media (Table 4) and reagents (Table 5).

2.1 *C. difficile* growth

C. difficile strains were routinely grown in Brain Heart Infusion (BHI) (Oxoid) supplemented with 5 g/L yeast extract (Sigma), 0.05% L-Cysteine (Sigma) (BHIS). Strains were grown at 37 °C under anaerobic conditions, in a Modular Atmosphere Control System 500 (Don Whitney Scientific) with all media undergoing at least four hours pre-equilibration prior to use. Selection of *C. difficile* strains with plasmids derived from pMTL84151 was achieved by growth with thiamphenicol at 15 µg/ml. Transformant *E. coli* strains were routinely grown at 37 °C in Luria-Bertani (LB) agar or broth with 25 µg/ml or 12.5 µg/ml chloramphenicol respectively. All strains used in this study are listed in appendix table 1.

2.2 DNA manipulation

2.2.1 Polymerase chain reaction

Primers used in this study are listed in appendix table 3. Before ordering, all primers were checked for appropriate annealing temperatures using Integrated DNA Technologies (IDT) Oligo Analyser tool (<https://eu.idtdna.com/pages/tools/oligoanalyzer>). Subsequently, all primers were ordered from IDT. DNA amplification for cloning was carried out using DNA Phusion (New England Biolabs) according to manufacturer's instructions with alterations made, where necessary, to annealing temperature and extension time. DNA amplification for screening was carried out using MyTaq™ DNA polymerase (Bioline) according to manufacturer's instructions. PCR reactions were carried out using DNA Engine Tetrad 2 thermocycler (Bio-Rad Laboratories Inc.) or where a gradient PCR was required, a GS4 thermocycler (G-Storm) was used.

DNA purification from these reactions was by either 1% agarose gel electrophoresis, as described below, followed by extraction with the Monarch DNA Gel extraction kit (NEB) or by using the Monarch PCR and DNA Cleanup kit (NEB) as per manufacturer's instructions.

2.2.2 Agarose Gel Electrophoresis

Gel electrophoresis was carried out using 1% w/v agarose with a 1:20000 dilution of GelRed (Biotium). The running buffer used was tris-acetic acid-ethylenediaminetetraacetic acid (EDTA) (TAE): 40 mM Tris, 20 mM acetic acid and 1 mM EDTA. Electrophoresis was run at 400 A, 100 V for 50 minutes. Samples were prepared with 6X loading dye (NEB). Visualisation of gels was done with either Gene Genius Bio Imaging System (Syngene) or G:Box F3 (Syngene).

2.2.3 DNA Extraction

Genomic DNA extractions for screening purposes was carried out by Chelex extraction as follows: 1 ml of an overnight culture was pelleted, and the supernatant discarded. The pellet was resuspended in 300 μ l 5% Chelex100 (Sigma) and then heated at 100 °C for 10 minutes before being placed on ice for 10 minutes. The sample was pelleted and 200 μ l of the supernatant was stored at -20 °C for future use.

Extraction of plasmid DNA from *E. coli* strains was done using the Monarch Miniprep kit (NEB) according to manufacturer's instructions.

Where high quality DNA or RNA were required phenol-chloroform extraction was used. Samples were grown for the required length of time to reach the desired growth stage for testing. Once this was reached, cultures were transferred to a 50 ml conical falcon tube. RNA Protect (Qiagen) equal to the culture volume was added before pelleting at 4500 x *g* at 4 °C for 20 minutes. The supernatant was removed, and the pellets stored immediately at -80 °C. All centrifugation steps that follow were carried out at 17000 x *g* at 4 °C for the indicated amount of time. Extraction was carried out by re-suspending the pellet in 1 ml of RNAPro solution (MP Biomedicals) for every 5 ml of culture used to

make the pellets. The pellet in RNAPro solution was transferred to a Lysing Matrix B tube (MP Biomedicals) and then ribolysed for 40 seconds at 6.0 m/s using a FastPrep-24 Classic Instrument (MP Biomedicals). Samples were immediately centrifuged for 10 minutes. 600 µl was removed from the Lysing Matrix B tube and put in to a fresh 1.5 ml microcentrifuge tube and 300 µl chloroform was added. The samples were vortexed for 15 seconds and then centrifuged for 15 minutes. The upper phase of the sample was removed and transferred to a fresh microcentrifuge tube containing 500 µl 100% ethanol before freezing at -20 °C overnight to precipitate the nucleic acids. Following precipitation, samples were centrifuged for 15 minutes and the supernatant discarded. The pellet was re-suspended in 70% ethanol before centrifugation for five minutes, the supernatant was removed, and the pellets allowed to air-dry for up to 10 minutes. Pellets were re-suspended in 50 µl of DEPC-treated water and stored at -80 °C pending further processing.

2.3 RNA extraction

Where RNA was required, phenol-chloroform extractions were carried out as described above. Following re-suspension in DEPC-treated water, the samples underwent DNase treatment. 50 or 25 µl of extracted material was added to 20 U Turbo DNase I (ThermoFisher), 40 U RNasin plus RNase inhibitor (Promega), 20 mM magnesium sulphate, 90 mM sodium acetate and DEPC-treated water up to 150 µl. Samples were incubated for one hour at 37 °C in a PCR thermocycler before addition of 20 U Turbo DNase and 40 U RNasin plus and a further incubation for one hour at 37 °C.

Following DNase treatment, samples underwent RNA purification using the RNeasy kit (Qiagen). Samples were added to 350 µl RLT buffer and 200 µl 100% Ethanol and then applied to a RNeasy column and centrifuged for 15 seconds at 12,800 x *g*. 500 µl RPE wash buffer was applied to the column and centrifuged, this wash step was repeated once. The column was dried by centrifugation for two minutes. The RNA was eluted by addition of 17 µl DEPC-treated water, the column was allowed to stand for three minutes before a one-minute centrifugation, the elution step was repeated twice to give an approximate total elution volume of 50 µl.

Eluted RNA was tested for quality and concentration using a nanodrop 1000 instrument (Thermo scientific). In addition, to ensure DNA had been fully removed, a 1:1000 dilution of the RNA was used to set up a PCR which was visualised via agarose gel electrophoresis, if a positive result was seen then the DNase treatment was repeated as above.

2.3.1 cDNA Synthesis

The first stage of cDNA synthesis was primer annealing; carried out by addition of up to 1 µg RNA to 5.5 µM random primers and made up to 11 µl with DEPC-treated water. Samples were incubated at 65 °C for 10 minutes before being allowed to cool at room temperature for 30 minutes.

The second stage of cDNA synthesis was reverse transcriptase (RT) treatment. Samples were treated with either superscript II or superscript IV RT. Reactions for superscript II were set up by addition of the samples to 1X first strand buffer, 10 mM dTT, 2.5 mM dNTPs, 20 U RNasin plus and 200 U superscript II. Reactions for superscript IV were set up by addition of the sample to: 1X SSIV buffer, 0.5 mM of each dNTP, 5 mM DTT, 20 U RNase plus and 200 U superscript IV. For each sample, two reactions were prepared, one with superscript II RT or superscript IV RT (ThermoFisher) and one without RT. For those treated with superscript II both RT positive and negative reactions were incubated for 1 hour at 42 °C after which samples were immediately frozen at -20 °C. For those samples treated with superscript IV samples were incubated at 23 °C for 10 minutes, 55 °C for 10 minutes and 80 °C for 10 minutes followed by immediate freezing of samples at -20 °C.

2.4 Cloning techniques

2.4.1 Restriction-ligation cloning

Restriction digest of plasmids and the sequence to be inserted into the plasmid was done using the appropriate enzyme incubated with the target sequences for 1 hour at 37 °C as per manufacturer's instructions. This was followed by dephosphorylation of the vector by Antarctic phosphatase (NEB) and ligation by T4 ligase (NEB) as per manufacturer's instructions.

2.4.2 Gibson assembly

Where Gibson assembly was used for plasmid construction primers were designed using the NEBuilder assembly tool (<https://nebuilder.neb.com/>). PCRs were carried out as per standard Phusion protocol described above, followed by agarose gel electrophoresis and extraction using the Monarch Gel extraction kit (NEB). Assembly used the Hifi kit (NEB) as per manufacturer's instructions with incubation of the mastermix with the PCR products for 1 hour at 50 °C. Reactions were used to transform *E. coli* via electroporation as described below.

2.4.3 Site directed mutagenesis

Site directed mutagenesis (SDM) was used to make alterations to a number of reporter plasmids. For SDM, primers were designed so that the base pair(s) to be altered were encoded on overlapping sections of both primers to maximise efficiency (225). PCR was carried out as per standard protocol and visualised by agarose gel electrophoresis. Successful reactions were treated by Kinase-Ligase-DpnI (NEB) as per manufacturer's instructions before transformation into *E. coli* via heat shock as described below.

2.4.4 DNA sequencing

All DNA sequencing was carried out by either Source Bioscience or Eurofins Scientific Sanger Sequencing services, with DNA concentrations prepared per the services requirements.

2.5 Cell Transformation

2.5.1 *E. coli*

Transformation of *E. coli* strains was carried by either electroporation or heat shock. Electroporation was carried out by addition of 2 µl of a prepared plasmid to an aliquot of electrocompetent cells. Electroporation was carried out at 2.5 kV using a Gene Pulser Xcell (Bio-Rad Laboratories Inc.) and cells recovered using 700 µl of super optimal broth with catabolite repression (SOC) (Thermofisher) for one hour at 37 °C, shaking incubation. Heat shock was performed by the addition of 2 µl of a prepared plasmid to an aliquot of competent cells, the cells were incubated on ice for 30 minutes followed by incubation in a water bath at 42 °C for 30 seconds and then cooling on ice for 5 minutes. After cooling, the cells were recovered by addition of 950 µl SOC to the cells and incubated with shaking at 37 °C for

one hour. Following recovery, 100 µl of cells in SOC were spread plated onto LB plates with chloramphenicol (25 µg/ml) to select for transformant colonies.

2.5.2 *C. difficile*

Transformation of *C. difficile* with plasmid DNA was carried out by conjugation. The plasmids were electroporated into the donor strain of *E. coli* CA434 as outlined above. Conjugation was first done by growing the donor CA434 strain aerobically overnight in LB with chloramphenicol (15 µg/ml) and the recipient *C. difficile* strain in BHIS broth anaerobically. 1 ml of the CA434 strain was pelleted at 4000 x g for 2 minutes, the supernatant discarded, and the pellet placed into the anaerobic workstation. 200 µl of the *C. difficile* culture was heated at 52 °C for 5 minutes to increase conjugation efficiency (226). The *C. difficile* culture was then used to re-suspend the pellet of CA434 and the cells were plated on to non-selective BHI plates and grown for 24 hours. After 24 hours, the cells were scraped from the plate and re-suspended in 500 µl phosphate buffered saline (PBS). 100 µl aliquots of the PBS was spread plated on to BHIS with cycloserine (250 mg/L), cefoxitin (8 mg/L) and thiamphenicol (15 mg/L) (CCTm) plates and left to grow for 72 hours to select for transformant *C. difficile* colonies. After 72 hours, if colonies had grown they were re-streaked on to fresh BHIS CCTm to confirm successful conjugation.

2.6 qRT-PCR

qRT-PCR reactions were set up using the Kapa Sybr Fast kit (Kapa Biosystems). Appropriate dilutions of cDNA were made for use in the PCR to ensure Ct values could be accurately measured. Reactions were set up with 2 µl sample, 1X Kapa Sybr Fast mix, 200 nM of each primer and nuclease free water up to 20 µl. All qRT-PCRs were run on the 7500-Real time PCR system (Applied Biosystems) with cycling conditions as per manufacturer's instructions for the Kapa Sybr Fast kit. Each sample was tested in technical triplicate.

2.7 Electromobility Shift Assays (EMSAs)

EMSAs were carried out using purified recombinant protein with recombinant SinR provided by Janina Muench, PhD student in the Isaacson laboratory at King's College London. PCR products were

amplified using IRDye700 labelled primers as per standard PCR protocols and purified by Monarch PCR cleanup kit (NEB) as per manufacturer's instructions. Binding reactions were carried out using the Odyssey® EMSA kit (LICOR) with the following set up: 1X binding buffer (10 mM Tris, 50 mM KCl, 1 mM DTT, pH 7.5), 5 mM DTT, 0.5% Tween20, 0.1 µg/µl Poly (dI•dC) (dissolved in 10 mM Tris, 1 mM EDTA, pH 7.5) and 1 µl of PCR product diluted 1:10 with nuclease free water, the remaining volume up to 20 µl used the maximum available protein concentration i.e., 20 ng/ml SinR or approximately 24 ng/ml CD2697. Reactions were incubated at room temperature for 30 minutes in the dark before addition of 1X orange loading dye and then 18 µl of the binding reaction was loaded on to a 6% DNA retardation gel (ThermoFisher) and then run on ice in the dark for 90-120 minutes at 70V in 1X tris-boric acid-EDTA (TBE) buffer being visualised using on an Odyssey 9120 imager (LICOR). All assays were carried out a minimum of three times.

2.8 Reporter assays

2.8.1 SNAP-tag assay

Culture volumes varied depending on requirements of the experiment, however, all samples subsequently underwent testing for SNAP-tag expression by the same method. Cultures were incubated in the dark for 30 minutes at 37 °C following the addition of the SNAP-Tag substrate TMR-Star (NEB) at 0.2 nM. These cultures were pelleted and washed with 5 ml PBS before re-pelleting, the pellet was then re-suspended in 600 µl PBS and transferred to a Lysing Matrix B tube. Samples were ribolysed for 40 s at 6.0 m/s twice using a FastPrep-24 Classic instrument (MP Biomedicals), before pelleting at 17,000 x *g* for 2 minutes to remove cell debris, 300 µl of supernatant was removed and frozen at -20 °C for later testing by SDS-PAGE.

SDS-PAGE for the SNAP-tag assays was carried out using 10% Bis-Tris gels (Thermofisher) using 3-(*N*-morpholino) propane sulfonic acid (MOPS) buffer (Thermofisher) for 40 minutes at 180 V and 400 A. Gels were imaged using a Typhoon Trio Variable Mode Imager System (GE Healthcare) set to detect fluorescence at 580 nm. The bands of interest were quantified using ImageQuant TL Image Analysis software (GE Healthcare). Limit of detection (LOD) was calculated as the average of the results of the

following formula for all samples across the biological replicates: background pixel volume/culture OD_{590 nm}. Background pixel volume was obtained by measuring pixels from an equal size gate directly above that of the SNAP-tag band. All assays were carried out with a minimum of three biological replicates.

2.8.2 *gusA* assay

630Δ*erm* P_{hpdB}-*gusA* were grown to the required stage and then 1.5 ml of culture was pelleted, the supernatant discarded and the pellets frozen for later testing as per the method outlined by Mordaka and Heap (175). The only modification made to the method was the OD_{590 nm} taken after the four-hour growth stage was used to adjust for growth instead of measurement from the re-suspended pellet. Assays were carried out with three biological replicates.

2.8.3 *phoZ* assay

Assays were performed as described by Edwards *et al.* (172). The only modification made was that OD_{590 nm} taken after the growth step was used to normalise the experiments, rather than measuring the optical density from the re-suspended pellet. *phoZ* reporter expression was quantified using the formula: $[\text{OD}_{420 \text{ nm}} - (\text{OD}_{550 \text{ nm}} \times 1.75)] \times 1000 / t(\text{min}) \times \text{OD}_{590 \text{ nm}} \times \text{vol. cells (ml)}$. OD_{420 nm} and OD_{550 nm} readings were made using a Spectramax M3 Multi-Mode plate reader (Molecular Devices). All assays were carried out with a minimum of three biological replicates.

2.9 Phenotypic assays

2.9.1 Growth curves

All growth curves were set up using the same initial method. Each strain or species to be tested were grown anaerobically overnight in BHIS broth. The following morning the cultures were back diluted to OD_{590 nm} of 0.5 in pre-equilibrated fresh BHIS media, thereafter 1 ml of the back diluted culture was added to 10 ml of the test conditions. OD_{590 nm} was taken every hour for eight hours with a final additional reading taken at 24 hours. Where the final pH of the media was required following completion of growth curves, the cultures were filter sterilised using 0.22 μm filters (Millipore) and

the supernatant's pH was tested using a Mettler Toledo Seveneasy pH Meter. All growth curves were carried out with a minimum of three biological replicates.

2.9.2 High pressure liquid chromatography (HPLC)

To determine the concentrations of tyrosine, *p*-HPA and *p*-cresol in cultures HPLC analysis of samples was carried out.

Samples analysed in chapter three were prepared as follows: 630 Δ *erm* was grown overnight after inoculation of a single colony in to 10 ml BHIS. The following morning 2.5 ml of the overnight BHIS culture was added to 7.5 ml of minimal media (MM) and incubated for three hours. MM was prepared using a combination of recipes from Karasawa *et al.* (227) and Cartman *et al.* (228). The recipe used the salt, trace salt, vitamin and iron sulphate solutions as per Cartman *et al.* (228) whilst the amino acid solution was prepared according to Karasawa *et al.* (227) with the exception of the glycine being increased to 400 mg/l, tyrosine being included only as indicated, and glucose solutions were omitted completely. After three hours cultures were pelleted (centrifugation at 4000 x *g* for 5 minutes) and washed using 5 ml fresh MM before re-pelleting. Pellets were resuspended in 1 ml of MM, and 100 μ l of this suspension was added to 10 ml of each of the following test conditions: MM only, MM plus 2 mg/ml *p*-HPA, and MM plus 400 mg/l tyrosine and grown for 10 h at which point 1 ml samples were taken as well as measurement of the OD_{590 nm} to allow for concentrations to be normalised to growth. Supernatants were obtained from the cultures by filter sterilisation of 1 ml of the sample using 0.22 μ m filters (Millipore) and frozen at -80 °C until testing by HPLC. Supernatants were prepared in three biological replicates.

In chapter four, samples prepared for testing the five strains representing each clade of *C. difficile* (630 Δ *erm*, R20291, CD305, M68 and M120) as well as the Δ *codY* mutant and its parent 630 Δ *erm* strain were grown overnight in MM which was made using a combination of recipes from Karasawa *et al.* (227) and Cartman *et al.* (228). The salt, trace salt, vitamin and iron sulphate solutions were made as per Cartman *et al.* (228) whilst the glucose and amino acid solutions were prepared according to

Karasawa *et al.* (227) with the exception of the *p*-tyrosine concentration which was increased to 400 mg/l. Each strain was grown overnight in MM and then back diluted the following day to an OD_{590 nm} of 0.5, 1 ml of which was used to inoculate 10 ml of MM and MM with 2 mg/ml *p*-HPA. Strains were grown for 8 hours with 1 ml samples taken after 4 and 8 hours and at each time point the OD_{590 nm} was taken to allow for *p*-HPA and *p*-cresol production to be normalised to growth. Samples for HPLC analysis were filter sterilised using 0.22 µm filters before freezing at -80 °C prior to analysis. Supernatants were prepared with a minimum of three biological replicates.

All HPLC quantification was carried out by Dr Harparkash Kaur as per the following methods: Defrosted culture supernatants were mixed in a 1:1 ratio with methanol-water, transferred to HPLC tubes, and processed immediately by HPLC. Separations were performed utilizing an Acclaim 120 (Thermo Fisher) C18 5 µm analytical (4.6 by 150 mm) column and a mobile phase consisting of ammonium formate (10 mM [pH 2.7]) and menthol (vol/vol; 40:60) at a flow rate of 2 ml/min. Tyrosine, *p*-HPA, and *p*-cresol were detected using a photodiode array detector (UV-PDA; DAD 3000) set at 280 nm. The peak identity was confirmed by measuring the retention times of commercially available tyrosine, *p*-HPA, and *p*-cresol and determination of the absorbance spectra using the UV-PDA detector. A calibration curve of each compound was generated by Chromeleon (Dionex Software) with known amounts of the reference standards (0 to 100 mg/ml) in methanol-water (1:1 [vol/vol]) injected onto the column, from which the concentrations in the samples were determined. Concentrations of the compounds were normalised to growth by division of the concentrations by the OD_{590 nm} of when the sample was taken.

2.9.3 Colony forming units

To assess sporulation and total cell counts in a bacterial culture in the presence of *p*-HPA (see chapter four), strains were grown overnight in BHIS before back dilution in fresh pre-equilibrated BHIS to an OD_{590 nm} of 0.5, of which 1 ml was added to 10 ml of BHIS with 0, 1, 2 and 3 mg/ml *p*-HPA and incubated for 24 hours after which two 1 ml aliquots of the culture removed. One aliquot was heated at 65 °C for 20 minutes to leave only spores present, the other aliquot was left untreated to provide a total cell

count. Both aliquots underwent serial ten-fold dilutions in 1X PBS and 10 µl of each serial dilution was plated in triplicate onto a quarter of a BHIS plate supplemented with 0.1% w/v sodium taurocholate (Sigma Aldrich). Plates were incubated for 24 hours anaerobically and then CFUs counted. The lowest dilution where colonies could be distinguished from one another was counted. The CFU was calculated by using the colony count and multiplying by the dilution factor, then multiplying by 100 to adjust counts to provide the CFU per ml. CFUs were performed in a minimum of three independent replicates.

2.9.4 Phosphate release assay

The phosphate release assay was carried out as per Passmore *et al.* (223). Briefly, strains to be tested were grown overnight before 5 ml of the culture underwent centrifugation for 5 minutes at 4500 x *g*. The supernatant was removed, and the pellets washed twice in 5 ml Tris-buffered saline (TBS) (50 mM Tris-HCl pH 7.5, 150 mM NaCl). The pellets were suspended in 5 ml TBS for a third and final time and then normalised to an OD_{590 nm} of 1.0. 500 µl aliquots were pelleted at 17000 x *g* for 5 minutes and then re-suspended in TBS supplemented with 0, 1, 2 or 3 mg/ml *p*-HPA. In addition, an aliquot of each bacteria was re-suspended in TBS and boiled for 15 minutes to provide a maximum phosphate release. 100 µl of each test condition was taken after 30, 60 and 90 minutes at which point they were pelleted at 17000 x *g* for 5 minutes. In a 96-well plate, 30 µl of the supernatant was added to 170 µl of water and 30 µl of phosphate reagent from the Phosphate Release kit (Abcam) with the plate being kept in the dark between sample additions. After the final 90 minute sample was exposed to the phosphate release reagent, the plate was incubated for 30 minutes at room temperature in the dark. Phosphate release was assessed by OD at 650 nm measured by an M3 Spectramax multi-mode plate reader. To determine if the effect of *p*-HPA's acidity on the tested strains led to phosphate release the pH of the TBS with 1, 2 and 3 mg/ml *p*-HPA was measured to be pH 6.6, 4.1 and 3.8 respectively, and the phosphate release assays were repeated in TBS with matched pHs to the *p*-HPA samples, with pH adjustment carried out with hydrochloric acid. Assays were carried out with a minimum of three biological replicates.

2.9.5 Confocal microscopy

Microscopy was carried out using $P_{\text{HpdB-CDS}}$ -SNAP conjugated into 630 Δ erm and 630 Δ erm *hpdB::CT*. This plasmid carried HpdB fused to a SNAP-tag via a linker under the control of the *hpdB* promoter. Confirmation of translation of HpdB fused to the SNAP-tag was carried out by anti-SNAP western using standard western blotting methods as described below. SDS-PAGE prior to western blotting was carried out in duplicate with one gel used for western blot and the second used following the western blot for excision of the appropriately sized band to be tested by mass spectrometry at King's College London's Centre of Excellence for Mass Spectrometry (MS). MS was done as previously (229), with the following modifications: 1) 75 μ M C18 column (50 cm length) was used rather than 75 μ M C18 column (15 cm length), 2) Xcalibur software used was v4.4.16.14, 3) Proteome Discoverer software used was v2.5, and 4) Scaffold 5 software used was v5.0.1. Analysis of MS results was carried out in Scaffold software v5.0.1 and compared to the Uniprot All Taxonomy database. Six peptides were identified as being from *C. difficile* strain 630's HpdB, four with 100% identity and two with 99% identity providing coverage of 5.9% of HpdB's amino acid sequence, peptides identified by MS are listed in chapter four, supplementary table 1.

To prepare slides for microscopy, overnight cultures of 630 Δ erm $P_{\text{HpdB-CDS}}$ -SNAP were back diluted to an $OD_{590 \text{ nm}}$ of 0.5 and 1 ml added to 10 ml BHIS +/- 2 mg/ml *p*-HPA. These cultures were grown for four hours before 500 μ l was removed and added to 1 μ l of 50 μ M TMR-Star and incubated anaerobically at 37 °C in the dark for 30 minutes. Following incubation, the samples were pelleted at 4000 x *g* for 2 minutes and washed twice with 500 μ l PBS, after the second wash approximately 20 μ l PBS was left in the tube and used to re-suspend the pellet. A 1 μ l loop was used to spread the culture on to a glass slide, which was allowed to air dry for 1 minute. 10 μ l of Vectashield with DAPI was added to the dried slide with a coverslip placed over the top and sealed with clear nail polish. The slides were imaged under oil immersion using a Zeiss LSM-800 microscope (100X objective). Excitation and emissions used for the dyes was 358 and 463 nm for DAPI and 578 and 603 nm for TMR-Star,

respectively. Image analysis was carried out using Zeiss Zen Blue software. A minimum of three fields of view per slide were imaged for each of the three biological replicates.

2.9.6 Preparation of samples for quantification of SinR and CD2215 by western blot

Samples were prepared by overnight culture of 630 Δ erm P_{sinR}-SinR-His and 630 Δ erm P_{sinR}-CD2215-His being back diluted to an OD_{590 nm} of 0.5, 3 ml of which was added to 30 ml BHIS. 10 ml of sample was taken when growth reached an OD_{590 nm} of 0.6 as well as after 24 hours, the sample was pelleted by centrifugation at 4500 x *g* for 15 minutes and the pellet frozen at -80 °C until His-tag purification carried out as follows: Pellets were resuspended in 2 ml of binding buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 25 mM imidazole) and transferred to a lysing matrix B tube and then ribolysed at 6.0 m/s for 40 s. Samples were centrifuged at 17,000 x *g* for 1 minute and then 700 μ l of supernatant was added to 100 μ l NiNTA resin followed by mixing for 30 minutes on a sample rotator at 4 °C. Samples were recentrifuged and then washed by addition of 1 ml binding buffer and mixed for 15 minutes, this wash step was repeated twice. After the third wash the resin was eluted with 100 μ l elution buffer (50 mM Tris, 300 mM NaCl, 300 mM imidazole) and this was stored at 4 °C until a western blot could be performed.

2.9.7 Western blotting

For each test condition 15 μ l of the prepared sample was added to 5 μ l 4X loading dye (ThermoFisher) and the sample heated for 5 minutes at 95 °C before loading on to a 10% Bis-Tris gel run at 180V 400A. Transfer to a nitrocellulose membrane was carried out using the iBlot system as per manufacturer's instructions. The membrane was washed in 1X phosphate buffered saline (PBS) Tween20 (0.1%) for 5 minutes before undergoing blocking with blocking buffer (5% whole milk in PBS Tween20 (0.1%)). 10 ml blocking buffer containing an appropriate dilution of primary antibody was added to the membrane and incubated for one hour at room temperature. The membrane was washed three times using PBS Tween20 (0.1%) for 5 minutes. 10 ml PBS Tween20 (0.1%) containing a 1:10,000 dilution of IRDye secondary antibody (LI-COR) was added to the membrane and incubated for one hour before being

washed in PBS Tween20 (0.1%) for 5 minutes, a total of three washes were carried out. The membrane was visualised and imaged at the 800 wavelength on a LI-COR Odyssey instrument.

2.9.8 Motility assays

Motility assays were set up as previously described (230). Briefly, using a toothpick a colony of *C. difficile* was pressed halfway down into *C. difficile* minimal media 0.3% agar plates which were incubated anaerobically for 24 hours. Where appropriate plates were supplemented with 15 µg/ml thiamphenicol for plasmid retention and 2 ng/ml anhydrotetracycline for expression of the complementation constructs. After 24 hours the diameter of the bacteria were measured in two directions, the average of which was used for analysis. Motility assays were performed across three biological replicates.

2.10 Statistical analysis

All graphs were produced using GraphPad Prism8. All linear regressions were performed in Stata16 software (StataCorp). All analysis by ANOVA was carried out in GraphPad Prism8 and Spearman's rank order correlation on www.vassarstats.net. For all statistical analyses $p < 0.05$ was considered significant.

Linear regression was used to determine if there were significant differences between:

- 1- Expression of reporter constructs (SNAP-tag, *gusA* and *phoZ*) in the various media conditions tested
- 2- Expression of reporter constructs following mutation of promoter sites or inverted repeat sites within the promoter sequences controlling expression of the reporters
- 3- Expression of reporter constructs in different strain backgrounds, e.g., wild-type and their respective mutants
- 4- Fold change, as determined by qRT-PCR, of *hpdC* expression between 630Δ*erm* and *sigL::erm* strains, and between the five clinical lineages in BHIS +/- 2 mg/ml *p*-HPA

- 5- Tyrosine, *p*-HPA and *p*-cresol production or conversion of tyrosine or *p*-HPA in either different media conditions or between different strains in identical media conditions
- 6- Total cell counts following growth of 630 Δ *erm* and 630 Δ *erm hpdC::CT* in *p*-HPA for 24 hours
- 7- The final pH of the media following growth of 630 Δ *erm* and 630 Δ *erm hpdC::CT* in pHs equivalent to 0, 1, 2, 3 and 4 mg/ml *p*-HPA
- 8- Motility of strains: 630 Δ *erm*, 630 Δ *erm sinR-CD2215::CT*, 630 Δ *erm sinR-CD2215::CT P_{tet}-sinR*, 630 Δ *erm sinR-CD2215::CT P_{tet} sinR-CD2215*, 630 Δ *erm CD2215::CT*, 630 Δ *erm CD2215::CT P_{tet}-CD2215*, 630 Δ *erm (Paris)*, 630 Δ *erm Δ sinR* and 630 Δ *erm Δ sinR P_{tet}-sinR*

Spearman's rank order correlation was used to determine if there were significant correlations between:

- 1- Expression from the *hpdBCA* promoter and increasing *p*-HPA concentration
- 2- Sporulation rates of 630 Δ *erm* and 630 Δ *erm hpdC::CT* when grown in increasing *p*-HPA concentrations for 24 hours

ANOVA analysis was used to determine if there were significant differences between:

- 1- Growth of bacterial strains grown in different *p*-HPA concentrations
- 2- Growth of bacterial strains in media with the pH altered to that of the *p*-HPA concentrations tested above
- 3- Inorganic phosphate release of *C. difficile* and *E. coli* when exposed to increasing *p*-HPA concentrations over the course of 90 minutes
- 4- Inorganic phosphate release of *C. difficile* and *E. coli* when exposed to the pHs matched to the *p*-HPA concentrations tested previously over the course of 90 minutes

Chapter 3: *Clostridioides difficile* para-cresol production is induced by the precursor para-hydroxyphenylacetate

3.1. Research paper cover sheet

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – STUDENT DETAILS

Student ID Number	LSH366976	Title	Mr
First Name(s)	Mark		
Surname/Family Name	Harrison		
Thesis Title	Characterising regulation of virulence factors affecting the interaction between <i>Clostridioides difficile</i> and the intestinal microbiome		
Primary Supervisor	Dr Lisa Dawson		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Journal of Bacteriology		
When was the work published?	August 2020		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	Not applicable		
Have you retained the copyright for the work?*	No	Was the work subject to academic peer review?	Yes

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

This article was published under Creative Commons Attribution 4.0 International license as stated on the first page of the article.


SECTION C – Prepared for publication, but not yet published


Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work



For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	The project was conceived by Dr Dawson who had the initial idea of using a reporter to identify triggers of <i>p</i> -cresol production, together we built on this and expanded the scope of the project to include a comparison of different reporters as well as a characterisation of the promoter of the <i>hpdBCA</i> operon. I carried out all of the lab work in this project with two exceptions: 1) The RNA extractions for the qRT-PCR as well as the qRT-PCR for <i>hpdC</i> expression itself (Fig 7G) was completed by Dr Dawson, 2) I prepared the samples for HPLC however the samples were analysed by Dr Harparkash Kaur. I initially wrote the first draft of the manuscript which was revised and edited by Dr Dawson into its finalised first draft form, this draft was sent to the other authors and subsequently updated prior to submission.
--	--

SECTION E

Student Signature	
Date	15/09/2021

Supervisor Signature	
Date	20/9/2021

Clostridioides difficile *para*-Cresol Production Is Induced by the Precursor *para*-Hydroxyphenylacetate

Mark A. Harrison,^a Alexandra Faulds-Pain,^a Harparkash Kaur,^a Bruno Dupuy,^b Adriano O. Henriques,^c  Isabelle Martin-Verstraete,^{b,d} Brendan W. Wren,^a  Lisa F. Dawson^a

^aDepartment of Infection Biology, London School of Hygiene and Tropical Medicine, London, United Kingdom ^bLaboratoire Pathogénèses des Bactéries Anaérobies, Institut Pasteur, Université de Paris, Paris, France ^cInstituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal ^dUniversité Paris Diderot, Sorbonne Paris Cité, Paris, France

ABSTRACT *Clostridioides difficile* is an etiological agent for antibiotic-associated diarrheal disease. *C. difficile* produces a phenolic compound, *para*-cresol, which selectively targets gammaproteobacteria in the gut, facilitating dysbiosis. *C. difficile* decarboxylates *para*-hydroxyphenylacetate (*p*-HPA) to produce *p*-cresol by the action of the HpdBCA decarboxylase encoded by the *hpdBCA* operon. Here, we investigate regulation of the *hpdBCA* operon and directly compare three independent reporter systems; SNAP-tag, glucuronidase *gusA*, and alkaline phosphatase *phoZ* reporters to detect basal and inducible expression. We show that expression of *hpdBCA* is upregulated in response to elevated *p*-HPA. *In silico* analysis identified three putative promoters upstream of *hpdBCA* operon—P₁, P₂, and P_{σ54}; only the P₁ promoter was responsible for both basal and *p*-HPA inducible expression of *hpdBCA*. We demonstrated that turnover of tyrosine, a precursor for *p*-HPA, is insufficient to induce expression of the *hpdBCA* operon above basal levels because it is inefficiently converted to *p*-HPA in minimal media. We show that induction of the *hpdBCA* operon in response to *p*-HPA occurs in a dose-dependent manner. We also identified an inverted palindromic repeat (AAAAAG-N₁₃-CTTTT) upstream of the *hpdBCA* start codon (ATG) that is essential for inducing transcription of the *hpdBCA* operon in response to *p*-HPA, which drives the production of *p*-cresol. This provides insights into the regulatory control of *p*-cresol production, which affords a competitive advantage for *C. difficile* over other intestinal bacteria, promoting dysbiosis.

IMPORTANCE *Clostridioides difficile* infection results from antibiotic-associated dysbiosis. *para*-Cresol, a phenolic compound produced by *C. difficile*, selectively targets gammaproteobacteria in the gut, facilitating dysbiosis. Here, we demonstrate that expression of the *hpdBCA* operon, encoding the HpdBCA decarboxylase which converts *p*-HPA to *p*-cresol, is upregulated in response to elevated exogenous *p*-HPA, with induction occurring between > 0.1 and ≤ 0.25 mg/ml. We determined a single promoter and an inverted palindromic repeat responsible for basal and *p*-HPA inducible *hpdBCA* expression. We identified turnover of tyrosine, a *p*-HPA precursor, does not induce *hpdBCA* expression above basal level, indicating that exogenous *p*-HPA was required for *p*-cresol production. Identifying regulatory controls of *p*-cresol production will provide novel therapeutic targets to prevent *p*-cresol production, reducing *C. difficile*'s competitive advantage.

KEYWORDS *Clostridium difficile*, *para*-cresol, tyrosine, *p*-HPA, reporter, SNAP-tag, PhoZ, GusA, transcription, σ₅₄, *Clostridioides difficile*, transcriptional regulation, transcriptional reporter

Citation Harrison MA, Faulds-Pain A, Kaur H, Dupuy B, Henriques AO, Martin-Verstraete I, Wren BW, Dawson LF. 2020. *Clostridioides difficile* *para*-cresol production is induced by the precursor *para*-hydroxyphenylacetate. J Bacteriol 202:e00282-20. <https://doi.org/10.1128/JB.00282-20>.

Editor Michael J. Federle, University of Illinois at Chicago

© Crown copyright 2020. This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Lisa F. Dawson, lisa.dawson@lshtm.ac.uk.

Received 12 May 2020

Accepted 24 June 2020

Accepted manuscript posted online 6 July 2020

Published 25 August 2020

Clostridioides difficile, previously classified as *Clostridium difficile*, is a major problem in health care systems and in the United States alone costs the economy up to \$5.4 billion per year (1). *C. difficile* is a spore-forming bacillus, whose spores resist disinfectants, heat, and desiccation (2, 3), as well as surviving in hospitals and the environment providing a reservoir for infection. *C. difficile* spores are transmitted via the oral-fecal route (2); they germinate in the gut in response to bile acid germinants (4), when the natural protective gut microbiota has been disrupted with antibiotics (5). Infected patients are treated with either metronidazole, vancomycin, or fidaxomicin. However, relapse of *C. difficile* infection is a serious concern, with up to 20 to 30% of patients requiring multiple rounds of antibiotics and in extreme cases fecal microbial transplants to treat disease (6).

C. difficile produces two toxins, TcdA and TcdB, which damage the intestinal barrier via glycosyltransferase activities targeting Rho, Rac, and Cdc42 in the cytoskeleton (7, 8). However, other putative virulence factors have been identified (9). *C. difficile* is among four intestinal bacteria that produce *para*-cresol above 100 μ M (10); *p*-cresol is an antimicrobial compound, which selectively targets Gram-negative bacteria in the host, reducing microbial diversity and increasing relapse in a murine model of infection (11). This selectively provides a competitive advantage for *C. difficile* over Gram-negative intestinal bacteria isolated from healthy human volunteers (11). Importantly, while 55 strains of intestinal bacteria from 152 species were shown to produce *p*-cresol, 51 of these strains produced between 10- and 1,000-fold lower levels than *C. difficile* (10, 12). The production levels of *p*-cresol unique to *C. difficile* combined with its microbiome altering effects make *p*-cresol production an important attribute for *C. difficile* during establishment of colonization and subsequent disease.

p-Cresol is produced from tyrosine fermentation with the intermediary formation of *p*-HPA, which is then decarboxylated by the HpdBCA decarboxylase to produce *p*-cresol (13). Disruption of any of the three genes of the *hpdBCA* operon encoding the decarboxylase renders *C. difficile* unable to produce *p*-cresol (14). Currently, very little is known regarding the genetic regulatory control of *p*-cresol production. Previously, we have shown that *p*-cresol production is inhibited in rich medium (brain heart infusion supplemented with yeast extract) and produced in less-rich medium (yeast peptone) (14). We can overcome the inhibition of *p*-cresol production in BHIS with the addition of the intermediate *p*-HPA (14), suggesting a novel regulatory mechanism of *p*-cresol formation.

There are a number of colorimetric and chemiluminescent reporters available for use in gene regulation studies (15). However, most transcriptional and translational reporters require the presence of oxygen to fully function and are therefore difficult to use in anaerobic bacteria (15, 16). Another complication is that *C. difficile* autofluoresces green, rendering the utilization of green fluorescent protein difficult (16). In recent years, a number of alternative reporters have been developed for use in *C. difficile* with various efficacies, including *phoZ* (15), SNAP and CLIP tags (17), and LOV domain (18), including iLOV and phiLOV (19), mCherryOpt (16), and FAST (20). SNAP-tag reporters, which are based on human *O*-6-methylguanine-DNA methyltransferase, have been successfully used in translational fusions to identify protein localization and for fluorescence microscopy (17, 21) but have received limited attention for quantification of promoter activity. Second, the glucuronidase reporter, *gusA*, from *Escherichia coli* has been used in a number of studies to assess expression of the *C. difficile* toxin genes; the promoter regions of *tcdA*, *tcdB*, and *tcdR* were fused to the *E. coli gusA* reporter gene and transferred into either *Clostridium perfringens* (22, 23) or *E. coli* (24) for analysis. More recently, the *E. coli gusA* reporter has been cloned with a constitutive (*cwp2*) or an inducible promoter (*ptet*) and assessed in *C. difficile* as a tool to compare the expression levels (25). Third, *phoZ*, an alkaline phosphatase from *Enterococcus faecalis* has been

shown to be a cost-effective and easily quantifiable reporter for transcriptional studies in *C. difficile* (15). These three reporters represent the most promising reporters in the study of *C. difficile* and therefore warranted further investigation as well as direct comparison of their sensitivity, cost, and ease of method.

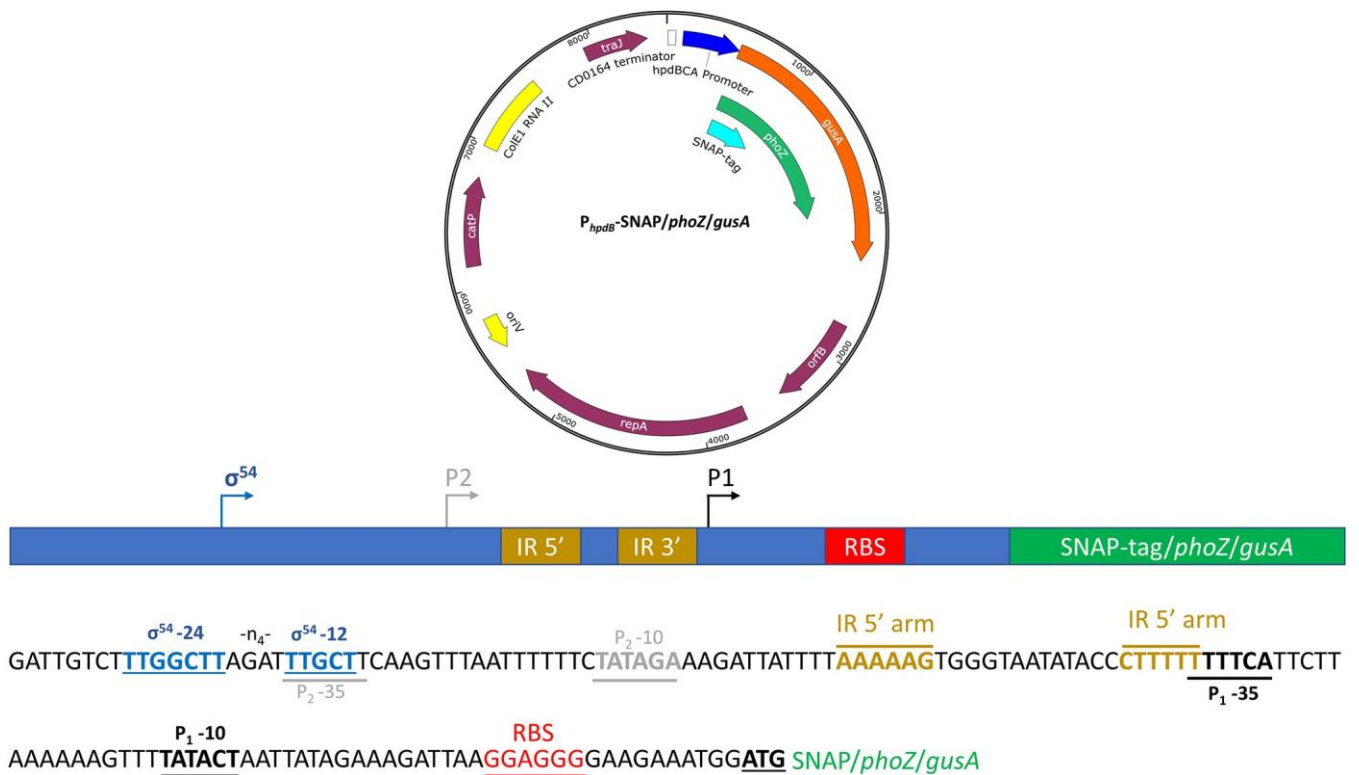


FIG 1 Overview of the SNAP, *phoZ*, and *gusA* reporter constructs. Each reporter was cloned into the *C. difficile* shuttle plasmid pMTL84151 and placed under transcriptional fusion with a 402-bp region (-399 to +3), including 399 bp directly upstream of *hpdB*CA, to the *hpdB* start codon (ATG).

In this study, we investigate the regulation of expression of the *hpdB*CA operon encoding the HpdBCA decarboxylase, which converts *p*-HPA to *p*-cresol. Therein, we compare three different reporter systems (SNAP-tag, *phoZ*, and *gusA*) to determine their efficacy in the detection of basal and inducible expression. We detected low-level basal expression from the 402-bp region upstream of *hpdB*CA using all three reporters. Importantly, the *gusA* and *phoZ* reporters produced similar quantitative trends in expression from the *hpdB*CA promoter region, with a significant decrease in the induction ratio of the SNAP-tag in the presence of *p*-HPA, which was partially overcome by increasing the culture volume sampled. Second, we identify three putative promoters, P₁, P₂, and P σ^{54} , and show by site-directed mutagenesis that the P₁ promoter alone is responsible for both basal and inducible expression of the reporter constructs. Furthermore, we identify that the turnover of tyrosine in minimal media to *p*-HPA is insufficient to significantly induce expression from the *hpdB*CA promoter region above basal level, suggesting that the conversion of tyrosine to *p*-HPA under these conditions may be inefficient. We found that operon expression was induced at a concentration of > 0.1 to \leq 0.25 mg/ml *p*-HPA. Finally, we have identified that *p*-HPA induced expression from the P₁ promoter requires the presence of an inverted palindromic repeat (AAAAAG-N₁₃-CTTTT) located upstream of the P₁ promoter.

RESULTS Detecting expression from the *hpdBCA* promoter region using different transcriptional fusions. Transcriptional control of the operon encoding the HpdBCA decarboxylase, which converts *p*-HPA to *p*-cresol, was assessed using three different reporter systems. A 402-bp DNA fragment corresponding to the promoter region located upstream of *hpdB* (-399 to +3) was fused, including the *hpdB* start codons (ATG) to the second codons of the SNAP-tag, *phoZ*, and *gusA* reporters, to produce in-frame reporter constructs (Fig. 1). Expression of the reporters was assessed in minimal medium (MM) in the presence or absence of *p*-HPA or tyrosine. We sought to detect and quantify basal and inducible expression from these reporters and to compare their efficiencies.

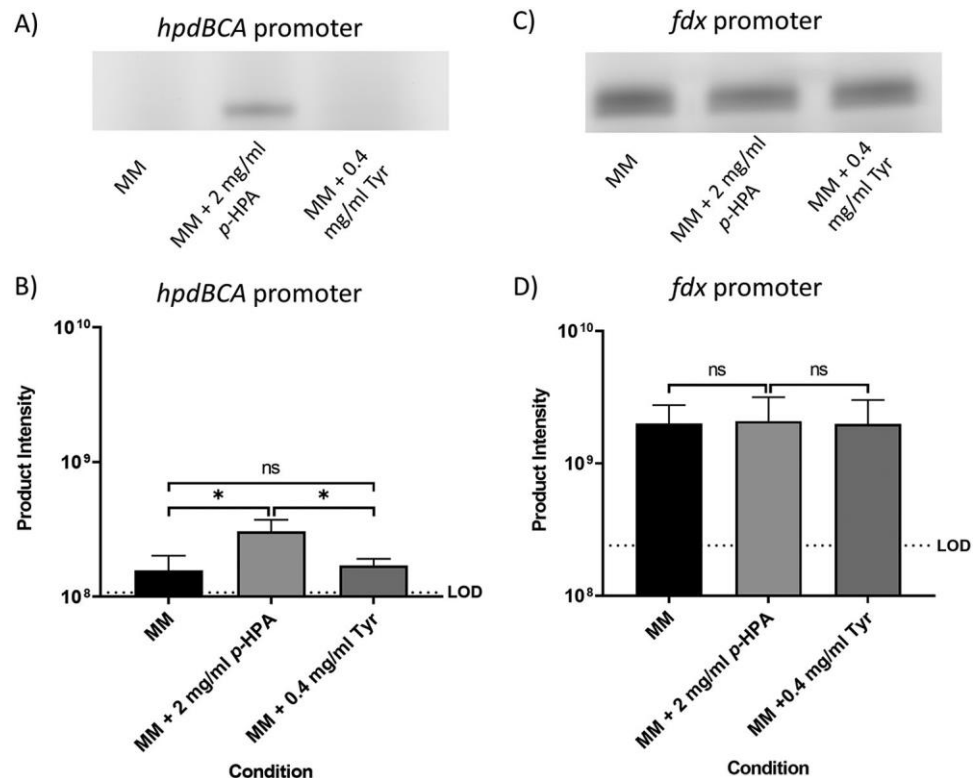


FIG 2 Visualization of the SNAP-tag reporter activity under the control of the *hpdBCA* promoter and a constitutive *fdx* promoter. Strains 630 Δ erm P_{hpdB} -SNAP (A and B) and 630 Δ erm P_{fdx} -SNAP (C and D) were grown for 4 h in MM, in MM plus 2 mg/ml *p*-HPA, or in MM plus 0.4 mg/ml tyrosine. Samples were run on an SDS-PAGE gel and processed with the fluorescent substrate TMR-Star prior to quantification of the SNAP-tag band using ImageQuant TL software. Three biological replicates were quantified using the following formula: product intensity = pixel volume/culture OD₅₉₀. Dotted lines represent the limit of detection (LOD) calculated as the average of the results of the following formula for all samples: background pixel volume/culture OD₅₉₀. Data represent means and standard errors. Statistical analysis was undertaken using linear regression to determine whether there is a significant effect of growth medium composition on the expression of the reporter construct in 630 Δ erm P_{hpdB} -SNAP (B) and 630 Δ erm P_{fdx} -SNAP (D). *, $P < 0.05$; ns, not significant. In panel D, no significant difference in expression was observed in MM alone compared to MM plus *p*-HPA and tyrosine.

SNAP-tag reporter fusion. A SNAP-tag reporter was fused to the *hpdBCA* promoter region (P_{hpdB} -SNAP), as well as to a constitutive promoter *fdx* (derived from the ferredoxin gene from *Clostridium sporogenes* NCIMB 10696) (P_{fdx} -SNAP) in *C. difficile* (Fig. 2). The expression from these constructs was analyzed using SDS-PAGE as recommended by Cassona *et al.* (26) with additional gating of the fluorescent bands imaged by a Typhoon fluorescence imager for quantification. We assessed transcription in MM compared to MM with the addition of either 2 mg/ml *p*-HPA or 0.4 mg/ml tyrosine (Fig. 2). In the

presence of *p*-HPA, transcription from the P_{hpdB} -SNAP fusion was found to be significantly higher than that in MM ($P = 0.036$, coefficient of variance [COV] = 0.297) or MM plus 0.4 mg/ml tyrosine ($P = 0.015$, COV = 0.248) (Fig. 2A and B). The P_{fdx} -SNAP control construct, with the SNAP-tag under the control of the constitutive *fdx* promoter, showed no significant difference in expression in minimal media compared to MM with the addition of *p*-HPA and tyrosine (Fig. 2C and D; see also File S1 in the supplemental material), therefore indicating that the effect of *p*-HPA on the induction of SNAP-tag production from the P_{hpdB} -SNAP fusion was specific (Fig. 2).

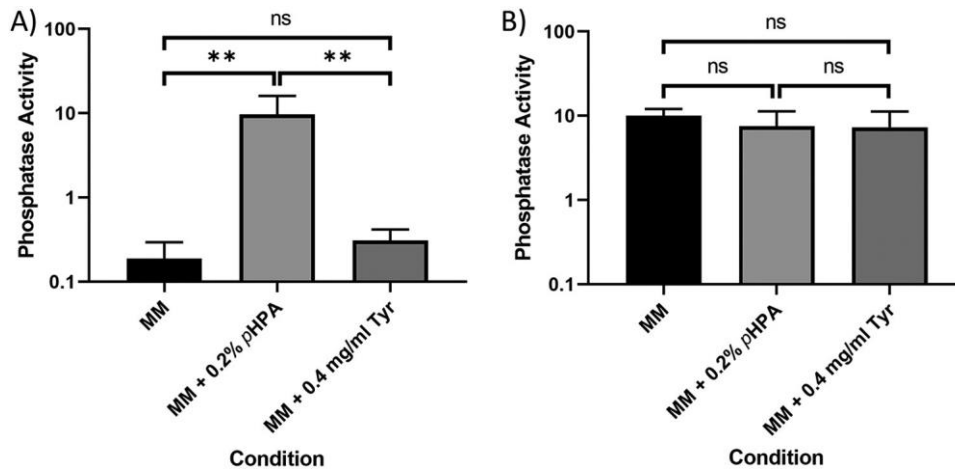


FIG 3 Visualization of the *phoZ* reporter activity under the control of the *hpdBCA* promoter and a constitutive *fdx* promoter. Strains $630\Delta erm$ P_{hpdBCA} -*phoZ* (A) and $630\Delta erm$ P_{fdx} -*phoZ* (B) were grown for 4 h in MM, in MM supplemented with 2 mg/ml *p*-HPA, or in MM supplemented with 0.4 mg/ml tyrosine. The expression of the *phoZ* reporters was quantified by using the following formula: $[OD_{420} - (OD_{550} \times 1.75)] \times 1,000/t$ (min) $\times OD_{590} \times$ volume of cells (ml). The data represent means and standard errors from three biological replicates. Statistical analyses were undertaken using linear regression to determine whether there is a significant effect on growth medium composition in $630\Delta erm$ P_{hpdBCA} -*phoZ* (A) and $630\Delta erm$ P_{fdx} -*phoZ* (B) strains (**, $P < 0.01$).

***phoZ* reporter fusion.** The promoter region of the *hpdBCA* operon was also fused to the chemiluminescent *phoZ* reporter gene (15). Using a P_{hpdBCA} -*phoZ* fusion and a control P_{fdx} -*phoZ* fusion (Fig. 3), we observed basal expression from the *hpdBCA* promoter when cells were grown in MM alone or in the presence of tyrosine, while the expression of the P_{hpdBCA} -*phoZ* construct is induced in MM in the presence of *p*-HPA via detection of phosphatase activity (Fig. 3A). We noted that to optimize detection of promoter activity, the cultures harvested at 4 h required an overnight incubation with the *phoZ* substrate *p*-NPP to optimize detect basal level expression in MM and MM plus tyrosine. The presence of tyrosine gave a slight, but not significant increase in the expression of *phoZ* from the *hpdBCA* promoter region (Fig. 3A) ($P = 0.207$, COV = 0.237), while the addition of *p*-HPA significantly increased expression from the *hpdBCA* promoter region (Fig. 3A) ($P < 0.01$, COV = 1.69). We found no significant differences in expression of *phoZ* from the constitutive promoter (P_{fdx}) under the different conditions tested (*p*-HPA comparison ($P < 0.290$, COV = -0.152) and tyrosine comparison ($P = 0.330$; COV = -0.187), suggesting that medium conditions had no effect on reporter detection and quantification using these methodologies (Fig. 3B) and that the effect of *p*-HPA was specific to the *hpdBCA* promoter.

***gusA* reporter fusions.** We also investigated whether *gusA*, an alternative chemiluminescent reporter to *phoZ*, could be used for such studies. Here, the same 402-bp upstream region of *hpdBCA* (-399 to +3) was cloned in a transcriptional fusion to *gusA*,

including the *hpdB* start codon, to the second codon of *gusA* ($P_{hpdB-gusA}$). We detected basal level expression from the $P_{hpdB-gusA}$ transcriptional fusion in MM, with a slight but not significant increase in the presence of tyrosine (Fig. 4) ($P = 0.271$, $COV = 0.033$). Once again, the addition of *p*-HPA resulted in a significant increase in expression from the *hpdBCA* promoter region (Fig. 4) ($P < 0.01$, $COV = 1.143$).

Comparative quantification of fold change induction of the *hpdBCA* operon using different techniques. When comparing reporter changes in the presence of *p*-HPA we see a (2.04 ± 0.61)-fold change when tested by the SNAP-tag (Fig. 2A), a (49.48 ± 8.71)-fold change by *phoZ* (Fig. 3A), and a (41.48 ± 23.35)-fold change by *gusA* (Fig. 4). Our analysis suggested that the SNAP-tag (Fig. 2) significantly underestimated the fold change compared to both the *phoZ* ($P < 0.001$, $COV = -4.650$) and *gusA* reporters ($P = 0.002$, $COV = -4.234$). Both *phoZ* and *gusA* reporters have the largest fold change of the three reporters in response to *p*-HPA, which may facilitate detection of small changes in expression.

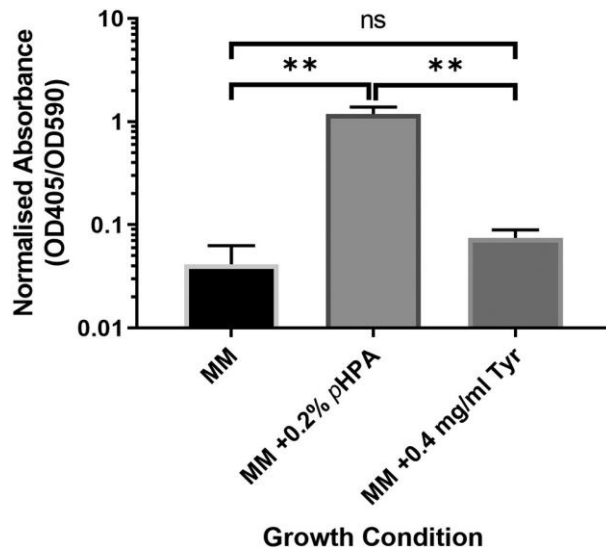


FIG 4 Visualization of the *gusA* reporter activity under the control of the *hpdBCA* promoter. $630\Delta erm$ *hpdBCA-gusA* was grown for 4 h in MM, in MM supplemented with 2 mg/ml *p*-HPA, or in MM supplemented with 0.4 mg/ml tyrosine. Expression was measured at OD_{405} and adjusted by growth using OD_{590} . Data represent the means and standard errors from three biological replicates (**, $P < 0.01$). Statistical analysis was undertaken using linear regression to determine whether there is a significant effect on growth medium composition in the $630\Delta erm$ $P_{hpdB-gusA}$.

Induction of *hpdBCA* expression correlates with exogenous *p*-HPA. Based on comparison of the fold change in response to *p*-HPA (Fig. 2A, 3A, and 4), we chose the $P_{hpdB-phoZ}$ fusion to examine the concentration-dependent induction of the *hpdBCA* promoter region in response to exogenous *p*-HPA. We showed a clear correlation between *p*-HPA concentration and *phoZ* expression ($R^2 = 0.8429$, $P < 0.001$) (Fig. 5A). A significant induction of the $P_{hpdB-phoZ}$ fusion was detected in the presence of ≥ 0.25 mg/ml *p*-HPA ($P = 0.025$, $COV = 0.406$) (Fig. 5B).

