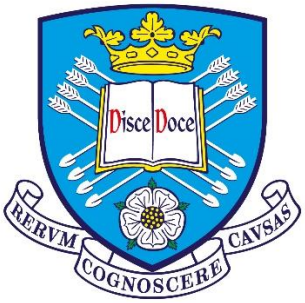



The University Of Sheffield.

# Life Without an S-layer:

# The Conditional Genome of *C. difficile*

**By:**

**Shauna O Beirne**

A dissertation presented for the degree of Doctor of Philosophy,

in the Faculty of Science

# Abstract

*Clostridium difficile* is a Gram-positive spore-forming obligate anaerobe and the most common cause of antibiotic-associated infectious diarrhoea worldwide. *C. difficile* infection (CDI) has increased in severity and incidence over the last decade, representing a huge economic burden to healthcare systems. Two toxins elaborated by *C. difficile* are widely regarded as the major causes of CDI symptoms. However, little research has focused on the assembly of the cell envelope, which is the first point of contact with the host. Vegetative *C. difficile* are covered in a proteinaceous paracrystalline array, known as the S-layer. Production of the S-layer comes at an enormous metabolic cost to the cell and functional analysis of the major surface layer protein (SlpA) has uncovered roles in pathogenicity and sporulation.

Using a *C. difficile* S-layer null strain, we examined the conditional genome of *C. difficile* in the absence of this major surface structure. This was achieved by creating a large library of mutants via insertional mutagenesis coupled to Transposon Directed Insertion Site Sequencing (TraDIS). We have identified 55 genes as conditionally essential and 20 genes as conditionally non-essential in the absence of SlpA. For example, we have demonstrated that *secA2*, encoding the SlpA translocase is dispensable for growth in an S-layer mutant. Interestingly, we have also demonstrated that in the absence of SlpA, several PSII biosynthetic genes, involved in S-layer surface attachment, are conditionally non-essential for growth *in vitro*.

As conditionally non-essential genes represent potential targets for disrupting S-layer assembly, we investigated the phenotypic effects of gene silencing via CRISPR interference of conditionally non-essential genes identified in this study. We provide evidence that knock-down of *secA2* in a WT background produces the same cell morphology as an S-layer mutant. Additionally, we show depletion of Era, a poly(A) polymerase and several PSII biosynthesis genes from cells has effects on cell shape, surface boundaries, membrane permeability and localisation of septas.

# Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Robert Fagan for his unwavering support and encouragement throughout my PhD. Thank you for letting me pursue my own projects but offering guidance when needed. I also owe a huge thank you to Dr. Joseph Kirk for all the advice he has given me along the way, and all he has taught me in the way of molecular biology. Thanks to both of you, and the endless fount of knowledge you possess, I am leaving a much more confident scientist.

I am hugely grateful for all members of the Fagan group, past and present. You all made it a fantastic place to work, and when times were challenging you all kept my spirits high. A million thanks to Hannah Fisher. We worked the infamous shift 2 pattern together, and you were always there for me to lean on through the tough times (and the great too), and for that I cannot thank you enough. A big thank you to Jason, another member of the shift 2 crew, you were extremely patient with me when I tried to explain my cloning woes, and for that I am extremely grateful. I would also like to thank Laurence, Jess, Sophie, Luz, Rob Smith, Jessie, and Josh. I have thoroughly enjoyed our obscure lunch time conversations, some of which I will never be able to erase from my mind. Ruairidh and Alex, we started our PhDs together, and there are no two other people I would rather finish them with. Without all the weekends spent playing Mario kart and drinking the strongest ever espresso martinis, I would have lost my mind a long time ago.

To Ian, you have shown infinite patience with me, sacrificing many weekends, especially all the Sunday nights (which I know is sacred NFL time) to drive me to the lab so I could set up overnights. Thank you for listening to all my ramblings about failed experiments. I don't think I could have bounced back from them if it wasn't for you.

I couldn't have gotten through all of this without my family. To my brother Robert, you have been a huge support for me throughout my writing process, saving me from many breakdowns with come dine with me marathons. To David, thank you for all the laughs you gave me as you danced to Maniac 2000 while I was studying. You forced me to not get too bogged down with writing, and for that I am grateful. Thank you to Marina and Karen, I am so happy you both joined our family. To my nephew Ethan, your infectious personality instantly rectified my mood after long writing days. I could always count on you to make me laugh, even after the most taxing days. I dedicate the work in this thesis to my parents. To my dad, thank you for all the sacrifices you have made, all the support you have given me, all the guidance you gave when I needed it most. Also a huge thank you for proofreading my thesis. And thank you Mum, whenever I felt overwhelmed you constantly reminded me on what is important in life, always lending a sympathetic ear. I couldn't have done this without all your support.

***"Wisdom begins in wonder"***

*-- Socrates*

# Table of Contents

# List of Figures

## List of Tables

# List of Abbreviations

| | |
|---|---|
| ATc | Anyhydrotetracycline |
| Av-CD | Avidocin-*C. difficile* |
| bEP | Bacterial enhancer proteins |
| BHI | Brain-heart infusion |
| Ca-DPA | Calcium Dipicolinic acid |
| *cas* | CRISPR-associated |
| CDI | *C. difficile* infection |
| CFU | Colony forming units |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| CRISPRa | CRISPR activation |
| CRISPRi | CRISPR interference |
| CROP | Combined repetitive oligopeptide |
| crRNA | CRISPR RNA |
| CTD | C-terminal domain |
| CWB2 | Cell wall binding domain 2 |
| CWP | Cell wall protein |
| DMSO | Dimethyl sulfoxide |
| DPA | Dipicolinic acid |
| DSB | Double stranded break |
| EDTA | Ethylene diaminetetraacetic acid |
| EtOH | Ethanol |
| FMT | Faecal microbiota transplantation |
| GDH | Glutamate dehydrogenase |
| gDNA | Genomic DNA |
| GI | Gastro intestinal |
| gRNA | Guide RNA |
| GTP | Guanosine triphosphate |
| HITS | High throughput insertion tracking by deep sequencing |
| HMW | High molecular weight |

| | |
|---|---|
| HRP | Horseradish peroxidase |
| IAP | Intracisternal A particle |
| INSeq | Insertion sequencing |
| KEGG | Kyoto encyclopedia of genes and genomes |
| LB | Luria-Bertani |
| LMW | Low molecular weight |
| NAAT | Nucleic acid amplification test |
| O/N | Overnight |
| OD | Optical density |
| ORF | Open reading frame |
| PaLoc | Pathogenicity locus |
| PAM | Protospacer adjacent motif |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PLG | Phase lock gel |
| PSII | Polysaccharide II |
| RBP | Receptor binding protein |
| RISC | RNA-induced silencing complex |
| RNAi | RNA interference |
| RNAP | RNA polymerase |
| SASP | Small acid soluble proteins |
| SCWP | Secondary cell wall polysaccharide |
| SDS-PAGE | Sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| sgRNA | Single guide RNA |
| S-layer | Surface layer |
| SLP | S-layer protein |
| SOC | Super optimal broth by catabolite repression |
| STM | Signature tagged mutagenesis |
| TALEN | Transcription activator-like effector nuclease |
| TEM | Transmission electron microscopy |
| TIR | Terminal inverted repeats |

| | |
|---|---|
| TLR4 | Toll-like receptor 4 |
| Tn-Seq | Transposon sequencing |
| tracrRNA | *trans*-acting CRISPR RNA |
| TraDIS | Transposon directed insertion site sequencing |
| TY | Tryptose, yeast extract |
| UTR | Untranslated region |
| WT | Wild-type |

# Chapter I. Introduction

## 1.1 *Clostridium difficile*

*Clostridium difficile,* a Gram-positive, spore-forming, obligate anaerobe, is the leading cause of antibiotic-associated infectious diarrhoea worldwide (Paredes-Sabja *et al*., 2014). This rod-shaped bacterium was originally named *Bacillus difficilis*, due to the difficulty isolating and culturing the microorganism (Hall and O'Toole, 1935). While *C. difficile* was initially identified as a commensal bacterium of the microflora of healthy infants in 1935, the bacterium was only recognised as a major causative agent of antibiotic associated diarrhoea in 1978 (Bartlett *et al*., 1978). Clinical presentations of *C. difficile* infection (CDI) are varied, ranging from mild self-limiting diarrhoea to more complicated life-threatening conditions such as pseudomembranous colitis and toxic megacolon, which are associated with high morbidity and mortality rates **(Fig 1.1)** (Hryckowian *et al*., 2017).

*C. difficile* has only gained the 'hospital superbug' moniker in the last few decades due to the emergence of epidemic strains (Bartlett *et al*., 1978; Hall and O'Toole, 1935; Smits *et al*., 2016). Until that point, understanding of *C. difficile* as a human pathogen remained limited. After 2003, high incidence of CDI and increasing mortality rates were observed in North America and across Europe attributed to the emergence of epidemic PCR ribotype 027 strains, and to a lesser extent PCR ribotype 078 (Goorhuis *et al*., 2008). The PCR ribotype 027 strains were associated with higher levels of toxin production, more severe diarrhoea, and higher levels of recurrent infections (O'Horo *et al*., 2014). Between 2004 and 2007, deaths in England and Wales caused by CDI rose from 2,238 to 8,324, more than five times the number of deaths caused by MRSA (Carter, 2009). From April 2003 to March 2006, 498 patients contracted CDI at Stoke Mandeville hospital in Buckinghamshire in one prolonged outbreak (Stabler *et al*., 2009). Of these 498, 127 patients died. In 2011, the US Centres for Disease Control and Prevention (CDC) estimated that there were approximately 500,000 cases of CDI annually in the US, resulting in 29,000 deaths (Banaei *et al*., 2015).

Generally, healthy individuals clear CDI and/or become asymptomatic carriers (Leffler and Lamont, 2015). It is estimated that asymptomatic carriage of *C. difficile* occurs in 2-18% of adults and this rate is higher in hospital patients (25-30%) (Shim *et al*., 1998). Under normal circumstances, the microbiota provides colonisation resistance against *C. difficile*, however, CDI cases can occur in healthcare settings where there are high rates of antibiotic usage (Smits *et al*., 2016). Broad-spectrum antibiotics will perturb the intestinal microflora, providing a niche for *C. difficile* to proliferate (Vincent and Manges, 2015). While all antibiotics can predispose individuals to CDI, some antibiotic classes pose greater risks than others, such as clindamycin, cephalosporins and fluoroquinolones (Slimings and Riley, 2014). However, even low risk antibiotics can perturb the intestinal microflora, especially when two or more antibiotics are prescribed to the patient (Smits *et al*., 2016). After antibiotic usage, the intestinal flora remains perturbed for ~ 3 months, meaning patients remain susceptible to developing CDI again (Smits *et al*., 2016). Unsurprisingly, recurrence of infection occurs in about 20-30% of patients (Tieu *et al*., 2019). Additional risk-factors for development of CDI include age, people with an inadequate immune response and poor hygiene practices (Eze *et al*., 2017).

## 1.2 *C. difficile* Pathogenesis

### 1.2.1 *C. difficile* Transmission

*C. difficile* is transmitted via the faecal-oral route (Czepiel *et al*., 2019). *C. difficile* can undergo a critical survival strategy known as sporulation, enabling the bacterium to persist in an aerobic environment (Huang and Hull, 2017). These spores are metabolically dormant particles that are highly resistant to environmental stressors, including disinfectants and many antimicrobials, which target metabolically active cells (Smits *et al*., 2016). Spores are ubiquitous in the environment and are regarded as the infectious vehicle as vegetative cells could not survive in an oxygenated environment or the harsh acidic conditions of the stomach (Deakin *et al*., 2012).

### 1.2.2 *C. difficile* Sporulation

Investigating the mechanism of *C. difficile* sporulation has relied on comparisons with the well-defined pathway of *Bacillus subtilis* (Shen *et al*., 2019). However, these studies have revealed that both species vary markedly in the process (Setlow *et al*., 2017). The decision to sporulate is governed by the master regulator, Spo0A, and is believed to be triggered by nutrient depletion and quorum sensing (Deakin *et al*., 2012). A classic phosphorelay pathway has been defined for *B. subtilis*, however, this pathway is absent in *C. difficile* (Al-Hinai *et al*., 2015). Instead, Spo0A is believed to be phosphorylated by five orphan histidine kinases that respond to unknown signals (Paredes-Sabja *et al*., 2014). Three of these, CD1492, CD1579 and CD2492, share sequence similarity with the phosphorelay sensor kinases of *B. subtilis*, and were initially proposed to be responsible for Spo0A phosphorylation (Underwood *et al*., 2009). However, only CD1579 was shown to directly phosphorylate Spo0A. While the role of CD2492 remains unclear, CD1492 was shown to inhibit sporulation initiation and to affect toxin production and motility through the regulatory proteins RstA and SigD (Childress *et al*., 2016).

*C. difficile* sporulation occurs in 4 stages **(Fig 1.1)**. First, asymmetric division of vegetative cells occurs creating a smaller compartment, known as the forespore, and a larger compartment, known as the mother cell (Crawshaw *et al*., 2014). Second, the septum separating the forespore and mother cell is degraded, and the mother cell engulfs the forespore in a phagocytic-like motion, resulting in a pre-spore in the mother cell cytoplasm that is enclosed within two membranes. The mother cell proteins SpoIID and SpoIIP provide part of the driving force for engulfment (Dembek *et al*., 2018). In the third morphogenic stage, the spore cortex and coat layers are assembled (Paredes-Sabja *et al*., 2014). Lastly, the mother cell lyses, releasing the mature spore (McKenney *et al*., 2013).

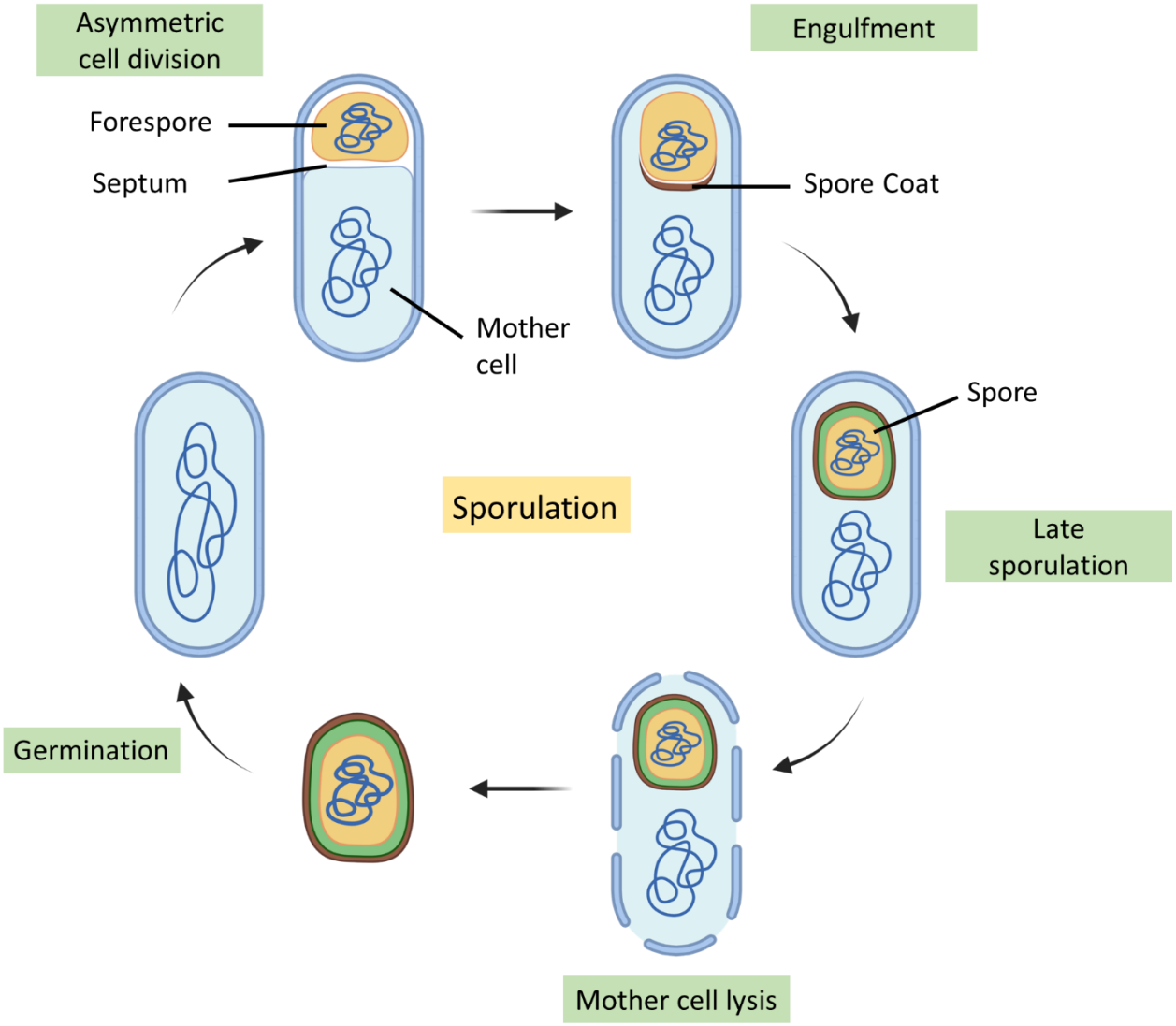



**Fig 1.1. Sporulation in *B. subtilis*.** Sporulation occurs in 4 stages. In the first stage, vegetative cells undergo asymmetric division, forming a smaller forespore and the larger mother cell. In the second step, the forespore is engulfed into the mother cell cytoplasm, and is enclosed within two membranes. The spore coat and cortex layers are assembled in the third stage. Lastly, the mother cell lyses, releasing the mature spore. When favourable conditions are encountered, spores will germinate.

*C. difficile* spores are composed of a series of concentric shells **(Fig 1.2)**. At the core of the *C. difficile* spores are spore DNA, RNA, ribosomes, and enzymes essential to return to metabolic activity during germination (Lawler *et al*., 2020). Within the core, DNA tightly associates with small acid soluble proteins (SASPs), altering the properties of the DNA, which contributes to spore resistance to

heat and chemicals (Paredes-Sabja *et al*., 2014). The core is also saturated with dipicolinic acid (DPA) and $Ca^{2+}$ ions (Setlow, 2007). DPA chelates calcium ions ($Ca^{2+}$-DPA), reducing the water content of the core and preserving the metabolically dormant state of the spore. The core is surrounded by a tight inner membrane, which exhibits low permeability to small molecules, including water. This feature provides resistance to DNA-damaging agents (Swick *et al*., 2016). External to the inner membrane is a layer of peptidoglycan (PG) comprised of two sub-layers: the cortex and germ cell wall, which vary in their PG compositions (Popham, 2002). The germ cell wall is derived from the asymmetric septum and mother cell envelope during engulfment and becomes the nascent cell wall during outgrowth (Lawler *et al*., 2020). Around this is a thick PG cortex and a second membrane. The primordial disaccharide of PG chains is alternating β-1,4-linked *N*-acetylglucosamine and *N*-acetylmuramic acid residues, cross-linked by 4,3 stem peptides (Atrih *et al*., 1999). *C. difficile* vegetative cell PG favours 3,3, linkages of catalysed by LD-transpeptidases (Peltier *et al*., 2011). For *C. difficile* cortex PG, every second *N*-acetylmuramic acid moiety is modified to muramic-δ-lactam (Coullon *et al*., 2018). Consequently, cortex PG has fewer stem-peptides, fewer cross-links and an overall more flexible structure. The cortex is devoted to maintaining the relatively dry spore state, through constriction of spore volume (Swick *et al*., 2016). The cortex is enclosed by a second membrane, which does not function as a permeability barrier in spore resistance (Lawler *et al*., 2020). The coat acts as a sieve, restricting passage of large molecules into the spore. In some *C. difficile* strains, the coat protein layers are surrounded by a multi-layered protein exosporium (Paredes-Sabja *et al*., 2014). The exosporium is believed to contribute to spore resistance as *C. difficile* lacking the exosporium-specific cysteine-rich protein CdeC display reduced resistance to ethanol, lysozyme, and heat treatment (Barra-Carrasco *et al*., 2013).

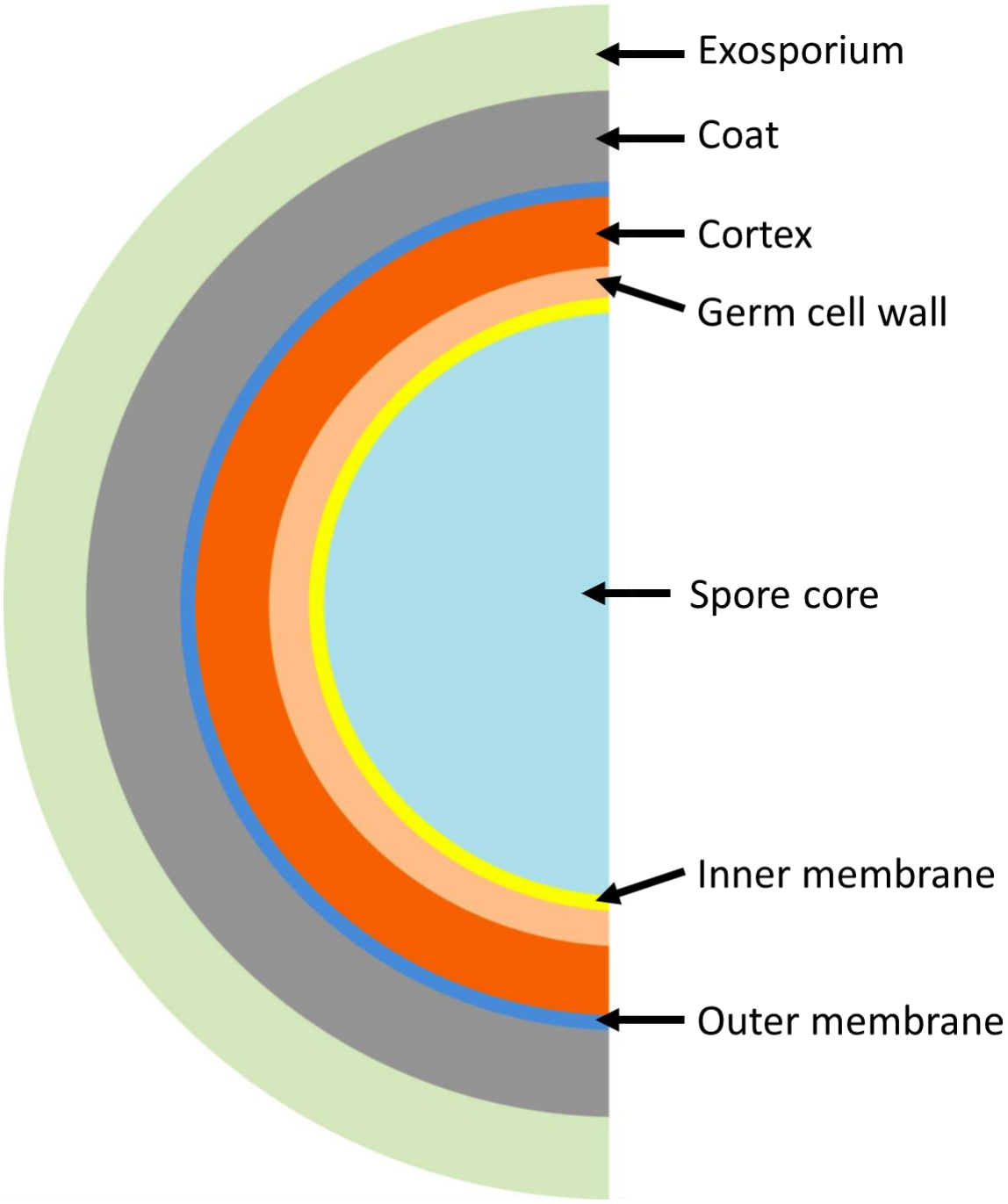



**Fig 1.2. *C. difficile* spore ultrastructure.** The core of *C. difficile* spores contains spore DNA, cations, dipicolinic acid (DPA) and calcium ions. Calcium ions and DPA chelate, constituting approx. 25% of the dry weight of the spore. External to the inner membrane is the germ cell wall, which is destined to become the cell wall of the resulting vegetative cell. Surrounding the germ cell wall, is the cortex, which is composed of a thick layer of modified PG and is required for spore germination. The outer-most spore layer is the exosporium. Figure taken with permission from Paredes-Sabja *et al*., 2014.

### 1.2.3 *C. difficile* Germination

*C. difficile* spores travel to the gastrointestinal (GI) tract and to the large intestine upon ingestion. Spores germinate in the GI tract and differentiate back to metabolically active dividing cells, which are capable of producing toxins and causing the distinct disease profile seen with CDI. For this to occur, spores interact with bile acids from the liver, initiating germination. This interaction results in the activation of a proteolytic cascade and degradation of the spore PG (Lawler *et al*., 2020). There are four types of bile salts: primary, secondary, conjugated and non-conjugated (Urdaneta and Casadesus, 2017). Primary bile salts, cholic acid and chenodeoxycholic acid are the immediate products of cholesterol degradation in the liver. After synthesis, these bile acids can be conjugated with taurine or glycine and are secreted into the duodenum (Ajouz *et al*., 2014). Bile acids then pass into the jejunum, where they are largely absorbed or recycled (Nagengast *et al*., 1995). Some bile acids pass through the anaerobic environment of the caecum, where some members of the microbiota express bile salt hydrolases, which remove the conjugated amino acid from the respective bile salt. Bacterial action, through 7α-dehydroxylation, on the two primary bile salts, results in the formation of secondary bile salts, namely deoxycholate and lithocholate (Ridlon *et al*., 2016). Notably, in the absence of some of the intestinal microbiota species, for example, elimination through use of broad-spectrum antibiotics, secondary bile-acid conversion is reduced, resulting in an increased concentration of *C. difficile* spore germinants (Lawler *et al*., 2020). Cholic acid derivatives positively modulate *C. difficile* spore germination, while chenodeoxycholic acid and its derivatives are inhibitors of germination (Sorg and Sonenshein, 2008, 2009).

*C. difficile* germination does not occur in the presence of a single germinant, with a co-germinant, glycine, also required (Sorg and Sonenshein, 2008). Recognition of germinants is mediated through the action of the germination receptor, CspC, located in the spore cortex or spore coat (Francis *et al*., 2013). Activation of CspC signals a second protease CspB to cleave the cortex hydrolase, pro-SleC to its active form, initiating cortex degradation (Adams *et al*., 2013). Following

this, release of $Ca^{2+}$-DPA occurs, spores are rehydrated, and outgrowth can occur (Francis *et al*., 2015). The incorporation of CspB and CspC into the mature spore is regulated by the activity of GerG (Lawler *et al*., 2020). Interestingly, deletion of CspB inhibits SleC processing, but does not abolish germination completely, suggesting other enzymes are involved in cortex hydrolysis (Lawler *et al*., 2020). Indeed, a second enzyme YabG converts prepro-SleC, to pro-SleC, and elimination of YabG prevents the incorporation of pro-SleC into the mature spore (Shrestha *et al*., 2019). Another protein, GerS regulates germination through modification to the PG cortex (Diaz *et al*., 2018). GerS mutants are unable to degrade the PG cortex and complete germination as these cortex-specific modifications are essential for recognition by SleC (Lawler *et al*., 2020).

### 1.2.4 Toxin Production

Once spores have germinated, vegetative cells can proliferate and release toxins that are widely regarded as the major causes of CDI disease symptoms (Smits *et al*., 2016). *tcdA* and *tcdB,* which encode two homologous toxins, are located within a 19.6 kb pathogenicity locus (PaLoc), alongside toxin regulatory genes *tcdC*, *tcdE* and *tcdR* (Dupuy *et al*., 2008; El Meouche *et al*., 2013; Hammond and Johnson, 1995; Monot *et al*., 2015; Tan *et al*., 2001). These toxins function as glucosyltransferases, transferring glucose from UDP-glucose to threonine residues of Rho and Ras guanosine triphosphatases (GTPases), such as Rho, Rac1 and Cdc42, in epithelial cells of the GI tract inhibiting the regulatory functions of its targets (Hall, 2012). TcdA and TcdB contain four domains: a Rho and Rac glucosyltransferase domain (GTD), a cysteine protease domain (CTD), a pore-forming delivery domain devoted to host membrane insertion, and a combined repetitive oligopeptide (CROP) domain, postulated to bind host cell surface receptors prior to endocytosis (Chandrasekaran and Lacy, 2017).

TcdA and TcdB enter host cells via receptor-mediated endocytosis using clathrin- and dynamin- dependent pathways (Papatheodorou *et al*., 2010). TcdA recognises carbohydrates and Gp96 on the apical surface of colonocytes (Na *et al*., 2008). Conversely, TcdB recognises poliovirus

receptor-like 3, expressed on human colonic epithelial cells, and chondroitin-sulfate proteoglycan-4 (LaFrance *et al*., 2015; Yuan *et al*., 2015). Once inside the cell, acidification of the endosome occurs, triggering a conformation change in the toxin, which inserts into the membrane and forms a pore (Awad *et al*., 2014). Cellular inositol hexakisphosphate triggers autocatalytic cleavage of TcdA and TcdB, which releases the GTD domain into the host cell cytosol (Di Bella *et al*., 2016). The GTD glucosylates the Rho GTPases; Rho, Rac and Cdc42 (Chen *et al*., 2015). These GTPases function as molecular switches controlling signal pathways involved in many host cellular processes (Hall, 2012). Glucosylation by these toxins locks the GTPase in its 'OFF' position, thereby inactivating them (Voth and Ballard, 2005). This results in actin depolymerisation, cell rounding and programmed cell death. Additionally, disruption to these cellular signalling pathways stimulates the release of pro-inflammatory cytokines (IL-1, IL-6 and TNF-α) and chemokines (CXCL1-2, CXCL5, IL-8, CCL2 and CCL7) from epithelial cells (Popoff, 2018). This causes disruption to cytoskeletal assembly, with inhibition leading to the disruption of tight junctions between epithelial cells, fluid secretion and programmed cell death (Shen, 2012).

TcdE, TcdR and TcdC are three regulators of toxin expression, also encoded by PaLoc (Martin-Verstraete *et al*., 2016). *tcdR* encodes an alternative sigma factor, required for transcription initiation of *tcdA*, *tcdB and tcdR*. TcdC is an anti-sigma factor, which negatively regulates toxin production, likely by sequestration of TcdR (Dupuy *et al*., 2008). In the epidemic ribotype 027, *tcdC* has an 18 bp deletion, and it is postulated that derepression of toxin production accounts for increased virulence observed with these strains (Carter *et al*., 2011). TcdE shares homology with holin proteins, however, its role in toxin production regulation is poorly understood. It is believed to be involved in toxin secretion (Govind and Dupuy, 2012).

## 1.3 Diagnosis of CDI

CDI can be diagnosed by detection of *C. difficile* products, including glutamate dehydrogenase (GDH), aromatic fatty acids and TcdA and/or TcdB (Bartlett and Gerding, 2008).

Additionally, toxigenic culture of *C. difficile*, nucleic acid amplification tests (NAAT) for 16S RNA, toxin genes and GDH genes are also regularly used. The European Society of Clinical Microbiology and Infectious Diseases, relies on a two-step algorithm for diagnosis, which involves testing for the presence of *C. difficile* as well as a test to detect free toxin in the faeces (Crobach *et al*., 2016). Absence of free toxin indicates that CDI is highly unlikely. However, toxin assays vary markedly in their sensitivity (Smits *et al*., 2016). As such, if *C. difficile* is present, but the free toxin is negative, CDI cannot be excluded, and clinical evaluation is required (Planche and Wilcox, 2015). Current therapy for CDI is complicated by the recurrence of infection, which occurs in up to 30% of patients (Johnson, 2009).

## 1.4 Treatment of CDI

### 1.4.1 Antibiotics

Multiple antimicrobials have been used for the treatment of CDI (Ooijevaar *et al*., 2018). Until recently, vancomycin and metronidazole were the front-line drugs for CDI treatment and have been prescribed orally for treatment since the 1970s (Fekety *et al*., 1989; Keighley *et al*., 1978; Silva *et al*., 1981). Metronidazole diffuses into *C. difficile* cells, blocking protein synthesis through interactions with DNA (Rineh *et al*., 2014). The consequence of this is breakdown of the helical DNA structure and strand breakage, resulting in cell death. This drug has shown to be effective in the treatment of initial non-severe cases of CDI (Cho *et al*., 2020). Due to the low cost, metronidazole was the mainstay drug for treatment of mild to moderate CDI, however, the majority of the drug is absorbed into the small intestine and only low concentrations make it back to the colon (Baines *et al*., 2008). Indeed, metronidazole concentrations in the faeces range from <0.25 to 9.5 mg/L, with concentrations decreasing to non-detectable levels once inflammation levels improve. Metronidazole is not suitable for all CDI patients and is not recommended for children and women during pregnancy or lactation (Passmore *et al*., 1988). However, now due to the association of

metronidazole with recurrent infection and increasing resistance, it is no longer recommended as the mainstay drug for treating mild CDI infections (McDonald *et al*., 2018).

Vancomycin is now the recommended first line drug for the treatment of mild to moderate CDI (Chiu *et al*., 2019). This glycopeptide antibiotic inhibits cell wall assembly, by blocking incorporation of the *N*-acetylglucosamine-*N*-acetylmuramic acid disaccharide into the PG matrix (Allen *et al*., 1996). Vancomycin is poorly absorbed in the GI tract, and reaches high concentration in the faeces, resolving *C. difficile* infection much quicker (Mullane, 2014). For non-severe cases of CDI, metronidazole and vancomycin have similar cure rates, 98% and 90%, respectively (Zar *et al*., 2007). The most marked difference was observed in patients with severe disease, with cure rates of 97% for vancomycin and 76% for metronidazole (Shen and Surawicz, 2008). While these drugs are effective in clearance of infection, both antimicrobials cause significant dysbiosis to the gut microflora, and contribute to the recurrence of infection observed in some patients (Smits *et al*., 2016). Thus, it is crucial that alternative methods of treatment are developed.