Inefficient turnover of tyrosine to *p*-HPA by *C. difficile*. In MM supplemented with tyrosine, we used high-pressure liquid chromatography (HPLC) to determine the levels of tyrosine, *p*-HPA, and *p*-cresol after 10 h of growth (late exponential phase), in which we observed maximum *C. difficile* growth (measured as the optical density at 595 nm [OD_{595}]), and 24 h of growth (stationary phase), at which we have previously observed maximum production of *p*-cresol (14). Here, we added the maximum soluble tyrosine

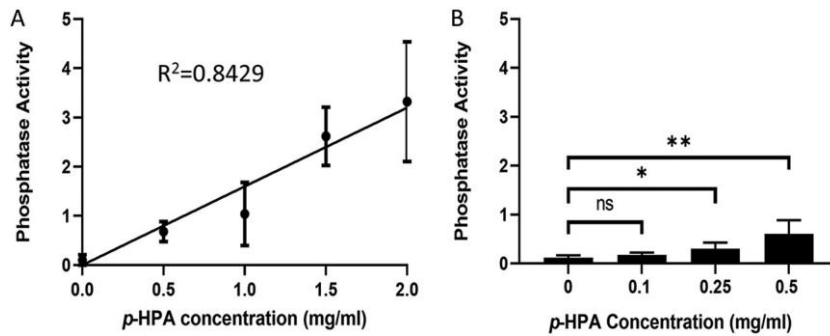


FIG 5 Use of the *phoZ* reporter to determine a concentration-dependent response to *p*-HPA for induction of the *hpdBCA* operon. (A) Expression from P_{hpdB} -*phoZ* was monitored in MM with increasing concentrations of *p*-HPA (0.5, 1, 1.5, and 2 mg/ml). (B) Expression from P_{hpdB} -*phoZ* was tested with *p*-HPA concentrations of 0.1, 0.25, and 0.5 mg/ml. The expression of the *phoZ* reporters was quantified by using the following formula: $[\text{OD}_{420} - (\text{OD}_{550} \times 1.75)] \times 1,000/t$ (min) \times $\text{OD}_{590} \times$ volume of cells (ml). Statistical analysis was performed. (A) Linear regression was used to determine any effect on phosphatase activity as *p*-HPA concentration increases ($R^2 = 0.8429$, $P < 0.001$); (B) linear regression was used to determine at what *p*-HPA concentration there was a significant effect on expression compared to the absence of *p*-HPA (*, $P < 0.05$; **, $P < 0.01$).

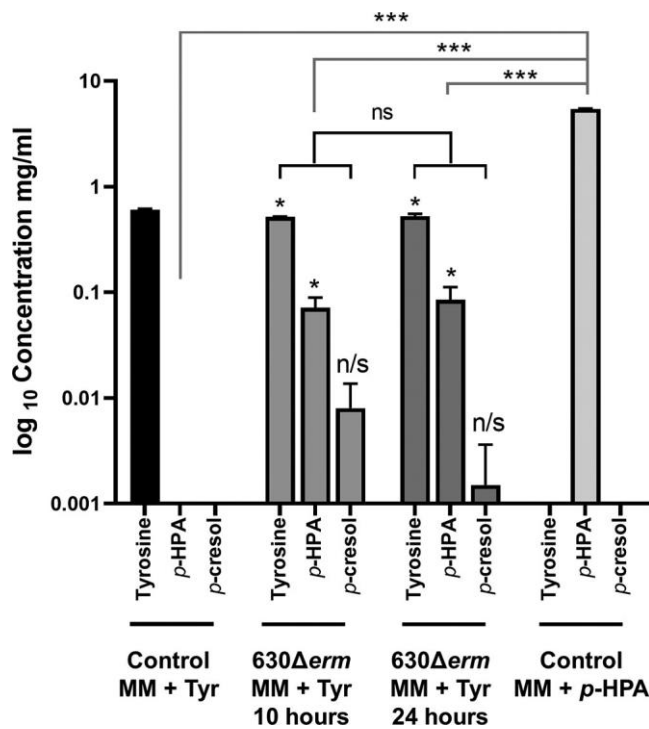


FIG 6 HPLC analysis to detect tyrosine, *p*-HPA, and *p*-cresol in spent culture media. The concentrations of tyrosine, *p*-HPA, and *p*-cresol in late-exponential-growth cultures of *C. difficile* strain 630 Δ erm were determined by HPLC. Minimal medium was supplemented with 0.4 mg/ml tyrosine or 2 mg/ml *p*-HPA. Statistical analysis using linear regression was undertaken (i) to determine whether there was a significant increase in *p*-HPA and *p*-cresol production in media containing tyrosine (in black) and (ii) to determine whether there were any significant differences in the amount of *p*-HPA in the 2 mg/ml supplemented media compared to the amount of *p*-HPA produced in growth media containing tyrosine (in gray) (*, $P < 0.05$; ***, $P < 0.001$).

concentration (0.4 mg/ml) to maximize the potential conversion of tyrosine to *p*-HPA. However, we saw no significant induction of expression of *hpdBCA* when tested by any of the reporters used in this study (Fig. 2 to 4). After 10 h, *C. difficile* utilized only 0.0895 mg/ml of the available tyrosine, leading to the production of 0.07 ± 0.02 mg/ml *p*-HPA (P

= 0.031, COV = 0.0715), representing a turnover of 11.8% from tyrosine to *p*-HPA. This, in turn, led to a slight but not significant detection of *p*-cresol 0.008 ± 0.006 mg/ml ($P = 0.106$, COV = 0.008) (Fig. 6). No significant differences were observed in the concentrations of tyrosine, *p*-HPA and *p*-cresol between the 10- and 24-h time points (see File S1 in the supplemental material). Upon testing by the *phoZ* assay, no significant difference was detected in the expression of the P_{hpdB} -*phoZ* fusion in the presence of 0.1 mg/ml *p*-HPA compared to no *p*-HPA; however, a significant increase in induction above baseline was detected at 0.25 mg/ml *p*-HPA ($P = 0.025$, COV = 0.406) (Fig. 5B). These results suggest that, under these conditions, tyrosine is very inefficiently used for the production of *p*-HPA and *p*-cresol. Theoretically, higher concentrations of tyrosine may induce more efficient turnover of tyrosine to *p*-HPA (by an as-yet-unidentified pathway) and therefore increase *p*-cresol production; however, tyrosine itself has no direct effect on the expression of the *hpdBCA* operon but only an indirect effect via the production of *p*-HPA.

Regulation of *hpdBCA* expression. We identified three putative promoters upstream of the *hpdBCA* operon in the promoter region used to obtain transcriptional fusions. Two of them (P_1 and P_2) showed similarity to the consensus of the housekeeping sigma factor, SigA (TTGACA-N₁₇-TATAAT) (27). The -10 box of the P_1 promoter corresponds to the sequence TATACT, while the -35 box TTTTCA is located 16 nucleotides upstream of the -10 box (Fig. 7A). The P_2 promoter contains -10 and -35 boxes TTGCTT-N₁₇-TATAGA similar to the consensus of SigA-dependent promoters (Fig. 7A). A third putative promoter is present in the promoter region of the *hpdBCA* operon with a consensus with similarities to a σ^{54} (SigL) promoter site (TTGGCAT-N₅-TTGCT) (28, 29). However, a reduced spacer between the -12 and -24 boxes (TTGG CTT-N₄-TTGCT) (Fig. 7A) is present.

We first carried out site-directed mutagenesis to modify key bases TAT>TGC of the P_1 promoter (Fig. 7A). Using a scaled-up SNAP-tag assay (see “Scaled-up SNAP-tag” below), we found improved performance since in the larger culture volume the fold change was found to be 18.48 ± 2.64 compared to the smaller culture’s fold change of 2.04 ± 0.61 . We analyzed the effects of these mutations on basal and inducible expression of *hpdBCA* operon, in MM, MM with *p*-HPA, and MM with tyrosine (Fig. 7B and C). Mutation of the -10 box in the P_1 promoter (TAT>TGC) abolished expression in MM ($P < 0.001$, COV = -0.357), MM supplemented with tyrosine ($P < 0.001$, COV = -0.440), and both basal and induced expression in the presence 2 mg/ml *p*-HPA ($P < 0.001$, COV = -1.574) (Fig. 7). In the presence of *p*-HPA, mutation of the P_2 promoter (TAT>TGC) had no effect on expression from the *hpdBCA* promoter region (Fig. 7D and E) ($P = 0.924$, COV = -0.011), whereas the double mutation of both P_1 and P_2 led to no detectable expression (Fig. 7D). These results strongly suggest that P_1 rather than P_2 is functional.

To assess whether the putative SigL promoter had an effect on expression, we transferred the SNAP-tag reporter in a $630\Delta erm$ *sigL::erm* background. After growth in BHIS (see Materials and Methods) supplemented with 100 mM glucose (BHISG), required by $630\Delta erm$ *sigL::erm* for successful culture, in the presence or absence of *p*-HPA, we observed that *sigL* inactivation did not decrease expression of the SNAP reporter in the presence ($P = 0.498$, COV = -0.265) or absence ($P = 0.071$, COV = -0.107) of *p*-HPA (Fig. 7F). We confirmed by quantitative reverse transcription-PCR (qRT-PCR) that the expression of *hpdC* was unchanged in the *sigL::erm* mutant compared to $630\Delta erm$ in BHISG broth ($P = 0.943$, COV = -0.121) and BHISG supplemented with *p*-HPA ($P = 0.641$, COV = 0.101) (Fig. 7G). These results confirm that the P_1 promoter is solely responsible for expression and induction of *hpdBCA* operon under the conditions tested (Fig. 7). In addition, we found that by qRT-PCR fold change for $630\Delta erm$ was 333.29 ± 169.26 . In comparison to the qRT-PCR, all three of the reporters tested significantly underestimated fold change (see File S1 in the supplemental material). The relative fold change for the scaled-up SNAP-tag assay

was 18-fold lower than the qRT-PCR, whereas the *phoZ* and *gusA* reporters were 6.8- and 8.2-fold lower, respectively, than the qRT-PCR.

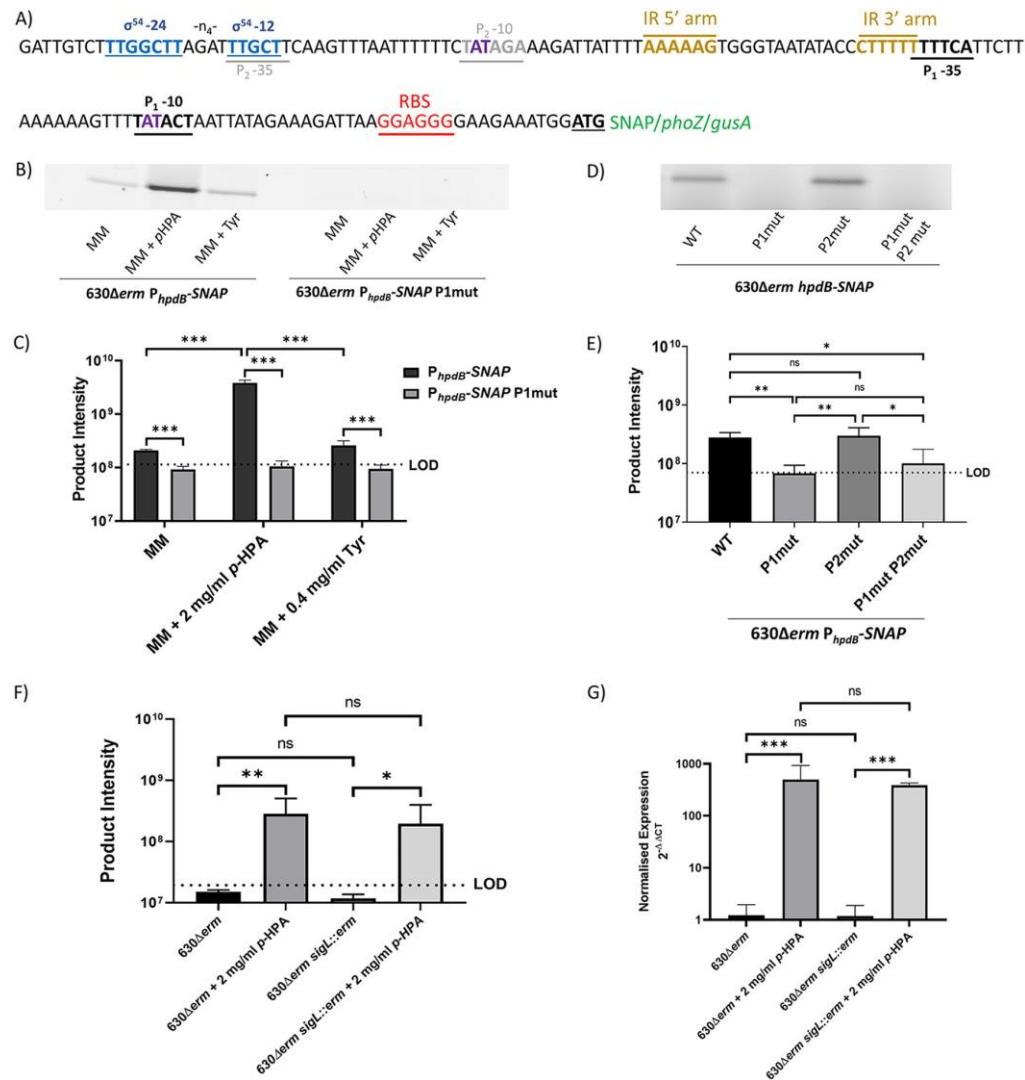


FIG 7 Mutation of the potential promoter regions in the *hpdBCA-SNAP*-tag reporter constructs. (A) Modifications by site-directed mutagenesis of the P₁ and P₂ promoter regions within the *hpdBCA-SNAP* reporter construct. The following putative promoter elements are marked on the diagram: putative ribosome binding site (red), putative P1 promoter (black), putative P2 promoter -10 region site (gray), putative SigL promoter (blue), putative inverted repeats (gold), and bases mutated by site-directed mutagenesis for functional analysis are indicated in purple. (B) SDS-PAGE gel image of the 630Δerm *P_{hpdB}-SNAP* alongside the mutated P₁ promoter (630Δerm *P_{hpdB}-SNAP P1mut*). (C) Quantification of expression from 630Δerm *P_{hpdB}-SNAP* and 630Δerm *P_{hpdB}-SNAP P1mut*. (D) SDS-PAGE expression from the 630Δerm *hpdBCA-SNAP* compared to the mutated P₁ promoter 630Δerm *P_{hpdB}-SNAP P1mut*, the mutated P₂ promoter the 630Δerm *P_{hpdB}-SNAP P2mut*, and the mutation of both P₁ and P₂ promoters in MM plus 2 mg/ml p-HPA. (E) Quantification of expression from 630Δerm *hpdBCA-SNAP* compared to the mutated P₁ promoter 630Δerm *P_{hpdB}-SNAP P1mut*, the mutated P₂ promoter the 630Δerm *P_{hpdB}-SNAP P2mut*, and the mutation of both P₁ and P₂ promoters in MM plus 2 mg/ml p-HPA. (F) Strains 630Δerm and 630Δerm *sigL::erm* were transformed with *P_{hpdB}-SNAP* construct. Expression of the SNAP-tag construct was assessed in BHISG medium compared to BHISG supplemented with 2 mg/ml p-HPA. Samples were run on an SDS-PAGE gel and processed with the fluorescent substrate TMR-Star prior to quantification of the SNAP-tag band using ImageQuant TL software. Three biological replicates were quantified using the following formula: product intensity = pixel volume/OD₅₉₀. The dotted lines in panels C, E, and F represent the limit of detection (LOD) calculated as the average of the following formula for all samples: background pixel volume/culture OD_{590nm}. (G) qRT-PCR was used to assess the expression of *hpdC* in late exponential phase normalized to the

16S rRNA internal control ($2^{-\Delta\Delta CT}$). All data represent means and standard errors. Statistical analysis was undertaken using linear regression to determine (i) whether there is a significant effect of growth medium composition on expression from the P_{hpdB} -SNAP in the 630 Δerm strain compared to the mutated P_1 promoter P_{hpdB} -SNAP P1mut, (ii) whether there is an effect on expression of P_{hpdB} -SNAP in which the P_1 and P_2 -10 boxes were mutated (TAT>TGC bases) in the P_1 and P_2 -10 sites, (iii) whether there is a significant increase in expression of the SNAP construct in media supplemented with *p*-HPA in both 630 Δerm and 630 Δerm *sigL::erm* strains, and (iv) whether there are any differences in the expression of *hpdC* in the 630 Δerm strain compared to the 630 Δerm *sigL::erm* strain (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

Response to *p*-HPA is dependent on the presence of an inverted repeat upstream of P_1 . We sought to determine whether the presence of an inverted palindromic repeat (AAAAAG-N₁₃-CTTTT) overlapping the P_1 -35 box was involved in induction of the P_{hpdB} -*phoZ* fusion in response to *p*-HPA. Site-directed mutagenesis was used to remove either the 5' arm of the inverted repeat or the entire inverted repeat from the P_{hpdB} -*phoZ* fusion reporter. Removal of the 5' arm (AAAAAG) or removal of the entire inverted repeat (AAAAAG-N₁₃-CTTTT) resulted in the abolition of inducible expression from the P_{hpdB} -*phoZ* fusion in the presence of *p*-HPA ($P < 0.001$, COV = -1.015; and $P = 0.003$, COV = -1.303, respectively) (Fig. 8), without interfering with basal transcription, suggesting that this palindromic repeat is essential for responding to elevations in intrinsic or extrinsic *p*-HPA.

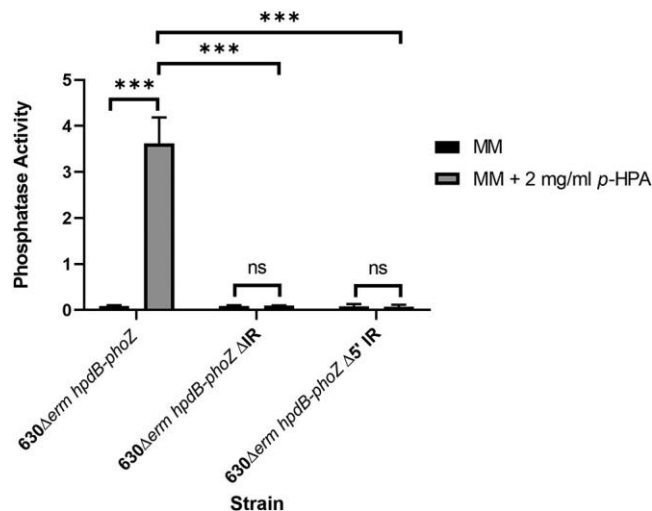


FIG 8 Mutation of the inverted repeat located within the *hpdBCA* upstream region leads to an inability to respond to *p*-HPA. Expression from the P_{hpdB} -*phoZ* reporter was assessed in the presence or absence of an inverted repeat (AAAAAG-N₁₃-CTTTT) located 82 bp upstream of the *hpdB* start codon. Inverse PCR was used to delete the entire inverted repeat, including the spacer or the 5' arm of the inverted repeat. Minimal medium was supplemented with 2 mg/ml *p*-HPA to determine whether any changes in expression were seen as a result of the mutation of the inverted repeat. ***, $P < 0.001$.

DISCUSSION Conditions that trigger *p*-cresol production by *C. difficile* have not been described to date. Here, we carried out a comparison of three different transcriptional reporters, as well as using them to ascertain the regulation of expression from the promoter region of the *hpdBCA* operon, which encodes the HpdBCA decarboxylase, responsible for the conversion of *p*-HPA to *p*-cresol. We demonstrate here that the *hpdBCA* operon has a single functional promoter (P_1) that controls both basal expression and *p*-HPA induced expression. Moreover, we showed that an inverted repeat (AAAAAG-N₁₃-CTTTT), located 45 bp upstream of the P_1 -10 site, seems to be essential for responding to the presence of *p*-HPA and inducing expression of the *hpdBCA* operon. Significant expression of *hpdBCA* above baseline is induced at a *p*-HPA concentration threshold in the range of > 0.1 to ≤ 0.25 mg/ml.

Transcriptional regulation in bacteria is tightly regulated to facilitate temporal and spatial gene expression (30). There are limited direct comparisons between transcriptional reporters in anaerobic bacteria (31), particularly in *C. difficile*. We showed transcriptional activity and induction of the promoter region of *hpdBCA* using all three reporters (SNAP-tag, *phoZ*, and *gusA*). However, the methods of detection of the reporters vary, as do their sensitivities when using relatively small culture volumes. The SNAP-tag can be measured by spectrophotometry; however, to increase accuracy, we used SDS-PAGE gel separation and subsequent quantification. This mitigates the high-level background, visualized as a low-molecular-weight nonspecific fluorescent species (see Fig. S1 in the supplemental material). This method of quantification is more labor-intensive in comparison to the other two reporters. SNAP-tag assays are also amenable to being scaled up to facilitate detection of low-level transcripts. An increase in culture volume facilitated the detection of basal transcription from the *hpdBCA*-SNAP fusion and resulted in detection of induction ratio in comparable to the other reporters and qRT-PCR. One advantage of the SNAP-tag reporter is that multiple substrates can be used with a wide range of wavelengths and has been widely used in translational fusions to investigate protein localization via microscopy, including down to single cell detection (17), a feature that neither *gusA* nor *phoZ* reporters offer. The *gusA* and *phoZ* reporter assays are relatively simple, inexpensive, and easy to perform. In the present study, cost differences occur in processing of the harvested cells carrying the different reporters. This is largely due to the expense of the reporter substrates; per single assay, the cost was found to be approximately £10 for the SNAP-tag compared to under £0.20 for the *gusA* and *phoZ* reporters, respectively. Both the *phoZ* and the *gusA* reporters can be used in smaller culture volumes than the SNAP-tag and still detect basal expression. Each of the three reporters underestimated the fold change of induced expression compared to transcription detected by qRT-PCR, although both *phoZ* and *gusA* reporters outperformed the SNAP-tag. SNAP-tag performance was improved using a scaled-up method, which improved the sensitivity from 166-fold lower than the qRT-PCR to only 18-fold lower. The scaled-up SNAP-tag worked well for differentiating promoter deletions; however, this came at a cost of additional complexity and expense.

A disadvantage of the *phoZ* reporter was that detection of basal expression required an overnight incubation, whereas both *gusA* and SNAP-tag reporters require a 30-min incubation in the presence of their substrates. Compared to the gold standard, qRT-PCR, we found that the *phoZ* reporter was the most efficient at detecting inducible expression while also being the most cost-effective system. It is possible in the future that new reporters will be more comparable to qRT-PCR in terms of detecting inducible expression.

Promoter consensus sequences have been identified for 9 of the 22 putative sigma factors in *C. difficile*, including SigA (27), SigH (27), SigL (σ^{54}) (28), SigB (32), SigD (33), and SigK, SigE, SigG, and SigF (34). In the 399-bp region upstream of the *hpdB* start codon (ATG), we show that among the three potential promoters identified *in silico* using consensus previously identified, only P_1 is functional and allows expression of the *hpdB* gene. The -10 box of P_1 (TATACT) and -35 box of P_1 (TTTTCA) have 83.3 and 66.7% homologies, respectively, to the consensus of σ^A -dependent promoters. Mutation of the putative P_1 -10 region (TAT>TCG) by site-directed mutagenesis abolishes expression, indicating that this promoter is crucial for *hpdBCA* expression, which is induced in the presence of *p*-HPA. Synthetic promoter biology has shown that the promoter (-10 and -35 sites), along with the spacer regions are essential for expression in clostridia (31). Expression of *hpdBCA* is regulated by an as-yet-unknown transcriptional regulator sensing *p*-HPA. In addition to identification of a single active promoter, we have shown that induction of this promoter by *p*-HPA is dependent on the presence of an inverted repeat located 82 bp upstream of the *hpdBCA* ATG directly upstream (AAAAAG-N₁₃-CTTTT).

Indeed, removal of this inverted repeat abolished inducible expression from the promoter while retaining basal expression levels.

In agreement with previous findings by Selmer and Andrei (13), we show that tyrosine does not directly affect expression of the *hpdBCA* operon in our conditions. A limiting factor in this study is the solubility of tyrosine (0.4 mg/ml), which may be below a threshold needed to induce conversion of tyrosine to *p*-HPA, by an as-yet-unidentified pathway. However, under the conditions tested in this study, with the maximum soluble tyrosine concentration, the conversion of tyrosine to *p*-HPA is extremely low (0.07 mg/ml). Even if basal level expression from the *hpdBCA* operon was sufficient to detect low level turnover of tyrosine to *p*-HPA by HPLC, the concentration of *p*-HPA within the cell is likely insufficient to observe an induction. In contrast, expression of the *hpdBCA* operon can be induced as a direct response to exogenous *p*-HPA. In the presence of exogenous *p*-HPA, an increased turnover of *p*-HPA to *p*-cresol is observed. Interestingly, a recent study has identified a novel tyrosine transporter (35), which facilitates the uptake of tyrosine for utilization in the cell. However, although the tyrosine transporter mutants showed reduced *p*-cresol production, this transporter is not involved in the transport of *p*-HPA (35). This suggests the presence of a novel and currently unidentified *p*-HPA transporter.

We have shown previously that *p*-cresol inhibited growth of Gram-negative gut commensal bacteria at a concentration of 1 mg/ml (11). Here, we demonstrate turnover from tyrosine to *p*-HPA and *p*-cresol; however, in this system the turnover is inefficient and results in the production of 0.01 mg/ml *p*-cresol, which is unlikely to be physiologically relevant in terms of having a modulatory effect on the intestinal microbiome. However, we show that exogenous *p*-HPA can induce the *hpdBCA* operon promoting *p*-cresol production. Exogenous *p*-HPA is detected in fecal samples, which is suggestive of its presence within the human gut (36), albeit at unknown concentrations. A range of intestinal bacteria (11) and human cells (36) produce *p*-HPA, potentially providing *C. difficile* with exogenous *p*-HPA, which then induces *p*-cresol production at concentrations necessary to adversely affect the diversity of the microbiome. The gene expression profile of *C. difficile* greatly varies depending on the environment, with genes involved in metabolism particularly affected by environmental changes in the gut (reviewed by Theriot and Young [37]); therefore, it is reasonable to hypothesize that *C. difficile* is able to respond efficiently to the availability of exogenous *p*-HPA.

Our results demonstrate new insights into the regulation of *p*-cresol production in *C. difficile*, highlighting that the HpdBCA decarboxylase is produced from a single promoter P₁, driving both basal and inducible expression of the *hpdBCA* operon. We have shown that *p*-HPA triggers expression of the *hpdBCA* operon in a concentration dependent manner via an inverted repeat sequence directly upstream of the P₁ promoter to enhance expression of the *hpdBCA* operon and upregulate *p*-cresol production. This strongly suggests that the *hpdBCA* operon is controlled by a positive uncharacterized regulator likely sensing *p*-HPA. A better understanding of the pathways leading to *p*-cresol production may help us develop strategies to inhibit *p*-cresol production and therefore reduce its deleterious effects on the microbiome.

TABLE 1 Strains and plasmids used in this study

Strain or plasmid	Relevant feature(s)	Source or reference
<i>C. difficile</i> strains		
630 Δ <i>erm</i>	Erythromycin sensitive strain of 630	42
630 Δ <i>erm</i> P _{hpdB} -SNAP	630 Δ <i>erm</i> with plasmid P _{hpdB} -SNAP	This study
630 Δ <i>erm</i> P _{fdx} -SNAP	630 Δ <i>erm</i> with plasmid P _{fdx} -SNAP	This study
630 Δ <i>erm</i> P _{hpdB} - <i>gusA</i>	630 Δ <i>erm</i> with plasmid P _{hpdB} - <i>gusA</i>	This study
630 Δ <i>erm</i> P _{hpdB} - <i>phoZ</i>	630 Δ <i>erm</i> with plasmid P _{hpdB} - <i>phoZ</i>	This study
630 Δ <i>erm</i> P _{fdx} - <i>phoZ</i>	630 Δ <i>erm</i> with plasmid P _{fdx} - <i>phoZ</i>	This study
CDIP217	630 Δ <i>erm</i> sigL:: <i>erm</i>	28
630 Δ <i>erm</i> sigL:: <i>erm</i> P _{hpdB} -SNAP	630 Δ <i>erm</i> sigL:: <i>erm</i> with plasmid P _{hpdB} -SNAP	This study
Plasmids		
P _{hpdB} -SNAP	pMTL84151 plasmid carrying a SNAP-tag under the control of the <i>hpdBCA</i> promoter region	26; this study
P _{hpdB} - <i>gusA</i>	pMTL84151 plasmid carrying <i>gusA</i> under the control of the <i>hpdBCA</i> promoter	This study
P _{hpdB} - <i>phoZ</i>	pMTL84151 plasmid carrying <i>phoZ</i> under the control of the <i>hpdBCA</i> promoter	This study
P _{fdx} -SNAP	pMTL84153 carrying SNAP-tag under control of the <i>fdx</i> promoter	This study
P _{fdx} - <i>phoZ</i>	pMTL84153 carrying <i>phoZ</i> under control of the <i>fdx</i> promoter	This study
P _{hpdB} -SNAP P1mut	pMTL84151 plasmid carrying a SNAP-tag under the control of the <i>hpdBCA</i> promoter region with a mutation of the P ₁ site from TAT to TGC	This study
P _{hpdB} -SNAP P2mut	pMTL84151 plasmid carrying a SNAP-tag under the control of the <i>hpdBCA</i> promoter region with a mutation of the P ₂ site from TAT to TGC	This study
P _{hpdB} -SNAP P1mut P2mut	pMTL84151 plasmid carrying a SNAP-tag under the control of the <i>hpdBCA</i> promoter region with mutation in both the P ₁ and P ₂ sites from TAT to TGC	This study
P _{hpdB} - <i>phoZ</i> Δ IR	630 Δ <i>erm</i> with plasmid P _{hpdB} - <i>phoZ</i> with the inverted repeat upstream of <i>hpdB</i> removed	This study
P _{hpdB} - <i>phoZ</i> Δ 5'IR	630 Δ <i>erm</i> with plasmid P _{hpdB} - <i>phoZ</i> with the 5' arm of the inverted repeat upstream of <i>hpdB</i> removed	This study
pMC358	Plasmid carrying <i>phoZ</i>	15
pRPF185	Plasmid carrying <i>gusA</i>	25

MATERIALS AND METHODS Bacterial strains and growth conditions. *C. difficile* strains used in this study are listed in Table 1. *C. difficile* strains were routinely grown on brain heart infusion (Oxoid) supplemented with 5 g liter⁻¹ yeast extract (Sigma) and 0.05% L-cysteine (Sigma) (BHIS) or in MM (38). Strains were grown under anaerobic conditions at 37°C in a Modular Atmosphere Control System 500 (Don Whitney Scientific). All media underwent at least 4 h pre-equilibration prior to inoculation. Thiamphenicol (Tm) was used at 15 μ g/ml to ensure retention of the reporter plasmids. Transformant *E. coli* strains were routinely grown at 37°C in Luria-Bertani (LB) agar or broth with 25 or 12.5 μ g/ml chloramphenicol, respectively.

Strain and plasmid construction. All oligonucleotides used in this study are listed in Table 2. The P_{fdx}-SNAP plasmid was constructed using a G-block of the SNAP coding sequence (Integrated DNA Technology), amplified by PCR (for the oligonucleotides, see Table 2), to add restriction sites NdeI and SacI and was ligated into the multiple cloning site (MCS) of pMTL84153. To generate the P_{hpdBCA}-SNAP plasmid, a 402-bp fragment from position -399 to +3 (ATG) of the *hpdB* start codon corresponding to the promoter region of *hpdBCA* (Fig. 1) was amplified from 630 Δ *erm* genomic DNA, while the SNAP-tag was amplified from P_{fdx}-SNAP. These two fragments were joined by SOE-PCR to create an in-frame construct, before cloning into the MCS of pMTL84151 using NdeI and SacI restriction enzymes and T4 ligase. To construct plasmids carrying P_{hpdB}-*phoZ* or P_{hpdB}-*gusA*, the SNAP-tag was replaced either *gusA* or *phoZ* amplified from pRPF185 (25) and pMC358 (15), with an inverse PCR to remove the SNAP-tag before the addition of PCR amplified *gusA* or *phoZ* using the NEbuilder HiFi assembly system (NEB). To generate the P_{fdx}-*phoZ* construct, the backbone of pMTL84153 was amplified by inverse PCR and the

insert, *phoZ*, was amplified from pMC358 and then assembled using an NEbuilder HiFi assembly system(NEB).

Transformation of chemically competent (NEB5 α ; NEB) or electrocompetent (Top10; Thermo Fisher) cells was performed according to the manufacturer's instructions. Reporter constructs were checked by sequencing, and the corresponding plasmids were transferred by electroporation into *E. coli* CA434 cells used for conjugation into *C. difficile* (39). Transconjugants were selected on BHIS with *C. difficile* supplement (CC; Oxoid) and Tm (15 μ g/ml) (CCTm) plates. Colonies were restreaked once more on BHIS CCTm plates to ensure plasmid transfer and stored at -80°C for future use. Plasmid maps were drawn with Snapgene software.

TABLE2 Oligonucleotides used in this study

Primer	Sequence (5'–3')	Purpose
OligoAFP316	TTCGTATGGATCCTCTACCCAAGTCTGGTTTC	Cloning of SNAP-tag in to pMTL84153 to create P _{fad} -SNAP; forward primer
oligoAFP325	AGTTCACATATGGATAAAGATTGTGAAATG	Cloning of SNAP-tag in to pMTL84153 to create P _{fad} -SNAP; forward primer
hpdB_SNAP_For_P5	GGAAGAAATGGATAAAGATTGTGAAATGAAGAGAAC	Amplification of SNAP-tag; forward primer
SNAP_V_rev_P6	CCCGGGTACCGAGCTCGAATTTACCCAAGTCTGGTTTC	Amplification of SNAP-tag and for SOE-PCR; reverse primer
hpdB_V_For_P3	CCATATGACCATGATTACGAAGATCTGAATTCGATAGGG	Amplification of <i>hpdBCA</i> promoter region and for SOE-PCR; forward primer
hpdB_SNAP_Rev_P4	AATCTTTATCCATTTCTCCCTCCTTAATC	Amplification of <i>hpdBCA</i> promoter region; reverse primer
<i>gusA</i> -F-P	TTACGTCTGTAGAAACCCC	Amplification of <i>gusA</i> coding sequence; forward primer
<i>gusA</i> -R-P	TCATTGTTTGCCTCCCTG	Amplification of <i>gusA</i> coding sequence; reverse primer
<i>gusA</i> vector forward	ATTCGAGCTCGGTACCCG	Inverse PCR of <i>hpdBCA</i> -SNAP for construction of <i>hpdBCA-gusA</i> ; forward primer
<i>gusA</i> vector reverse	CATTTCTCCCTCCTTAATC	Inverse PCR of <i>hpdBCA</i> -SNAP for construction of <i>hpdBCA-gusA</i> ; reverse primer
PhoZ hpdB Vec F	AAAAGCAGAAATTCGAGCTCGGTACCCG	Inverse PCR of <i>hpdBCA</i> -SNAP for construction of <i>hpdBCA-phoZ</i> ; forward primer
PhoZ hpdB Vec R	ACATTGACGGCATTCTCCCTCCTTAATCTTTC	Inverse PCR of <i>hpdBCA</i> -SNAP for construction of <i>hpdBCA-phoZ</i> ; reverse primer
PhoZ F	GGAAGAAATGCCGCTCAATGTATGGGTAG	Amplification of <i>phoZ</i> coding sequence; forward primer
PhoZ R	GAGCTCGAATTTCTGCTTTTCTTCATTTTG	Amplification of <i>phoZ</i> coding sequence; reverse primer
84153 PhoZ F	AAAAGCAGAAAGATCCTCTAGAGTCGAC	Inverse PCR of pMTL84153 with overhangs for <i>phoZ</i> ; forward primer
84153 PhoZ R	ACATTGACGGCATATGTAACACACCTCC	Inverse PCR of pMTL84153 with overhangs for <i>phoZ</i> ; reverse primer
PhoZ 84153 F	GTTACATATGCCGCTCAATGTATGGGTAG	Amplification of <i>phoZ</i> from pMC358 with overhangs for ligation in to pMTL84153; forward primer
PhoZ 84153 R	TAGAGGATCCTTCTGCTTTTCTTCATTTTG	Amplification of <i>phoZ</i> from pMC358 with overhangs for ligation in to pMTL84153; reverse primer
hpdB P1 SDM F	TTTGCACTAATTATAGAAAGATTAAGGA	For mutation of the P ₁ site on any of the reporter plasmids; forward primer
hpdB P1 SDM R	TAGTCGAAAACCTTTTAAGAATGAAAAA	For mutation of the P ₁ site on any of the reporter plasmids; reverse primer
hpdB P2 SDM F	TTCTCGAGAAAGATTATTTAAAAAAGT	For mutation of the P ₂ site on any of the reporter plasmids; forward primer
hpdB P2 SDM R	TTCTCGAGAAAAATTAACCTTGAA	For mutation of the P ₂ site on any of the reporter plasmids; reverse primer
hpdB IR Rem F	TTTCATTTTAAAAAGTTTTATACTAATTATAGAAAG	Removal of entire inverted repeat upstream of <i>hpdB</i> ; forward primer
hpdB 5'IR Rem F	TAATATACCCTTTTTCATTTCTAAAAAG	Removal of 5' arm of inverted repeat upstream of <i>hpdB</i> ; forward primer
hpdB IR Rem R	hpdB IR Rem R	Removal of 5' arm or entire inverted repeat upstream of <i>hpdB</i> ; reverse primer

Reporter comparison. The strains containing different fusions were grown overnight in BHIS with thiamphenicol (15 μ g/ml). A 2.5-ml portion of the overnight sample was added to 7.5 ml of MM (38) containing 100 mM glucose and grown for 3 h. After 3 h of growth, the cultures were pelleted and washed using 5 ml of fresh MM before repelleting. Pellets were resuspended in 1 ml of MM, and 100 μ l of this suspension was added to 10 ml of each of the following test conditions: MM only, MM plus 2 mg/ml *p*-HPA, and MM plus 0.4 mg/ml tyrosine. Strains were grown in biological triplicates for 4 h before processing required for each of the reporters as described below:

SNAP-tag reporter. Next, 1 ml of culture was used for the OD₅₉₀ reading to assess growth. A 5-ml portion was removed for testing the SNAP-tag, the SNAP-tag substrate TMR-Star (NEB) was added at 0.2 nM to 5-ml portions of cultures, and the mixture was incubated in the dark for 30 min at 37°C. These samples were pelleted and washed with 5 ml of phosphate-buffered saline (PBS). After resuspension of the pellet in 600 μ l of PBS, these samples were transferred to a Lysis Matrix B tube (MP Biomedicals) and lysed for 40 s at 6.0 m/s twice using a FastPrep-24 Classic instrument (MP Biomedicals). After centrifugation, the supernatant was used to quantify SNAP-tag production via SDS-PAGE (10% Bis-Tris protein gel). Gels were imaged using Typhoon Trio Variable Mode Imager System (GE Healthcare), and fluorescence was detected at 580 nm. Analysis was carried out using ImageQuant TL image analysis software. Expression was detected by gating

bands corresponding to the SNAP-tag and a control that was subtracted by measuring pixels from an equal size gate directly above that of the SNAP-tag band. Background controls were calculated based on the background pixel volume/OD₅₉₀; the average was taken for all samples per experiment to give an average background fluorescence that was considered to be the limit of detection (LOD). SNAP-tag production was quantified according to the following formula: product intensity = pixel volume/OD₅₉₀. These experiments were undertaken with at least triplicate biological replicates. The expression from the SNAP-tag fusion with the upstream region of *hpdBCA* containing the wild-type promoter region or deletion of putative promoters, P₁ and/or P₂, was carried out as above; however, all strains were grown in MM in the presence of 2 mg/ml *p*-HPA.

Scaled-up SNAP-tag. The mutated P₁ promoter was tested further with initial growth and 3-h growth steps in BHIS and MM as described above; however, 100 µl was inoculated into 45 ml of MM, without glucose, in each of the test conditions: MM only, MM plus 2 mg/ml *p*-HPA, and MM plus 0.4 mg/ml tyrosine. After 24 h of growth, 1 ml was removed for OD₅₉₀ determination to assess growth, and the remaining culture was pelleted and resuspended in 4 ml of fresh MM. Prepared triplicate biological samples underwent SNAP-tag assay testing as described above. To investigate a putative role of SigL (σ^{54}) in *hpdBCA* expression, the SNAP-tag reporter fusion was transferred into the 630 Δ *erm sigL::erm* strain (28). Due to the growth limitations of strain 630 Δ *erm sigL::erm* strain in minimal medium, BHIS supplemented with 100 mM glucose (BHISG) was chosen for all analyses undertaken. For these experiments, initial overnight cultures were carried out in 10 ml of BHISG before 100 µl of the overnight culture was inoculated into 45 ml of fresh BHISG and grown for 24 h before pelleting and resuspension in 4 ml of BHISG. Prepared biological triplicate samples underwent SNAP-tag assay testing as described above.

***gusA* reporter.** Triplicate biological samples were prepared as described above. After 4 h of growth in MM, 1 ml of the culture was removed to detect the OD₅₉₀, and 1.5-ml samples were pelleted and frozen for later testing as described previously by Mordaka and Heap (31). The OD₅₉₀ after 4 h was used to normalize for growth so that growth was taken into account in the same way as the other reporters. Testing was carried out by resuspension of the pellets in 0.8 ml of buffer Z (60 mM sodium phosphate dibasic heptahydrate, 10 mM KCl, 40 mM sodium phosphate monobasic, 1 mM magnesium sulfate heptahydrate; the pH was adjusted to 7.0, and 50 mM 2-mercaptoethanol was added freshly). Then, 600 µl was transferred to a fresh microcentrifuge tube, and 6 µl of toluene was added, followed by vortexing for 1 min and incubation on ice for 10 min. Samples were heated at 37°C for 30 min with the microcentrifuge caps open. After heating, 120 µl of 6 mM *p*-nitro phenyl- β -D-glucuronide in buffer Z was added to start the assay reaction, followed by incubation for 30 min at 37°C, before the addition of 1 M sodium carbonate to stop the reaction. Samples were centrifuged for 10 min at 10,000 rpm, and supernatants were transferred to a 24-well plate. Absorbance was measured at 405 nm using a Spectramax M3 Multi-Mode plate reader (Molecular Devices) with *gusA* activity calculated using the following formula: final OD₄₀₅/culture OD₅₉₀.

***phoZ* reporter.** Triplicate biological samples were prepared as described above with sample testing carried out according to the method of Edwards *et al.* (15). After 4 h of growth in MM, 1 ml of the culture was taken for OD₅₉₀ measurement, and 2-ml samples were pelleted and frozen for assays. The OD₅₉₀ after 4 h was used to ensure growth was taken into account identically to the other reporters. Testing was carried out by resuspension of the pellets in 500 µl of cold wash buffer (10 mM Tris-HCl [pH 8], 10 mM magnesium sulfate) prior to pelleting and resuspension in 1 ml of assay buffer (1 M Tris-HCl [pH 8], 0.1 mM zinc chloride). Then, 500 µl of the cell suspension was added to 300 µl of fresh assay buffer, 50 µl of 0.1% SDS, and 50 µl of chloroform. Samples were vortexed for 1 min, incubated in a water bath at 37°C for 5 min, and then chilled on ice for 5 min. Samples were then preheated to 37°C, and 100 µl of 0.4% *p*-nitrophenyl phosphate in 1 M Tris-HCl (pH 8) was added. The samples were incubated at 37°C until a yellow color

developed, at which point 100 μ l of stop solution (1 M potassium phosphate monobasic) was added and the time taken for the color to develop was recorded. Samples underwent pelleting by centrifugation at maximum speed for 5 min, and the supernatants were transferred to a spectrophotometer cuvette. The optical densities were measured at 420 and 550 nm using a Spectramax M3 Multi-Mode plate reader with activity quantified as follows: $[\text{OD}_{420} - (\text{OD}_{550} \times 1.75)] \times 1,000 / t \text{ (min)} \times$

$\text{OD}_{590} \times \text{cell volume (ml)}$.

HPLC. Supernatants were obtained from *C. difficile* cultures grown for 10 h in MM and in MM supplemented with either 2 mg/ml *p*-HPA or 0.4 mg/ml tyrosine. The MM was altered from that used above to contain defined amino acids as described by Karasawa *et al.* (40), with the exception of glycine being increased to 0.4 mg/ml, and tyrosine was only included in samples as indicated. In addition, glucose was not included to ensure maximum usage of amino acids such as tyrosine. Supernatants were filter sterilized using 0.2 μ m filters and stored at -80°C. Defrosted culture supernatants were mixed in a 1:1 ratio with methanol-water, transferred to HPLC tubes, and processed immediately by HPLC. Separations were performed utilizing an Acclaim 120 (Thermo Fisher) C_{18} 5- μ m analytical (4.6 by 150 mm) column and a mobile phase consisting of ammonium formate (10 mM [pH 2.7]) and menthol (vol/vol; 40:60) at a flow rate of 2 ml/min. Tyrosine, *p*-HPA, and *p*-cresol were detected using a photodiode array detector (UV-PDA; DAD 3000) set at 280 nm. The peak identity was confirmed by measuring the retention times of commercially available tyrosine, *p*-HPA, and *p*-cresol and determination of the absorbance spectra using the UV-PDA detector. A calibration curve of each compound was generated by Chromeleon (Dionex Software) with known amounts of the reference standards (0 to 100 mg/ml) in methanol-water (1:1 [vol/vol]) injected onto the column, from which the concentrations in the samples were determined. Samples from three independent biological replicates were analyzed compared to medium controls and standard curves. The data were analyzed in GraphPad Prism7, and statistical analysis was performed in Stata16 using linear regression analysis $P < 0.05$ was considered statistically significant.

RNA extractions and qRT-PCR. Total RNA was isolated from *C. difficile* 630 Δ *erm* and *sigL* mutant strains grown in BHISG medium to an OD_{590} of 1.2 ± 0.2 . RNAprotect (Qiagen) was added to exponential or early-stationary-phase cultures, cells were pelleted at 4°C, and pellets were stored at -80°C. After defrosting on ice, the pellets were processed using the RNAPro extraction kit (MP Biomedicals). Samples were DNase treated twice with 10 U of Turbo DNase (Invitrogen) and 40 U of RNase inhibitor (NEB). Samples were purified by using a Qiagen RNase kit. Samples were checked for DNA contamination by PCR. RNA integrity was determined by using an Agilent RNA Nano Bioanalyser Chip (Agilent), followed by cDNA synthesis via reverse transcription and quantitative real-time PCR analysis. cDNA was synthesized from 200 ng of total RNA. Random primers were annealed to the RNA by heating at 80°C for 5 min, followed by cooling slowly to room temperature. A 10 mM concentration of each deoxynucleoside triphosphate and 200 U of reverse transcriptase III in 1X RT buffer (Thermo Fisher) were added in reaction mixtures to a final volume of 25 μ l. Reverse transcriptase-negative samples were included as controls (RT-). Real-time quantitative PCR was performed in a 20 μ l reaction volume containing 10 ng (for *hpdC*) or 200 pg (for 16S rRNA) of cDNAs, along with 12.75 μ l of the SYBR PCR master mix (Applied Biosystems), and 10 μ M concentrations of each gene-specific primer. Amplification and detection were performed by using an ABI 7500 thermocycler. In each sample, the quantity of cDNA of a gene was determined by subtracting the RT- and then normalized to the quantity of cDNA of the 16S rRNA gene. The relative change in gene expression was recorded as the ratio of normalized target concentrations (determined using the threshold cycle [$\Delta\Delta C_T$] method) as previously described (27). The relative change in gene expression was recorded as the ratio of normalized target concentrations ($\Delta\Delta C_T$), expressed as $2^{-\Delta\Delta C_T}$ (41) for triplicate technical and duplicate biological replicates. The data

were analyzed in GraphPad Prism7, and statistical analysis was performed in Stata16 using linear regression analysis ($P < 0.05$ was considered statistically significant).

Statistical analysis. The expression data from the transcriptional reporter fusions were analyzed for significant differences in expression under different conditions. The data were transformed using \log_{10} to approximate a normal distribution, and then linear regression analysis was used to determine significant differences (i) between growth conditions, including the addition of tyrosine and *p*-HPA, compared to the minimal medium control; (ii) between fold changes in expression between reporters and in comparison to qRT-PCR in the presence of *p*-HPA; (iii) between the wild-type and mutated -10 promoter regions; (iv) between the wild-type and the *sigL* mutant; and (v) in the fold change ($2^{-\Delta\Delta CT}$) between the control media and the addition of *p*-HPA, as well as between the wild-type and the *sigL* mutant strain. HPLC analysis was analyzed by linear regression to determine whether there were any differences in the concentration of tyrosine, *p*-HPA, and *p*-cresol detected in the cultures, and the correlation between the *phoZ* activity and *p*-HPA concentration was analyzed using linear regression. Statistical differences are indicated in the figures. The coefficient of variance (COV) indicates whether the difference between the test samples is higher (positive number) or lower (negative number) than the reference. Analysis was performed using Stata15 (StataCorp); statistical test summaries are available in File S1 in the supplemental material.

SUPPLEMENTAL MATERIAL Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 2, PDF file, 0.2 MB.

ACKNOWLEDGEMENTS We thank Shonna McBride for providing pMC358 for use in the study.

This study was supported by the Medical Research Council (LSHTM studentship MR/N013638/1). Funding for B.W.W. and A.F.-P. was provided Medical Research Council grant MR/K000551/1. Funding for L.F.D. was provided by an ISSF Fellowship from the Wellcome Trust (105609/Z/14/Z) and an Athena Swan Career Restart Fellowship (from London School of Hygiene and Tropical Medicine).

We declare that no competing interests exist.

REFERENCES

1. Desai K, Gupta SB, Dubberke ER, Prabhu VS, Browne C, Mast TC. 2016. Epidemiological and economic burden of *Clostridium difficile* in the United States: estimates from a modeling approach. *BMC Infect Dis* 16:303. <https://doi.org/10.1186/s12879-016-1610-3>.
2. Deakin LJ, Clare S, Fagan RP, Dawson LF, Pickard DJ, West MR, Wren BW, Fairweather NF, Dougan G, Lawley TD. 2012. The *Clostridium difficile* spo0A gene is a persistence and transmission factor. *Infect Immun* 80:2704–2711. <https://doi.org/10.1128/IAI.00147-12>.
3. Dawson LF, Valiente E, Donahue EH, Birchenough G, Wren BW. 2011. Hypervirulent *Clostridium difficile* PCR-ribotypes exhibit resistance to widely used disinfectants. *PLoS One* 6:e25754. <https://doi.org/10.1371/journal.pone.0025754>.
4. Rohlfing AE, Eckenroth BE, Forster ER, Kevorkian Y, Donnelly ML, Benitode la Puebla H, Doublie S, Shen A. 2019. The CspC pseudoprotease regulates germination of *Clostridioides difficile* spores in response to multiple environmental signals. *PLoS Genet* 15:e1008224. <https://doi.org/10.1371/journal.pgen.1008224>.

5. Britton RA, Young VB. 2014. Role of the intestinal microbiota in resistance to colonization by *Clostridium difficile*. *Gastroenterology* 146: 1547–1553. <https://doi.org/10.1053/j.gastro.2014.01.059>.
6. Vardakas KZ, Polyzos KA, Patouni K, Rafailidis PI, Samonis G, Falagas ME. 2012. Treatment failure and recurrence of *Clostridium difficile* infection following treatment with vancomycin or metronidazole: a systematic review of the evidence. *Int J Antimicrob Agents* 40:1–8. <https://doi.org/10.1016/j.ijantimicag.2012.01.004>.
7. Hecht G, Pothoulakis C, LaMont JT, Madara JL. 1988. *Clostridium difficile* toxin A perturbs cytoskeletal structure and tight junction permeability of cultured human intestinal epithelial monolayers. *J Clin Invest* 82: 1516–1524. <https://doi.org/10.1172/JCI113760>.
8. Just I, Fritz G, Aktories K, Giry M, Popoff MR, Boquet P, Hegenbarth S, vonEichel-Streiber C. 1994. *Clostridium difficile* toxin B acts on the GTP binding protein Rho. *J Biol Chem* 269:10706–10712.
9. Vedantam G, Clark A, Chu M, McQuade R, Mallozzi M, Viswanathan VK. 2012. *Clostridium difficile* infection. *Gut Microbes* 3:121–134. <https://doi.org/10.4161/gmic.19399>.
10. Saito Y, Sato T, Nomoto K, Tsuji H. 2018. Identification of phenol- and *p*-cresol-producing intestinal bacteria by using media supplemented with tyrosine and its metabolites. *FEMS Microbiol Ecol* 94:fiy125. <https://doi.org/10.1093/femsec/fiy125>.
11. Passmore IJ, Letertre MPM, Preston MD, Bianconi I, Harrison MA, Nasher F, Kaur H, Hong HA, Baines SD, Cutting SM, Swann JR, Wren BW, Dawson LF. 2018. *para*-Cresol production by *Clostridium difficile* affects microbial diversity and membrane integrity of Gram-negative bacteria. *PLoS Pathog* 14:e1007191. <https://doi.org/10.1371/journal.ppat.1007191>.
12. Russell WR, Duncan SH, Scobbie L, Duncan G, Cantlay L, Calder AG, Anderson SE, Flint HJ. 2013. Major phenylpropanoid-derived metabolites in the human gut can arise from microbial fermentation of protein. *Mol Nutr Food Res* 57:523–535. <https://doi.org/10.1002/mnfr.201200594>.
13. Selmer T, Andrei PI. 2001. *p*-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. *Eur J Biochem* 268:1363–1372. <https://doi.org/10.1046/j.1432-1327.2001.02001.x>.
14. Dawson LF, Donahue EH, Cartman ST, Barton RH, Bundy J, McNerney R, Minton NP, Wren BW. 2011. The analysis of *para*-cresol production and tolerance in *Clostridium difficile* 027 and 012 strains. *BMC Microbiol* 11:86. <https://doi.org/10.1186/1471-2180-11-86>.
15. Edwards AN, Pascual RA, Childress KO, Nawrocki KL, Woods EC, McBride SM. 2015. An alkaline phosphatase reporter for use in *Clostridium difficile*. *Anaerobe* 32:98–104. <https://doi.org/10.1016/j.anaerobe.2015.01.002>.
16. Ransom EM, Ellermeier CD, Weiss DS. 2015. Use of mCherry Red fluorescent protein for studies of protein localization and gene expression in *Clostridium difficile*. *Appl Environ Microbiol* 81:1652–1660. <https://doi.org/10.1128/AEM.03446-14>.
17. Cassona CP, Pereira F, Serrano M, Henriques AO. 2016. A fluorescent reporter for single cell analysis of gene expression in *Clostridium difficile*, p 69–90. In Roberts AP, Mullany P (ed), *Clostridium difficile: methods and protocols*. Springer, New York, NY.
18. Drepper T, Eggert T, Circolone F, Heck A, Krauss U, Guterl J-K, Wendorff M, Losi A, W, Jaeger K-E. 2007. Reporter proteins for *in vivo* fluorescence without oxygen. *Nat Biotechnol* 25:443–445. <https://doi.org/10.1038/nbt1293>.
19. Buckley AM, Petersen J, Roe AJ, Douce GR, Christie JM. 2015. LOV-based reporters for fluorescence imaging. *Curr Opin Chem Biol* 27:39–45. <https://doi.org/10.1016/j.cbpa.2015.05.011>.
20. Streett HE, Kalis KM, Papoutsakis ET. 2019. A Strongly fluorescing anaerobic reporter and protein-tagging system for clostridium organisms based on the fluorescence-activating and absorption-shifting tag protein (FAST). *Appl Environ Microbiol* 85:e00622-19. <https://doi.org/10.1128/AEM.00622-19>.
21. Pereira FC, Saujet L, Tome AR, Serrano M, Monot M, Couture-Tosi E, Martin-Verstraete I, Dupuy B, Henriques AO. 2013. The spore differentiation pathway in the

- enteric pathogen *Clostridium difficile*. PLoS Genet 9:e1003782. <https://doi.org/10.1371/journal.pgen.1003782>.
22. Karlsson S, Dupuy B, Mukherjee K, Norin E, Burman LG, Akerlund T. 2003. Expression of *Clostridium difficile* toxins A and B and their sigma factor TcdD is controlled by temperature. Infect Immun 71:1784–1793. <https://doi.org/10.1128/iai.71.4.1784-1793.2003>.
 23. Dupuy B, Sonenshein AL. 1998. Regulated transcription of *Clostridium difficile* toxin genes. Mol Microbiol 27:107–120. <https://doi.org/10.1046/j.1365-2958.1998.00663.x>.
 24. Govind R, Vedyappan G, Rolfe RD, Dupuy B, Fralick JA. 2009. Bacteriophage-mediated toxin gene regulation in *Clostridium difficile*. J Virol 83:12037–12045. <https://doi.org/10.1128/JVI.01256-09>.
 25. Fagan RP, Fairweather NF. 2011. *Clostridium difficile* has two parallel and essential secretion systems. J Biol Chem 286:27483–27493. <https://doi.org/10.1074/jbc.M111.263889>.
 26. Cassona CP, Pereira F, Serrano M, Henriques AO. 2016. A fluorescent reporter for single cell analysis of gene expression in *Clostridium difficile*. Methods Mol Biol 1476:69–90. https://doi.org/10.1007/978-1-4939-6361-4_6.
 27. Saujet L, Monot M, Dupuy B, Soutourina O, Martin-Verstraete I. 2011. The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*. J Bacteriol 193:3186–3196. <https://doi.org/10.1128/jb.00272-11>.
 28. Dubois T, Dancer-Thibonnier M, Monot M, Hamiot A, Bouillaut L, Soutourina O, Martin-Verstraete I, Dupuy B. 2016. Control of *Clostridium difficile* physiopathology in response to cysteine availability. Infect Immun 84:2389–2405. <https://doi.org/10.1128/IAI.00121-16>.
 29. Nie X, Yang B, Zhang L, Gu Y, Yang S, Jiang W, Yang C. 2016. PTS regulation domain-containing transcriptional activator CelR and sigma factor σ^{54} control cellobiose utilization in *Clostridium acetobutylicum*. Mol Microbiol 100:289–302. <https://doi.org/10.1111/mmi.13316>.
 30. Bervoets I, Charlier D. 2019. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. FEMS Microbiol Rev 43:304–339. <https://doi.org/10.1093/femsre/fuz001>.
 31. Mordaka PM, Heap JT. 2018. Stringency of synthetic promoter sequences in *Clostridium* revealed and circumvented by tuning promoter library mutation rates. ACS Synth Biol 7:672–681. <https://doi.org/10.1021/acssynbio.7b00398>.
 32. Soutourina OA, Monot M, Boudry P, Saujet L, Pichon C, Sismeiro O, Semenova E, Severinov K, Le Bouguenec C, Coppee JY, Dupuy B, Martin-Verstraete I. 2013. Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. PLoS Genet 9:e1003493. <https://doi.org/10.1371/journal.pgen.1003493>.
 33. El Meouche I, Peltier J, Monot M, Soutourina O, Pestel-Caron M, Dupuy B, Pons JL. 2013. Characterization of the SigD regulon of *C. difficile* and its positive control of toxin production through the regulation of *tcdR*. PLoS One 8:e83748. <https://doi.org/10.1371/journal.pone.0083748>.
 34. Saujet L, Pereira FC, Serrano M, Soutourina O, Monot M, Shelyakin PV, Gelfand MS, Dupuy B, Henriques AO, Martin-Verstraete I. 2013. Genome wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*. PLoS Genet 9:e1003756. <https://doi.org/10.1371/journal.pgen.1003756>.
 35. Steglich M, Hofmann JD, Helmecke J, Sikorski J, Spröer C, Riedel T, Bunk B, Overmann J, Neumann-Schaal M, Nübel U. 2018. Convergent loss of ABC transporter genes from *Clostridioides difficile* genomes is associated with impaired tyrosine uptake and *p*-cresol production. Front Microbiol 9:901. <https://doi.org/10.3389/fmicb.2018.00901>.
 36. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, Sayeeda Z, Lo E, Assempour N, Berjanskii M, Singhal S, Arndt D, Liang Y, Badran H, Grant J, Serra-Cayuela A, Liu Y, Mandal R, Neveu V, Pon A, Knox C, Wilson M, Manach C, Scalbert A. 2018. HMDB 4.0: the human metabolome

- database for 2018. *Nucleic Acids Res* 46:D608–D617. <https://doi.org/10.1093/nar/gkx1089>.
37. Theriot CM, Young VB. 2014. Microbial and metabolic interactions between the gastrointestinal tract and *Clostridium difficile* infection. *Gut Microbes* 5:86–95. <https://doi.org/10.4161/gmic.27131>.
 38. Cartman ST, Minton NP. 2010. A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. *Appl Environ Microbiol* 76:1103–1109. <https://doi.org/10.1128/AEM.02525-09>.
 39. Purdy D, O’Keeffe TAT, Elmore M, Herbert M, McLeod A, Bokori-Brown M, Ostrowski A, Minton NP. 2002. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol Microbiol* 46:439–452. <https://doi.org/10.1046/j.1365-2958.2002.03134.x>.
 40. Karasawa T, Ikoma S, Yamakawa K, Nakamura S. 1995. A defined growth medium for *Clostridium difficile*. *Microbiology* 141:371–375. <https://doi.org/10.1099/13500872-141-2-371>.
 41. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>.
 42. Hussain HA, Roberts AP, Mullany P. 2005. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630 Δ erm) and demonstration that the conjugative transposon Tn916 Δ E enters the genome of this strain at multiple sites. *J Med Microbiol* 54:137–141. <https://doi.org/10.1099/jmm.0.45790-0>.