One such drug is fidaxomicin, a bactericidal macrocyclic antibiotic, which blocks RNA synthesis (Zhanel *et al*., 2015). Less disruption to the intestinal microbiome has been reported for this drug, and lower rates of recurrent infection have been reported following treatment with fidaxomicin (Louie *et al*., 2012). Fidaxomicin is now the standard second line antibiotic for CDI and is recommended for severe cases and for recurrent infections (Cho *et al*., 2020).

### 1.4.2 Faecal Microbiota Transplantation

The development of new antimicrobial therapies should not adversely affect the gut microflora and the colonisation resistance provided. One such treatment is faecal microbiota transplantation (FMT), with high potential for resolving CDI (Rao and Safdar, 2016). This approach has been used since 1958 to resolve pseudomembranous colitis by administration of faeces via enemas (Eiseman *et al*., 1958). FMT involves the transplantation of processed stool from a healthy donor to a patient with relapsing disease, to replenish the normal colonic flora. This rescue

treatment is recommended for patients with >2 recurrences of infection, and those who have failed to respond to other treatments (Smits *et al*., 2016). High success rates of over 90% have been reported and FMTs are highly promising and effective (~81%) in preventing recurrent infections (Bakken *et al*., 2011). However, the long-term consequences to the colonic microflora are unknown.

While FMTs represent a very promising therapy for resolving CDI, squeamishness can be a barrier to wider adoption. Additionally, several risks are associated with FMT including the transfer of infectious pathogens from donor to patients, making FMT unsuitable for immunocompromised individuals (Gupta *et al*., 2016). Notably, the FDA issued a safety alert regarding FMTs, after two immunocompromised patients developed invasive infections with ESBL -producing *E. coli* (DeFilipp *et al*., 2019). These patients received the transplant from the same donor, with donated sample containing the infectious bacterium. It was later reported that this donor had not been screened for the presence of multi-drug resistant bacteria prior to donating. Moreover, gut microbiome composition has effects on human health and metabolism, and several patients have reported weight gain after FMT treatment (Alang and Kelly, 2015). Additionally, while FMTs have been used since the 1950s, standardised methods of sample retrieval, storage and donor screening were only established in 2015 by the Netherlands Donor Faeces Bank (Terveer *et al*., 2017).

### 1.4.3 Diffocins and Avidocins-CDs

As mentioned previously, broad-spectrum antibiotics perturb the colonic microflora, and the reduction in microbiome diversity is then exploited by *C. difficile*. As such, there should be a push for the development of narrow-spectrum antibacterial agents, that do not significantly alter the intestinal microflora. Avidocins represent one such agent. *C. difficile* naturally produces R-type bacteriocins, named diffocins, which resemble the R-type Pyocins produced by *Pseudomonas aeruginosa* (Gebhart *et al*., 2015). Avidocins are modified R-type bacteriocins, displaying bactericidal activity against competing *C. difficile* strains (Kirk *et al*., 2017b).

Several genes encoded within the diffocin locus (ORF 1359-1376) resemble a typical *Myoviridae* phage, including a sheath, baseplate, tail fibre and tail length determining proteins (Gebhart *et al*., 2012). For killing, contraction of the sheath drives a nanotube core through the bacterial cell envelope, creating a small pore. The resulting ion leakage dissipates membrane potential, leading to cell death. Killing specificity of Diffocins is determined by receptor binding proteins (RBPs) located on the tail fibre, which trigger sheath contraction upon recognition of the cognate receptor on the bacterial cell surface. Killing can be targeted by replacing the RBP with homologues of other strains (Kirk *et al*., 2017b). These modified bacteriocins are termed Avidocins-CDs. One such Avidocin is Av-CD292.2, with the capacity to kill all hypervirulent 027 strains. Av-CD292.2 specifically targets the surface layer (S-layer) of *C. difficile* cells, with the S-layer protein, SlpA identified as its cognate receptor (Kirk *et al*., 2017b) **(Fig 1.3)**.

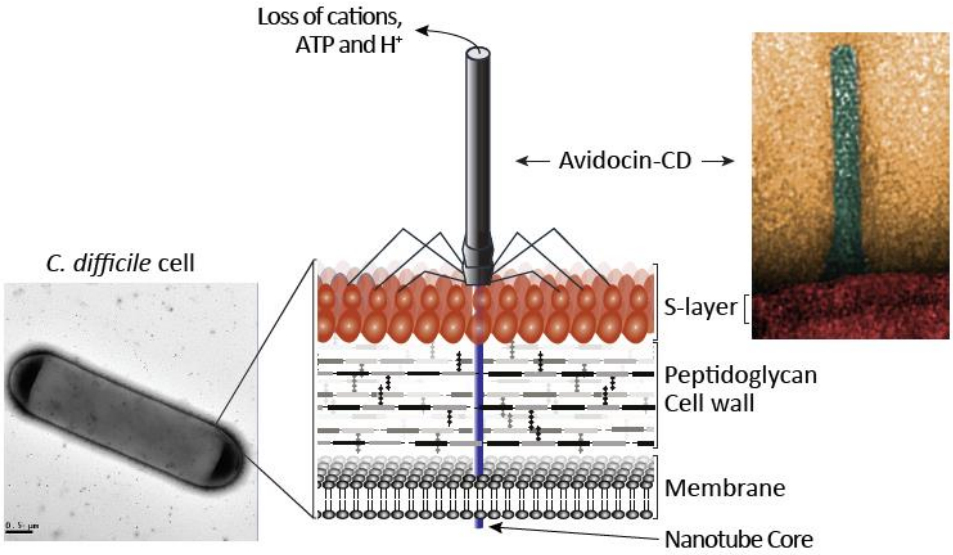



**Fig 1.3. *C. difficile* Avidocin-CD killing mechanism.** Avidocin-CDs are modified R-type bacteriocins, displaying bactericidal activity against competing *C. difficile* strains. Upon recognition of the cognate receptor, SlpA, a nanotube core is injected into the competing bacterial cell, penetrating the cell envelope, resulting in pore formation. Consequently, loss of cations and cell membrane potential occurs at the target, leading to cell death. Figure taken from Dr. Robert Fagan with permission.

## 1.5 Bacterial S-layers

Many bacterial species will elaborate a proteinaceous coat on the cell surface, known as an S-layer. These 2D paracrystalline arrays are formed by one or more S-layer proteins (SLPs), which are often glycosylated (Fagan and Fairweather, 2014). SLPs undergo self-assembly to form a regularly spaced array which covers the entire cell. S-layers are diverse in nature, and SLPs can self-assemble into oblique (p1, p2), square (p4) or hexagonal (p6) planar symmetry groups (Sleytr *et al*., 2014). S-layer ultrastructure revealed thickness of between 5 and 25nm, with the distance between symmetry subunits ranging from 2.5 to 35 nm (Sara and Sleytr, 2000). S-layers were first recognised in the 1950s, however, the lack of this protein layer in model organisms such as *Escherichia coli* and *B. subtilis*, hampered understanding of their functionality (Houwink, 1953). As mentioned previously,

S-layers are diverse in nature, and no unifying function for S-layers has been described. However, S-layer functions have been described for specific species, and it is clear that these proteinaceous arrays are crucial in cell physiology. The S-layer of *Campylobacter fetus* is anchored to the cell surface via lipopolysaccharide and functions as a key virulence factor (Thompson, 2002). Additionally, S-layers have been proposed to function as molecular sieves, enabling the passage of small ions and molecules into the cell, and as receptors for bacteriophage (Fagan and Fairweather, 2014).

## 1.6 The S-layer of *C. difficile*

Colonisation of the GI tract is central to *C. difficile* pathogenesis; however, knowledge of the process is limited. One structure has been implicated in adherence to enteric cells, the S-layer of *C. difficile (*Calabi *et al.*, 2002*)*. First identified in 1984, the *C. difficile* S-layer is composed of two subunits: the high-molecular weight (HMW) and the low-molecular weight (LMW) SLPs, both of which are derived from the pre-protein SlpA (Kawata *et al*., 1984). SLPs are the most abundant proteins in the cell requiring large amounts of energy to produce. Remarkably, the S-layer of *C. difficile* is estimated to require 590,000 subunits, with 164 subunits to be produced and translocated across the cell envelope per second during exponential growth to maintain the mature S-layer (Kirk *et al*., 2017a). Surprisingly, despite the metabolic pressure S-layer production places on the cell *slpA* is essential for growth, as evidenced by an inability of *slpA* to tolerate transposon insertions (Dembek *et al*., 2015).

SlpA has three morphologically distinct domains: an N-terminal signal peptide, which directs S-layer secretion, a highly variable LMW domain (~35 kDa), which has an immunostimulatory role and a more conserved HMW domain (~40 kDa), containing three tandem cell wall binding 2 (CWB2) motifs, involved in cell surface binding (Willing *et al*., 2015). The high variability and immunodominance of the LMW suggests it is exposed on the surface of the cell, while the cell wall

binding motifs in the HMW are consistent with a position closer to the cell wall (Calabi *et al*., 2001; Willing *et al*., 2015).

Further S-layer functionality is provided by 28 minor cell wall proteins (CWPs), all of which contain three tandem CWB2 motifs (Fagan and Fairweather, 2014). Located at either the amino or carboxy terminus of the CWPs, CWB2 motifs were originally identified in CwlB, an autolysin produced by *B. subtilis* (Kuroda *et al*., 1992). CWB2 motifs mediate the attachment of the CWPs to the cell surface via interactions with the polysaccharide II (PSII), although the exact nature of this interaction needs to be studied in more detail (Willing *et al*., 2015).

### 1.6.1 S-layer secretion

Typically, bacteria have secretion systems devoted for S-layer secretion (Fagan and Fairweather, 2014). *C. difficile* is no exception with S-layer secretion mediated by the accessory Sec system. *C. difficile* possesses two SecA homologues: SecA1 and SecA2 (Fagan and Fairweather, 2011). For *C. difficile*, SecA2 functions as the S-layer translocase. SecA2 is an energising ATPase, which interacts with SecYEG and serves as an ATP-driven molecular motor for translocation of SlpA across the membrane (Oatley *et al*., 2020). Post-secretion, SlpA undergoes proteolytic cleavage, a process catalysed by the highly conserved cysteine protease, Cwp84, generating the two SLPs (Kirby *et al*., 2009). The HMW and LMW self-assemble to form tightly packed H/L heterodimers, with the LMW external to the cell surface (Fagan *et al*., 2009). These H/L heterodimers incorporate into the S-layer at areas of newly synthesised PG (Oatley *et al*., 2020). S-layer cleavage is crucial for the cell, as evidenced by full length SlpA being poorly tolerated by the cell, leading to a substantial reduction in cell wall integrity (Dang *et al*., 2010). SecA2 is also responsible for the secretion of the major phase variable cell wall protein, CwpV (Fagan and Fairweather, 2011) **(Fig 1.4)**.

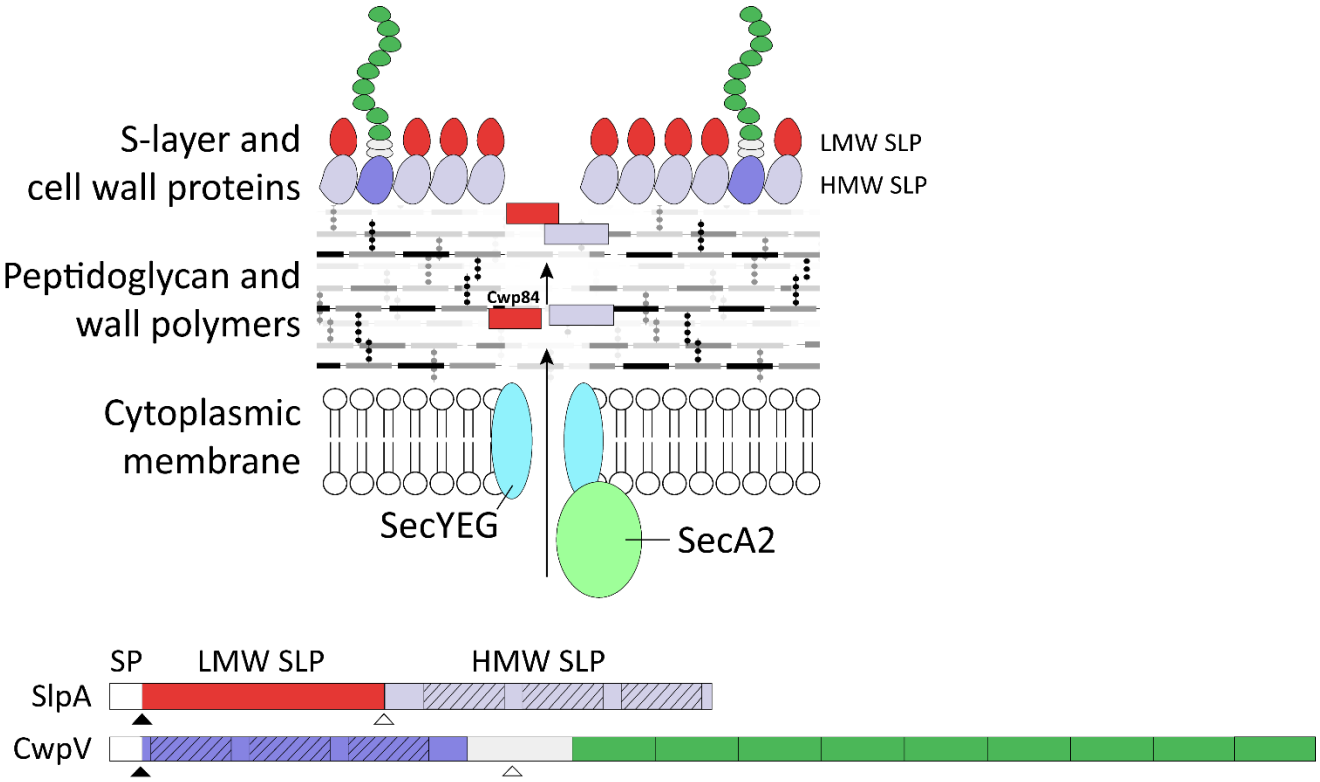



**Fig 1.4. *C. difficile* S-layer secretion and biogenesis.** SlpA is comprised of an N-terminal signal peptide, directing secretion, a variable low molecular weight (LMW) domain and a conserved HMW domain, composed of three cell wall binding 2 (CWB2) motifs for attachment to the cell surface. For SlpA secretion, the protein is secreted through a channel in the cytoplasmic membrane formed by SecYEG, a process driven by the ATPase SecA2. SlpA is post-translationally cleaved to form the mature S-layer proteins (SLPs), the HMW and LMW subunits, a process catalysed by the cysteine protease Cwp84. The HMW and LMW form heterodimers, which are tightly packed to form the mature S-layer, with the LMW subunit directed towards the environment. CwpV is also secreted by SecA2. Unlike SlpA, the CWB2 motifs of CwpV are located at the N-terminus of the protein. CwpV is post-translationally cleaved in a manner analogous to SlpA. Figure taken from Dr. Robert Fagan with permission.

### 1.6.2 Functions of the *C. difficile* S-layer

The *C. difficile* S-layer plays a pivotal role in immune signalling. SlpA has been implicated in the induction of the innate immune response via Toll-like receptor 4 (TLR4) (Fagan and Fairweather, 2014). Recently, by using Avidocins that specifically target the *C. difficile* S-layer, mutants which resisted killing arose at a frequency of 1 x $10^{-9}$ (Kirk *et al*., 2017b). These mutants had single

nucleotide polymorphisms in *slpA* that were predicted to truncate the protein at a site N-terminal to the post-translation cleavage site, preventing S-layer formation. One of these mutant strains, FM2.5, lacks detectable SlpA subunits. Isolation of S-layer null mutants revealed additional roles for SlpA in sporulation, toxin production and resistance to innate immunity effectors (Kirk *et al*., 2017b).

# 1.7 S-layer Surface Attachment via PSII

## 1.7.1 Secondary Cell Wall Polymers of *C. difficile*

*C. difficile* elaborates three anionic polymers on its surface, PSI, PSII and PSIII **(Fig 1.5)** (Ganeshapillai *et al*., 2008; Reid *et al*., 2012). Both PSI and PSII have been likened to wall teichoic acids, although they do not exhibit the classic poly-ribitol or poly-glycerol phosphate repeat structures (Ganeshapillai *et al*., 2008). PSI is composed of penta-glycosylphosphate repeats, while PSII is composed of hexa-glycosylphosphate repeats (Ganeshapillai *et al*., 2008; Weidenmaier and Peschel, 2008). The exact function of PSI has yet to be elucidated as PSI is not readily detected *in vitro*, suggesting this antigen is phase-variable (Bertolo *et al*., 2012). PSIII is a water-insoluble polymer, comparable to lipoteichoic acids (Percy and Grundling, 2014; Reid *et al*., 2012). PSIII is expressed in all strains examined to date and is comprised of phosphate, glucose, *N*-acetylglucosamine and glycerol (Monteiro, 2016).

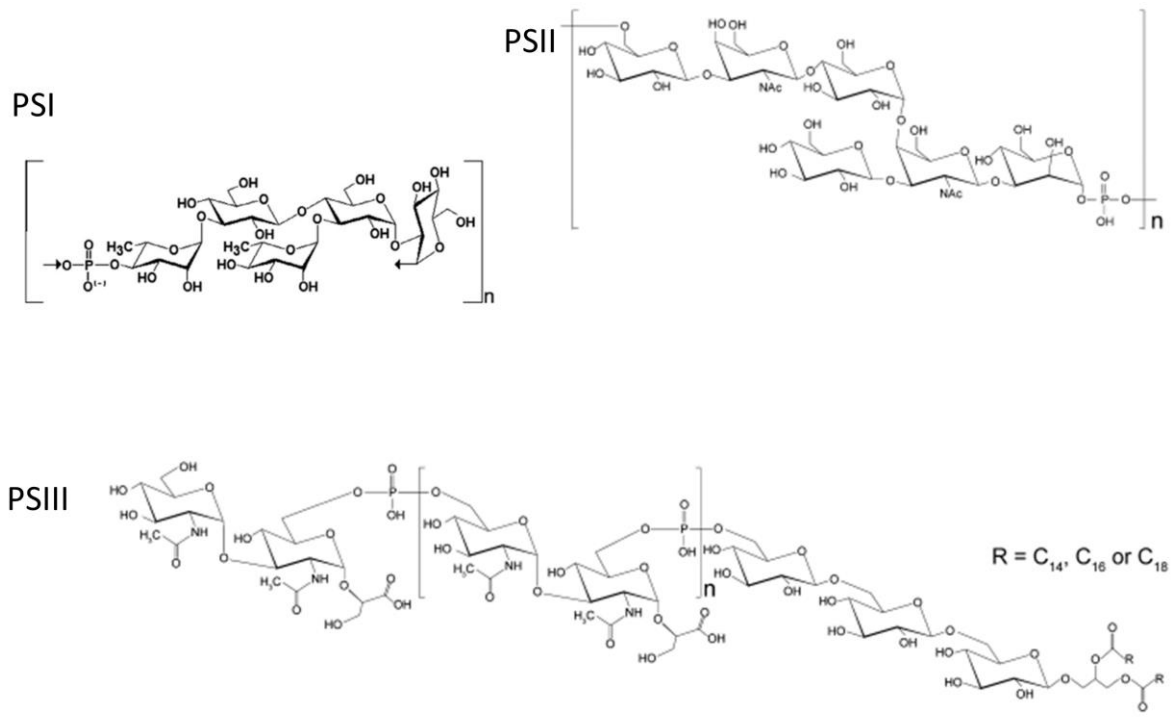



**Fig 1.5. Structure of the *C. difficile* cell wall polymers.** *C. difficile* elaborates three cell wall polymers on the cell surface: PSI, PSII, PSIII. PSI is composed of penta-glycosylphosphate repeats, with a core repeating unit of [→4)-α-l-Rha*p*-(1→3)-β-D-Glc*p*-(1→4)-[α-l-Rha*p*-(1→3]-α-D-Glc*p*-(1→2)-α-D-Glc*p*-(1→P]. PSII: a conserved cell wall polymer is composed of hexa-glycosylphosphate repeats, with the core repeating unit: [→6)-β-D-Glc*p*-(1→3)-β-D-Gal*p*NAc-(1→4)-α-D-Glc*p*-(1→4)-[β-D-Glc*p*-(1→]-β-D-Gal*p*NAc-(1→3)-α-D-Man*p*-(1→P]. PS-III is a conserved lipid-anchored cell wall polysaccharide in the extended lipoteichoic acid family with a core repeating unit of [→6)-α-D-Glc*p*NAc-(1→3)-[→P-6]-α-D-Glc*p*NAc-(1→2)-D-GroA]. This repeat unit is linked to →6)-β-D-Glc*p*-(1→6)-β-D-Glc*p*-(1→6)-β-D-Glc*p*-(1→1)-Gro, with the terminal glycerol esterified with C14, C16, or C18 saturated or mono-unsaturated fatty acids.

### 1.7.2 PSII Structure

PSII hexa-glycosyl repeats contain one mannose, two *N*-acetylglucosamine and three glucose residues with a phosphate group linking mannose and glucose-I **(Fig 1.4)** (Ganeshapillai *et al*., 2008; Ma *et al*., 2020). This is consolidated by the carbohydrate composition analysis of *C. difficile* NCTC 11223, which were shown to contain glucose, mannose, galactosamine and phosphate, similar to the composition of PSII (Poxton and Cartmill, 1982). PSII is highly conserved and has been identified in all

strains examined to date, making it an attractive target for a *C. difficile* vaccine (Bertolo *et al*., 2012). While the exact functions of PSI and PSIII have yet to be elucidated, PSII serves as a ligand for the attachment of CWPs to the cell wall (Willing *et al*., 2015). For example, SlpA and Cwp2 have been shown to bind PG-PSII but not PG alone, suggesting it is the polysaccharide and not the PG alone functioning as the ligand. Attachment is mediated by three tandem CWB2 motifs present in all *C. difficile* CWPs, although the exact nature of the interaction has yet to be elucidated (Kirk *et al*., 2017a).

### 1.7.3 PSII Biosynthesis

While the exact pathway of PSII synthesis has yet to be determined, a model has been proposed for PSII biosynthesis (Chu *et al*., 2016) **(Fig 1.6)**. In the cytoplasm, UppS, an undecaprenyl pyrophosphate synthase synthesises undecaprenyl pyrophosphate (UndPP). Undecaprenyl pyrophosphate phosphatase (UppP) activity on UndPP, yields undecaprenyl phosphate (UndP), which acts as the lipid carrier upon which PG, teichoic acids and PSII are built, revealing the complex interplay between many biosynthetic pathways. TagO is the initiating transferase for wall teichoic acids in Gram-positive bacteria, however, *C. difficile* does not possess a TagO homolog, suggesting that PSII is linked to PG by a unique linkage (Willing *et al*., 2015).

The transferase CD2783 likely initiates PSII synthesis. While Chu *et al*., 2016 propose that CD2783 functions as a glycosyltransferase involved in the transfer of a glucose residue onto the UndP lipid carrier, it is now believed that *cd2783* encodes a GalNAc-1-P transferase, with PSII initiating with the addition of GalNAc-I or GalNAc-II (Ma *et al*., 2020). This also challenges the hypothesis that the second step in the pathway is the addition of a GalNAc residue onto the Glucose-UndP PSII unit, which was believed to be the committed step in PSII biosynthesis. Regardless, it is widely believed that PSII adopts a Wzy-dependent pathway for synthesis, whereby the repeat unit is built on the inner face of the cytoplasmic membrane, transported across the membrane by a Wzx flippase and polymerised by a Wzy polymerase (Chu *et al*., 2016). Although the contribution of

individual genes to PSII synthesis is in its infancy, *cd2783* encodes the initiating transferase, MviN is the proposed flippase of the system, responsible for transporting the synthesised hexamer across the membrane, *cd2777* encodes the polysaccharide polymerase, responsible for polymerising PSII extracellularly and LcpA/LcpB are surface anchoring proteins involved in the attachment of PSII to PG (**Fig 1.5)**

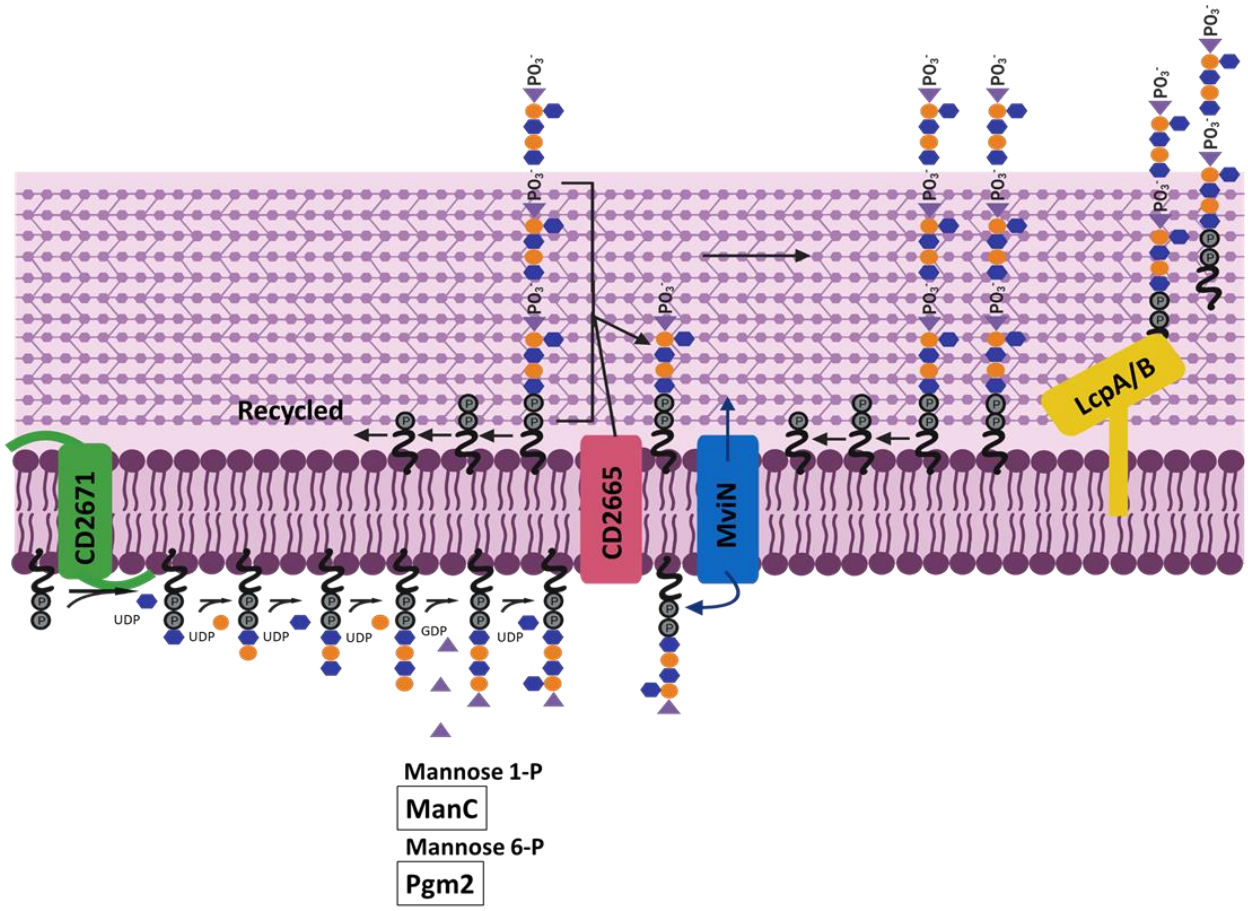



**Fig 1.6. Proposed model for PSII biosynthesis in *C. difficile*.** PS-II is assembled in the cytoplasm on the lipid carrier molecule, UDP. UDP-PS-II is transported to the cell surface where it is polymerised. Fully synthesised PSII interacts with the CWB2 motifs of the cell wall proteins, for their attachment to PG. Locus tags for *C. difficile* R20291 are shown. CD2671 (green) serves as the initiating transferase, transferring glucose to the PS-II unit. A second glycosyltransferase adds *N*-acetylglucosamine onto the PS-II unit, held by a β 1,3 linkage. ManC and Pgm2, two cytoplasmic glycosyltransferases complete the synthesis of the hexaglycosyl PS-II unit. MviN (blue), a flippase, transports PS-II across the cell membrane. CD2665 (pink), polymerises PS-II extracellularly.

LcpA/LcpB, are two surface anchoring proteins, which catalyse the transport and anchoring of the fully synthesised PS-II unit to PG (Chu *et al*., 2016).

Recently, CD2775 was identified as a putative mannosyl-1-phosphotransferase, responsible for the transfer of mannose-1-phosphate onto Glc-β1,3-GalNAc-UndP, resulting in the formation of a unique Man-1-P-6-Glc linkage (Ma *et al*., 2020). CD2775 is the first mannosyl-1-phosphotransferase identified in all living systems. In this study, it was also found that *cd2774*, another gene involved in PSII biosynthesis shares 65% sequence identity with a β1,3 glucosyltransferase, WbdN. All of these biosynthetic genes are contained within an anionic polymer (AP) locus, believed to direct PSII synthesis (Willing *et al*., 2015). The locus which also contains genes for glycosyltransferases with homology to other SCWP biosynthetic genes, a phosphomannomutase (*pgm2*), a mannose-1-phosphateguanylyltransferase (*manC*), and an undecaprenyl pyrophosphate synthase (*uppS*), providing the lipid carrier for many essential biosynthetic pathways **(Fig 1.7)**. Moreover, this AP locus is located immediately downstream of *slpA*, *secA2* and gene encoding other major CWPs, Cwp84 and Cwp66. Taken together, this suggests that the AP locus directs synthesis of PSII (Willing *et al*., 2015).

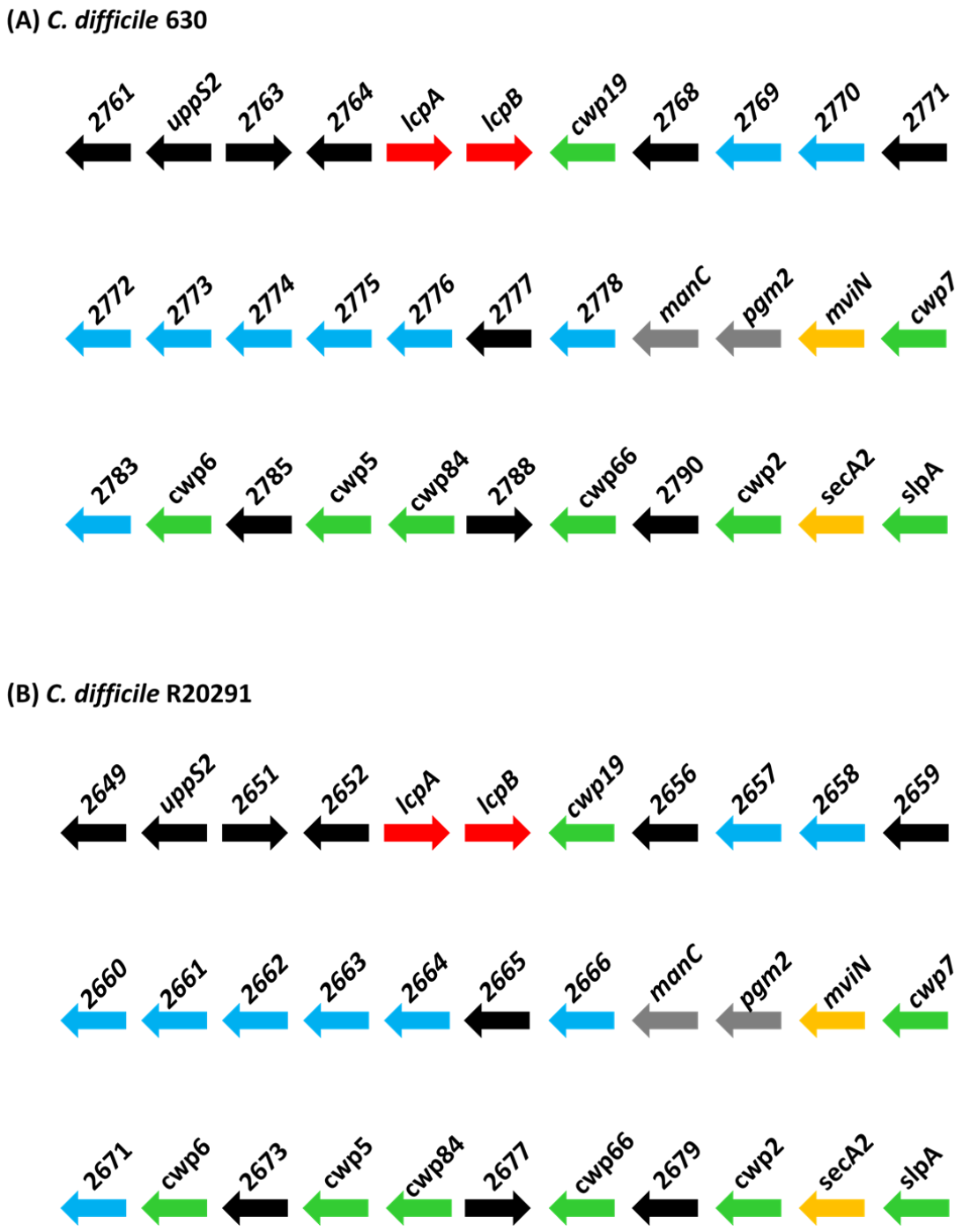



**Fig 1.7. Anionic polymer locus of *C. difficile* 630 and *C. difficile* R20291.** Schematic illustration of the genetic locus of the anionic polymer and cell wall proteins of *C. difficile* strains 630 (A) and R20291 (B). Gene colours indicate putative functions. S-layer and cell wall proteins are shown in green, glycosyltransferases in blue, and genes for attachment to the cell wall PG are depicted in red. Mannose biosynthetic genes are shown in grey, and genes devoted to protein translocation are shown in yellow. Other genes or genes of unknown function are depicted in black (Willing *et al*., 2015).

### 1.7.3 Functions of PSII biosynthetic genes

Transposon mutagenesis data from both *C. difficile* strains 630 and R20291 revealed that the genes involved in PSII are essential, not tolerating a single transposon insertion across the locus. This is unsurprising, due to the fact that PSII functions as the ligand for S-layer attachment to the surface, a structure which is vital for *C. difficile* (Dembek *et al*., 2015). However, PSII may be essential in its own right due to an unknown function unrelated to the S-layer. Furthermore, disrupting later steps in secondary cell wall polymer (SCWP) pathways can lead to the accumulation of toxic intermediates (Brown *et al*., 2013) . This is evidenced in *C. difficile* by an inability to insertionally inactivate *cd2775, manC, pgm2, mviN, uppS, cd2769* (Chu *et al*., 2016; Ma *et al*., 2020; Willing *et al*., 2015). As a result, antisense RNA methods have been employed to knock-down the expression of these genes so their contribution to PSII synthesis could be determined.

Knock-down of the flippase *mviN* results in a slight growth defect and a curvature of cells (Chu *et al*., 2016). A similar growth defect was observed for knock-downs of *pgm2* with cell shape being elongated. Additionally, depletion of *pgm2* and *mviN* altered the CWP profile of *C. difficile* with less Cwp66 present on the cell surface. Interestingly, depletion of *uppS* resulted in the accumulation of unprocessed full-length SlpA on the cell surface. This did not coincide with the accumulation of full-length SlpA in the cytoplasm, suggesting knock-down of this gene is disrupting S-layer assembly (Willing *et al*., 2015).

Lcp proteins play a role in cell wall biosynthesis in many Gram-positive organisms, specifically in wall teichoic acid anchoring, however their roles in *C. difficile* were unknown (Stefanovic *et al*., 2021). *lcpA* and *lcpB* were previously inactivated by ClosTron mutagenesis, allowing their contribution to PSII anchoring, if any, to be determined (Chu *et al*., 2016; Heap *et al*., 2010). Despite showing 64% amino acid identity, LcpA and LcpB do not exhibit functional redundancy. Only modest defects were observed for the *lcpA*$^{-}$ mutants, which were believed to be polar effects of *lcpB* inactivation. However, a myriad of phenotypes was described for *lcpB*$^{-}$ mutants,

including increased shedding of PSII and impaired anchoring of polymers, suggesting a role for LcpB in maintenance of deposited PSII. *lcpB*$^{-}$ mutants exhibit severe growth defects, most notably thicker, curved elongated cells with multiple mislocalised septa (Chu *et al*., 2016). Most striking, *lcpB*$^{-}$ mutants are hypervirulent in a hamster model of infection with the average time of death decreasing from 152 hours for hamsters injected with wild-type *C. difficile* to 65 hours for hamsters injected with the *lcpB*$^{-}$ mutants. This is perhaps due to the increased shedding of PSII being hypo-immunostimulatory.

## 1.8 *C. difficile* Cell Wall Proteins

SlpA and all 28 CWPs contain three tandem CWB2 motifs (Pfam 04122) and are attached to the cell surface via PSII (Willing *et al*., 2015). In addition to the CWB2 motifs, these proteins generally harbour an additional domain, conferring additional functionality to the S-layer **(Fig 1.8)**. While specific functions have not been identified for all minor CWPs, several have been elucidated. For example, Cwp66 and Cwp2 are involved in host cell adhesion (Bradshaw *et al*., 2017; Waligora *et al*., 2001). Cwp22, a novel PG modification enzyme, functions as an L, D transpeptidase and plays pleiotropic roles in *C. difficile* (Zhu *et al*., 2019). Interestingly, a ClosTron generated *cwp22* mutant displayed reduced TcdA and TcdB production during early growth, and increased cell permeability and autolysis. Recently, Cwp19 has been described as a novel PG hydrolase which mediates toxin release in a glucose-specific manner (Wydau-Dematteis *et al*., 2018). It is believed that Cwp19 and TcdE co-exist, acting as independent toxin release proteins in response to different environmental conditions with a potential role in *C. difficile* pathogenesis. This work was carried out in *C. difficile* 630, however, Cwp19 has been identified in all sequenced *C. difficile* genomes to date and further work must be done to see if this function extends to the hypervirulent 027 ribotypes.

As described previously Cwp84 is a cysteine protease responsible for S-layer processing (Kirby *et al*., 2009). Additionally, this protein shows activity against fibronectin and type IV collagen, indicating a potential role during infection (Janoir *et al*., 2007). However, loss of *cwp84* does not

result in loss of virulence in an animal model of infection. Interestingly, a second cysteine protease, Cwp13, is encoded by *C. difficile*, sharing 63% sequence identity with Cwp84. This protein can partially substitute the SlpA processing, however, cleavage occurs at a site distinct from the Cwp84 target (de la Riva *et al*., 2011).

Encoded by *cwpV* and subject to phase-variable expression, CwpV is the largest member of the CWP family and is highly abundant on the cell surface when expressed. Specifically, it comprises approx. 13.3% of the total surface layer proteins (Reynolds *et al*., 2011). CwpV is post-translationally cleaved to produce two subunits in a manner analogous to SlpA (Dembek *et al*., 2012). However, this cleavage is autocatalytic and does not require an additional protease. CwpV and SlpA may interact, providing structural integrity to the S-layer (Fagan and Fairweather, 2011). CwpV is expressed by ~5% of *C. difficile* cells in a given population under normal laboratory conditions, however, in FM2.5, the strain that lacks an S-layer, 95+% of cells express CwpV (Fagan and Fairweather, 2011). CwpV phase variation is mediated by an invertible genetic switch located upstream of the gene (Emerson *et al*., 2009). Unlike similar systems that have been studied in other bacteria, the *cwpV* promoter is not itself located within this genetic switch, rather it is found upstream of the invertible sequence. In the 'ON' orientation, transcriptional read-through occurs and CwpV is expressed. In the 'OFF' orientation a Rho-independent transcriptional terminator forms, preventing transcriptional read-through (Emerson *et al*., 2009). In FM2.5 the *cwpV* genetic switch is biased towards the 'ON' orientation but the underlying mechanism is not understood (Dr. Joseph Kirk, personal communication).

CwpV exhibits auto-aggregation activity, which has been attributed to the C-terminal region of the protein. This auto-aggregation activity has been described in cells and liquid media (Reynolds *et al*., 2011). Five antigenically distinct types of the C-terminal region have been identified. CwpV is cleaved in two: a 42 kDa N-terminal fragment, containing the CWB2 motifs, and a 90-120 kDa C-terminal fragment, depending on CwpV type. The C-terminal region of each CwpV type comprises 4-

9 tandem repeats, each of which contains between 79 and 120 amino acids. An additional role for CwpV in protection against phage infection systems has been described (Sekulovic *et al*., 2015). All CwpV types confer protection against siphophages, while types I, III and V also provide protection against myophages.

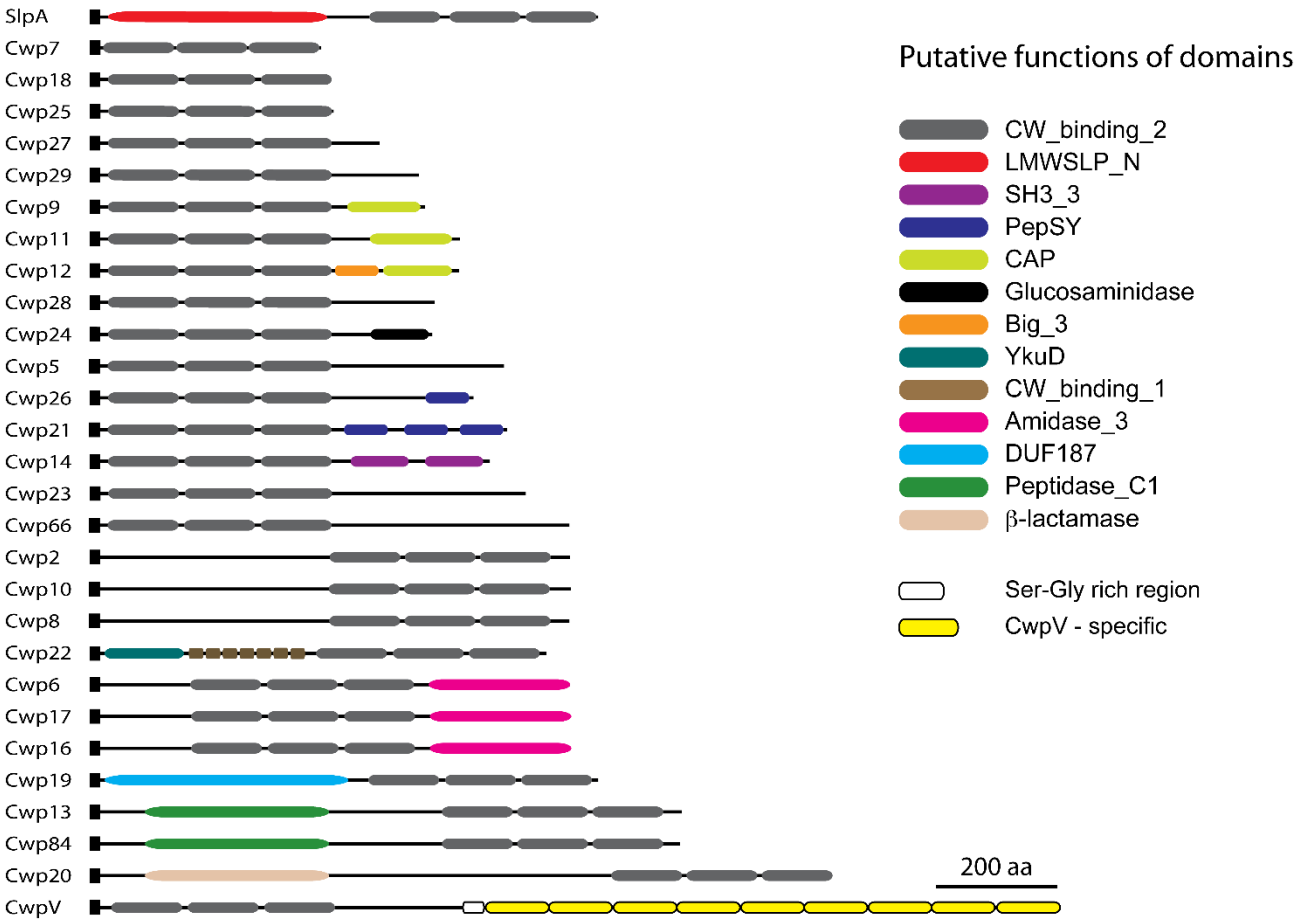



**Fig 1.8. *C. difficile* cell wall proteins.** Domain organisation and identification of the 29 cell wall proteins (CWPs) elaborated by *C. difficile*. Common to all CWPs is the presence of three CWB2 motifs, located at either the amino- or carboxy- terminus of the protein. Functionality of CWPs is provided by additional domains, located within the protein. Figure taken with permission from Dr. Robert Fagan.

## 1.9 Genetic Tools for *C. difficile* Studies

Historically, *C. difficile* has been regarded as a genetically intractable organism, with the first method for conjugative transfer of plasmids only published in 2002 (Purdy *et al*., 2002). The first technique for *C. difficile* genome manipulation to be described, ClosTron, involves inactivation of genes through an insertion of a group II intron from the *ltrB* gene of *Lactococcus lactis* (Heap *et al*., 2007; Mills *et al*., 1997). Unfortunately, polar effects can be caused at the site of insertion, and clean knockouts cannot be made using this technique (Heap *et al*., 2007). This system was superseded by a homologous recombination using *codA*- and *pyrE*-based negative selection, which allow deletion, insertion, or substitution of sequences (Cartman *et al*., 2012; Ng *et al*., 2013). The *codA* allele exchange technique uses pseudo-suicide vectors, which have sub-optimal origins of replication. This means that the rate of growth under selection is limited by the rate of plasmid replication. A single recombination event results in the plasmid integrating into the chromosome. This allows the plasmid to replicate as the chromosome replicates, visualised as enhanced growth. In a second recombination event, the plasmid is excised from the genome. The recombination vector has a *codA* gene, encoding a cytosine deaminase. CodA converts cytosine to uracil. Double recombinants are screened on media containing 5'-fluorocytosine, and only bacteria that have lost the plasmid can grow as the conversion of 5'-fluorocytosine to 5'-fluorouracil is toxic. Bacteria that grow are either reverted wild-type or the mutant bearing the deletion of interest. However, the *codA* technique is labour intensive, and the frequency of mutants obtained is low. Recently, transposon mutagenesis coupled with high-throughput sequencing for obtaining saturated libraries of mutants has been described for *C. difficile* (Dembek *et al*., 2015). This and the development of CRISPR and CRISPR interference (CRISPRi) technologies have expanded the genetic toolbox for precise manipulation of the *C. difficile* genome (Muh *et al*., 2019).

## 1.10 Transposon Directed Insertion Site Sequencing

Transposon mutagenesis involves the creation of random insertional mutants, which can be readily scaled to generate large libraries in which all non-essential genes will tolerate transposon insertions (Munoz-Lopez and Garcia-Perez, 2010). Transposons are short DNA fragments which insert randomly into the genome and studies involving these DNA fragments are useful for a wide-variety of applications, including identification of virulence genes for bacterial pathogens, identification of the essential genome of an organism and the identification of genetic interaction networks (Goodall *et al*., 2018; Hensel *et al*., 1995; van Opijnen and Camilli, 2010). Transposons can produce various genetic alterations by excision, insertions, or duplication at the site of transposon integration (Hamer *et al*., 2001). Transposon insertion within the open reading frame or regulatory region of the gene often disrupts the function or expression of the gene and by observing the phenotype of mutants harbouring these inserts, a link between phenotype and genotype can be drawn (Cain *et al*., 2020). Additionally, with the recent global push to discover new antimicrobials, subjecting mutant libraries to high concentrations of antibiotics, would enable genes contributing to resistance to be identified, which could be specifically targeted in future drug development.

### 1.10.1 Discovery of the Transposon

Colloquially known as 'jumping genes', transposons were first described by Barbara McClintock in 1948 through experimentation with *Zea mays* (McClintock, 1950). McClintock noticed unusual streaks and spots of colour in a strain of corn, which was attributed to breakage on chromosome 9, and concluded that a genetic element caused the mutation, with an additional element controlling the activity of the first element. These were identified as two independent genetic loci which she named the 'Dissociator' and the 'Activator'. The Dissociator (Ds) was the element that caused the mutation and was located on the short arm of chromosome 9. The Activator (Ac) element is autonomous and found to control transposition of the Dissociator element.

The most important finding was these elements sometimes appeared in other chromosomal locations. The ability of a linear DNA fragment to move to different positions of the chromosome, reversibly inactivating the target gene in which it was inserted, dissipated the notion that a genome was a stationary entity.

Since their discovery, transposons have now been identified in all kingdoms of life. In the 1970s two groups independently discovered that transposons could transpose genes conferring resistance to antibiotics, with the Tn*10* transposon containing a tetracycline resistance gene and the Tn*5* element conferring resistance to kanamycin, streptomycin, and bleomycin, making it easy to select for mutants harbouring inserts (Berg *et al*., 1975; Kleckner *et al*., 1975). It is now recognised that there are two major classes of transposons, which differ in terms of transposition mechanisms (Finnegan, 1992). Class I elements follow the retrotransposition mechanism and include *Alu* elements in primates, *Ty* elements in yeast, intra-cisternal A particles (IAPs) in rodents and *gypsy* and *copia*-like elements in *Drosophila* (Hamer *et al*., 2001). Class II transposons generally transpose by a 'cut and paste' mechanism directly from DNA to DNA. Typically, these DNA transposons encode a transposase gene flanked by two terminal inverted repeats (TIRs). The transposase recognises the TIRs, excising the transposon from the DNA at these points. The transposon is reinserted into a new genomic location, and upon insertion the target site DNA is duplicated. Class II elements include the *Tc-1/mariner* superfamily of transposons and bacterial *Tn* elements (Hamer *et al*., 2001). Class II elements have been used for transposon mutant library generation. Many transposon mutagenesis studies use a Tn*5* derivative, as it does not show any bias for DNA sequence and is active in a wide array of bacterial species (Barquist *et al*., 2013). However, since the transposon is normally delivered by electroporation of a fragment of DNA containing the transposon, the number of transposon mutants varies according to the transformation efficiency of the host organism. As *C. difficile* is extremely AT rich (~70% AT), the preference for insertional mutagenesis is a *Mariner/Himar I* transposase, which has an absolute necessity for AT dinucleotides at the insertion site (Dembek *et al*., 2015). All Tc1/*mariner* elements share two common domains, an amino-terminal region

containing a Helix-Turn-Helix motif for recognition and binding of TIRs, and a carboxy-terminal catalytic domain, consisting of three amino acids, DDE for *Tc-1* elements and DDD for *mariner* elements (Munoz-Lopez and Garcia-Perez, 2010) **(Fig 1.9)**.

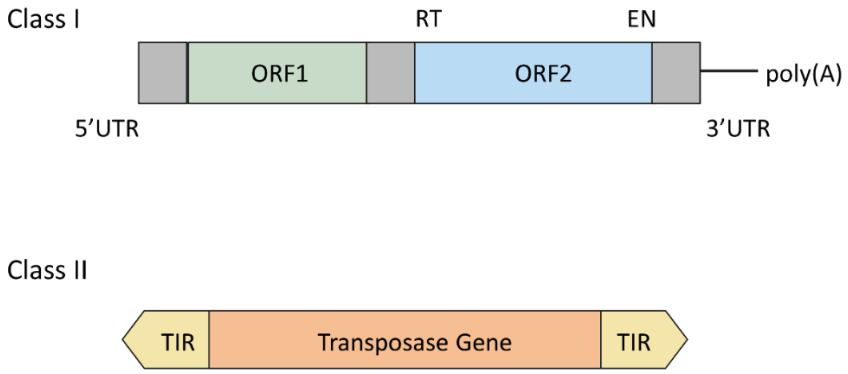



**Fig 1.9. Classes of transposons.** Two major classes of transposons have been described. The type I transposon consists of a 5' untranslated region (UTR), with internal promoter activity. The first open reading frame (ORF) encodes a nucleic acid binding protein while the second ORF encodes a protein that displays both reverse transcriptase (RT) and endonuclease activity. Class I transposons end in a 3' UTR and a poly(A) tail. Class II transposons consist of two terminal inverted repeats (TIRs), flanking a transposase gene (Lopez 2010).

### 1.10.2 Transposon Mutagenesis

Signature tagged-mutagenesis (STM) was the first evolution of traditional transposon mutagenesis techniques which aimed to identify transposon insertion mutants required for infection within an animal host or host cell (Hensel *et al*., 1995). This technique was first described by Hensel *et al*. in 1995, who created 96 barcodes to tag the transposon. This allowed mutants to be phenotypically characterised in pools of 96, drastically increasing the throughput. The mixed pool of tagged mutants is inoculated into the host, the gDNA extracted and the 96 tags pulled down by DNA hybridisation. The first of its kind, this technique allowed the discovery of virulence genes without prior notion of an individual genes function. Unfortunately, although significantly better than screening individual mutants, STM is still a very labour-intensive technique, limited to smaller pool sizes involving the screening of each gene within a bacterial genome.

Until the invention of next generation sequencing technologies, no significant further developments were made in transposon mutagenesis studies. However, with the advent of Illumina sequencing, it became possible to perform a genome-wide screen of transposon insertions. The Illumina sequencing platform was adapted for sequencing the transposon-gDNA junctions of pooled mutants, enabling the precise location of transposon insertions to be determined without needing to isolate individual mutants. Since this, major transposon directed insertion site sequencing experiments have been described (Gawronski *et al*., 2009; Goodman *et al*., 2009; Langridge *et al*., 2009; van Opijnen *et al*., 2009). The raw sequencing data generated is processed and aligned to the target genome to identify insertion sites across the genome. With a sufficiently dense library, the frequency of transposon insertions across the genome displays a bimodal distribution (Langridge *et al*., 2009). The first mode represents genes unable to survive without a functioning copy of the gene and are classified as essential, while those in the second mode that can be disrupted by insertions along the coding sequence are deemed non-essential **(Fig 1.10)**. However, there are some instances where genes have essential and dispensable domains and are classified as being of ambiguous essentiality (Christen *et al*., 2011).

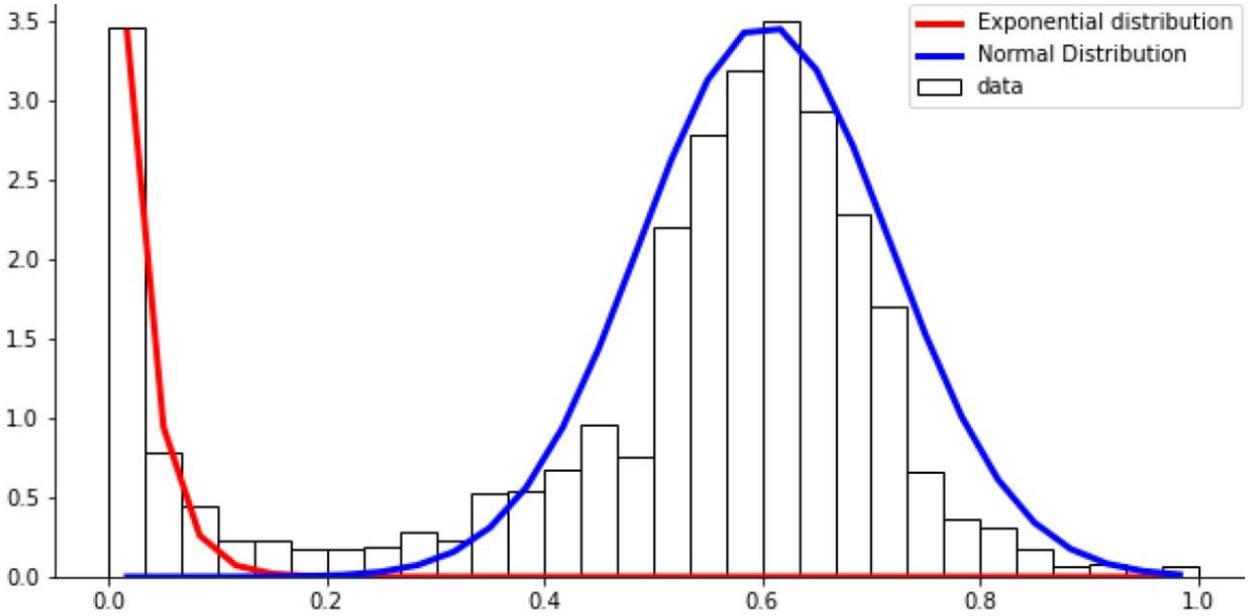



**Fig 1.10. Bimodal distribution of an insertion mutagenesis library**. The histogram depicts the frequency of insertion indexes for a TraDIS dataset. For the left mode (red), an exponential distribution is fit to essential genes, which cannot survive without a functioning copy of the gene. The right peak (blue) a gamma distribution is fitted, which represents all non-essential genes which can be disrupted by transposon insertions. Figure taken with permission from Larivière *et. al* (2021).

Since 2009, four variations of transposon insertion sequencing have been described. These include Transposon Sequencing (Tn-Seq), Insertion Sequencing (IN-Seq), Transposon Directed Insertion Site Sequencing (TraDIS) and High-throughput Insertion Tracking by Deep Sequencing (HITS) (Cain *et al.*, 2020). TraDIS and Tn-Seq experiments were used to determine the core essential genome for growth of organisms in rich medium and the fitness cost associated with gene disruptions (Charbonneau *et al.*, 2017; Poulsen *et al.*, 2019; van Opijnen and Camilli, 2013). On the other hand, HITS and IN-Seq were used to test human infection models (Bachman *et al.*, 2015; Gawronski *et al.*, 2009; Wang *et al.*, 2014). All four methods follow the same core methodology. Each technique requires a dense pooled library of transposon mutants. The transposon is delivered into cells by transformation or conjugation and allowed to insert randomly throughout the genome (Artiguenave *et al.*, 1997; Dembek *et al.*, 2015) **(Fig 1.11)**. Mutants are pooled and the genomic DNA (gDNA) is extracted (Langridge *et al.*, 2009). Ideally, each mutant in the pool will harbour a single

transposon insertion, and within the pool, each gene that can be disrupted will be disrupted multiple times at different sites (Chao *et al*., 2016).

Next, to separate the gDNA harbouring the transposon insertions from the large array of wild-type DNA, the DNA is fragmented through restriction digestion or through physical shearing of the DNA (Cain *et al*., 2020). Tn-Seq and IN-Seq both use the type II restriction enzyme MmeI for DNA fragmentation. MmeI recognition sites are close to the end of the transposon, and fragmentation results in uniform short fragments (van Opijnen and Camilli, 2013). MmeI makes a 2 bp staggered cut 20 bp laterally to the recognition site, generating fragments comprising the left and right transposon ends plus 16 bp of flanking DNA, sufficient to accurately pinpoint the location of the transposon insertion (Goodman *et al*., 2011). HITS and TraDIS utilise physical shearing of the gDNA through sonication (Cain *et al*., 2020). The HITS protocol also includes an additional affinity purification step to remove contaminating DNA before sequencing (van Opijnen and Camilli, 2013). TraDIS and HITS have the advantage of being applicable to any transposon as fragmentation does not rely on a restriction site within the inverted repeats. However, shearing for TraDIS and HITS generates randomly sized PCR fragments, potentially allowing for PCR biases as shorter fragments will be preferentially amplified over longer ones. To conclude, the optimum protocol to use is dependent on the strain and the lab resources available.

Following shearing, sequencing adaptors containing flow cell adaptors, barcodes and sequencing primer sites are ligated to the end of blunt-end DNA, providing a priming site for the samples to bind to the Illumina flow cell during sequencing. PCR is performed on these fragments using a transposon-specific primer and a sequencing adaptor-specific primer, which enriches for fragments spanning the transposon-gDNA junction **(Fig 1.11)**. The resulting DNA fragments can then be sequenced using standard methods.

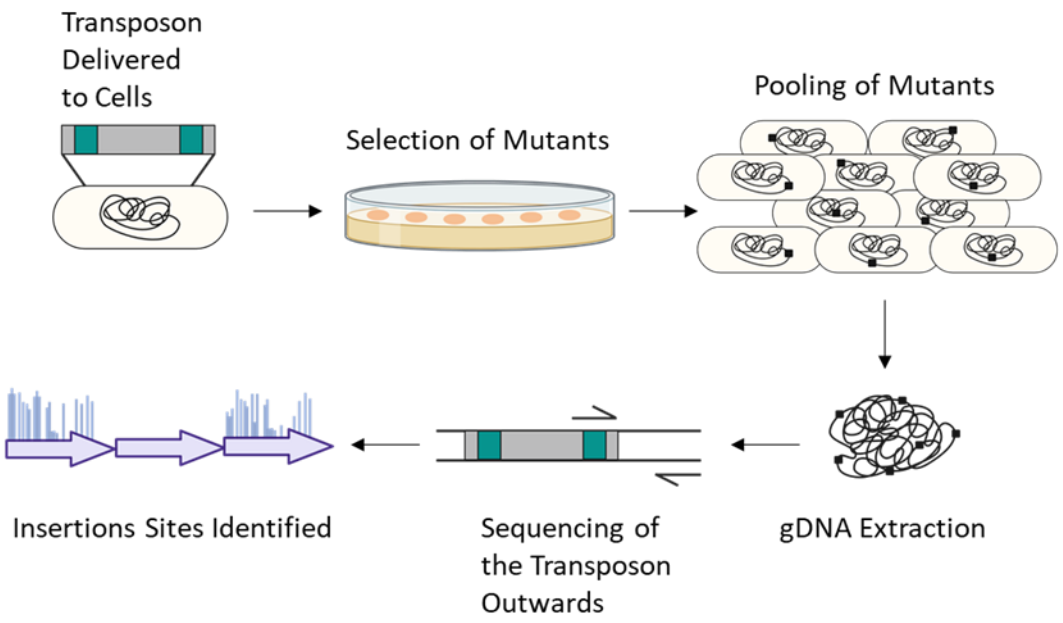



**Fig. 1.11. Workflow of a transposon sequencing experiment.** The transposon is delivered to cells via transformation or conjugation and allowed to insert randomly throughout the genome. Mutants are pooled and the gDNA is extracted and processed as outlined above. Transposon insertion sites are identified by sequencing from the transposon-gDNA junction. Insertions are mapped to a reference genome to identify essential (harbouring no insertions) and non-essential genes.

### 1.10.3 A High-Density Transposon Library for *C. difficile*.

The method chosen for the work in this thesis is TraDIS, a negative selection assay capable of screening thousands of mutants simultaneously. The first transposon mutant library for *C. difficile* was generated using a mariner-based transposon (Cartman and Minton, 2010). However, the size of this library was limited by inefficiency of plasmid delivery into *C. difficile* and lack of effective control of the timing of transposition. It wasn't until 2015 that the first comprehensive high-density transposon mutant library was constructed using a novel conditional and inducible mariner delivery vector (Dembek *et al*., 2015). This library was constructed in the hypervirulent ribotype 027 strain R20291. By TraDIS, Dembek *et. al* reported a set of 404 genes as essential for growth in *vitro*. This library was then put through the sporulation process, required for transmission of the organism, and

from this, 798 genes were identified as likely to affect sporulation, presenting attractive targets for drug development (Dembek *et al*., 2015).

### 1.10.4 Beyond Essentiality: Applications of Transposon Sequencing

Beyond identifying the essential genome of an organism, TraDIS can be used to identify conditionally essential genes and synthetically lethal gene pairs (Gawronski *et al*., 2009; Santa Maria *et al*., 2014; Wong *et al*., 2016). Genes can be described as conditionally essential if they are required for growth under test conditions, for example, under different stressors or colonisation of host tissue in an infection model. Assessing conditional essentiality involves growth of the library under permissive or stressful conditions and comparing the input and output pool of mutants **(Fig 1.12).** For example, analysis of a mutant library of the human symbiont *Bacteroides thetaiotamicron* identified several genes required for survival in the colon, and that colonisation was partly influenced by the gut microbiome composition and competition for various nutrients, including Vitamin B12 (Goodman *et al*., 2009).

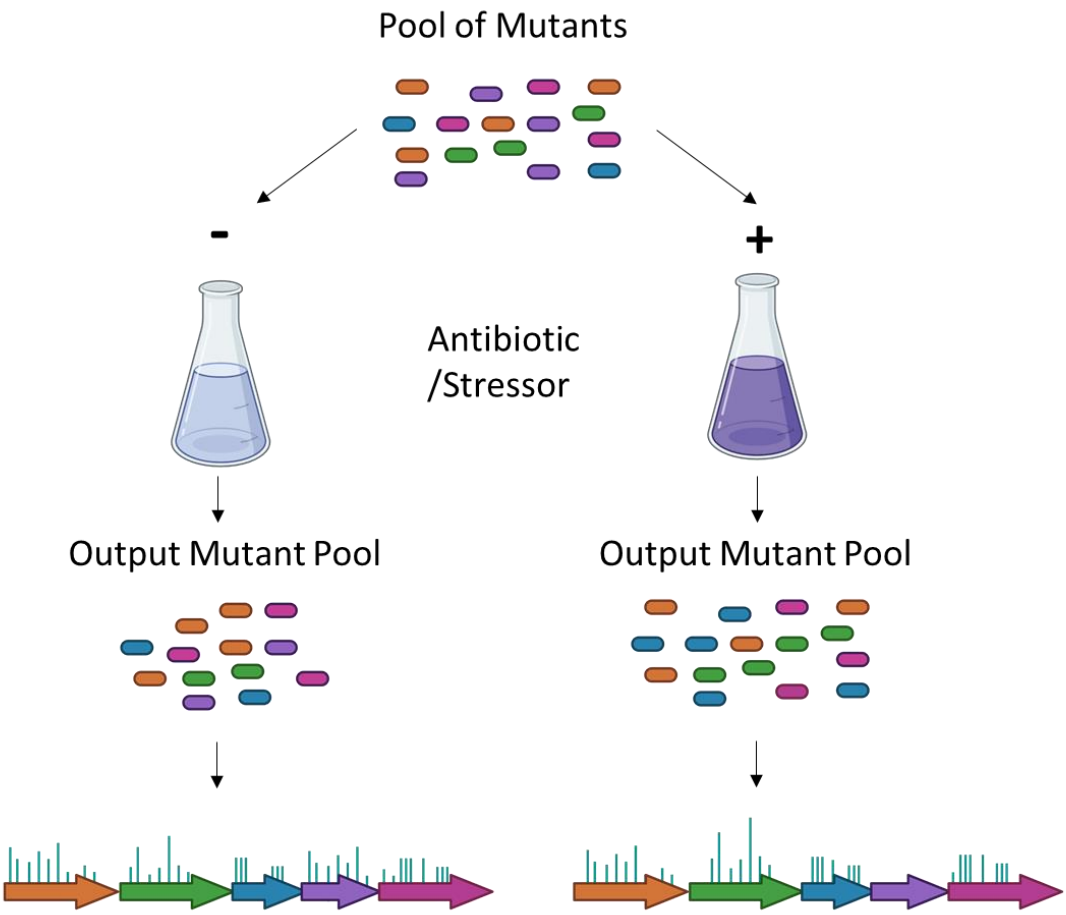



**Fig 1.12. Identification of conditionally essential genes using transposon sequencing.** An input library of mutants is created in which all non-essential genes will harbour transposon inserts. This pool of mutants is subjected to permissive (lacking antibiotic or a stressor) or stressful conditions (media supplemented with antibiotic or a stressor). The selection pressure will result in some mutants being lost or out-competed from the population. Comparison of mapped transposon insertions for the output mutant pool, lacking the antibiotic or stressor, and the mutant pool grown in the presence of the antibiotic or stressor, can identify genes which are conditionally essential (e.g., gene in purple) for growth under the condition of interest.

As mentioned previously, transposon mutagenesis has been used to identify synthetically lethal gene pairs. Gene pairs are described as synthetically lethal if deletion of either gene does not have any impact on cell viability, but the combined deletion of both genes results in a non-viable mutant. In the context of TraDIS it is possible to identify all synthetically lethal gene pairs by

constructing a transposon mutant library in a strain previously mutated to lack the gene of interest. For example, *Streptococcus pneumoniae* mutant libraries were created in single-gene deletion strains lacking individual carbohydrate uptake genes (van Opijnen *et al*., 2009). TraDIS was then used to identify the complex genetic interaction network that occurs during regulation of carbohydrate uptake. Identification of synthetically lethal pairs for a *C. difficile* S-layer mutant will be discussed further in Chapter IV.