SUPPLEMENTAL MATERIAL

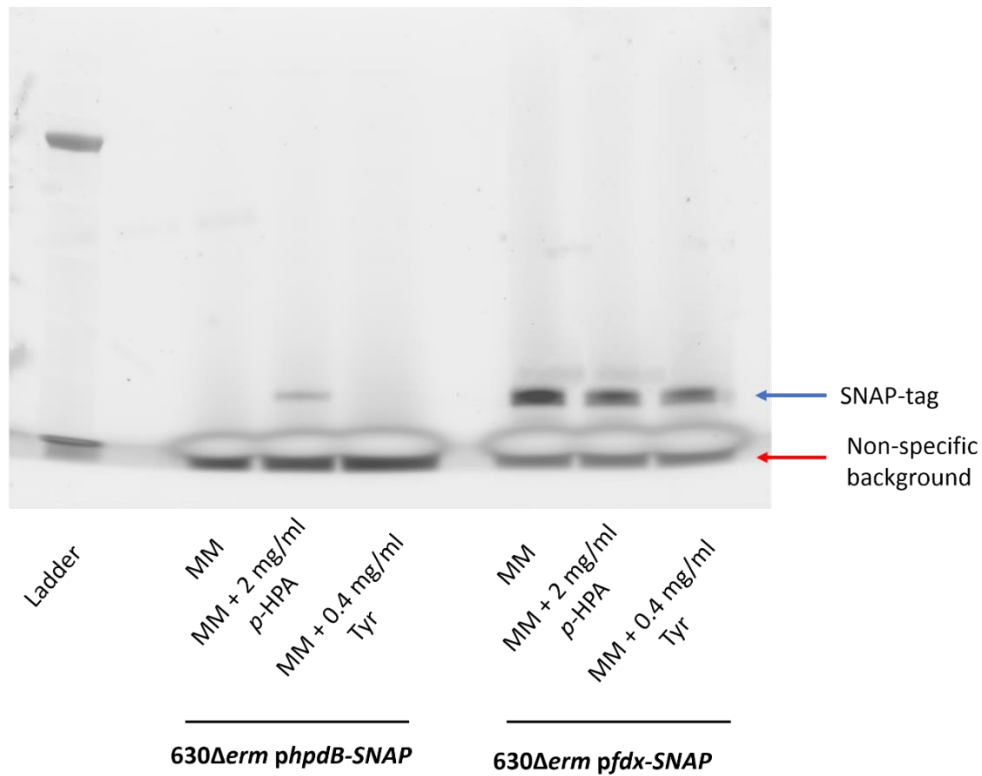


Figure S1. Uncropped SNAP-tag gel. Uncropped SNAP-tag gel. Strains 630 Δ erm *P_{phpdB}*-SNAP and 630 Δ erm *P_{pfdx}*-SNAP were grown for four hours in MM alone, or MM supplemented with 2 mg/ml *p*-HPA or 0.4 mg/ml Tyrosine. Samples were processed with the fluorescent substrate TMR-Star prior to being run on an SDS-PAGE gel and imaged using a Typhoon Trio Variable Mode Imager System (GE Healthcare) with fluorescence detected at 580 nm. Background fluorescence is labelled with a red arrow. The blue arrow indicates the SNAP tag signal.

3.2 Additional results and discussion

In addition to the publication, I assessed the fold-change in expression of *hpdC* by qRT-PCR with samples prepared identically to those used for the reporters, i.e., 630 Δ *erm* grown with and without *p*-HPA in MM. I extracted RNA and then carried out qRT-PCR using primers for *hpdC* as well as for *16S* for normalisation by $2^{-\Delta\Delta Ct}$ analysis (231). An average fold change of 47.6 (± 15.9) was found for qRT-PCR (Fig. 9), whilst the fold changes for the reporters were as follows: *phoZ*: 49.5 \pm 8.7, *gusA*: 41.5 \pm 23.4, and SNAP-tag: 2.0 \pm 0.6. When these fold changes were compared by linear regression to the qRT-PCR both *phoZ* and *gusA* were not found to be significantly different (*gusA*; $p=0.623$, *phoZ*; $p=0.781$) whereas all three were significantly higher than the SNAP-tag ($p<0.001$). This suggests that both the *gusA* and *phoZ* reporters, under these conditions, are of a similar sensitivity to the gold standard qRT-PCR as well as having much simpler methods and being significantly cheaper. Therefore, this demonstrates clear advantages of *phoZ* and *gusA* as transcriptional fusions compared to the SNAP-tag.

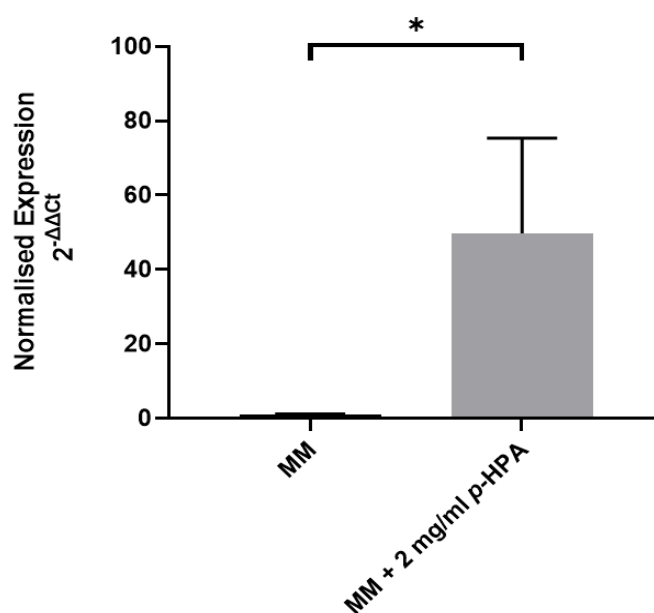


Fig 9. qRT-PCR of 630 Δ *erm* *hpdC* expression in response to exogenous *p*-HPA. 630 Δ *erm* was grown in minimal media for four hours with and without 2 mg/ml *p*-HPA before undergoing RNA extraction. qRT-PCR was carried out for expression of *hpdC* which was normalised to the *16S rRNA* internal control ($2^{-\Delta\Delta Ct}$) (231). Statistical analysis by linear regression was used to determine whether expression of *hpdC* was significantly higher in the presence of *p*-HPA than in the absence of it. Data represents three independent replicates; errors bars represent standard deviation. * $p<0.05$

Finally, bioinformatics searches of the 630 Δ *erm* genome were carried out to identify the locations of binding sites that matched that of the inverted repeat found upstream of *hpdBCA*. The aim of this was to identify the regulator(s) involved in the induction of *hpdBCA* expression in response to *p*-HPA as transcriptional regulators are often autoregulatory (232) so it was hypothesised the same binding site would be upstream of the regulator in question. Searches were carried out by Dr Mark Preston using the motif AAAAAG-n13-CTTTTT. Dr Preston ran searches looking for either an exact match to the motif or a single base pair mismatch located within 200 base pairs of the start codon of any gene. These searches identified six exact motif hits, of which three were located upstream of a start codon and one of which was the motif located upstream of *hpdB*. The other two binding sites were located upstream of genes: *CD3256*, valyl-tRNA synthetase, and *CD1951*, a putative Acyl-CoA N-acyltransferase. In addition, to the exact matches only a single hit was found when searching for motifs with a one base pair mismatch within 200 bps upstream of the start codon and this was also located upstream of *CD1951*. Unfortunately, neither of these genes are good candidates for acting as regulators of *hpdBCA* expression and it was not possible to run wider searches for motifs with more than a single base pair mismatch in this study, therefore this should be a priority for future research.

Chapter 4: Production of *p*-cresol by decarboxylation of *p*-HPA by all five lineages of *Clostridioides difficile* provides a growth advantage.

4.1 Research paper cover sheet

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – STUDENT DETAILS

Student ID Number	LSH366976	Title	Mr
First Name(s)	Mark		
Surname/Family Name	Harrison		
Thesis Title	Characterising regulation of virulence factors affecting the interaction between <i>Clostridioides difficile</i> and the intestinal microbiome		
Primary Supervisor	Dr Lisa Dawson		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	an	Was the work subject to academic peer review? Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.


SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Frontiers in Cellular and Infection Microbiology
Please list the paper's authors in the intended authorship order:	Mark A Harrison, Harparkash Kaur, Brendan W Wren and Lisa F Dawson
Stage of publication	Undergoing revision

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	The project was conceived by Dr Dawson, then together we developed a plan to test the hypotheses explored in this article. The methodologies were set up by myself, building on our earlier paper (Chapter 3). All of the lab work was carried out by myself with the exception of the HPLC which was carried out by Dr Harparkash Kaur. I initially wrote the first draft of the manuscript which was revised and edited by Dr Dawson into its finalised first draft form, this draft was sent to other authors and subsequently updated prior to submission
--	---

SECTION E

Student Signature	
Date	15/09/2021

Supervisor Signature	
Date	20/9/2021

Production of *p*-cresol by decarboxylation of *p*-HPA by all five lineages of *Clostridioides difficile* provides a growth advantage.

Mark A Harrison¹, Harparkash Kaur², Brendan W Wren¹ and Lisa F Dawson^{1*} Affiliations:

1: Department of Infection Biology, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT

2: Department of Clinical Research, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT

Corresponding Author (*): Dr Lisa Dawson, email address: lisa.dawson@lshtm.ac.uk

Keywords: *Clostridioides difficile*, *para*-cresol, *p*-cresol, *p*-HPA, virulence factors, infection pathogenesis, dysbiosis, microbiome

Running title: *p*-cresol production in *Clostridioides difficile*

Number of tables: 2

Number of figures: 7

Number of supplementary tables: 1

Number of supplementary figures: 6

Abstract

Clostridioides difficile is the leading cause of antibiotic associated diarrhoea and is capable of causing severe symptoms such as pseudomembranous colitis and toxic megacolon. An unusual feature of *C. difficile* is the distinctive production of high levels of the antimicrobial compound *para*-cresol. *p*-cresol production provides *C. difficile* with a competitive colonisation advantage over gut commensal species, in particular Gram-negative species. *p*-cresol is produced by the conversion of *para*-hydroxyphenylacetic acid (*p*-HPA) via the actions of HpdBCA decarboxylase coded by the *hpdBCA* operon. *C. difficile* can produce *p*-cresol either from *p*-tyrosine metabolism, via *p*-HPA, or from exogenous *p*-HPA. Host cells and certain bacterial species produce *p*-HPA, however, the effects of *p*-HPA on viability of *C. difficile* and other gut microbiota is unknown. We show here that production of *p*-cresol is controlled at the transcriptional level by representative strains from all five *C. difficile* clades, whereby expression of the HpdBCA decarboxylase is induced by *p*-HPA leading to high level *p*-cresol production. However, strain specific differences in *p*-cresol production were attributed to modulations in *p*-tyrosine fermentation; representatives of clade 3 (CD305) and clade 5 (M120) produced the highest levels of *p*-cresol via this pathway. The HpdBCA decarboxylase in *C. difficile* populations is produced ubiquitously in response to exogenous *p*-HPA, however, deletion of both the HpdBCA decarboxylase and the global transcriptional regulator CodY reduced *p*-cresol production *in vitro*. Excess exogenous *p*-HPA (≥ 2 mg/ml) negatively impacts the growth of representative Gram-negative gut bacteria as well as growth of *C. difficile*, potentially by disruption of membrane permeability. This study provides insights into the importance of the HpdBCA decarboxylase in *C. difficile* pathogenesis, both in terms of *p*-cresol production and detoxification of *p*-HPA, highlighting its importance to cell survival and as a highly specific therapeutic target for inhibition of *p*-cresol production across the *C. difficile* species.

Introduction

Clostridioides difficile is a major nosocomial pathogen that causes significant mortality and morbidity globally, with an estimated 12,800 deaths in 2017 in the USA alone (1). *C. difficile* primarily affects patients who have been treated with broad spectrum antibiotics for an unrelated condition, resulting in damage to the gut microbiome and a resultant loss of colonisation resistance to *C. difficile*, with patients who receive longer courses of therapy at greater risk than those receiving short courses (2). Whilst treatment of *C. difficile* is effective with either metronidazole, vancomycin or fidaxomicin, a major feature of *C. difficile* infection (CDI) is the high proportion of patients (20-30%) who suffer from recurrence (often multiple recurrences) either as a result of reinfection or relapse (3). As a result, treatments that are highly specific for *C. difficile* that prevent recurrence are a current imperative.

Para-cresol (4-methylphenol) is an antimicrobial compound produced by *C. difficile* either through the fermentation of *p*-tyrosine (4-hydroxyphenylalanine) via the intermediate *para*-hydroxyphenylacetic acid (*p*-HPA) (4, 5), or via the conversion of exogenous *p*-HPA to *p*-cresol (6). Conversion of *p*-HPA to *p*-cresol occurs via the actions of the HpdBCA decarboxylase (7), encoded by the *hpdBCA* operon (4). This operon is formed of three genes, each encoding a subunit of the decarboxylase, with all three subunits being required for *p*-cresol production (4). *C. difficile* is highly tolerant to *p*-cresol, with the hypervirulent R20291 (Ribotype (RT) 027) strain significantly more tolerant than 630 strain (RT012) (4). *p*-cresol selectively inhibits growth of Gram-negative bacteria of the Gammaproteobacteria family including; *Escherichia coli*, *Proteus mirabilis* and *Klebsiella oxytoca*, whilst Gram-positive species such as *Lactobacillus fermentum*, *Enterococcus faecium* and *Bifidobacterium adoscelentis* are significantly more tolerant to *p*-cresol (8). Furthermore, a *p*-cresol null mutant has a fitness defect in a mouse relapse model of CDI, (8) highlighting the importance of this pathway to *C. difficile* during infection. *p*-cresol production among the resident gut bacteria is unusual with only three other identified bacterial species produce *p*-cresol to relatively high levels; *Blautia hydrogenotrophica*, *Olsenella uli* and *Romboutsia lituseburensis* (9). These gut commensals produced more *p*-cresol than *C. difficile* when

cultured in rich media over a period of six days (9), however *C. difficile* is inefficient at metabolising *p*-tyrosine to produce *p*-cresol in rich media (4), and requires the addition of *p*-HPA, an intermediate in the pathway (4, 6). Therefore, the conditions used by Saito *et al.* likely hindered the production of *p*-cresol by *C. difficile* (9). Whilst we have demonstrated the importance of *p*-cresol to *C. difficile* virulence, the effect of *p*-HPA on growth and pathogenesis is unknown. *p*-HPA is produced via *p*-tyrosine metabolism by a variety of bacteria including *Acinetobacter*, *Klebsiella* and *Clostridium* species as well as being produced by mammalian cells and is detected in all human tissues and biofluids (10). We recently showed that exogenous *p*-HPA induces expression of the *hpdBCA* operon, which in turn induces high level *p*-cresol production (6). Therefore, exogenous *p*-HPA produced either by human cells, the gut microbiome, or a combination of both, may provide *C. difficile* with a reservoir of *p*-HPA to enhance *p*-cresol production and thus provide *C. difficile* with a competitive advantage over a number of commensal species which are vital for colonisation resistance.

The *C. difficile* species is encompassed by five distinct clades (11, 12). Clade 1 contains strain 630, a ribotype (RT) 012 strain, isolated from a patient in the 1980's with severe pseudomembranous colitis and the first strain to be genome sequenced, however this clade also contains a diverse range of both toxigenic and nontoxigenic strains (13, 14). Clade 2 includes hypervirulent strains such as R20291 (RT027), which was responsible for significant outbreaks worldwide from 2004 (15, 16) and remains a prevalent global ribotype (17-20). Clade 3 includes strain CD305 (RT023), a recently emerged hypervirulent lineage prevalent in the UK (21). Clade 4 strains including M68, a toxin A negative RT017 strain (17), are prevalent in Asia (22-25). Finally, Clade 5 including strain M120 (RT078) have been commonly isolated in animals, however, more recently RT078 caused human outbreaks in the Netherlands (26) and are one of the most common community-acquired ribotype in Europe (17, 27). There have been suggestions of a further three clades of *C. difficile* (C-I, C-II and C-III), with C-I identified as being able to cause CDI (28), however, these cryptic clades fall well below the average nucleotide identity (ANI) analysis threshold, indicating that these strains are in fact likely to be another species (29).

In this study, we demonstrate that representative strains from all five clades of *C. difficile* were found to produce high levels of *p*-cresol, with expression of the *hpdBCA* operon induced in the presence of *p*-HPA. Expression of the HpdBCA decarboxylase was ubiquitous across all cells within a population of *C. difficile* in response to *p*-HPA. Exogenous *p*-HPA inhibits *C. difficile* growth and enhances sporulation, as well as inhibiting growth of a number of representative commensal bacterial strains from the gut, especially Gram-negative species. Potentially, *p*-HPA induces perturbations in the cell membrane integrity of these bacteria, through a combination of reduced pH and cell membrane disruption. Together, these findings demonstrate the importance of conversion of *p*-HPA to *p*-cresol and validates it as a specific therapeutic target that should be effective against all *C. difficile* strains.

Materials and Methods

Bacterial strains and growth conditions

All bacterial strains used in this study are listed in table 1. Routine growth of all species was carried out using Brain Heart Infusion (Oxoid) supplemented with 5 gL⁻¹ yeast extract (Sigma) and 0.05% L-Cysteine (Sigma) (BHIS). All strains were grown in anaerobic conditions in a Modular Atmosphere Control System 500 (Don Whitney Scientific) at 37°C. All media underwent a minimum of 4 hour pre-equilibration in anaerobic conditions prior to inoculation.

Growth analysis Growth of *C. difficile* strains (630Δ*erm* and 630Δ*erm hpdC::CT*) and intestinal species (*Escherichia coli*, *Klebsiella oxytoca*, *Proteus mirabilis*, *Bifidobacterium adoscelentis*, *Lactobacillus fermentum* and *Enterococcus faecium*) was assessed for a minimum of 8 hours. Each strain was grown overnight in BHIS before being back diluted to an OD_{590nm} of 0.5, 1 ml of which was added to 10 ml of BHIS supplemented with 0, 1, 2, 3 and 4 mg/ml *p*-HPA. To establish whether *p*-HPA's toxicity was related to acidification of the media, the pH of BHIS media supplemented with *p*-HPA (0, 1, 2, 3 and 4 mg/ml) was measured using a Mettler Toledo Seveneasy pH Meter. To mimic the acidity that results from *p*-HPA supplementation, the pH of BHIS was adjusted using hydrochloric acid to match the pH of

media supplemented with 0, 1, 2, 3 and 4 mg/ml *p*-HPA, these correspond to pH: 6.6, 6.2, 5.8 and 5.4, respectively. Furthermore, following completion of growth curves at this pH the *C. difficile* cultures were filter sterilised using 0.22 µm filter (Millipore) and the supernatant's pH was tested. All growth curves were carried out in a minimum of biological triplicate with the OD_{590nm} determined every hour for 8 hours with a final reading at 24 hours with the exception of *C. difficile* grown in the presence of *p*-HPA, which was not read at 24 hours due to *C. difficile* cultures clumping in the presence of *p*-HPA making an accurate OD_{590nm} reading impossible.

Sporulation assays

Sporulation assays were carried out by growth of 630Δ*erm* and 630Δ*erm hpdC::CT* in BHIS media supplemented with 0, 1, 2 and 3 mg/ml *p*-HPA. Each strain was grown overnight in BHIS and then backdiluted to an OD_{590nm} of 0.5, 1 ml of which was added to 10 ml of BHIS supplemented with 0, 1, 2 or 3 mg/ml *p*-HPA. These cultures were grown for 24 hours. Where cultures had aggregated in the presence of *p*-HPA these aggregations were dispersed by vigorous pipetting and vortexing. Colony forming units per millilitre (CFU/ml) were determined for the total cell count and the spore count using 1 ml of the culture for each. Spores counts were determined from the total cell counts by heat killing of vegetative cells at 65°C for 25 minutes. Dilutions were plated on to BHIS supplemented with 0.1% sodium taurocholate in technical triplicate and counted the following day. Sporulation assays were carried out in a biological triplicate. Sporulation percentages were calculated and analysed by Spearman rank order correlation. Statistical analysis by regression analysis was carried out to determine significant differences between total cell counts in each condition tested. $p < 0.05$ was considered significantly different.

Phosphate release assay

This assay was carried out as per Passmore *et al.* (8) using the phosphate assay kit (Abcam). Briefly, species to be tested were grown overnight in 10 ml BHIS, 5 ml of which was pelleted and washed twice

in 5 ml Tris-buffered saline (TBS, 50 mM Tris-HCl pH 7.5, 150 mM NaCl) before suspension in 5 ml TBS. OD_{590nm} was measured and suspensions were back-diluted to an OD_{590nm} of 1.0. 500 μ l aliquots were taken and pelleted for 2 minutes at 17000 x *g* and the supernatant removed. Pellets were then re-suspended in 500 μ l of TBS with 0, 1, 2 or 3 mg/ml *p*-HPA, additionally extra suspensions were prepared in TBS alone to determine the maximum intracellular phosphate release which was found by boiling the sample for 15 minutes. After 30, 60 and 90 minutes 100 μ l was removed from the anaerobic chamber, pelleted and 30 μ l of the supernatant was added to 170 μ l H₂O and 30 μ l ammonium molybdate and malachite green reagent in a 96 well plate. After the final 90 minute samples were added to the reagent the plate was incubated at room temperature for 30 minutes before the absorbance at OD_{650nm} was read by a Spectramax M3 plate reader. Final results were calculated by subtraction of the OD_{650nm} from the media control from each of the media conditions. The pH of the TBS with 1, 2 and 3 mg/ml *p*-HPA was measured to be pH 6.6, 4.1 and 3.8 respectively. The phosphate release assays were repeated in TBS with matched pH (to the *p*-HPA samples, using hydrochloric acid). All assays were performed in technical duplicate and biological triplicate. Statistical analysis was undertaken by ANOVA using GraphPad Prism 8 software to compare phosphate release in each condition to the TBS control. $p < 0.05$ was considered significantly different.

RNA extraction and qRT-PCR

Total RNA was isolated from all five representative *C. difficile* strains grown in BHIS media up to an OD_{590nm} of 0.6-0.7, RNA protect was added in a 1:1 ratio and incubated at 37°C for 5 minutes. Cells were pelleted by centrifugation at 17000 x *g* at 4°C and pellets were stored immediately at -80°C. All centrifugations were carried out at 4°C and 17000 x *g* unless otherwise indicated. Pellets were thawed and processed using the RNAPro kit (MP biomedical), 1 ml of RNAPro solution was used to re-suspend the pellets and was transferred to Lysis Matrix B tubes (MP Biomedicals) before undergoing ribolysis for 40 s at 6.0 m/s using a FastPrep-24 Classic instrument (MP Biomedicals). The sample was transferred to 300 μ l chloroform and centrifuged for 15 minutes and the upper phase transferred to

100% ethanol and underwent DNA precipitation overnight at -20°C. Samples were then centrifuged for 5 minutes, washed with 70% ethanol (made with nuclease free dH₂O) and recentrifuged. Finally, the pellets were resuspended in 50 µl of DEPC-treated water which was stored at -80°C until further processing. 25 µl of extracted material was added to 20 U Turbo DNase I (ThermoFisher), 80 U RNasin plus RNase inhibitor (Promega), 50 mM magnesium sulphate, 90 mM sodium acetate, and DEPC-treated water up to 150 µl. Samples were incubated for one hour at 37°C in a PCR thermocycler before addition of a further 20 U Turbo DNase and 80 U RNasin plus followed by incubation for another hour at 37°C.

DNase treated samples underwent RNA purification using the RNeasy kit (Qiagen). Samples were added to 350 µl RLT buffer and 200 µl 100% Ethanol and then applied to a RNeasy column and centrifuged for 15 seconds at 12,800 x *g*. Two wash steps using 500 µl RPE wash buffer were carried out followed by drying of the column by centrifugation for two minutes. The RNA was eluted in 17 µl DEPC-treated water, the column was allowed to stand for three minutes before a one-minute centrifugation, the elution step was repeated twice to give an approximate volume of 50 µl eluted RNA. Eluted RNA was tested for quality (260 : 280 nm ratio) and concentration using a Denovix DS-11 FX instrument (Thermo scientific). Synthesis of cDNA was carried out on 1 µg of RNA using Superscript IV (Thermo scientific) as per manufacturer's instructions. qRT-PCR was carried out using the Kapa Sybr Fast kit (Roche) as per manufacturer's instructions on an ABI-7500 Fast system (Applied Biosystems) using 2 µl of cDNA diluted 1:10 for *hpdC* and 1:100 for *16S rRNA* (internal control). Fold change in gene expression was calculated by the $\Delta\Delta C_t$ method (30) using triplicate technical and five biological replicates. Regression analysis was carried out using Stata16, to determine if fold changes were significant for each strain in the presence of *p*-HPA compared to the BHIS control and if fold changes in the presence of *p*-HPA were significantly different.

High Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD)

The five strains representing each clade of *C. difficile* (630 Δ erm, R20291, CD305, M68 and M120) as well as 630 Δ erm and the Δ codY strains were grown overnight in MM which was made using a combination of recipes from Karasawa *et al.* (31) and Cartman *et al.* (32). The salt, trace salt, vitamin and iron sulphate solutions were made as per Cartman *et al.* (32) whilst the glucose and amino acid solutions were prepared according to Karasawa *et al.* (31) with the exception of the *p*-tyrosine concentration which was increased to 400 mg/l. Each strain was grown overnight in MM and then back diluted the following day to an OD_{590nm} of 0.5, 1 ml of which was used to inoculate 10 ml of MM and MM with 2 mg/ml *p*-HPA. Strains were grown for 8 hours with 1 ml samples taken after 4 and 8 hours and at each time point the OD_{590nm} was also taken to allow for *p*-HPA and *p*-cresol production to be normalised for growth. Samples for HPLC-DAD were filter sterilised using 0.22 μ m filters before freezing at -80°C prior to HPLC analysis.

Filter sterilised samples were transferred to HPLC vials and analysed immediately by injecting onto the HPLC column. Separations were achieved utilising an Acclaim 120 (ThermoFisher), C₁₈, 5 μ m (4.6 x 150 mm), with the mobile phase consisting of ammonium formate (10 mM, pH 2.7) and menthol (v/v; 40:60) at a flow rate of 2 ml/min. *p*-HPA and *p*-cresol were detected by the diode array detector (PDA; DAD 3000) set at 280 nm. Peak identity was confirmed by measuring the retention time, of commercially available *p*-HPA and *p*-cresol and determination of absorbance spectra using the DAD. A calibration curve of each compound was generated by Chromeleon (Dionex software) using known amounts of the reference standards (0–100 mg/ml) dissolved in media, injected onto the column and amount of *p*-HPA and *p*-cresol in the samples were determined. Samples from three independent biological replicates were analysed compared to media controls and standard curves. The data was analysed in GraphPad Prism 8 and statistical analysis was performed in Stata16 using linear regression analysis $p < 0.05$ was considered statistically significant.

phoZ assays

To determine if reduced *hpdBCA* promoter activity was the mechanism for reduced *p*-HPA turnover in 630 Δ *erm* Δ *codY* a *phoZ* transcriptional reporter fused to the *hpdBCA* promoter sequence was used (P_{hpdB} -*phoZ*) (6). The reporter plasmid was conjugated in to 630 Δ *erm* and 630 Δ *erm* Δ *codY*. Assays were carried out exactly as per Harrison *et al.* (6). Briefly, each strain was grown overnight in BHIS before back dilution to OD_{590nm} of 0.5 and growth in 3:1 MM:BHIS for three hours. Cultures were washed twice in fresh MM before a final resuspension in 1 ml of MM, 100 μ l of which was added to 10 ml MM +/- 2 mg/ml *p*-HPA and grown for four hours before 2 ml samples were taken, pelleted and frozen before processing for *phoZ* activity as per (6).

Analysis of hpdB expression and localisation

To investigate expression localisation of HpdB, a translational fusion was constructed with the promoter region of *hpdBCA* and the HpdB coding sequence joined via a linker to a SNAP-tag carried in a pMTL84151 vector to give plasmid P_{hpdB} -CDS-SNAP. Construction of the plasmid was carried out in a two-stage process. Firstly, Gibson cloning was used to insert the *hpdBCA* promoter and coding sequence into plasmid pMTL84151 (Table 1) followed by insertion of the linker and SNAP-tag into the pMTL84151 carrying the *hpdBCA* promoter and coding sequence (Table 1). All PCRs were carried out using Phusion (NEB) as per manufacturer's instructions, PCR products were run on a 1% agarose gel and the bands were cut out then purified using the Monarch gel extraction kit (NEB), and ligation of PCR products was carried out using Hifi master mix (NEB) as per manufacturer's instructions. Primers used are listed in table 2. Construction of the plasmid was confirmed by Sanger sequencing before electroporation into the conjugation strain of *E. coli* CA434 (33). Conjugation was carried out using heat shock for 5 minutes at 52°C as per Kirk *et al.* (34) with transconjugants selected by growth on BHIS with *C. difficile* supplement (Oxoid) and thiamphenicol (15 μ g/ml) (CCTm). Transconjugants were re-streaked once more on to BHIS CCTm plates to ensure plasmid retention.

Confocal microscopy

To prepare slides for microscopy, overnight cultures of 630 Δ erm P_{hpdB-CDS}-SNAP were back diluted to an OD_{590nm} of 0.5 and 1 ml added to BHIS +/- 2 mg/ml *p*-HPA. These cultures were grown for four hours before 500 μ l was removed and added to 2.5 μ l 50 nM TMR-Star and incubated anaerobically in the dark for 30 minutes. Following incubation, the samples were pelleted at 4000 x *g* for 2 minutes and washed twice in 500 μ l PBS, after the second wash approximately 20 μ l PBS was left in the tube and used to re-suspend the pellet. A 1 μ l loop was used to spread the culture on to a glass slide, which was allowed to air dry for 1 minute. 10 μ l of Vectashield with DAPI was added to the dried cells with a coverslip placed over the top and sealed with clear nail polish. The slides were imaged under oil immersion using a Zeiss LSM-800 microscope (100X objective). Excitation and emissions used for the dyes was 358 nm and 463 nm for DAPI and 578 nm and 603 nm for TMR-Star, respectively. A minimum of three fields of view per slide were imaged. Images were processed in Zeiss Zen Blue software.

Western Blot and Mass Spectrometry

Confirmation of the fusion of the SNAP-tag to HpdB was carried out by anti-SNAP-tag western blotting and mass spectrometry (MS). Samples were prepared by overnight culture of 630 Δ erm P_{hpdB-CDS}-SNAP being back diluted to an OD_{590nm} of 0.5 and 1 ml added to BHIS +/- 2 mg/ml *p*-HPA and grown for four hours. Following growth 10 ml of the cultures were pelleted and frozen at -80°C. Pellets were resuspended in 1 ml of binding buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 25 mM imidazole) and transferred to a lysis matrix B tube and then ribolyzed at 6.0 m/s for 40 s. Samples were centrifuged and the supernatant saved. In duplicate for each culture sample 15 μ l of the supernatant was added to 5 μ l of 4X loading dye (ThermoFisher) and the sample heated at 95°C for 5 minutes before loading on to two separate 10% Bis-Tris gels which were run at 180V 400A. One gel then underwent transfer to a nitrocellulose membrane using the IBlot system as per manufacturer's instructions, the other gel was saved for excision of the appropriate band for mass-spectrometry. The membrane was washed in 1X phosphate buffered saline (PBS) Tween20 (0.1%) for 5 minutes before undergoing blocking with

blocking buffer (5% whole milk in PBS Tween 20 (0.1%)). 10 ml blocking buffer containing a 1:1000 dilution of Anti-SNAP antibody (New England Biolabs) was added to the membrane and incubated at room temperature for one hour. The membrane was washed three times in PBS Tween20 (0.1%) for 5 minutes. 10 ml PBS Tween20 (0.1%) containing a 1:10,000 dilution of IRDye CW800 goat anti-rabbit antibody (Li-cor) was added to the membrane and incubated for one hour before three wash 5 minute wash steps in PBS Tween20 (0.1%). The membrane was visualised and imaged on an Odyssey Li-cor instrument at the 800 wavelength. The HpdB-SNAP-tag band was identified at 121 kDa, which was excised from the second Bis-Tris gel and sent for mass spectrometry analysis at the Centre of Excellence for Mass Spectrometry at Kings College London. MS was done as previously (35), with the following modifications: 1) 75 μ M C18 column (50 cm length) was used rather than 75 μ M C18 column (15 cm length), 2) Xcalibur software used was v4.4.16.14, 3) Proteome Discoverer software used was v2.5, and 4) Scaffold 5 software used was v5.0.1. Analysis of MS results was carried out in Scaffold software v5.0.1 and compared to the Uniprot All Taxonomy database. Six peptides were identified as being from *C. difficile* strain 630's HpdB, four with 100% identity and two with 99% identity providing coverage of 5.9% of HpdB's amino acid sequence, peptides identified by MS are listed in Supplementary Table 1.

Results

Conservation of p-HPA turnover and p-cresol production in the five C. difficile lineages

We show a significant induction of the *hpdBCA* operon in the presence of *p*-HPA, which is conserved in representative strains from all five *C. difficile* lineages; Clade 1; 630 Δ *erm* (RT012), Clade 2; R20291 (RT027), Clade 3; CD305 (RT023), Clade 4; M68 (RT017) and Clade 5; M120 (RT078) (Fig 1a). The largest fold change in expression of the *hpdC* gene in response to *p*-HPA was observed in strains R20291: 879.9 ± 331.0 and CD305 792.7 ± 215.3 , compared to 630 Δ *erm* (469.4 ± 120.7), M68 (447.7 ± 153.0) and M120 (312.3 ± 98.9) (Fig. 1a).

We used high performance liquid chromatography with diode array detection (HPLC-DAD) to quantify *p*-HPA and *p*-cresol production from the turnover of *p*-tyrosine (without exogenous *p*-HPA), as well as production of *p*-cresol in response to exogenous *p*-HPA. We found significant differences in *p*-HPA and *p*-cresol generation via *p*-tyrosine fermentation between the 5 clades (Fig. 1). M120, R20291 and CD305 produced the highest levels of *p*-HPA at both 4 and 8 hours (Fig. 1b&c). After 8 hours growth, M120 produced significantly more *p*-HPA than 630 Δ *erm* ($p=0.004$), CD305 ($p=0.018$) and M68 ($p<0.001$) (Fig. 1c). Comparatively, *p*-cresol production tracked with *p*-HPA production (Fig. 1). Strain CD305 (RT023) produced the highest level of *p*-cresol at both 4 and 8 hours (Fig. 1d&e). After 4 hours, strain CD305 produced significantly more *p*-cresol (0.0046 ± 0.0015 mg/ml) than 630 Δ *erm* (0.0035 ± 0.0020 mg/ml, $p<0.001$), R20291 (0.0025 ± 0.0017 mg/ml, $p=0.017$) and M68 (0.0014 ± 0.0001 , $p<0.001$). After 8 hours growth, strains CD305 (0.0046 ± 0.001 mg/ml) and M120 (0.0058 ± 0.0022 mg/ml) produced the highest levels of *p*-cresol with both significantly higher than M68 ($p<0.005$) (Fig. 1e). As anticipated, the addition of exogenous *p*-HPA (2 mg/ml) significantly increased *p*-cresol production by over 30-fold in all five strains (Fig. 1). No significant differences in *p*-cresol concentration were observed under *p*-HPA induction, indicating that all five strains sense, take up and convert *p*-HPA with similar efficiency.

The HpdBCA decarboxylase is produced ubiquitously in C. difficile cells

To investigate cellular localisation of the HpdBCA decarboxylase, a plasmid based HpdB-SNAP-tag translational fusion was constructed using a *C. difficile* compatible plasmid (pMTL84151) under control of the *hpdBCA* promoter. The *hpdB* coding sequence omitting the stop codon, was fused via a linker to a SNAP-tag (Table 1). Confirmation of HpdB linked to the SNAP-tag was carried out via western blot (Supplementary Fig. 1), and mass spectrometry, in which we identified six peptides, four with 100% identity and two with 99% identity unique to HpdB (Supplementary Table 1). Localisation of HpdB was visualised by confocal microscopy in 630 Δ *erm* and the *p*-cresol deficient mutant (*hpdC::CT*) (carrying the HpdB-SNAP-tag fusion (P_{hpdB-CDS}-SNAP). Visualisation of the HpdB-SNAP-tag fusion was undertaken

in the presence and absence of *p*-HPA (supplemented into BHIS growth media) to induce expression of *hpdBCA* operon. The SNAP-tag substrate TMR-Star (excitation/emission:578 nm/603 nm) was added to visualise the HpdB-SNAP-tag and DAPI (excitation/emission: 358 nm/463 nm) was used to visualise DNA within the cells. Confocal microscopy revealed that all cells grown in the presence of *p*-HPA expressed the HpdB-SNAP-tag in response to *p*-HPA, whereas those grown without *p*-HPA showed very little or no visible expression (Fig. 2). Furthermore, the HpdB-SNAP-tag fusion was located throughout the cytoplasm of the cell in the wild type strain (Fig. 2), however, in contrast, HpdB-SNAP-tag in the *hpdC* knockout strain was localised in aggregates within the cells. This provides some evidence that the HpdB-SNAP is most likely forming a HpdBCA decarboxylase, and that all three subunits of the HpdBCA decarboxylase are required for diffusion throughout the cell.

p-HPA adversely affected growth of *C. difficile* and induced sporulation

C. difficile strains 630 Δ *erm* and a 630 Δ *erm hpdC* inactivation mutant (*hpdC::CT*) were grown in 0, 1, 2, 3 and 4 mg/ml *p*-HPA and monitored over an 8 hour time course (Fig. 3). A significant growth defect was observed at ≥ 2 mg/ml in both wild-type and *hpdC::CT* mutant, showing that *p*-HPA is deleterious to *C. difficile* (Fig. 3). Furthermore, in the presence of 2 mg/ml *p*-HPA, growth of the *hpdC* mutant is significantly decreased compared to its wild type counterpart ($p < 0.01$) (Fig. 3), showing that the efficient turnover of *p*-HPA to *p*-cresol enhances growth of *C. difficile* over the 8 hour time course. After 24 hours growth, we observed a significant decrease in total viable counts (vegetative cells and spores) in both the wild-type 630 Δ *erm* (Fig 4a) and the *p*-cresol deficient mutant (Fig 4b), in response to increasing concentrations of *p*-HPA. The decrease in total cell counts was significantly more pronounced in the wild type, than the *hpdC* mutant. Analysis of growth in the presence of *p*-HPA as a percentage of the BHIS only control showed that at 1, 2 and 3 mg/ml *p*-HPA, growth was reduced by 90.35% ($\pm 1.52\%$), 99.06% ($\pm 0.41\%$), and 99.71% ($\pm 0.22\%$), respectively, in the wild type, respectively, compared to 24.61% ($\pm 6.53\%$), 48.43% ($\pm 5.17\%$) and 75.97% ($\pm 7.71\%$) in the *hpdC::CT* strain. Alongside this drop in total viable counts, we observed a significant increase in sporulation frequency, with a

positive correlation between *p*-HPA concentration and sporulation rate (Fig. 4c), with a stronger correlation in the wild type ($R^2=0.9193$, $p=0.000012$) than *hpdC::CT* strain ($R^2=0.8868$, $p=0.00006$), which suggests that both *p*-HPA and *p*-cresol derived from *p*-HPA turnover induce sporulation in *C. difficile*

p-HPA adversely affects growth of representative Gram-negative gut bacteria

p-HPA also inhibits growth of commensal gut bacteria, with Gram-negative species more sensitive to *p*-HPA than Gram-positive species (Fig. 5). The growth of the Gammaproteobacteria *E. coli* and *K. oxytoca* was significantly inhibited by *p*-HPA (≥ 1 mg/ml), whilst *P. mirabilis* was significantly inhibited at ≥ 2 mg/ml (Fig. 5). In contrast the Gram-positive species *B. adoscelentis* and *L. fermentum* were significantly more tolerant to *p*-HPA than the Gram-negative bacteria. *L. fermentum* was only significantly inhibited at 4 mg/ml and *B. adoscelentis* only exhibited a growth defect at *p*-HPA concentrations of ≥ 3 mg/ml (Fig. 5). Surprisingly, *E. faecium* was significantly more sensitive to *p*-HPA than the other Gram-positive bacteria, exhibiting a significant growth defect at 1 mg/ml *p*-HPA (Fig. 5). Unlike the other Gram-positive bacteria, *C. difficile* was unable to grow in *p*-HPA ≥ 4 mg/ml (Fig. 3) showing that exogenous *p*-HPA is more deleterious to *C. difficile* growth than other Gram-positive bacteria.

Exploring the toxicity of *p*-HPA

The addition of *p*-HPA to growth media reduces the pH in a concentration dependent manner, we identified that at *p*-HPA concentrations of 1, 2, 3 and 4 mg/ml, the resultant media pH was 6.6, 6.2, 5.8 and 5.4, respectively. Therefore, we assessed the effect of lower pH on growth of *C. difficile* and the representative gut bacteria. A minimal growth defect in *C. difficile* was observed at the lowest pH; pH 5.4 (630 Δ *erm* $p=0.035$, *hpdC::CT* $p=0.020$ compared to control), which is equivalent to the pH of media supplemented with 4 mg/ml *p*-HPA, however, no significant effects were seen at any other pH tested including 6.2 and 5.8, which are equivalent to the pH of media containing 2 and 3 mg/ml *p*-HPA

respectively (Supplementary Fig. 2). Additionally, at the end of the growth analyses the media was filter sterilised and the pH measured. This showed that whilst *C. difficile* was able to buffer small pH changes (maximum of 0.52 ± 0.08), the changes were insufficient to restore the pH of the media to pH 7, which corresponds to the pH of media not supplemented with *p*-HPA (Supplementary figure. 3).

The effect of reduced pH was assessed on the representative gut commensal species (Supplementary Fig. 4). Most strains, with the exception of *L. fermentum* had a slight growth defect in acidic media, however, this effect was not as dramatic as the effect of growth in the presence of *p*-HPA (Supplementary Fig. 4). These analyses show that whilst pH likely has a role in inhibiting growth it is not the only mechanism of inhibition. After 8 hours, 2 mg/ml *p*-HPA had a significant negative impact on growth of *C. difficile* ($p=0.0064$), *K. oxytoca* ($p=0.0073$), *P. mirabilis* ($p=0.0073$) and *E. faecium* ($p=0.0086$), compared to growth in media without *p*-HPA but at a matched pH (pH 6.2). The growth of *E. coli* at 3 mg/ml *p*-HPA was significantly lower compared to growth at media without *p*-HPA, pH matched to pH 5.8 ($p<0.001$) (Supplementary figure 1). However, the growth of both *B. adoscelentis* and *L. fermentum* was similarly affected in both media containing *p*-HPA and media with a matched pH (compared to the *p*-HPA), thus indicating the observed effects on growth were potentially pH driven.

p-HPA affects cell membrane permeability of *C. difficile*

To further explore the mechanisms of *p*-HPA toxicity we carried out phosphate release assays, as previously described (8), with measurements every 30 minutes up to 90 minutes exposure to *p*-HPA. We compared phosphate release in *C. difficile* strain 630 Δ *erm* and *E. coli* at 0, 1, 2 and 3 mg/ml *p*-HPA to the maximum phosphate release. We found significant increases in phosphate release at 2 mg/ml ($p=0.0013$) for *C. difficile* (Fig. 6a), and at 1 mg/ml for *E. coli* ($p=0.010$) (Fig. 6b). To investigate whether this phosphate release was due to the acidification of the media in the presence of *p*-HPA, phosphate release assays were undertaken in control media (0 mg/ml *p*-HPA) with the pH matched to that observed in the presence of *p*-HPA, i.e., pH's 6.6, 4.1 and 3.8 were equivalent to that for 1, 2 and 3

mg/ml *p*-HPA respectively (Fig 6 c&d). In *C. difficile*, we observed a significant increase in phosphate release at pH 3.8 ($p=0.0027$), but not at pH 6.6 ($p=0.545$) or pH 4.1 ($p=0.3039$) when compared to the control. Furthermore, the highest percentage of maximum phosphate release for *C. difficile* at low pH was only 86.5 ($\pm 4.0\%$) after 90 minutes at pH 3.8, in contrast to > 95% after just 30 minutes in the presence of 2 mg/ml *p*-HPA, rising to >99% at 90 minutes (Fig 6 a&c). Similarly, in *E. coli*, significant phosphate release was seen at ≥ 1 mg/ml *p*-HPA, whereas when the pH equivalent was tested (pH 6.6) phosphate release was actually significantly reduced ($p=0.0184$), and furthermore, no significant differences were observed in release at either pH 4.1 ($p=0.5757$) or pH 3.8 ($p=0.2155$) when compared to the control.

Mutation of CodY leads to reduced p-HPA turnover to p-cresol

CodY is a global transcriptional regulator that responds to the presence of branched chain amino acids (36) and GTP (37) to suppress or activate transcription. We sought to determine whether mutation of *codY* altered production of *p*-cresol both in the presence and absence of exogenous *p*-HPA. A CodY mutant ($\Delta codY$) and its wild-type parent strain, 630 Δerm , were grown in minimal media (MM) with and without 2 mg/ml *p*-HPA. At 4 and 8 hours *p*-HPA and *p*-cresol concentrations were measured by HPLC. We identified a significant deficiency in the turnover of exogenous *p*-HPA to *p*-cresol in the CodY mutant, which was more pronounced at the later growth stage (8 hours) (Fig 7), with 27.42% ($\pm 2.17\%$) turnover of *p*-HPA in the *codY* mutant compared to 37.75% ($\pm 1.18\%$) in the wild-type ($p=0.004$). No significant differences in *p*-cresol production were observed in MM without *p*-HPA (Supplementary Fig. 5) Using a transcriptional fusion of the native *hpdBCA* promoter fused to a *phoZ* reporter ($P_{hpdB-phoZ}$), we show that reduced turnover of *p*-HPA to *p*-cresol in the *codY* mutant was a direct result of a 32.9% \pm 13.7% reduction in expression from the *hpdBCA* operon promoter (Fig. 7c). It is likely that the effect of CodY on expression of *hpdBCA* is focused on induction of the operon rather than basal expression, and elevated *p*-HPA is sensed and transduced indirectly by CodY due to the absence of a potential CodY binding motif in the *hpdBCA* promoter region.

Discussion

The importance of *p*-cresol production for *C. difficile* colonisation and pathogenesis combined with the recent finding that exogenous *p*-HPA induces expression of the *hpdBCA* operon, highlights the potential regulatory role for *p*-HPA in the virulence of *C. difficile* through modulation of *p*-cresol production. In this study we sought to identify whether this phenotype is conserved throughout all the *C. difficile* lineages, and the effect exogenous *p*-HPA has on the viability of *C. difficile* and other representative gut bacteria present in the healthy microbiome.

The production of *p*-cresol is an important virulence attribute for *C. difficile*, providing a competitive advantage over other commensal gut bacteria to promote dysbiosis (8), however, the importance of its precursor *p*-HPA is yet to be determined. *p*-HPA, is produced by a range of bacteria found in the gut, such as *Clostridium* and *Klebsiella* species, as well as by mammalian cells (10), which raises the possibility of a pool of accessible *p*-HPA in the gut for conversion by *C. difficile* to *p*-cresol. Here we present new evidence that *p*-HPA is inhibitory to *C. difficile* growth and furthermore that a *C. difficile* mutant unable to produce the HpdBCA decarboxylase that converts *p*-HPA to *p*-cresol is significantly more susceptible to growth inhibition by *p*-HPA. This shows that the benefits of producing *p*-cresol are two-fold: i) *p*-cresol production enables *C. difficile* to outcompete other gut bacteria and ii) *p*-cresol production facilitates detoxification of *p*-HPA, which has a deleterious effect on *C. difficile* growth.

There appears to be a selective advantage to drive *p*-cresol production by *C. difficile*, however, we identified there is also a link between elevated *p*-HPA and increased sporulation rates in both the wild type and HpdBCA deficient mutant strain. In addition, a greater reduction in total cell count in the wild-type strain with increasing *p*-HPA concentrations was observed, which suggests that there is a fine balance between *p*-HPA turnover and *p*-cresol production. We have shown in a previous study that *p*-cresol is deleterious to *C. difficile* growth when produced at levels ≥ 9.5 mM, and here we show that *p*-HPA is inhibitory to growth at high concentrations (2 mg/ml, 13.1 mM), therefore tight regulation of *p*-cresol production is advantageous to *C. difficile*. Determination of *p*-HPA availability in

the gut over the course of CDI has not been assessed and would be difficult to achieve due to the invasive nature of sample collection, however, there is evidence that *p*-HPA is present in the human colon and 19 μ M *p*-HPA was detected in human faecal samples (38). In addition, we previously showed that a *p*-cresol null mutant was at a significant disadvantage compared to its wild type in a mouse relapse model of CDI (8). This is strong evidence that sufficient *p*-tyrosine, *p*-HPA or both are available to *C. difficile* to produce *p*-cresol to maintain dysbiosis, evidenced by our observation of a concurrent fall in Gammaproteobacteria in mice exposed to the wild type compared to the *p*-cresol null mutant (8). We observed previously (6), as well as in this study, that *p*-tyrosine fermentation to produce *p*-HPA and *p*-cresol is extremely inefficient *in vitro*, therefore this suggests that utilisation of exogenous *p*-HPA is an important source of *p*-cresol production.

C. difficile is a genetically diverse species, which consists of five clades (Clades 1-5) (11, 12), and whilst three further cryptic clades have been identified these are likely to be a separate species (29, 39). Variation exists both within and between clades in major virulence factors, such as toxin production (40, 41), motility (42) and sporulation (41). Significant examples of this include strains of clade 4 that do not have functional toxin A (43, 44) and clade 5 isolates which are non-motile, such as M120 (42). Yet, despite this diversity in virulence attributes between clades, the *hpdBCA* operon is highly conserved between the clades 1-5, and interestingly, the more genetically divergent cryptic clades also carry *hpdBCA*-like operons. Alignments of the DNA sequences of these operons to *hpdBCA* from strain 630, showed identities of 91.1%, 95.0% and 92.2% for cryptic clades C-I, C-II and C-III respectively (Supplementary Fig. 6). Therefore, our finding that strains representative of clades 1-5 all show high level induction of the *hpdBCA* operon, and high levels of *p*-cresol production in the presence of *p*-HPA, suggests this response is very well conserved and therefore of importance to *C. difficile*, and by extension, possibly CDI pathogenesis. Furthermore, given that conversion of exogenous *p*-HPA to *p*-cresol was consistent among all representatives of clades 1-5 these findings imply that these lineages have similar capacities to transport *p*-HPA into the cell and convert it to *p*-cresol.

p-cresol can be produced via two pathways, firstly from the metabolism of *p*-tyrosine, and secondly from exogenous *p*-HPA. We identified that strains CD305 (RT023) and M120 (RT078) were the most efficient producers of *p*-HPA from *p*-tyrosine fermentation, whereas strain M68 (RT017) was the least efficient. There was a direct correlation between the ability to produce *p*-HPA from *p*-tyrosine fermentation and the subsequent conversion to *p*-cresol, which indicates strain specific differences in proficiency of *p*-tyrosine utilisation. Interestingly, these higher *p*-cresol producing strains, RT023 and RT078 are amongst the most prevalent ribotypes identified in the UK (17). It is noteworthy that fermentation of *p*-tyrosine to *p*-HPA under these conditions is inefficient, which limits the ability to produce *p*-cresol. The level of *p*-cresol produced from *p*-tyrosine fermentation in the conditions tested is a minimum of 72.6-fold lower after 8 hours compared to *p*-cresol produced from exogenous *p*-HPA. Whilst these differences provide evidence for variation in *p*-tyrosine fermentation *in vitro*, the *in vivo* utilisation of *p*-tyrosine may differ depending on the local availability of nutrients. In support of this, we have previously shown that environmental conditions affect *p*-cresol production *in vitro*, where we observed that production is significantly lower in rich media (BHIS) than less rich media (Yeast Peptone) (4). This further suggests that exogenous *p*-HPA is a major source of *p*-cresol production.

When looking at the impact of *p*-cresol and *p*-HPA on representative strains of gut species we found *p*-HPA had a similar effect to our published data on the effect of *p*-cresol (8), with both compounds being generally more toxic to Gram-negative species than Gram-positive. Both *K. oxytoca* and *E. coli* were significantly inhibited by 1 mg/ml (6.6 mM) *p*-HPA whereas most Gram-positive bacteria were highly resistant to both *p*-HPA and *p*-cresol, with the exception of *E. faecium* that was significantly inhibited at this concentration. Interestingly, *p*-HPA is acidic therefore we sought to determine whether the effect of *p*-HPA on growth of *C. difficile* and the Gram-negative gut bacteria was a result of *p*-HPA's acidification of the environment. Further analysis by testing acidic pH's via growth curves, as well as utilisation of a phosphate release assay provided evidence that the mechanism of *p*-HPA's toxicity is not limited to its acidification of the environment, as in addition, *p*-HPA also disrupts cell envelope integrity, a further similarity to *p*-cresol (8).

The mechanisms controlling *p*-cresol production by *C. difficile* centre around transcriptional regulation of *hpdBCA* operon, in response to *p*-HPA. We identified that *p*-HPA activates regulatory factors to initiate transcription. Here, we provide evidence of the first such regulator to be involved in this process; the global regulator CodY. Deletion of *codY* renders *C. difficile* less able to convert *p*-HPA to *p*-cresol as a result of a reduced expression of the *hpdBCA* operon. This is likely to be an indirect effect as no CodY binding site has been identified directly upstream of the *hpdBCA* operon (45). However, CodY is a global regulator with over 350 binding sites identified through the genome, 37 of which are located near to regulatory genes (45), one of which could be the factor directly responsible for high level induction of the *hpdBCA* operon.

Importantly, we have identified that every cell in a population responded strongly to the presence of exogenous *p*-HPA via transcription and translation of a SNAP-tagged HpdB subunit, however, *p*-tyrosine fermentation by *C. difficile* *in vitro* produced insufficient *p*-HPA to observe induction of HpdB. A mutant deficient in the HpdC subunit of the HpdBCA decarboxylase, provided evidence that all three subunits of HpdBCA are responsible for facilitating trafficking of the enzyme throughout the cell. Given that every cell expressed HpdBCA in response to exogenous *p*-HPA, this clearly demonstrates that *p*-cresol production is not a virulence factor expressed heterogeneously within a population such as pneumolysin in *Streptococcus pneumoniae* (46) or type three secretion systems in *Salmonella typhimurium* (47), instead, this indicates that the response to *p*-HPA is ubiquitous throughout a population as well as conserved within the species. This may suggest that conversion of *p*-HPA to *p*-cresol by *C. difficile* is important to individual cell's survival as well as the wider population.