## 1.11 CRISPR

### 1.11.1 CRISPR-Cas Systems

In nature, bacteria and archaea dominate a myriad of highly competitive environments (Hibbing *et al*., 2010). To survive, these microbes have evolved an RNA-mediated adaptive immune response, known as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), to fend off incoming foreign genetic material (Jinek *et al*., 2012). In 1987 Ishino *et al,* first described the cloning and sequencing of *E. coli iap,* whose product is required for isozyme conversion of alkaline phosphatase (Ishino *et al*., 1987). While characterising this locus, a series of repeats of unknown function were identified. These are now known as CRISPR arrays, composed of 30-40 bp short partially palindromic repetitive sequences interspersed by short non-homologous spacer sequences **(Fig 1.13)** (Peters *et al*., 2015). These loci are typically flanked by CRISPR-associated (*cas*) genes (Shmakov *et al*., 2020). The sequence upstream of the CRISPR array is known as the leader and includes the promoter for transcription of the array (Kieper *et al*., 2019). The partial palindromic nature of the repeat regions enables hairpin structure formation, required for binding of Cas proteins (Barrangou and Marraffini, 2014).

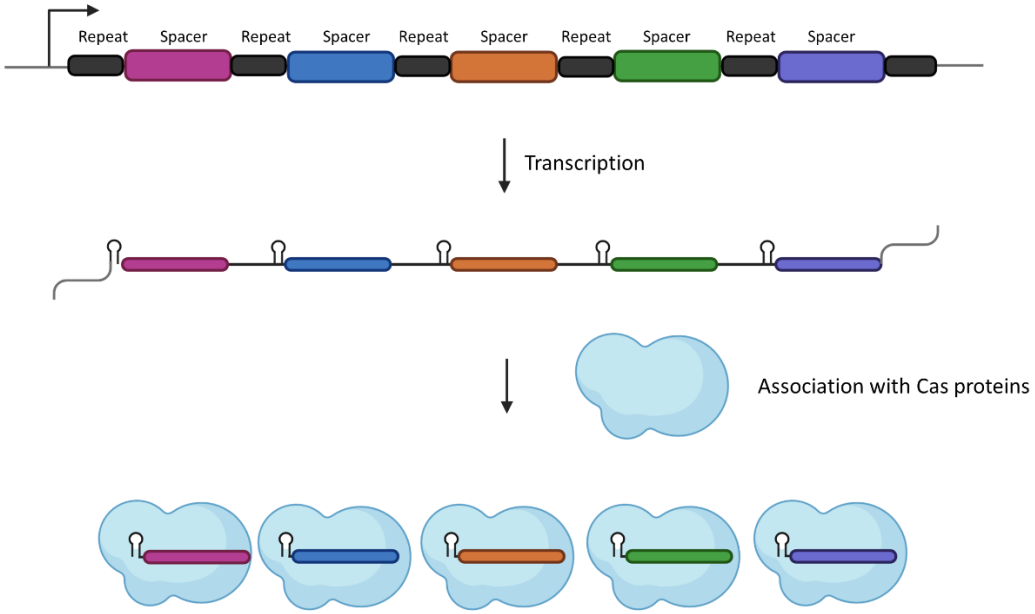



**Fig 1.13. CRISPR RNA biogenesis.** The repeat-spacer elements, or CRISPR array, are transcribed as a long pre-cursor CRISPR RNA (pre-crRNA). The repeat elements form hairpin structures, which are necessary for associations with Cas proteins. The pre-crRNA is cleaved by repeat-specific endoribonucleases, forming mature crRNAs which are loaded onto Cas proteins for targeted DNA cleavage.

With this, came the recognition that these CRISPR arrays shared homology to bacteriophage genomes and other mobile genetic elements, and that these spacer sequences could be captured by the host genome, creating a memory of previous encounters with foreign genetic elements (Mojica *et al*., 2005). The first biological evidence of a microbial nucleic-acid based adaptive immune system came in 2007, when *Streptococcus thermophilus* was infected with an invasive bacteriophage (Barrangou *et al*., 2007). *S. thermophilus* acquired new spacers identical to the genome of the invading phage and these spacers provided immunity when the bacterium was subsequently infected with the same phage type.

### 1.11.2 Mechanism of Immunity

Since their discovery, CRISPR-Cas systems have been found in ~40% bacterial and 90% archaeal genomes (Grissa *et al*., 2007). Generally, CRISPR-Cas systems involve transcription and processing of the CRISPR arrays, yielding small CRISPR RNAs (crRNAs), which bind and target Cas proteins. Immunity to a particular genetic invader occurs in three steps **(Fig 1.14)**. The first step is 'adaptation' or 'spacer acquisition', in which bacteria with one or more CRISPR loci respond to incoming genetic material by integrating a short fragment of incoming sequence, known as a protospacer into the host chromosome at the proximal end of the CRISPR array. This incorporation of new spacers into the CRISPR locus during infection creates a memory. These sequences will direct the CRISPR-Cas machinery if the same genetic material is encountered again, protecting the host (Marraffini and Sontheimer, 2010). In the second step immunity is executed. The CRISPR array is transcribed as a long precursor, or pre-crRNA (Jinek *et al*., 2012). During crRNA biogenesis, pre-crRNAs are cleaved at the repeat sequence, by a repeat-specific endoribonuclease, forming short mature crRNAs (Anzalone *et al*., 2020). The final step is recognition where the mature crRNA remains bound to Cas nucleases as a guide RNA (gRNA) to find the protospacer in the genome of the invader, which is then cleaved or inactivated by the Cas nuclease. This requires the presence of a short DNA sequence (two to five nucleotides) called the protospacer adjacent motif (PAM) near the DNA target site (Gleditzsch *et al*., 2019). This cleavage stops the infection and confers immunity to the host (McGinn and Marraffini, 2016). The pathogen can evolve mutations in the target site, enabling the invader to evade CRISPR-Cas immunity and re-establish infection (Anzalone *et al*., 2020).

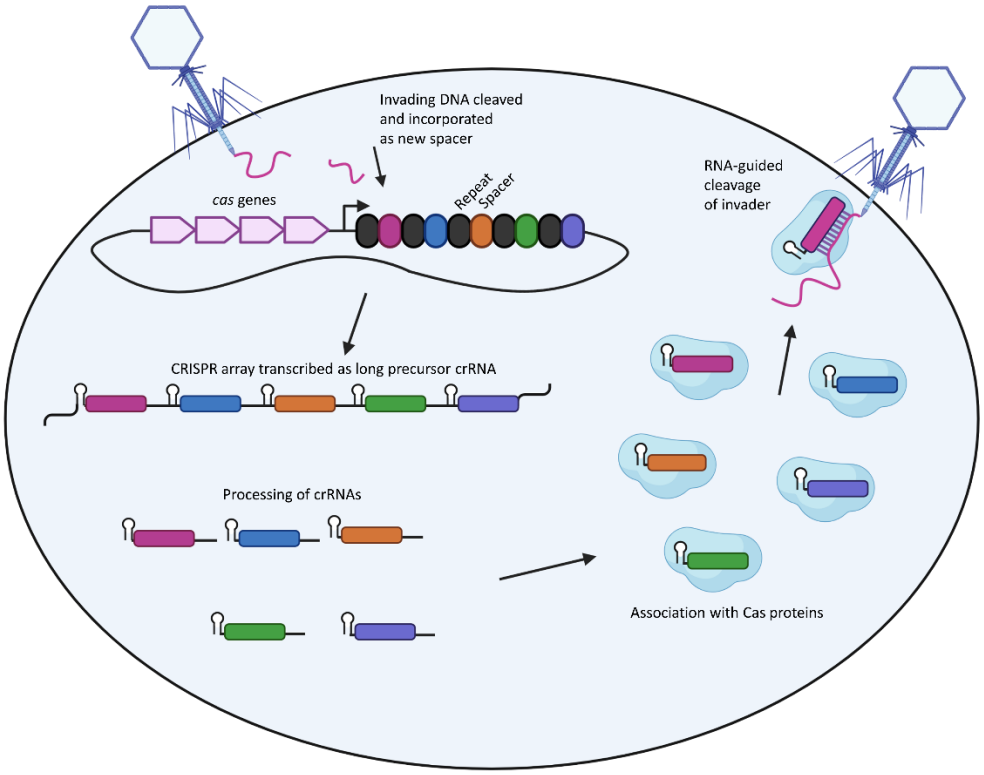



**Fig 1.14. Nucleic-acid based immunity by CRISPR-Cas systems.** Invading DNA is cleaved and incorporated into the CRISPR array as a new spacer at the proximal end. The CRISPR array is composed of short, partially palindromic repeat sequences (black), interspersed by non-homologous spacer sequences, acquired from previous infections. Immediately upstream of the CRISPR array are the CRISPR-associated (*cas*) genes, which encode proteins for cleavage of the invader. The CRISPR array is transcribed as a long precursor CRISPR RNA (pre-crRNA), which undergoes processing to form the short mature crRNAs. These mature crRNAs tightly associate with Cas proteins. In the final step, the crRNA remains loaded on the Cas protein, functioning as a guide RNA (gRNA) to find a protospacer adjacent motif (PAM) sequence in the genome of the invader. Once this sequence on the invader has been found, the Cas proteins direct cleavage on the invading genetic material, stopping the infection.

### 1.11.3 Class 1 CRISPR Interference

Since their discovery, two classes (Class 1 and 2) and six sub-types (Type I-VI) of CRISPR-Cas systems have been described, which exhibit diversity in the arrangement of the *cas* genes and the

subunits of the effector complexes (Nidhi *et al*., 2021) **(Fig 1.15)**. Class 1 systems are the most prevalent in bacteria and archaea, while Class 2 systems only make up about 10% of all CRISPR-Cas systems (Shmakov *et al*., 2017). Seven types of Class 1 systems have been described, Type I-A to I-G. Although less common, Class 2 CRISPR-Cas systems are the best studied, utilising single effector proteins (Cas9, Cas12 or Cas13) for invading nucleic acid interference (Yan *et al*., 2019). Class 2 systems are further divided into Type II (Cas9), Type V (Cas12) and Type VI (Cas13) (Nidhi *et al*., 2021). Class 1 systems involve multi-subunit CRIPSR-associated complex for antiviral defence complex (Cascade) for DNA binding, crRNA and a Cas3 nuclease (Barrangou and Marraffini, 2014) **(Fig 1.16)**. For *E. coli* (type I-E) Cascade is formed by Cse1, Cse2, Cas7, Cas5 and Cas6e proteins, while for *Pseudomonas aeruginosa* (type I-F), Cascade is formed by Csy1, Csy2, Csy3 and Cas6f proteins (Wiedenheft *et al*., 2011).

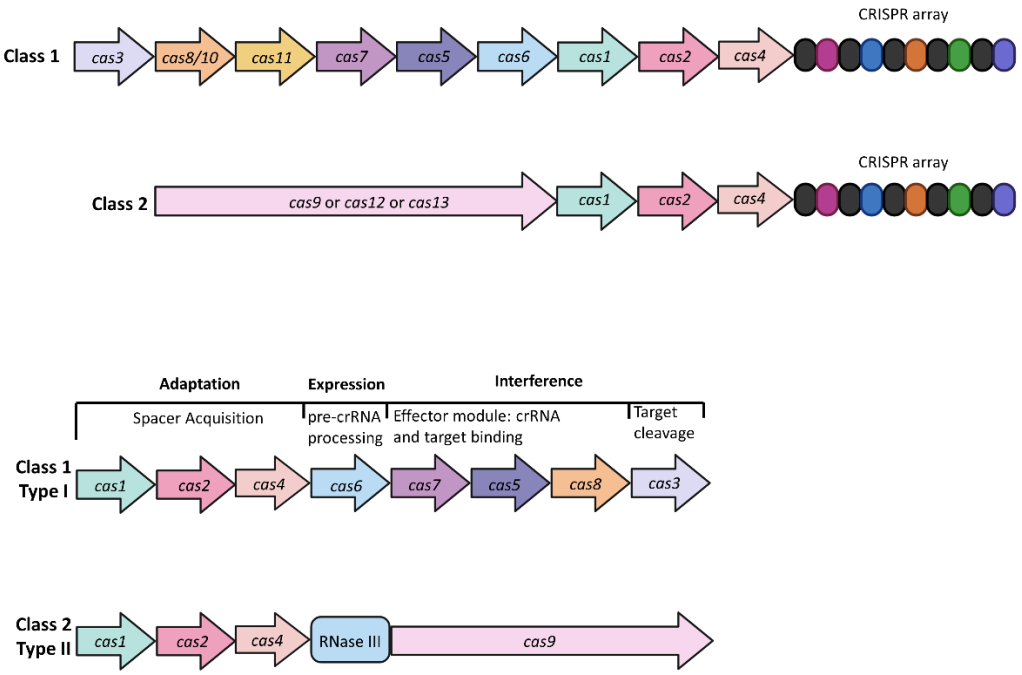



**Fig 1.15. Spatial organisation of class 1 and class 2 *cas* operons.** Two CRISPR classes have been described, Class 1 and Class 2. For both classes, the CRISPR associated (*cas*) genes are located immediately upstream of the CRISPR array. Each class is further divided into different subtypes. Class 1 type I and Class 2 type II are shown here. The *cas* loci contain the genes required for spacer acquisition (*cas1, cas2* and *cas4*), for the adaptation phase, genes for pre-crRNA processing (*cas6*, or alternatively host RNase III), and genes required for interference of invading DNA (*cas7, cas5, cas8* and *cas3* for Class 1 systems and *cas9* for Class 2 systems).

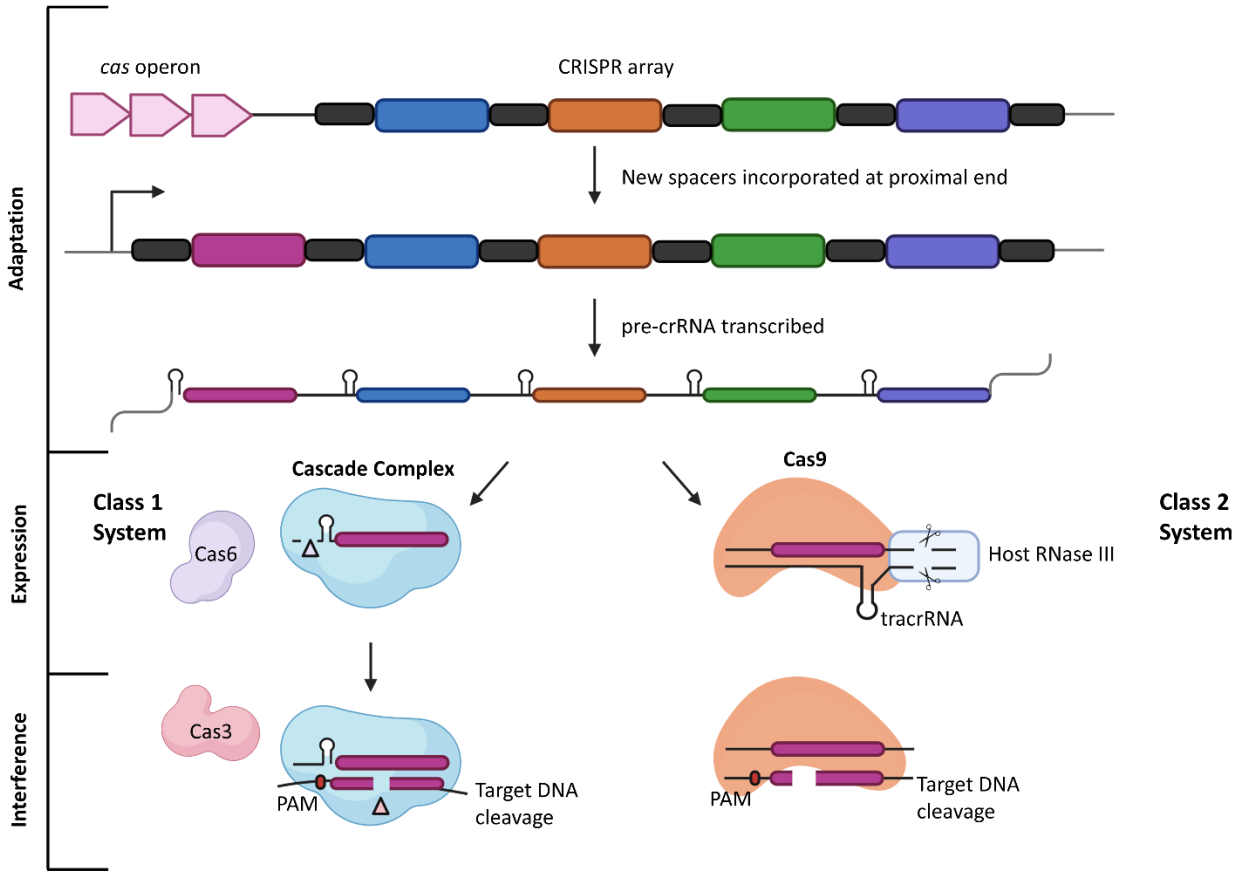



**Fig 1.16. CRISPR class 1 and class 2 interference mechanisms.** During the adaptation phase, new spacers are incorporated at the proximal end of the CRISPR array, presumably by Cas1 and Cas2. The CRISPR array is then transcribed as a long primary crRNA transcript. At this point, Class 1 and Class 2 differ in the expression and interference mechanisms. For Class I systems, the Cascade complex binds the pre-crRNA, which is subsequently cleaved by Cas6, releasing mature crRNAs. Class 2 systems utilise a small non-coding *trans*-acting RNA (tracrRNA), which pairs with the repeat regions of the pre-crRNA. This is followed by cleavage within the repeats by host RNase III. During the interference phase, invading DNA is cleaved. In class 1 systems the crRNA directs the Cascade complex to the target DNA, with the Cas3 nuclease presumably responsible for DNA cleavage. In class 2 systems, no Cas3 orthologue has been described, and Cas9 loaded with the mature crRNAs is probably directly responsible for cleaving invading DNA.

For these systems, pre-crRNA is processed at the repeat regions (Barrangou and Marraffini, 2014). Other Cas proteins bind the crRNA, assembling into a functional Cascade (Brouns *et al*., 2008). To discriminate between self- and non-self-DNA, Cascade recognises the PAM sequence immediately

upstream of the protospacer, which triggers a conformational change upon binding, allowing Cas3 to be recruited (Barrangou and Marraffini, 2014). Cas3 consists of a helicase domain, which unwinds the target DNA and a nuclease domain responsible for degradation of the target DNA (He *et al*., 2020). The mechanism of spacer acquisition is not fully understood but it does require the metal-dependent Cas1 and Cas2 proteins, as evidenced by mutations in these proteins abolishing spacer acquisition (Datsenko *et al*., 2012).

While Class 2 systems have typically been used for genetic engineering due to their simplicity, in recent years Class I systems have also been adapted for genetic engineering, for example with *Clostridium pasteurianum* (I-B) and *Pseudomonas aeruginosa* (I-F) (Pyne *et al*., 2016; Xu *et al*., 2020). For *E. coli* K-12, two distinct CRISPR array flank the *cas* operon, however, the type I-E system is regarded as non-functional owed to strong repression by the global regulator H-NS (Pul *et al*., 2010). This helps explain why the *E. coli* CRISPR-Cas system wasn't discovered until relatively recently.

### 1.11.4 Class 2 CRISPR Interference

In 2012, two groups independently discovered that *S. thermophilus* and *Streptococcus pyogenes* utilise an RNA-guided endonuclease, Cas9, for interference and that this protein could be programmed for targeted genome editing (Gasiunas *et al*., 2012; Jinek *et al*., 2012). *S. pyogenes* Cas9 complexes with two small RNAs, inducing a conformational change in Cas9 converting the protein to an active state (Le Rhun *et al*., 2019). The effector complex then scans the DNA for PAM sequences in the target DNA (Qi *et al*., 2013). If encountered, the complex unwinds dsDNA that is complementary to a 20 bp sequence downstream of the PAM, generating a double stranded break (DSB) at this point. Breakage is achieved through the concerted action of two conserved endonuclease domains in Cas9, RuvC-like and HNH-like domains, which nick the DNA in close proximity on each strand. DNA cleavage occurs 3 nucleotides upstream of the PAM, which is essential for discriminating between self and non-self DNA (Marraffini and Sontheimer, 2010).

Maturation of crRNAs is pivotal to the CRISPR-Cas system, as this molecule complexes with the Cas9 protein for DNA targeting (Jinek *et al*., 2012). In class 2 systems, crRNA processing requires binding of a *trans*-acting RNA (tracrRNA) to complementary repeat sequences in the pre-cRNA, forming an active complex. This complex is cleaved by host RNase III, which liberates small mature crRNAs that remain bound to Cas9 via tight associations with the tracrRNA. This processing creates Cas9 molecules loaded with a gRNA ready to search invading DNA molecules for its targets (McGinn and Marraffini, 2016). To maintain the optimal number of spacers in the array, old spacers located at the distal end are eliminated, potentially through homologous recombination with old CRISPR repeat sequences (Garrett, 2021). An inability to anneal results in quick release of Cas9, which samples other sequences (McGinn and Marraffini, 2016). Once the target DNA is found, Cas9 initiates the DSB through its nuclease activity (Jinek *et al*., 2012).

### 1.11.5 CRISPR-Cas9 for Genetic Engineering

The potential to utilise a DNA-specific nuclease, which could be targeted to specific DNA sequences by the expression of small RNAs, was immediately recognised (Sternberg *et al.,* 2015). While Zinc finger proteins and transcription activator-like effector nucleases (TALEN), can be programmed to bind specific DNA sequences, the genetic engineering required is time-consuming and expensive (Gaj *et al*., 2013). The idea that a myriad of DNA sequences could be targeted by simply altering the spacer sequence within the gRNA was a very attractive concept. Class 2 systems have been classically used for genetic engineering due to their simplicity, with most work focussed on the *S. pyogenes* CRISPR-Cas9 system (Qi *et al*., 2013). Typically, genome engineering by CRISPR-Cas9 requires a Cas9 protein, a tracrRNA and crRNA and a region of donor DNA to make the desired mutation (Sternberg and Doudna, 2015). This approach was further simplified by fusing the coding regions of the tracrRNA-crRNA duplex to form a chimeric single guideRNA (sgRNA), which retained the secondary structures necessary to target Cas9 to specified DNA sequences (Peters *et al*., 2015). The sgRNA is a 102-nucleotide long chimeric non-coding RNA, composed of a 20-nucleotide target

specific complementary region, a 42- nucleotide Cas9 binding 'handle' and a 40-nucleotide transcriptional terminator derived from *S. pyogenes* (Hawkins *et al*., 2015).

If the gRNA and target DNA display complementarity, the RuvC and HNH domains will cleave the target DNA (Jinek *et al*., 2012). The basis of CRISPR-Cas9 mutagenesis relies on the fact that DSBs caused by Cas9 cleavage are fatal in most bacterial genomes (Peters *et al*., 2015). sgRNAs are designed to target the 'unedited' genome and the Cas9-sgRNA is directed to the target DNA where nuclease activity in the Cas9 protein generates a lethal DSB, acting as a selection against wild-type sequences without the need to introduce a resistance cassette into the genome (Arroyo-Olarte *et al*., 2021). In this way, the DSB drives genome editing through homologous recombination, or more rarely in bacteria through non-homologous end joining (Shuman and Glickman, 2007). Thus, it is the host species DNA damage and repair systems that perform the editing. In most bacterial species, RecA mediated homologous recombination is induced to repair DNA damage by DSBs (Rosenberg *et al*., 2012). In bacteria with a low intrinsic frequency for homologous recombination, expression of recombinases (e.g., the lambda red recombinase in *E. coli*), in addition to the CRISPR-Cas machinery enhances recovery of bacteria that have undergone the desired editing (Jiang *et al*., 2013).

A conserved type II-B CRISPR-Cas locus has been reported in *C. difficile* (Sebaihia *et al*., 2006). The most prominent feature of the *C. difficile* type II-B system is the unusually high number of CRISPR arrays per genome compared to other bacterial species, averaging 8.5 arrays per genome, while other systems tend to have 1-3 arrays (Andersen *et al*., 2016). Utilising the native *C. difficile* Cas locus for genome editing is unfavourable due to a requirement for multiple protein effectors (Boudry *et al*., 2015). Rather, to allow genome editing, the *S. pyogenes* CRISPR-Cas9 has been adapted for *C. difficile* (McAllister *et al*., 2017). As the CRISPR-Cas system of *C. difficile* recognises a different PAM sequence, 5'CCW'3, compared to *S. pyogenes*, the native system does not interfere with this genetic tool.

### 1.11.6 CRISPR Interference

While the traditional CRISPR-Cas system can be used for targeted genome editing, this system cannot be extended to studies of essential genes as inactivation is lethal (Qi *et al*., 2013). Methods for manipulating essential genes are limited, and include isolation of conditional mutants (e.g., temperature sensitive mutants), or placing essential genes under the control of regulatable promoters in combination with knockout mutants (Blomfield *et al*., 1991). Unfortunately, both of these approaches are time-consuming and labour-intensive. Another method is RNA interference (RNAi), which utilises vectors expressing small antisense RNA molecules which associate with and activate protein complexes, most notably the RNA-induced silencing complex (RISC). Once bound, the complex can bind the target mRNA. Consequently, ribosomes are prevented from binding, abolishing protein synthesis and the mRNA is marked for destruction (Neumeier and Meister, 2020). Again, RNAi requires substantial engineering, and the lack of these systems in prokaryotes meant methods for global regulation of gene expression were limited (Bikard *et al*., 2013).

In bacteria, more attention has been focussed on utilising a catalytically inactive variant of Cas9, dCas9, to downregulate transcription, which provides many advantages over existing methods (Qi *et al*., 2013). This system is known as CRISPR interference (CRISPRi). For CRISPRi, the dCas9 harbours point mutations in both the RuvC-like (D10A) and HNH (H840A) domains (Todor *et al*., 2021). CRISPRi was first described by Qi *et al.* who showed that co-expression of dCas9 with an sgRNA designed with a 20 bp complementary region to a gene of interest could effectively silence gene transcription with up to 99.9% repression (Qi *et al*., 2013). As CRISPRi repression depends on base-pairing between a short segment of the sgRNA and DNA target, new targets can be specified by altering the 20 bp region, making this inexpensive for regulating gene expression. Specificity of CRISPRi is owed to a 12 bp 'seed region' within the 20 bp region for binding and a PAM sequence on the target DNA (Qi *et al*., 2013). With CRISPRi, dCas9 is only defective in cleavage but it still functions as an RNA-guided DNA binding complex, targeting the non-template DNA strand, and blocking RNA

polymerase elongation (Larson *et al*., 2013). For CRISPRi, sgRNAs targeting the non-template strand exhibit 10-300-fold repression, while sgRNAs targeting the template strand show little repression, potentially due to RNAP being able to read through the dCas9-sgRNA complex in this orientation. In this case, the sgRNA faces the RNAP and may be unzipped by the helicase activity of RNAP (Qi *et al*., 2013).

The sgRNA can target the promoter or coding region of a gene (Larson *et al*., 2013). When directed to the promoter region, the dCas9-sgRNA complex can sterically hinder association between *cis*-DNA motifs and their cognate *trans*-acting transcription factors, thereby blocking transcription initiation. Notably, when targeting promoter sequences, sgRNAs can be designed for the template or non-template strand. The silencing by the dCas9-sgRNA complex is inducible, fully reversible, and highly specific in bacterial cells. Indeed, specificity of the system can be tuned by introducing single or multiple mismatches into the 20 bp region (Larson *et al*., 2013).

While CRISPRi presents an inexpensive means for targeted gene silencing, there are several potential limitations to the method. First, targeted gene silencing using the *S. pyogenes* dCas9 requires the presence of an NGG PAM motif, which limits the availability of target sites, particularly in AT-rich genomes (Larson *et al*., 2013). Additionally, it has been shown that the *S. pyogenes* Cas9 partially recognises an NAG PAM motif, which could expand the number of targetable sites but also the off-target effects (Collias and Beisel, 2021). Unfortunately, it has been shown that CRISPRi does exhibit polar effects on downstream genes, and to a lesser extent, upstream genes (Peters *et al*., 2016).

Recently, a CRISPRi system was developed for *C. difficile*, utilising a xylose-inducible promoter to control the timing and extent of gene silencing (Muh *et al*., 2019). To validate the system, Muh *et. al* designed an sgRNA for targeted, titratable repression of the red-fluorescent protein gene (*rfp* gene). Additionally, *slpA*, *ftsz* and *cdr20291_0712* knock-down vectors were

constructed. The phenotypic findings for *slpA* silencing were consistent with those obtained previously for an S-layer null strain (Kirk *et al*., 2017b; Muh *et al*., 2019)

### 1.11.7 CRISPR Activation

In addition to gene silencing, dCas9 has been used to direct transcriptional activators to specific genomic sites for programmable targeted gene activation, known as CRISPR activation (CRISPRa) **(Fig 1.17)** (Ho *et al*., 2020). Many CRISPRa systems have been described for eukaryotes, however, far fewer examples of CRISPRa for bacteria have been shown (Chavez *et al*., 2016).

Bacterial sigma factors play a pivotal role in engaging the transcriptional initiation machinery. Sigma factors interact with the core RNA polymerase (RNAP) enzyme ($\alpha_2\beta\beta'\omega$) and bind to specific promoter sequences (Browning and Busby, 2004). Typically, transcriptional activators bind to specific components of the RNAP complex and direct the complex to the promoter sequences, enhancing transcription at these sites (Browning and Busby, 2016). However, transcriptional activators in bacteria are poorly characterised, and have not been shown to mediate transcription activation when coupled synthetically to DNA binding domains. To this end, only a handful of imperfect CRISPRa systems have been described in bacterial systems. The first, involved fusion of the dCas9 complex to the ω subunit of the RNAP complex, however, this system was only functional in a ω deletion mutant (Bikard *et al*., 2013). The second, used bacterial enhancer binding proteins (bEBPs) as the fused activation domains, but these proteins were only compatible with $\sigma^{54}$ promoters and deletion of the bEPs was required (Liu *et al*., 2019). The necessity to modify the bacterial genome limited the genetic tractability of these approaches. A third study came from Dong *et. al*, who used a scaffold RNA containing the gRNA and an MS2 domain, capable of binding an MS2-fused transcription factor SoxS, enabling transcription enhancement. While this system exhibited higher activity, a narrow targeting range within the promoter elements was reported (Fontana *et al*., 2020).

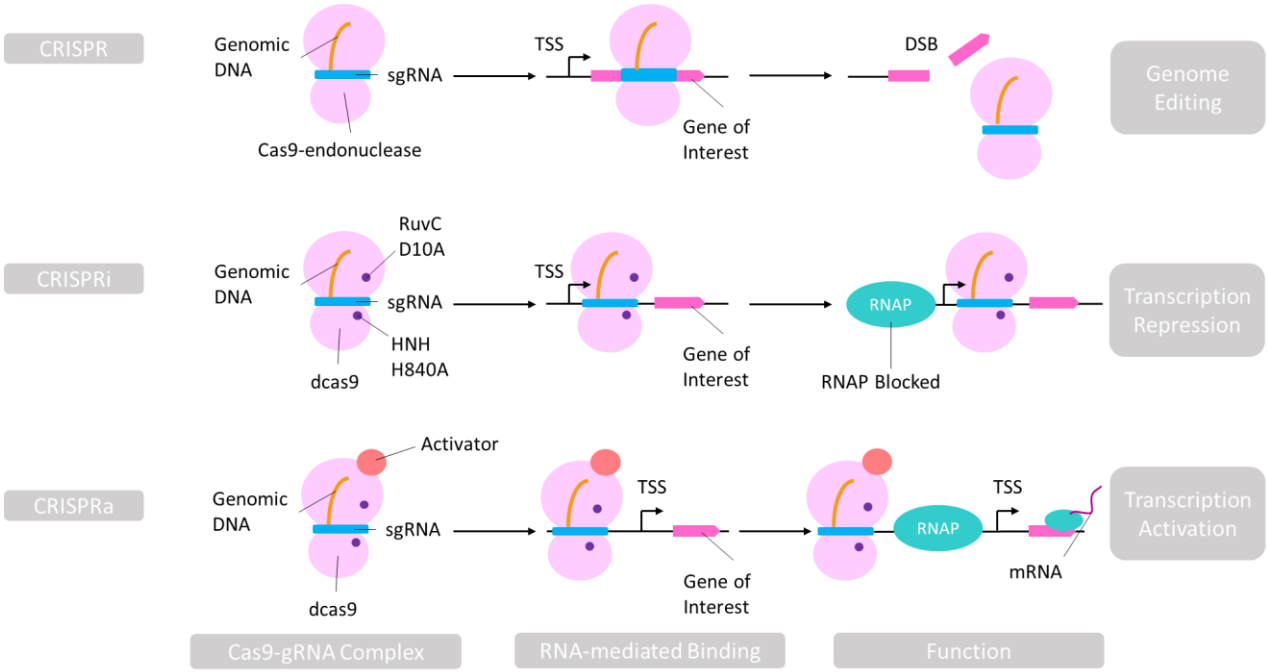



**Fig 1.17. Mechanisms of CRISPR, CRISPRi and CRISPRa systems.** For genome editing, CRISPR systems are composed of a Cas9 endonuclease, which complexes with a single guide RNA (sgRNA). The Cas9 complex is directed to the target genomic DNA sequence, where a double stranded break (DSB) is introduced through the nuclease activity of Cas9. CRISPRi and CRISPRa use a catalytically deactivated Cas9 (dCas9), with point mutations: D10A in the RuvC-like domain and H840 HNH nuclease domains. For dCas9, nuclease activity is abolished but the protein still maintains RNA-guided DNA-binding activity. For CRISPRi, the dCas9-sgRNA complex will bind to the target sequence, blocking access of the RNA polymerase (RNAP) to the underlying DNA, resulting in transcriptional repression. For CRISPRa, the dCas9 is fused to transcriptional activators, binding upstream of RNAP, resulting in up-regulation of gene expression.

## 1.12 Project Aims

The *C. difficile* S-layer represents a metabolically costly structure, and as such may provide vital cellular functions. This is reinforced by *slpA* being essential for growth and pathogenicity, unlike other bacterial species. However, selection with bacteriocins that specifically target the *C. difficile* S-layer, led to isolation of low frequency resistant mutants that had no evident S-layer due to point mutations in *slpA* (Kirk *et al.,* 2017b). With this *C. difficile* S-layer null strain, we now have the opportunity to examine which genes become conditionally essential or conditionally non-essential in the absence of this major surface structure. This experiment would help us to (i) identify genes which could compensate for the lack of S-layer and (ii) identify genes which could prevent formation of the S-layer, aiding in the development of prophylactic agents against this pathogen. To identify these changes, we attempted to construct a high-density transposon mutant library in this strain for subsequent analysis using transposon-directed insertion site sequencing (TraDIS).

With target genes obtained from the TraDIS analysis, we wanted to validate essentiality and assess function of these genes. However, traditional gene knockout methods are incompatible with study of essential genes, as inactivation results in cell death. As such, alternative methods for interrogating these genes were required. For this thesis, a CRISPRi system was developed for *C. difficile*, utilising an inducible catalytically inactive variant of Cas9, dCas9, and a programmable gRNA which binds to target sequences and disrupts transcription, resulting in a knock-down of gene expression levels.

The work presented in this thesis describes the construction of a transposon mutant library for an S-layer null strain of *C. difficile* (Chapter III), identification of conditionally essential and non-essential genes in the absence of this structure (Chapter IV), development of a CRISPRi tool for *C. difficile* to study gene function of conditionally essential and non-essential genes (Chapter V), and

use of the CRISPRi tool for studying the effects of gene silencing on targets obtained from this study, namely via transmission electron microscopy of thin sections (Chapter VI).

# Chapter II. Materials and Methods

## 2.1 Handling of Bacterial Strains

### 2.1.1 Handling of *C. difficile* Strains and Culture

*C. difficile* strains used in this study are listed in table A1. *C. difficile* strains were grown in pre-reduced tryptone yeast (TY) extract broth, composed of 3% tryptose and 2% yeast extract (Bacto) at 37°C under anaerobic conditions in a Don Whitley anaerobic workstation, with an atmosphere pressure composed of 80% $N_2$, 10% $CO_2$ and 10% $H_2$. *C. difficile* strains were grown on Brain-Heart Infusion (BHI) agar (Sigma Aldrich). Media was supplemented with thiamphenicol (15 μg/ml), colistin (50 μg/ml), lincomycin (20 μg/ml) and anhydrotetracycline (ATc) (20 ng/ml) as required.

### 2.1.2 Handling of *E. coli* Strains and Culture

*E. coli* strains used in this study are listed in table A2. *E. coli* strains were grown in Luria-Bertani broth (LB) (Fisher) or LB agar (Sigma Aldrich). Cultures were supplemented with chloramphenicol (15 μg/ml) and carbenicillin (50 μg/ml) as required and grown aerobically at 37°C with shaking. *E. coli* Neb5α were used routinely for cloning and plasmid propagation. CA434 cells were used as conjugation donors for plasmid transfer into *C. difficile*.

### 2.1.3 Production of Chemically Competent *E. coli*

An O/N of *E. coli* CA434 was set up in 4 ml LB broth. The O/N was subcultured 1/100 into 400 ml LB broth and grown with shaking at 37°C. Growth was monitored hourly until cells reached mid-exponential growth ($OD_{600nm}$ = 0.4-0.6). Cells were harvested by centrifugation at 4,000 *x g* for 10 min at 4°C. The supernatant was discarded, and the pellet resuspended in 5 ml ice-cold 100 mM $CaCl_2$. Cells were incubated on ice for 15 min. Cells were harvested by centrifugation at 4,000 *x g* for 10 min at 4°C. The supernatant was discarded, and the pellet was resuspended in 1 ml ice-cold 100

mM $CaCl_2$, 15% glycerol (v/v). Cells were incubated on ice for 2 h. 50 μl aliquots were stored at -80°C.

### 2.1.4 Transformation of Plasmid DNA into Competent *E. coli* cells.

*E. coli* Neb5α or CA434 strains were transformed with the plasmids outlined in table A3. For the transformation, vials of competent cells (stored at -80°C) were thawed on ice for 5 min. 50 μl of thawed cells were aliquoted into pre-cooled microcentrifuge tubes. To this, 4 μl of Gibson or ligation reactions or approximately 0.5 μl of purified plasmid was added to thawed cells. Cells were placed on ice for 30 min. Cells were heat-shocked at 42°C for 30 s, allowing for the uptake of exogenous DNA. The cells were then placed on ice for 2 min. 450 μl of Super Optimal broth with Catabolic repression (SOC) outgrowth media was added to each reaction and incubated at 37°C with gentle shaking for 1h. Cells were plated on LB containing the appropriate antibiotic and incubated O/N at 37°C.

### 2.1.5 Conjugative Transfer of Plasmid DNA into *C. difficile*

O/N cultures of the *E. coli* conjugation donor, CA434, and the *C. difficile* conjugation recipient were grown in LB broth supplemented with chloramphenicol and TY broth, respectively. 200 μl of each *C. difficile* culture was heat-shocked at 50°C for 10 min and then cooled to 37°C. 1 ml of *E. coli* CA434 cultures were harvested by centrifugation at 4,000 x *g* for 2 min, and the supernatant discarded. *E. coli* pellets were gently resuspended in 200 μl of heat-treated *C. difficile*. The suspension was spotted onto BHI plates and incubated at 37°C for 8-24 h under anaerobic conditions. The following day, conjugation mixtures were harvested from the BHI plates with TY broth and serially diluted onto BHI plates containing colistin and thiamphenicol. The next day, an isolated colony was restreaked to purity on colistin plus thiamphenicol plates, and subsequently on agar containing thiamphenicol alone, preventing *E. coli* outgrowth.

### 2.1.6 Storage of Strains

1 ml of *E. coli* and *C. difficile* strains were stored in cryovials containing 20% (v/v) glycerol. Strains were stored at -80°C. It should be noted that transposon libraries were stored in 10% (v/v) glycerol as we have previously shown that higher glycerol concentrations reduce library resuscitation efficiency (Dr. Nadia Fernandes, personal communication).

## 2.2 DNA Manipulation

### 2.2.1 Isolation of Plasmid DNA

*E. coli* harbouring the plasmid of interest were grown in 5 ml LB broth with the appropriate antibiotic. Cells were harvested by centrifugation at 4,000 x *g* for 10 min. Plasmid DNA was extracted using the GeneJET Plasmid Miniprep kit (Thermo) as per the manufacturer's instructions. Plasmid DNA was eluted in 60 μl of nuclease-free water. Plasmid DNA concentration and quality was quantified using $A_{260}$ spectrophotometry.

### 2.2.2 Purification of Genomic DNA from *C. difficile*

Genomic DNA was isolated from *C. difficile* using a phenol-chloroform method (Dembek et *al*., 2015). *C. difficile* strains were grown in 5 ml TY broth. 1 ml of culture was harvested by centrifugation at 4,000 x *g* for 2 min. The supernatant was discarded, and the pellet was resuspended in 200 μl PBS. 10 μl of CD27L, a *C. difficile* specific endolysin, was added to the resuspension (Peltier *et al*., 2015). The endolysin cleaves intact PG. Samples were incubated for 1 h at 37°C. 10 μl of pronase (final concentration 1 mg/ml) was added to degrade cellular proteins, followed by incubation at 55°C for 1 h. 80 μl of N-lauroylsarcosine (final concentration 2%) was added to disrupt cellular membranes, followed by incubation at 37°C for 1 h. Lastly, 200 μl RNase (final concentration 0.2 mg/ml) was added to the sample and incubated at 37°C for 1 h. The sample was transferred to a Phase Lock Gel (PLG) tube. 500 μl of Phenol:Chloroform:Isoamyl alcohol (25:24:1) was added to the PLG tube and mixed by inversion. The sample was centrifuged at 13,000

x *g* for 2 min. This allows for the separation of the DNA in the aqueous phase from the organic phase. The aqueous phase was transferred to a new PLG and treated with Phenol:Chloroform:Isoamyl alcohol (25:24:1) as before. The sample was centrifuged at 13,000 x *g* for 2 min. The DNA-containing aqueous phase was removed and transferred to a new PLG tube and treated with 500 μl chloroform:isoamyl alcohol (24:1) to remove excess phenol. The sample was centrifuged at 13,000 x *g* for 2 min. The DNA-containing aqueous phase was transferred to a new PLG tube and treated with chloroform:isoamyl alcohol (24:1) as before. The aqueous phase was transferred to a new microcentrifuge tube and an equal volume of isopropanol was added. The sample was incubated at -20°C O/N to precipitate the gDNA. The gDNA was harvested by centrifugation at 4,000 x *g* for 15 min. The DNA was washed with 70% ethanol. gDNA was harvested by centrifugation at 4,000 x *g* for 10 min, the ethanol was removed, and samples were air-dried to remove any residual ethanol. gDNA was resuspended in 50 μl of nuclease-free $H_2O$. DNA concentration and quality was quantified using an $A_{260}/A_{280}$ and $A_{260}/A_{230}$ spectrophotometer.

### 2.2.3 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) is an *in vitro* technique, which allows for the amplification of a specific region of DNA of known sequence. 2x Phusion High-Fidelity master mix (NEB) was used as per manufacturer's instructions. Each reaction (20 μl) was composed of 10 μl of commercial Phusion (2x), 1 μl each of forward and reverse primers (10 nM), 1-2 ng of template DNA and nuclease-free $H_2O$. Primers used in this study are listed in table A4. 1-5% DMSO (v/v) was added to PCR reactions when required. DMSO functions to disrupt secondary structure formation in the DNA template and was used when primers contained AT-rich regions. For *C. difficile* colony PCR, selected colonies were resuspended in 100 μl of nuclease-free water with a small quantity of chelex resin, boiled at 100°C for 10 min, briefly centrifuged to sediment the beads and 2 μl of the supernatant was used as the DNA template for reactions. Alternatively, 1 ng of plasmid DNA acted as the DNA template. Reactions began with a 30 s denaturation step at 98°C, followed by 35 cycles of 98°C (10

s), 56-58°C (10 s), 72°C (30 s/kb product). The final extension step was carried out at 72°C for 5-10 min. Sizes of PCR products were estimated by agarose gel electrophoresis.

### 2.2.4 Agarose Gel Electrophoresis

PCR products can be separated according to size using agarose gel electrophoresis. Gel electrophoresis relies on the speed at which DNA products will migrate through an agarose matrix in an electric field. Smaller products will migrate further and resolve quicker, whereas larger fragments move slowly through the gel. Sizes of PCR products were estimated by comparison with a 1 kb Gene Ruler DNA ladder (Thermo) containing fragments of known sizes. Unless otherwise stated, PCR products were electrophoresed through 0.8% agarose gels (w/v), with tris-acetate-EDTA (TAE) buffer plus SYBR Safe stain (Invitrogen), or UView dye (BioRad) for 30 min at 110 V.

### 2.2.5 Purification of PCR Products

PCR products were purified using a GeneJET PCR purification kit (Thermo) as per manufacturer's instructions. Purified PCR products were eluted in 20 μl of nuclease-free water. DNA purity and concentration was quantified using $A_{260}$ spectrophotometry. Samples were stored at -20°C until needed.

### 2.2.6 Gel Extraction of DNA

Gel extraction allows for the isolation and purification of a desired DNA fragment from an agarose gel. PCR or restriction digestion products, previously resolved on an agarose gel with UView stain and loading dye were extracted from the agarose gel. DNA samples were visualised under UV light using a transilluminator. DNA fragments of the desired size were excised from the gel using a scalpel. DNA was extracted from the agarose gel using a GeneJET gel extraction kit (Thermo) as per the manufacturer's instructions. DNA was eluted in 20 μl of nuclease-free water and DNA purity and concentration was quantified using $A_{260}$ spectrophotometry.

### 2.2.7 Restriction Endonuclease Digestion of DNA

DNA is held together by phosphodiester bonds joining adjacent nucleotides. The enzymes capable of hydrolysing these bonds are known as exonucleases and endonucleases. Exonucleases digest nucleotides from the ends of the DNA molecules whereas endonucleases hydrolyse bonds within the DNA molecule. Each enzyme has its own characteristic recognition sequence and will cut the DNA at these points. The single-stranded overhangs (sticky-ends) generated by some restriction endonuclease digests can be used for subcloning or cloning of DNA fragments into the vector of interest. All restriction endonucleases were supplied by NEB and used as per the manufacturer's instructions with the appropriate buffers. Digests for cloning were incubated at 37°C for 1-2 h. After digestion, DNA was purified as outlined in section 2.2.6. Diagnostic restriction digests were carried out to confirm the vector contained the product of interest. Reactions were incubated at 37°C for 1 h. For the processing of gDNA for Illumina sequencing, a restriction digestion step was incorporated to reduce sequencing reads being wasted on the plasmid. For this, samples were incubated O/N with the appropriate enzymes and buffers.

### 2.2.8 Ligation of DNA

Ligation reactions involve the joining of two DNA fragments. The ends of these fragments are joined together by the formation of a phosphodiester bond between the 3' hydroxyl group of one DNA fragment and the 5' phosphoryl group of another fragment, through the action of an enzyme. DNA fragments were ligated using T4 ligase (NEB) as per manufacturer's instructions. Each reaction contained 20 ng of plasmid and a 1-10--fold molar excess of insert in 1x T4 ligase buffer and nuclease-free water to make the total volume 10 μl. Reactions were incubated at room temperature for 1 h.

### 2.2.9 Gibson Assembly of DNA Fragments

This technique allows for the joining of multiple overlapping DNA fragments in a single reaction. For Gibson assembly, the NEB Hifi Assembly kit was used, composed of an exonuclease, a

polymerase, and a DNA ligase. The exonuclease creates 3' single stranded overhangs, facilitating the annealing of overhangs that are complementary at the overhanging end. The polymerase fills in the gaps within each annealed fragment while the DNA ligase repairs nicks in the assembled DNA. As a result, a double stranded DNA molecule is formed, serving as a template for transformations. For a 20 μl reaction, 50 ng of plasmid was required. The reaction was assembled in a molar ratio of 1:2:2 or 1:3, plasmid to insert. Reactions were held at 50°C for 1 h.

### 2.2.10 DNA Sequencing

DNA sequencing was carried out by Genewiz (Takeley, United Kingdom). For plasmid DNA, a 10 μl sequencing reaction contained 500 ng plasmid in nuclease-free $H_2O$ and 2.5 μl sequencing primer. For purified PCR products, the 10 μl reaction was composed of 20-80 ng/μl of the product in nuclease-free $H_2O$ and 2.5 μl of the sequencing primer. For Illumina MiSeq sequencing, 8 nM of the library samples were required. Sequencing was carried out at the Sheffield Children's Hospital. TraDIS data for FM2.5 was analysed by Dr. Roy Chaudhuri (University of Sheffield, UK).

## 2.3 Insertion Mutagenesis Library Generation for *C. difficile*

### 2.3.1 Transposition Frequency of an S-layer Null Strain

Calculating transposition frequency is required to understand the number of cells required to obtain a highly dense insertion mutagenesis library. This is achieved through comparison of CFU counts between induced and non-induced samples harbouring the transposition plasmid, pRPF215 **(Fig 2.1)**.

Transposition frequency of FM2.5/pRPF215 was deciphered by the following protocol: strains harbouring the transposition plasmid were restreaked to BHI-thiamphenicol, two days before the transposition assay. FM2.5/pRPF215 was grown overnight in 5 ml TY. The following morning, the $OD_{600nm}$ of FM2.5/pRPF215 was measured. FM2.5/pRPF215 was subcultured to 0.05 in triplicate in 5 ml TY. The $OD_{600nm}$ was monitored hourly. Cultures were grown to log phase ($OD_{600nm}$ 0.4-0.8). At log

phase, cultures were serially diluted in TY in a 96 well plate (270 µl TY and 30 µl FM2.5/pRPF215). 100 µl of culture was plated onto both BHI-thiamphenicol (for CFUs of the non-induced cells) and BHI supplemented with ATc (20 ng/ml) and lincomycin (20 µg/ml). ATc will induce transposition in pRPF215, while lincomycin will select for any bacteria harbouring the *ermB* transposon.

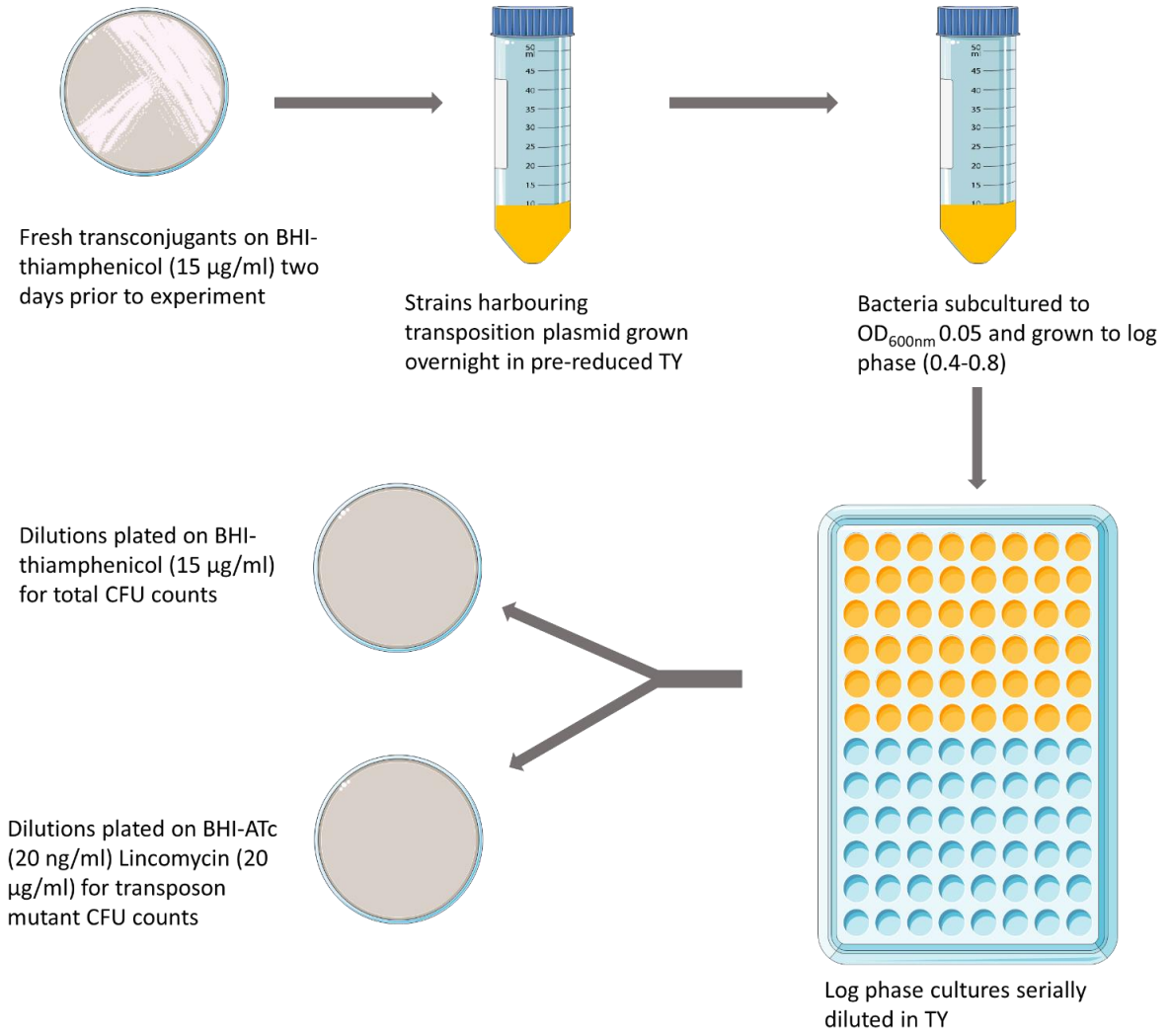



**Fig 2.1. Method for determining transposition frequency for large scale mutagenesis library generation**. pRPF215, was conjugated into *C. difficile* strains by the method outlined in section 2.1.5. Transconjugants were restreaked to fresh BHI supplemented with the plasmid selection marker, thiamphenicol (15 µg/ml). *C. difficile* strains harbouring pRPF215, were grown overnight in pre-reduced TY media. Overnights were subcultured to $OD_{600nm}$ 0.05. The $OD_{600nm}$ was monitored hourly until cultures reached log phase ($OD_{600nm}$ 0.4-0.8). At log phase, cultures were serially diluted in TY and plated onto BHI-thiamphenicol and BHI supplemented with the transposition inducer, ATc (20 ng/ml) and lincomycin (20 µg/ml).

### 2.3.2 Transposon mutant libraries constructed in liquid media

Dr. Nadia Fernandes (The University of Sheffield) constructed libraries for FM2.5 in liquid culture, by inoculating 400 ml TY with 0.5% glucose with a defined volume of freshly conjugated FM2.5/pRPF215 overnight culture to an $OD_{600nm}$ of 0.03. Once cultures had reached $OD_{600nm}$ of 0.05 (~ 1 h), cultures were supplemented with the transposition inducer ATc (20 ng/ml) and the transposon selectable marker, lincomycin (20 µg/ml). After 8 h, culture was taken for sample storage and gDNA extraction. At this point the libraries were carried over for processing by me. Libraries were processed as outlined in section 2.4 and analysed on an Illumina MiSeq.

### 2.3.3 Transposon mutant libraries constructed on solid agar

Insertion mutagenesis libraries for FM2.5 were constructed on solid agar from overnight cultures. Given the transposition frequency from overnights is $6.72 \times 10^{-2}$ (~1/700), 700,000 bacteria were plated onto each plate to give rise to an estimated 1,000 transposon mutants. When grown overnight FM2.5 reaches an $OD_{600nm}$ range of 1.2-1.8. From $OD_{600nm}$ vs. CFU/ml experiments carried out by Dr. Joseph Kirk, The University of Sheffield, we knew that at an $OD_{600nm}$ of 1.2 there are ~ $1.7 \times 10^8$ bacteria per ml. 300 µl was plated onto the transposition inducer plates, which is $5.1 \times 10^7$ bacteria, requiring an 85x dilution to obtain 1,000 transposon mutants per plate. Each overnight culture was diluted in TY-thiamphenicol one at a time, with each diluted overnight plated onto one stack of 15 plates, giving 150 plates total. Bacteria were incubated overnight. After 24 h, culture was taken for sample storage and gDNA extraction. Libraries were processed as outlined in section 2.4 and analysed on an Illumina MiSeq.

## 2.4 Processing Transposon Mutagenesis Library Samples

### 2.4.1 Genomic DNA Extraction and Shearing

gDNA was processed as per the method described by Goodall *et. al.,* (2018) **(Fig 2.2)**. 1 µg of purified gDNA (resuspended in 130 µl 10 mM Tris-HCl, pH 8.5) was required for shearing. Shearing was carried out by Sheffield Children's Hospital with the 300 bp protocol on the Covaris sonicator. The Covaris sonicator requires 130 µl, but the next processing step only accommodates 55.5 µl. As such, sample volume was reduced through freeze-drying at -80°C. Freeze drying does not consistently sublimate at the same rate every time, so samples were thawed every 20 min to ensure the volume was at or below 55.5 µl. The volume of samples was determined and topped up to 55.5 µl with nuclease free $H_2O$ if required.

### 2.4.2 NEBNext End Prep Kit Ultra I

Shearing generates blunt-ended DNA, which is repaired, and to this, adaptors are ligated. DNA ends were repaired with the NEBNext End Prep Kit Ultra I. The reaction was set up as follows with all reagents being kept on ice: 3.0 µl End Prep Enzyme Mix, 6.5 µl End Repair Reaction Buffer (10x) and 55.5 µl fragmented DNA.

All reagents were kept on ice, except enzymes which were stored in a cold-block. Reactions were mixed by pipetting. Samples were heated on a thermocycler as follows: 20°C for 30 min, then 65°C for 30 min. Samples were held at 4°C until needed for subsequent steps.

### 2.4.3 Adaptor Ligation

Adaptors were ligated to the ends of the newly repaired DNA fragments. The following reagents were added directly to the End Prep reaction mix: 15 µl Blunt/TA Ligase Master Mix, 2.5 µl NEBNext Adaptor for Illumina Sequencing, 1 µl Ligation Enhancer, 65 µl End Prep Reaction Mix.

Reactions were mixed by pipetting and samples were incubated at 20°C for 15 min in a thermocycler. 3 µl of USER™, a uracil-specific excision reagent, was added to the ligation mixture from the NEBNext Singleplex or Multiplex Oligos for Illumina kit. NebNext adaptors contain a hairpin loop structure, which is removed by the USER enzyme, opening up the loop, making it an available PCR substrate. Reactions were mixed and incubated at 37°C for 15 min.

### 2.4.4 Size Selection of Adaptor Ligated DNA

AMPure XP beads were removed from the fridge and allowed to come to room temperature. Beads were resuspended thoroughly by vortexing. Ligation reactions were topped up to 100 µl with nuclease-free water. Ligation reaction mixtures were resuspended with 55 µl beads. The volume of beads determines the size of fragments to be selected, with this volume selecting for fragments of ~ 300 bp. The mixture was held at room temperature for 5 min before transferring to a magnetic stand, separating the beads from the supernatant. After 5 min, once the solution is clear the supernatant was removed and transferred to a fresh microcentrifuge tube. The beads contain the unwanted large DNA fragments. Caution was taken to not disturb the beads. 45 µl of beads were added to the supernatant and mixed by pipetting. The mixture was held at room temperature for 5 min and transferred to the magnetic stand for another 5 min. Once the supernatant was clear (~5 min), it was removed and discarded as the beads now contain the fragments at ~300 bp. Caution was taken not to disturb the beads.

Beads were washed with 200 µl freshly prepared 80% ethanol. The tube was incubated for 30 s at room temperature, after which the ethanol was removed and discarded. The ethanol wash was repeated for each sample. Once the ethanol was carefully removed, beads were air-dried for 5 min and residual ethanol was removed. Beads were resuspended in 17 µl 10 mM Tris-HCl and held at room temperature for 5 min before placing the beads on the magnetic stand for a further 5 min. Once the supernatant had been separated from the beads, 15 µl was transferred to a PCR tube.

### 2.4.5 PCR Enrichment of the Transposon Junction

The next step is an enrichment PCR for the transposon junction fragments using a custom forward primer, RF1520, which anneals to the transposon end, and a custom reverse primer, RF1522, which anneals to the ligated adaptor, amplifying the gDNA-transposon junction. The following components were added to the PCR reaction: 15 µl Adaptor Ligated DNA Fragments (our samples), 25 µl KAPA HiFi Polymerase, 2.5 µl 10 µM PCR forward primer (RF1520), 2.5 µl 10 µM PCR reverse primer (RF1522) and 5 µl nuclease-free water.

The PCR was placed on a thermocycler on the following program: 98°C (3 min), 98°C (15 s), 65°C (30 s) and 72°C (30 s) for 10 cycles. A final extension of 1 min at 72°C was carried out. Samples were held at 4°C until required. Samples were digested with EcoRI-HF to circumvent plasmid contamination. EcoRI was the most suitable candidate enzyme, which cuts the plasmid once. Consequently, reads amplified from the plasmid are extremely short and negligible. Additionally, this enzyme cuts infrequently throughout the genome, 96 times. The reaction was as follows: 50 µl of the PCR reaction, 6 µl of Cutsmart buffer and 4 µl of EcoRI-HF. Restriction digests were carried out to completion overnight at 37°C. EcoRI-HF was heat-inactivated at 65°C for 15 min.

### 2.4.6 Clean-up of PCR Amplification

The PCR reaction was transferred to a fresh microcentrifuge tube. Vortexed AMPure XP beads (45 µl) were added to the PCR reactions and mixed by pipetting. The mixture was held at room temperature for 5 min and subsequently added to the magnetic stand. Once the supernatant had separated from the beads (~5 min), the supernatant containing unwanted DNA targets was removed and discarded. Beads were washed with 80% freshly prepared ethanol as described in section 2.3.4. Beads were air dried for 5 min. Caution was taken not to over dry the beads, lowering recovery of target DNA. Beads were resuspended in 17 µl of 10 mM Tris-HCl and held at room temperature for 5 min. Once the supernatant was removed from the beads (~5 min), 15 µl of the supernatant was removed to a PCR tube.

### 2.4.7 Second PCR Amplification for Library Preparation

The second PCR prepares the library for sequencing through the addition of flow cell adaptors, P5 and P7, which are on the outermost ends of the fragments **(Fig 2.3)**. These provide a priming site for the samples to bind to on the MiSeq flow cell. Just inside from these are illumina adaptors, providing sequencing primer binding sites. Next are two barcodes: the illumina index, increasing index capacity while staggering introduction of the transposon and the illumina barcode for sample identification. The following were added to the PCR reaction: 15 µl Adaptor Ligated DNA Fragments (our samples), 25 µl KAPA HiFi Polymerase, 2.5 µl 10 µM inline index custom forward primer, 2.5 µl 10 µM illumina index primer and 5 µl nuclease-free water.

The inline index custom forward primer is unique to each sample being processed while the illumina index primer is unique to each condition, i.e., *C. difficile* strain. Reactions began with a 98°C initial denaturation step (3 min) followed by 20 cycles of 98°C (15 s), 65°C (30 s), 72°C (30 s). The final extension step was carried out at 72°C for 1 min and samples were held at 4°C until required.

### 2.4.8 Clean-up of Second PCR Amplification

PCR reactions were purified as outlined in section 2.3.6. The beads were resuspended in 33 µl of 10 mM Tris-HCl and held at room temperature for 5 min. The mixture was placed on a magnetic stand until the supernatant had separated from the beads (~ 5 min). 32 µl of the prepped libraries were transferred to a fresh microcentrifuge tube. Libraries were stored at -20°C until sequenced.

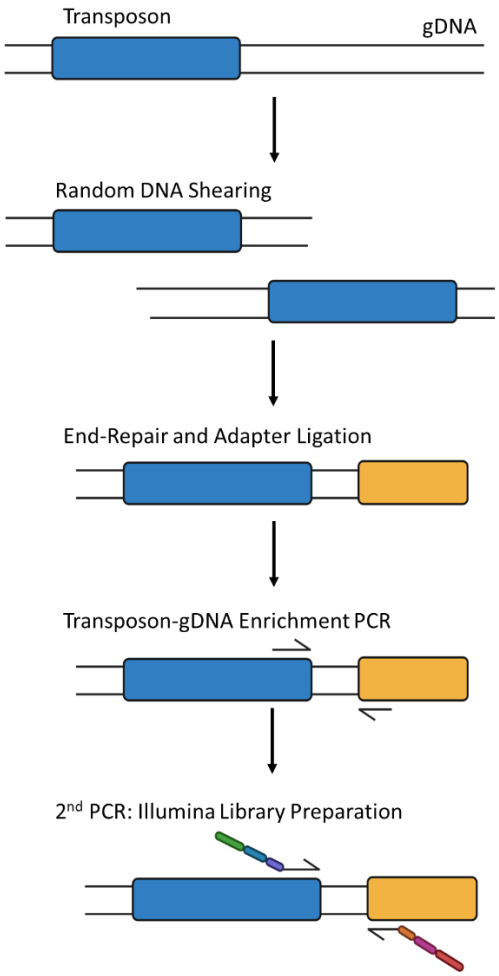



**Fig 2.2. Preparation of genomic DNA for sequencing.** gDNA is extracted from the pooled library of mutants and fragmented by sonication, generating blunt-end DNA. The DNA is repaired, and to this, adaptors are ligated. Next, an enrichment PCR of the gDNA-transposon junction selects for fragments harbouring transposon insertions, using a transposon specific and ligated adaptor specific primer. In a second PCR step, fragments are prepared by addition of sequencing specific adaptors and barcodes.

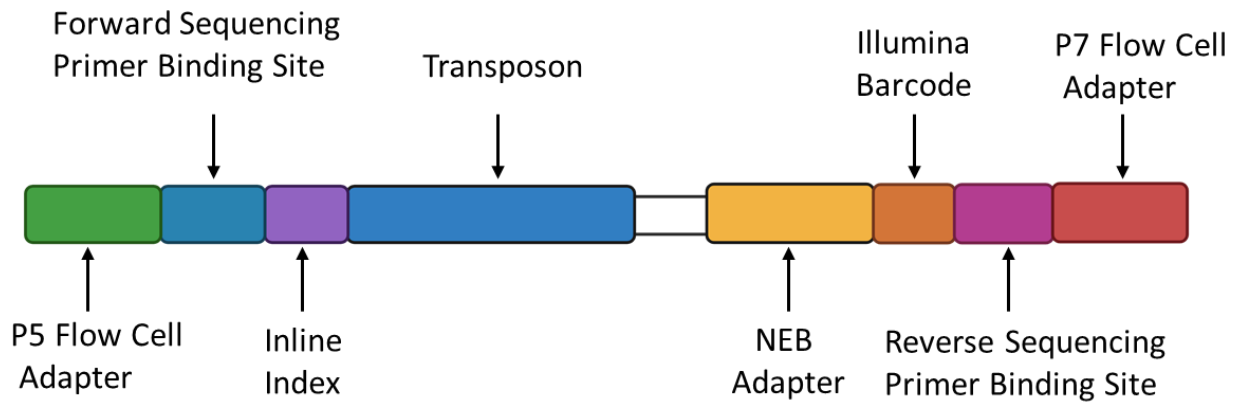



**Fig 2.3. gDNA fragments for Illumina sequencing.** The P5 and P7 adaptors are required for binding to the Illumina flow cell (green and red). Illumina forward and reverse adaptors, provide the sequencing primer binding sites (blue and pink). The inline index staggers the introduction of the transposon (purple), while the illumina barcode is used to identify samples (orange). The NEB adaptor (yellow) increases ligation efficiency.