In conclusion, our results provide new insights into the impact of *p*-HPA on viability of *C. difficile* and other bacterial species in the gut microbiome. We found that in addition to *C. difficile* benefiting from *p*-HPA induction of *p*-cresol contributing to dysbiosis, it also benefits from the efficient removal of *p*-HPA from the immediate environment, as *p*-HPA is deleterious to *C. difficile* growth. Our findings underscore the importance of the response to *p*-HPA by demonstrating that every cell exposed to *p*-

HPA responds with high level induction of the HpdBCA decarboxylase, which is conserved in the *C. difficile* species. Rationally designed inhibitors of HpdBCA could be both highly specific and effective target to reduce problematic *C. difficile*.

References

1. CDC. Antibiotic Resistance Threats in the United States, 2019. Atlanta, GA: U.S. Department of Health and Human Services, CDC; 2019. 2019.
2. Brown KA, Langford B, Schwartz KL, Diong C, Garber G, Daneman N. Antibiotic prescribing choices and their comparative *C. difficile* infection risks: a longitudinal case-cohort study. *Clin Infect Dis*. 2020.
3. Vardakas KZ, Polyzos KA, Patouni K, Rafailidis PI, Samonis G, Falagas ME. Treatment failure and recurrence of *Clostridium difficile* infection following treatment with vancomycin or metronidazole: a systematic review of the evidence. *Int J Antimicrob Agents*. 2012;40(1):1-8.
4. Dawson LF, Donahue EH, Cartman ST, Barton RH, Bundy J, McNerney R, *et al*. The analysis of *para*-cresol production and tolerance in *Clostridium difficile* 027 and 012 strains. *BMC Microbiology*. 2011;11(1):86.
5. Scheline RR. Metabolism of Phenolic Acids by the Rat Intestinal Microflora. *Acta Pharmacologica et Toxicologica*. 1968;26(2):189-205.
6. Harrison MA, Faulds-Pain A, Kaur H, Dupuy B, Henriques AO, Martin-Verstraete I, *et al*. *Clostridioides difficile para*-Cresol Production Is Induced by the Precursor *para*-Hydroxyphenylacetate. *Journal of Bacteriology*. 2020;202(18):e00282-20.
7. Selmer T, Andrei PI. *p*-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glyceryl radical enzyme catalysing the formation of *p*-cresol. *Eur J Biochem*. 2001;268(5):1363-72.
8. Passmore IJ, Letertre MPM, Preston MD, Bianconi I, Harrison MA, Nasher F, *et al*. *Para*-cresol production by *Clostridium difficile* affects microbial diversity and membrane integrity of Gram-negative bacteria. *PLOS Pathogens*. 2018;14(9):e1007191.
9. Saito Y, Sato T, Nomoto K, Tsuji H. Identification of phenol- and *p*-cresol-producing intestinal bacteria by using media supplemented with tyrosine and its metabolites. *FEMS Microbiology Ecology*. 2018;94(9).
10. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, *et al*. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*. 2018;46(D1):D608-d17.
11. Stabler RA, Dawson LF, Valiente E, Cairns MD, Martin MJ, Donahue EH, *et al*. Macro and micro diversity of *Clostridium difficile* isolates from diverse sources and geographical locations. *PloS one*. 2012;7(3):e31559-e.
12. Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, *et al*. Multilocus Sequence Typing of *Clostridium difficile*. *Journal of Clinical Microbiology*. 2010;48(3):770-8.
13. Wüst J, Sullivan NM, Hardegger U, Wilkins TD. Investigation of an outbreak of antibiotic associated colitis by various typing methods. *Journal of clinical microbiology*. 1982;16(6):1096-101.
14. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and Evolution in the Genome of *Clostridium difficile*. *Clinical Microbiology Reviews*. 2015;28(3):721-41.
15. Loo VG, Poirier L, Miller MA, Oughton M, Libman MD, Michaud S, *et al*. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *N Engl J Med*. 2005;353(23):2442-9.
16. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, *et al*. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet (London, England)*. 2005;366(9491):1079-84.
17. PHE. *Clostridium difficile* Ribotyping Network (CDRN) for England and Northern Ireland 2015-2018. 2019.

18. Marujo V, Arvand M. The largely unnoticed spread of *Clostridioides difficile* PCR ribotype 027 in Germany after 2010. *Infection Prevention in Practice*. 2020;2(4):100102.
19. Tamez-Torres KM, Torres-González P, Leal-Vega F, García-Alderete A, López García NI, Mendoza-Aguilar R, *et al.* Impact of *Clostridium difficile* infection caused by the NAP1/RT027 strain on severity and recurrence during an outbreak and transition to endemicity in a Mexican tertiary care center. *International Journal of Infectious Diseases*. 2017;65:44-9.
20. Freeman J, Vernon J, Pilling S, Morris K, Nicholson S, Shearman S, *et al.* The ClosER study: results from a three-year pan-European longitudinal surveillance of antibiotic resistance among prevalent *Clostridium difficile* ribotypes, 2011-2014. *Clin Microbiol Infect*. 2018;24(7):724-31.
21. Shaw HA, Preston MD, Vendrik KEW, Cairns MD, Browne HP, Stabler RA, *et al.* The recent emergence of a highly related virulent *Clostridium difficile* clade with unique characteristics. *Clinical Microbiology and Infection*. 2020;26(4):492-8.
22. Kim J, Kim Y, Pai H. Clinical Characteristics and Treatment Outcomes of *Clostridium difficile* Infections by PCR Ribotype 017 and 018 Strains. *PLoS One*. 2016;11(12):e0168849.
23. Kim H, Jeong SH, Roh KH, Hong SG, Kim JW, Shin MG, *et al.* Investigation of toxin gene diversity, molecular epidemiology, and antimicrobial resistance of *Clostridium difficile* isolated from 12 hospitals in South Korea. *Korean J Lab Med*. 2010;30(5):491-7.
24. Jin D, Luo Y, Huang C, Cai J, Ye J, Zheng Y, *et al.* Molecular Epidemiology of *Clostridium difficile* Infection in Hospitalized Patients in Eastern China. *J Clin Microbiol*. 2017;55(3):801-10.
25. Van den Berg RJ, Claas EC, Oyib DH, Dijkshoorn L, Brazier JS, *et al.* Characterization of toxin A-negative, toxin B-positive *Clostridium difficile* isolates from outbreaks in different countries by amplified fragment length polymorphism and PCR ribotyping. *J Clin Microbiol*. 2004;42(3):1035-1041
26. Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, *et al.* Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. *Clin Infect Dis*. 2008;47(9):1162-70.
27. Bauer MP, Notermans DW, van Benthem BHB, Brazier JS, Wilcox MH, Rupnik M, *et al.* *Clostridium difficile* infection in Europe: a hospital-based survey. *The Lancet*. 2011;377(9759):63-73.
28. Ramírez-Vargas G, Rodríguez C. Putative Conjugative Plasmids with *tcdB* and *cdtAB* Genes in *Clostridioides difficile*. *Emerging Infectious Disease journal*. 2020;26(9):2287.
29. Knight DR, Imwattana K, Kullin B, Guerrero-Araya E, Paredes-Sabja D, Didelot X, *et al.* Major genetic discontinuity and novel toxigenic species in *Clostridioides difficile* taxonomy. *eLife*. 2021;10:e64325.
30. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(-Delta C(T)) Method. *Methods*. 2001;25(4):402-8.
31. Karasawa T, Ikoma S, Yamakawa K, Nakamura S. A defined growth medium for *Clostridium difficile*. *Microbiology (Reading)*. 1995;141 (Pt 2):371-5.
32. Cartman ST, Minton NP. A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. *Appl Environ Microbiol*. 2010;76(4):1103-9.
33. Purdy D, O'Keefe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, *et al.* Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol Microbiol*. 2002;46(2):439-52.
34. Kirk JA, Fagan RP. Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe*. 2016;42:1-5.
35. Dawson LF, Peltier J, Hall CL, Harrison MA, Derakhshan M, Shaw HA, *et al.* Extracellular DNA, cell surface proteins and c-di-GMP promote biofilm formation in *Clostridioides difficile*. *Scientific Reports*. 2021;11(1):3244.

36. Shivers RP, Sonenshein AL. Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. *Mol Microbiol.* 2004;53(2):599-611.
37. Ratnayake-Lecamwasam M, Serror P, Wong KW, Sonenshein AL. *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev.* 2001;15(9):1093-103.
38. Jenner AM, Rafter J, Halliwell B. Human fecal water content of phenolics: the extent of colonic exposure to aromatic compounds. *Free Radic Biol Med.* 2005;38(6):763-72.
39. Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW, Stoesser N, *et al.* Evolutionary history of the *Clostridium difficile* pathogenicity locus. *Genome biology and evolution.* 2014;6(1):36-52.
40. Merrigan M, Venugopal A, Mallozzi M, Roxas B, Viswanathan VK, Johnson S, *et al.* Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as robust toxin production. *J Bacteriol.* 2010;192(19):4904-11.
41. Akerlund T, Persson I, Unemo M, Norén T, Svenungsson B, Wullt M, *et al.* Increased sporulation rate of epidemic *Clostridium difficile* Type 027/NAP1. *J Clin Microbiol.* 2008;46(4):15303.
42. Valiente E, Dawson LF, Cairns MD, Stabler RA, Wren BW. Emergence of new PCR ribotypes from the hypervirulent *Clostridium difficile* 027 lineage. *J Med Microbiol.* 2012;61(Pt 1):49-56.
43. Kuijper EJ, de Weerd J, Kato H, Kato N, van Dam AP, van der Vorm ER, *et al.* Nosocomial outbreak of *Clostridium difficile*-associated diarrhoea due to a clindamycin-resistant enterotoxin A negative strain. *Eur J Clin Microbiol Infect Dis.* 2001;20(8):528-34.
44. Drudy D, Harnedy N, Fanning S, Hannan M, Kyne L. Emergence and control of fluoroquinolone-resistant, toxin A-negative, toxin B-positive *Clostridium difficile*. *Infect Control Hosp Epidemiol.* 2007;28(8):932-40.
45. Dineen SS, McBride SM, Sonenshein AL. Integration of metabolism and virulence by *Clostridium difficile* CodY. *J Bacteriol.* 2010;192(20):5350-62.
46. Surve MV, Bhutda S, Datey A, Anil A, Rawat S, Pushpakaran A, *et al.* Heterogeneity in pneumolysin expression governs the fate of *Streptococcus pneumoniae* during blood-brain barrier trafficking. *PLoS Pathog.* 2018;14(7):e1007168.
47. Sturm A, Heinemann M, Arnoldini M, Benecke A, Ackermann M, Benz M, *et al.* The cost of virulence: retarded growth of *Salmonella Typhimurium* cells expressing type III secretion system 1. *PLoS Pathog.* 2011;7(7):e1002143.
48. Hussain HA, Roberts AP, Mullany P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630 Δ erm) and demonstration that the conjugative transposon Tn916 Δ E enters the genome of this strain at multiple sites. *Journal of Medical Microbiology.* 2005;54(2):137-41.
49. Tremblay YDN, Durand BAR, Hamiot A, Martin-Verstraete I, Oberkampf M, Monot M, *et al.* Metabolic adaption to extracellular pyruvate triggers biofilm formation in *Clostridioides difficile*. *The ISME Journal.* 2021.
50. Heap JT, Pennington OJ, Cartman ST, Minton NP. A modular system for *Clostridium* shuttle plasmids. *J Microbiol Methods.* 2009;78(1):79-85.

Tables and figures

Strain or plasmid	Relevant features	Source or reference
C. difficile strains		
630 Δ erm	Erythromycin sensitive strain of 630 - Clade 1	(48)
630 Δ erm P _{hpdB-CDS} -SNAP	630 Δ erm carrying SNAP-tag reporter fused via a linker to <i>hpdB</i> coding sequence	This study
630 Δ erm P _{hpdB-phoZ}	630 Δ erm knockout carrying <i>hpdB phoZ</i> transcriptional reporter	(6)
630 Δ erm Δ codY	Δ codY mutant of 630 Δ erm	(49)
630 Δ erm Δ codY P _{hpdB-phoZ}	<i>codY</i> knockout carrying <i>hpdB phoZ</i> transcriptional reporter	This study
R20291	Representative strain – Clade 2	(11)
CD305	Representative strain – Clade 3	(11)
M68	Representative strain – Clade 4	(11)
M120	Representative strain – Clade 5	(42)
630 Δ erm P _{hpdB-CDS} -SNAP	630 Δ erm carrying a translational fusion of the promoter and coding sequence of <i>hpdB</i> fused to a SNAP-tag	This study
Gut commensal strains		
<i>Escherichia coli</i>		Dr Simon Baines
<i>Klebsiella oxytoca</i>		Dr Simon Baines
<i>Proteus mirabilis</i>		Dr Simon Baines
<i>Enterococcus faecium</i>		Dr Simon Baines
<i>Lactobacillus fermentum</i>		Dr Simon Baines
<i>Bifidobacterium adoscelentis</i>		Dr Simon Baines
E. coli cloning strains		
<i>E. coli</i> NEB5 α	General cloning	New England Biolabs
<i>E. coli</i> CA434	Conjugation donor	(48)
Plasmids		
pMTL84151		(50)
P _{hpdB-CDS} -SNAP	pMTL84151 carrying translational fusion of the promoter and coding sequence of <i>hpdB</i> joined to a SNAP-tag	This study
P _{hpdB-phoZ}	pMTL84151 plasmid carrying transcriptional fusion of the native <i>hpdBCA</i> promoter to <i>phoZ</i> reporter.	(6)

Table 1. Strains and plasmids used in this study

Primer name	Sequence (5' to 3')	Use
hpdB CDS F	ATGATTACGAAGATCTGAATTCGATAGGGTGTGC	Amplification of hpdB promoter and coding sequence
hpdB CDS R	GAGCTCGAATTTACACCCCTTCATACTCTGTTCTAGC	Amplification of hpdB promoter and coding sequence
84151 hpdB+CDS F	AGGGGTGTAAATTCGAGCTCGGTACCCG	Amplification of pMTL84151 to clone hpdB promoter and coding sequence
84151 hpdB+CDS R	ATTCAGATCTTCGTAATCATGGTCATATGGATACAG	Amplification of pMTL84151 to clone hpdB promoter and coding sequence
hpdB Linker Vec F	TGGGTAAATTCGAGCTCGGTACCCGGGGATCCTCT	Amplification of pMTL84151 hpdB
hpdB Linker Vec R	CACCACCAAGCACCCCTTCATACTCTGTTCTAGCAATTAC	Amplification of pMTL84151 hpdB
Linker-SNAP F	TGAAGGGGTGCTTGGTGGTGGAGGTTTCAG	Amplification of the linker and SNAP-tag sequence from P _{fdx} SNAP
Linker-SNAP R	GAGCTCGAATTTACCCAAGTCTGGTTTC	Amplification of the linker and SNAP-tag sequence from P _{fdx} SNAP
hpdC qRT-PCR F	GGATGCAACCAAAGGAATTTGT	Used for qRT-PCR of <i>hpdC</i>
hpdC qRT-PCR R	ACCCAGTCTTCTTTCTCTAGGC	Used for qRT-PCR of <i>hpdC</i>
16S qRT-PCR F	GGCAGCAGTGGGGAATATTG	Used for qRT-PCR of <i>16S</i>
16S qRT-PCR R	CCGTAGCCTTTCACTCCTGA	Used for qRT-PCR of <i>16S</i>

Table 2. Oligos used in this study

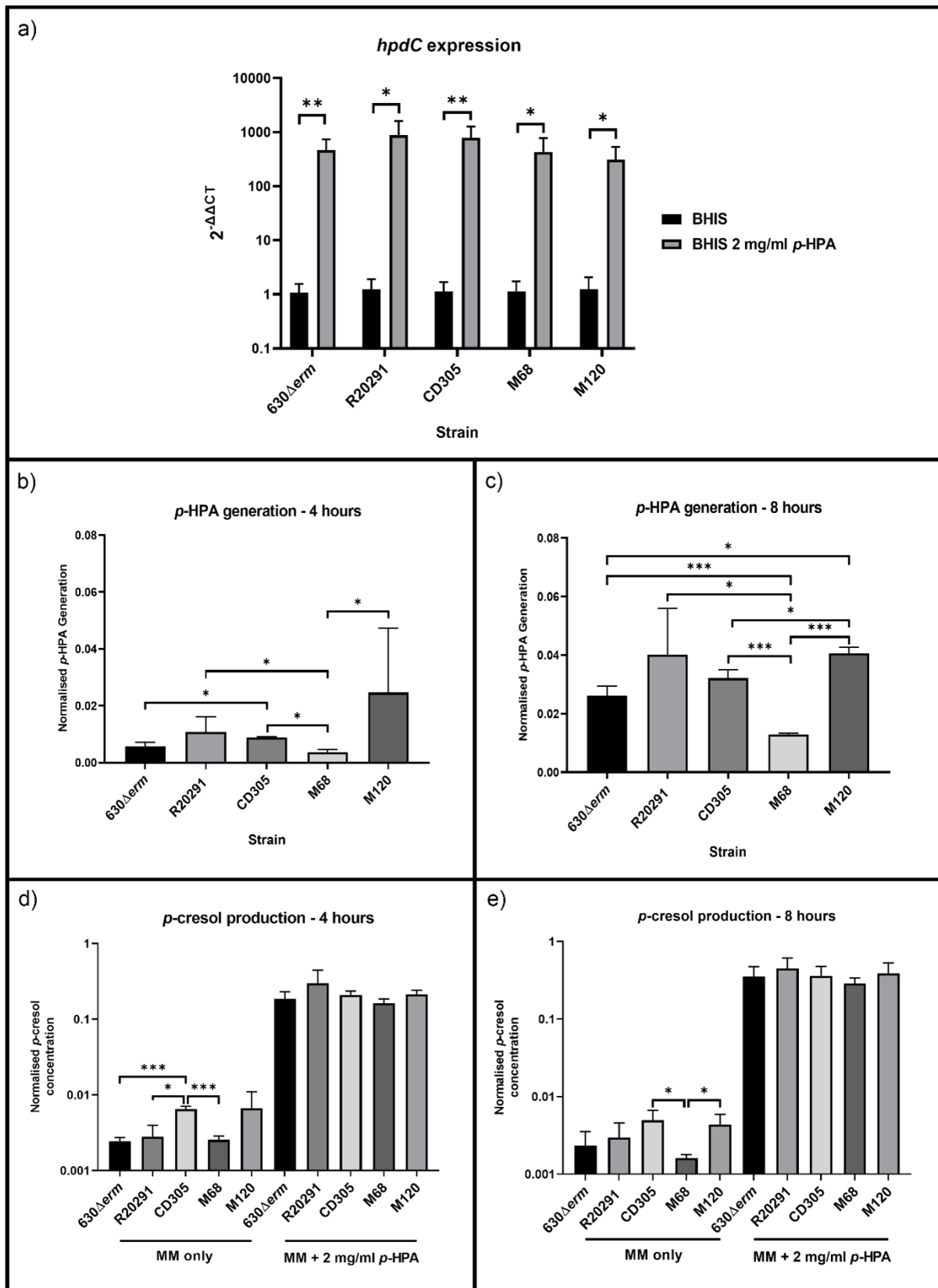


Figure 1. *hpdBCA* expression, *p*-HPA and *p*-cresol production by strains representing each of *C. difficile* clades 1-5. A) qRT-PCR was used to determine whether *p*-HPA induces expression of the *hpdBCA* operon across all five clades of *C. difficile*. Representative strains 630 Δ erm (Clade 1), R20291 (Clade 2), CD305 (Clade 3), M68 (Clade 4) and M120 (Clade 5) were grown to exponential phase (OD_{590nm} 0.6-0.7) in BHIS +/- 2 mg/ml *p*-HPA. Expression was normalised to 16S rRNA control and

analysed by $2^{-\Delta\Delta C_t}$. Data represents mean of five biological replicates in technical triplicate and their standard deviation with statistical analysis was carried out by linear regression. *p*-HPA and *p*-cresol concentrations were determined by HPLC on samples from each strain grown in minimal media (MM) with or without 2 mg/ml exogenous *p*-HPA, with samples taken after 4 hours (B&D) and 8 hours (C&E). B&C) *p*-HPA generation was determined by addition of *p*-HPA and *p*-cresol detected, and D&E) *p*-cresol concentration was determined. Concentrations were normalised to growth (by division of the concentration by the OD_{590nm} at the time the sample was taken). Data represents mean and standard deviation of three independent replicates. Statistical analysis was carried out by linear regression and used to determine significant differences between normalised *p*-HPA generation and *p*-cresol concentrations for each strain. Statistically significant differences are indicated with: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

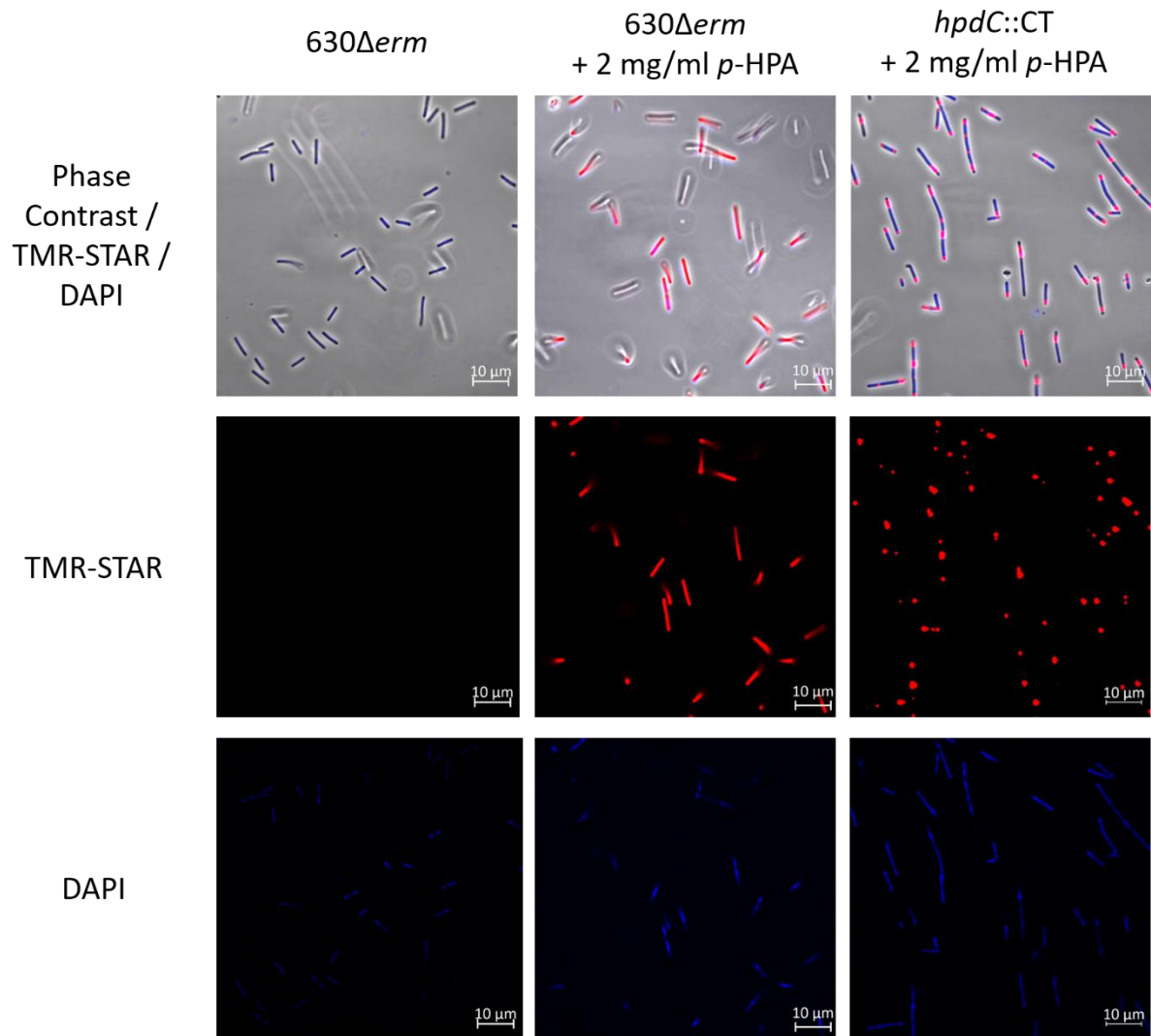


Figure 2. Confocal microscopy to confirm expression and determine localisation of HpdB. A plasmid based translational fusion of HpdB was constructed in *C. difficile* strain 630 Δ erm, using the native *hpdBCA* promoter and *hpdB* coding sequence (without a stop codon) fused via a linker to the SNAP-tag to produce P_{hpdB-CDS}-SNAP. Expression of HpdB linked to the SNAP-tag was confirmed by anti-SNAP western and mass spectrometry. Cultures of *C. difficile* 630 Δ erm and *hpdC*::CT each carrying P_{hpdB-CDS}-SNAP were grown for four hours in BHIS +/- 2 mg/ml *p*-HPA and then harvested in the presence of the SNAP-tag substrate TMR-Star before staining with Vectashield DAPI and were imaged on a Zeiss LSM880 confocal microscope at 578 nm excitation and 603 nm emission (for TMR-Star) and 358 nm excitation and 463 nm emission for DAPI. Images are representative of three independent replicates.

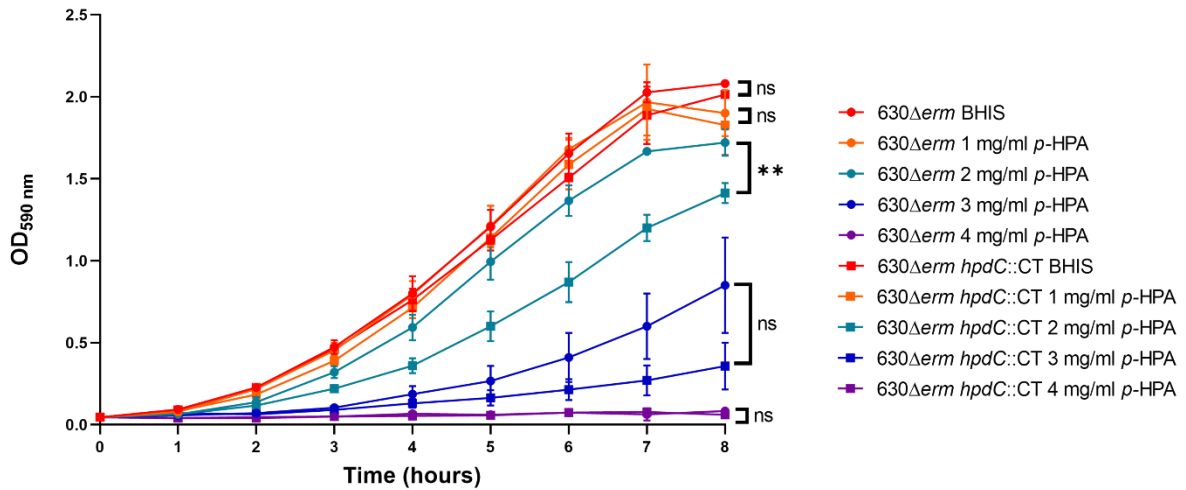


Figure 3. Growth analysis of 630Δerm and 630Δerm hpdC::CT in the presence of p-HPA. Growth curves were undertaken in BHIS media supplemented with 0, 1, 2, 3, and 4 mg/ml p-HPA over an 8 hour time course. *C. difficile* strains 630Δerm (circles) and the p-cresol null mutant hpdC::CT (squares) were compared. Data represents mean and standard deviation of three independent replicates. Statistical analysis was carried out by ANOVA and significant differences are indicated with * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

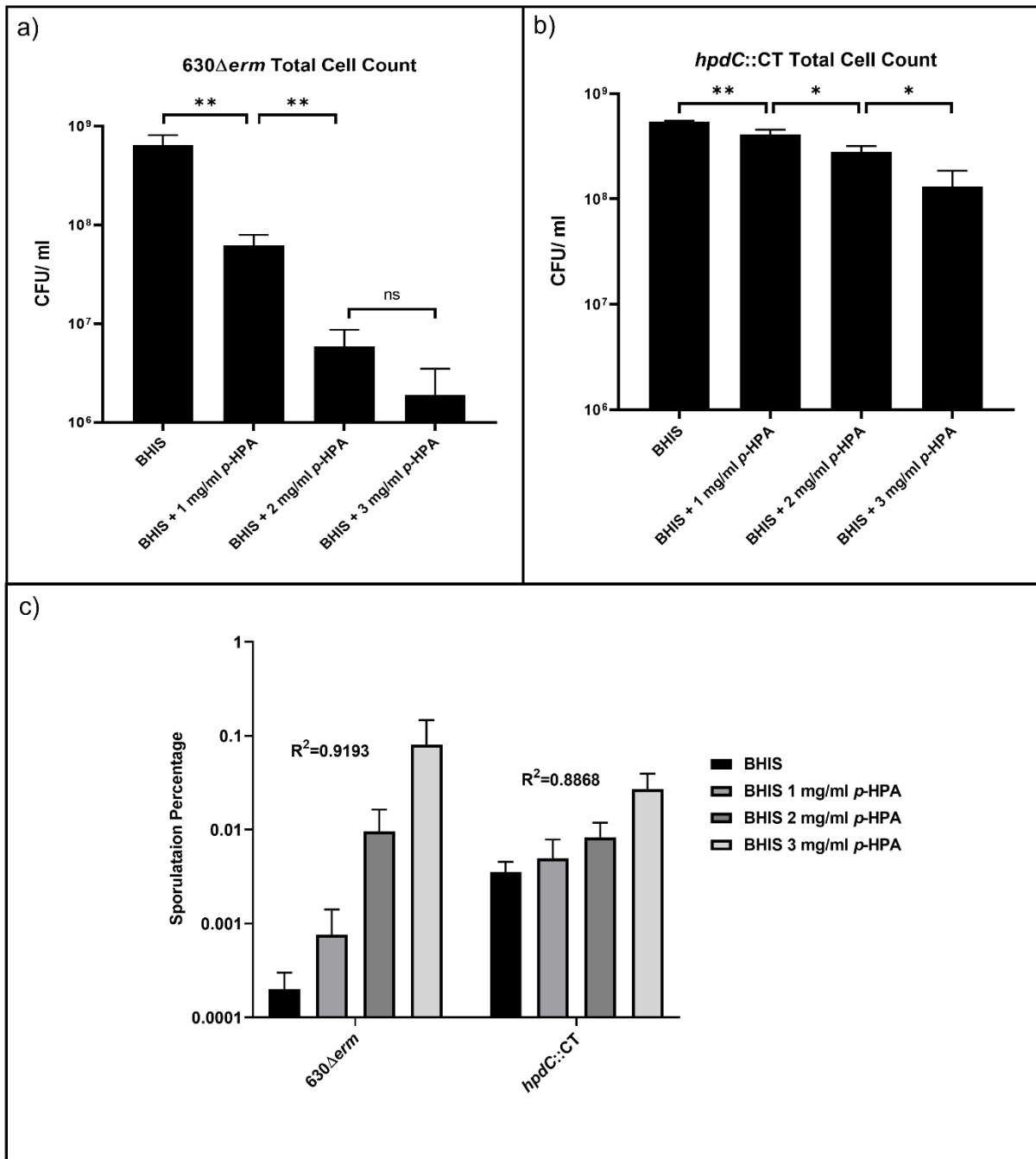


Figure 4. Effect of *p*-HPA and *p*-cresol on sporulation rate and total cell count. 630 Δ erm and hpdC::CT were grown for 24 hours in BHIS in the presence of 0, 1, 2 or 3 mg/ml *p*-HPA. CFUs were performed for total cell counts and for spore count (spores obtained by heating culture at 65°C for 20 minutes) by plating on to BHIS plates containing 0.1% sodium taurocholate. Statistical analysis for a correlation between *p*-HPA concentration and sporulation rate was carried out by a Spearman one tailed rank order correlation. Statistical analysis to determine differences in total cell count was carried out by regression. Data represents a minimum of three independent replicates, error bars represent standard deviation, significant differences are indicated with * p <0.05, ** p <0.01, *** p <0.001

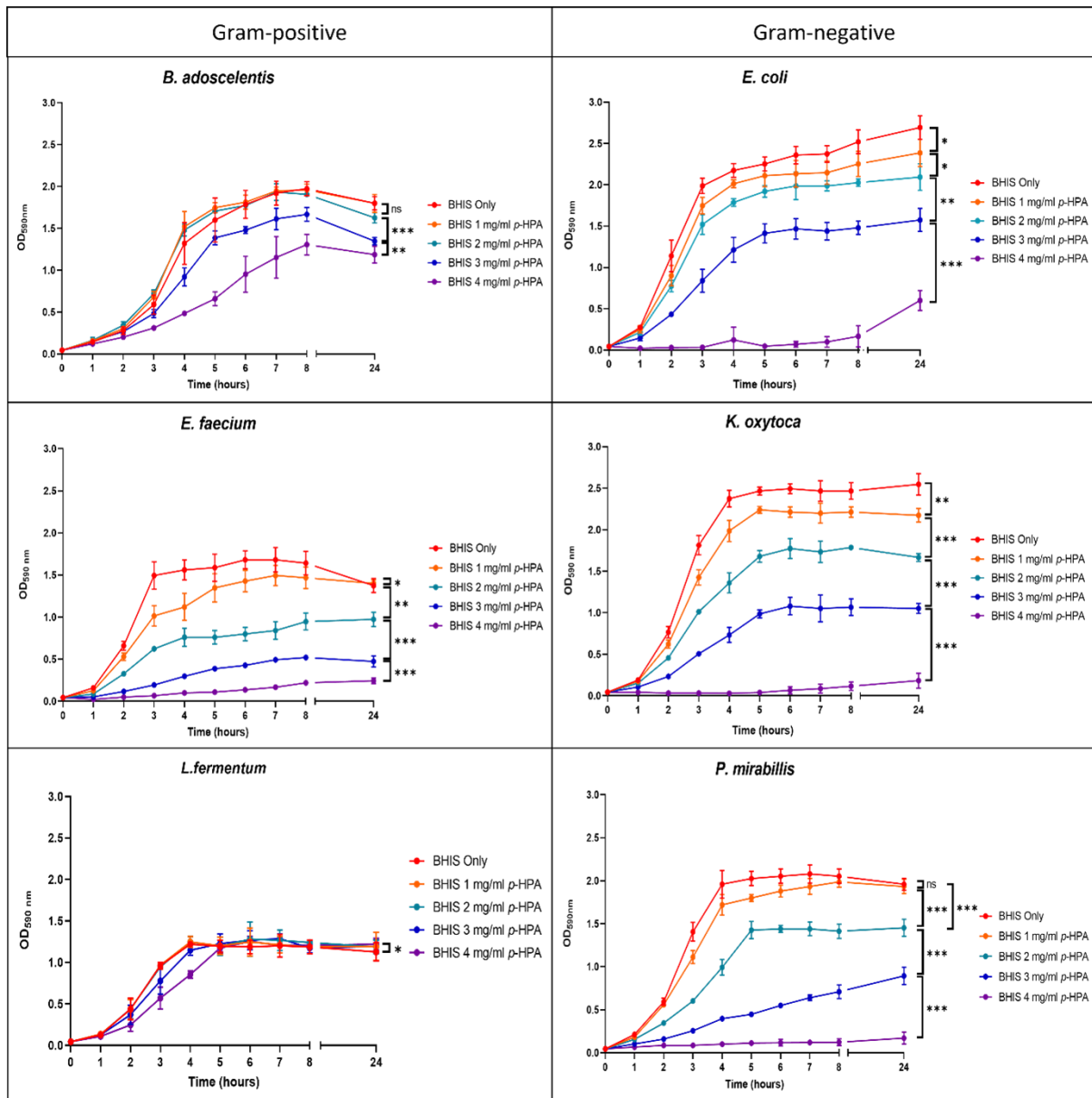


Figure 5. Analysis of growth of *C. difficile* and gut commensal species in the presence of p-HPA. Six representative gut commensal species, three Gram-positive (*Enterococcus faecium*, *Lactobacillus fermentum* and *Bifidobacterium adoscelentis*) and three Gram-negative (*Escherichia coli*, *Klebsiella oxytoca* and *Proteus mirabillis*) strains, were grown in BHIS media with 0, 1, 2, 3 and 4 mg/ml p-HPA. Growth curves represent three biological replicates. ANOVA analysis was used to determine significant differences between growth curves. Error bars represent standard deviation and significant differences are indicated with * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

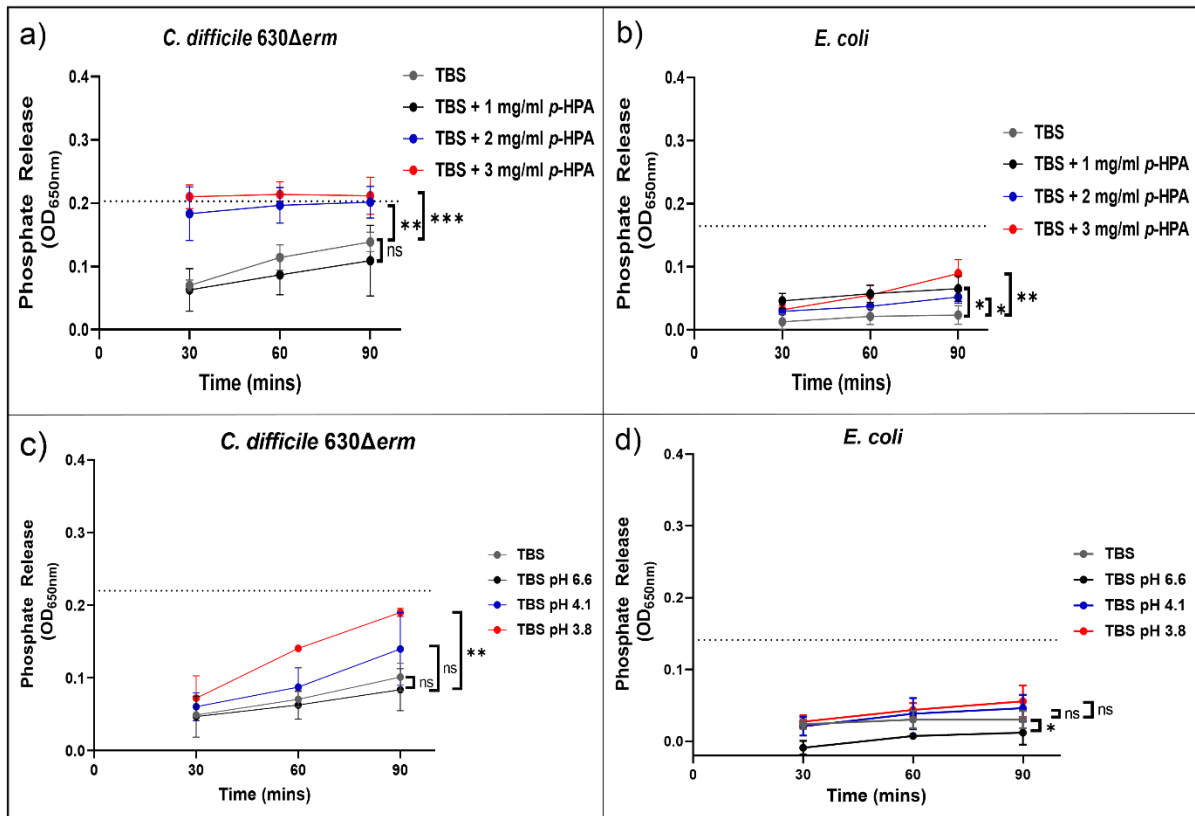


Figure 6. Assessing the effect of p-HPA compared to pH on membrane integrity of *C. difficile* compared to *E. coli*. a & b) Phosphate release assays were carried out in TBS with 0, 1, 2, and 3 mg/ml p-HPA in *C. difficile* and *E. coli* respectively. c & d) TBS was acidified to pH 6.6, 4.1 and 3.8 respectively, with the addition of hydrochloric acid, to match the pH of the phosphate release assays undertaken in the presence of p-HPA. Statistical analysis was undertaken using ANOVA to compare phosphate release in p-HPA or acidified TBS to the TBS control. Data represents the mean and standard deviation of three independent replicates. Significant differences are indicated with * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

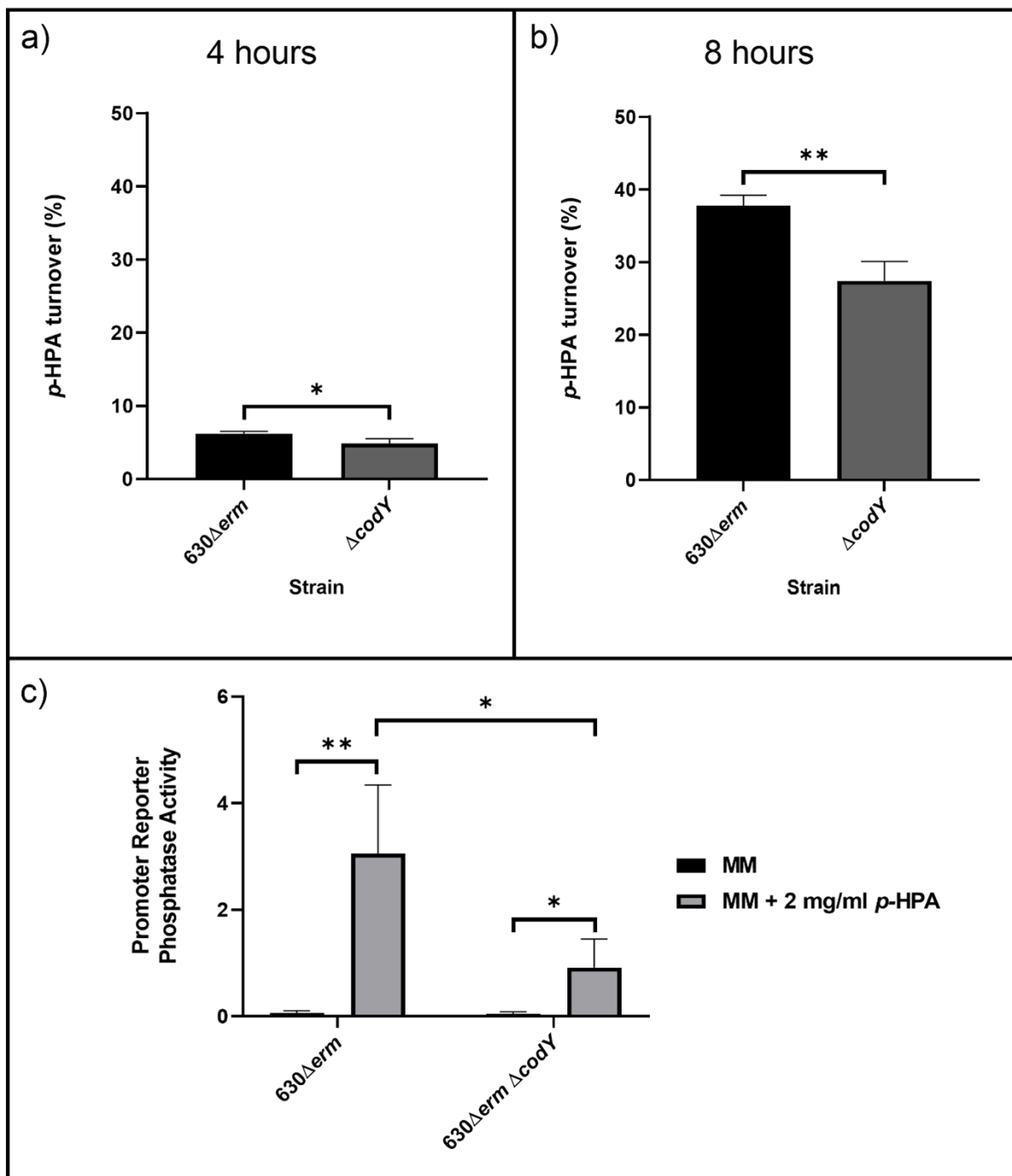
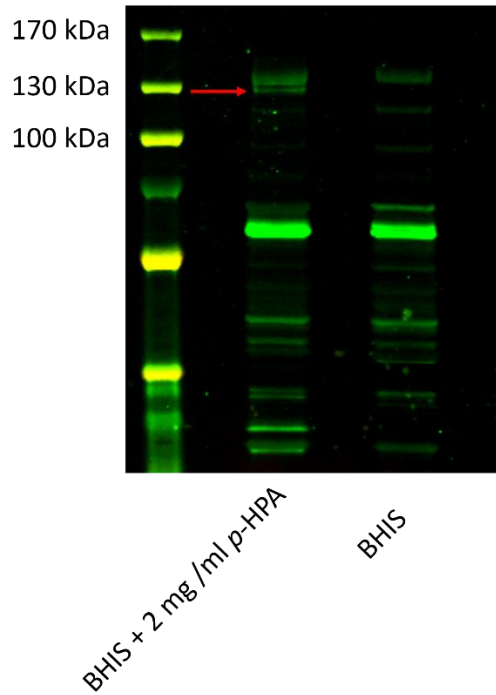
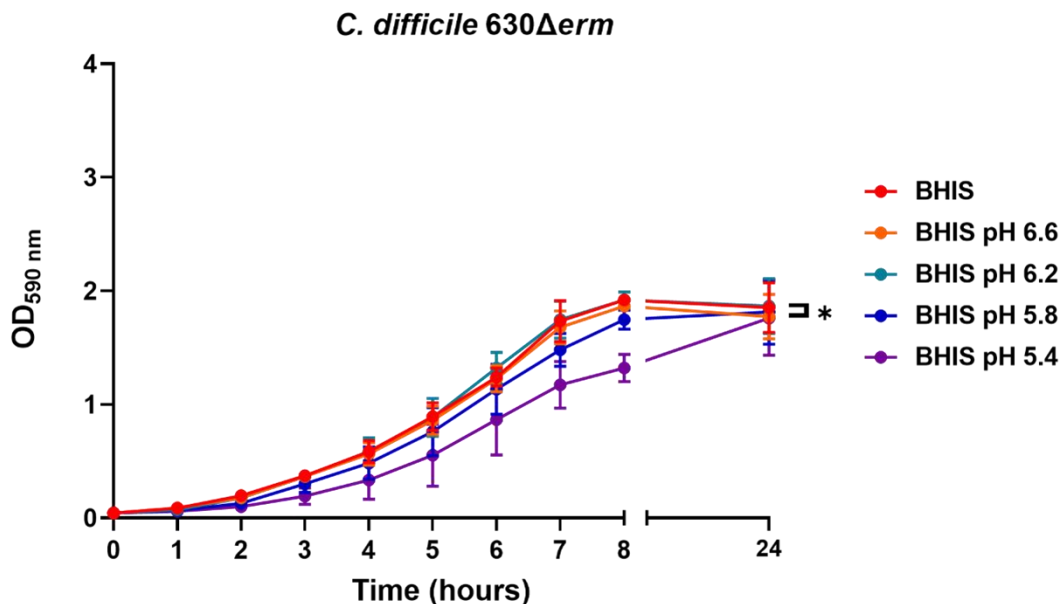


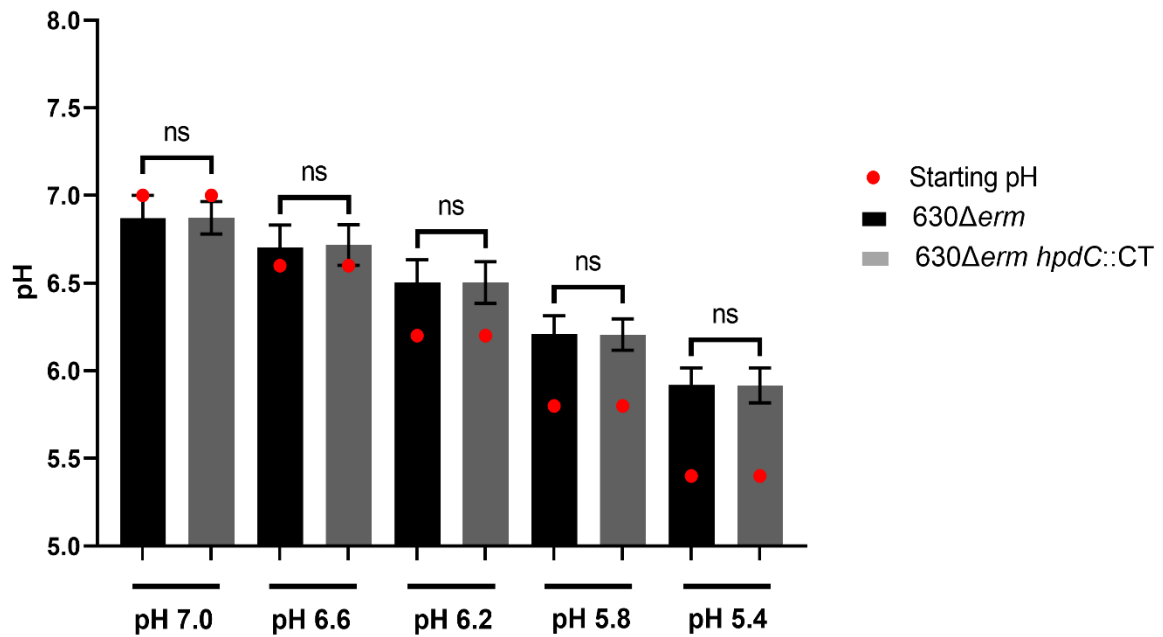
Figure 7. Mutation of *codY* reduces *p*-HPA conversion to *p*-cresol via a reduction in expression of the *hpdBCA* operon. a&b) 630Δerm and 630Δerm Δ*codY* were grown in defined minimal media (MM) with 2 mg/ml *p*-HPA for 8 hours with samples taken at 4 and 8 hours for HPLC analysis of *p*-HPA and *p*-cresol concentration. Turnover percentage was calculated as (*p*-cresol / *p*-HPA + *p*-cresol) normalised to growth by OD_{590 nm}. c) Expression from the *hpdBCA* promoter was assessed by phosphatase activity in 630Δerm and 630Δerm Δ*codY* carrying P_{*hpdB*}-*phoZ* grown for four hours in MM +/- 2 mg/ml *p*-HPA. Data represents mean and standard deviation of three independent replicates. Statistical analysis was carried out by linear regression and used to determine significant differences between normalised *p*-HPA turnover and promoter reporter activity for each strain. **p*<0.05, ***p*<0.01, ****p*<0.001



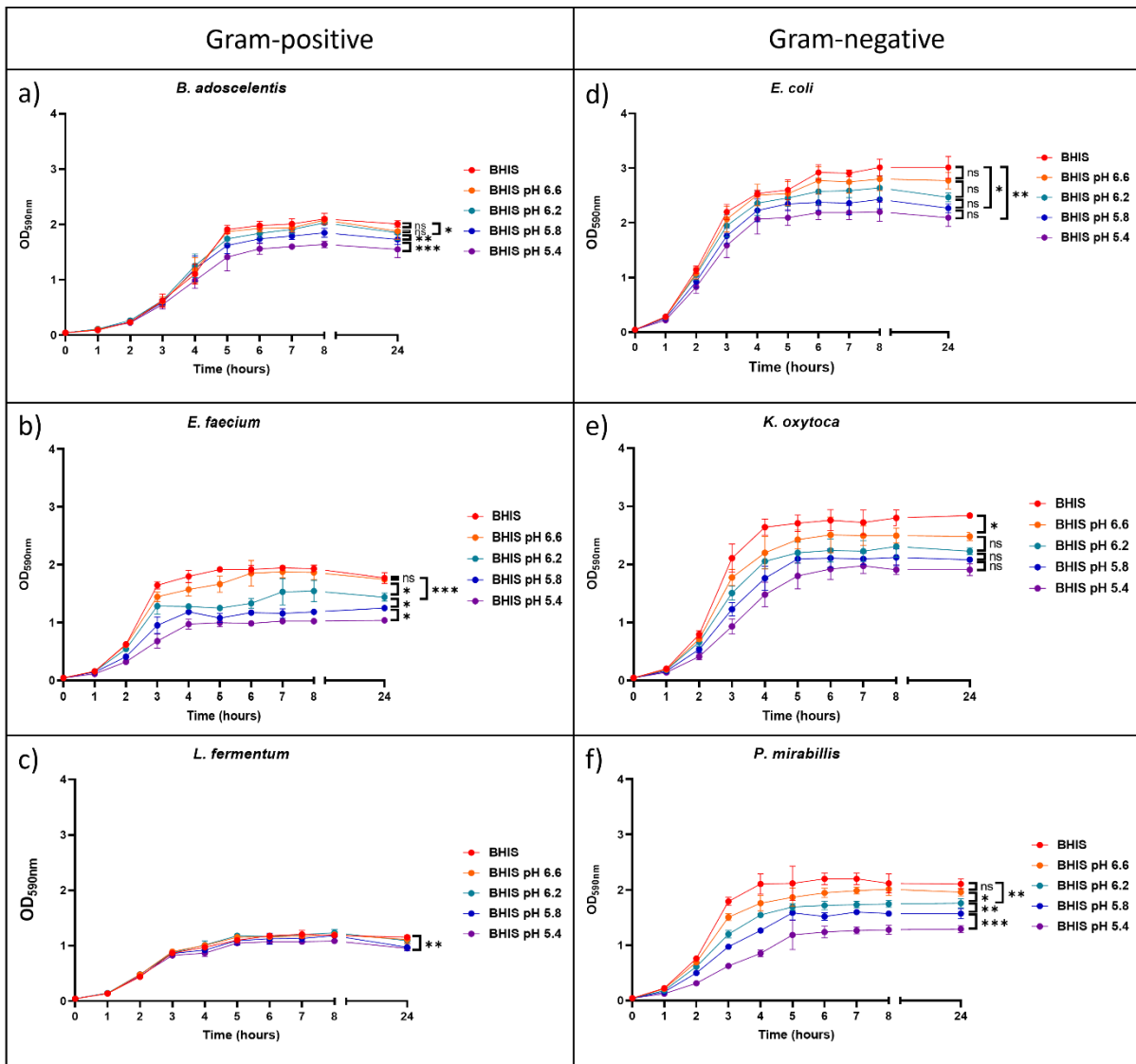
Supplementary Figure 1. Confirmation of HpdB-SNAP fusion. Samples were grown in the presence and absence of 2 mg/ml *p*-HPA before undergoing western blot analysis. The indicated band was approximately the right size for the HpdB-SNAP fusion (121 kDa). The band was excised from a duplicate gel and the presence of the HpdB-SNAP fusion was confirmed by mass-spectrometry.



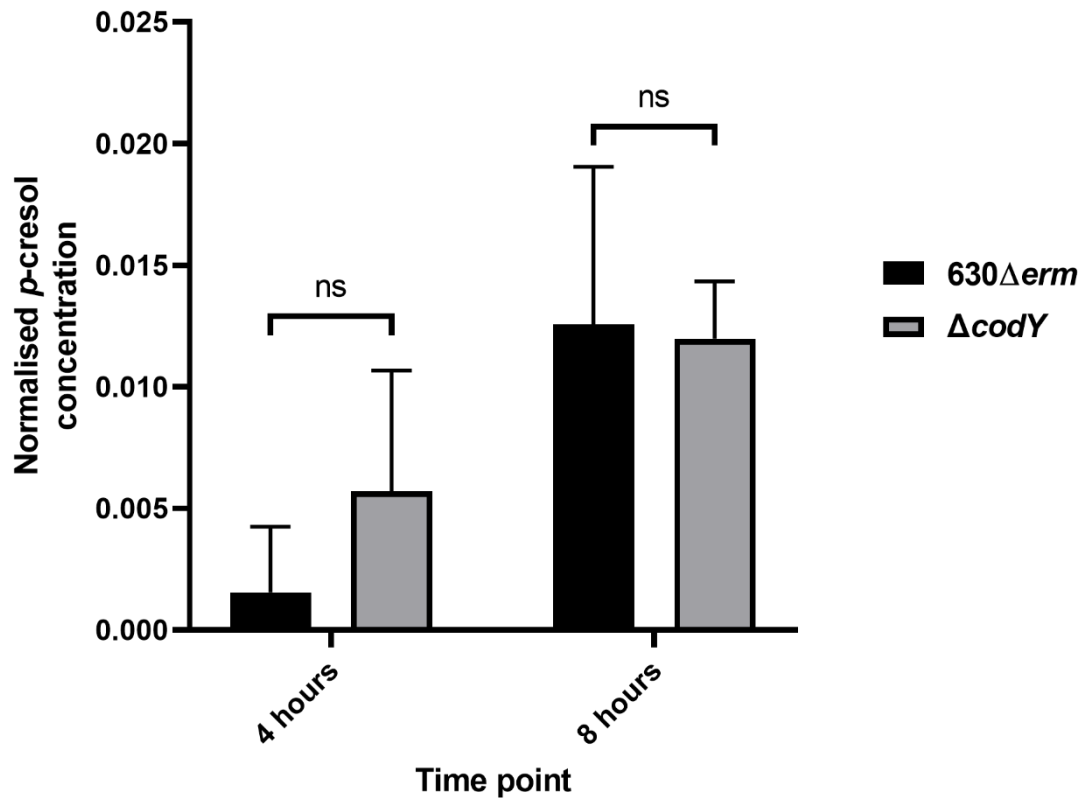
Supplementary figure 2. Growth of *C. difficile* in pH matched to the presence of *p*-HPA. *C. difficile* strain was grown in BHIS alongside BHIS with the pH lowered to 6.6, 6.2, 5.8, and 5.4 to match the pH found at 1, 2, 3 and 4 mg/ml *p*-HPA respectively. Statistical analysis by ANOVA was used to determine differences in growth at the different pH's. Error bars represent standard deviation and data represents a minimum of three independent replicates, significant differences are indicated with: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.



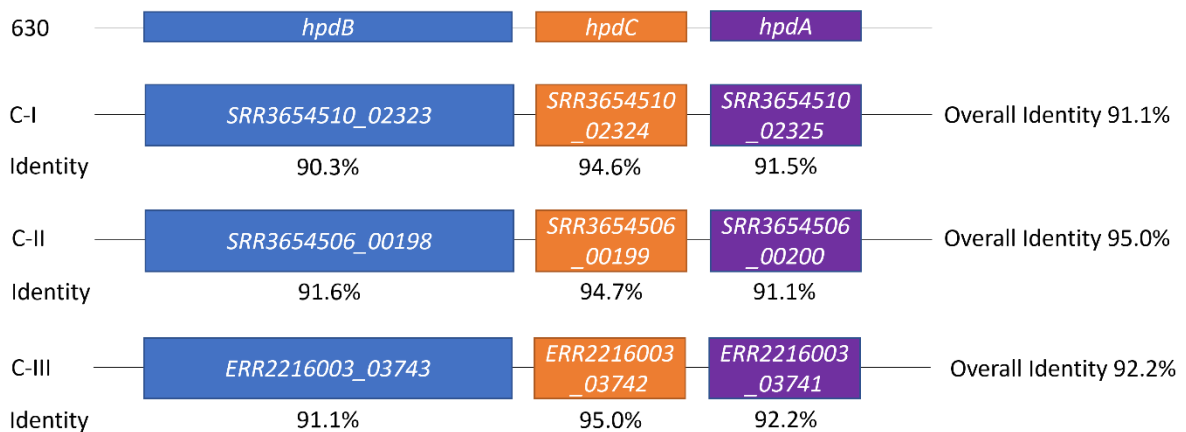
Supplementary figure 3. Growth of *C. difficile* in acidic BHIS media. 630Δerm and 630Δerm hpdC::CT strains were grown in BHIS alongside BHIS with the pH lowered to 6.6, 6.2, 5.8, and 5.4 to match the pH found at 1, 2, 3 and 4 mg/ml *p*-HPA respectively with the final pH measured after 24 hours growth. Regression analysis was used to determine whether there were any significant differences in final pH between the two strains at each starting pH. Error bars represent standard deviation and data represents a minimum of three independent replicates. ns: non-significant



Supplementary figure 4. Growth of gut commensals in pH matched to the presence of *p*-HPA. Each strain was grown in BHIS alongside BHIS with the pH lowered to 6.6, 6.2, 5.8, and 5.4 to match the pH found at 1, 2, 3 and 4 mg/ml *p*-HPA respectively. Representative Gram-positive (a, b & c) and representative Gram-negative (d, e & f) gut bacteria were assessed. Statistical analysis by ANOVA was used to determine differences in growth at the different pH's. Error bars represent standard deviation and data represents a minimum of three independent replicates, significant differences are indicated with: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.



Supplementary figure 5. *p*-cresol production from 630Δ*erm* and 630Δ*erm* Δ*codY*. Each strain was grown in minimal media for 8 hours, samples were taken after 4 and 8 hours which were analysed by HPLC for *p*-cresol concentration. *p*-cresol concentration was normalised to growth (as measured by OD_{590nm}) at the time the sample was taken. Regression analysis was used to determine significant differences between strains in *p*-cresol production.



Supplementary Figure 6. Alignment of *hpdB*CA-like operons carried by cryptic clades C-I, C-II and C-III to *hpdB*CA operon from 630. Alignments were carried out using the Emboss Needle software from EMBL-EBI.

Peptide	Identity to HpdB
KSDGDIPVVR	99%
SDGDIPVVR	99%
LASNTADELTK	100%
QFADEGMTVEEAR	100%
QAINVLER	100%
VCEEAQSLYAK	100%

Supplementary table 1. Peptides identified by mass spectrometry of HpdB-SNAP-tag. An anti-SNAP-tag western blot was run to probe identify the presence of HpdB fused to the SNAP-tag. Mass spectrometry was used to confirm the identity of the tested protein. Analysis was carried out in Scaffold v5.0.1 searching Uniprot's All Taxonomy database. Peptides identified as HpdB are listed. The peptides provide 5.9% of the complete HpdB protein sequence.

Author contributions

LFD and MH conceived and designed the study. MH and HK carried out the experimental work and data analysis. MH and LFD wrote the manuscript with contributions from BW and HK. All authors approved the manuscript prior to submitted manuscript.

Funding statement

Funding for MAH was provided by the Medical Research Council (LSHTM studentship MR/N013638/1). Funding for BWW was provided Medical Research Council grant: MR/K000551/1. Funding for LFD was provided by ISSF fellowship from the Wellcome Trust: 105609/Z/14/Z (LFD), <https://wellcome.ac.uk/funding>) and an Athena Swan Career Restart Fellowship (from London School of Hygiene and Tropical Medicine).

Declaration of interest

The authors declare that no competing interests exist.

Acknowledgements

The authors would like to thank Dr Isabelle Martin-Verstraete and Dr Bruno Dupuy for providing the CodY mutant and its parent strain and Dr Simon Baines for providing the gut commensal species used in this study. We would like to thank Dr Henrik Strahl for useful discussions and expertise relating to

cell envelope structures and stability, alongside Dr. Elizabeth McCarthy at LSHTM for providing training on the confocal microscope.

Chapter 5: Characterisation of regulation and expression of the *C. difficile sinR-CD2215* locus

5.1 Introduction

The control of lifestyle switching for pathogenic bacteria is a key virulence trait. In the model spore forming organism *B. subtilis*, there are three distinct options for lifestyle choice: sporulation, biofilm formation and cannibalism, which are all important traits that support niche adaptation (233-235). These interrelated pathways are under the control of the master sporulation regulator Spo0A. Spo0A is activated by a phosphor-relay system with phosphorylation of Spo0A allowing it to form a dimer, which is its active form (188). With a high level of phosphorylation Spo0A drives sporulation, whilst at low levels it drives biofilm formation (235). A significant part of Spo0A's role in driving these lifestyle choices is its interaction with the *sin* locus. The *sin* locus in *B. subtilis* is comprised of two genes: *sinI_{BS}* and *sinR_{BS}* (236) (Fig. 5.1). SinR_{BS} has a helix-turn-helix (HTH) domain as well as a multimerisation domain whilst SinI_{BS} carries only a multimerisation domain (191). SinR_{BS} is a repressor of sporulation, whilst SinI_{BS} is an antagonist of SinR_{BS} which functions by directly binding SinR_{BS} to form a heterodimer thus preventing the formation of the active SinR_{BS} tetramer (192). The active SinR_{BS} tetramer represses key genes related to lifestyle choice, such as *spo0A* (237), *spolIA*, *spolIE* and *spolIG* involved in sporulation (238), and the genes of the exopolysaccharide synthesis operon *epsA-O* (189) involved in biofilm formation. Therefore, through the direct protein-protein interaction with SinR_{BS}, SinI_{BS} plays a key role in lifting sporulation repression and determining lifestyle choice (191). Three promoters are involved in controlling expression of this locus in *B. subtilis* (Fig. 5.1). During vegetative growth SinR_{BS} expression is maintained at a constant level through constitutive expression from the P₃ promoter located between the two genes (236) (Fig. 5.1). Two further promoters, P₁ and P₂, are located upstream of both genes (236). Expression from the P₁ promoter is induced by the binding of phosphorylated Spo0A to a site directly upstream of the promoter (239) and despite the co-transcription of both genes from this promoter, SinI_{BS} expression is ten-fold higher than that of SinR_{BS} thus SinI_{BS} is able to significantly inhibit SinR_{BS} resulting in derepression of sporulation (240). The P₂

site also controls transcription of both genes, however, there is no detectable transcription from it during either early vegetative growth or early sporulation, with expression only detectable following the initiation of sporulation (236), therefore, it is considered less relevant in the control of lifestyle choice. *C. difficile* has two genes which are orthologous to *sinR_{BS}* (*CD2214* (*sinR*) and *CD2215*) but it does not have an ortholog to *sinI_{BS}* (194). Both SinR and CD2215 have multimerisation domains as well as HTH domains (194), which is in direct contrast to *B. subtilis*' SinI_{BS} which does not carry a HTH domain, only a multimerisation domain, so likely only functions via its interaction with SinR_{BS}.

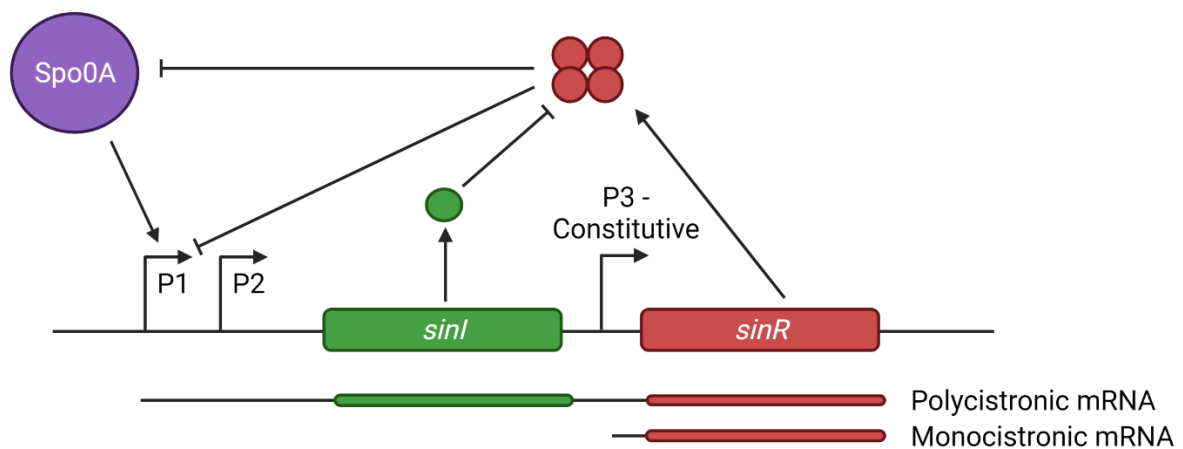


Figure 5.1. Structure of the *sin* locus in *B. subtilis*. Figure created with Biorender.com

In *C. difficile*, Girinathan *et al.* showed that both SinR and CD2215 are pleiotropic regulators with regulatory effects on virulence factors including toxin production, sporulation and motility (194). When Girinathan *et al.* investigated a double mutant of both *sinR* and *CD2215* in the R20291 strain they observed that this mutation resulted in an asporogenic phenotype whilst a single mutant of just *CD2215* led to a hyper-sporulating phenotype (194). However, an important caveat to these findings was that the authors were unable to successfully complement their sporulation phenotype in the double *sinR-CD2215* mutant (194). Despite this, Girinathan *et al.* concluded that, in direct contrast to *B. subtilis*, that SinR had a positive effect on sporulation whilst CD2215 was a repressor of sporulation (194). A further publication by Ciftci *et al.* showed, using pulldown experiments that the

multimerisation domain of CD2215 binds directly to SinR, indicating that this is the mechanism by which CD2215 prevents SinR from binding to DNA and exerting its regulatory effects on gene expression (195). Therefore, this system shows some similarity to the *sin* locus in *B. subtilis* in that one protein, using its multimerisation domain, binds to the second protein to inhibit it.

The mechanisms controlling expression of SinR_{BS} and SinI_{BS} in *B. subtilis* are well characterised, however, in *C. difficile*, little is known about either the features of the *sinR-CD2215* promoter or other factors involved in transcriptional regulation of *sinR* and/or *CD2215*. Girinathan *et al.*, provided evidence that expression of these genes occurs as an operon (194), however, the promoter site had yet to be identified. Recent evidence by Dhungel and Govind identified the presence of two Spo0A binding sites, located 219 and 276 base pairs upstream of the *sinR* start codon, which are used by Spo0A to repress expression of the two genes (196). Using the CodY consensus sequence, Dineen *et al.* identified a putative CodY binding site located upstream of the *sinR-CD2215* operon (180). The role of CodY in *sinR-CD2215* expression was investigated by Nawrocki *et al.*, who found that the role of CodY varied between the strains tested, with an increase in *sinR* transcription in a 630 Δ *erm codY* mutant compared to its parent strain but a decrease in *sinR* transcription in a *codY* mutant of the UK1 strain when compared to its parent strain (185). In the same work, Nawrocki *et al.* investigated the putative CodY binding site upstream of *sinR* using a *phoZ* transcriptional reporter fused to the upstream region of the *sinR-CD2215* operon (185). The reporter was altered by mutation of a single base pair of the putative CodY binding site, however, the authors found this mutation did not lead to any changes in expression from the promoter suggesting that this putative CodY site is not functional, however, the authors stated that mutation of a single base pair may not have been enough to prevent CodY from binding to this site (185). Further experiments into the role of CodY were carried out by Girinathan *et al.* who found, via EMSA, that recombinant CodY bound a 59 base pair probe of the upstream region of the *sinR-CD2215* operon which included the putative CodY binding site, as well as showing, via western blot, that CodY represses expression of the both SinR and CD2215 (194). Therefore, the role of both CodY and its putative binding site located upstream of *sinR* remains to be

fully determined. In addition to being regulated by both Spo0A and CodY, a number of transcriptomics studies have identified other regulators which exert effects on *sinR-CD2215* expression including CcpA (126), SigH (241) and SigD (242) which suggests that a complex network governs the expression of these key regulators.

The aim of this study was to characterise regulatory elements involved in expression of *sinR* and *CD2215*, including identification of functional promoters controlling either one or both of these genes. Furthermore, the work described in this study sought to further characterise the role of important *C. difficile* regulators Spo0A, CodY and CcpA in expression of the *sinR-CD2215* operon. The purpose of this research was to provide insight into the importance of this operon to *C. difficile*'s lifestyle switching and expression of key virulence factors, with a particular focus on sporulation given that spores are the agents of transmission for *C. difficile* (138).

5.2 Results

5.2.1 Transcriptional reporter construction

Using 5' RACE (Rapid Amplification of cDNA Ends) to identify putative transcriptional start sites the Dawson laboratory had evidence that suggested the presence of multiple promoter sites controlling both *sinR* and *CD2215* expression as well as *CD2215* alone (Personal communication, Dr Lisa Dawson). I sought to characterise these putative promoters to better determine the factors involved in the control of the expression of these two genes. Initially, experiments were undertaken to see if there was a promoter present that controlled expression of *CD2215* only. Two putative promoters were identified that were located within the coding sequence of *sinR* and therefore directly upstream of *CD2215*. To investigate whether there was a functional promoter upstream of *CD2215* and therefore controlling *CD2215* expression alone two reporters P_{2215} -Long and P_{2215} -Short were used (Fig. 5.2). These two reporters differed in that the Long reporter included a putative Spo0A binding site (TTAGACAT) not carried by the Short reporter. Two further reporters were designed; P_{sinR} -Short-SNAP and P_{sinR} -Long-SNAP which carried the sequence directly upstream of *sinR* to identify promoters controlling the expression of both genes (Fig. 5.2). The P_{sinR} -Long-SNAP reporter contained four additional putative Spo0A binding sites as well as a putative CodY binding site (Fig. 5.2). At the time of design, the two confirmed Spo0A binding sites had yet to be identified, however the site directly downstream of the CodY binding site was one of those confirmed by Dhungel and Govind (TTCTACA) (196). Each of the four described promoter regions controlled expression of a SNAP-tag reporter. The reporters were initially conjugated in to $630\Delta erm$, $630\Delta erm \ sinR$ -*CD2215*::CT, $630\Delta erm \ CD2215::CT and $630\Delta erm \ spo0A::CT. Importantly, insertional inactivation of *sinR* leads to a frame shift that also inactivates *CD2215* resulting in the double *sinR*-*CD2215* mutant (243).$$

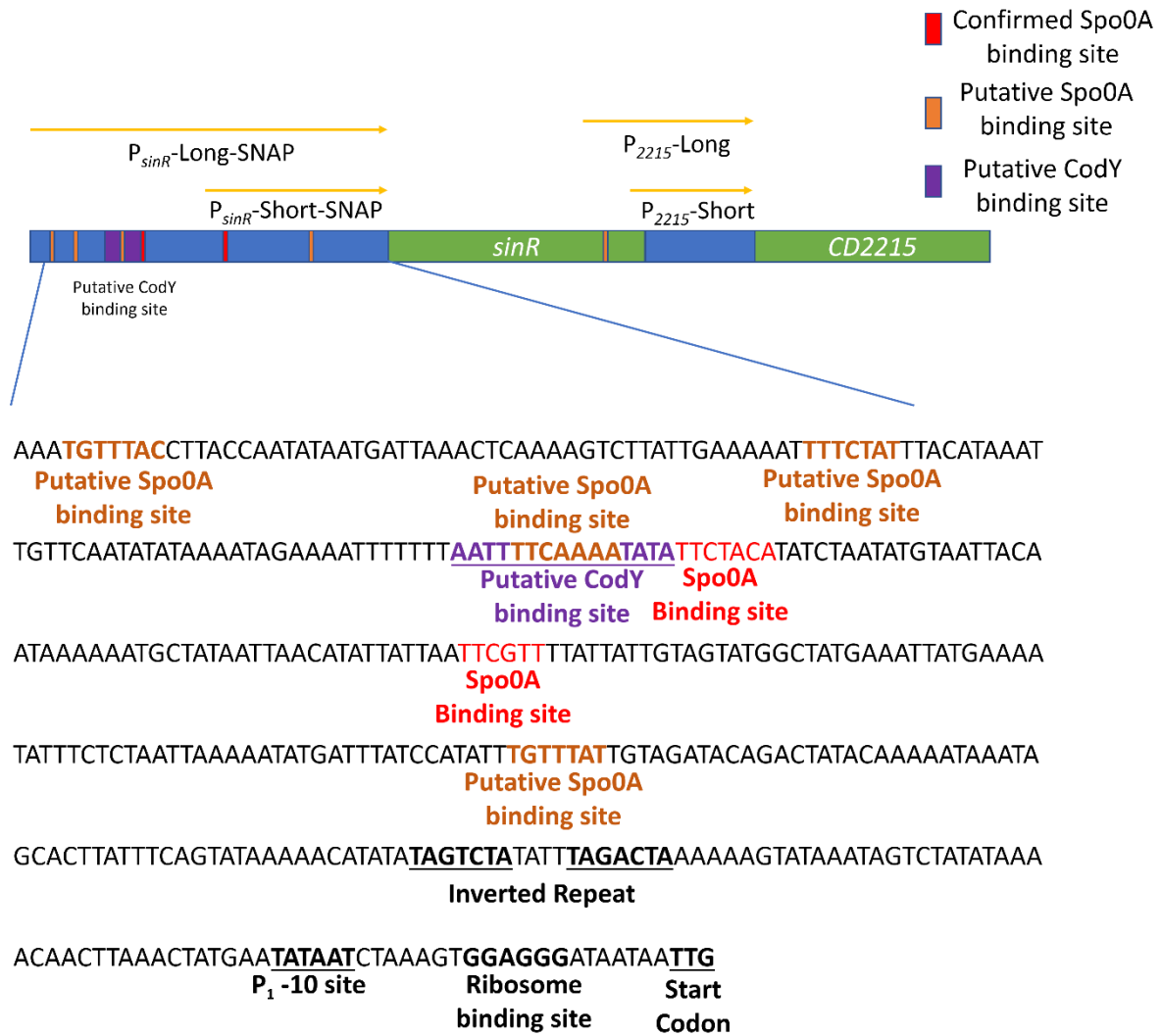


Figure 5.2. Diagram showing design of four reporter constructs for characterisation of regulatory controls of *sinR* and *CD2215*. Spo0A binding sites in red confirmed by Dhungel and Govind (196). The CodY site in purple and Spo0A binding sites marked in orange are putative binding sites identified based on their sequence.

5.2.2 Identification of a single promoter controlling *sinR*-*CD2215* expression

The initial experiments focused on the two P_{2215} reporters (testing as per methods detailed in chapter 2.8.1). Following growth to either $OD_{590\text{ nm}}$ 0.6 (exponential phase) or after 24 hours (stationary phase) expression could not be seen from either of the reporters, when run on SDS-PAGE, in any of the strains tested. This suggested that, under these conditions, there is no promoter that controls expression of *CD2215* alone and that both genes are controlled from the sequence upstream of *sinR* only (Fig. 5.3).

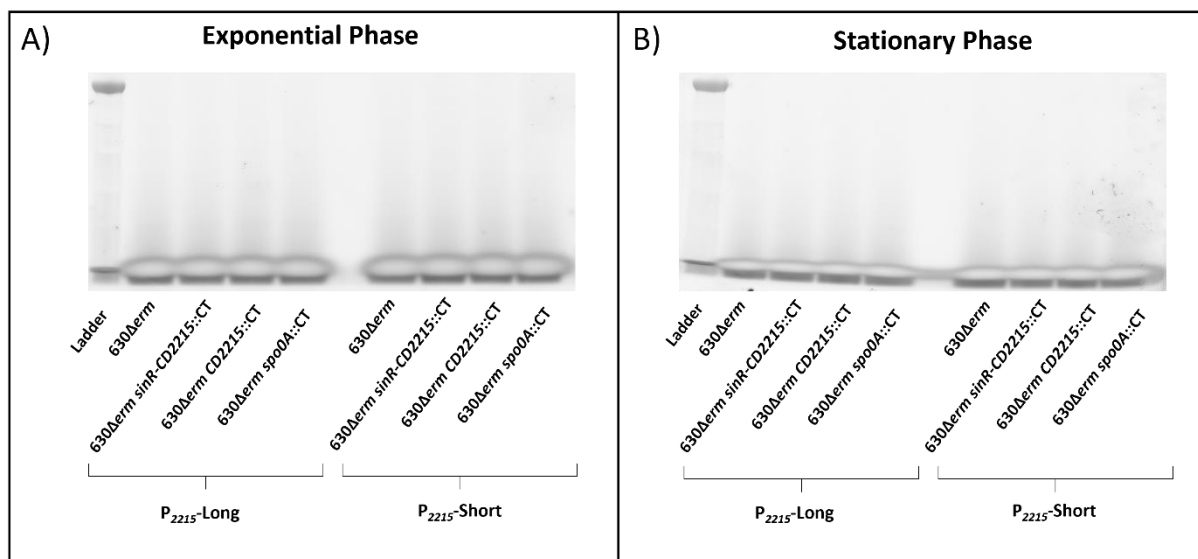


Figure 5.3. Visualisation of SNAP-tag reporter activity under control of putative promoters of *CD2215*. Strains 630Δerm, 630Δerm *sinR*-*CD2215*::CT, 630Δerm *CD2215*::CT and 630Δerm *spo0A*::CT carrying either the P_{2215} -Short or P_{2215} -Long SNAP-tag reporters were grown to (A) exponential or (B) stationary phase of growth. Samples were processed by addition of the fluorescent SNAP-tag substrate TMR-STAR before being run on SDS-PAGE.

A putative -10 site (P_1) was identified directly upstream of *sinR* (Fig. 5.2) with homology to the -10 site consensus sequence of the housekeeping sigma factor SigA (TATAAT) (241). To determine if this site was active the P_{sinR} -Short-SNAP reporter was used alongside a reporter with the putative P_1 -10 site mutated (P_{sinR} -Short-SNAP P_1 mut) by site directed mutagenesis (from TATAAT to TGCAAT) (chapter 2.4.3). These two reporters were conjugated into each of the four aforementioned strains and grown to exponential phase before visualisation of expression on SDS-PAGE and quantification using ImageQuant TL software. From the wild type reporter (P_{sinR} -Short-SNAP) expression was observed in 630 Δ *erm*, 630 Δ *erm* CD2215::CT and 630 Δ *erm* *spo0A*::CT, with quantification showing no significant differences in expression between these strains at this growth phase (Fig. 5.4). Notably, it was not possible to identify expression in the *sinR*-CD2215::CT strain from the P_{sinR} -Short-SNAP reporter, a finding characterised further below. Mutation of the P_1 -10 site led to no detectable expression from any of the strains tested, suggesting that under these conditions it is the only active promoter. Work by Harrison *et al.* using RT-PCR with primers that generate a product that encompasses parts of both *sinR* and CD2215 mRNA has shown that the two genes are co-transcribed providing evidence that the two genes are expressed as an operon (243). This suggests that control over different ratios of SinR:CD2215 may take place at a post-transcriptional level. This is in contrast to *B. subtilis*, where regulation of the *sin* locus occurs via multiple promoters (236).

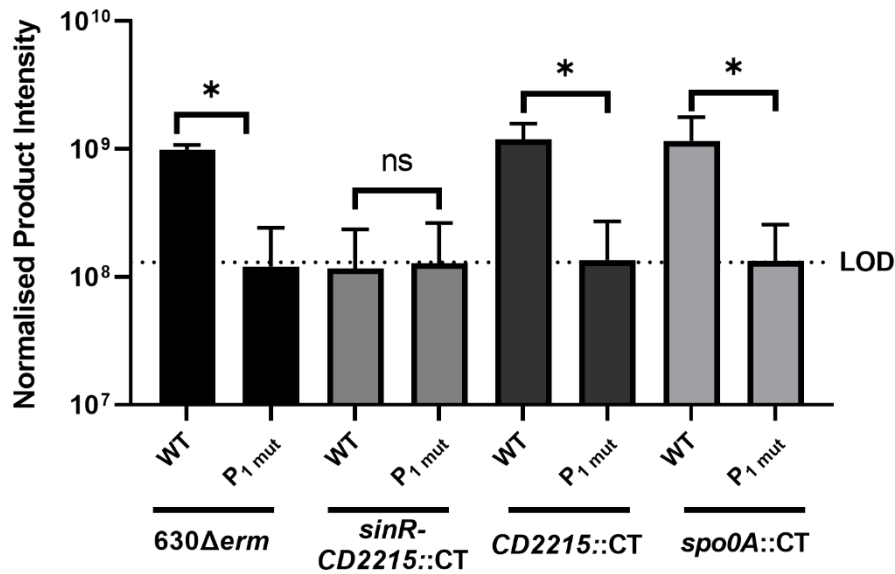


Figure 5.4. Expression from P_{sinR} -Short-SNAP and P_{sinR} -Short-SNAP $P_{1 mut}$. Strains $630\Delta erm$, $630\Delta erm sinR-CD2215::CT$, $630\Delta erm CD2215::CT$ and $630\Delta erm spo0A::CT$ carrying either the P_{sinR} -Short-SNAP or P_{sinR} -Short-SNAP $P_{1 mut}$ SNAP-tag reporters were grown to exponential phase of growth ($OD_{590 nm}$ 0.6). Samples were processed by addition of TMR-Star prior to running on SDS-PAGE and quantification of the SNAP-tag band was carried out using ImageQuant TL software. Normalised Product Intensity = Pixel volume / $OD_{590 nm}$. Dotted lines represent the limit of detection (LOD) calculated as the average of the results of the following formula for all samples: background pixel volume/culture $OD_{590 nm}$. Data represents means and standard deviation of three biological replicates. Statistical analysis was undertaken using linear regression to determine any significant differences in SNAP-tag expression between the two reporters carried by each strain. ns not significant * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

5.2.3 Expression from the *sinR-CD2215* promoter in *sinR-CD2215*, *CD2215* and *spo0A* mutants

To identify regulatory proteins involved in transcriptional regulation of *sinR* and *CD2215* the P_{sinR} -Long-SNAP and Short-SNAP reporters were used to determine if mutation of *sinR*, *CD2215* or *spo0A* led to altered expression of the reporters at either exponential or stationary phases of growth. The P_{sinR} -Long-SNAP reporter included a putative CodY binding site as well as four putative Spo0A binding sites not found in the P_{sinR} -Short-SNAP reporter. Expression was tested with both reporters at $OD_{590\text{ nm}} 0.6$ and after 24 hours so that samples were tested from both exponential and stationary phases of growth. At both growth phases expression of the SNAP-tag could not be measured in the *sinR-CD2215::CT* strain from either reporter but was present from the *CD2215::CT* strain (Fig. 5.5), indicating that SinR enhances expression of the operon and acts as an autoregulator. No significant differences were observed between expression in the wild type strain and the *CD2215::CT* or *spo0A::CT* with either reporter at either time point (Fig. 5.5), suggesting that neither *CD2215* or *Spo0A* up or down regulate expression of the operon.

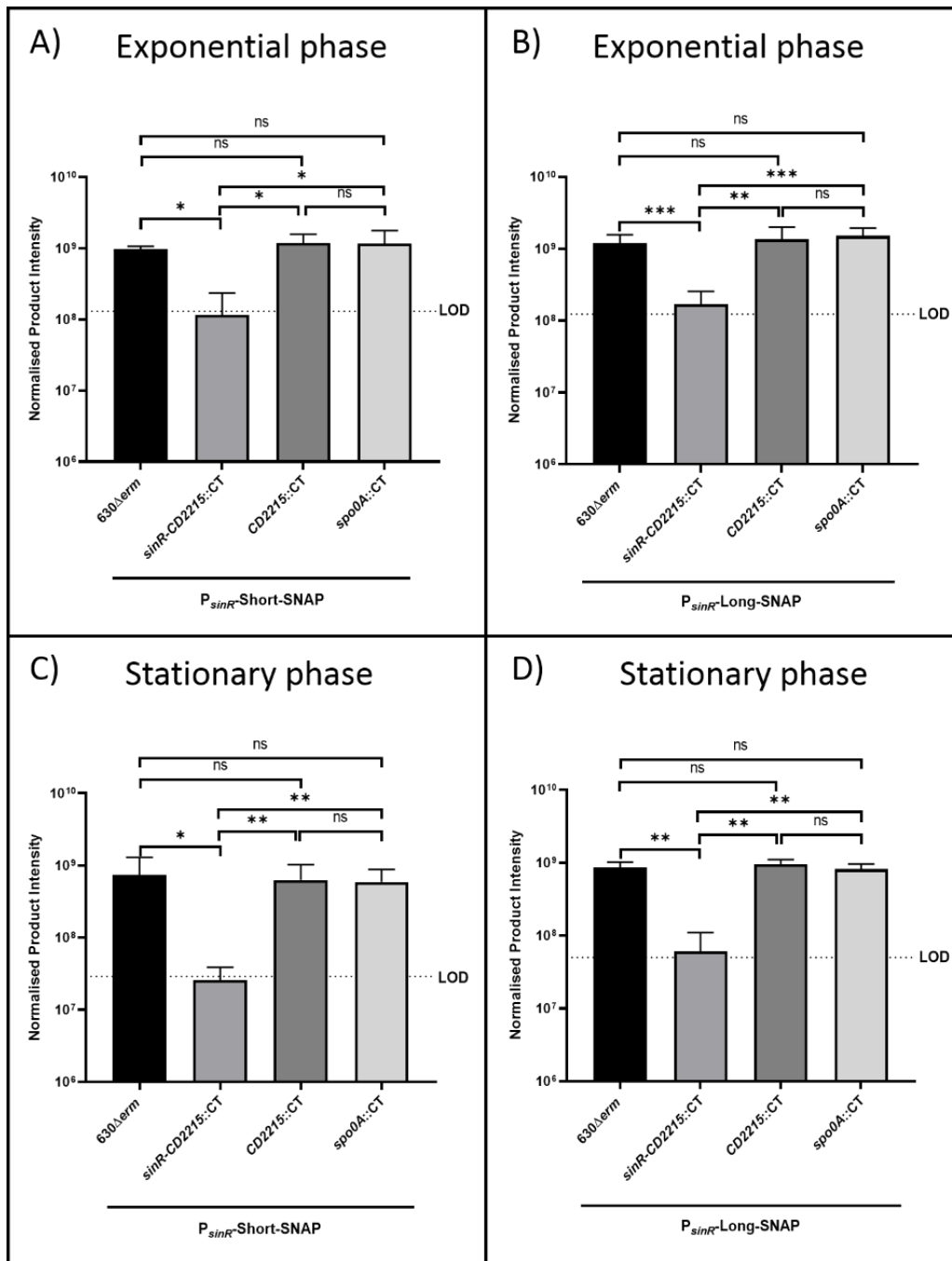


Figure 5.5. Quantification of SNAP-tag expression from reporters P_{sinR} -Short-SNAP and P_{sinR} -Long-SNAP at exponential and stationary phases of growth. Strains 630Δerm, 630Δerm sinR-CD2215::CT, 630Δerm CD2215::CT and 630Δerm spo0A::CT were transformed with either the P_{sinR} -Short-SNAP reporter (A&C) or the P_{sinR} -Long reporter (B&D). Each strain was grown to either approximately OD_{590 nm} 0.6 (exponential phase) or grown for 24 hours (stationary phase) and processed by addition of TMR-Star prior to running on SDS-PAGE and quantification of the SNAP-tag band using ImageQuant TL software. Normalised Product Intensity = Pixel volume / OD_{590 nm}. Dotted lines represent the limit of detection (LOD) calculated as the average of the results of the following formula for all samples: background pixel volume/culture OD_{590 nm}. Data represents means and standard deviation of three biological replicates. Statistical analysis was undertaken using linear regression to determine any significant differences in SNAP-tag expression between strains. ns not significant * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

The SNAP-tag reporter can lack sensitivity for the detection of small to modest changes in expression, as described in chapter 3.1 (244). Additionally, recent evidence demonstrated that Spo0A regulates expression of the *sinR-CD2215* operon with Dhungel and Govind using a *gusA* reporter as well as pull-down experiments to show that Spo0A represses expression of the *sinR-CD2215* operon by binding to two different sites located 276 and 219 base pairs upstream of the start codon (196). As a result, a more sensitive reporter was used and based on the study in chapter 3.1 the *phoZ* reporter was selected as it was easier to use than the *gusA* reporter. Therefore, the SNAP-tag reporter carried by P_{*sinR*}-Long-SNAP was replaced with the more sensitive *phoZ* reporter to generate construct P_{*sinR*}-Long-*phoZ* which was used to determine if the SNAP-tag's lack of sensitivity led to missing a regulatory effect of *spo0A*. As before, strains carrying the reporter were grown to exponential and stationary phases before quantification of reporter expression (chapter 2.8.3). The more sensitive *phoZ* reporter showed that there was significantly higher expression in the *spo0A::CT* strain at exponential phase, whilst no significant differences were identified at stationary phase (Fig. 5.6). This shows that Spo0A does indeed repress expression of the *sinR-CD2215* operon, in agreement with Dhungel and Govind (196), however, this study demonstrates that this is a growth phase dependent effect, as well as providing further evidence that the SNAP-tag is limited in its utility as a transcriptional reporter.

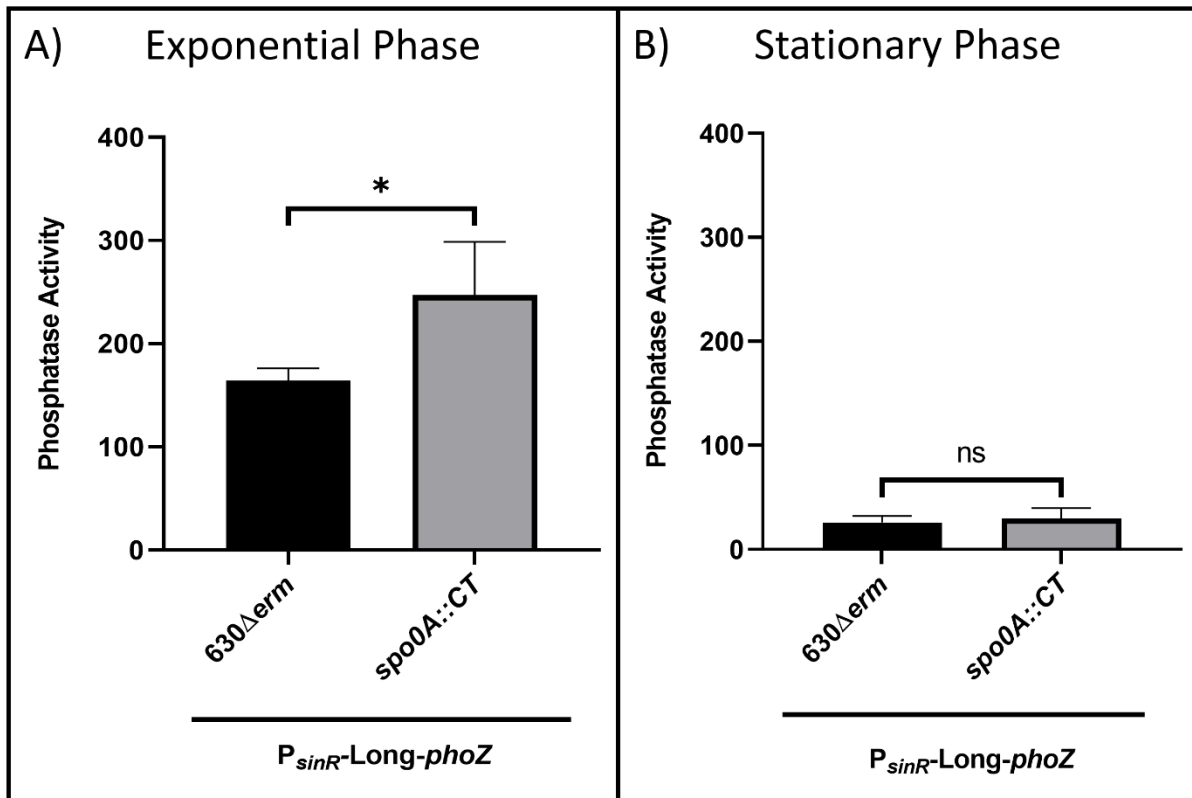


Figure 5.6. *phoZ* reporter activity under the control of the *sinR*-CD2215 promoter in 630Δerm and 630Δerm *spo0A*::CT. 630Δerm and 630Δerm *spo0A*::CT carrying P_{sinR}-Long-*phoZ* were grown to exponential phase (OD_{590 nm} 0.6) and stationary phase (24 hours growth). Expression was determined by the formula $[OD_{420} - (OD_{550} \times 1.75)] \times 1000 / t(\text{min}) \times OD_{590 \text{ nm}} \times \text{vol. cells (ml)}$ as per Edwards *et al.* (172), the OD_{590 nm} reading used was that taken at the time of sample harvesting. Data represents means and standard deviations of a minimum of three biological replicates. Statistical analysis was undertaken by linear regression to determine differences between the strains. ns not significant, * $p < 0.05$

5.2.4 Promoter expression in a single *sinR* deletion mutant

After generating the results with the Clostron knockout strains described above, a clean deletion mutant of *sinR* and its parent strain (630 Δ *erm* (Paris)) were obtained from collaborators at the Institut Pasteur (Dr Isabelle Martin-Verstraete and Dr Bruno Dupuy). This strain was used to confirm if deletion of *sinR* only leads to a significant drop in expression from the *sinR*-CD2215 promoter using the P_{sinR} -Long-SNAP reporter. At both exponential and stationary phases of growth, expression of the SNAP-tag reporter was below the limit of detection in the Δ *sinR* mutant (Fig. 5.7), confirming that *sinR* enhances expression of the operon.

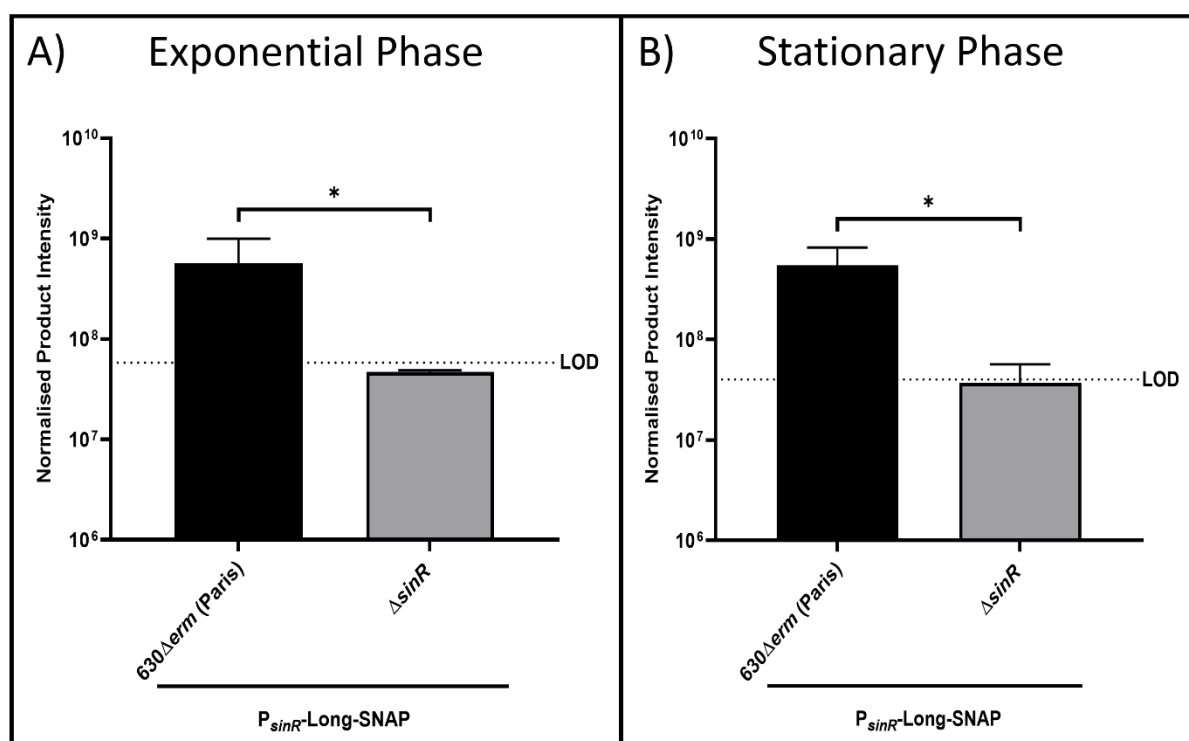


Figure 5.7. Quantification of SNAP-tag expression from reporter P_{sinR} -Long-SNAP in 630 Δ *erm* (Paris) and 630 Δ *erm* Δ *sinR*. Strains 630 Δ *erm* (Paris) and Δ *sinR* (in the 630 Δ *erm* Paris background) carrying the P_{sinR} -Long-SNAP reporter were grown to exponential (A) or stationary phase of growth (B). Samples were processed by addition of TMR-Star prior to running on SDS-PAGE and quantification of the SNAP-tag band using ImageQuant TL software using the formula Normalised Product Intensity = Pixel volume / OD_{590 nm}. Dotted lines represent the limit of detection (LOD) calculated as the average of the results of the following formula for all samples: background pixel volume/culture OD_{590 nm}. Data represents means and standard deviation of three biological replicates. Statistical analysis was undertaken using linear regression to determine any significant differences in SNAP-tag expression between strains. ns not significant * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

5.2.5 Regulators affecting expression from the *sinR-CD2215* promoter

A number of key regulators, including CcpA and CodY, are known to be involved in gene expression relating to key virulence factors including toxin production and sporulation. Using microarray analysis it has been shown that in a mutant of *ccpA*, expression of both *sinR* and *CD2215* significantly increases (126). Similarly, CodY has been shown to have significant effects on expression of *sinR* and *CD2215*, albeit with contrasting results across strains (185, 194). Therefore, to investigate whether part of these regulators mechanisms of action may be via the *sinR-CD2215* operon, the P_{sinR} -Long-SNAP reporter was conjugated into knockout strains of each of these two regulators ($630\Delta erm \Delta codY$ and $630\Delta erm \Delta ccpA$) acquired from collaborators at the Institut Pasteur (Dr Isabelle Martin-Verstraete and Dr Bruno Dupuy). The $\Delta codY$ strain was of particular interest as a putative CodY binding site had been identified directly upstream of the operon promoter (180), and CodY is a known repressor of sporulation (185). Here, the P_{sinR} -Long-SNAP reporter showed that there was a significant decrease in expression at both exponential and stationary phases of growth in the *codY* mutant compared to the wild type strain (Fig. 5.8A&B) suggesting that CodY enhances expression of *sinR* and *CD2215*.

In their study, Nawrocki *et al.*, produced a transcriptional reporter fusion with *phoZ* under the control of the *sinR-CD2215* upstream region. As part of this investigation the authors generated a mutant of this reporter with a single base pair alteration made to the putative CodY binding site found upstream of *sinR* (AATTTTCAAATATA to AATTTTAAAATATA). The authors found that this mutation did not have any effect on expression from the promoter, however, they did note that the single base pair alteration made to the putative CodY binding site may not have been sufficient to prevent CodY binding (185). To determine the role of this putative binding site it was removed in full from the P_{sinR} -Long-SNAP reporter by inverse PCR. When tested, the SNAP-tag assay showed that there was no difference in expression between the wild type P_{sinR} -Long-SNAP reporter compared to the reporter lacking the CodY binding site: P_{sinR} -Long-SNAP Δ CodY (Fig. 5.8C&D) at either growth phase. Therefore, whilst it appears CodY does upregulate expression of *sinR-CD2215*, this effect is either indirect or there is another yet to be identified CodY binding site upstream of the operon.

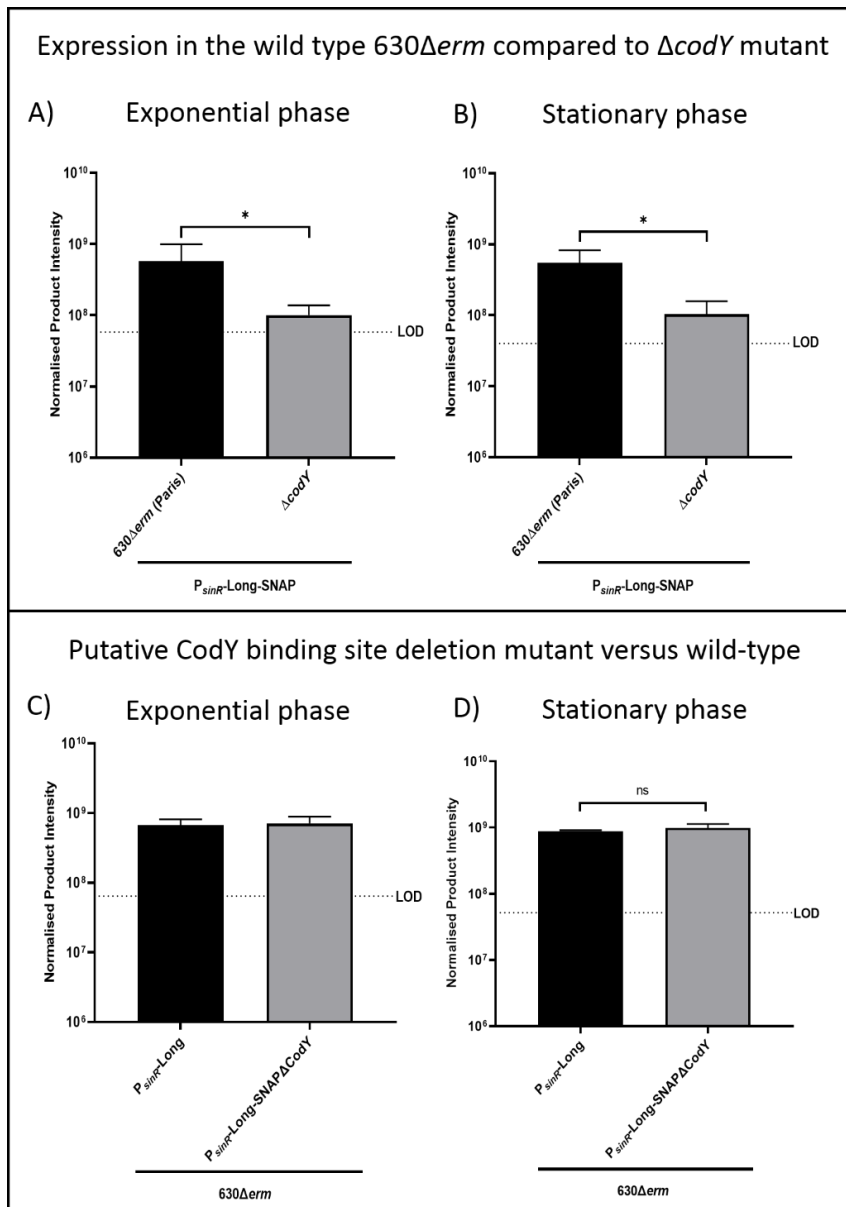


Figure 5.8. The effect of mutation of *codY* and a putative *codY* binding site on expression from the *sinR-CD2215* operon promoter. 630 Δ erm (Paris) and 630 Δ erm Δ codY were transformed with P_{sinR} -Long-SNAP and were grown to exponential phase (A) or stationary phase of growth (B). 630 Δ erm carrying the P_{sinR} -Long-SNAP reporter with the wild type upstream promoter region and P_{sinR} -Long-SNAP Δ CodY lacking the putative CodY binding site were tested at exponential (C) and stationary (D) phases of growth. Samples were processed by addition of TMR-Star prior to running on SDS-PAGE and quantification of the SNAP-tag band using the formula Normalised Product Intensity = Pixel volume / OD_{590 nm}. Dotted lines represent the limit of detection (LOD) calculated as the average of the results of the following formula for all samples: background pixel volume/culture OD_{590 nm}. Data represents means and standard deviation of three biological replicates. Statistical analysis was undertaken using linear regression to determine any significant differences in SNAP-tag expression between strains. ns not significant * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

Another global transcriptional regulator of interest was CcpA as it has been shown to repress expression of both *sinR* and *CD2215* as well as repressing the transcription of a number of key sporulation genes including the master sporulation regulator *spo0A* (126). The P_{sinR} -Long-SNAP reporter was conjugated in to the $\Delta ccpA$ strain to explore whether CcpA has any effect on expression from the *sinR-CD2215* promoter. The SNAP-tag results showed that expression was increased at both growth phases in the $\Delta ccpA$ knockout, however, this increase was statistically insignificant (Fig. 5.9A&B). Therefore, this was repeated with the more sensitive P_{sinR} -Long-*phoZ* reporter in the $\Delta ccpA$ strain. Using the *phoZ* reporter, it was found that at exponential phase there was no difference in expression (Fig. 5.9C) between wild type and $\Delta ccpA$ strains, however, at stationary phase the $\Delta ccpA$ strain had significantly higher expression of *phoZ* (Fig. 5.9D). This indicates that CcpA acts as a repressor of operon expression at stationary phase. This is likely to be an indirect effect mediated by another gene under the control of CcpA as there is no putative CcpA binding site located directly upstream of *sinR* (126).

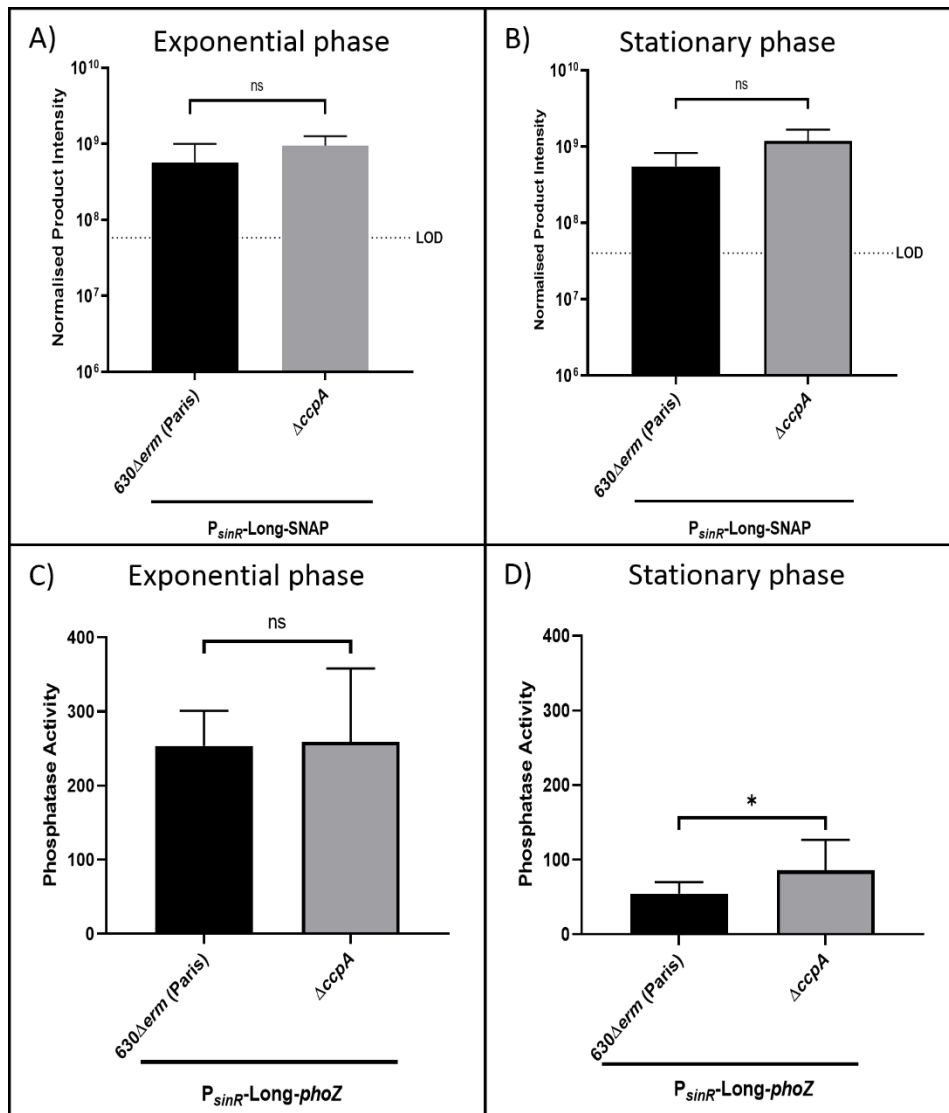


Figure 5.9. Quantification of expression from the *sinR-CD2215* promoter in *630Δerm (Paris)* and *ΔccpA* using *P_{sinR}-Long-SNAP* and *P_{sinR}-Long-phoZ* reporters. Strains *630Δerm (Paris)* and *630Δerm ΔccpA* carrying the *P_{sinR}-Long-SNAP* reporter were grown to exponential phase (A) or stationary phase of growth (B). Samples were processed by addition of TMR-Star prior to running on SDS-PAGE and quantification of the SNAP-tag band using the formula Normalised Product Intensity = Pixel volume / OD_{590 nm}. Dotted lines represent the limit of detection (LOD) calculated as the average of the results of the following formula for all samples: background pixel volume/culture OD_{590 nm}. Strains *630Δerm (Paris)* and *630Δerm ΔccpA* carrying the *P_{sinR}-Long-phoZ* reporter were grown to exponential phase (C) or stationary phase of growth (D) with PhoZ activity determined by the formula $[\text{OD}_{420} - (\text{OD}_{550} \times 1.75)] \times 1000 / t(\text{min}) \times \text{OD}_{590 \text{ nm}} \times \text{vol. cells (ml)}$ as per Edwards *et al.* (172), the OD_{590 nm} reading used was that taken at the time of sample harvesting. Data represents means and standard deviation of three biological replicates. Statistical analysis was undertaken using linear regression to determine any significant differences in SNAP-tag expression between strains. ns not significant * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

5.2.6 Identification of the *sinR* binding site

As shown by the data from the P_{sinR} -Long-SNAP reporter, it is likely that SinR is an enhancer of expression from the *sinR-CD2215* promoter. Building on work by Dr Dawson and Maria Derakhshan which identified that SinR bound to the upstream region of *sinR* this study sought to identify whether the precise residues bound by SinR were those of an inverted repeat (TAGTCTA-n4-TAGACTA), identified by Dr Dawson, located 71 base pairs upstream of the *sinR* start codon, especially as regulatory factors often bind to genetic features such as inverted repeats. To investigate this, mutants of the P_{sinR} -Long-SNAP reporter were made by inverse PCR (chapter 2.4.3), which was used to either fully remove the inverted repeat, or just the 5' arm or 3' arm of the inverted repeat to give plasmids P_{sinR} -Long-SNAP Δ IR, P_{sinR} -Long-SNAP Δ 5'IR and P_{sinR} -Long-SNAP Δ 3'IR, respectively. These reporters were used in SNAP-tag assays, and as templates to generate PCR products, using infrared dyed primers, for EMSAs (chapter 2.7). Purified recombinant SinR was produced by collaborators at King's College London (Isaacson laboratory - Janina Muench). Additionally, recombinant CD2697 was produced by Dr Catherine Hall at LSHTM and used as a negative control as CD2697 is a hypothetical membrane protein with no predicted ability to bind DNA. The EMSAs demonstrated that the SinR protein bound to the wild type upstream sequence, however, it did not bind to any of the inverted repeat mutants suggesting that enhancement of expression of *sinR-CD2215* is as a result of a direct interaction between SinR and this inverted repeat, with each arm of the inverted repeat required for binding (Fig. 5.10). In support of the EMSAs, the SNAP-tag data clearly demonstrated that removal of the inverted repeat led to significantly reduced expression levels in the wild type $630\Delta erm$, *CD2215::CT* and *spo0A::CT* strains to below or at the limit of detection, making them comparable to that of the *sinR-CD2215::CT* and $\Delta sinR$ strains where expression is low due to the mutation of *sinR* (Fig. 5.11). Therefore, these two results taken together show that SinR binds directly to this inverted repeat leading to enhancement of expression from the *sinR-CD2215* operon promoter.

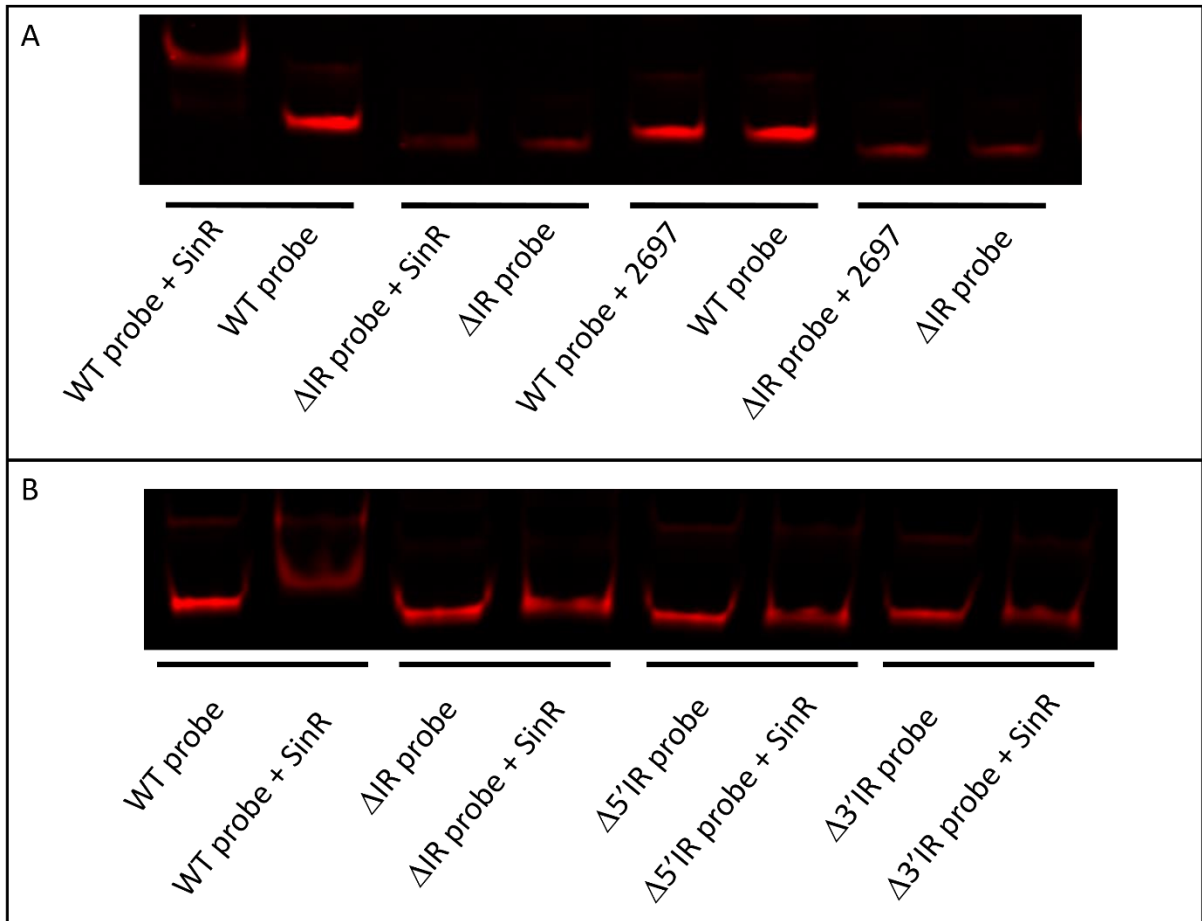


Figure 5.10. Electromobility shift assay to determine binding of SinR to the *sinR*-CD2215 upstream region. EMSAs were used to determine if purified recombinant SinR bound to the *sinR*-CD2215 upstream region (210 bp). IR-dyed primers were used to amplify the *sinR*-CD2215 upstream region from reporters: P_{*sinR*}-Long-SNAP, P_{*sinR*}-Long-SNAP Δ IR, P_{*sinR*}-Long-SNAP Δ 5'IR or P_{*sinR*}-Long-SNAP Δ 3'IR to generate the WT, Δ IR, Δ 5'IR and Δ 3'IR probes respectively. A) EMSAs were carried out to determine binding of: A) recombinant SinR or CD2697 to WT or Δ IR probes, B) recombinant SinR to WT, Δ IR, Δ 5'IR and Δ 3'IR probes. Each assay was carried out in triplicate.