### 2.4.9 Illumina MiSeq Sequencing

Prepped library samples were adjusted to 8 nM and 1.5 µl of each prepped library was pooled together in a fresh microcentrifuge tube. Libraries were loaded onto the Illumina MiSeq with 150 bp paired-end reads, sequencing both ends of the transposon-containing gDNA fragments. Sequencing was carried out at the Sheffield Children's hospital. Demultiplexing of TraDIS samples was carried out by Dr. Roy Chaudhuri, The University of Sheffield. Reads from the Illumina sequencing run came back as fastq files from Sheffield Children's Hospital. Fastq is a text-based file which contains the nucleotide sequences and the quality score. Bio::TraDIS was run on the individual FM2.5 libraries and one lane of the R20291 Dembek *et. al* data [ERR377408] from one sample (Dembek *et al*.,2015; Barquist *et al*., 2016). The Bio::TraDIS pipeline provides software utilities for processing, mapping, and analysing transposon insertion sequencing data. At the core of the Bio::TraDIS pipeline is the bacteria_tradis script, which filters reads in the fastq format for transposon tags, removing the tag. The modified reads were then mapped to the *C. difficile* R20291 genome and the transposition plasmid, pRPF215, using the bwa short read mapper. To produce an

initial prediction of gene essentiality, the output from bacteria_tradis was run through the tradis_essentiality.R script, which quantifies the number of insertions per gene **(Table 4.2)**. The insertion frequency for each gene was normalised against gene length to generate an insertion index where genes with low insertion scores are predicted to be essential for growth. The tradis_essentiality.R script attempts to fit distributions to the two modes observed for insertion indexes **(Fig 1.9)**. This produces a table of putative essential and ambiguous genes, along with plots that can evaluate the predictions (Barquist *et al*., 2013).

https://hactar.shef.ac.uk/TraDIS_201119

### 2.4.10 Tools for Computational Analysis of the Conditional Genome of FM2.5

Conditionally essential and non-essential genes identified in this study were translated to amino acid sequences. Amino acid sequences were run through BLAST, Pfam, TMHMM and SignalP servers to assess protein similarity to other known sequences and cellular location of proteins, providing insights into the function of these identified genes and how these genes may translate into targets for blocking S-layer assembly. Descriptions of these servers are outlined below:

**BLAST:** Finds regions of local similarly between nucleotide, or protein sequences against a database of known genome sequences (Johnson *et al*., 2008). For the FM2.5 TraDIS analysis, protein sequences were compared to other microbial genomes via BLAST.

**Pfam:** is a widely used database that provides classification of protein families and domains (Finn *et al*., 2014). Pfam contains multiple sequence alignments for each protein family, as well as profile hidden Markov models for finding these domains in new protein sequences. Presented are Pfam matches obtained for conditionally essential and conditionally non-essential genes of FM2.5, and the regions these domains span.

**TMHMM:** For any given protein sequence, TMHMM, gives the number of predicted transmembrane helices, the expected number of amino acids in the transmembrane helices (if the number is greater

than 18 then it is likely to be a transmembrane helix) (Krogh *et al*., 2001). Additionally, this software provides the probability that the N-terminus is cytoplasmic. The algorithm is based on statistical analysis of TMbase, a database of naturally occurring transmembrane proteins. Predictions are made using several weight-matrices for scoring. Protein sequences from the FM2.5 TraDIS data set were analysed for the presence of transmembrane helices.

**SignalP 5.0:** This server predicts the presence of signal peptides in Archaea, Gram-positive bacteria, Gram-negative bacteria, and Eukaryotes (Almagro Armenteros *et al*., 2019). For bacteria, SignalP 5.0 can distinguish between Sec/SPI signal peptides, in which proteins are secreted by the Sec translocon, Tat signal peptides, secreted by the Tat translocon and Sec/SPII lipoprotein signal peptides. For the Signal5-P software, a probability close to 1 means a highly reliable prediction. For Archaea, Gram-positive and Gram-negative bacteria, the probability threshold is 0.25, as there are four possible classes of signal peptide (Sec/SPI, Tat/SPI, SPII and Other). A probability close to this threshold means a very unreliable prediction. Protein sequences from the FM2.5 TraDIS data set were analysed for the presence of signal peptides.

**KEGG Pathway:** Kyoto Encyclopaedia of Genes and Genomes (KEGG) Pathway shows a compilation of molecular interactions and biochemical reactions based on manually verified pathways (Kanehisa and Goto, 2000).

## 2.5 Protein Manipulation

### 2.5.1 Preparation of Cell Wall Protein Extracts

*C. difficile* strains were grown O/N in 5 ml pre-reduced TY. Cells were harvested by centrifugation at 4,000 x *g* for 10 min. The supernatant was discarded, and cells were normalised to $OD_{600nm}$ 50 in 0.2 M glycine (pH 2.2). Suspensions were incubated at room temperature for 30 min with rotation. Samples were centrifuged at 20,000 x *g* for 2 min and the supernatant, containing the cell wall proteins were transferred to a fresh microcentrifuge tube. The pH was adjusted to 6-8 with 2 M Tris.

### 2.5.2 SDS-PAGE

Proteins were separated according to size using denaturing polyacrylamide gels as outlined in table 1.

**Table 1. Recipe for SDS-PAGE resolving gel. Sufficient for two 1.0 mm mini-gels.**

| | 6% | 8% | 10% | 12% | 15% |
|---|---|---|---|---|---|
| 30% acrylamide/bis-acrylamide | 2 ml | 2.7 ml | 3.3 ml | 4 ml | 5 ml |
| 1.5 M Tris-HCl, pH 8.8 | 2.5 ml | 2.5 ml | 2.5 ml | 2.5 ml | 2.5 ml |
| $H_2O$ | 5.4 ml | 4.7 ml | 4.1 ml | 3.4 ml | 2.4 ml |
| 10% SDS | 100 µl | 100 µl | 100 µl | 100 µl | 100 µl |
| 10% APS | 50 µl | 50 µl | 50 µl | 50 µl | 50 µl |
| TEMED | 10 µl | 10 µl | 10 µl | 10 µl | 10 µl |

The resolving gel was poured and set with a layer of isopropanol and allowed to polymerise for 1 h. The isopropanol was removed, and the stacking gel was poured as outlined in table 2.

**Table 2. Recipe for SDS-PAGE stacking gel. Sufficient for two 1.0 mm mini-gels.**

| | 5% |
|---|---|
| 30% acrylamide/bis-acrylamide | 833 µl |
| 0.5 M Tris-HCl, pH 6.8 | 1.25 ml |
| $H_2O$ | 2.87 ml |
| 10% SDS | 50 µl |
| 10% APS | 25 µl |
| TEMED | 5 µl |

Samples were combined with an equal volume of Laemmli buffer (150 mM Tris-HCl pH 6.8, 1.5% (w/v) SDS, 15% (v/v) β-mercaptoethanol, 30% (v/v) glycerol and 2 μg/ml bromophenol blue) (Laemmli, 1970). Electrophoresis was performed using the mini-protean tetra cell apparatus (Bio Rad). A colour prestained broad range marker was used (11-245 kDa). 10 μl of sample was loaded to each well. Gels were run at a constant voltage of 190 V until the dye front reached the bottom of the gel ~ 60 min.

### 2.5.3 Coomassie Blue Staining

When required, gels were stained O/N in Coomassie blue (45% methanol, 10% acetic acid and 1 mg/ml brilliant blue R-250). Gels were destained using 45% methanol, 10% acetic acid.

### 2.5.4 Western Immunoblotting

Western blotting allows for detection of a specific protein from samples, using an antibody probe that is specific for the protein. Protein samples were separated on a denaturing gel as outlined in section 2.4.2. Gels were equilibrated in Cathode buffer (25 mM Tris-HCl, pH 9.4, 40 mM glycine, 10% (v/v) methanol) for 10 min at room temperature. An Immobilon-P PVDF membrane was briefly soaked in methanol, incubated in Milli Q water for 2 min and then equilibrated in Anode II buffer (25 mM Tris-HCl, pH 10.4, 10% (v/v) methanol) for 5 min. Transfer stacks were assembled from anode to cathode as follows: two pieces of Whatmann filter paper soaked in Anode I buffer (0.3 M Tris-HCl, pH 10.4, 10% (v/v) methanol), followed by a piece of filter paper soaked in Anode II buffer, the Immobilon-P PVDF membrane, the equilibrated gel and finally three pieces of filter paper soaked in cathode buffer. Transfers were run at 15 V for 15 min for one gel or 15 V for 25 min for two gels. After transfer, membranes were stained with Poncaeu S (0.5% (w/v) Ponceau S, 1% acetic acid) for 2 min at room temperature to assess the quality of protein transfer. Membranes were washed with Milli Q water until protein bands were visible. Membranes were dehydrated with methanol and dried by incubation at room temperature O/N. The primary antibody was diluted appropriately in PBS with 3% skimmed milk. Membranes were incubated with the primary antibody at room

temperature for 1 h with gentle agitation. Membranes were washed four times with PBS and incubated with secondary horseradish peroxidase (HRP) conjugated antibodies in PBS with 3% skimmed milk for 1 h at room temperature with gentle agitation. Membranes were washed four times with PBS. Blots were developed using an enhanced chemiluminescent (ECL) substrate (Bio-Rad), which covered the membrane for 2 min at room temperature. Blots were visualised using the Bio-Rad ChemiDoc imaging system. A white image was also taken to visualise the molecular marker.

## 2.6 Growth Analysis of *C. difficile* Strains

*C. difficile* strains were grown O/N in 5 ml pre-reduced TY supplemented with antibiotic as required. The optical density of cells ($OD_{600nm}$) was adjusted to 0.01-0.05 with cultures being diluted in fresh pre-reduced TY. Growth was monitored by measuring the $OD_{600nm}$ at 1 h intervals over an 8-10 h period.

## 2.7 Thin-Sectioning of *C. difficile* samples

Thin sectioning was performed by Christopher Hill at the electron microscopy unit at The University of Sheffield. After O/N fixation with 2.5% Glutaraldehyde, cells were further fixed with 2% Osmium Tetroxide. Samples were dehydrated with increasing concentrations of ethanol, followed by Propylene Oxide. Cells were embedded in an araldite resin and sectioned at 85 nm on a Leica U6 ultramicrotome. Sections were transferred to copper coated grids and stained with uranyl acetate and lead citrate. Samples were visualised with a FEI Technai BioTWIN TEM at 80 kV fitted with a Gatan MS600CW camera.

## 2.8 Bioinformatic Analysis

Geneious software 8.1.9 was used routinely for DNA sequence alignment, *in silico* cloning, sgRNA target design for CRISPRi and analysing the RNA-fold of the sgRNAs. Primers used for Gibson assembly were designed on Nebbuilder (http://nebbuilder.neb.com).

# Chapter III. Optimising Transposon Mutagenesis Library Generation in *C. difficile*.

## 3.1 Introduction

Directed gene inactivation often relies on assumptions about the contribution of an individual gene to a phenotype. Instead, it may be advantageous to generate large pools of mutants across an entire genome, which can be screened simultaneously. Transposon directed insertion site sequencing (TraDIS) couples transposon mutagenesis with short DNA fragment sequencing, generating large random libraries of mutants, which can be phenotypically screened, allowing a direct link between genotype and phenotype to be drawn. This technique requires the construction of a high-density library, in which all non-essential genes contain the transposon insert, followed by growth of the library through defined conditions.

Identifying the core genome of *C. difficile* would enable all essential genes for growth and disease to be identified. Often, essential genes will encode components of novel metabolic or biosynthetic pathways, aiding in the development of therapeutics against this pathogen. TraDIS has been applied to *C. difficile* R20291 using transposon mutants generated on solid rich agar media to identify genes essential for growth (Dembek *et al.*, 2015). For R20291, a library containing more than 77,000 unique mutants was obtained. From this, a set of 404 genes were identified as essential for growth *in vitro*, and 798 genes were identified, which were likely to affect sporulation, a process critical for *C. difficile* transmission. Generation of transposon mutants relies on a conditional mariner delivery vector, pRPF215 **(Fig 3.1)**. This plasmid displays replicational instability when induced with the tetracycline nonantibiotic analogue, anhydrotetracycline. The inducible expression system is composed of a pair of divergent promoters, each with overlapping tet operator sequences (*tetO*). $P_{tetR}$ drives expression of *tetR*, encoding the system repressor and $P_{tet}$ drives expression of the mariner transposase. $P_{tetR}$ is oriented towards the pCD6 replication origin. There is no transcriptional terminator following *tetR*, and, as a result, transcriptional read-through disrupts the origin of

replication, blocking further plasmid replication. Moreover, unlike *C. difficile* 630, R20291 does not harbour *tetM*, a tetracycline resistance gene, making pRPF215 a good candidate for the selection of transposon mutants (Stabler *et al*., 2009). The plasmid also contains an *E. coli* ColE1 origin of replication, a *catP* gene conferring chloramphenicol/thiamphenicol resistance and a transposon containing the *ermB* gene (conferring erythromycin resistance). Upon addition of tetracycline, the transposase is expressed, and recognises the inverted terminal repeats flanking the transposon, cutting the sequence from the DNA at these points. The transposon is re-inserted into any sequence containing the recognition sequence, TA, which is prevalent across the *C. difficile* genome as it is 70% AT.

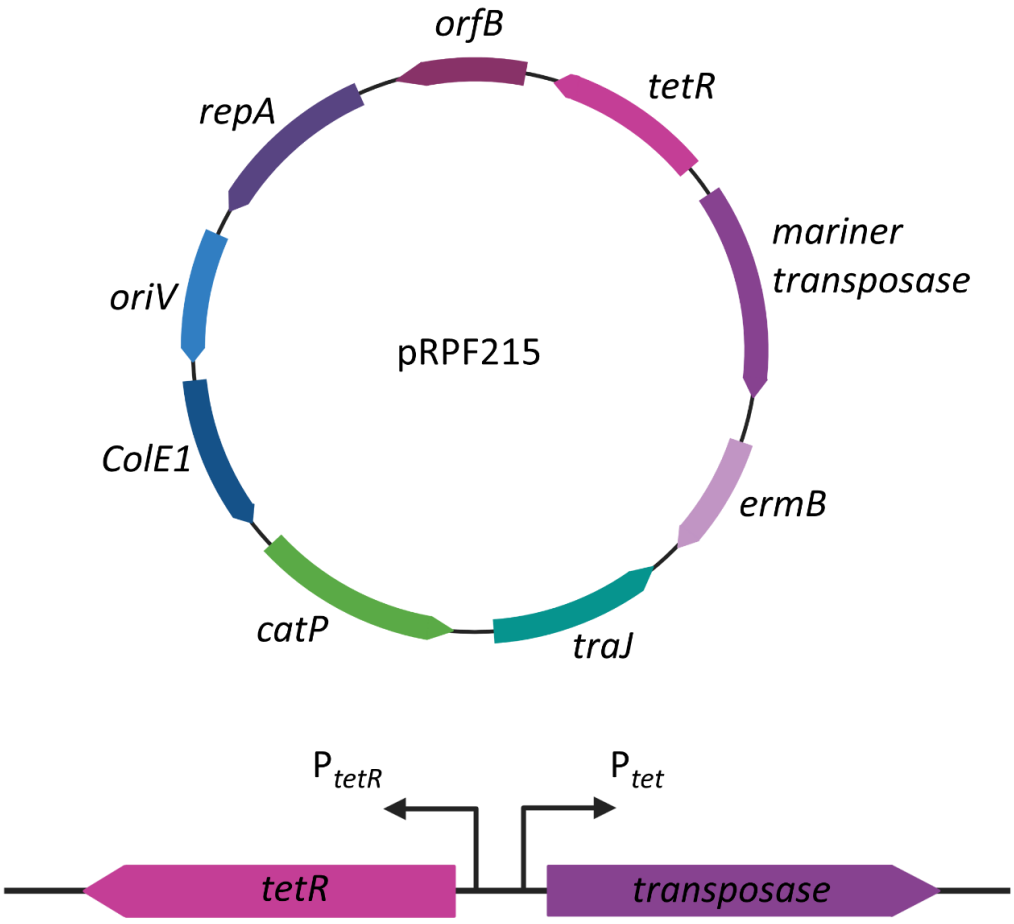



**Fig 3.1. *C. difficile* transposon mutagenesis system.** A shuttle vector containing a tetracycline-inducible mariner transposase and an *ermB* transposon.

One of the genes identified as essential for growth *in vitro* encodes the surface layer protein, SlpA. This major surface structure has been suggested to be involved in a myriad of cellular processes, including sporulation, toxin production and resistance to innate immunity effectors (Kirk *et al.*, 2017b). Additionally, the S-layer is postulated to be involved in induction of the host immune response and colonisation of host tissues (Fagan and Fairweather, 2014).

Previous attempts to knock-out *slpA* have been unsuccessful. However, using bacteriocins that specifically target the S-layer, mutants which resisted killing by avidocin AV-CD291.2 due to a mutation in *slpA* arose at a frequency of $< 1 \times 10^{-9}$ (Kirk *et al.*, 2017b). This mutation was predicted to truncate the protein at a site N-terminal to the post-translation cleavage site, preventing S-layer formation. One of these mutant strains, FM2.5, lacks detectable SlpA subunits. The *C. difficile* S-layer is a major surface structure, which is metabolically costly to produce. As we now have a strain that doesn't produce an S-layer, we speculate that other genes might become conditionally essential or non-essential in the absence of this major surface protein. For example, in the absence of the S-layer we may postulate that the protein responsible for S-layer secretion, encoded by *secA2*, may become conditionally non-essential. Additionally, we speculate that genes involved in PSII biosynthesis, the anionic polymer responsible for S-layer surface attachment may become non-essential in the absence of SlpA.

Identification of changes in gene essentiality between R20291 and FM2.5 was achieved through the construction of a high-density transposon mutagenesis library for the S-layer null strain. Ideally, scaling up library production could obtain a higher number of unique mutants than the 77,000 obtained from the R20291 TraDIS dataset, giving rise to a higher resolution data set with fewer number of base pairs between transposon insertions. This improved data resolution would allow essentiality of several genes that were previously ambiguous to be ascertained. TraDIS was applied to the FM2.5 mutagenesis library, to identify genes that become conditionally essential and conditionally non-essential in the absence of the *C. difficile* S-layer.

## 3.2 Results

### 3.2.1 Anhydrotetracycline Tolerance of an S-layer Null Strain

Before generating large scale transposon mutagenesis libraries, the fitness of the S-layer mutant, FM2.5, to the transposition inducer needed to be assessed. For the published *C. difficile* insertion mutagenesis libraries, transposition was induced with ATc at a concentration of 100 ng/ml. R20291 and 630 both have an S-layer, which to some degree may provide a barrier of protection against ATc. FM2.5 does not produce this surface layer, and as such there could be differences in sensitivity to ATc in this strain. When generating these transposon libraries on a larger scale, a balance between transposition induction and a steady growth rate is desirable. To assess the tolerance of FM2.5 to ATc, the S-layer null strain and wild-type, R20291, were grown in increasing concentrations of the inducer and the $OD_{600nm}$ was monitored hourly for each concentration **(Fig 3.2)**. Overnight cultures of R20291 and FM2.5 were grown in 5 ml of pre-reduced TY. Overnights were subcultured to $OD_{600nm}$ 0.05 in TY supplemented with increasing concentrations of ATc: 20 ng/ml, 40 ng/ml, 80 ng/ml and 160 ng/ml ATc. TY lacking the inducer was used as a control. Growth rate of each strain was monitored hourly over 8 h by measuring the $OD_{600nm}$. Experiments were carried out in triplicate using biological duplicates. Mild sensitivity to ATc was observed at 40 ng/ml for R20291 and FM2.5, and the growth defects were more profound with higher concentrations of inducer. All strains grew as expected at 20 ng/ml ATc. Consequently 20 ng/ml ATc was used to induce transposition for our libraries.

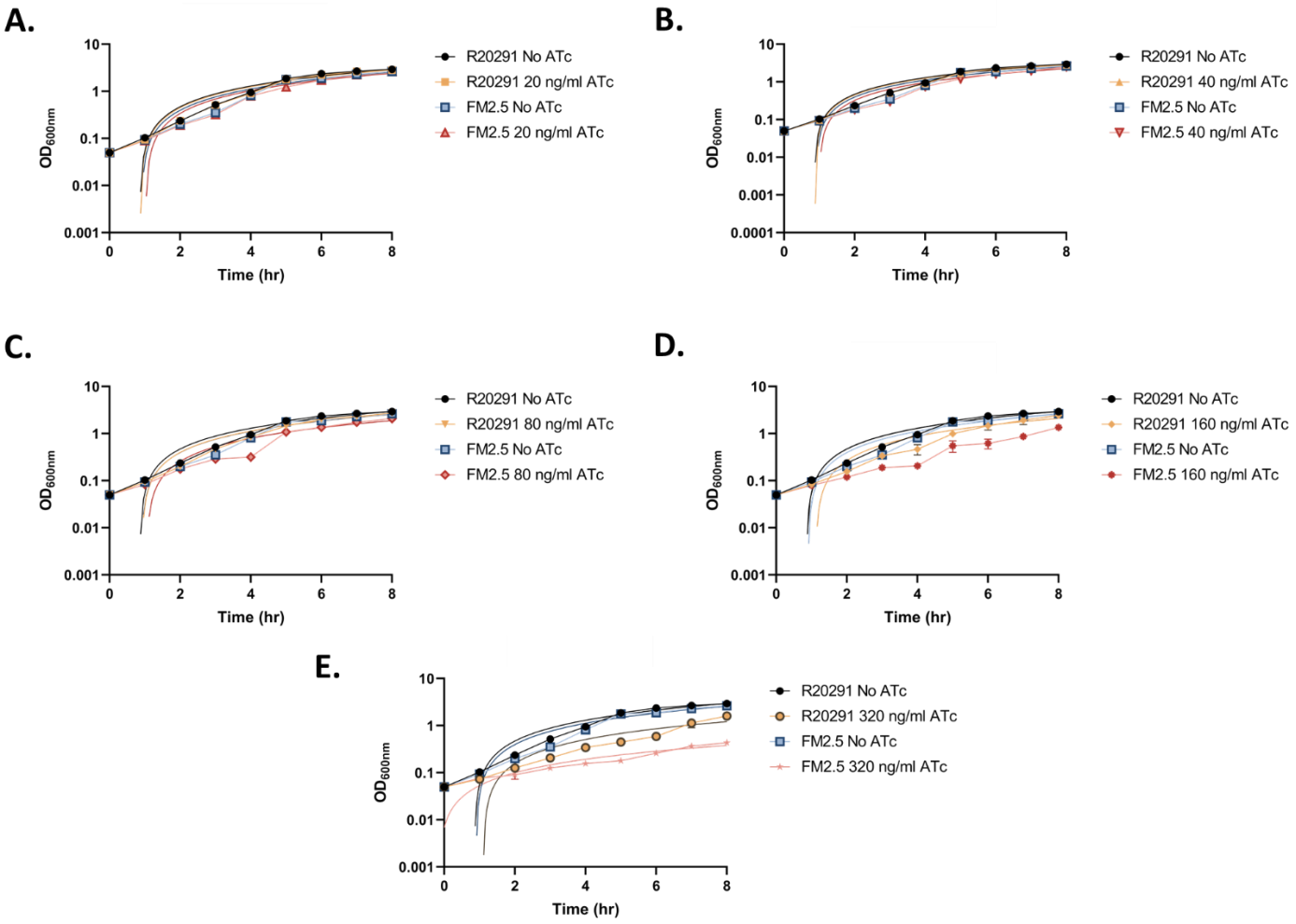



**Fig 3.2. An S-layer null strain is less tolerant to the transposition inducer than its wild-type counterpart.** *C. difficile* strains R20291 (wild-type) and FM2.5 (S-layer null strain) were grown overnight in 5 ml TY broth. Overnights were subcultured to an $OD_{600nm}$ of 0.05 in fresh pre-reduced TY containing increasing concentrations of anhydrotetracycline (ATc): 20 ng/ml, 40 ng/ml, 80 ng/ml and 160 ng/ml. $OD_{600nm}$ was monitored hourly for 8 h. Experiments were carried out in triplicate using biological duplicates. Shown are the mean, standard deviation and a linear regression model fitted to the growth data to determine if FM2.5 is less tolerant to the transposition inducer than its wild-type counterpart.

### 3.2.2 Transposition Frequency of an S-layer Null Strain

Calculating transposition frequency is required to understand the number of cells required to obtain a highly dense insertion mutagenesis library. This is achieved through comparison of CFU counts between induced and non-induced samples harbouring pRPF215 **(Fig 3.3)**. Transposition frequency of FM2.5/pRPF215 was deciphered by the protocol outlined in section 2.3.1**.**

Colony morphology was monitored after 24 h. Varying colony morphology was observed on BHI plates supplemented with ATc and lincomycin, indicating successful transfer of the transposon. CFUs were counted for induced and non-induced samples and tabulated **(Table 3.1)**. Transposition frequency was calculated by the formula below:

$$\text{Transposition Frequency} = \frac{\text{No. of colonies on BHI-thiamphenicol}}{\text{No. of colonies on BHI- ATc-Lincomycin}}$$

**Table 3.1. CFU/ml of induced and non-induced FM2.5 exponentially growing cultures**

| Strain | CFU/ml BHI-thiamphenicol | CFU/ml BHI-ATc-lincomycin |
|---|---|---|
| **FM2.5/pRPF215 replicate 1** | 6.00E+07 | 1.18E+05 |
| **FM2.5/pRPF215 replicate 2** | 6.80E+07 | 1.12E+05 |
| **FM2.5/pRPF215 replicate 3** | 8.2E+07 | 1.71E+05 |

**Table 3.2. Transposition frequency of induced FM2.5 exponentially growing cultures**

| Strain | Transposition Frequency | Mean Transposition Frequency |
|---|---|---|
| **FM2.5/pRPF215 replicate 1** | 1.97E-03 | |
| **FM2.5/pRPF215 replicate 2** | 1.65E-03 | 1.9E-03 |
| **FM2.5/pRPF215 replicate 3** | 2.09E-03 | |

### 3.2.3 Large Scale Transposon Mutagenesis Libraries in Liquid Media

Previously, insertion mutagenesis libraries were constructed in liquid media by Dr. Nadia Fernandes, The University of Sheffield. Unlike solid agar, in liquid media, cells are given ~ 16 h in a competitive environment. The best compromise for minimising competition effects would be to grow the library through fewer generations (~8). Additionally, when reducing generations of growth there is a trade-off as we want to avoid bottlenecks and a lack of mutant diversity while also sufficiently diluting the plasmid from the population. Theoretically, 8 generations should be sufficient to reduce plasmid contamination, with pRPF215 pCD6 replication origin assuming 6 plasmid copies per cell.

From the preliminary data set, insertion mutagenesis library in liquid media generated a highly saturated mutant pool with 1.2 million chromosomal reads. However, no essential genes were identified in the data set, which is likely due to the library being grown through too few generations to be able to isolate essential genes. Furthermore, a large percentage of sequencing reads came from the transposition plasmid, wasting the majority of sequencing reads. Given the apparent limitations of this approach, we instead decided to further optimise the methodology of making mutant libraries on agar with the aim of improving the data resolution of previous published TraDIS data for *C. difficile*.

### 3.2.4 Improving Data Resolution for Agar Based Biological Libraries

Previously, approximately 750,000 colonies were harvested in liquid broth, yielding 77,000 unique mutants (Dembek *et al*., 2015). Scaling up to 1.5 million colonies could potentially generate a greater number of unique mutants. A higher insertion density would allow the essentiality of smaller genes lost in the previous analysis to be ascertained. For agar based biological libraries for FM2.5, pRPF215 was conjugated into FM2.5 as outlined in section 2.1.5. Five overnights of FM2.5/pRPF215 were grown in 10 ml TY-thiamphenicol (15 µg/ml). Overnights were subcultured to $OD_{600nm}$ 0.05 in 10 ml TY-thiamphenicol and grown to $OD_{600nm}$ 0.3. To obtain 1.5 million lincomycin resistant

mutants, subcultures were diluted 1:10 in TY as per the transposition frequency of FM2.5 in section 2.3.1. 300 µl of culture was plated onto BHI supplemented with ATc (20 ng/ml) and lincomycin (20 µg/ml). Bacteria were harvested in TY with ATc and lincomycin for sample storage and gDNA extraction. Libraries were processed as outlined in section 2.4.

Illumina sequencing runs are costly, therefore we wanted to ensure that there was no issue with plasmid contamination and wasted sequencing reads. Genewiz offers an Amplicon-EZ service. With this service, 50,000 reads are provided, sufficient to detect significant plasmid contamination. For this, the custom inline indices are removed and replaced with partial adaptors by PCR with oligonucleotides RF1522 and RF1640. PCR products were column purified as outlined in section 2.2.5. DNA concentration was measured by Qubit. 500 ng of DNA were prepared at a concentration of 20 ng/µl. Samples were sent for sequencing in duplicate.

For the FM2.5 agar based biological library, 88% of sequencing reads were mapped to the plasmid and only 12% to the chromosome, indicating plasmid contamination was a major issue in our library preparations **(Table 3.3) (Fig 3.3)**.

**Table 3.3. Number of reads mapped to the R20291 chromosome and plasmid pRPF215 from the first solid agar library attempt**

| Sample | Number of reads with the transposon tag | Number of reads mapped to the chromosome | % Reads mapped to the chromosome | Number reads mapped to the plasmid | % Reads mapped to the plasmid |
|---|---|---|---|---|---|
| **FM2.5 (1)** | 69,183 | 8,663 | 12.52 | 60,520 | 87.47 |
| **FM2.5 (2)** | 75,509 | 8,983 | 11.89 | 66,526 | 88.1 |
| **Mean** | | | 12.2 | | 87.79 |

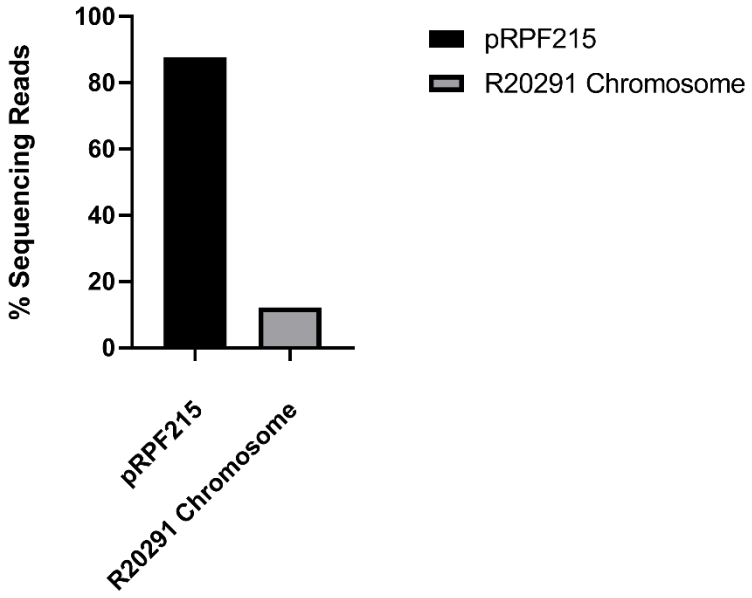



**Fig 3.3. Percentage reads mapped to R20291 chromosome and transposition plasmid pRPF215.** The mean number of sequencing reads mapped to the transposition plasmid, pRPF215, and the R20291 chromosome, were plotted as a percentage of the total number of mappable reads from duplicate samples.

### 3.2.5 Enzymatic Digestion Effectively Removes Plasmid Contaminants During gDNA Processing.

With ~ 87% sequencing reads being lost to the plasmid, a restriction digest step using EcoRI-HF was incorporated into our library processing preparation, as outlined in section 2.4.5. Digested and undigested samples were run in duplicate on a HiSeq lane to compare the efficiency of plasmid curing by this method. A HiSeq lane should give up to 300 million sequencing reads instead of 30 million from a MiSeq lane. This should circumvent the issue of plasmid contamination as the number of remaining reads should be sufficient to do an essential/non-essential gene analysis. Chromosomal mapping of sequencing reads increased from 12% to 62%, when the digestion was incorporated **(Table 3.4) (Fig 3.4)**.

**Table 3.4. Comparison of the number of reads mapped to chromosome and plasmid pRPF215 following EcoRI digestion.**

| Sample | Number of reads with the transposon tag | Number of reads mapped to the chromosome | % Reads mapped to the chromosome | Number of reads mapped to the pRPF215 plasmid | % Reads mapped to the plasmid |
|---|---|---|---|---|---|
| **FM2.5 Undigested (1)** | 965,135 | 83,111 | 9.5 | 795,780 | 90.5 |
| **FM2.5 Undigested (2)** | 3,497,283 | 254,596 | 8.1 | 2,874,906 | 91.9 |
| **Mean** | | | 8.8 | | 90.95 |
| **FM2.5 Digested (1)** | 7,857,796 | 3,082,119 | 71.2 | 1,247,989 | 28.8 |
| **FM2.5 Digested (2)** | 9,300,784 | 3,133,403 | 52.9 | 2,792,593 | 47.1 |
| **Mean** | | | 62.05 | | 37.95 |

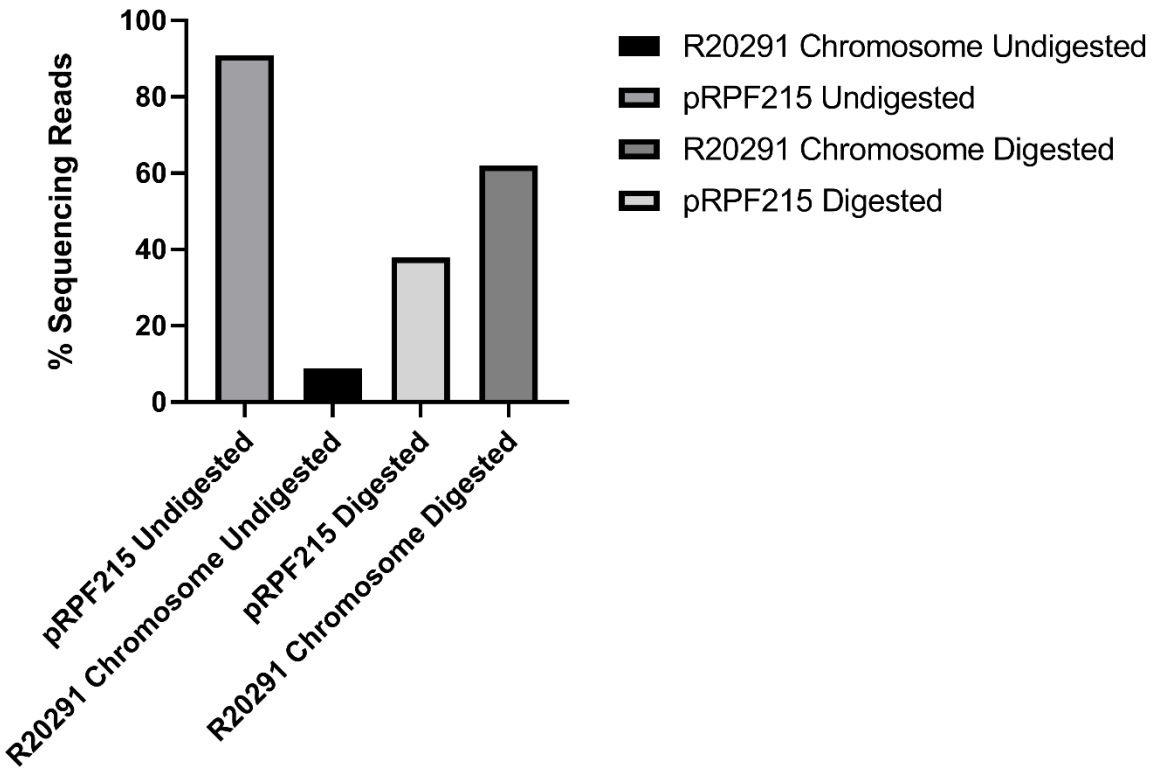



**Fig 3.4. Enzymatic digestion reduces plasmid contamination during gDNA library processing.** Undigested and digested gDNA samples were run on an illumina HiSeq in duplicate. The digested samples were digested with EcoRI after the ligation of adaptors to the blunt-end DNA. The mean number of sequencing reads mapped to the transposition plasmid, pRPF215, and the R20291 chromosome, were plotted as a percentage of the total number of mappable reads containing the transposon tag from duplicate samples.