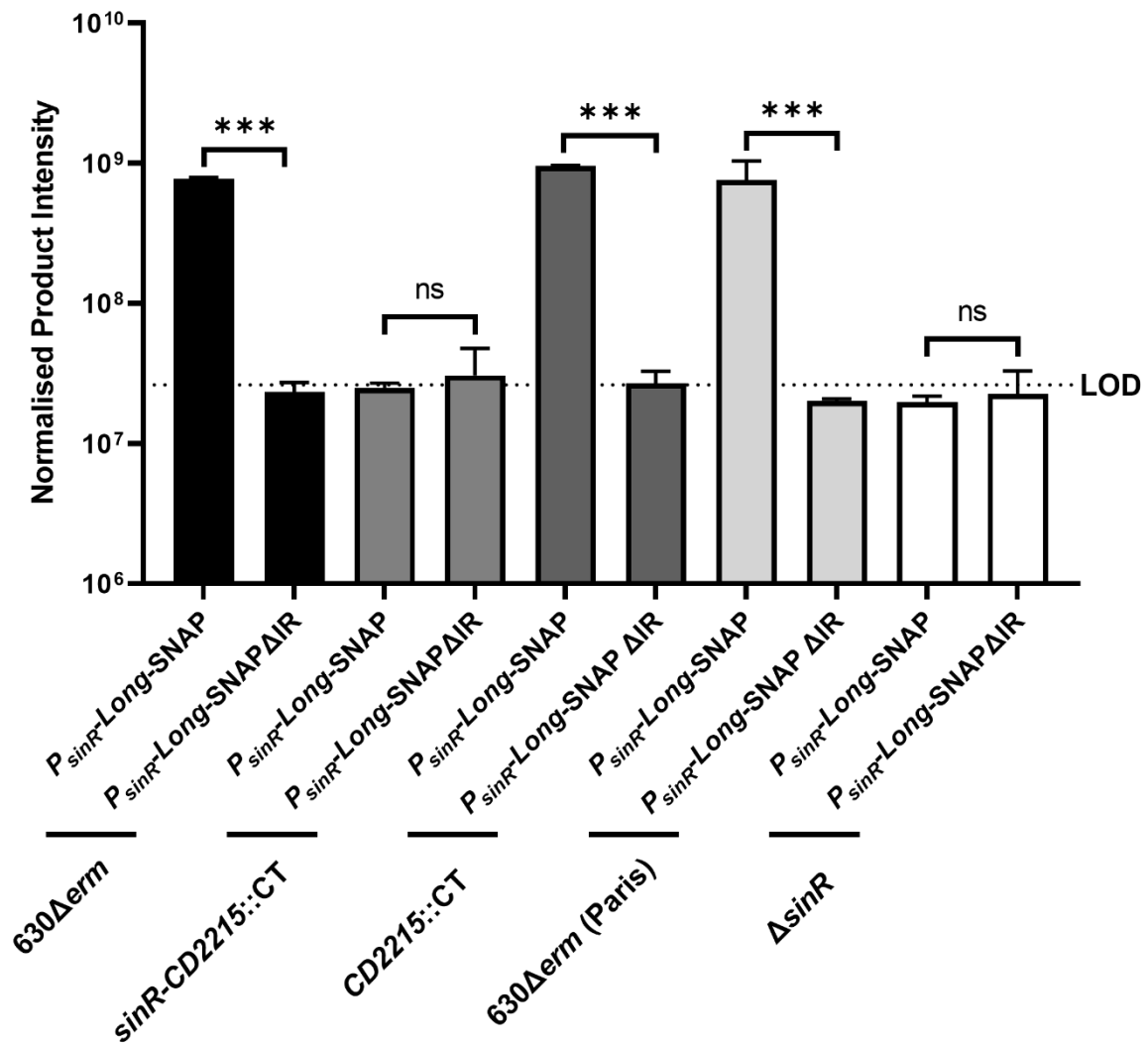


Figure 5.11. Quantification of SNAP-tag expression from reporters P_{sinR}-Long-SNAP and P_{sinR}-Long-SNAP ΔIR. 630Δerm, 630Δerm sinR-CD2215::CT, 630Δerm CD2215::CT, 630Δerm (Paris) and 630ΔermΔsinR were transformed with P_{sinR}-Long-SNAP or P_{sinR}-Long-SNAP ΔIR and were grown to stationary phase. Samples were processed by addition of TMR-Star prior to running on SDS-PAGE and quantification of the SNAP-tag band using the formula Normalised Product Intensity = Pixel volume / OD_{590 nm}. Dotted lines represent the limit of detection (LOD) calculated as the average of the results of the following formula for all samples: background pixel volume/culture OD_{590 nm}. Data represents means and standard deviation of three biological replicates. Statistical analysis was undertaken using linear regression to determine any significant differences in SNAP-tag expression between strains. ns not significant * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

5.2.7 SinR and CD2215 protein quantification

After finding that SinR and CD2215 are transcribed from a single promoter, it suggested that to avoid the proteins being at a constant level with one another, and therefore CD2215 always inhibiting SinR, there may be post-transcriptional processing of the mRNA to ensure concentrations of the protein differ depending on the cell's needs. To compare protein levels the coding sequences of both *sinR* and *CD2215* were inserted in to the P_{sinR} -Long-SNAP construct in place of the SNAP-tag, so that the genes were under the control of their own promoter. This was followed by the generation of two new constructs; one with SinR with a 3' His-tag (P_{sinR} -SinR-His) and one with CD2215 with a 3' His-tag (P_{sinR} -CD2215-His). The constructs were conjugated in to 630 Δ *erm* and then strains carrying these plasmids were grown to exponential and stationary phases of growth before undergoing processing, including a His-purification step, for a western blot (chapter 2.9.7). A positive control of a recombinantly produced and purified His-tagged CD2697 was included to ensure the functionality of the anti-His antibody used. Unfortunately, whilst clear binding to the positive control (CD2697) was identified, neither the His-tagged SinR or CD2215 could be detected by this method from the cell lysates produced (Fig. 5.12), suggesting that the proteins may be expressed at a level too low to detect by this method.

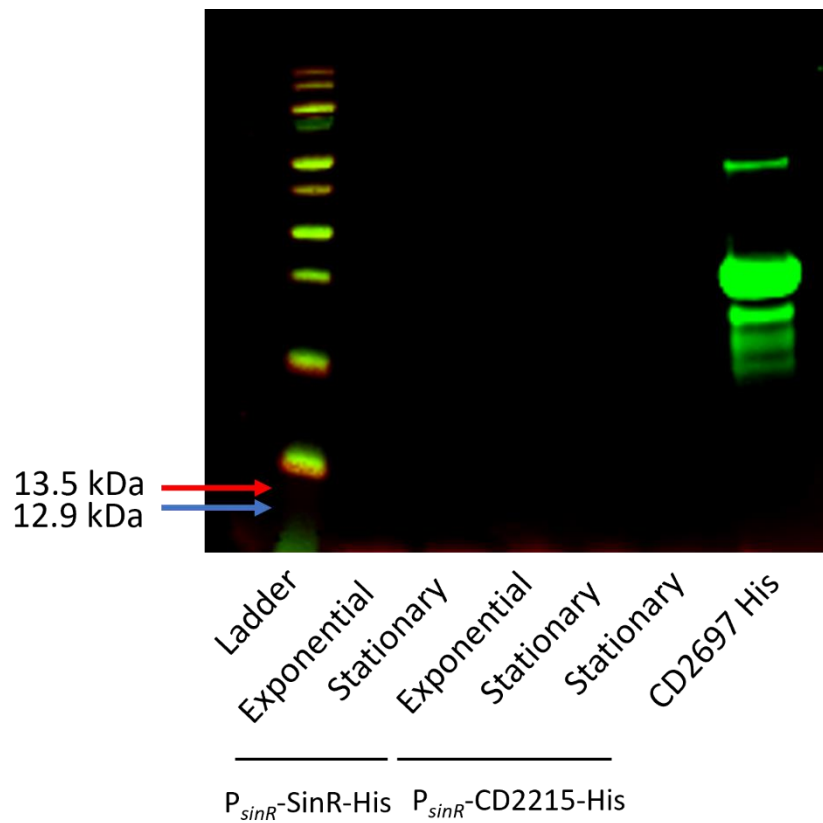


Figure 5.12. Western blot with anti-His for His-tagged SinR and CD2215. *630Δerm* was transformed with plasmids carrying the *sinR-CD2215* promoter and coding sequences with either *sinR* (P_{sinR} -SinR-His) or *CD2215* (P_{sinR} -CD2215-His) His-tagged. Culture samples were taken at exponential and stationary phases before cell lysis and protein purification. Protein samples were run on a 10% Bis-Tris SDS-PAGE before probing with anti-His antibody (green). Expected band size for SinR (13.5 kDa) marked with red arrow, expected band size for CD2215 (12.9 kDa) marked with blue arrow.

5.2.8 Effect of mutation of *sinR*-*CD2215* on motility and toxin production

RNA-seq data generated by Harrison *et al.* showed that in the *CD2215::CT* strain genes encoding the flagella, such as *fliC* and *flgB*, were expressed to a significantly higher level at the stationary phase of growth, whilst no changes were found in the *sinR-CD2215::CT* strain (243). To determine if the changes seen in the *CD2215::CT* strain translated to phenotypic changes, the knockout strains and their complements were grown on *C. difficile* minimal media 0.3% agarose and their motility measured 24 hours after inoculation (chapter 2.9.8). The *CD2215::CT* strain was found to have significantly greater motility than the wild type ($p=0.022$), a phenotype which was successfully complemented (Fig. 5.13). No significant changes were observed in either the double *sinR-CD2215* or single $\Delta sinR$ knockouts. This makes it likely that the increased expression of *sinR* observed in the *CD2215::CT* strain is involved in increased motility. This data is corroborated by the transcriptomics and qRT-PCR datasets presented by Harrison *et al.* (243) who showed that the drop in *sigD*, the key sigma factor that positively regulates flagella operon expression, observed from early to late exponential phases of growth in the wild type does not occur in the *CD2215::CT* strain and *sigD* expression is actually increased. Furthermore, in the double *sinR-CD2215::CT* strain at late exponential phase *sigD* is expressed at similar levels to the wild type. Therefore, this suggests that the *CD2215* indirectly represses motility by preventing the overexpression of *sinR*, which elevates expression of *sigD* (243).

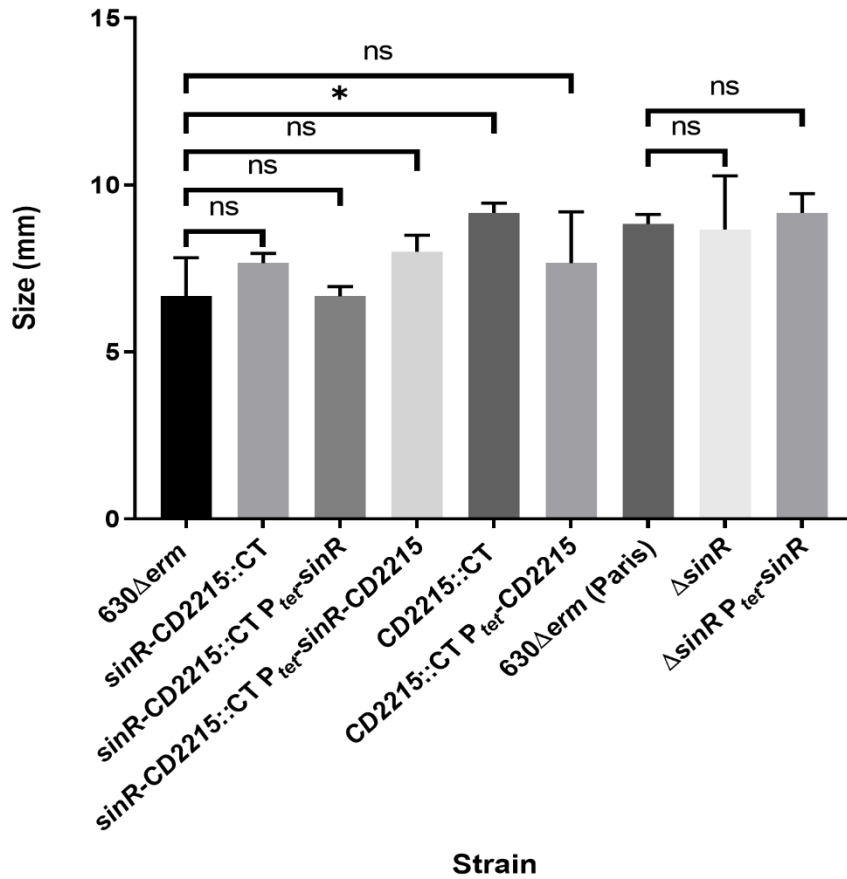


Figure 5.13. Swimming motility of *sinR* and *CD2215* knockout strains. Individual colonies of 630Δerm, 630Δerm *sinR*-CD2215::CT, 630Δerm *sinR*-CD2215::CT P_{tet-sinR}, 630Δerm *sinR*-CD2215::CT P_{tet-sinR}-CD2215, 630Δerm CD2215::CT, 630Δerm CD2215::CT P_{tet-CD2215}, 630Δerm (Paris), 630ΔermΔ*sinR* and 630ΔermΔ*sinR* P_{tet-sinR} were inoculated on to 0.3% minimal media agar plate supplemented with 2 ng/ml ATc and 15 μg/ml thiamphenicol where appropriate. Strains were grown for 24 hours and the diameter of growth was measured in two directions with the average taken. Linear regression was undertaken to assess differences in motility between strains. Data represents means and standard deviation of three biological replicates. ns not significant, * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

5.3 Discussion

The process of lifestyle switching in *C. difficile*, in particular the entry into sporulation, has been relatively well characterised, however, there is a scarcity of knowledge on regulatory networks that control these switches. In *B. subtilis*, often used as a model organism for *C. difficile*, a key feature of the control of lifestyle choice is the *sin* locus, formed of two genes: *sinR_{BS}* and *sinI_{BS}* (236, 245). SinR_{BS} has been found to repress sporulation during exponential growth, with the activities of SinR_{BS} inhibited later in the life cycle by SinI_{BS}, which binds directly to SinR_{BS} to form a heterodimer and prevents SinR_{BS} from assembling into its active form: a tetramer (191). Control of these genes occurs from three promoters (Fig. 5.1); expression of *sinR_{BS}* occurs constitutively from the P₃ promoter, with expression of both genes from the P₁ promoter being induced later in the lifecycle by phosphorylated Spo0A, and despite the two genes being co-transcribed, SinI_{BS} is expressed at a level 10-fold higher than SinR_{BS} (236). Expression of high levels of SinI_{BS} from the P₁ promoter results in the inhibition of sporulation carried out by SinR_{BS} being lifted (191). Whilst both genes are also co-transcribed from the P₂ promoter this does not occur until after the initiation of sporulation (236) and is therefore considered less relevant in lifestyle switching. In *C. difficile*, there is no ortholog to *sinI_{BS}* but there are two homologs of *sinR_{BS}* (*CD2214* (*sinR*) and *CD2215*) (194). That led to the hypothesis that these orthologs may also have roles in the control of lifestyle choice, and sporulation in particular, akin to that of the *sin* locus in *B. subtilis*.

Prior work by Girinathan *et al.* and by Harrison *et al.* (243) has shown that both SinR and CD2215 are pleiotropic regulators of a wide range of genes involved in virulence factors including genes involved in sporulation, motility and toxin production (194). In this chapter, it has been shown that both *sinR* and *CD2215* are controlled by a single -10 promoter site, which is in contrast to the *sin* locus found in *B. subtilis*, which has three promoters, with constitutive expression from P₃ leading to high level expression of SinR_{BS} and induction of P₁ by Spo0A leading to a high level of SinI_{BS} (240). In *C. difficile*, as both *sinR* and *CD2215* genes are transcribed together (243) it is currently unclear how the two proteins are expressed to different levels in order to exert their effects on gene regulation (243).

Transcriptomics analysis has shown that expression of these two genes is altered to differing levels by a number of regulators such as CcpA (126) and Spo0A (127) but the mechanism for this is currently unknown. It is possible that there is post-transcriptional processing of the co-transcript which leads to a higher concentration of either protein depending on the cell's needs. Evidence is emerging for a significant post-transcriptional regulatory network in *C. difficile* via RNA-chaperones such as Hfq, which has also been found to significantly up-regulate both *sinR* and *CD2215* (246). Alternatively, there may be another system involved in the function of these two proteins, for example, in *B. subtilis*, SlrR binds to SinR forming a complex of the two which acts to repress genes involved in motility and cell separation (193).

In this study it has been shown that expression of the operon is significantly enhanced by SinR itself which binds to an inverted repeat located 71 base pairs upstream of the start codon, with both arms of the inverted repeat essential for SinR binding. Autoregulation is present amongst other factors involved in lifestyle choice, with Spo0A also suggested to be autoregulatory (144). No significant effect was found for CD2215 in regulation of the *sinR-CD2215* operon by the SNAP-tag reporter in this work, however, as shown in chapter 3.1, the SNAP-tag lacks sensitivity, and despite multiple attempts, for unknown reasons the P_{sinR} -Long-*phoZ* reporter was not successfully conjugated into the *CD2215::CT* strain. Importantly however, transcriptomics and qRT-PCR data generated by Harrison *et al.* found that CD2215 represses expression of *sinR* (243). This provides further evidence for the SNAP-tag's inadequacies as a transcriptional reporter. The single *CD2215* mutant strain has been shown to be both more motile (Chapter 5.2.8 and by Harrison *et al.* (243)) and produces significantly more Toxin A and B (243) suggesting that in addition to inhibiting *sinR*, CD2215 may also help to make the switch to sporulation more energy efficient through repression of virulence factors not involved in sporulation. These results are corroborated by Girinathan *et al.* who found that mutation of *CD2215* led to increased motility and toxin production (194). Currently, it is unknown whether the observed phenotypic changes are as a result of direct repression by CD2215, or if it is indirect through CD2215 binding *sinR* preventing its autoregulatory activities, or other indirect effects through other regulators.

Spo0A is the master regulator of sporulation in *C. difficile* (127, 143, 144), therefore its interaction with the *sinR-CD2215* operon in *C. difficile* is key to lifestyle switching. This work has shown that in the exponential phase of growth Spo0A represses expression from the *sinR-CD2215* operon promoter. Repression of the operon by Spo0A is not observed at stationary phase, this is perhaps because SinR's enhancement of operon expression is inhibited by CD2215 and therefore any activity by Spo0A is not detectable as expression of the operon is already significantly lower at stationary phase and Spo0A does not have any additional effects on top of operon repression by CD2215. Dhungel and Govind also found that Spo0A represses expression from the *sinR-CD2215* operon through two binding sites upstream of the *sinR* promoter (196). These binding sites were identified using a biotinylated 340 base pair sequence of the upstream region of *sinR-CD2215* as bait for pulldown experiments followed by western blotting for Spo0A, however, it is unclear at which stage of the lifecycle this was found at as this is not well described by the authors (196). The finding of Spo0A repressing expression of *sinR* and *CD2215* is also corroborated by transcriptomics data from Pettit *et al.* where expression of both genes was increased in a *spo0A* mutant (127). As well as Spo0A modulating expression of the *sinR-CD2215* operon, the *sinR-CD2215* operon also has a regulatory effect on the expression of *spo0A*, as previous work by Harrison *et al.* found that Spo0A transcript, when tested by qRT-PCR, and Spo0A protein, tested using a Spo0A specific antibody, are significantly reduced in the *CD2215::CT* strain (243). In addition, in the same study, a hypersporulation phenotype was observed in the *sinR-CD2215::CT* strain and a reduced sporulation rate was seen in the *CD2215::CT* strain providing strong evidence that *sinR* is a repressor of sporulation, via repression of *spo0A*, and CD2215 is responsible for derepression of sporulation (243). This is in stark contrast to Girinathan *et al.* who reported in the R20291 strain that mutation of both *sinR* and *CD2215* led to an asporogenic phenotype, whilst a single mutation of just *CD2215* led to a hyper-sporulating phenotype, a phenotype which could not be complemented (194), and therefore this leads to the reasonable hypothesis that a non-specific genetic mutation may have led to the observed asporogenic phenotype.

CodY is a global regulator that is active in the presence of BCAAs (122) and GTP (123) and has been shown to repress sporulation (185). The data from the SNAP-tag reporter clearly demonstrates a drop in expression from the *sinR-CD2215* promoter in the $\Delta codY$ strain suggesting that CodY's negative effects on sporulation may at least be in part mediated by its induction of SinR. There is, however, contrasting data on CodY's effect on this operon. Nawrocki *et al.* found conflicting results between strains with their $630\Delta erm codY$ mutant strain having increased *sinR* transcription compared to its parent strain whilst their *codY* mutant in the UK1 strain background had *sinR* expression five-fold higher than its parent strain (185). Girinathan *et al.* found via EMSA that recombinant CodY binds to a 59 base pair sequence found within the upstream region of *sinR* and includes the putative CodY binding site, however, the binding was relatively weak in comparison to its binding of the upstream region of *tcdC* (194), with binding to *tcdC* upstream observed at 100 nM CodY whereas it was not observed until 400 nM with the *sinR* upstream sequence. Despite this the data demonstrated that CodY negatively regulates expression of *sinR* and *CD2215* (194). Additionally, via western blot analysis, the same group found that both SinR and CD2215 proteins were significantly higher in the UK1 CodY mutant (194). As well as utilisation of different strains, direct comparisons between the studies are further complicated by utilisation of different culture media. In this study BHIS was used, Girinathan *et al.* used tryptose yeast (194) and Nawrocki used 70:30 sporulation media (185). Given the different nutrients contained in each of these media it is unknown how active CodY is in each of the different conditions which may contribute to the varying results seen. Nawrocki. *et al.* investigated the putative CodY binding site found upstream of *sinR* by fusing the upstream promoter region of *sinR* to a *phoZ* reporter before mutating a single base pair within the CodY binding site (185). The authors found that this had no effect on expression suggesting the CodY binding site is not active, which complements the data presented here, however, it is unknown whether a single base pair mutation was sufficient to prevent CodY binding (185). In this chapter, the CodY binding site was fully removed and no difference in expression of the SNAP-tag was found suggesting that the binding site is not active and that CodY's effects on expression of *sinR* is likely indirect.

A third regulator which was tested for its effects on *sinR-CD2215* expression was CcpA. CcpA has been found to repress sporulation via the downregulation of genes involved in the early stages of sporulation such as *spo0A* and *sigF*, as well as having been shown to repress both *sinR* and *CD2215* (126). Similarly, to the results seen with the *spo0A::CT* strain in this study, initially, no significant differences were found when tested by the SNAP-tag reporter. However, as previously reported, the SNAP-tag can lack sensitivity so when the strain was re-tested using the P_{sinR} -Long-*phoZ* reporter a significant increase in expression was found at stationary phase in the $\Delta ccpA$ strain. In support of this finding, Antunes *et al.* found that both *sinR* and *CD2215* were significantly upregulated in a mutant of CcpA at the entry to stationary phase (126). One potential reason for CcpA's downregulation of the operon is that CcpA reduces expression of *CD2215* at stationary phase in order to repress *CD2215*'s derepression of sporulation, and therefore CcpA maintains its negative effects on sporulation. It has been shown previously that CcpA mediated repression of sporulation is long lasting with Antunes *et al.* finding that after 48 hours in sporulation media, a CcpA mutant still had higher sporulation rates than the wild type (126), so repression of *CD2215* maybe one such mechanism by which CcpA achieves this long lasting effect. *CD2215* also has a significant negative effect on CcpA expression suggesting an important interaction between these two regulators that may help to balance the drive to sporulation. Transcriptomics data generated by Harrison *et al.* showed that CcpA is not induced in the wild type strain in the transition from early to late exponential growth whereas it is in both the *sinR-CD2215::CT* and *CD2215::CT* strains indicating that under normal conditions *CD2215* downregulates CcpA expression (243). This means *CD2215* has a dual effect on derepressing sporulation genes, firstly, by repressing and binding SinR, and secondly, by downregulation of CcpA.

When combining all these data sets a model emerges of the regulation of *sinR* and *CD2215* and how this impacts sporulation (Fig. 5.14). In the exponential stages of growth SinR is strongly expressed, with expression driven by CodY and SinR itself, leading to repression of sporulation. SinR expression at this stage is partially repressed by Spo0A, perhaps to prevent complete repression of sporulation. This is followed by increased *CD2215* expression which represses CcpA, represses SinR expression and

binds to SinR preventing the formation of its active tetrameric form. This results in a significant drop in SinR expression and activity allowing Spo0A to drive the cells to sporulate. This chapter highlights the regulation of the *sinR-CD2215* operon, providing insight into the effects it has on lifestyle switching in *C. difficile*. Further characterisation of this operon and its regulon offers additional opportunities for greater understanding of the regulatory networks that control key *C. difficile* virulence factors.

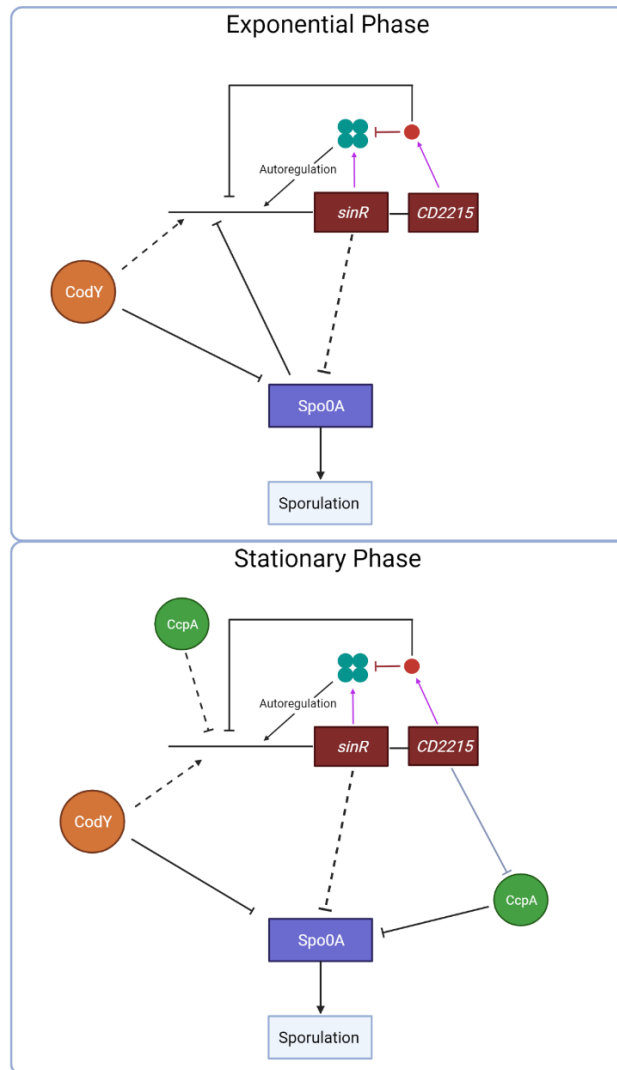


Fig 5.14. Summary of findings of regulatory factors involved in *sinR-CD2215* expression. Triangular arrowhead represents upregulatory activities, flat headed arrow represents downregulatory activities, solid line indicates direct regulation, dashed line indicates indirect regulation, pale blue line indicates if regulation is unknown to be direct or indirect, red line indicates direct protein:protein interaction and purple lines transcription and translation of the gene to form its protein product. Regulation of: *sinR-CD2215* by *CD2215*, *ccpA* by *CD2215* and *spo0A* by *sinR* shown by Harrison *et al.* (243). Direct protein:protein binding of SinR by *CD2215* shown by Ciftci *et al.* (195). Regulation of *spo0A* by *CcpA* shown by Antunes *et al.* (126). Regulation of *spo0A* by *CodY* shown by Nawrocki *et al.* (185). Figure created with BioRender.com

Chapter 6: Discussion

C. difficile infection causes significant morbidity and mortality, placing a heavy burden on healthcare systems globally, with the estimated annual cost in the USA alone being \$5.4 billion (25). The emergence of hypervirulent strains alongside the widespread use of antibiotics combined to cause a number of significant outbreaks in the early 2000s (9-11). Whilst the number of infections in the UK has significantly fallen from its peak, approximately 12,000 cases are still reported each year (5). With little prospect of a highly efficacious vaccine becoming available in the short term, improvements in treatment that reduce infection recurrence remain a priority. Treatments directed against *C. difficile* virulence factors offer opportunities to target the organism more specifically, limiting the effect on the beneficial microbiota and therefore promoting recolonisation of the gut by bacteria that provide colonisation resistance against *C. difficile*.

p-cresol production is an important virulence factor associated with infection relapse in a mouse model of CDI as a result of *p*-cresol selectively inhibiting Gram-negative species, specifically of the Gammaproteobacteria class (223). This study aimed to determine factors involved in expression of the HpdBCA decarboxylase responsible for the conversion of the precursor *p*-HPA to *p*-cresol, as well as to characterise the regulation of expression of the *hpdBCA* operon encoding the decarboxylase. In addition, this study also aimed to further investigate the role of SinR and CD2215 due to their homology to the major lifestyle choice regulator SinR found in *B. subtilis*, with a particular focus on factors that affect expression of these two regulators. Identification of regulatory controls of both *p*-cresol production and factors involved in *sinR-CD2215* expression will provide insight into two of the mechanisms by which *C. difficile* adapts to its environment in the gut as well as possibly identifying novel strategies for targeting key *C. difficile* virulence factors.

6.1 *P*-cresol production

6.1.1 Utilisation of exogenous *p*-HPA by *C. difficile* for *p*-cresol production

Initial investigations into *p*-cresol production in chapter three sought to determine environmental factors involved in the regulation of the *hpdBCA* operon. To address this, transcriptional reporters

were used which also allowed for a comparison of three separate reporter systems: SNAP-tag, *phoZ* and *gusA*. Each of these reporters showed that in minimal media with exogenous tyrosine or *p*-HPA that the presence of 2 mg/ml *p*-HPA led to a significant increase in expression from the promoter of the *hpdBCA* operon, whilst tyrosine had no significant effect on the operon's expression. These results correspond with previous studies which showed that *p*-HPA is required as a supplement to media for *p*-cresol to be readily detected (219, 223). Further investigation into the *hpdBCA* operon revealed that expression was *p*-HPA concentration dependent, showing that *C. difficile* is able to import *p*-HPA into the cell and respond proportionally to the imported *p*-HPA. *p*-HPA is produced by a number of bacterial species including *Klebsiella* and *Clostridium* species, as well as being produced by host cells with *p*-HPA being detectable in all tested human biological fluids and excreta including faeces (247). This suggests that there is a pool of *p*-HPA available in the gut for *C. difficile* to take up and convert to *p*-cresol, which is likely a more efficient process for *C. difficile* in comparison to tyrosine fermentation. The HPLC in chapter three showed that turnover of tyrosine to *p*-HPA and *p*-cresol represented just 11.8% of the available tyrosine. Furthermore, whilst tyrosine is utilised for Stickland fermentation by *C. difficile* it has been shown to be one of the least efficiently used amino acids suggesting that its metabolism is limited and therefore production of *p*-HPA would be low as a result (248). However, whilst tyrosine utilisation has been shown to be inefficient *in vitro*, this may not be the case *in vivo* so it is not possible to rule it out as an important source of *p*-cresol production without further study. Passmore *et al.* demonstrated in a mouse model of CDI relapse that following the induction of relapse a *p*-cresol null mutant was found at a significantly reduced load compared to the wild type. This disparity between the mutant and wild type strains implies that there is sufficient tyrosine, *p*-HPA or a combination of the two in the gut available for *C. difficile* to convert to *p*-cresol. A priority for future work in this area is the direct sampling from the gut over the course of CDI to determine what level of *p*-HPA and/or tyrosine is necessary to produce physiologically relevant quantities of *p*-cresol. Given the invasive nature of the sampling required for such a study this would need to be carried out in an

animal model, with the mouse model of CDI relapse first developed by Theriot *et al.* (249) and used by Passmore *et al.* (223) an excellent option for such a study.

6.1.2 Comparison of transcriptional reporters

Prior to this study little data had been published comparing transcriptional reporters commonly used in *C. difficile*. Using these three reporters under the control of an identical promoter and using an identical method for each provided an opportunity to carry out a direct comparison of these reporters. All three of the reporters were able to detect a significant increase in expression from the *hpdBCA* promoter in response to exogenous *p*-HPA. When comparing the methods of the reporters the *phoZ* and *gusA* reporters were significantly less time consuming and complex to use than the SNAP-tag. Furthermore, for both of these reporters the per single assay cost of the reporter substrates was approximately £0.20 compared to £10 for the SNAP-tag.

An important comparison of the reporters was of their sensitivity and to determine this the fold-change in *hpdBCA* promoter activity identified by the reporters was compared to the fold-change in *hpdC* expression observed when using the gold standard qRT-PCR. Following an initial comparison of fold-change in *hpdC* expression in 630 Δ *erm* grown in the presence of *p*-HPA in BHIS, as seen in chapter three, I prepared new samples of 630 Δ *erm* grown in minimal media with and without *p*-HPA following the same method as that used for the reporters and extracted RNA for qRT-PCR. Analysis of fold-change in response to *p*-HPA between qRT-PCR and the three reporters showed that the fold-change observed was not significantly different between the qRT-PCR and the *gusA* and *phoZ* reporters, however, all three found significantly higher fold-changes than that of the SNAP-tag. This demonstrates that *gusA* and *phoZ* reporters can perform with a similar sensitivity to the gold standard qRT-PCR whilst also being easier to use and offering significantly reduced costs. A limitation of the comparison of fold-changes is that the fold-change identified by the reporters is in promoter activity in response to environmental signals whilst qRT-PCR identifies the change in actual transcript level, therefore the two methods identify changes in two slightly different aspects of gene expression and as such this is a potential source of variation in the fold-changes identified between the two methods.

It should be noted that the SNAP-tag whilst performing significantly less well than the other reporters in terms of sensitivity, ease of method and cost, does offer the ability to carry out translational fusions observable by confocal microscopy to identify protein localisation or determination of whether protein expression is observable in a subset of cells or across the entire population, such as that seen in chapter four, something the other reporters cannot be used for.

6.1.3 Regulation of *hpdBCA* expression

Characterisation of the *hpdBCA* operon offers the potential for identification of pathways involved in the production of *p*-cresol which could be viable therapeutic targets. Here it was shown that following mutation of two putative -10 sites only one of these sites was functional. Furthermore, an inverted repeat was found upstream of the promoter and was identified as being the binding site for a transcriptional regulator that induced high level expression of the operon in response to *p*-HPA. Transcriptional regulators can act as activators and/or repressors of gene expression and function by binding to specific binding sites to modulate the binding of RNA polymerase to the gene's promoter (Reviewed by Balleza *et al.* (250)). Transcriptional regulators are often autoregulatory (232), and as such it was predicted that the regulator responsible for induction of *hpdBCA* expression would have had the identified binding site (AAAAAG-N₁₃-CTTTT) located upstream of their own coding region. Unfortunately, a bioinformatic search, carried out by Dr Mark Preston, identified only two genes with either the exact motif or a single base pair mismatch upstream of the start codon: CD3256, valyl-tRNA synthetase, and CD1951, a putative Acyl-CoA N-acyltransferase. Unfortunately, neither of these genes are good candidates for being the regulator involved in *hpdBCA* expression given their known or predicted functions. Therefore, future work should investigate carrying out further searches, with more flexibility in the motif, i.e., more than a single base pair mismatch and also a different number of base pairs between the two arms of the motif. Identification of this regulator will provide an alternative therapeutic target to inhibit *p*-cresol production. Another approach for identification of the regulator would be to use the upstream region of the *hpdBCA* promoter for pulldown experiments with the bound samples analysed by mass spectrometry. Based on these findings it suggests that

expression of the operon is a relatively simple system whereby expression is very low when no *p*-HPA is present but when *p*-HPA is available a regulatory factor binds to the inverted repeat inducing high level expression of the operon via the identified -10 site.

Identification of regulatory factors involved in the control of the *hpdBCA* operon can provide insight into other factors involved in the process of tyrosine or *p*-HPA uptake as well as regulatory control over the operon itself. A key regulator of a range of virulence genes, including sporulation and toxin genes, in *C. difficile* is CodY (180, 185). CodY is bound by BCAAs and GTP which act as co-factors allowing CodY to bind to DNA (122, 123). CodY mostly functions as a repressor of gene expression, with Dineen *et al.* showing that in a CodY mutant 146 genes were overexpressed by greater than four-fold compared to only 19 genes that were under-expressed by four-fold (180). Genes under the control of CodY are involved in a range of metabolism pathways including nutrient transport and fermentation pathways (180). Importantly, CodY has been found to repress genes, *CD0873-CD0878*, that form two neighbouring ABC transporter systems involved in tyrosine import into the 630 strain of *C. difficile* (180). Steglich *et al.* showed in R20291 that *CDR20291_805* and *_806*, which share 99.4 and 99.3% homology to *CD0876* and *CD0877* in 630 respectively, are involved in tyrosine import and their mutation leads to reduced *p*-cresol production due to reduced tyrosine uptake (218). Whilst in the neighbouring transporter, Bradshaw *et al.* hypothesised that their mutant of *CD0873* in 630 was less fit than the wild type in a mouse model of infection due to reduced *p*-cresol production (101). Therefore, it was hypothesised that CodY may play a role in the regulation of the *hpdBCA* operon given its role in regulation of virulence factors as well as in the repression of tyrosine import systems. Surprisingly, mutation of CodY led to a significantly reduced turnover of *p*-HPA to form *p*-cresol at both time points tested. Furthermore, when the P_{hpdB} -*phoZ* reporter was employed in the $\Delta codY$ strain a significant reduction in expression from the *hpdBCA* promoter was observed in the *codY* mutant. This drop in expression was not comparable to the drop identified upon removal of the inverted repeat found upstream of the operon or to conditions without exogenous *p*-HPA. Additionally, Dineen *et al.* did not identify a CodY binding site upstream of the operon by pulldown analysis (180). This strongly

suggests that CodY's effect is indirect and that it is not the regulatory factor responsible for induction of the *hpdBCA* operon expression in the presence of *p*-HPA. Over 350 binding sites for CodY have been identified, 37 of these sites are located near to regulatory genes (180), therefore, future experiments should investigate these regulators with a CodY binding site to see if they are differentially regulated in the presence of *p*-HPA and therefore could be the regulator directly responsible for *hpdBCA* induction. Alternatively, CodY may have a role in the control of factors responsible for the sensing or uptake of *p*-HPA and its mutation led to a reduced ability to sense or take up *p*-HPA therefore leading to the observed reduction in *hpdBCA* induction as this is *p*-HPA concentration dependent. A number of potential transporters have been identified by pulldown analysis by Dineen *et al.* with CodY found to bind to transporter proteins such as CD1505, an ABC transporter, and CD1737, a permease (180), and it is possible systems such as these may be involved in uptake of *p*-HPA. Unfortunately, neither of the ABC transporters investigated by Bradshaw *et al.* (101) or Steglich *et al.* (218) were tested for a role in *p*-HPA uptake and as such they should also be followed up to determine whether they are involved in *p*-HPA uptake. A combined transcriptomics and metabolomics approach comparing the $\Delta codY$ and *hpdBCA* knockout strains to their wild type grown in the presence of *p*-HPA may provide more insight into the role CodY has in responding to *p*-HPA and identify key factors in *p*-HPA utilisation and *p*-cresol production.

6.1.4 Inhibitory effects of *p*-HPA

After showing that *hpdBCA* expression is controlled by sensing and responding to *p*-HPA this led to the hypothesis that *C. difficile* may convert *p*-HPA to *p*-cresol due to potential inhibitory effects of *p*-HPA, in addition to the competitive advantage *C. difficile* gains by production of *p*-cresol. A key finding in this study was that *p*-HPA significantly inhibited *C. difficile* growth at concentrations ≥ 2 mg/ml. Furthermore, inactivation of *hpdC*, which renders *C. difficile* incapable of conversion of *p*-HPA to *p*-cresol, led to an enhanced inhibition of growth. This suggests that in addition to production of *p*-cresol another driving factor behind *p*-HPA turnover is the growth advantage that it provides to *C. difficile*. *p*-HPA was also found to inhibit the growth of Gram-negative gut commensal species significantly

more than Gram-positive species with two of the Gram-negative species significantly inhibited at 1 mg/ml and one at 2 mg/ml compared to the Gram-positives where only one species was inhibited at 1 mg/ml with the other two species inhibited at 3 and 4 mg/ml. These findings are similar to that of *p*-cresol which was found to be more deleterious to Gram-negative species than Gram-positive (223). It is notable that the concentrations of *p*-HPA found to be inhibitory to both *C. difficile* and gut commensal species are significantly higher than that of *p*-cresol. Passmore *et al.* found that 0.1% v/v (9.5 mM) *p*-cresol was almost completely inhibitory to the growth of the Gram-negative species and significantly inhibitory to *C. difficile* (223), whereas here, 0.1% w/v *p*-HPA (1 mg/ml, 13.3 mM) whilst being inhibitory still allowed relatively strong growth. This demonstrates that *p*-cresol is more toxic to these species than *p*-HPA. A further demonstration of the increased toxicity of *p*-cresol compared to *p*-HPA was observed in the sporulation assays. *p*-HPA was found to increase the sporulation rate of both the wild type and *hpdC::CT* strains, however, the positive correlation between sporulation rate increase and *p*-HPA concentration was stronger in the wild type ($R^2=0.9193$, $p=0.000012$) than the *hpdC::CT* strain ($R^2=0.8868$, $p=0.00006$). Furthermore, analysis of the total cell counts after 24 hours growth revealed that the reduction in total cell count observed at 1, 2 and 3 mg/ml *p*-HPA compared to the BHIS only control were significantly greater in the wild type strain (>90% reduction in each concentration) than the mutant strain (maximum reduction of 75.97% observed at 3 mg/ml), which cannot convert *p*-HPA to *p*-cresol. Therefore, whilst the growth curves clearly show that turnover of *p*-HPA is initially advantageous to *C. difficile*, the increasing concentration of *p*-cresol over time becomes far more detrimental to *C. difficile* than *p*-HPA alone, providing more evidence of the increased toxicity of *p*-cresol in comparison to *p*-HPA. These results show that *C. difficile* benefits in a competitive environment from the conversion of *p*-HPA to *p*-cresol due to the greater toxicity of *p*-cresol and its more significant inhibition of Gram-negative species, as well as showing that *p*-cresol leads to a higher sporulation rate which may help the bacteria in infection persistence, and therefore relapse, as well as infection transmission.

6.1.5 Mechanism of *p*-HPA toxicity

Elucidation of *p*-HPA's mechanism of toxicity was carried out using a combination of growth curve analysis and phosphate release assay. *p*-HPA is an acidic compound which when added to BHIS media lowers the pH by approximately 0.4 for every 1 mg/ml added. Lowering of environmental pH is well known to affect the growth of bacteria, including *C. difficile* (251), and for causing damage to bacterial cell membranes (reviewed by Lund *et al.* (252)). As such to determine if the mechanism of toxicity for *p*-HPA was simply driven by its acidity, growth curves were carried out at the pH equivalents to 0, 1, 2, 3 and 4 mg/ml *p*-HPA i.e., 7.0, 6.6, 6.2, 5.8 and 5.4. At pH 5.4, which is equivalent to the pH of media supplemented with 4 mg/ml *p*-HPA, a small growth inhibition was observed but at no other pH was growth significantly inhibited. Furthermore, following completion of the growth curves the media was sterilised and the pH tested, this showed that whilst *C. difficile* could buffer small changes to the pH (maximum pH change of 0.52 ± 0.08). These small changes to pH could not explain the differences in how *C. difficile* growth was less inhibited by the low pHs than in their matched *p*-HPA concentrations. When the same concentrations of *p*-HPA and their matched pHs were tested in gut commensal species a similar pattern was observed where the acidic pH could not fully account for the drop in growth in the matched *p*-HPA concentrations. This suggested that an alternative mechanism of *p*-HPA toxicity existed. When tested by phosphate release assay *p*-HPA caused significant disruption to the cell membrane of both *C. difficile* and *E. coli* well above that of the disruption seen in their matched pHs in the same assays. Taken together this shows that *p*-HPA's mechanism of toxicity is a combination of its acidic pH, which causes membrane disruption, as well as its ability to disrupt the cell membrane via another yet to be determined mechanism. There are a variety of different mechanisms by which *p*-HPA could affect cell membrane integrity such as through the formation of pores or affecting membrane fluidity and there are a number of microscopy techniques available for the study of membrane disruption including methods that can determine the size of pores formed in bacterial membranes (253) or membrane fluidity (254). Further work on this area could utilise some of the

aforementioned techniques to explore the mechanism of *p*-HPA toxicity with a focus on the effects it has on the cell membrane.

6.1.6 Cellular response to *p*-HPA

The induction of HpdB in response to *p*-HPA was found to be ubiquitous within a *C. difficile* culture, as when observed by confocal microscopy a translational fusion of HpdB to a SNAP-tag showed that all of the cells strongly expressed the fusion in response to *p*-HPA. Additionally, little to no fluorescence was observed without the presence of exogenous *p*-HPA in the growth media providing further evidence to support that from the reporters and qRT-PCR (see chapters three and four), that without the presence of *p*-HPA the *hpdBCA* operon is expressed to an extremely low level. The observation that the HpdB-SNAP-tag was homogeneously expressed by all the *C. difficile* cells shows that *p*-cresol production is not a heterogeneously expressed virulence factor such as pneumolysin, a pore forming toxin, in *S. pneumoniae* (255) or a type three secretion system in *Salmonella typhimurium*, which inject effector proteins into host epithelial cells leading to diarrhoeal disease (256). This indicates that *p*-HPA turnover to *p*-cresol is induced by each individual cell, leading to the hypotheses that: i) this conversion is essential for survival of the cell and possibly acts as a trigger for lifestyle switching to sporulation and ii) all cells within the *C. difficile* population are required to produce *p*-cresol in order to generate physiologically relevant quantities. When tested in the *hpdC::CT* strain the HpdB-SNAP fusion was found to aggregate within the cells whereas in the wild type strain the fusion was found throughout the individual bacterium. Inactivation of *hpdC* by Clostron has polar effects on *hpdA* so the microscopy suggests that *hpdC* and/or *hpdA* are required for the correct folding and transition of the full HpdB_{CA} decarboxylase through the cell. Therefore, future work using single deletion mutants of *hpdC* and *hpdA*, made by a technique such as allele exchange, combined with this confocal microscopy technique, may reveal if it is one or both of these subunits that are necessary for free movement of the decarboxylase.

6.1.7 Conservation of *hpdBCA* induction by *p*-HPA leads to high level *p*-cresol production across all five *C. difficile* clades

Previously it had been shown that the hypervirulent R20291 (RT027) strain produced more *p*-cresol than 630 following growth in yeast peptone media (220). In this work I sought to characterise strains representative of each of the five clades of *C. difficile* in terms of their response to *p*-HPA as well as their *p*-cresol production. Clade 1 strains include the historic 630 strain (RT012) and is formed of a diverse range of both toxigenic and non-toxigenic strains (257). Clade 2 strains include R20291, a RT027 strain, with RT027 strains remaining one of most prevalent global ribotypes (258-261). Clade 3 strains, including CD305 (RT023), are a recently emerged hypervirulent lineage which has been found to be prevalent in the UK (262). Clade 4 strains include M68 (RT017) and are prevalent across Asia (114-117), with a notable feature of this clade being that the strains are toxin A negative (114, 258). Clade 5 strains, including M120 (RT078), have recently caused significant outbreaks in the Netherlands and are one of the most common community acquired ribotypes in Europe (258, 263). In this work it was identified that all five clades of *C. difficile* respond to the presence of *p*-HPA with a strong induction of the *hpdBCA* operon leading to high levels of *p*-cresol production. No significant differences were observed between the fold-change of *hpdBCA* operon induction nor the final *p*-cresol concentration between any of the lineages in the presence of exogenous *p*-HPA. This demonstrates that the turnover of *p*-HPA is an extremely well conserved pathway within the *C. difficile* lineages. There are major differences between these lineages in other significant virulence factors including toxin production (264, 265) and sporulation (265), with an extreme example of variation between clades being that M120 is non-motile (266). Therefore, the identification that the induction of *hpdBCA* leads to high levels of *p*-cresol production across all of the clades highlights the importance of this pathway to *C. difficile*. Further evidence of how well conserved this pathway is, is found when looking at the cryptic clades of *C. difficile*. Recently, Knight *et al.* identified that *C. difficile* has three cryptic clades: C-I, C-II and C-III, which whilst being able to cause symptomatic disease fall below the threshold of sequence identity to be considered part of the same species as *C. difficile* (267). However, strains from all three cryptic clades carry *hpdBCA*-like operons with 91.1-95.0% homology to the *hpdBCA*

operon found in 630, this is despite the common ancestor of clades 1-5 and the cryptic clades being present over a million years ago (267). Importantly, the level of conservation of this pathway observed in clades 1-5 demonstrates that therapies that inhibit the HpdBCA decarboxylase would be effective against all *C. difficile* strains. Furthermore, as no significant differences were observed in *p*-cresol production or *hpdBCA* expression in response to *p*-HPA this also suggests that each strain has a similar ability to sense and take up *p*-HPA as well as release *p*-cresol. This further presses the case to identify the sensor and transport mechanisms of *p*-HPA as these will likely offer additional therapeutic targets for reduction of *p*-cresol production.

6.1.8 *p*-cresol production from tyrosine metabolism

Production of *p*-cresol from tyrosine is yet to be fully defined in *C. difficile* with the mechanisms by which tyrosine is converted to *p*-HPA unknown, and indeed to date, the mechanism of conversion of tyrosine to *p*-HPA has not been identified in any bacteria. A number of possible intermediates have been suggested including 4-hydroxyphenylpropionic acid and *para*-hydroxyphenylpyruvate due to their production in *Clostridium sporogenes* and *Clostridium botulinum* (224, 268), however a full pathway to *p*-HPA has yet to be elucidated in any species. In *E. coli*, the tyrosine lyase ThiH, involved in the thiamine synthesis pathway, has been shown to produce *p*-cresol directly from tyrosine by cleaving tyrosine to form dehydroglycine which releases *p*-cresol as a by-product (269). *C. difficile* carries an ortholog to ThiH however it only shares 36% identity to that found in *E. coli* and it is unknown if the ortholog is active in *C. difficile* (224). Whilst the pathway from tyrosine to *p*-HPA is unknown, previous data has shown that *p*-HPA generation from tyrosine varies between the clinical lineages with Dawson *et al.* finding that R20291 turns over more tyrosine to produce *p*-cresol than 630 in yeast peptone media (270). In this study strains representative of all five lineages were assessed for their ability to convert tyrosine to *p*-cresol. After both four and eight hours growth in minimal media, the M68 strain, representing clade 4, was found to generate significantly less *p*-HPA than all of the other lineages. M68 is a RT017 strain which are rarely found in Europe (261) and in the UK it is not amongst the top 10 most commonly identified ribotypes (18). Therefore, it is reasonable to hypothesise that

infection with this strain is less commonly observed, in part, due to a significantly inferior ability to produce *p*-cresol to levels necessary for influencing the composition of the microbiome. Moreover, CD305 and M120 which represent two of the most prevalent ribotypes identified in the UK (RT023 and RT078 respectively) (258), were also the two strains which produced the highest levels of *p*-cresol at both timepoints without exogenous *p*-HPA present. It is currently unknown how efficiently tyrosine is utilised by *C. difficile* *in vivo*, however, *in vitro* data from this study has shown it is inefficiently utilised and this is supported by Neumann-Schaal *et al.* who identified it as amongst the least efficiently used amino acids for Stickland fermentation (248). Therefore, it is possible that exogenous *p*-HPA may be the more important source of *p*-cresol production rather than tyrosine and, as described, strains from each clade were found to have a similar ability to utilise exogenous *p*-HPA. Additionally, whilst infections with RT017 strains are rare in the UK and Europe they are commonly identified in Asia (115-117). Moreover, there are significant differences in expression of other virulence factors, such as toxin production (264, 271) and motility (266), between these clades which may account for the differences in their ability to cause infection. *p*-cresol production is associated with infection relapse, however there is little data available on the relative risk of the ribotypes to cause infection relapse so it is not possible to make an association between *p*-cresol production ability in these ribotypes and infection relapse at this time. As discussed above, further work is required to determine the concentrations of tyrosine and exogenous *p*-HPA utilised by *C. difficile* *in vivo* to produce physiologically relevant concentrations of *p*-cresol in order to determine their relative importance to primary infection and infection relapse.

6.2 Regulation of *sinR-CD2215* expression

6.2.1 Characterisation of the *sinR-CD2215* promoter and the roles of SinR and CD2215 in *sinR-CD2215* expression

sinR-CD2215 has been shown to be a vital operon in the control of a range of key *C. difficile* virulence factors such as sporulation, toxin production and motility (194, 243). A major focus of research into the *sin* locus in *B. subtilis* and the *sinR-CD2215* operon in *C. difficile* has been their effects on lifestyle

switching especially with regards to their respective roles in sporulation. Prior to this study however, there had been little focus on factors that control expression of *sinR-CD2215*.

In *B. subtilis* the *sin* operon is formed of two genes: *sinR_{BS}* and *sinI_{BS}* (236). SinR_{BS} is a repressor of sporulation and SinI_{BS} functions as an antagonist to SinR_{BS} by binding it to form a heterodimer thus preventing SinR_{BS} from assembling into its active form; a tetramer (192). Expression of these two genes occurs from three promoter sites (236). In the initial stages of growth SinR_{BS} is constitutively expressed from the P₃ promoter, located between the two genes, in order to repress sporulation (236). Later in the lifecycle Spo0A is phosphorylated which allows it to form a dimer which is its active form (188). When Spo0A is phosphorylated to a low level it drives cells to biofilm formation, however, when Spo0A is phosphorylated to a high level it induces expression of the *sin* locus from the P₁ promoter located upstream of both genes and despite the two genes being co-transcribed SinI_{BS} is expressed to a level ten-fold higher than that of SinR_{BS}, and as a result SinI_{BS} is able to inhibit SinR_{BS} and the cell enters into sporulation (236, 239, 240). A third promoter site, P₂, is also functional however expression is only detectable from this promoter after the cell has initiated sporulation and therefore it is considered less relevant to lifestyle choice (236). *C. difficile* has two orthologs to *sinR_{BS}*: *sinR* and *CD2215* (194). In chapter five, I showed that in *C. difficile* expression of *sinR-CD2215* takes place from a single -10 promoter site suggesting that it is expressed as an operon, this was confirmed by Harrison *et al.* with qRT-PCR using primers to form a product which encompasses both *sinR* and *CD2215* (243), and therefore this is in agreement with Girinathan *et al.* who provided evidence the two genes are expressed as an operon as well (194). Expression of the two genes is in direct contrast to expression to the *sin* locus in *B. subtilis* which, as described, takes place from three separate promoters. This raises questions about how the two genes are differentially expressed in *C. difficile*, as if both are expressed at a constant ratio then this may not allow either to exert effects on their respective regulons efficiently. This makes it possible that there is post-transcriptional processing of the mRNA, possibly by RNA-chaperone proteins such as Hfq, which has been found to positively regulate both *sinR* and *CD2215* (246), however, the exact mechanism for *sinR-CD2215* co-transcript processing has

yet to be elucidated. Quantification of the SinR and CD2215 proteins was attempted in this work but unfortunately the addition of a His-tag on to the proteins encoded on a plasmid could not be detected. The addition of a His-tag on to the genes on the chromosome may be a successful alternative approach or, ideally, the use of SinR and CD2215 specific antibodies would likely be an effective method. An alternative strategy for the control of SinR or CD2215 activity may be a system similar to that in *B. subtilis* whereby SinR is bound by a third factor which leads to regulation of a different subset of genes. In *B. subtilis* this factor is SlrR and binding of SinR by SlrR leads to downregulation of genes involved in motility and cell separation (193). Therefore, investigation into whether other factors bind SinR in *C. difficile* should be a priority as this would be a significant factor in the regulation of genes related to lifestyle choice in *C. difficile*.

Through the use of a combination of the SNAP-tag reporter and EMSA binding assays it is clear that SinR enhances expression of the *sinR-CD2215* operon. This is in contrast to SinR_{BS} in *B. subtilis* which largely acts as a repressor of gene expression (189, 237, 238, 245) and, to date, there is no evidence to suggest that it is autoregulatory. The identification of the SinR binding site using both of the aforementioned assays is an important finding as by undertaking degenerate mutagenesis it will be possible to identify the full binding motif sequence for SinR. Using the binding motif bioinformatic searches combined with the transcriptomics data would allow for the identification of genes which are either directly or indirectly regulated by SinR. Additionally, analysis by mutation of individual nucleotides within the inverted repeat as well as mutation of specific amino acids within the HTH domain of SinR would provide useful data on the nature of the interaction between SinR and DNA. If it is possible to generate specific antibodies to SinR and CD2215 then ChIP-seq would be an excellent option to identify genes that are part of either the direct or indirect regulons for each of these proteins. In addition, this technique would allow for identification of the SinR and CD2215 binding consensus sequences, which could be confirmed using DNase I footprinting or EMSA analysis. Similarly to SinR, mutation of individual nucleotides to determine how CD2215 interacts with DNA would be valuable information. A better appreciation of the respective regulons of these genes would provide

a significantly better understanding of the regulatory network that controls lifestyle choice of the cells and perhaps offer therapeutic targets to manipulate *C. difficile* cells into a lifestyle choice that is not conducive to prolonged infection or infection relapse.