However, a large portion of reads were still wasted on the transposition plasmid. To assess if the enzymatic digestion step could be improved, the step was moved after the enrichment PCR of the transposon junction, which amplifies only fragments harbouring the transposon. Samples were then processed as outlined from section 2.4.6 onward.

**Table 3.5. Number of reads mapped to the chromosome after enzymatic digestion.**

| Sample | Number of reads with the transposon tag | Number of reads mapped to the chromosome | % Reads mapped to the chromosome |
|---|---|---|---|
| **FM2.5 Digested (1)** | 868,361 | 845,528 | 97.37 |
| **FM2.5 Digested (2)** | 2,007,420 | 1,942,291 | 96.76 |
| **Mean** | | | 97.06 |

Altering the position of the restriction digest resulted in ~97% mapping of sequencing reads to the chromosome. We now had a defined method for removal of contaminating plasmid DNA when making transposon libraries.

Upon removal of the plasmid from our samples, it then became apparent that our libraries also contained dominating mutants that monopolised sequencing reads, with insufficient mapping in the remaining reads to do the desired essential/non-essential gene analysis. We postulated that the presence of a dominating mutant is due to the leakiness of the $P_{tet}$ promoter, allowing low frequency early transposition prior to mutant selection, followed by amplification of the resulting mutant through the library selection procedure.

### 3.2.6 A Multifactorial Approach to Prevent an Early Transposition Event

Generation of transposon mutants took 5 days from initial delivery of pRPF215 into *C. difficile*, involving multiple restreaks without induction, increasing the likelihood of an early transposition event occurring, especially if the $P_{tet}$ promoter which drives transcription of the transposase is leaky. The first precaution to be taken was reducing the time the strain containing the plasmid was cultured prior to library generation. This was achieved by harvesting the conjugation after 8 h instead of 24 h, and plating bacteria on BHI-colistin thiamphenicol instead of BHI-

cycloserine thiamphenicol. Cycloserine is bacteriostatic, preventing *E. coli* growth whereas colistin is bactericidal, killing *E. coli*. Swapping these antibiotics reduces the restreaks required to eliminate the *E. coli* from two to one. Combined, these refinements halved the entire process from 5 to 2.5 days. Additionally, previous transposition experiments used exponentially growing cells, potentially selecting for a dominating mutant in samples and adding yet more growing time. Selecting for mutants directly from overnight cultures would reduce the number of generations the bacteria go through, reducing the chances of selecting for bacteria that have undergone an early transposition event. A final precaution that was taken was increasing the number of transconjugants used to construct the library from five to ten. Each overnight culture, representing an independent transconjugant, was plated onto 15 plates, giving 150 plates total. This process was then repeated, allowing 1.5 million colonies to be harvested. To use overnight cultures to induce transposition, the transposition frequency of overnight cultures was determined as outlined in section 2.3.1 **(Fig 3.5)**.

**Table 3.6. CFU/ml of *C. difficile* FM2.5 mid-log and overnight cultures harbouring the transposition plasmid pRPF215.**

| | Strain | CFU/ml BHI-thiamphenicol | CFU/ml BHI-ATc-lincomycin |
|---|---|---|---|
| **Mid-Log Cultures** | **FM2.5/pRPF215 replicate 1** | 2.43E+06 | 1.50E+05 |
| | **FM2.5/pRPF215 replicate 2** | 1.19E+06 | 1.30E+05 |
| | **FM2.5/pRPF215 replicate 3** | 8.00E+05 | 5.00E+04 |
| **Overnight Cultures** | **FM2.5/pRPF215 replicate 1** | 7.80E+06 | 2.00E+05 |
| | **FM2.5/pRPF215 replicate 2** | 1.13E+07 | 6.00E+05 |
| | **FM2.5/pRPF215 replicate 3** | 5.70E+06 | 7.00E+05 |

**Table 3.7. Transposition frequency of *C. difficile* FM2.5 mid-log and overnight cultures.**

| | Strain | Transposition Frequency | Mean Transposition Frequency |
|---|---|---|---|
| **Mid-Log Cultures** | **FM2.5/pRPF215 replicate 1** | 6.17E-02 | |
| | **FM2.5/pRPF215 replicate 2** | 1.09E-01 | 7.78E-02 |
| | **FM2.5/pRPF215 replicate 3** | 6.25E-02 | |
| **Overnight Cultures** | **FM2.5/pRPF215 replicate 1** | 2.56E-02 | 6.72E-02 |
| | **FM2.5/pRPF215 replicate 2** | 5.31E-02 | |
| | **FM2.5/pRPF215 replicate 3** | 1.23E-01 | |

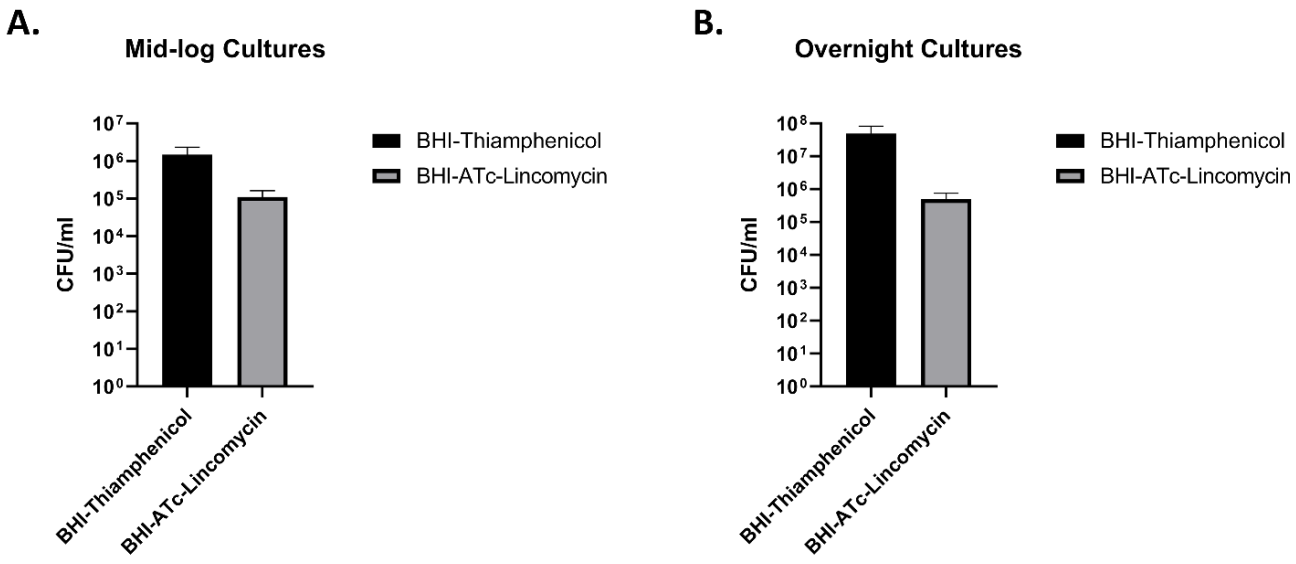



**Fig 3.5. Comparison of transposition frequency from mid-log and overnight cultures.** *C. difficile* strains harbouring pRPF215, were grown overnight in pre-reduced TY media. To assess the transposition frequency of mid-log cultures, overnights were subcultured to $OD_{600nm}$ 0.05. The $OD_{600nm}$ was monitored hourly until cultures reached log phase ($OD_{600nm}$ 0.4-0.8). At log phase, cultures were serially diluted in TY and plated onto BHI-thiamphenicol for total CFU counts and BHI supplemented with the transposition inducer, ATc (20 ng/ml) and lincomycin (20 µg/ml). To assess transposition frequency of overnight cultures, bacteria were serially diluted in TY and plated onto BHI-thiamphenicol and BHI supplemented with ATc and lincomycin. Cultures were spotted in triplicate and the CFU/ml for induced and non-induced samples was determined. Shown are the means and standard deviations.

The next day, plates were examined to determine if an early transposition event had occurred in any of the 10 independent starters, indicated by a set of plates with higher colony density than expected. Each stack was then pooled one at a time in TY-ATc-lincomycin. Samples from each pool were taken for long-term sample storage at -80°C and gDNA extraction. gDNA was extracted from 10 stacks as outlined in section 2.2.2.

To assess if the issue of an early transposition event had occurred, sample QC was performed using the Amplicon-EZ service described above to assess the number of reads containing the transposon tag. For the FM2.5 TraDIS, no dominating mutants were observed. 30-40,000

insertions were obtained for the library, with 20-30,000 of those insertions supported by multiple reads. This was sufficient coverage to do an essential/non-essential gene analysis between R20291 and FM2.5. The optimised transposition method developed in this chapter is described in section 2.4.

## 3.3 Discussion

In summary, we have now generated a transposon mutant library in *C. difficile* FM2.5, to observe changes in gene essentiality in the absence of an S-layer. The method of making transposon libraries in liquid media vs solid agar, was first adopted to reduce the amount of space used in the anaerobic cabinets and the labour involved with harvesting mutants from plates. Library generation in liquid media has several potential advantages: it is less costly, quicker and less labour intensive, with the possibility of generating several libraries for different strains in parallel. Generating libraries by this method may give rise to higher density transposon libraries, as mutant density per ml is greater in liquid media than the mutant density per square cm on agar. Ideally, this would allow clarification of essentiality in both directions. However, library generation in liquid media presented several challenges. First, *C. difficile* cannot be transformed, as such the transposon must be delivered on a plasmid, resulting in a trade-off between efficient plasmid curing and obtaining a highly saturated library. Secondly, in liquid media, cells are given ~ 16 h of growth in a competitive environment. Consequently, , experimental aims change from identifying the core essential genome in an S-layer null strain to identifying changes in gene essentiality when *C. difficile* lacking this structure are competing for the same nutrients and resources. To minimise competition effects we reduced the number of cell generations that libraries were grown through. However, by this method we were unable to ascertain any essential genes as the library was growth through too few generations. Also, growing the library through too few generations would not sufficiently dilute the transposition plasmid from the library, which would waste valuable sequencing reads in the downstream processes. Hesitancy to grow the libraries through further generations made us decide to go back and adjust the protocol for generating insertion libraries on solid agar.

For these libraries, transposition was induced using 20 ng/ml ATc, due to concerns of ATc sensitivity of FM2.5. Higher concentrations of ATc (20-500 ng/ml) have been used previously for transposition induction in *C. difficile* R20291 and 630 (Dembek 2015). However, varied colony

morphology was observed after induction with 20 ng/ml, indicative of the occurrence of a transposition event. For the first solid agar mutant library, a large number of sequencing reads came from the transposition plasmid, which could be due to the low concentrations of ATc used in the experiment and a consequent reduction in the efficiency of plasmid curing. To ensure there was no problem with wasted sequencing reads from the plasmid a restriction digestion step was incorporated after the enrichment PCR. For FM2.5, EcoRI was chosen, which cuts a site present on the plasmid once, adjacent to the transposon inverted terminal repeat, and cuts infrequently throughout the genome. Digestion of samples with EcoRI meant that reads amplified from the plasmid were short and negligible. Introduction of a restriction digestion step reduced plasmid sequencing reads from 87% to 3%. Additionally, from the preliminary FM2.5 dataset, dominating mutants were also observed, which monopolised sequencing reads. These mutants left insufficient reads in the remaining genes to do the desired essentiality analysis. The appearance of these mutants was likely due to leakiness of the $P_{tet}$ promoter on the transposition plasmid. This would allow a low frequency early transposition event to occur prior to mutant selection. The resulting mutant would then be enriched for by amplification through the library selection procedure. To overcome this, several adjustments were made to the library generation protocol. The first precaution to be taken was reducing the time *C. difficile* harbouring the plasmid was cultured prior to induction. Secondly, previous transposition experiments used exponentially growing cells, adding more growing time and potentially selecting for a dominating mutant in the samples. Instead, we selected for mutants using overnight cultures, reducing the number of generations the bacteria have gone through, and reducing the likelihood that bacteria selected have gone through an early transposition event. The final precaution was increasing the number of transconjugants used in the experiment from 5 to 10. Each overnight culture represented an independent transconjugant and these were plated onto 15 plates, giving 150 plates total. Repeating this process allowed 1.5 million colonies to be harvested, each potentially representing a unique transposon mutant. Any stacks of plates which appeared to have a higher density than expected, suggesting that a premature

transposition event had occurred in that lineage,  were discarded. All of these refinements lead to the library generated for this thesis, containing 30-40,000 insertions with 20-30,000 insertions supported by multiple reads. One potential way of improving coverage for subsequent library generations would involve inducing transposition with a higher concentration of ATc. 6 copies of the transposition plasmid are assumed per cell. Consequently, per cell, the transposon can jump into the genome a maximum of 6 times. It is possible that usage of low concentrations of ATc, may not result in transposition occurring in all copies of the plasmid, reducing coverage, as observed in the sequencing stages. An optimised method for library generation in *C. difficile* is described in section 2.3, however, future insertion mutagenesis libraries should be constructed with higher concentrations of ATc to improve coverage. Methodology from this experiment has now been extended to another *Clostridia,* generating a library in *Clostridium saccarhoperbutylacetonicum*.

# Chapter IV. Conditional Gene Essentiality in an S-layer Null Mutant

## 4.1 Introduction

Generally, TraDIS is used to study the core genetic elements required for growth. Interrogating genomes in this way is pivotal to understanding key cellular processes, from identifying novel metabolic components to potential antimicrobial targets. The S-layer of *C. difficile* is essential, for reasons that remain unknown (Dembek *et al*., 2015). Despite this, and as outlined in Chapter III, we now have an S-layer null strain, FM2.5. This strain is avirulent and displays severe sporulation defects. Extending a TraDIS analysis to FM2.5, could allow the identification of alternative protein targets to disrupt S-layer formation. Random transposon mutagenesis of FM2.5, gave 30-40,000 unique transposon insertions, with 20-30,000 of those insertions supported by multiple reads. This was sufficient coverage to do an essential/non-essential gene comparison between FM2.5 and WT, R20291.

Here, we define a gene as essential if the gene encodes functions vital for bacterial growth under the conditions of the assay. Genes identified as conditionally essential in the absence of SlpA represent synthetically lethal gene pairs as this library was constructed in a mutant strain lacking the S-layer. On the other hand, conditionally non-essential genes in the absence of the S-layer potentially represent targets for disrupting S-layer assembly. We may speculate that in the absence of this surface structure, *slpA* may become dispensable for growth. We may also speculate that the protein responsible for S-layer secretion, SecA2, may become conditionally non-essential in this strain (Fagan and Fairweather, 2011). This may also extend to biosynthetic genes for PSII, the anionic polymer responsible for S-layer surface attachment (Willing *et al*., 2015).

Here we present successful application of TraDIS to identify a repertoire of genes either dispensable or vital for *C. difficile* growth in the absence of an S-layer. Additionally, we provide a functional analysis of all the candidate genes identified in this study.

## 4.2 Results

### 4.2.1 Sequencing of FM2.5 TraDIS Libraries

As outlined in chapter III, TraDIS libraries were sequenced in biological duplicates with technical duplicates, FM2.5 1a and 1b and FM2.5 2a and 2b. Samples were prepped as outlined in section 2.3. Libraries were loaded onto the MiSeq flow-cell and run on an Illumina MiSeq with 150 bp paired-end reads. This involves sequencing both ends of the transposon-containing gDNA fragments and aligning the forward and reverse reads as read pairs, enabling more accurate read alignment than single reads. Demultiplexing TraDIS samples was carried out by Dr. Roy Chaudhuri, The University of Sheffield as outlined in section 2.3.10 **(Table 4.1)**. The mean number of unique insertions sites for all FM2.5 samples, was 33,022 resulting in an average of 1 insertion every 126 bp, unlike the 54 bp observed for R20291 (Dembek *et al*., 2015).

**Table 4.1. Summary of TraDIS results from FM2.5 sequenced libraries.**

| Sample | Total reads | No. reads with transposon tag (%) | No. reads mapped to reference genome (%) | No. unique insertion sites |
|---|---|---|---|---|
| **FM2.5 1a** | 4,849,448 | 4,796,794 (98.9) | 4,338,542 (90.4) | 35,252 |
| **FM2.5 1b** | 4,766,458 | 4,706,317 (98.7) | 4,255,282 (90.4) | 28,220 |
| **FM2.5 2a** | 3,201,625 | 3,115,029 (97.3) | 2,871,088 (92.2) | 28,652 |
| **FM2.5 2b** | 4,562,453 | 4,470,657 (98) | 4,048,634 (90.6) | 39,965 |

### 4.2.2 Identification of the putative essential genome of FM2.5 by TraDIS

In an ideal library with a high density of insertions, this analysis should generate two peaks **(Fig 4.1)**. The first node contains all the genes with a low number of transposon inserts, the essential genes, while the second node contains all the genes that tolerated higher numbers of insertions above the threshold of essentiality, the non-essential genes. Highly complex libraries will contain multiple unique insertions in every gene, with a clear separation between the two peaks, allowing the differences in essentiality to be clearly ascertained, as observed for the R20291 TraDIS data **(Fig 4.1)** (Dembek *et al*., 2015).

Analysis of samples via Bio::TraDIS revealed 488 genes as essential in the R20291 data **(Table 4.2)**. For R20291, the highly complex library contained multiple unique insertions in every gene, with a clear separation between the two peaks, allowing the differences in gene essentiality to be clearly ascertained (Dembek *et al*., 2015). However, for each individual FM2.5 data set, overlapping peaks are observed, meaning that many non-essential genes, especially small genes, or genes that contain few potential insertions are incorrectly characterised as essential **(Fig 4.1).** From the TraDIS analysis, there are many insertion locations with only a small number of reads, and these locations are not consistent across all individual libraries. While each library has 30-40,000 unique insertion sites, only 7709 of these are consistently found across all FM2.5 samples.

**Table 4.2 Gene essentiality in an S-layer mutant.**

| Sample | Essential genes | Non- essential genes | Ambiguous genes | Conditionally essential genes | Conditionally non-essential genes |
|---|---|---|---|---|---|
| **FM2.5 1a** | 379 | 3154 | 37 | 55 | 20 |
| **FM2.5 1b** | 824 | 2386 | 360 | 55 | 20 |
| **FM2.5 2a** | 460 | 3028 | 82 | 55 | 20 |
| **FM2.5 2b** | 325 | 3176 | 69 | 55 | 20 |
| **R20291** | 488 | 2904 | 178 | - | - |

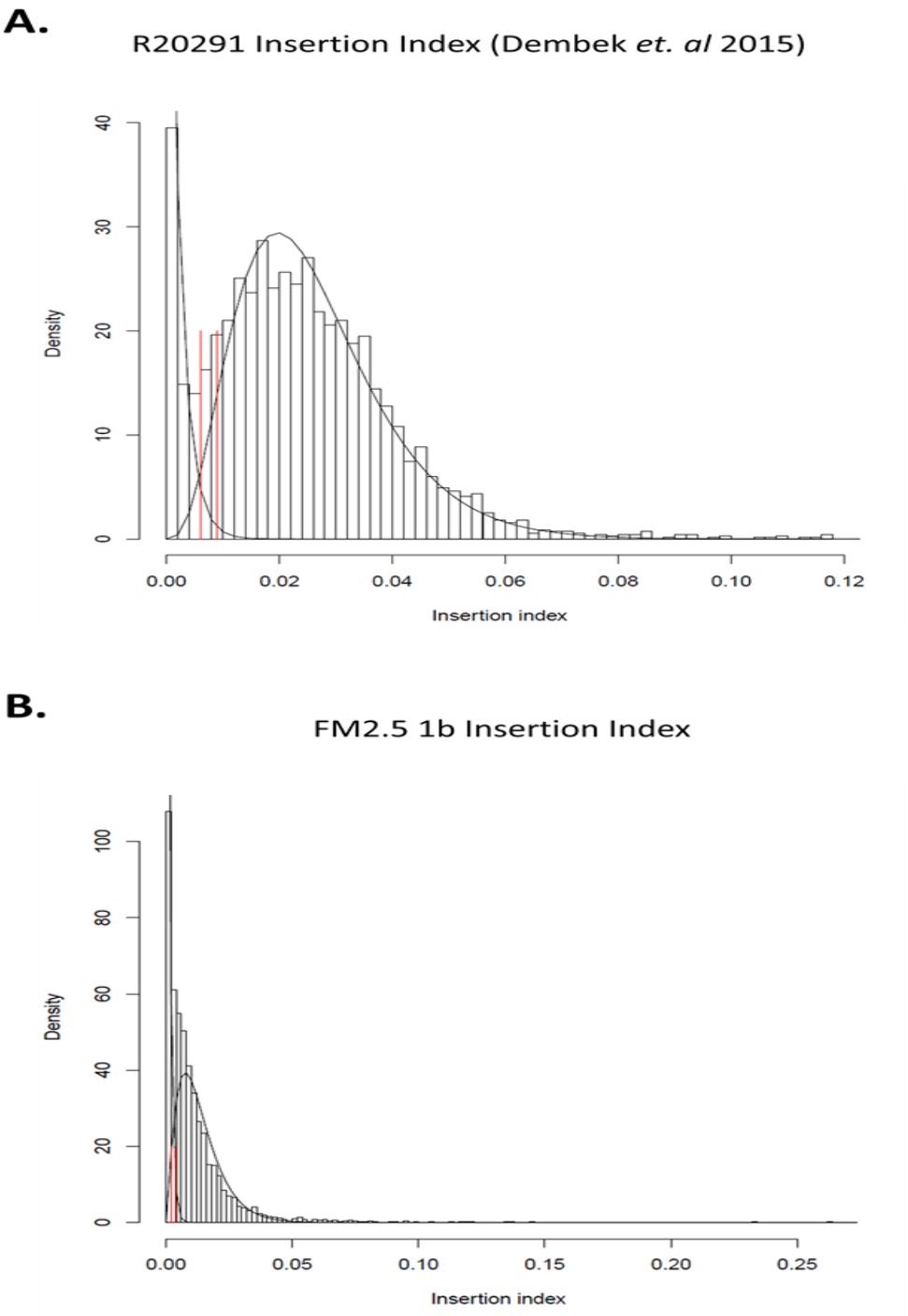



**Fig 4.1. Insertion Indices of *C. difficile* R20291 and *C. difficile* FM2.5. A.** Gene insertion index scores of WT *C. difficile* R20291 and **B.** FM2.5. The frequency distribution of insertion index scores for R20291 is bimodal. An exponential distribution model is fitted to the first mode, including all essential genes (red) and a gamma distribution model is fitted to the right, indicating all genes that tolerate higher number of transposon insertions, the non-essential genes. Dissimilarly, the insertion index for FM2.5 sample 1b has overlapping modes, making differences in essentiality harder to ascertain. All FM2.5 samples followed a similar distribution.

### 4.2.3 Identification of the conditional genome in an S-layer null mutant

Merging the data sets made it harder to specifically decipher the conditional genome of the S-layer mutant **(Table 4.2)**. As such, we will define conditional essentiality as genes which are essential in all four data sets of FM2.5 and non-essential in the published R20291 TraDIS analysis and vice versa for conditionally non-essential genes.

Genes identified as non-essential in all FM2.5 libraries were compared against the R20291 data to identify initial changes in essentiality. In a second step, any differences in essentiality were compared against the published Dembek *et. al* data to ascertain the genes dispensable for growth in the absence of an S-layer. By this method, 20 genes were identified as conditionally non-essential in an S-layer null mutant **(Table 4.3)**. Additionally, 55 genes were identified as vital for growth in the absence of this major surface structure **(Table 4.5)**.

**Table 4.3. Conditionally non-essential genes of an S-layer mutant**

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_0038 (*proS*)** | 13 | 12 | 15 | 17 | 0.007576 | 0.006993 | 0.008741 | 0.009907 | 1716 | proylyl-tRNA synthetase |
| **CDR20291_0481** | 10 | 8 | 4 | 8 | 0.013228 | 0.010582 | 0.005291 | 0.014898 | 756 | putative endonuclease |
| **CDR20291_0513** | 83 | 67 | 53 | 89 | 0.018954 | 0.0153 | 0.012103 | 0.020324 | 4379 | hypothetical protein |
| **CDR20291_0896** | 19 | 17 | 10 | 20 | 0.005479 | 0.004902 | 0.002884 | 0.005767 | 3468 | ATP-dependent nuclease subunit B |
| **CDR20291_1130** | 4 | 4 | 2 | 3 | 0.0199 | 0.0199 | 0.00995 | 0.014925 | 201 | putative small acid-soluble spore protein |
| **CDR20291_1213** | 2 | 2 | 1 | 1 | 0.006116 | 0.006116 | 0.003058 | 0.003058 | 327 | putative phage protein |

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_1930** | 4 | 8 | 8 | 8 | 0.010753 | 0.021505 | 0.021505 | 0.021505 | 372 | putative phage regulatory protein |
| **CDR20291_2329 (*era*)** | 5 | 5 | 6 | 5 | 0.005593 | 0.005593 | 0.006711 | 0.005593 | 894 | GTP-binding protein |
| **CDR20291_2355 *(grpE*)** | 2 | 4 | 3 | 2 | 0.003221 | 0.006441 | 0.004831 | 0.003221 | 621 | heat shock protein |
| **CDR20291_2463** | 7 | 5 | 1 | 4 | 0.009485 | 0.006775 | 0.001355 | 0.00542 | 738 | probable short chain dehydrogenase |
| **CDR20291_2603 (*gtaB*)** | 7 | 10 | 4 | 3 | 0.007407 | 0.010582 | 0.004233 | 0.003175 | 945 | UTP—glucose-1-phosphate uridylyltransferase |
| **CDR20291_2659 (*rkpK*)** | 11 | 9 | 3 | 10 | 0.008185 | 0.006696 | 0.002232 | 0.00744 | 1344 | putative UDP-glucose-6-dehydrogenase |
| **CDR20291_2660 (*tuaG*)** | 3 | 3 | 1 | 3 | 0.003984 | 0.003984 | 0.001328 | 0.003984 | 758 | putative teichuronic acid biosynthesis glycosyltransferase |

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_2681 (*secA2*)** | 38 | 32 | 56 | 86 | 0.016198 | 0.01364 | 0.02387 | 0.036658 | 2346 | preprotein translocase SecA subunit |
| **CDR20291_2694 (*ruvB*)** | 6 | 4 | 9 | 11 | 0.005882 | 0.003922 | 0.008824 | 0.010784 | 1020 | Holliday junction DNA helicase |
| **CDR20291_3187A** | 1 | 1 | 1 | 1 | 0.007092 | 0.007092 | 0.007092 | 0.007092 | 138 | autoinducer prepeptide |
| **CDR20291_3217 (*pfkA*)** | 19 | 10 | 16 | 27 | 0.019792 | 0.010417 | 0.016667 | 0.028125 | 960 | 6-phosphofructokinase |
| **CDR20291_3399 (*murI*)** | 2 | 4 | 8 | 3 | 0.002478 | 0.004957 | 0.009913 | 0.003717 | 807 | glutamate racemase |
| **CDR20291_3535 (*gidA*)** | 21 | 23 | 4 | 18 | 0.011076 | 0.012131 | 0.00211 | 0.009494 | 1896 | glucose inhibited division protein A |

### 4.2.4 Tools for computational analysis of the conditional genome of FM2.5

Conditionally non-essential genes represent good targets for disrupting S-layer assembly in the WT strain. As such, computational analysis was performed on the conditional genome to gain further insights into the function of individual genes. Additionally, several genes from the analysis could be mapped to KEGG metabolic pathways for *C. difficile*, enabling identification of dispensable metabolic reactions in the absence of SlpA (Kanehisa and Goto, 2000). These reactions, and perhaps metabolic intermediates, would be essential in the WT background, and perhaps good targets for the development of antimicrobials. Tools for computational analysis of the conditional genome of an S-layer mutant are outlined in section 2.4.10.

### 4.2.5 Conditionally non-essential genes in the absence of an S-layer.

From the FM2.5 TraDIS analysis 20 genes were recognised as conditionally non-essential **(Table 4.3)**. Bioinformatics analysis revealed that none of these genes encode transmembrane proteins **(Table 4.4)**. Moreover, it is extremely likely that all of these proteins are intracellular, as evidenced by low probabilities of the proteins having a signal peptide sequence. From these 20 genes, two genes were annotated to encode hypothetical proteins (*cdR20291_0513* and *cdR20291_1213*) **(Table 4.4)**, however, bioinformatics analysis revealed little information about the functions of these two proteins.

*cdR20291_2329* was identified as conditionally non-essential. This gene encodes the Era GTPase involved in 30S ribosome subunit biogenesis (Wood *et al*., 2019). Six genes were identified as involved in sugar metabolism: *cdR20291_0481*, *cdR20291_2603* (*gtaB*), *cdR20291_2659* (*rkpK*), *cdR20291_2660* (*tuaG*)*, cdR20291_3217* (*pfkA*) and *cdR20291_3535*. Of these six, two (*gtaB* and *rkpK*) encoded proteins found within the same metabolic pathway **(Fig 4.3)**. GtaB, a 314 amino acid UTP--glucose-1-phosphate uridylyltransferase is involved in the conversion of glucose-1P to UDP-glucose. In many bacterial species UDP-glucose serves as a substrate for the synthesis of diacylglycerol (DAG) in lipoteichoic acid synthesis (Grundling and Schneewind, 2007). Subsequently,

RkpK, a putative UDP-glucose 6-dehydrogenase catalyses the conversion of UDP-glucose to UDP-glucoronate. RkpK is 447 amino acids in length and shares ~72% sequence similarity with a UDP-glucose/GDP-mannose dehydrogenase from *Bacillus* sp. From the bioinformatics analysis, RkpK lacks transmembrane domains and is potentially Sec secreted (0.3744 probability). For the Signal5-P software, a probability close to 1 means a highly reliable prediction. For Archaea, Gram-positive and Gram-negative bacteria, the probability threshold is 0.25, as there are four possible classes of signal peptide (Sec/SPI, Tat/SPI, SPII and Other). A probability close to this threshold means a very unreliable prediction.

Additionally, several proteins were identified as part of the anionic polymer, PSII, biosynthetic locus: *cdR20291_2659* and *cdR20291_2660* (Willing *et al*., 2015) **(Fig 1.6)**. CDR20291_2660 (TuaG) is 250 amino acids, and a putative teichuronic acid glycosyltransferase (Soldo *et al*., 1999). TuaG of *C. difficile* shares 59.2% amino acid sequence identity with that of *C. perfringens.* This protein is intracellular with a 0.9922 probability it is not secreted. Notably, *cdR20291_2661* and *cdR20291_2665*, which are also PSII biosynthetic genes were identified as conditionally non-essential in 3 out of 4 data sets and of ambiguous essentiality in the 4$^{th}$. This confirmed the hypothesis that several PSII biosynthesis genes would become conditionally non-essential in the absence of SlpA.

Another major gene identified as conditionally non-essential was *cdR20291_2681*. This gene is annotated to encode the preprotein translocase SecA2 (Fagan and Fairweather, 2011). SecA2, is 781 amino acids in length and is comprised of a DEAD-like domain, a SecA Wing, a scaffold domain, and a SecA preprotein cross-linking domain. It is this protein that is responsible for S-layer precursor secretion. Like *Bacillus anthracis*, *C. difficile secA2* is contained within an S-layer genomic locus (Minkovsky *et al*., 2002). However, unlike in *B. anthracis*, *secA2* is essential for viability in *C. difficile* (Dembek *et al*., 2015). This is potentially due to SlpA being a SecA2-dependent substrate. However, in the absence of the S-layer, we have identified *secA2* as expendable for growth.

Two genes identified are involved in sporulation **(Table 4.4)**. The first is small acid soluble spore (SASP) protein, *cdR20291_1130*, one of four SASPs described for *C. difficile*: *cdR20291_2576* (SspA), *cdR20291_3107* (SspB) and *cdr20291_3080* (Nerber and Sorg, 2021). These SASPs are located within core of the spore, tightly associating with, and altering the properties of the DNA, contributing to spore resistance to heat and chemicals (Paredes-Sabja *et al*., 2014). While the major SspA and SspB are involved in UV resistance, no such role has been described for *cdR20291_1130* (Nerber and Sorg, 2021). The second spore protein identified as conditionally non-essential is *cdR20291_3086*, or *dapB1*, encoding a dihydrodipicolinate reductase **(Fig 4.4)**. In *Clostridium perfringens dapA* and *dapB* encode the precursor enyzmes for dipicolinic acid, which preserves the metabolically dormant spore state (Paredes-Sabja *et al*., 2014). Interestingly, *cdR20291_3355* (*spoIIIF*), encoding the stage V sporulation protein B was identified as conditionally non-essential in 3 out of 4 data sets, and of ambiguous essentiality in the 4$^{th}$ (not shown).