Data presented by Harrison *et al.* showed, using transcriptomics and qRT-PCR, that inactivation of CD2215 led to a significant increase in SinR expression (243). This derepression effect was not observed using the SNAP-tag reporter in the *CD2215::CT* strain most likely due to the insensitivity of the SNAP-tag. Unfortunately, for unknown reasons the P_{sinR} -*phoZ* reporter was not successfully conjugated in to the *CD2215::CT* strain despite repeated conjugations including the utilisation of heat shock to increase conjugation efficiency as described by Kirk *et al.* (226). In spite of this, the transcriptomics and the qRT-PCR provide strong evidence that mutation of *CD2215* significantly derepresses SinR. The mechanism for how this occurs is unclear as from this data it is not possible to determine if this is due to direct repression of the operon by CD2215 or due to CD2215 being unavailable to bind SinR which then allows SinR to upregulate its own expression more efficiently, or a combination of these two mechanisms.

6.2.2 Roles of Spo0A, CodY and CcpA in *sinR-CD2215* expression

In addition to SinR's autoregulatory activities, Spo0A, CodY and CcpA were all shown to have regulatory roles in expression of *sinR-CD2215*. Recent data by Dhungel and Govind demonstrated the presence of two Spo0A binding sites used by Spo0A to repress expression of *sinR-CD2215*, however, it was unclear at what growth phase this was observed (196). Data presented in this study corroborates that from Dhungel and Govind (196) as using the P_{sinR} -*phoZ* reporter I was able to show that expression from the *sinR-CD2215* promoter significantly increases when in a *spo0A::CT* background in comparison to the wild type at the exponential phase of growth. This shows both similarities and differences to the role of Spo0A in *sin* locus expression in *B. subtilis*. In contrast to *B. subtilis* where Spo0A induces expression of *sinI_{BS}*, in *C. difficile* Spo0A represses expression of the operon. However, in *B. subtilis* and *C. difficile* Spo0A's respective effects on *sinIR* or *sinR-CD2215* expression are both growth phase dependent, in *C. difficile* Spo0A represses expression at exponential

phase, possibly to ensure SinR is not overexpressed which could lead to complete repression of sporulation as well as overexpression of other virulence factors such as the toxins, whilst in *B. subtilis* expression of *sinI_{BS}* is induced by Spo0A later in growth to ensure the transition to sporulation takes place (236). Therefore, in both bacteria Spo0A's effects on *sinIR* or *sinR-CD2215* expression acts to drive sporulation but through different mechanisms of action.

Using the P_{*sinR*}-long-SNAP transcriptional reporter it was found that in the CodY knockout strain there is a clear and significant drop in expression from the *sinR-CD2215* promoter when compared to the wild type strain at both exponential and stationary phases of growth. This was followed by the finding that removal of the CodY binding site upstream of *sinR* led to no significant differences in expression from the SNAP-tag reporters. This would suggest that whilst CodY is an inducer of *sinR-CD2215* expression this effect is indirect or at the very least not mediated via the putative CodY binding site identified by Dineen *et al.* (180). The finding regarding the CodY binding site is in agreement with Nawrocki *et al.* who found that a single base pair alteration of the binding site had no effect on expression from the *sinR-CD2215* upstream region (185). However, Nawrocki *et al.*'s findings in the 630Δ*erm* strain are in direct contrast to those presented here, as the authors of that study presented data showing that a CodY mutant of the 630Δ*erm* strain showed significantly increased expression of *sinR-CD2215* in comparison to its wild-type parent strain (185). Conversely, in a CodY mutant of the UK1 strain Nawrocki *et al.* found significantly lower expression of *sinR* compared to its parent strain (185). Whereas in direct contrast, Girinathan *et al.* found that in a CodY knockout of the UK1 strain that expression of both *sinR* and CD2215 was significantly increased (194). In the same study, Girinathan *et al.* also found via EMSA that recombinant CodY binds a 59 base pair sequence of the upstream region of *sinR* which includes the putative CodY binding site, although the binding was relatively weak (194). In both Nawrocki *et al.* (185) and Girinathan *et al.* (194) the mutants of *codY* were generated using insertional inactivation which can lead to polar effects on downstream gene expression whereas in this study the mutant was generated using allele exchange which avoids these effects, therefore, it is possible this played a role in the different findings between the studies and

strains. As discussed in chapter five, each of the different studies used a different media for growth of their strains further complicating how active CodY may have been due to differences in the available concentrations of CodY co-factors GTP and BCAAs (122, 123). A key thing to note is that CodY is a known repressor of sporulation, therefore the data presented in this study corresponds with this as SinR is a repressor of sporulation and therefore it follows that CodY is an activator of *sinR-CD2215* expression. In order to determine if the effect of CodY on *sinR-CD2215* is strain and/or media dependent, further experiments using each of the different strains carried out in identical media and sampled at the same stages of growth with measurements of CodY expression taken would provide valuable data.

CcpA is a regulator that mediates *C. difficile*'s response to glucose, expression of 18% of *C. difficile*'s genes are regulated by the availability of glucose, with approximately half of this 18% being regulated by CcpA (126). In addition to the changes made in expression to pathways involved in metabolism, CcpA is also a repressor of sporulation via downregulation of genes such as *spo0A* and *sigF* (126) as well as having regulatory effects on other virulence factors such as toxin production (125). Here I demonstrate that CcpA is a repressor of expression from the promoter of *sinR-CD2215* at the stationary phase of growth. This corroborates evidence from Antunes *et al.*, who found that CcpA downregulates expression of *sinR* and *CD2215* (126). A possible explanation for CcpA repressing expression of *sinR-CD2215* at stationary phase but not exponential phase is to repress *CD2215*'s derepression of sporulation and therefore CcpA maintains its downregulation of sporulation. This is consistent with Antunes *et al.* who found that CcpA's repression of sporulation is long lasting, with a *ccpA* mutant showing higher sporulation even after 48 hours in sporulation media (126). Interestingly, SinR is the only one of 30 regulators regulated by CcpA independently of glucose (126), which suggests that regulation of SinR by CcpA occurs indirectly. Further evidence for indirect regulation of *sinR-CD2215* by CcpA is demonstrated by the lack of a CcpA binding site directly upstream of *sinR* (126). Transcriptomics data generated by Harrison *et al.* showed that in the absence of *CD2215*, CcpA's expression is induced in the transition from the exponential phase to the stationary phase of growth

suggesting that CD2215 normally represses *ccpA* induction (243). Thus, in addition to inhibition of SinR, CD2215 further derepresses sporulation by repressing *ccpA*, a known repressor of sporulation. Therefore, these two results show that CcpA and CD2215 have an important interaction as they both act as repressors of one another, perhaps to prevent either under or over sporulation phenotypes.

6.3 Final conclusions

The aim of this work was to characterise the regulation of both virulence factors themselves as well as key regulatory factors which in turn influence virulence factor expression. The identification of the key factor of *hpdBCA* expression being *p*-HPA itself as well as the utilisation of exogenous *p*-HPA by *C. difficile* has opened up new lines of investigation for identification of therapeutic targets. The use of inhibitors to decrease *p*-cresol turnover by inhibition of HpdBCA is a viable strategy that would be effective against all lineages of *C. difficile* and may help to prevent CDI relapse. This work has provided further evidence that *C. difficile* is able to take up exogenous *p*-HPA for *p*-cresol production and as such identification of the *p*-HPA import system could provide an alternative treatment target should it not be possible to effectively inhibit HpdBCA decarboxylase. A third strategy identified by this work results from discovery of the inverted repeat necessary for *hpdBCA* induction which has demonstrated that there is a regulatory system that may be targeted in order to inhibit *p*-cresol production. Similarly, identification of key factors that control expression of *sinR-CD2215* and the factors they control may allow for precise manipulation of *C. difficile* lifestyle choice. This would allow for approaches that would push *C. difficile* to a lifestyle that is the least viable for infection relapse or transmission, such as driving cells away from sporulation pathways and towards vegetative growth where cells are susceptible to antimicrobial therapy.

6.4 Future work

6.4.1 Further research into *p*-cresol production by *C. difficile*

6.4.1.1 Characterisation of *C. difficile*'s response to *p*-HPA

In addition to the induction of *hpdBCA* in response to *p*-HPA a number of other systems are required for the sensing, import and utilisation of exogenous *p*-HPA. Two key areas for further investigation

are: i) transcriptional regulators that are upregulated in response to *p*-HPA and ii) transporter systems that may be involved in the uptake of *p*-HPA from the environment.

- i. Identification of transcriptional regulators
 - a. The use of RNA-seq data generated from samples grown in the presence and absence of *p*-HPA would identify regulators that are upregulated in response to *p*-HPA. Regulators significantly upregulated in the presence of *p*-HPA could be investigated by mutagenesis to see if inactivation of these regulators led to *C. difficile* being unable to respond to *p*-HPA by induction of *hpdBCA*. The mutants could be tested for induction of *hpdBCA* using transcriptional reporters, such as those described in chapter three, or through the use of qRT-PCR for expression of any of the *hpdBCA* genes.
 - b. An additional strategy for identification of the transcriptional regulator responsible for *hpdBCA* induction would be to use the promoter region of *hpdBCA* as bait for pulldown experiments. The upstream region of *hpdBCA* used could be generated with biotinylated primers for binding to streptavidin magnetic beads in order to specifically capture the *hpdBCA* upstream DNA sequence. Preparation of cell free extracts from *C. difficile* cultures grown in the presence and absence of *p*-HPA could be tested against the upstream region of *hpdBCA* to determine any proteins that bind specifically in the presence of *p*-HPA. Analysis of the captured DNA would be carried out by mass spectrometry to identify the protein(s) that are bound.
- ii. Identification of *p*-HPA transport systems could be assessed with three strategies:
 - a. Similarly to the identification of transcriptional reporters, RNA-seq with samples generated in the presence and absence of *p*-HPA could provide target genes for mutagenesis to investigate their role in *p*-HPA uptake. Mutants of the transporters would be followed up by testing their ability to utilise *p*-HPA for conversion to *p*-cresol

using HPLC. If *p*-HPA utilisation and *p*-cresol generation were decreased in the mutant strains this would suggest they play a role in *p*-HPA uptake.

- b. Two neighbouring ABC transporters (*CD0873-CD0878*) have been identified as tyrosine transporters by Bradshaw *et al.* (272) and Steglich *et al.* (218). Acquisition of the mutant strains used in these studies would allow for assessment of the role of these transporters in *p*-HPA uptake. Generation of double mutants of both of these transporters would allow for determination of any redundancy in *p*-HPA uptake carried out by these transporters. Assessment of *hpdBCA* operon induction could be used as a marker for *p*-HPA uptake, as induction of *hpdBCA* is *p*-HPA concentration dependent (chapter three, fig. 5A) therefore any decrease in induction would suggest reduced *p*-HPA uptake. Measurement of *hpdBCA* induction could be achieved using either the transcriptional reporters generated in chapter three or using qRT-PCR.
- c. An alternative strategy would be growth of the transporter mutants and their wild type in the presence of exogenous *p*-HPA and then use of HPLC to determine if *p*-HPA utilisation or *p*-cresol production was reduced compared to the wild type strain. If a reduction in either *p*-HPA utilisation or *p*-cresol generation was identified then this would indicate the cells were less able to import *p*-HPA.

6.4.1.2 Sampling of the gut for tyrosine, *p*-HPA and *p*-cresol concentrations

To build on previous work using faecal samples, sampling of tyrosine, *p*-HPA and *p*-cresol concentrations directly from the gut would confirm the available concentrations of these compounds *in vivo*. As discussed above, utilisation of the mouse model of CDI relapse developed by Theriot *et al.* (249) and used by Passmore *et al.* (223) would be an excellent option to determine how much tyrosine and *p*-HPA are available to *C. difficile* as well as how much *p*-cresol is produced throughout the stages of infection from initial colonisation to post-infection relapse. This would help to determine the relative importance of tyrosine and *p*-HPA to *p*-cresol production which would be useful knowledge for the development of inhibitors of *p*-cresol generation.

6.4.1.3 Determination of pathway from tyrosine to *p*-HPA

A key part of *p*-cresol production that has yet to be determined is the conversion of tyrosine to *p*-HPA.

One strategy to determine the pathway would be to use ¹³C labelled tyrosine combined with NMR to determine the intermediates between tyrosine and *p*-HPA. With the knowledge of the intermediates, it would be significantly easier to define the pathways used by *C. difficile* to produce these intermediates and then target them to reduce *p*-cresol production.

6.4.1.4 Investigating a correlation between *p*-cresol production and ribotype prevalence

To build on the work in chapter four, which looked at *p*-cresol production across strains representative of the five clades of *C. difficile*, investigation into *p*-cresol production from both tyrosine and exogenous *p*-HPA utilisation across *C. difficile* ribotypes with varying prevalence would help to determine if differences in higher *p*-cresol production may play a role in the ribotypes that are more prevalent. These experiments could be carried out as per chapter four, i.e. growth of *C. difficile* strains in defined media with and without *p*-HPA with analysis of *p*-HPA and *p*-cresol concentrations by HPLC.

6.4.1.5 *p*-cresol production by *C. difficile* cryptic clades and other species with orthologs of *hpdBCA*

As shown in chapter four, the three cryptic clades of *C. difficile* (C-I, C-II and C-III) all carry *hpdBCA*-like operons which share greater than 90% homology to the *hpdBCA* operon carried by 630. Furthermore, a number of other species (*R. lituseburensis*, *O. uli* and *B. hydrogenotrophica*) identified by Saito *et al.* were found to produce *p*-cresol and also carry orthologs of the *hpdBCA* operon (224). In Saito *et al.*, production of *p*-cresol was not determined from utilisation of exogenous *p*-HPA (224), therefore, I suggest growing these strains in the presence of *p*-HPA and then using HPLC to determine if they too are capable of utilisation of exogenous *p*-HPA for the production of high concentrations of *p*-cresol. This would give insight into whether other species are capable of *p*-HPA utilisation in the same manner as clades 1-5 of *C. difficile* and if they can produce *p*-cresol to the same concentrations. The cryptic clades of *C. difficile* have been shown to be capable of causing symptomatic infection (267), so as with clades 1-5 these strains could be candidates for treatment by strategies to reduce *p*-cresol production.

6.4.1.6 *p*-HPA and *p*-cresol mechanism of toxicity

Previous work by Passmore *et al.* (223) showed that *p*-cresol disrupts cell membranes, whilst work described in this study shows that *p*-HPA also causes cell membrane perturbations (chapter four). There are a number of microscopy techniques available to explore the effects of *p*-HPA on the cell membrane. These techniques include being able to determine: 1) if pores are formed by *p*-cresol or *p*-HPA and if so the size of the pores (253) and 2) if *p*-cresol or *p*-HPA has effects on membrane fluidity (254). I would recommend using these techniques as elucidation of the mechanisms by which these two compounds cause toxicity would provide insight into the commensal species that would be more tolerant to these compounds. In the case of *p*-cresol this would be particularly useful as species identified as particularly *p*-cresol tolerant could help to determine the bacteria utilised in either probiotic or FMT treatments to give colonisation resistance to *C. difficile*.

6.4.2. Further characterisation of *sinR* and *CD2215* and factors involved in their expression

6.4.2.1 Characterisation of SinR and CD2215 binding sites

The generation of specific antibodies for the detection of SinR and CD2215 would allow for ChIP-seq to be used to identify the genes directly regulated by these two proteins. In addition, ChIP-seq would allow for identification of the SinR and CD2215 binding sites consensus' and distributions, which could be followed up by using DNase I footprinting or EMSA to find the exact sites for binding. Furthermore, mutation of individual nucleotides in SinR and CD2215's respective binding sites to determine the essential nucleotides would provide valuable information about how the regulators interact with DNA. An improved understanding of the genes regulated by SinR and CD2215 would offer further insight into the regulatory networks responsible for the controlling lifestyle choice and exploitation of these networks may allow for treatments that push *C. difficile* to a lifestyle choice that prevents sporulation and therefore may reduce the opportunities for infection relapse or transmission.

6.4.2.2 Further characterisation of the roles of CodY and CcpA in *sinR-CD2215* expression

As described in chapter five, the role of CodY in *sinR-CD2215* expression has been found to vary significantly between the different strains tested, however it is unknown if this could be due to

variations in the strains themselves or the different conditions used between studies. Therefore, carrying out experiments using these strains in identical media with sampling of *sinR-CD2215* expression, by use of transcriptional reporter or qRT-PCR, at different phases of growth alongside determination of CodY expression either by qRT-PCR or western blot would help to explain the differences in CodY's role in *sinR-CD2215* expression identified by this study and by others between the tested strains (185, 194).

It is likely that the effects on expression of *sinR-CD2215* for both of these regulators is indirect therefore investigation of regulators under the control of CodY and CcpA would help to elucidate how these two regulators are involved in *sinR-CD2215* expression as well as to further characterise the complex network of regulators involved in *C. difficile* lifestyle choice. This could be achieved by generation of mutants of regulators under CodY or CcpA's control with their effect on expression of *sinR-CD2215* investigated using the transcriptional reporters used in chapter five or by qRT-PCR.

7. References

1. Lawson PA, Rainey FA. Proposal to restrict the genus *Clostridium Prazmowski* to *Clostridium butyricum* and related species. *International Journal of Systematic and Evolutionary Microbiology*. 2016;66(2):1009-16.
2. Bartlett JG, Gerding DN. Clinical recognition and diagnosis of *Clostridium difficile* infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2008;46 Suppl 1:S12-8.
3. Rousseau C, Poilane I, De Pontual L, Maherault A-C, Le Monnier A, Collignon A. *Clostridium difficile* Carriage in Healthy Infants in the Community: A Potential Reservoir for Pathogenic Strains. *Clinical Infectious Diseases*. 2012;55(9):1209-15.
4. Vardakas KZ, Polyzos KA, Patouni K, Rafailidis PI, Samonis G, Falagas ME. Treatment failure and recurrence of *Clostridium difficile* infection following treatment with vancomycin or metronidazole: a systematic review of the evidence. *International journal of antimicrobial agents*. 2012;40(1):1-8.
5. PHE. Annual epidemiological commentary: Gram-negative bacteraemia, MRSA bacteraemia, MSSA bacteraemia and *C. difficile* infections, up to and including financial year April 2019 to March 2020. 2020.
6. Desai K, Gupta SB, Dubberke ER, Prabhu VS, Browne C, Mast TC. Epidemiological and economic burden of *Clostridium difficile* in the United States: estimates from a modeling approach. *BMC Infectious Diseases*. 2016;16:303.
7. George WL, Sutter VL, Goldstein EJ, Ludwig SL, Finegold SM. Aetiology of antimicrobial-agent-associated colitis. *Lancet (London, England)*. 1978;1(8068):802-3.
8. Bartlett JG, Moon N, Chang TW, Taylor N, Onderdonk AB. Role of *Clostridium difficile* in antibiotic-associated pseudomembranous colitis. *Gastroenterology*. 1978;75(5):778-82.
9. Loo VG, Poirier L, Miller MA, Oughton M, Libman MD, Michaud S, et al. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *The New England journal of medicine*. 2005;353(23):2442-9.
10. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, et al. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet (London, England)*. 2005;366(9491):1079-84.
11. Kuijper EJ, Barbut F, Brazier JS, Kleinkauf N, Eckmanns T, Lambert ML, et al. Update of *Clostridium difficile* infection due to PCR ribotype 027 in Europe, 2008. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*. 2008;13(31).
12. Drudy D, Kyne L, O'Mahony R, Fanning S. gyrA mutations in fluoroquinolone-resistant *Clostridium difficile* PCR-027. *Emerging infectious diseases*. 2007;13(3):504-5.
13. McDonald LC, Killgore GE, Thompson A, Owens RC, Jr., Kazakova SV, Sambol SP, et al. An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N Engl J Med*. 2005;353(23):2433-41.
14. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, et al. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat Genet*. 2013;45(1):109-13.
15. Burns K, Morris-Downes M, Fawley WN, Smyth E, Wilcox MH, Fitzpatrick F. Infection due to *C. difficile* ribotype 078: first report of cases in the Republic of Ireland. *The Journal of hospital infection*. 2010;75(4):287-91.
16. Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, et al. Emergence of *Clostridium difficile* Infection Due to a New Hypervirulent Strain, Polymerase Chain Reaction Ribotype 078. *Clinical Infectious Diseases*. 2008;47(9):1162-70.

17. Vitucci JC, Pulse M, Tabor-Simecka L, Simecka J. Epidemic ribotypes of *Clostridium* (now *Clostridioides*) *difficile* are likely to be more virulent than non-epidemic ribotypes in animal models. *BMC microbiology*. 2020;20(1):27.
18. Wilcox MH, Shetty N, Fawley WN, Shemko M, Coen P, Birtles A, et al. Changing epidemiology of *Clostridium difficile* infection following the introduction of a national ribotyping-based surveillance scheme in England. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2012;55(8):1056-63.
19. CDC. Severe *Clostridium difficile*-associated disease in populations previously at low risk-four states, 2005. *MMWR Morbidity and mortality weekly report*. 2005;54(47):1201-5.
20. Terhes G, Urban E, Soki J, Hamid KA, Nagy E. Community-acquired *Clostridium difficile* diarrhea caused by binary toxin, toxin A, and toxin B gene-positive isolates in Hungary. *Journal of clinical microbiology*. 2004;42(9):4316-8.
21. Wilcox MH, Mooney L, Bendall R, Settle CD, Fawley WN. A case-control study of community-associated *Clostridium difficile* infection. *Journal of Antimicrobial Chemotherapy*. 2008;62(2):388-96.
22. Khanna S, Pardi DS, Aronson SL, Kammer PP, Orenstein R, St Sauver JL, et al. The epidemiology of community-acquired *Clostridium difficile* infection: a population-based study. *The American journal of gastroenterology*. 2012;107(1):89-95.
23. Lambert PJ, Dyck M, Thompson LH, Hammond GW. Population-based surveillance of *Clostridium difficile* infection in Manitoba, Canada, by using interim surveillance definitions. *Infection control and hospital epidemiology*. 2009;30(10):945-51.
24. Zanichelli V, Garenc C, Villeneuve J, Moisan D, Frenette C, Loo V, et al. Increased Community-Associated *Clostridioides difficile* Infections in Quebec, Canada, 2008-2015(1). *Emerging infectious diseases*. 2020;26(6):1291-4.
25. Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, et al. Burden of *Clostridium difficile* infection in the United States. *The New England journal of medicine*. 2015;372(9):825-34.
26. Chitnis AS, Holzbauer SM, Belflower RM, Winston LG, Bamberg WM, Lyons C, et al. Epidemiology of community-associated *Clostridium difficile* infection, 2009 through 2011. *JAMA internal medicine*. 2013;173(14):1359-67.
27. Dumford DM, 3rd, Nerandzic MM, Eckstein BC, Donskey CJ. What is on that keyboard? Detecting hidden environmental reservoirs of *Clostridium difficile* during an outbreak associated with North American pulsed-field gel electrophoresis type 1 strains. *American journal of infection control*. 2009;37(1):15-9.
28. Pepin J, Saheb N, Coulombe MA, Alary ME, Corriveau MP, Authier S, et al. Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-associated diarrhea: a cohort study during an epidemic in Quebec. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2005;41(9):1254-60.
29. Bignardi GE. Risk factors for *Clostridium difficile* infection. *The Journal of hospital infection*. 1998;40(1):1-15.
30. Brown KA, Khanafer N, Daneman N, Fisman DN. Meta-analysis of antibiotics and the risk of community-associated *Clostridium difficile* infection. *Antimicrobial agents and chemotherapy*. 2013;57(5):2326-32.
31. Deshpande A, Pasupuleti V, Thota P, Pant C, Rolston DD, Sferra TJ, et al. Community-associated *Clostridium difficile* infection and antibiotics: a meta-analysis. *The Journal of antimicrobial chemotherapy*. 2013;68(9):1951-61.

32. Slimings C, Riley TV. Antibiotics and hospital-acquired *Clostridium difficile* infection: update of systematic review and meta-analysis. *The Journal of antimicrobial chemotherapy*. 2014;69(4):881-91.
33. Brown KA, Langford B, Schwartz KL, Diong C, Garber G, Daneman N. Antibiotic prescribing choices and their comparative *C. difficile* infection risks: a longitudinal case-cohort study. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2020.
34. Furuya-Kanamori L, Stone JC, Clark J, McKenzie SJ, Yakob L, Paterson DL, et al. Comorbidities, Exposure to Medications, and the Risk of Community-Acquired *Clostridium difficile* Infection: a systematic review and meta-analysis. *Infection control and hospital epidemiology*. 2015;36(2):132-41.
35. Vardakas KZ, Konstantelias AA, Loizidis G, Rafailidis PI, Falagas ME. Risk factors for development of *Clostridium difficile* infection due to BI/NAP1/027 strain: a meta-analysis. *International journal of infectious diseases: IJID : official publication of the International Society for Infectious Diseases*. 2012;16(11):e768-73.
36. Hopkins MJ, Sharp R, Macfarlane GT. Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. *Gut*. 2001;48(2):198-205.
37. Claesson MJ, Cusack S, O'Sullivan O, Greene-Diniz R, de Weerd H, Flannery E, et al. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108 Suppl 1:4586-91.
38. Clabots CR, Johnson S, Olson MM, Peterson LR, Gerding DN. Acquisition of *Clostridium difficile* by hospitalized patients: evidence for colonized new admissions as a source of infection. *J Infect Dis*. 1992;166(3):561-7.
39. Abou Chakra CN, Pepin J, Sirard S, Valiquette L. Risk factors for recurrence, complications and mortality in *Clostridium difficile* infection: a systematic review. *PLoS One*. 2014;9(6):e98400.
40. Kwok CS, Arthur AK, Anibueze CI, Singh S, Cavallazzi R, Loke YK. Risk of *Clostridium difficile* infection with acid suppressing drugs and antibiotics: meta-analysis. *The American journal of gastroenterology*. 2012;107(7):1011-9.
41. Razik R, Rumman A, Bahreini Z, McGeer A, Nguyen GC. Recurrence of *Clostridium difficile* Infection in Patients with Inflammatory Bowel Disease: The RECIDIVISM Study. *The American journal of gastroenterology*. 2016;111(8):1141-6.
42. Phatharacharukul P, Thongprayoon C, Cheungpasitporn W, Edmonds PJ, Mahaparn P, Bruminhent J. The Risks of Incident and Recurrent *Clostridium difficile*-Associated Diarrhea in Chronic Kidney Disease and End-Stage Kidney Disease Patients: A Systematic Review and Meta-Analysis. *Digestive diseases and sciences*. 2015;60(10):2913-22.
43. VanInsberghe D, Elsherbini JA, Varian B, Poutahidis T, Erdman S, Polz MF. Diarrhoeal events can trigger long-term *Clostridium difficile* colonization with recurrent blooms. *Nature microbiology*. 2020;5(4):642-50.
44. Mefferd CC, Bhute SS, Phan JR, Villarama JV, Do DM, Alarcia S, et al. A High-Fat/High-Protein, Atkins-Type Diet Exacerbates *Clostridioides (Clostridium) difficile* Infection in Mice, whereas a High-Carbohydrate Diet Protects. *mSystems*. 2020;5(1).
45. Riggs MM, Sethi AK, Zabarsky TF, Eckstein EC, Jump RL, Donskey CJ. Asymptomatic carriers are a potential source for transmission of epidemic and nonepidemic *Clostridium difficile* strains among long-term care facility residents. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2007;45(8):992-8.

46. Shim JK, Johnson S, Samore MH, Bliss DZ, Gerding DN. Primary symptomless colonisation by *Clostridium difficile* and decreased risk of subsequent diarrhoea. *Lancet* (London, England). 1998;351(9103):633-6.
47. McDonald LC, Gerding DN, Johnson S, Bakken JS, Carroll KC, Coffin SE, et al. Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clinical Infectious Diseases*. 2018;66(7):e1-e48.
48. Dubberke ER, Han Z, Bobo L, Hink T, Lawrence B, Copper S, et al. Impact of clinical symptoms on interpretation of diagnostic assays for *Clostridium difficile* infections. *Journal of clinical microbiology*. 2011;49(8):2887-93.
49. Planche TD, Davies KA, Coen PG, Finney JM, Monahan IM, Morris KA, et al. Differences in outcome according to *Clostridium difficile* testing method: a prospective multicentre diagnostic validation study of *C. difficile* infection. *The Lancet Infectious diseases*. 2013;13(11):936-45.
50. Burgner D, Siarakas S, Eagles G, McCarthy A, Bradbury R, Stevens M. A prospective study of *Clostridium difficile* infection and colonization in pediatric oncology patients. *The Pediatric infectious disease journal*. 1997;16(12):1131-4.
51. Al-Jumaili IJ, Shibley M, Lishman AH, Record CO. Incidence and origin of *Clostridium difficile* in neonates. *Journal of clinical microbiology*. 1984;19(1):77.
52. Enoch DA, Butler MJ, Pai S, Aliyu SH, Karas JA. *Clostridium difficile* in children: Colonisation and disease. *Journal of Infection*. 2011;63(2):105-13.
53. Drall KM, Tun HM, Morales-Lizcano NP, Konya TB, Guttman DS, Field CJ, et al. *Clostridioides difficile* Colonization Is Differentially Associated With Gut Microbiome Profiles by Infant Feeding Modality at 3-4 Months of Age. *Front Immunol*. 2019;10:2866-.
54. Terveer EM, Crobach MJT, Sanders IMJG, Vos MC, Verduin CM, Kuijper EJ. Detection of *Clostridium difficile* in Feces of Asymptomatic Patients Admitted to the Hospital. *Journal of clinical microbiology*. 2017;55(2):403-11.
55. Ozaki E, Kato H, Kita H, Karasawa T, Maegawa T, Koino Y, et al. *Clostridium difficile* colonization in healthy adults: transient colonization and correlation with enterococcal colonization. *Journal of Medical Microbiology*. 2004;53(2):167-72.
56. Sheth PM, Douchant K, Uyanwune Y, Larocque M, Anantharajah A, Borgundvaag E, et al. Evidence of transmission of *Clostridium difficile* in asymptomatic patients following admission screening in a tertiary care hospital. *PLOS ONE*. 2019;14(2):e0207138.
57. Guerrero DM, Becker JC, Eckstein EC, Kundrapu S, Deshpande A, Sethi AK, et al. Asymptomatic carriage of toxigenic *Clostridium difficile* by hospitalized patients. *The Journal of hospital infection*. 2013;85(2):155-8.
58. Nissle K, Kopf D, Rösler A. Asymptomatic and yet *C. difficile*-toxin positive? Prevalence and risk factors of carriers of toxigenic *Clostridium difficile* among geriatric in-patients. *BMC Geriatrics*. 2016;16(1):185.
59. Gilboa M, Houri-Levi E, Cohen C, Tal I, Rubin C, Feld-Simon O, et al. Environmental shedding of toxigenic *Clostridioides difficile* by asymptomatic carriers: A prospective observational study. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2020;26(8):1052-7.
60. PHE. 30-day all-cause fatality subsequent to MRSA, MSSA and Gram-negative bacteraemia and *C. difficile* infections, 2017/18. London: Public Health England. 2018.

61. Luo R, Barlam TF. Ten-year review of *Clostridium difficile* infection in acute care hospitals in the USA, 2005-2014. *Journal of Hospital Infection*. 2018;98(1):40-3.
62. Marsh JW, Arora R, Schlackman JL, Shutt KA, Curry SR, Harrison LH. Association of relapse of *Clostridium difficile* disease with BI/NAP1/027. *Journal of clinical microbiology*. 2012;50(12):4078-82.
63. McFarland LV, Surawicz CM, Rubin M, Fekety R, Elmer GW, Greenberg RN. Recurrent *Clostridium difficile* Disease: Epidemiology and Clinical Characteristics. *Infection Control Hospital Epidemiology*. 1999;20(1):43-50.
64. Kim J, Seo MR, Kang JO, Kim Y, Hong SP, Pai H. Clinical characteristics of relapses and re-infections in *Clostridium difficile* infection. *Clinical Microbiology and Infection*. 2014;20(11):1198-204.
65. Alfayyadh M, Collins DA, Tempone S, McCann R, Armstrong PK, Riley TV, et al. Recurrence of *Clostridium difficile* infection in the Western Australian population. *Epidemiology and Infection*. 2019;147:e153.
66. Cho J, Cunningham S, Pu M, Lennon RJ, Dens Higanos J, Jeraldo P, et al. *Clostridioides difficile* Whole Genome Sequencing Differentiates Relapse with the Same Strain from Reinfection with a New Strain. *Clinical Infectious Diseases*. 2020;72(5):806-13.
67. PHE. Updating guidance on the management and treatment of *Clostridium difficile* Infection London: Public Health England. 2013.
68. Louie TJ, Cannon K, Byrne B, Emery J, Ward L, Eyben M, et al. Fidaxomicin preserves the intestinal microbiome during and after treatment of *Clostridium difficile* infection (CDI) and reduces both toxin expression and recurrence of CDI. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012;55 Suppl 2(Suppl 2):S132-S42.
69. Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, et al. Decreased diversity of the fecal Microbiome in recurrent *Clostridium difficile*-associated diarrhea. *J Infect Dis*. 2008;197(3):435-8.
70. Finegold SM, Molitoris D, Vaisanen ML, Song Y, Liu C, Bolanos M. *In vitro* activities of OPT-80 and comparator drugs against intestinal bacteria. *Antimicrobial agents and chemotherapy*. 2004;48(12):4898-902.
71. Louie TJ, Emery J, Krulicki W, Byrne B, Mah M. OPT-80 eliminates *Clostridium difficile* and is sparing of *Bacteroides* species during treatment of *C. difficile* infection. *Antimicrobial agents and chemotherapy*. 2009;53(1):261-3.
72. Crook DW, Walker AS, Kean Y, Weiss K, Cornely OA, Miller MA, et al. Fidaxomicin versus vancomycin for *Clostridium difficile* infection: meta-analysis of pivotal randomized controlled trials. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012;55 Suppl 2:S93-103.
73. Cornely OA, Miller MA, Louie TJ, Crook DW, Gorbach SL. Treatment of first recurrence of *Clostridium difficile* infection: fidaxomicin versus vancomycin. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012;55 Suppl 2(Suppl 2):S154-S61.
74. Gallagher JC, Reilly JP, Navalkele B, Downham G, Haynes K, Trivedi M. Clinical and economic benefits of fidaxomicin compared to vancomycin for *Clostridium difficile* infection. *Antimicrobial agents and chemotherapy*. 2015;59(11):7007-10.
75. Vickers RJ, Tillotson GS, Nathan R, Hazan S, Pullman J, Lucasti C, et al. Efficacy and safety of ridinilazole compared with vancomycin for the treatment of *Clostridium difficile* infection: a phase 2, randomised, double-blind, active-controlled, non-inferiority study. *The Lancet Infectious diseases*. 2017;17(7):735-44.

76. Goldstein EJ, Citron DM, Tyrrell KL, Merriam CV. Comparative *in vitro* activities of SMT19969, a new antimicrobial agent, against *Clostridium difficile* and 350 gram-positive and gram-negative aerobic and anaerobic intestinal flora isolates. *Antimicrobial agents and chemotherapy*. 2013;57(10):4872-6.
77. Baines SD, Crowther GS, Freeman J, Todhunter S, Vickers R, Wilcox MH. SMT19969 as a treatment for *Clostridium difficile* infection: an assessment of antimicrobial activity using conventional susceptibility testing and an *in vitro* gut model. *The Journal of antimicrobial chemotherapy*. 2015;70(1):182-9.
78. Mitra S, Chilton C, Freeman J, Wood H, Quirke P, Taylor M, et al. Preservation of Gut Microbiome Following Ridinilazole vs. Fidaxomicin Treatment of *Clostridium difficile* Infection. *Open Forum Infectious Diseases*. 2017;4(suppl_1):S526-S7.
79. Leslie JL, Vendrov KC, Jenior ML, Young VB. The Gut Microbiota Is Associated with Clearance of *Clostridium difficile* Infection Independent of Adaptive Immunity. *mSphere*. 2019;4(1).
80. Moayyedi P, Yuan Y, Baharath H, Ford AC. Faecal microbiota transplantation for *Clostridium difficile*-associated diarrhoea: a systematic review of randomised controlled trials. *The Medical journal of Australia*. 2017;207(4):166-72.
81. Hocquart M, Lagier J-C, Cassir N, Saidani N, Eldin C, Kerbaj J, et al. Early Fecal Microbiota Transplantation Improves Survival in Severe *Clostridium difficile* Infections. *Clinical Infectious Diseases*. 2018;66(5):645-50.
82. Varier RU, Biltaji E, Smith KJ, Roberts MS, Kyle Jensen M, LaFleur J, et al. Cost-effectiveness analysis of fecal microbiota transplantation for recurrent *Clostridium difficile* infection. *Infection control and hospital epidemiology*. 2015;36(4):438-44.
83. Zowall H, Brewer C, Deutsch A. PIN71 - Cost-Effectiveness of Fecal Microbiota Transplant in Treating *Clostridium difficile* Infection in Canada. *Value in Health*. 2014;17(7):A676.
84. You JHS, Jiang X, Lee WH, Chan PKS, Ng SC. Cost-effectiveness analysis of fecal microbiota transplantation for recurrent *Clostridium difficile* infection in patients with inflammatory bowel disease. *Journal of gastroenterology and hepatology*. 2020.
85. Merlo G, Graves N, Brain D, Connelly LB. Economic evaluation of fecal microbiota transplantation for the treatment of recurrent *Clostridium difficile* infection in Australia. *Journal of gastroenterology and hepatology*. 2016;31(12):1927-32.
86. Zipursky JS, Sidorsky TI, Freedman CA, Sidorsky MN, Kirkland KB. Patient attitudes toward the use of fecal microbiota transplantation in the treatment of recurrent *Clostridium difficile* infection. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2012;55(12):1652-8.
87. Solari PR, Fairchild PG, Noa LJ, Wallace MR. Tempered Enthusiasm for Fecal Transplant. *Clinical Infectious Diseases*. 2014;59(2):319-.
88. FDA. Information Pertaining to Additional Safety Protections Regarding Use of Fecal Microbiota for Transplantation – Screening and Testing of Stool Donors for Multi-drug Resistant Organisms. 2019.
89. Wilcox MH, Gerding DN, Poxton IR, Kelly C, Nathan R, Birch T, et al. Bezlotoxumab for Prevention of Recurrent *Clostridium difficile* Infection. *The New England Journal of Medicine*. 2017;376(4):305-17.
90. Lowy I, Molrine DC, Leav BA, Blair BM, Baxter R, Gerding DN, et al. Treatment with Monoclonal Antibodies against *Clostridium difficile* Toxins. *New England Journal of Medicine*. 2010;362(3):197-205.
91. Oksi J, Aalto A, Säilä P, Partanen T, Anttila VJ, Mattila E. Real-world efficacy of bezlotoxumab for prevention of recurrent *Clostridium difficile* infection: a retrospective study of 46 patients in five university hospitals in Finland. *Eur J Clin Microbiol Infect Dis*. 2019;38(10):1947-52.

92. Alonso CD, Mahoney MV. Bezlotoxumab for the prevention of *Clostridium difficile* infection: a review of current evidence and safety profile. *Infect Drug Resist.* 2018;12:1-9.
93. Leuzzi R, Spencer J, Buckley A, Brettoni C, Martinelli M, Tulli L, et al. Protective efficacy induced by recombinant *Clostridium difficile* toxin fragments. *Infection and immunity.* 2013;81(8):2851-60.
94. Kotloff KL, Wasserman SS, Losonsky GA, Thomas W, Jr., Nichols R, Edelman R, et al. Safety and immunogenicity of increasing doses of a *Clostridium difficile* toxoid vaccine administered to healthy adults. *Infection and immunity.* 2001;69(2):988-95.
95. Torres JF, Lyerly DM, Hill JE, Monath TP. Evaluation of formalin-inactivated *Clostridium difficile* vaccines administered by parenteral and mucosal routes of immunization in hamsters. *Infection and immunity.* 1995;63(12):4619-27.
96. Sougioultzis S, Kyne L, Drudy D, Keates S, Maroo S, Pothoulakis C, et al. *Clostridium difficile* toxoid vaccine in recurrent *C. difficile*-associated diarrhea. *Gastroenterology.* 2005;128(3):764-70.
97. Medicine USNLo. Study of a Candidate *Clostridium difficile* Toxoid Vaccine in Subjects at Risk for *C. difficile* Infection. 2019 [Available from: <https://clinicaltrials.gov/ct2/show/NCT01887912>].
98. Kitchin N, Remich SA, Peterson J, Peng Y, Gruber WC, Jansen KU, et al. A Phase 2 Study Evaluating the Safety, Tolerability, and Immunogenicity of Two 3-Dose Regimens of a *Clostridium difficile* Vaccine in Healthy US Adults Aged 65 to 85 Years. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America.* 2020;70(1):1-10.
99. Valneva. Valneva Announces Successful Completion of Phase II for *Clostridium Difficile* Vaccine Candidate. 2016.
100. Siddiqui F, O'Connor JR, Nagaro K, Cheknis A, Sambol SP, Vedantam G, et al. Vaccination With Parenteral Toxoid B Protects Hamsters Against Lethal Challenge With Toxin A–Negative, Toxin B–Positive *Clostridium difficile* but Does Not Prevent Colonization. *The Journal of Infectious Diseases.* 2011;205(1):128-33.
101. Bradshaw WJ, Bruxelle J-F, Kovacs-Simon A, Harmer NJ, Janoir C, Péchiné S, et al. Molecular features of lipoprotein CD0873: A potential vaccine against the human pathogen *Clostridioides difficile*. *Journal of Biological Chemistry.* 2019;294(43):15850-61.
102. Ghose C, Eugenis I, Edwards AN, Sun X, McBride SM, Ho DD. Immunogenicity and protective efficacy of *Clostridium difficile* spore proteins. *Anaerobe.* 2016;37:85-95.
103. Hammond GA, Johnson JL. The toxigenic element of *Clostridium difficile* strain VPI 10463. *Microbial pathogenesis.* 1995;19(4):203-13.
104. Matamouros S, England P, Dupuy B. *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Mol Microbiol.* 2007;64(5):1274-88.
105. Mani N, Dupuy B. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proceedings of the National Academy of Sciences of the United States of America.* 2001;98(10):5844-9.
106. Govind R, Dupuy B. Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. *PLoS Pathog.* 2012;8(6):e1002727.
107. Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP. The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature.* 2010;467(7316):711-3.
108. Just I, Selzer J, Wilm M, von Eichel-Streiber C, Mann M, Aktories K. Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature.* 1995;375(6531):500-3.
109. Just I, Wilm M, Selzer J, Rex G, von Eichel-Streiber C, Mann M, et al. The enterotoxin from *Clostridium difficile* (ToxA) monoglucosylates the Rho proteins. *The Journal of biological chemistry.* 1995;270(23):13932-6.

110. Ottlinger ME, Lin S. *Clostridium difficile* toxin B induces reorganization of actin, vinculin, and talin in cultured cells. *Exp Cell Res*. 1988;174(1):215-29.
111. Chaves-Olarte E, Weidmann M, Eichel-Streiber C, Thelestam M. Toxins A and B from *Clostridium difficile* differ with respect to enzymatic potencies, cellular substrate specificities, and surface binding to cultured cells. *The Journal of clinical investigation*. 1997;100(7):1734-41.
112. Ciesla WP, Jr., Bobak DA. *Clostridium difficile* toxins A and B are cation-dependent UDP-glucose hydrolases with differing catalytic activities. *The Journal of biological chemistry*. 1998;273(26):16021-6.
113. Carter GP, Chakravorty A, Pham Nguyen TA, Mileto S, Schreiber F, Li L, et al. Defining the Roles of TcdA and TcdB in Localized Gastrointestinal Disease, Systemic Organ Damage, and the Host Response during *Clostridium difficile* Infections. *mBio*. 2015;6(3):e00551.
114. van den Berg RJ, Claas EC, Oyib DH, Klaassen CH, Dijkshoorn L, Brazier JS, et al. Characterization of toxin A-negative, toxin B-positive *Clostridium difficile* isolates from outbreaks in different countries by amplified fragment length polymorphism and PCR ribotyping. *Journal of clinical microbiology*. 2004;42(3):1035-41.
115. Kim H, Jeong SH, Roh KH, Hong SG, Kim JW, Shin MG, et al. Investigation of toxin gene diversity, molecular epidemiology, and antimicrobial resistance of *Clostridium difficile* isolated from 12 hospitals in South Korea. *Korean J Lab Med*. 2010;30(5):491-7.
116. Kim J, Kim Y, Pai H. Clinical Characteristics and Treatment Outcomes of *Clostridium difficile* Infections by PCR Ribotype 017 and 018 Strains. *PLoS One*. 2016;11(12):e0168849.
117. Jin D, Luo Y, Huang C, Cai J, Ye J, Zheng Y, et al. Molecular Epidemiology of *Clostridium difficile* Infection in Hospitalized Patients in Eastern China. *Journal of clinical microbiology*. 2017;55(3):801-10.
118. Hundesberger T, Braun V, Weidmann M, Leukel P, Sauerborn M, von Eichel-Streiber C. Transcription analysis of the genes *tcdA-E* of the pathogenicity locus of *Clostridium difficile*. *European journal of biochemistry*. 1997;244(3):735-42.
119. Dupuy B, Govind R, Antunes A, Matamouros S. *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. *J Med Microbiol*. 2008;57(Pt 6):685-9.
120. Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype and toxin production. *Applied and environmental microbiology*. 2012;78(13):4683-90.
121. Bakker D, Smits WK, Kuijper EJ, Corver J. TcdC Does Not Significantly Repress Toxin Expression in *Clostridium difficile* 630Δ*erm*. *PLOS ONE*. 2012;7(8):e43247.
122. Shivers RP, Sonenshein AL. Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. *Molecular microbiology*. 2004;53(2):599-611.
123. Ratnayake-Lecamwasam M, Serror P, Wong KW, Sonenshein AL. *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev*. 2001;15(9):1093-103.
124. Dineen SS, Villapakkam AC, Nordman JT, Sonenshein AL. Repression of *Clostridium difficile* toxin gene expression by CodY. *Molecular microbiology*. 2007;66(1):206-19.
125. Antunes A, Martin-Verstraete I, Dupuy B. CcpA-mediated repression of *Clostridium difficile* toxin gene expression. *Molecular microbiology*. 2011;79(4):882-99.
126. Antunes A, Camiade E, Monot M, Courtois E, Barbut F, Sernova NV, et al. Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. *Nucleic Acids Research*. 2012;40(21):10701-18.
127. Pettit LJ, Browne HP, Yu L, Smits WK, Fagan RP, Barquist L, et al. Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. *BMC genomics*. 2014;15:160.

128. Mackin KE, Carter GP, Howarth P, Rood JJ, Lyras D. Spo0A Differentially Regulates Toxin Production in Evolutionarily Diverse Strains of *Clostridium difficile*. PLOS ONE. 2013;8(11):e79666.
129. Edwards AN, Anjuwon-Foster BR, McBride SM. RstA Is a Major Regulator of *Clostridioides difficile* Toxin Production and Motility. mBio. 2019;10(2):e01991-18.
130. Androga GO, Hart J, Foster NF, Charles A, Forbes D, Riley TV. Infection with Toxin A-Negative, Toxin B-Negative, Binary Toxin-Positive *Clostridium difficile* in a Young Patient with Ulcerative Colitis. Journal of clinical microbiology. 2015;53(11):3702-4.
131. Eckert C, Emirian A, Le Monnier A, Cathala L, De Montclos H, Goret J, et al. Prevalence and pathogenicity of binary toxin-positive *Clostridium difficile* strains that do not produce toxins A and B. New microbes and new infections. 2015;3:12-7.
132. Spigaglia P, Mastrantonio P. Comparative analysis of *Clostridium difficile* clinical isolates belonging to different genetic lineages and time periods. J Med Microbiol. 2004;53(Pt 11):1129-36.
133. Stewart DB, Berg A, Hegarty J. Predicting recurrence of *C. difficile* colitis using bacterial virulence factors: binary toxin is the key. Journal of gastrointestinal surgery: official journal of the Society for Surgery of the Alimentary Tract. 2013;17(1):118-24; discussion p.24-5.
134. Bacci S, Molbak K, Kjeldsen MK, Olsen KE. Binary toxin and death after *Clostridium difficile* infection. Emerging infectious diseases. 2011;17(6):976-82.
135. Schwan C, Kruppke AS, Nölke T, Schumacher L, Koch-Nolte F, Kudryashev M, et al. *Clostridium difficile* toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence. Proceedings of the National Academy of Sciences. 2014;111(6):2313.
136. Schwan C, Stecher B, Tzivelekidis T, van Ham M, Rohde M, Hardt WD, et al. *Clostridium difficile* toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. PLoS Pathog. 2009;5(10):e1000626.
137. Kuehne SA, Collery MM, Kelly ML, Cartman ST, Cockayne A, Minton NP. Importance of toxin A, toxin B, and CDT in virulence of an epidemic *Clostridium difficile* strain. J Infect Dis. 2014;209(1):83-6.
138. Deakin LJ, Clare S, Fagan RP, Dawson LF, Pickard DJ, West MR, et al. The *Clostridium difficile* spo0A Gene Is a Persistence and Transmission Factor. Infection and immunity. 2012;80(8):2704-11.
139. Baines SD, O'Connor R, Saxton K, Freeman J, Wilcox MH. Activity of vancomycin against epidemic *Clostridium difficile* strains in a human gut model. The Journal of antimicrobial chemotherapy. 2009;63(3):520-5.
140. Dawson LF, Valiente E, Donahue EH, Birchenough G, Wren BW. Hypervirulent *Clostridium difficile* PCR-Ribotypes Exhibit Resistance to Widely Used Disinfectants. PLOS ONE. 2011;6(10):e25754.
141. Oughton MT, Loo VG, Dendukuri N, Fenn S, Libman MD. Hand hygiene with soap and water is superior to alcohol rub and antiseptic wipes for removal of *Clostridium difficile*. Infection control and hospital epidemiology. 2009;30(10):939-44.
142. Pickering DS, Vernon JJ, Freeman J, Wilcox MH, Chilton CH. Investigating the transient and persistent effects of heat on *Clostridium difficile* spores. Journal of Medical Microbiology. 2019;68(10):1445-54.
143. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. The ClosTron: a universal gene knock-out system for the genus *Clostridium*. Journal of microbiological methods. 2007;70(3):452-64.
144. Rosenbusch KE, Bakker D, Kuijper EJ, Smits WK. *C. difficile* 630 Δ erm Spo0A regulates sporulation, but does not contribute to toxin production, by direct high-affinity binding to target DNA. PLoS One. 2012;7(10):e48608.
145. Dawson LF, Valiente E, Faulds-Pain A, Donahue EH, Wren BW. Characterisation of *Clostridium difficile* Biofilm Formation, a Role for Spo0A. PLOS ONE. 2012;7(12):e50527.

146. Dapa T, Leuzzi R, Ng YK, Baban ST, Adamo R, Kuehne SA, et al. Multiple Factors Modulate Biofilm Formation by the Anaerobic Pathogen *Clostridium difficile*. *Journal of bacteriology*. 2013;195(3):545.
147. Wörner K, Szurmant H, Chiang C, Hoch JA. Phosphorylation and functional analysis of the sporulation initiation factor Spo0A from *Clostridium botulinum*. *Molecular microbiology*. 2006;59(3):1000-12.
148. Steiner E, Dago AE, Young DI, Heap JT, Minton NP, Hoch JA, et al. Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. *Molecular microbiology*. 2011;80(3):641-54.
149. Underwood S, Guan S, Vijayasubhash V, Baines SD, Graham L, Lewis RJ, et al. Characterization of the Sporulation Initiation Pathway of *Clostridium difficile* and Its Role in Toxin Production. *Journal of bacteriology*. 2009;191(23):7296-305.
150. Childress KO, Edwards AN, Nawrocki KL, Anderson SE, Woods EC, McBride SM. The Phosphotransfer Protein CD1492 Represses Sporulation Initiation in *Clostridium difficile*. *Infection and immunity*. 2016;84(12):3434-44.
151. Oliveira PH, Ribis JW, Garrett EM, Trzilova D, Kim A, Sekulovic O, et al. Epigenomic characterization of *Clostridioides difficile* finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. *Nature microbiology*. 2020;5(1):166-80.
152. Srikhanta YN, Hutton ML, Awad MM, Drinkwater N, Singleton J, Day SL, et al. Cephamycins inhibit pathogen sporulation and effectively treat recurrent *Clostridioides difficile* infection. *Nature microbiology*. 2019;4(12):2237-45.
153. Oka K, Osaki T, Hanawa T, Kurata S, Okazaki M, Manzoku T, et al. Molecular and microbiological characterization of *Clostridium difficile* isolates from single, relapse, and reinfection cases. *Journal of clinical microbiology*. 2012;50(3):915-21.
154. Freeman J, Vernon J, Morris K, Nicholson S, Todhunter S, Longshaw C, et al. Pan-European longitudinal surveillance of antibiotic resistance among prevalent *Clostridium difficile* ribotypes. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2015;21(3):248.e9-.e16.
155. Karlowsky JA, Zhanel GG, Hammond GW, Rubinstein E, Wylie J, Du T, et al. Multidrug-resistant North American pulsotype 2 *Clostridium difficile* was the predominant toxigenic hospital-acquired strain in the province of Manitoba, Canada, in 2006-2007. *J Med Microbiol*. 2012;61(Pt 5):693-700.
156. Lee JH, Lee Y, Lee K, Riley TV, Kim H. The changes of PCR ribotype and antimicrobial resistance of *Clostridium difficile* in a tertiary care hospital over 10 years. *J Med Microbiol*. 2014;63(Pt 6):819-23.
157. Norman KN, Scott HM, Harvey RB, Norby B, Hume ME. Comparison of antimicrobial susceptibility among *Clostridium difficile* isolated from an integrated human and swine population in Texas. *Foodborne pathogens and disease*. 2014;11(4):257-64.
158. Thorpe CM, McDermott LA, Tran MK, Chang J, Jenkins SG, Goldstein EJC, et al. U.S.-Based National Surveillance for Fidaxomicin Susceptibility of *Clostridioides difficile*-Associated Diarrheal Isolates from 2013 to 2016. *Antimicrobial agents and chemotherapy*. 2019;63(7).
159. Sebahia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, et al. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet*. 2006;38(7):779-86.
160. Obuch-Woszczatynski P, Lachowicz D, Schneider A, Mol A, Pawlowska J, Ozdzenska-Milke E, et al. Occurrence of *Clostridium difficile* PCR-ribotype 027 and its closely related PCR-ribotype 176 in hospitals in Poland in 2008-2010. *Anaerobe*. 2014;28:13-7.

161. Senoh M, Kato H, Fukuda T, Niikawa A, Hori Y, Hagiya H, et al. Predominance of PCR-ribotypes, 018 (smz) and 369 (trf) of *Clostridium difficile* in Japan: a potential relationship with other global circulating strains? *J Med Microbiol*. 2015;64(10):1226-36.
162. Reil M, Hensgens MP, Kuijper EJ, Jakobiak T, Gruber H, Kist M, et al. Seasonality of *Clostridium difficile* infections in Southern Germany. *Epidemiol Infect*. 2012;140(10):1787-93.
163. Adler A, Miller-Roll T, Bradenstein R, Block C, Mendelson B, Parizade M, et al. A national survey of the molecular epidemiology of *Clostridium difficile* in Israel: the dissemination of the ribotype 027 strain with reduced susceptibility to vancomycin and metronidazole. *Diagnostic microbiology and infectious disease*. 2015;83(1):21-4.
164. Boekhoud IM, Hornung BVH, Sevilla E, Harmanus C, Bos-Sanders IMJG, Terveer EM, et al. Plasmid-mediated metronidazole resistance in *Clostridioides difficile*. *Nature communications*. 2020;11(1):598.
165. Dong D, Zhang L, Chen X, Jiang C, Yu B, Wang X, et al. Antimicrobial susceptibility and resistance mechanisms of clinical *Clostridium difficile* from a Chinese tertiary hospital. *International journal of antimicrobial agents*. 2013;41(1):80-4.
166. Freeman J, Vernon J, Pilling S, Morris K, Nicolson S, Shearman S, et al. Five-year Pan-European, longitudinal surveillance of *Clostridium difficile* ribotype prevalence and antimicrobial resistance: the extended ClosER study. *Eur J Clin Microbiol Infect Dis*. 2020;39(1):169-77.
167. Goldstein EJ, Citron DM, Sears P, Babakhani F, Sambol SP, Gerding DN. Comparative susceptibilities to fidaxomicin (OPT-80) of isolates collected at baseline, recurrence, and failure from patients in two phase III trials of fidaxomicin against *Clostridium difficile* infection. *Antimicrobial agents and chemotherapy*. 2011;55(11):5194-9.
168. Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences*. 1994;91(26):12501-4.
169. Mayer MJ, Garefalaki V, Spoerl R, Narbad A, Meijers R. Structure-based modification of a *Clostridium difficile*-targeting endolysin affects activity and host range. *Journal of bacteriology*. 2011;193(19):5477-86.
170. Pereira FC, Saujet L, Tomé AR, Serrano M, Monot M, Couture-Tosi E, et al. The Spore Differentiation Pathway in the Enteric Pathogen *Clostridium difficile*. *PLOS Genetics*. 2013;9(10):e1003782.
171. Fagan RP, Fairweather NF. *Clostridium difficile* Has Two Parallel and Essential Sec Secretion Systems. *The Journal of biological chemistry*. 2011;286(31):27483-93.
172. Edwards AN, Pascual RA, Childress KO, Nawrocki KL, Woods EC, McBride SM. An alkaline phosphatase reporter for use in *Clostridium difficile*. *Anaerobe*. 2015;32:98-104.
173. Ransom EM, Weiss DS, Ellermeier CD. Use of mCherryOpt Fluorescent Protein in *Clostridium difficile*. *Methods in molecular biology (Clifton, NJ)*. 2016;1476:53-67.
174. Buckley AM, Jukes C, Candlish D, Irvine JJ, Spencer J, Fagan RP, et al. Lighting Up *Clostridium difficile*: Reporting Gene Expression Using Fluorescent Lov Domains. *Scientific Reports*. 2016;6(1):23463.
175. Mordaka PM, Heap JT. Stringency of Synthetic Promoter Sequences in *Clostridium* Revealed and Circumvented by Tuning Promoter Library Mutation Rates. *ACS Synthetic Biology*. 2018;7(2):672-81.
176. Cassona CP, Pereira F, Serrano M, Henriques AO. A Fluorescent Reporter for Single Cell Analysis of Gene Expression in *Clostridium difficile*. *Methods in molecular biology (Clifton, NJ)*. 2016;1476:69-90.
177. Gautier A, Juillerat A, Heinis C, Corrêa IR, Jr., Kindermann M, Beaufils F, et al. An engineered protein tag for multiprotein labeling in living cells. *Chem Biol*. 2008;15(2):128-36.

178. Bruno D, L. SA. Regulated transcription of *Clostridium difficile* toxin genes. *Molecular microbiology*. 1998;27(1):107-20.
179. Ransom EM, Ellermeier CD, Weiss DS, Kivisaar M. Use of mCherry Red Fluorescent Protein for Studies of Protein Localization and Gene Expression in *Clostridium difficile*. *Applied and environmental microbiology*. 2015;81(5):1652-60.
180. Dineen SS, McBride SM, Sonenshein AL. Integration of Metabolism and Virulence by *Clostridium difficile* CodY. *Journal of bacteriology*. 2010;192(20):5350-62.
181. Slack FJ, Mueller JP, Strauch MA, Mathiopoulos C, Sonenshein AL. Transcriptional regulation of a *Bacillus subtilis* dipeptide transport operon. *Molecular microbiology*. 1991;5(8):1915-25.
182. Slack FJ, Serror P, Joyce E, Sonenshein AL. A gene required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. *Molecular microbiology*. 1995;15(4):689-702.
183. Pohl K, Francois P, Stenz L, Schlink F, Geiger T, Herbert S, et al. CodY in *Staphylococcus aureus*: a Regulatory Link between Metabolism and Virulence Gene Expression. *Journal of bacteriology*. 2009;191(9):2953-63.
184. Dupuy B, Sonenshein AL. Regulated transcription of *Clostridium difficile* toxin genes. *Molecular microbiology*. 1998;27(1):107-20.
185. Nawrocki KL, Edwards AN, Daou N, Bouillaut L, McBride SM. CodY-Dependent Regulation of Sporulation in *Clostridium difficile*. *Journal of bacteriology*. 2016;198(15):2113-30.
186. Edwards AN, Krall EG, McBride SM. Strain-Dependent RstA Regulation of *Clostridioides difficile* Toxin Production and Sporulation. *Journal of bacteriology*. 2020;202(2):e00586-19.
187. Chen K-Y, Rathod J, Chiu Y-C, Chen J-W, Tsai P-J, Huang I-H. The Transcriptional Regulator Lrp Contributes to Toxin Expression, Sporulation, and Swimming Motility in *Clostridium difficile*. *Frontiers in Cellular and Infection Microbiology*. 2019;9(356).
188. Lewis RJ, Scott DJ, Brannigan JA, Ladds JC, Cervin MA, Spiegelman GB, et al. Dimer formation and transcription activation in the sporulation response regulator Spo0A. *Journal of molecular biology*. 2002;316(2):235-45.
189. Chu F, Kearns DB, Branda SS, Kolter R, Losick R. Targets of the master regulator of biofilm formation in *Bacillus subtilis*. *Molecular microbiology*. 2006;59(4):1216-28.
190. Fujita M, González-Pastor JE, Losick R. High- and Low-Threshold Genes in the Spo0A Regulon of *Bacillus subtilis*. *Journal of bacteriology*. 2005;187(4):1357-68.
191. Bai U, Mandic-Mulec I, Smith I. SinI modulates the activity of SinR, a developmental switch protein of *Bacillus subtilis*, by protein-protein interaction. *Genes Dev*. 1993;7(1):139-48.
192. Scott DJ, Leejeerajumnean S, Brannigan JA, Lewis RJ, Wilkinson AJ, Hoggett JG. Quaternary re-arrangement analysed by spectral enhancement: the interaction of a sporulation repressor with its antagonist. *Journal of molecular biology*. 1999;293(5):997-1004.
193. Chai Y, Norman T, Kolter R, Losick R. An epigenetic switch governing daughter cell separation in *Bacillus subtilis*. *Genes Dev*. 2010;24(8):754-65.
194. Girinathan BP, Ou J, Dupuy B, Govind R. Pleiotropic roles of *Clostridium difficile* *sin* locus. *PLOS Pathogens*. 2018;14(3):e1006940.
195. Ciftci Y, Girinathan BP, Dhungel BA, Hasan MK, Govind R. *Clostridioides difficile* SinR' regulates toxin, sporulation and motility through protein-protein interaction with SinR. *Anaerobe*. 2019;59:1-7.
196. Dhungel BA, Govind R, Ellermeier CD. Spo0A Suppresses *sin* Locus Expression in *Clostridioides difficile*. *mSphere*. 2020;5(6):e00963-20.
197. Sonenshein AL. CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. *Current opinion in microbiology*. 2005;8(2):203-7.

198. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology*. 2014;146(6):1489-99.
199. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59-65.
200. Buffie CG, Jarchum I, Equinda M, Lipuma L, Gobourne A, Viale A, et al. Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infection and immunity*. 2012;80(1):62-73.
201. Zaura E, Brandt BW, Teixeira de Mattos MJ, Buijs MJ, Caspers MP, Rashid MU, et al. Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces. *mBio*. 2015;6(6):e01693-15.
202. Antharam VC, Li EC, Ishmael A, Sharma A, Mai V, Rand KH, et al. Intestinal dysbiosis and depletion of butyrogenic bacteria in *Clostridium difficile* infection and nosocomial diarrhea. *Journal of clinical microbiology*. 2013;51(9):2884-92.
203. Schubert AM, Rogers MAM, Ring C, Mogle J, Petrosino JP, Young VB, et al. Microbiome Data Distinguish Patients with *Clostridium difficile* Infection and Non-*C. difficile*-Associated Diarrhea from Healthy Controls. *mBio*. 2014;5(3):e01021-14.
204. Milani C, Ticinesi A, Gerritsen J, Nouvenne A, Lugli GA, Mancabelli L, et al. Gut microbiota composition and *Clostridium difficile* infection in hospitalized elderly individuals: a metagenomic study. *Scientific Reports*. 2016;6:25945.
205. Vincent C, Manges AR. Antimicrobial Use, Human Gut Microbiota and *Clostridium difficile* Colonization and Infection. *Antibiotics*. 2015;4(3):230-53.
206. Skraban J, Dzeroski S, Zenko B, Mongus D, Gangl S, Rupnik M. Gut Microbiota Patterns Associated with Colonization of Different *Clostridium difficile* Ribotypes. *PLoS ONE*. 2013;8(2):e58005.
207. Manges AR, Labbe A, Loo VG, Atherton JK, Behr MA, Masson L, et al. Comparative metagenomic study of alterations to the intestinal microbiota and risk of nosocomial *Clostridium difficile*-associated disease. *J Infect Dis*. 2010;202(12):1877-84.
208. Khanna S, Montassier E, Schmidt B, Patel R, Knights D, Pardi DS, et al. Gut microbiome predictors of treatment response and recurrence in primary *Clostridium difficile* infection. *Alimentary pharmacology & therapeutics*. 2016;44(7):715-27.
209. Crobach MJT, Ducarmon QR, Terveer EM, Harmanus C, Sanders I, Verduin KM, et al. The Bacterial Gut Microbiota of Adult Patients Infected, Colonized or Noncolonized by *Clostridioides difficile*. *Microorganisms*. 2020;8(5).
210. Lee AA, Rao K, Limsrivilai J, Gilliland M, III, Malamet B, Briggs E, et al. Temporal Gut Microbial Changes Predict Recurrent *Clostridioides difficile* Infection in Patients With and Without Ulcerative Colitis. *Inflammatory Bowel Diseases*. 2020.
211. Ferreyra Jessica A, Wu Katherine J, Hryckowian Andrew J, Bouley Donna M, Weimer Bart C, Sonnenburg Justin L. Gut Microbiota-Produced Succinate Promotes *C. difficile* Infection after Antibiotic Treatment or Motility Disturbance. *Cell Host & Microbe*. 2014;16(6):770-7.
212. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, et al. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature*. 2015;517(7533):205-8.
213. Reed AD, Nethery MA, Stewart A, Barrangou R, Theriot CM. Strain-dependent inhibition of *Clostridioides difficile* by commensal Clostridia encoding the bile acid inducible (*bai*) operon. *Journal of bacteriology*. 2020.

214. Goldenberg JZ, Yap C, Lytvyn L, Lo CK-F, Beardsley J, Mertz D, et al. Probiotics for the prevention of *Clostridium difficile*-associated diarrhea in adults and children. *Cochrane Database Syst Rev*. 2017;12(12):CD006095-CD.
215. McFarland LV, Ship N, Auclair J, Millette M. Primary prevention of *Clostridium difficile* infections with a specific probiotic combining *Lactobacillus acidophilus*, *L. casei*, and *L. rhamnosus* strains: assessing the evidence. *The Journal of hospital infection*. 2018;99(4):443-52.
216. Lau CS, Chamberlain RS. Probiotics are effective at preventing *Clostridium difficile*-associated diarrhea: a systematic review and meta-analysis. *Int J Gen Med*. 2016;9:27-37.
217. Scheline RR. Metabolism of phenolic acids by the rat intestinal microflora. *Acta pharmacologica et toxicologica*. 1968;26(2):189-205.
218. Steglich M, Hofmann JD, Helmecke J, Sikorski J, Sproer C, Riedel T, et al. Convergent Loss of ABC Transporter Genes From *Clostridioides difficile* Genomes Is Associated With Impaired Tyrosine Uptake and *p*-cresol Production. *Frontiers in microbiology*. 2018;9:901.
219. Selmer T, Andrei PI. *p*-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glyceryl radical enzyme catalysing the formation of *p*-cresol. *European journal of biochemistry*. 2001;268(5):1363-72.
220. Dawson LF, Donahue EH, Cartman ST, Barton RH, Bundy J, McNerney R, et al. The analysis of *para*-cresol production and tolerance in *Clostridium difficile* 027 and 012 strains. *BMC microbiology*. 2011;11:86.
221. Dawson LF, Stabler RA, Wren BW. Assessing the role of *p*-cresol tolerance in *Clostridium difficile*. *J Med Microbiol*. 2008;57(Pt 6):745-9.
222. Levett PN. Production of *p*-cresol by *Clostridium difficile* on different basal media. *Letters in Applied Microbiology*. 1987;5(4):71-3.
223. Passmore IJ, Letertre MPM, Preston MD, Bianconi I, Harrison MA, Nasher F, et al. *Para*-cresol production by *Clostridium difficile* affects microbial diversity and membrane integrity of Gram-negative bacteria. *PLOS Pathogens*. 2018;14(9):e1007191.
224. Saito Y, Sato T, Nomoto K, Tsuji H. Identification of phenol- and *p*-cresol-producing intestinal bacteria by using media supplemented with tyrosine and its metabolites. *FEMS Microbiology Ecology*. 2018;94(9).
225. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*. 2008;8(1):91.
226. Kirk JA, Fagan RP. Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe*. 2016;42:1-5.
227. Karasawa T, Ikoma S, Yamakawa K, Nakamura S. A defined growth medium for *Clostridium difficile*. *Microbiology (Reading)*. 1995;141 (Pt 2):371-5.
228. Cartman ST, Minton NP. A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. *Applied and environmental microbiology*. 2010;76(4):1103-9.
229. Dawson LF, Peltier J, Hall CL, Harrison MA, Derakhshan M, Shaw HA, et al. Extracellular DNA, cell surface proteins and c-di-GMP promote biofilm formation in *Clostridioides difficile*. *Scientific Reports*. 2021;11(1):3244.
230. Faulds-Pain A, Twine SM, Vinogradov E, Strong PC, Dell A, Buckley AM, et al. The post-translational modification of the *Clostridium difficile* flagellin affects motility, cell surface properties and virulence. *Mol Microbiol*. 2014;94(2):272-89.
231. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻($\Delta\Delta C(T)$) Method. *Methods*. 2001;25(4):402-8.

232. Rosenfeld N, Elowitz MB, Alon U. Negative Autoregulation Speeds the Response Times of Transcription Networks. *Journal of molecular biology*. 2002;323(5):785-93.
233. Hamon MA, Lazazzera BA. The sporulation transcription factor Spo0A is required for biofilm development in *Bacillus subtilis*. *Mol Microbiol*. 2001;42(5):1199-209.
234. Ellermeier CD, Hobbs EC, Gonzalez-Pastor JE, Losick R. A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. *Cell*. 2006;124(3):549-59.
235. Burbulys D, Trach KA, Hoch JA. Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. *Cell*. 1991;64(3):545-52.
236. Gaur NK, Cabane K, Smith I. Structure and expression of the *Bacillus subtilis* *sin* operon. *Journal of bacteriology*. 1988;170(3):1046-53.
237. Mandic-Mulec I, Doukhan L, Smith I. The *Bacillus subtilis* SinR protein is a repressor of the key sporulation gene *spo0A*. *J Bacteriol*. 1995;177(16):4619-27.
238. Mandic-Mulec I, Gaur N, Bai U, Smith I. Sin, a stage-specific repressor of cellular differentiation. *J Bacteriol*. 1992;174(11):3561-9.
239. Shafikhani SH, Mandic-Mulec I, Strauch MA, Smith I, Leighton T. Post-exponential Regulation of *sin* Operon Expression in *Bacillus subtilis*. *Journal of Bacteriology*. 2002;184(2):564-71.
240. Voigt CA, Wolf DM, Arkin AP. The *Bacillus subtilis* *sin* operon: an evolvable network motif. *Genetics*. 2005;169(3):1187-202.
241. Saujet L, Monot M, Dupuy B, Soutourina O, Martin-Verstraete I. The Key Sigma Factor of Transition Phase, SigH, Controls Sporulation, Metabolism, and Virulence Factor Expression in *Clostridium difficile*. *Journal of Bacteriology*. 2011;193(13):3186-96.
242. El Meouche I, Peltier J, Monot M, Soutourina O, Pestel-Caron M, Dupuy B, et al. Characterization of the SigD Regulon of *C. difficile* and Its Positive Control of Toxin Production through the Regulation of *tcdR*. *PLOS ONE*. 2013;8(12):e83748.
243. Harrison MA, Muench JH, Shaw HA, Preston MD, Derakhshan M, Passmore I, et al. SinR tetramer positively controls expression of the *Clostridioides difficile* *sin* operon, by protein-DNA interactions to adapt to environmental signals by modulating sporulation, motility, chemotaxis, and toxin production in strain 630 Δ *erm*. Submitted to *PLOS Biology*. 2021.
244. Harrison MA, Faulds-Pain A, Kaur H, Dupuy B, Henriques AO, Martin-Verstraete I, et al. *Clostridioides difficile* *para*-cresol Production Is Induced by the Precursor *para*-Hydroxyphenylacetate. *Journal of Bacteriology*. 2020;202(18):e00282-20.
245. Gaur NK, Dubnau E, Smith I. Characterization of a cloned *Bacillus subtilis* gene that inhibits sporulation in multiple copies. *Journal of bacteriology*. 1986;168(2):860-9.
246. Boudry P, Gracia C, Monot M, Caillet J, Saujet L, Hajnsdorf E, et al. Pleiotropic Role of the RNA Chaperone Protein Hfq in the Human Pathogen *Clostridium difficile*. *Journal of Bacteriology*. 2014;196(18):3234-48.
247. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*. 2018;46(D1):D608-d17.
248. Neumann-Schaal M, Hofmann JD, Will SE, Schomburg D. Time-resolved amino acid uptake of *Clostridium difficile* 630 Δ *erm* and concomitant fermentation product and toxin formation. *BMC Microbiol*. 2015;15:281.
249. Theriot CM, Koumpouras CC, Carlson PE, Bergin II, Aronoff DM, Young VB. Cefoperazone-treated mice as an experimental platform to assess differential virulence of *Clostridium difficile* strains. *Gut Microbes*. 2011;2(6):326-34.

250. Balleza E, López-Bojorquez LN, Martínez-Antonio A, Resendis-Antonio O, Lozada-Chávez I, Balderas-Martínez YI, et al. Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol Rev.* 2009;33(1):133-51.
251. Wetzel D, McBride SM, Ercolini D. The Impact of pH on *Clostridioides difficile* Sporulation and Physiology. *Applied and environmental microbiology.* 2020;86(4):e02706-19.
252. Lund P, Tramonti A, De Biase D. Coping with low pH: molecular strategies in neutralophilic bacteria. *FEMS Microbiol Rev.* 2014;38(6):1091-125.
253. Jahn N, Brantl S, Strahl H. Against the mainstream: the membrane-associated type I toxin BsrG from *Bacillus subtilis* interferes with cell envelope biosynthesis without increasing membrane permeability. *Molecular Microbiology.* 2015;98(4):651-66.
254. Wenzel M, Vischer NOE, Strahl H, Hamoen LW. Assessing Membrane Fluidity and Visualizing Fluid Membrane Domains in Bacteria Using Fluorescent Membrane Dyes. *Bio-protocol.* 2018;8(20):e3063.
255. Surve MV, Bhutda S, Datey A, Anil A, Rawat S, Pushpakaran A, et al. Heterogeneity in pneumolysin expression governs the fate of *Streptococcus pneumoniae* during blood-brain barrier trafficking. *PLoS Pathog.* 2018;14(7):e1007168.
256. Sturm A, Heinemann M, Arnoldini M, Benecke A, Ackermann M, Benz M, et al. The cost of virulence: retarded growth of *Salmonella Typhimurium* cells expressing type III secretion system 1. *PLoS Pathog.* 2011;7(7):e1002143.
257. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and Evolution in the Genome of *Clostridium difficile*. *Clinical Microbiology Reviews.* 2015;28(3):721-41.
258. PHE. *Clostridium difficile* Ribotyping Network (CDRN) for England and Northern Ireland 2015-2018. 2019.
259. Marujo V, Arvand M. The largely unnoticed spread of *Clostridioides difficile* PCR ribotype 027 in Germany after 2010. *Infection Prevention in Practice.* 2020;2(4):100102.
260. Tamez-Torres KM, Torres-González P, Leal-Vega F, García-Alderete A, López García NI, Mendoza-Aguilar R, et al. Impact of *Clostridium difficile* infection caused by the NAP1/RT027 strain on severity and recurrence during an outbreak and transition to endemicity in a Mexican tertiary care center. *International journal of infectious diseases: IJID : official publication of the International Society for Infectious Diseases.* 2017;65:44-9.
261. Freeman J, Vernon J, Pilling S, Morris K, Nicholson S, Shearman S, et al. The ClosER study: results from a three-year pan-European longitudinal surveillance of antibiotic resistance among prevalent *Clostridium difficile* ribotypes, 2011-2014. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases.* 2018;24(7):724-31.
262. Shaw HA, Preston MD, Vendrik KEW, Cairns MD, Browne HP, Stabler RA, et al. The recent emergence of a highly related virulent *Clostridium difficile* clade with unique characteristics. *Clinical Microbiology and Infection.* 2020;26(4):492-8.
263. Bauer MP, Notermans DW, van Benthem BH, Brazier JS, Wilcox MH, Rupnik M, et al. *Clostridium difficile* infection in Europe: a hospital-based survey. *Lancet.* 2011;377(9759):63-73.
264. Merrigan M, Venugopal A, Mallozzi M, Roxas B, Viswanathan VK, Johnson S, et al. Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as robust toxin production. *J Bacteriol.* 2010;192(19):4904-11.
265. Akerlund T, Persson I, Unemo M, Norén T, Svenungsson B, Wullt M, et al. Increased sporulation rate of epidemic *Clostridium difficile* Type 027/NAP1. *J Clin Microbiol.* 2008;46(4):1530-3.
266. Valiente E, Dawson LF, Cairns MD, Stabler RA, Wren BW. Emergence of new PCR ribotypes from the hypervirulent *Clostridium difficile* 027 lineage. *J Med Microbiol.* 2012;61(Pt 1):49-56.

267. Knight DR, Imwattana K, Kullin B, Guerrero-Araya E, Paredes-Sabja D, Didelot X, et al. Major genetic discontinuity and novel toxigenic species in *Clostridioides difficile* taxonomy. *eLife*. 2021;10:e64325.
268. Dodd D, Spitzer MH, Van Treuren W, Merrill BD, Hryckowian AJ, Higginbottom SK, et al. A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature*. 2017;551(7682):648-52.
269. Kriek M, Martins F, Challand MR, Croft A, Roach PL. Thiamine biosynthesis in *Escherichia coli*: identification of the intermediate and by-product derived from tyrosine. *Angew Chem Int Ed Engl*. 2007;46(48):9223-6.
270. Dawson LF, Donahue EH, Cartman ST, Barton RH, Bundy J, McNerney R, et al. The analysis of *para*-cresol production and tolerance in *Clostridium difficile* 027 and 012 strains. *BMC Microbiology*. 2011;11(1):86.
271. Drudy D, Harnedy N, Fanning S, Hannan M, Kyne L. Emergence and control of fluoroquinolone-resistant, toxin A-negative, toxin B-positive *Clostridium difficile*. *Infection control and hospital epidemiology*. 2007;28(8):932-40.
272. Bradshaw WJ, Bruxelle JF, Kovacs-Simon A, Harmer NJ, Janoir C, Péchiné S, et al. Molecular features of lipoprotein CD0873: A potential vaccine against the human pathogen *Clostridioides difficile*. *J Biol Chem*. 2019;294(43):15850-61.
273. Hussain HA, Roberts AP, Mullany P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630 Δ erm) and demonstration that the conjugative transposon Tn916 Δ E enters the genome of this strain at multiple sites. *J Med Microbiol*. 2005;54(Pt 2):137-41.
274. Dubois T, Dancer-Thibonnier M, Monot M, Hamiot A, Bouillaut L, Soutourina O, et al. Control of *Clostridium difficile* Physiopathology in Response to Cysteine Availability. *Infection and immunity*. 2016;84(8):2389-405.
275. Stabler RA, Dawson LF, Valiente E, Cairns MD, Martin MJ, Donahue EH, et al. Macro and micro diversity of *Clostridium difficile* isolates from diverse sources and geographical locations. *PLoS One*. 2012;7(3):e31559.
276. Tremblay YDN, Durand BAR, Hamiot A, Martin-Verstraete I, Oberkamp M, Monot M, et al. Metabolic adaption to extracellular pyruvate triggers biofilm formation in *Clostridioides difficile*. *The ISME Journal*. 2021.
277. Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, et al. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular microbiology*. 2002;46(2):439-52.

Appendices

Table 1. Bacterial strains used in this study

Strain	Relevant features	Source
<i>Clostridium difficile</i>		
630 Δ <i>erm</i>	Erythromycin sensitive strain of 630 and a Clade 1 strain	(273)
630 Δ <i>erm</i> P _{hpdB} -SNAP	630 Δ <i>erm</i> with plasmid P _{hpdB} -SNAP	This study
630 Δ <i>erm</i> P _{hpdB} -SNAP P1mut	630 Δ <i>erm</i> with plasmid P _{hpdB} -SNAP with the P1 -10 site mutated by SDM	This study
630 Δ <i>erm</i> P _{hpdB} -SNAP P2mut	630 Δ <i>erm</i> with plasmid P _{hpdB} -SNAP with the P2 -10 site mutated by SDM	This study
630 Δ <i>erm</i> P _{hpdB} -SNAP P1mut P2mut	630 Δ <i>erm</i> with plasmid P _{hpdB} -SNAP with the P1 and P2 -10 sites mutated by SDM	This study
630 Δ <i>erm</i> P _{fdx} -SNAP	630 Δ <i>erm</i> with plasmid P _{fdx} -SNAP	This study
630 Δ <i>erm</i> P _{hpdB} - <i>gusA</i>	630 Δ <i>erm</i> with plasmid P _{hpdB} - <i>gusA</i>	This study
630 Δ <i>erm</i> P _{hpdB} - <i>phoZ</i>	630 Δ <i>erm</i> with plasmid P _{hpdB} - <i>phoZ</i>	This study
630 Δ <i>erm</i> P _{hpdB} - <i>phoZ</i> Δ IR	630 Δ <i>erm</i> P _{hpdB} - <i>phoZ</i> with the inverted repeat upstream of <i>sinR</i> removed	This study
630 Δ <i>erm</i> P _{hpdB} - <i>phoZ</i> Δ 5'IR	630 Δ <i>erm</i> P _{hpdB} - <i>phoZ</i> with the 5' arm of inverted repeat upstream of <i>sinR</i> removed	This study
630 Δ <i>erm</i> P _{fdx} - <i>phoZ</i>	630 Δ <i>erm</i> with plasmid P _{fdx} - <i>phoZ</i>	This study
CDIP217	630 Δ <i>erm</i> <i>sigL::erm</i>	(274)
R20291	Clade 2 strain	(275)
CD305	Clade 3 strain	(275)
M68	Clade 4 strain	(275)
M120	Clade 5 strain	(262)
630 Δ <i>erm</i> <i>sigL::erm</i> P _{hpdB} -SNAP	630 Δ <i>erm</i> <i>sigL::erm</i> with plasmid P _{hpdB} -SNAP	This study
630 Δ <i>erm</i> P _{sinR} -short-SNAP	630 Δ <i>erm</i> with plasmid P _{sinR} -short-SNAP	This study
630 Δ <i>erm</i> P _{sinR} -short-SNAP P ₁ mut	630 Δ <i>erm</i> with plasmid P _{sinR} -short-SNAP with the P1 -10 site mutated by SDM	This study
630 Δ <i>erm</i> P _{sinR} -long-SNAP	630 Δ <i>erm</i> with plasmid P _{sinR} -long-SNAP	This study
630 Δ <i>erm</i> P _{sinR} -long-SNAP Δ IR	630 Δ <i>erm</i> with plasmid P _{sinR} -short-SNAP with the inverted repeated upstream of <i>sinR</i> removed	This study
630 Δ <i>erm</i> P _{sinR} -long- <i>phoZ</i>	630 Δ <i>erm</i> with plasmid P _{sinR} -long- <i>phoZ</i>	This study
630 Δ <i>erm</i> P ₂₂₁₅ -short	630 Δ <i>erm</i> with plasmid P ₂₂₁₅ -short	This study
630 Δ <i>erm</i> P ₂₂₁₅ -long	630 Δ <i>erm</i> with plasmid P ₂₂₁₅ -long	This study
630 Δ <i>erm</i> Δ <i>codY</i> P _{sinR} -long-SNAP Δ <i>codY</i>	630 Δ <i>erm</i> Δ <i>codY</i> with P _{sinR} -long-SNAP-putative <i>CodY</i> binding site removed from reporter by SDM	This study

630 Δ erm <i>sinR</i> -CD2215::CT	Insertional inactivation of <i>sinR</i> leading to inactivation of <i>CD2215</i> to generate a double mutant	This study
630 Δ erm <i>sinR</i> -CD2215::CT P _{<i>sinR</i>} -short-SNAP	630 Δ erm <i>sinR</i> -CD2215::CT with plasmid P _{<i>sinR</i>} -short-SNAP	This study
630 Δ erm <i>sinR</i> -CD2215::CT P _{<i>sinR</i>} -long-SNAP	630 Δ erm <i>sinR</i> -CD2215::CT with plasmid P _{<i>sinR</i>} -long-SNAP	This study
630 Δ erm <i>sinR</i> -CD2215::CT P _{<i>sinR</i>} -short-SNAP P ₁ mut	630 Δ erm <i>sinR</i> -CD2215::CT with plasmid P _{<i>sinR</i>} -short-SNAP with the P ₁ -10 site mutated by SDM	This study
630 Δ erm <i>sinR</i> -CD2215::CT P _{<i>sinR</i>} -long-SNAP Δ IR	630 Δ erm <i>sinR</i> -CD2215::CT with plasmid P _{<i>sinR</i>} -long-SNAP with the inverted repeated upstream of <i>sinR</i> removed	This study
630 Δ erm <i>sinR</i> -CD2215::CT P ₂₂₁₅ -short	630 Δ erm <i>sinR</i> -CD2215::CT with plasmid P ₂₂₁₅ -short	This study
630 Δ erm <i>sinR</i> -CD2215::CT P ₂₂₁₅ -long	630 Δ erm <i>sinR</i> -CD2215::CT with plasmid P ₂₂₁₅ -long	This study
630 Δ erm <i>sinR</i> -CD2215:CT P _{tet} - <i>sinR</i>	630 Δ erm <i>sinR</i> -CD2215:CT complemented by a plasmid carrying ATc inducible <i>sinR</i>	This study
630 Δ erm <i>sinR</i> -CD2215::CT P _{tet} - <i>sinR</i> -CD2215	630 Δ erm <i>sinR</i> -CD2215::CT complemented by a plasmid carrying ATc inducible <i>sinR</i> and <i>CD2215</i>	This study
630 Δ erm <i>CD2215</i> ::CT	Insertional inactivation of <i>CD2215</i>	This study
630 Δ erm <i>CD2215</i> ::CT P _{<i>sinR</i>} -short-SNAP	630 Δ erm <i>CD2215</i> ::CT with plasmid P _{<i>sinR</i>} -short-SNAP	This study
630 Δ erm <i>CD2215</i> ::CT P _{<i>sinR</i>} -long-SNAP	630 Δ erm <i>CD2215</i> ::CT with plasmid P _{<i>sinR</i>} -long-SNAP	This study
630 Δ erm <i>CD2215</i> ::CT P _{<i>sinR</i>} -short-SNAP P ₁ mut	630 Δ erm <i>CD2215</i> ::CT with plasmid P _{<i>sinR</i>} -short-SNAP with the P ₁ -10 site mutated by SDM	This study
630 Δ erm <i>CD2215</i> ::CT P _{<i>sinR</i>} -long-SNAP Δ IR	630 Δ erm <i>CD2215</i> ::CT with plasmid P _{<i>sinR</i>} -long-SNAP with the inverted repeated upstream of <i>sinR</i> removed	This study
630 Δ erm <i>CD2215</i> ::CT P _{tet} - <i>CD2215</i>	630 Δ erm <i>CD2215</i> ::CT complemented by a plasmid carrying ATc inducible <i>CD2215</i>	This study
630 Δ erm <i>CD2215</i> ::CT P ₂₂₁₅ -short	630 Δ erm <i>CD2215</i> ::CT with plasmid P ₂₂₁₅ -short	This study
630 Δ erm <i>CD2215</i> ::CT P ₂₂₁₅ -long	630 Δ erm <i>CD2215</i> ::CT with plasmid P ₂₂₁₅ -long	This study
630 Δ erm <i>spo0A</i> ::CT	Insertional inactivation of <i>spo0A</i>	This study
630 Δ erm <i>spo0A</i> ::CT P _{<i>sinR</i>} -short-SNAP	630 Δ erm <i>spo0A</i> ::CT with plasmid P _{<i>sinR</i>} -short-SNAP	This study

630 Δ erm <i>spo0A</i> ::CT P _{sinR} -short-SNAP P ₁ mut	630 Δ erm <i>spo0A</i> ::CT with plasmid P _{sinR} -short-SNAP with the P ₁ -10 site mutated by SDM	This study
630 Δ erm <i>spo0A</i> ::CT P _{sinR} -long-SNAP	630 Δ erm <i>spo0A</i> ::CT with plasmid P _{sinR} -long-SNAP	This study
630 Δ erm <i>spo0A</i> ::CT P _{sinR} -long- <i>phoZ</i>	630 Δ erm <i>spo0A</i> ::CT with plasmid P _{sinR} -long- <i>phoZ</i>	This study
630 Δ erm <i>spo0A</i> ::CT P ₂₂₁₅ -short	630 Δ erm <i>spo0A</i> ::CT with plasmid P ₂₂₁₅ -short	This study
630 Δ erm <i>spo0A</i> ::CT P ₂₂₁₅ -long	630 Δ erm <i>spo0A</i> ::CT with plasmid P ₂₂₁₅ -long	This study
630 Δ erm <i>spo0A</i> ::CT P _{sinR} -long- <i>phoZ</i>	630 Δ erm <i>spo0A</i> ::CT carrying plasmid P _{sinR} -long- <i>phoZ</i>	This study
630 Δ erm (Paris)	Parent strain for Δ <i>sinR</i> , Δ <i>codY</i> and Δ <i>ccpA</i>	(276)
630 Δ erm (Paris) P _{sinR} -long-SNAP	630 Δ erm (Paris) carrying plasmid P _{sinR} -long-SNAP	This study
630 Δ erm (Paris) P _{sinR} -long- <i>phoZ</i>	630 Δ erm (Paris) carrying plasmid P _{sinR} -long- <i>phoZ</i>	This study
630 Δ erm Δ <i>sinR</i>	Deletion mutant of <i>sinR</i> made by allele exchange (CDIP615)	This study
630 Δ erm Δ <i>sinR</i> P _{sinR} -long-SNAP Δ IR	630 Δ erm Δ <i>sinR</i> carrying plasmid P _{sinR} -long-SNAP Δ IR	This study
630 Δ erm Δ <i>sinR</i> P _{tet} - <i>sinR</i>	630 Δ erm Δ <i>sinR</i> carrying complement plasmid P _{tet} - <i>sinR</i> with ATc inducible <i>sinR</i>	This study
630 Δ erm Δ <i>codY</i>	Deletion of <i>codY</i> (CDIP1341)	(276)
630 Δ erm Δ <i>codY</i> P _{sinR} -long-SNAP	630 Δ erm Δ <i>codY</i> with plasmid P _{sinR} -long-SNAP	This study
630 Δ erm Δ <i>codY</i> P _{sinR} -long-SNAP Δ <i>codY</i>	630 Δ erm Δ <i>codY</i> with plasmid CT P _{sinR} -long-SNAP with putative CodY binding site removed	This study
630 Δ erm Δ <i>ccpA</i>	Deletion of <i>ccpA</i> (CDIP1335)	(276)
630 Δ erm Δ <i>ccpA</i> P _{sinR} -long- <i>phoZ</i>	630 Δ erm Δ <i>ccpA</i> with plasmid P _{sinR} -long- <i>phoZ</i>	This study
630 Δ erm P _{sinR} -SinR-His	630 Δ erm with plasmid P _{sinR} -SinR-His	This study
630 Δ erm P _{sinR} -CD2215-His	630 Δ erm with plasmid P _{sinR} -CD2215-His	This study
<i>E. coli</i>		
CA434	Conjugation donor	(277)
Top10	Cloning	Thermofisher
C2987H	Cloning	NEB
NEB5 α P _{hpdB} -SNAP	NEB5 α carrying plasmid P _{hpdB} -SNAP	This study
NEB5 α P _{hpdB} -SNAP P1mut	NEB5 α carrying plasmid P _{hpdB} -SNAP P1mut	This study
NEB5 α P _{hpdB} -SNAP P2mut	NEB5 α carrying plasmid P _{hpdB} -SNAP P2mut	This study

NEB5α P _{hpdB} -SNAP P1mut P2mut	NEB5α carrying plasmid P _{hpdB} -SNAP P1mut P2mut	This study
NEB5α P _{hpdB} -gusA	NEB5α carrying plasmid P _{hpdB} -gusA	This study
NEB5α P _{hpdB} -phoZ	NEB5α carrying plasmid P _{hpdB} -phoZ	This study
NEB5α P _{fdx} -phoZ	NEB5α carrying plasmid P _{fdx} -phoZ	This study
NEB5α P _{sinR} -short-SNAP	NEB5α carrying plasmid P _{sinR} -short-SNAP	This study
NEB5α P _{sinR} -long-SNAP	NEB5α carrying plasmid P _{sinR} -long-SNAP	This study
NEB5α P _{sinR} -long-phoZ	NEB5α carrying plasmid P _{sinR} -long-phoZ	This study
NEB5α P ₂₂₁₅ -short	NEB5α carrying plasmid P _{int-1}	This study
NEB5α P ₂₂₁₅ -long	NEB5α carrying plasmid P _{int-2}	This study
NEB5α P _{sinR} -long-SNAPΔIR	NEB5α carrying plasmid P _{sinR} -long-SNAPΔIR	This study
NEB5α P _{sinR} -long-SNAPΔ5'IR	NEB5α carrying plasmid P _{sinR} -long-SNAPΔ5'IR	This study
NEB5α P _{sinR} -long-SNAPΔ3'IR	NEB5α carrying plasmid P _{sinR} -long-SNAPΔ3'IR	This study
NEB5α P _{sinR} -long-SNAPΔcodY	NEB5α carrying plasmid P _{sinR} -long-SNAPΔcodY	This study
NEB5α P _{sinR} -SinR-His	NEB5α carrying plasmid P _{sinR} -SinR-His	This study
NEB5α P _{sinR} -CD2215-His	NEB5α carrying plasmid P _{sinR} -CD2215-His	This study
CA434 P _{hpdB} -SNAP	CA434 carrying plasmid P _{hpdB} -SNAP	This study
CA434 P _{hpdB} -SNAP P1mut	CA434 carrying plasmid P _{hpdB} -SNAP P1mut	This study
CA434 P _{hpdB} -SNAP P2mut	CA434 carrying plasmid P _{hpdB} -SNAP P2mut	This study
CA434 P _{hpdB} -SNAP P1mut P2mut	CA434 carrying plasmid P _{hpdB} -SNAP P1mut P2mut	This study
CA434 P _{hpdB} -gusA	CA434 carrying plasmid P _{hpdB} -gusA	This study
CA434 P _{hpdB} -phoZ	CA434 carrying plasmid P _{hpdB} -phoZ	This study
CA434 P _{fdx} -phoZ	CA434 carrying plasmid P _{fdx} -phoZ	This study
CA434 P _{sinR} -short-SNAP	CA434 carrying plasmid P _{sinR} -short-SNAP	This study
CA434 P _{sinR} -long-SNAP	CA434 carrying plasmid P _{sinR} -long-SNAP	This study
CA434 P _{sinR} -long-phoZ	CA434 carrying plasmid P _{sinR} -long-phoZ	This study
CA434 P ₂₂₁₅ -short	CA434 carrying plasmid P _{int-1}	This study
CA434 P ₂₂₁₅ -long	CA434 carrying plasmid P _{int-2}	This study
CA434 P _{sinR} -long-SNAPΔIR	CA434 carrying plasmid P _{sinR} -long-SNAPΔIR	This study
CA434 P _{sinR} -short-SNAP ΔP ₁	CA434 carrying plasmid P _{sinR} -short-SNAP ΔP ₁	This study
CA434 P _{sinR} -long-SNAPΔcodY	CA434 carrying plasmid P _{sinR} -long-SNAPΔcodY	This study
CA434 P _{sinR} -SinR-His	CA434 carrying plasmid P _{sinR} -SinR-His	This study
CA434 P _{sinR} -CD2215-His	CA434 carrying plasmid P _{sinR} -CD2215-His	This study
Commensal strains from gut soup model		
<i>Escherichia coli</i>	Gram-negative gut commensal	(223)

<i>Enterococcus faecium</i>	Gram-positive gut commensal	(223)
<i>Lactobacillus fermentum</i>	Gram-positive gut commensal	(223)
<i>Klebsiella oxytoca</i>	Gram-negative gut commensal	(223)
<i>Bifidobacterium adoscelentis</i>	Gram-positive gut commensal	(223)
<i>Proteus mirabilis</i>	Gram-negative gut commensal	(223)

Table 2. Plasmids used in this study

Plasmids		
<i>P_{hpdB}-SNAP</i>	pMTL84151 plasmid carrying a SNAP-tag under the control of the <i>hpdBCA</i> promoter region	This study
<i>P_{hpdB}-gusA</i>	pMTL84151 plasmid carrying <i>gusA</i> under the control of the <i>hpdBCA</i> promoter	This study
<i>P_{hpdB}-phoZ</i>	pMTL84151 plasmid carrying <i>phoZ</i> under the control of the <i>hpdBCA</i> promoter	This study
<i>P_{fdx}-SNAP</i>	pMTL84153 carrying SNAP-tag under control of the <i>fdx</i> promoter	Dr Alexandra Faulds-Pain
<i>P_{fdx}-phoZ</i>	pMTL84153 carrying <i>phoZ</i> under control of the <i>fdx</i> promoter	This study
<i>P_{hpdB}-SNAP P1mut</i>	pMTL84151 plasmid carrying a SNAP-tag under the control of the <i>hpdBCA</i> promoter region with a mutation of the P1 site from TAT to TGC	This study
<i>P_{hpdB}-SNAP P2mut</i>	pMTL84151 plasmid carrying a SNAP-tag under the control of the <i>hpdBCA</i> promoter region with a mutation of the P2 site from TAT to TGC	This study
<i>P_{hpdB}-SNAP P1mut P2mut</i>	pMTL84151 plasmid carrying a SNAP-tag under the control of the <i>hpdBCA</i> promoter region with mutation in both the P1 and P2 sites from TAT to TGC	This study
<i>P_{hpdB}-phoZΔIR</i>	630Δ <i>erm</i> with plasmid <i>P_{hpdB}-phoZ</i> with the inverted repeat upstream of <i>hpdB</i> removed	This study
<i>P_{hpdB}-phoZΔ5'IR</i>	630Δ <i>erm</i> with plasmid <i>P_{hpdB}-phoZ</i> with the 5' arm of the inverted repeat upstream of <i>hpdB</i> removed	This study
pMC358	Plasmid carrying <i>phoZ</i>	(172)
pRPF185	Plasmid carrying <i>gusA</i>	(171)
<i>P_{sinR}-short-SNAP</i>	pMTL84151 carrying a SNAP-tag under the control <i>sinR</i> Short upstream region	This study
<i>P_{sinR}-long-SNAP</i>	pMTL84151 carrying a SNAP-tag under the control <i>sinR</i> Long upstream region	This study
<i>P_{sinR}-long-phoZ</i>	pMTL84151 carrying a <i>phoZ</i> reporter under the control <i>sinR</i> Long upstream region	This study
<i>P₂₂₁₅-short</i>	pMTL84151 carrying a SNAP-tag under the control <i>CD2215</i> Short upstream region	This study

P ₂₂₁₅ -long	pMTL84151 carrying a SNAP-tag under the control CD2215 Long upstream region	This study
P _{sinR} -long-SNAP ΔIR	P _{sinR} -long-SNAP Long with the inverted repeat upstream of the start codon removed	This study
P _{sinR} -long-SNAP Δ5'IR	P _{sinR} -long-SNAP with the 5'arm of the inverted repeat upstream of the start codon removed	This study
P _{sinR} -long-SNAP Δ3'IR	P _{sinR} -long-SNAP with the 3'arm of the inverted repeat upstream of the start codon removed	This study
P _{sinR} -short-SNAP P ₁ mut	P _{sinR} -short-SNAP with the P1 -10 altered by SDM	This study
P _{sinR} -long-SNAP ΔcodY	P _{sinR} -long-SNAP with the putative codY binding site removed by inverse PCR	This study
P _{sinR} -SinR-His	pMTL84151 plasmid carrying the long promoter region and coding sequences of SinR and CD2215 with a His-tag at the 3' end of SinR	This study
P _{sinR} -CD2215-His	pMTL84151 plasmid carrying the long promoter region and coding sequences of SinR and CD2215 with a His-tag at the 3' end of CD2215	This study

Table 3. Primers used in this study

Primer	Sequence (5'-3')	Purpose
<i>hpdB</i> _SNAP_For_P5	GGAAGAAATGGATAAAGATTGTGAAA TGAAGAGAAC	Amplification of SNAP-tag– forward primer
SNAP_V_rev_P6	CCCGGGTACCGAGCTCGAATTACCCA AGTCCTGGTTTC	Amplification of SNAP-tag and for SOE-PCR - reverse primer
<i>hpdB</i> _V_For_P3	CCATATGACCATGATTACGAAGATCTG AATTCGATAGGG	Amplification of <i>hpdBCA</i> promoter region and SOE-PCR – forward primer
<i>hpdB</i> _SNAP_Rev_P4	AATCTTTATCCATTCTTCCCCTCCTTA ATC	Amplification of <i>hpdBCA</i> promoter region – reverse primer
<i>gusA</i> -F- P	TTACGTCCTGTAGAAACCCC	Amplification of <i>gusA</i> coding sequence– forward primer
<i>gusA</i> -R- P	TCATTGTTTGCCTCCCTG	Amplification of <i>gusA</i> coding sequence – reverse primer
<i>gusA</i> vector forward	ATTCGAGCTCGGTACCCC	Inverse PCR of P _{<i>hpdB</i>} -SNAP for construction of P _{<i>hpdB</i>} - <i>gusA</i> – forward primer
<i>gusA</i> vector reverse	CATTTCTTCCCCTCCTTAATC	Inverse PCR of P _{<i>hpdB</i>} -SNAP for construction of P _{<i>hpdB</i>} - <i>gusA</i> – reverse primer
<i>phoZ hpdB</i> Vec F	AAAAGCAGAAATTCGAGCTCGGTACC CG	Inverse PCR of P _{<i>hpdB</i>} -SNAP for construction of P _{<i>hpdB</i>} - <i>phoZ</i> – forward primer
<i>phoZ hpdB</i> Vec R	ACATTGACGGCATTCTTCCCCTCCTTA ATCTTTC	Inverse PCR of P _{<i>hpdB</i>} -SNAP for construction of P _{<i>hpdB</i>} - <i>phoZ</i> – reverse primer
<i>phoZ</i> F	GGAAGAAATGCCGTCAATGTATGGGT AG	Amplification of <i>phoZ</i> coding sequence– forward primer
<i>phoZ</i> R	GAGCTCGAATTTCTGCTTTTTCTTCATT TTG	Amplification of <i>phoZ</i> coding sequence – reverse primer
84153 <i>phoZ</i> F	AAAAGCAGAAGGATCCTCTAGAGTCG AC	Inverse PCR of pMTL84153 with overhangs for <i>phoZ</i> –forward primer
84153 <i>phoZ</i> R	ACATTGACGGCATATGTAACACACCTC C	Inverse PCR of pMTL84153 with overhangs for <i>phoZ</i> –reverse primer
<i>phoZ</i> 84153 F	GTTACATATGCCGTCAATGTATGGGTA G	Amplification of <i>phoZ</i> from pMC358 with overhangs for ligation in to pMTL84153 – forward primer

<i>phoZ</i> 84153 R	TAGAGGATCCTTCTGCTTTTCTTCATT TTG	Amplification of <i>phoZ</i> from pMC358 with overhangs for ligation in to pMTL84153 – reverse primer
<i>hpdB</i> P1 SDM F	TTGCACTAATTATAGAAAGATTAAGG A	For mutation of the P1 site on any of the reporter plasmids – forward primer
<i>hpdB</i> P1 SDM R	TAGTCGAAAACCTTTTAAAGAATGAAAA A	For mutation of the P1 site on any of the reporter plasmids –reverse primer
<i>hpdB</i> P2 SDM F	TTCTGCAGAAAGATTATTTTAAAAGT	For mutation of the P2 site on any of the reporter plasmids – forward primer
<i>hpdB</i> P2 SDM R	TTCTCGAGAAAAAATTAACCTTGAA	For mutation of the P1 site on any of the reporter plasmids – reverse primer
<i>sinR</i> P2F	GTAGTATGGCTATGAAATTATGAAAAT ATTT	Primer used to amplify <i>sinR</i> short sequence - forward
<i>sinR</i> Pro Rev	TTATTATCCCTCCACTTTAGATTATATT C	Primer used to amplify <i>sinR</i> -short sequence - reverse
<i>sinR</i> Long SNAP vec F	GGGATAATAAATGGATAAAGATTGTG AAATGAAG	Vector amplification for <i>sinR</i> -long - forward
<i>sinR</i> Long SNAP vec R	AGGTAAACATTCGTAATCATGGTCATA TGG	Vector amplification for <i>sinR</i> -long - reverse
<i>sinR</i> Long Gib F	ATGATTACGAATGTTTACCTTACCAAT ATAATGATTAAC	Used to <i>sinR</i> -long sequence – forward
<i>sinR</i> Long Gib R	CTTTATCCATTTATTATCCCTCCACTTTA GATTATATTC	Used to <i>sinR</i> -long sequence – reverse
<i>CD2215</i> Short Gib F	ATGATTACGAGATAACTCTGTACCTTT AGAATG	P ₂₂₁₅ -short insert sequence - forward
<i>CD2215</i> Short Gib R	CTTTATCCATTATTGTTACTATTTCCCTC AC	P ₂₂₁₅ -short insert sequence – reverse
<i>CD2215</i> Short Vec F	AGTAACAATAATGGATAAAGATTGTG AAATGAAG	P ₂₂₁₅ -short vector sequence – forward
<i>CD2215</i> Short Vec R	CAGAGTTATCTCGTAATCATGGTCATA TG	P ₂₂₁₅ -short vector sequence – reverse
<i>CD2215</i> Long Gib F	ATGATTACGACTACTGCTTTAGACATA CC	P ₂₂₁₅ -long insert sequence – forward
<i>CD2215</i> Long Gib R	CTTTATCCATTATTGTTACTATTTCCCTC AC	P ₂₂₁₅ -long insert sequence – reverse
<i>CD2215</i> Long Vec F	AGTAACAATAATGGATAAAGATTGTG AAATGAAG	P ₂₂₁₅ -long vector sequence – forward
<i>CD2215</i> Long Vec R	AAAGCAGTAGTCGTAATCATGGTCATA TG	P ₂₂₁₅ -long vector sequence – reverse
<i>sinR</i> P1 SDM F	GAATGCAATCTAAAGTGGAGGG	SDM of <i>sinR</i> -CD2215 P1 -10 site - forward

<i>sinR</i> P1 SDM R	GATTGCATTCATAGTTTAAAGTTGTTTTA TA	SDM of <i>sinR</i> -CD2215 P1 -10 site - reverse
<i>sinR</i> CodY Rem F	TTCTACATATCTAATATGTAATTACAAT AAAAAATG	Removal of the putative codY binding site from P _{<i>sinR</i>} -long - forward
<i>sinR</i> CodY Rem R	AAAAAAATTTTCTATTTTATATATTGAA CAATTTATG	Removal of the putative codY binding site from P _{<i>sinR</i>} -long - forward
<i>sinR</i> 5' IR removal F	TATTTAGACTAAAAAAGTATAAATAGT CTATA	Removal of the 5' arm upstream of <i>sinR</i> start codon – Forward
<i>sinR</i> 5' IR removal R	TATATGTTTTTATACTGAAATAAGTGC	Removal of the 5' arm or the entire inverted repeat upstream of <i>sinR</i> 's start codon – Reverse
<i>sinR</i> 3' IR removal F	AAAAAGTATAAATAGTCTATATAAAC AACTTAAAC	Removal of the 3' arm or the entire inverted repeat upstream of <i>sinR</i> 's start codon – Forward
<i>sinR</i> 3' IR removal Rev	AATATAGACTATATATGTTTTTATACTG AAATAAGTGC	Removal of the 3' arm upstream of <i>sinR</i> 's start codon – Reverse
<i>hpdB</i> IR Rem F	TTTCATTCTTAAAAAGTTTTATACTAAT TATAGAAAG	Removal of entire inverted repeat upstream of <i>hpdB</i> – Forward primer
<i>hpdB</i> 5' IR Rem F	TAATATACCCTTTTTTTTCATTCTTAAA AAG	Removal of 5' arm of inverted repeat upstream of <i>hpdB</i> – Forward primer
<i>hpdB</i> IR Rem R	AAAATAATCTTTCTATAGAAAAAATTA AACTTGAAGC	Removal of 5'arm or entire inverted repeat upstream of <i>hpdB</i> – Reverse
16S qRT-PCR F	GGCAGCAGTGGGGAATATTG	Used for qRT-PCR of 16S
16S qRT-PCR R	CCGTAGCCTTTCACTCCTGA	Used for qRT-PCR of 16S
<i>hpdC</i> qRT-PCR F	GGATGCAACCAAAGGAATTTGT	Used for qRT-PCR of <i>hpdC</i>
<i>hpdC</i> qRT-PCR R	ACCCAGTCTTCTTTCTCTAGGC	Used for qRT-PCR of <i>hpdC</i>

Table 4. Components of growth media used in this study

Growth media component	Supplier
Brain Heart Infusion	Oxoid
Yeast extract	Sigma Aldrich
L-Cysteine	Sigma Aldrich
Thiamphenicol	Sigma Aldrich
Chloramphenicol	Sigma Aldrich
Luria-Bertani broth	Oxoid
Luria-Bertani agar	Oxoid
SOC Medium	Thermofisher
<i>C. difficile</i> supplement (Cefoxitin / D-Cycloserine)	Oxoid
Cas-amino acids	VWR
Histidine	Sigma Aldrich
Tryptophan	Sigma Aldrich
Glycine	Sigma Aldrich
Arginine	Sigma Aldrich
Phenylalanine	Sigma Aldrich
Methionine	Sigma Aldrich
Threonine	Sigma Aldrich
Alanine	Sigma Aldrich
Lysine	Sigma Aldrich
Serine	Sigma Aldrich
Valine	Sigma Aldrich
Isoleucine	Sigma Aldrich
Aspartic acid	Sigma Aldrich
Leucine	Sigma Aldrich
Proline	Sigma Aldrich
Glutamic acid	Sigma Aldrich
Glucose	Sigma Aldrich
Potassium phosphate monobasic	Sigma Aldrich
Sodium chloride	Sigma Aldrich
Sodium phosphate dibasic	Sigma Aldrich
Sodium bicarbonate	Sigma Aldrich
Calcium chloride dehydrate	Sigma Aldrich
Manganese chloride tetrahydrate	Sigma Aldrich
Cobalt chloride hexahydrate	Sigma Aldrich
Magnesium chloride hexahydrate	Sigma Aldrich
Ammonium sulphate	Sigma Aldrich
Iron sulphate heptahydrate	Sigma Aldrich
Ca-D-pantothenate	Sigma Aldrich
Pyridoxine	Sigma Aldrich
d-biotin	Sigma Aldrich
Sodium taurocholate	Sigma Aldrich
Anhydrotetracycline	Sigma Aldrich

Table 5. Reagents used in this study

Reagents	Supplier
DNA Phusion	NEB
dNTPs	NEB
MyTaq DNA polymerase	Bioline
Agarose	Bioline
TAE buffer (50X)	Thermofisher Scientific
Monarch DNA Gel Extraction Kit	NEB
Monarch PCR and DNA Cleanup kit	NEB
GelRed	Biotium
6X DNA loading dye	NEB
DNA Hyperladder 1KB	Bioline
Chelex 100	Sigma Aldrich
Monarch Miniprep kit	NEB
RNA Protect	Qiagen
RNAPro	MPBio
Chloroform	Sigma Aldrich
Lysing Matrix B Tubes	MPBio
Ethanol	Fisher Scientific
DEPC-treated water	Ambion
Turbo DNase I	Thermofisher
RNasin plus RNase inhibitor	Promega
Magnesium sulphate	Sigma Aldrich
Sodium acetate	Ambion
RNeasy kit	Qiagen
Random primers	NEB
Superscript II	Thermofisher scientific
Superscript IV	Thermofisher scientific
DTT	Sigma Aldrich
Restriction enzymes	NEB
Antarctic phosphatase	NEB
T4 Ligase	NEB
Hifi assembly mastermix	NEB
Kinase-Ligase-DpnI kit	NEB
Kapa sybr fast kit	Kapa Biosystems
Odyssey® EMSA kit	LICOR
TBE Buffer	Sigma Aldrich
TMR-Star	NEB
MOPS buffer	Thermofisher
10% Bis-Tris Gels	Thermofisher
Sodium phosphate dibasic	Sigma Aldrich
Sodium phosphate monobasic	Sigma Aldrich
Potassium chloride	Sigma Aldrich
Magnesium sulphate heptahydrate	Sigma Aldrich
Beta-mercaptoethanol	Sigma Aldrich
Toluene	Sigma Aldrich
<i>p</i> -nitrophenyl- β -d-glucuronide	Sigma Aldrich
Tris base	Sigma Aldrich
Trizma	Sigma Aldrich
Magnesium sulphate	Sigma Aldrich
Zinc chloride	Sigma Aldrich
Hydrochloric acid	Sigma Aldrich

SDS	Sigma Aldrich
<i>p</i> -nitrophenyl phosphate	Sigma Aldrich
Potassium phosphate monobasic	Sigma Aldrich
<i>p</i> -HPA	Sigma Aldrich
Tyrosine	Sigma Aldrich
<i>p</i> -cresol	Sigma Aldrich
Phosphate assay kit	Abcam
Phosphate buffered saline tablets	VWR
Vectashield with DAPI	Vectorlabs
Clear nail polish	Superdrug
Imidazole	Sigma Aldrich
NiNTA resin	Qiagen
Whole milk powder	Sigma Aldrich
Mouse anti-His antibody	LICOR
Goat anti-mouse antibody	LICOR
Tween20	Sigma Aldrich

Glossary

Dysbiosis – an imbalance between the types and/or abundance of organisms present in an individual's average microbiome

***gusA* reporter** – reporter gene encoding a glucuronidase gene from *E. coli*

Microbiome – a characteristic microbial community occupying a reasonable well-defined habitat which has distinct physio-chemical properties

***para*-cresol** – antibacterial compound produced in high concentrations by *Clostridioides difficile* through the decarboxylation of the precursor *p*-HPA by the *p*-HPA decarboxylase enzyme encoded on the *hpdBCA* operon

***para*-hydroxyphenylacetate** – metabolite of tyrosine produced by a range of bacterial species including *Clostridium* and *Klebsiella* species as well as by mammalian cells

***p*-HPA decarboxylase** – enzyme encoded by the *hpdBCA* operon carried by *C. difficile* that converts *p*-HPA to *p*-cresol

***phoZ* reporter** – transcriptional reporter gene encoding alkaline phosphatase gene from *E. faecalis*

Pseudomembranous colitis – Inflammation of the colon characterised by raised plaques forming pseudomembranes on the mucosa

SNAP-tag reporter – gene that can be used as either a transcriptional or translational reporter

Spore – dormant, non-reproductive form of a bacteria that is resistant to a range of stresses, such as heat, oxygen stress or chemicals such as ethanol

Sporulation – process of vegetative bacteria forming into a spore

Transcriptional reporter – a reporter gene is fused to the promoter sequence of interest and used to determine expression from the promoter

Translational reporter – a reporter gene is fused to the protein coding sequence of interest and used to determine expression of the protein and/or its location within a cell