*cdR20291_3399,* or *murI,* encodes a glutamate racemase, MurI, responsible for conversion of L-glutamate to D-glutamate, which is then incorporated into the PG cell wall. *murI*, is essential in *E. coli* and *C. difficile* expressing SlpA (Dembek *et al*., 2015; Doublet *et al*., 1993). However, in the absence of this surface structure, *murI* is dispensable for growth. Interestingly, while two *murI* homologues have been identified in many bacterial species, no such *murI* homologue is present in the *C. difficile* genome (Duvall *et al*., 2017). MurI exhibits species-specific structural and enzymatic properties and as such may represent a good target for the development of narrow-spectrum antibiotics to treat CDI (Lundqvist *et al*., 2007).

Additional proteins include one involved in proline metabolism **(Fig 4.2)**, CDR20291_0038 (ProS), CDR20291_2355 (GrpE), encoding a heat-shock protein, postulated to be an S-layer chaperone (Peter Oatley, personal communication), a phage transcriptional regulator, CDR20291_1930 and the the Holliday junction branch migration DNA helicase, RuvB.

**Table 4.4 Bioinformatic analysis of conditionally non-essential genes**

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Pred | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Signal Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_0038** | prolyl-tRNA-synthase | proline tRNA-ligase | tRNA synthetase class II core domain (G, H, P S and T) (94-454), anticodon binding domain (470-562) | 0 | 0.0047 | 0.0006 | 0.0004 | 0.9943 |
| **CDR20291_0481** | putative endonuclease | sugar phosphate isomerase/epimerase | xylose isomerase-like TIM barrel (29-210) | 0 | 0.0115 | 0.0002 | 0.0022 | 0.9861 |
| **CDR20291_0513** | hypothetical protein | hypothetical protein | no significant Pfam matches | 0 | 0.1494 | 0.0022 | 0.0101 | 0.8384 |
| **CDR20291_0896** | ATP-dependent nuclease subunit B | helicase-exonuclease AddAB subunit AddB | exodeoxyribonuclease V, gamma subunit (299-439), PD-(D/E)XK nuclease superfamily (781-1125) | 0 | 0.0051 | 0.0003 | 0.0012 | 0.9933 |
| **CDR20291_1130** | putative small acid soluble spore protein | spore protein | no significant Pfam matches | 0 | 0.0074 | 0.001 | 0.0009 | 0.9907 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Pred | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Signal Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_1213** | hypothetical protein | DUF2577 family protein | protein of unknown function DUF2577 (8-104) | 0 | 0.0014 | 0.0001 | 0.0003 | 0.9982 |
| **CDR20291_1930** | putative phage regulatory protein | Helix-turn-helix transcriptional regulator | Helix-turn-helix domain (8-62) | 0 | 0.0039 | 0.0005 | 0.0005 | 0.9952 |
| **CDR20291_2329** | GTP-binding protein | GTPase Era | 50S ribosome-binding GTPase (6-121), KH domain (205-283) | 0 | 0.0081 | 0.0005 | 0.0014 | 0.99 |
| **CDR20291_2355** | heat shock protein | nucleotide exchange factor GrpE | GrpE (2-206) | 0 | 0.0067 | 0.0123 | 0.002 | 0.979 |
| **CDR20291_2463** | probable short chain dehydrogenase | 3-oxoacyl-ACP reductase FabG, SDR family NAD(P) dependent oxidoreductase | Enoyl-(Acyl carrier protein) reductase (10-244) | 0 | 0.1118 | 0.0034 | 0.0053 | 0.8795 |
| **CDR20291_2603** | UTP--glucose-1-phosphate uridylyltransferase | UTP--glucose-1-phosphate uridyltransferase | Nucleotidyl transferase (30-293) | 0 | 0.0792 | 0.002 | 0.0089 | 0.9099 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Pred | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Signal Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_2659** | putative UDP-glucose 6-dehydrogenase | UDP-glucose/GDP-mannose dehydrogenase family protein | UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain (17-201), UDP-glucose/GDP-mannose dehydrogenase family, central domain (215-308), UDP-glucose/GDP-mannose dehydrogenase family, UDP binding domain (330-432) | 0 | 0.377 | 0.0045 | 0.0699 | 0.5482 |
| **CDR20291_2660** | putative teichuronic acid biosynthesis glycosyltransferase | Glycosyltransferase family 2 protein | Glycosyltransferase family 2 (7-166) | 0 | 0.006 | 0.0001 | 0.0699 | 0.9922 |
| **CDR20291_2681** | preprotein translocase SecA subunit | Preprotein translocase subunit SecA | SecA DEAD-like domain (4-381), SecA Wing and scaffold domain (574-781), SecA preprotein cross-linking domain (229-337) | 0 | 0.0016 | 0.0002 | 0.0003 | 0.9979 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Pred | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Signal Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_2694** | Holliday junction DNA helicase | Holliday junction branch migration DNA helicase RuvB | Holliday junction DNA helicase RuvB P-loop domain (26-184), RuvB AAA lid domain (187-260), RuvB C-terminal winged helix domain (262-332) | 0 | 0.0034 | 0.0006 | 0.0005 | 0.9954 |
| **CDR20291_3086** | dihydrodipicolinate reductase | 4-hydroxy-tetrahyrdodipicolinase reductase | dihydrodipicolinate reductase, N-terminus (2-113), dihydrodipicolinate, C-terminus (116-248) | 0 | 0.0049 | 0.0002 | 0.0006 | 0.9944 |
| **CDR20291_3187A** | autoinducer prepeptide | cyclic lactone autoinducer prepeptide | no significant Pfam Matches | 0 | 0.7163 | 0.03 | 0.1704 | 0.0833 |
| **CDR20291_3217** | 6-phosphofructokinase | 6-phosphofructokinase | phosphofructokinase (3-275) | 0 | 0.0849 | 0.0009 | 0.0111 | 0.9031 |
| **CDR20291_3399** | glutamate racemase | glutamate racemase | Asp/Glu/hydantoin racemase (6-214) | 0 | 0.0046 | 0.0002 | 0.0007 | 0.9945 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Pred | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Signal Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_3535** | Glucose inhibited division protein A | tRNA uridine-5-carboxyaminomethyl synthesis enzyme MnmG, tRNA uridine-5-carboxymethylaminomethyl modification protein GidA | glucose inhibited division protein A 10-400, GidA associated domain (402-616) | 0 | 0.0513 | 0.0084 | 0.153 | 0.7873 |

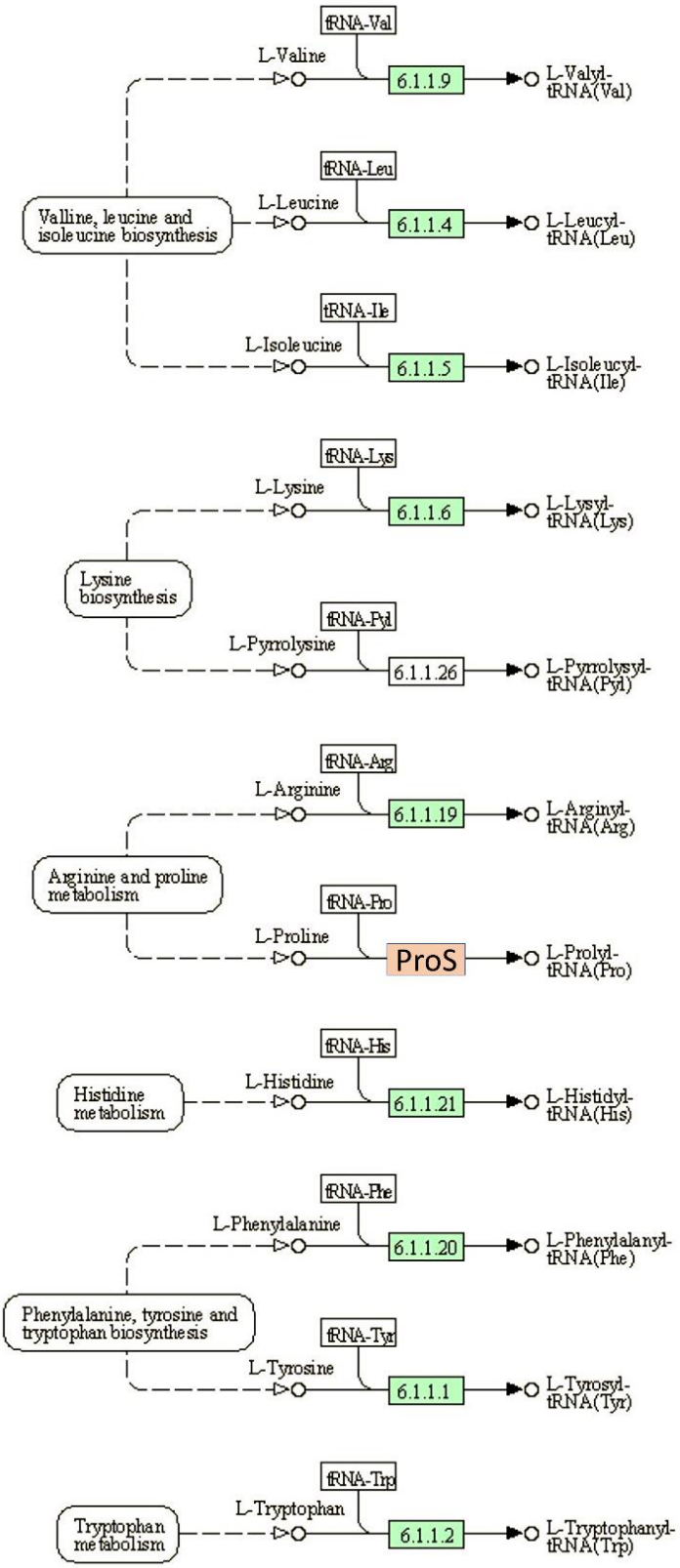



**Fig 4.2. *C. difficile* amino acyl-tRNA biosynthesis pathways.** Orange boxes indicate conditionally non-essential genes in the absence of an S-layer. Green boxes indicate organism-specific pathways.

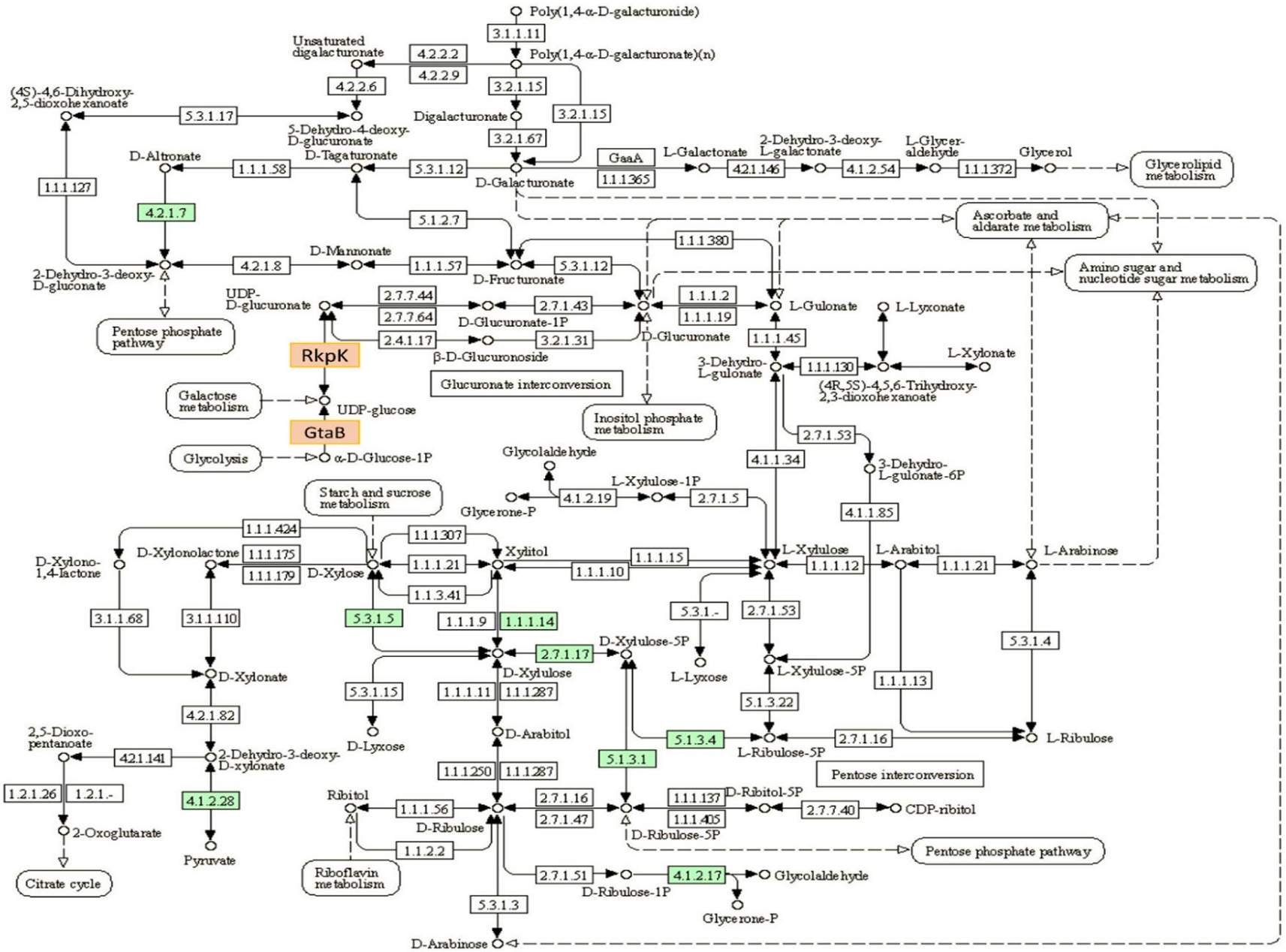



**Fig 4.3. Pentose and glucoronate interconversions in *C. difficile*.** Orange boxes indicate conditionally non-essential genes in an S-layer mutant. Green boxes indicate organism-specific pathways.

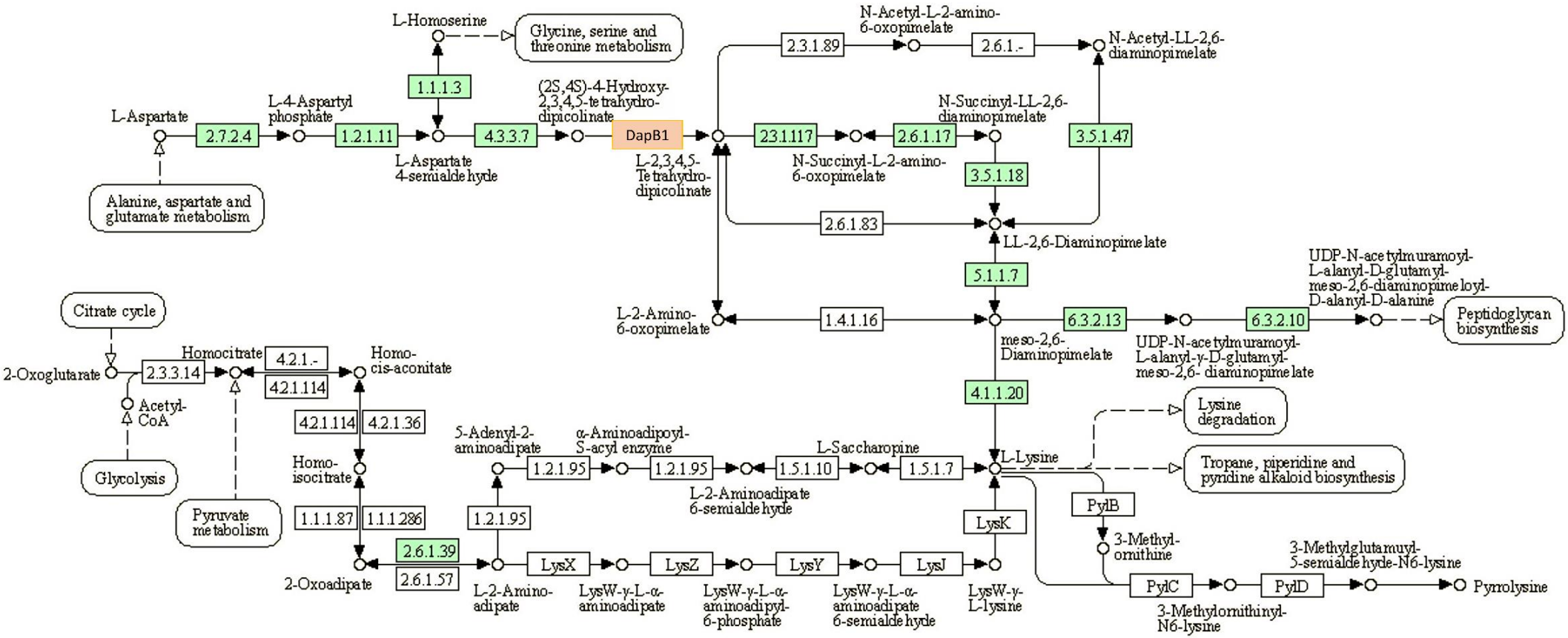



**Fig 4.4. *C. difficile* lysine biogenesis KEGG pathway.** Orange boxes indicate conditionally non-essential genes in an S-layer null background. Green boxes indicate organism-specific pathways.

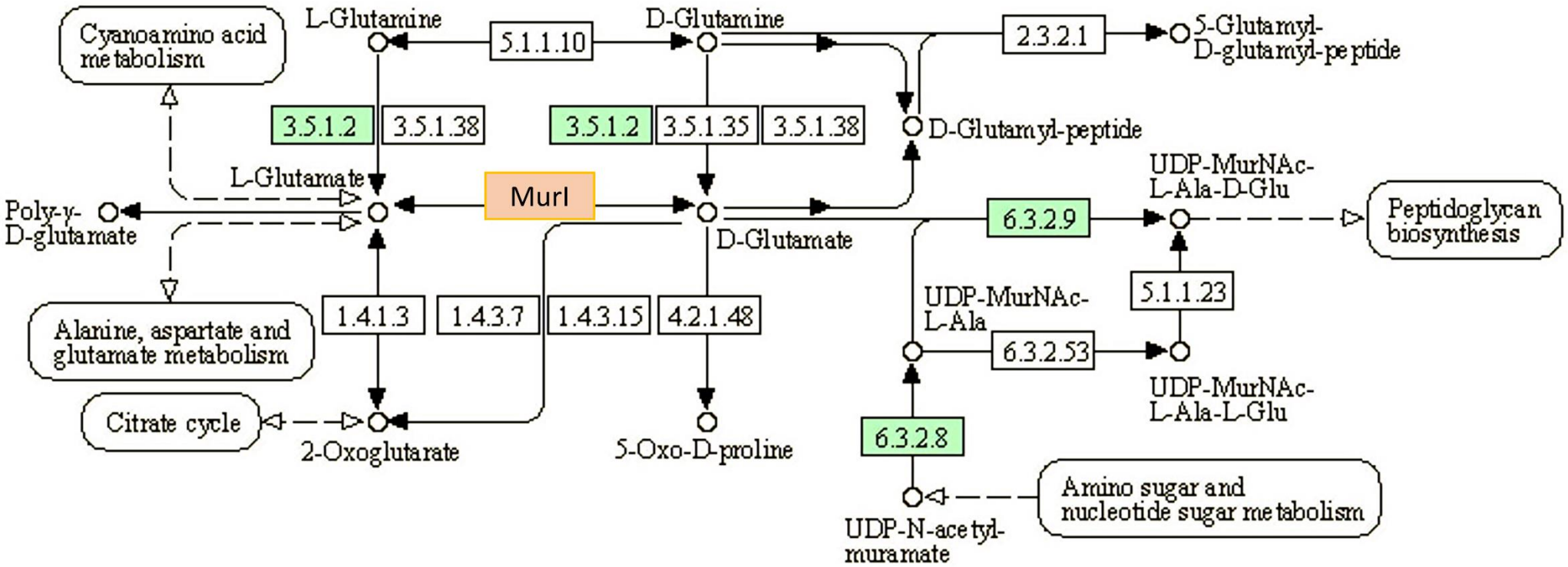



**Fig 4.5. D-glutamate and D-glutamine *C. difficile* metabolic KEGG pathway.** Orange boxes indicate conditionally non-essential genes in an S-layer null background. Green boxes indicate organism-specific pathways.

### 4.2.6 Genes vital for growth in the absence of SlpA

From the FM2.5 TraDIS analysis, 55 candidate genes were identified as conditionally essential for growth in the absence of the *C. difficile* S-layer **(Table 4.5)**. Conditionally essential genes in the absence of the S-layer are considered synthetically lethal gene pairs, as this insertion mutagenesis library was constructed in an S-layer mutant. For these gene pairs deletion of either gene does not have any impact on cell viability, but the combined deletion of both genes results in a non-viable mutant. This gene set did not tolerate a single transposon insertion, revealing that these genes encode functions vital for cell growth, with bacteria requiring a functioning copy of the gene. Of the 55 candidate genes, 21 were annotated to encode conserved hypothetical proteins. For the majority of these, Pfam and BLAST analyses revealed little about the function, however, CDR20291_2416 was identified as an antibiotic biosynthesis monooxygenase and CDR20291_2810 was identified as the CRISPR-associated endonuclease, Cas2. For the native *C. difficile* CRISPR-Cas system, Cas2 is required for integration of new spacers to fend off invading genetic material (Boudry *et al*., 2015). Additionally, while annotated as a conserved hypothetical protein, bioinformatic analysis revealed that *cdR20291_3099* encodes a proline reductase.

TMHMM analysis of the 55 conditionally essential genes, revealed that 15 genes encoded proteins with putative transmembrane helices: *cdR20291_0320, cdR20291_0537*, *cdR20291_0921, cdR20291_0922, cdR20291_0978, cdR20291_0979, cdR20291_1272, cdR20291_1283, cdR20291_1445, cdR20291_1816, cdR20291_1823, cdR20291_2251, cdR20291_ cdR20291_2617, cdR20291_2656*. Using Signal-P, 4 genes were likely to encode proteins destined for secretion: *cdR20291_0537*, *cdR20291_0978, cdR20291_1445, cdR20291_1661, cdR20291_2656*. Two iron uptake proteins were identified: CDR20291_0114, ferrodoxin, and CDR20291_0978 (RnfB), the B subunit of the RnfABCDGE electron transfer complex. Of the 55 conditionally essential genes, two encoded proteins of phosphotransferase systems: *cdR0291_0792* for lactose/cellobiose uptake and *cdR20291_2851* a phosphoenolpyruvate-dependent sugar phosphotransferase protein. Interestingly,

*cdR20291_2656* is vital for growth in the absence of SlpA. CDR20291_2656 is a cell wall hydrolase located within the PSII polymer locus. Conditionally essential genes and bioinformatic analysis of these genes are shown in tables 4.5 and 4.6 below.

**4.5 Conditionally essential genes in the absence of an S-layer**

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_0114** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 216 | ferredoxin |
| **CDR20291_0230** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 276 | negative regulator of flagellin synthesis |
| **CDR20291_0320** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 213 | hypothetical protein |
| **CDR20291_0537** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 177 | conserved hypothetical protein |
| **CDR20291_0544** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 141 | hypothetical protein |
| **CDR20291_0587** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 270 | hypothetical protein |
| **CDR20291_0695** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 366 | PTS system, glucitol/sorbitol-specific IIa component |

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_0792** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 330 | PTS system, IIa component |
| **CDR20291_0822** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 201 | cold shock protein |
| **CDR20291_0846** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 516 | putative thioesterase |
| **CDR20291_0921** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 180 | hypothetical protein |
| **CDR20291_0922** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 201 | hypothetical protein |
| **CDR20291_0978** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 972 | electron transport complex protein |
| **CDR20291_0979** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 168 | hypothetical protein |
| **CDR20291_1104** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 249 | hypothetical protein |
| **CDR20291_1111** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 345 | conserved hypothetical protein |

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_1132** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 423 | conserved hypothetical protein |
| **CDR20291_1272** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 675 | putative membrane protein |
| **CDR20291_1275** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 135 | hypothetical protein |
| **CDR20291_1283** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 291 | putative membrane protein |
| **CDR20291_1308** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 468 | putative 5-nitroimidazole reductase |
| **CDR20291_1327** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 222 | putative ferrous iron transport protein A |
| **CDR20291_1347** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 477 | hypothetical protein |
| **CDR20291_1440** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 438 | phage protein |
| **CDR20291_1445** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 162 | hypothetical phage protein |

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_1467** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 639 | MarR-family transcriptional regulator |
| **CDR20291_1512** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 795 | putative sodium extrusion protein ABC transporter, ATP binding protein |
| **CDR20291_1661** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 465 | putative lipoprotein |
| **CDR20291_1707** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 399 | putative two-component system response regulator |
| **CDR20291_1740** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 309 | pseudogene |
| **CDR20291_1746** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 309 | putative uncharacterised protein |
| **CDR20291_1816** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 369 | putative lipoprotein |
| **CDR20291_1822** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 171 | pseudogene |

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_1823** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 207 | putative lipoprotein signal peptidase |
| **CDR20291_1858** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 261 | stage V sporulation protein S |
| **CDR20291_1953** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 216 | putative transcriptional regulator |
| **CDR20291_2043** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 204 | hypothetical protein |
| **CDR20291_2095** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 681 | beta-phosphoglucomutase |
| **CDR20291_2190** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 378 | putative regulatory protein |
| **CDR20291_2192** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 219 | pseudogene |
| **CDR20291_2251** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 225 | putative membrane protein |
| **CDR20291_2405** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 381 | putative translation inhibitor endoribonuclease |

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_2416** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 312 | conserved hypothetical protein |
| **CDR20291_2617** | 0 | 1 | 0 | 0 | 0 | 0.00211 | 0 | 0 | 474 | putative membrane protein |
| **CDR20291_2644** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 261 | PTS system, phosphocarrier protein |
| **CDR20291_2656** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 708 | putative cell wall hydrolase |
| **CDR20291_2810** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 279 | conserved hypothetical protein |
| **CDR20291_2851** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 462 | PTS system, IIa component |
| **CDR20291_3041** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 540 | conserved hypothetical protein |
| **CDR20291_3099** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 468 | conserved hypothetical protein |
| **CDR20291_3134** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 219 | putative ferrous iron transport protein A |

| Locus_tag | Insertion Count 1a | Insertion Count 1b | Insertion Count 2a | Insertion Count 2b | Insertion Index 1a | Insertion Index 1b | Insertion Index 2a | Insertion Index 2b | Gene Length (bp) | Annotated Function |
|---|---|---|---|---|---|---|---|---|---|---|
| **CDR20291_3201** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 585 | putative protease |
| **CDR20291_3373** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 936 | putative phosphonate metabolism protein |
| **CDR20291_3461** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 645 | chloramphenicol o-acetyltransferase |

**Table 4.6 Bioinformatics of conditionally essential genes of FM2.5**

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Pred | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Signal Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_0114** | ferredoxin | 4Fe-4S binding protein | 4Fe-4S dicluster domain (12-65) | 0 | 0.0384 | 0.0006 | 0.0026 | 0.9585 |
| **CDR20291_0230** | negative regulator of flagellin synthesis anti sigma D factor | flagellar biosynthesis anti-sigma factor FlgM | anti-sigma 28 factor, FlgM (35-87) | 0 | 0.0061 | 0.0003 | 0.0006 | 0.9931 |
| **CDR20291_0320** | conserved hypothetical protein | hypothetical protein | no significant Pfam matches | 1 | 0.0014 | 0.0002 | 0.001 | 0.9974 |
| **CDR20291_0537** | conserved hypothetical protein | hypothetical protein | no significant Pfam matches | 2 | 0.0121 | 0.0004 | 0.2966 | 0.6908 |
| **CDR20291_0544** | conserved hypothetical protein | hypothetical protein | no significant Pfam matches | 0 | 0.1507 | 0.0035 | 0.0056 | 0.8403 |
| **CDR20291_0587** | hypothetical protein | DUF3795 domain-containing protein | Protein of unknown function DUF3795 (5-61) | 0 | 0.1506 | 0.0037 | 0.0235 | 0.8222 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Prediction | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_0695** | PTS system, glucitol/sorbitol-specific IIA component | PTS glucitol/sorbitol transporter subunit IIA | PTS system glucitol/sorbitol-specific IIA component (4-115) | 0 | 0.0059 | 0.0001 | 0.0008 | 0.9932 |
| **CDR20291_0780** | conserved hypothetical protein | NifU-like domain protein | NifU-like domain (5-71) | 0 | 0.0567 | 0.0125 | 0.0037 | 0.9271 |
| **CDR20291_0792** | PTS system, IIA component | PTS lactose/cellobiose transporter subunit IIA | PTS system lactose/cellobiose specific IIA subunit (8-100) | 0 | 0.006 | 0.0011 | 0.0011 | 0.9918 |
| **CDR20291_0822** | cold shock protein | cold shock protein | cold shock DNA-binding domain (3-65) | 0 | 0.0101 | 0.0006 | 0.002 | 0.9872 |
| **CDR20291_0846** | putative thioesterase | PaaI family thioesterase | thioesterase superfamily, (57-128) | 0 | 0.0088 | 0.0003 | 0.0011 | 0.9897 |
| **CDR20291_0921** | hypothetical protein | hypothetical protein | no significant Pfam matches | 2 | 0.0338 | 0.0028 | 0.0327 | 0.9306 |
| **CDR20291_0922** | hypothetical protein | hypothetical protein | no significant Pfam matches | 2 | 0.0361 | 0.001 | 0.0384 | 0.9245 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Prediction | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_0978** | electron transfer complex protein | RnfABCDGE subunit B | putative FeS cluster (43-75), 4Fe-4S binding domain (163-184), 4Fe-4S dicluster domain (214-257), 4Fe-4S dicluster domain (269-321) | 1 | 0.4143 | 0.0082 | 0.0127 | 0.5648 |
| **CDR20291_0979** | hypothetical protein | hypothetical protein | no significant Pfam matches | 1 | 0.014 | 0.0003 | 0.0171 | 0.9687 |
| **CDR20291_1104** | conserved hypothetical protein | hypothetical protein | no significant Pfam matches | 0 | 0.0024 | 0.0003 | 0.0003 | 0.997 |
| **CDR20291_1111** | conserved hypothetical protein | YraN family protein | uncharacterised protein family UPF0102 (10-100) | 0 | 0.0066 | 0.0003 | 0.0009 | 0.9923 |
| **CDR20291_1132** | conserved hypothetical protein | DUF2000 domain-containing protein | protein of unknown function DUF2000 (7-140) | 0 | 0.0257 | 0.0023 | 0.0039 | 0.968 |
| **CDR20291_1272** | putative membrane protein | MgtC/SapB family protein | MgtC family (16-140) | 4 | 0.0124 | 0.0006 | 0.0136 | 0.9733 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Prediction | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_1275** | hypothetical protein | hypothetical protein | no significant Pfam matches | 0 | 0.0825 | 0.0072 | 0.0115 | 0.8988 |
| **CDR20291_1283** | putative membrane protein | hypothetical protein | no significant Pfam matches | 2 | 0.0208 | 0.0003 | 0.0222 | 0.9566 |
| **CDR20291_1308** | putative 5-nitroimidazole reductase | pyridoxamine 5'-phosphate oxidase family protein | pyridoxamine 5'-phosphate oxidase (14-152) | 0 | 0.0583 | 0.0012 | 0.0019 | 0.9386 |
| **CDR20291_1327** | putative ferrous iron transport protein A | ferrous iron transport protein A | FeoA domain (3-73) | 0 | 0.0051 | 0.0013 | 0.0011 | 0.9926 |
| **CDR20291_1347** | hypothetical protein | type I restriction enzyme HsdR N-terminal domain-containing protein | type I restriction enzyme R protein N-terminus (4-108) | 0 | 0.0022 | 0.0001 | 0.0001 | 0.9976 |
| **CDR20291_1440** | phage protein | hypothetical protein | no significant Pfam matches | 0 | 0.0093 | 0.0004 | 0.0017 | 0.9886 |
| **CDR20291_1445** | hypothetical phage protein | hypothetical protein | no significant Pfam matches | 1 | 0.2811 | 0.0079 | 0.1406 | 0.5704 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Prediction | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_1467** | MarR-like transcriptional regulator | MarR transcriptional regulator | MarR family (62-121) | 0 | 0.0017 | 0.0001 | 0.0002 | 0.998 |
| **CDR20291_1512** | putative sodium extrusion ABC transporter, ATP binding protein | ABC transporter, ATP-binding protein | ABC transporter (40-183) | 0 | 0.0435 | 0.0004 | 0.0006 | 0.955 |
| **CDR20291_1661** | putative lipoprotein | DUF4624 domain-containing lipoprotein | Domain of unknown function DUF4624 (22-153) | 0 | 0.0008 | 0.0001 | 0.9987 | 0.0003 |
| **CDR20291_1681** | putative hypothetical protein | BtrH N-terminal domain containing protein | butirosin biosynthesis protein H, N-terminal (25-160) | 0 | 0.0433 | 0.0005 | 0.0047 | 0.9516 |
| **CDR20291_1707** | putative two-component system response regulator | putative two component system response regulator | LytTr DNA-binding domain (33-123) | 0 | 0.01 | 0.0002 | 0.0011 | 0.9887 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Prediction | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_1746** | putative uncharacterised protein | hypothetical protein | no significant Pfam matches | 0 | 0.0078 | 0.0001 | 0.0005 | 0.9916 |
| **CDR20291_1816** | putative lipoprotein | putative lipoprotein | no significant Pfam matches | 1 | 0.0014 | 0.0003 | 0.1109 | 0.8873 |
| **CDR20291_1822** | putative transposase (fragment) | transposase | no significant Pfam matches | 0 | 0.0079 | 0.0038 | 0.0012 | 0.9871 |
| **CDR20291_1823** | putative lipoprotein signal peptidase | signal peptidase | no significant Pfam matches | 1 | 0.0231 | 0.0003 | 0.0019 | 0.9747 |
| **CDR20291_1858** | Stage V sporulation protein S | Stage V sporulation protein S | Stage V sporulation protein S SpoVS (2-84) | 0 | 0.0372 | 0.0386 | 0.0016 | 0.9225 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Prediction | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_1953** | putative transcriptional regulator | Helix-turn-helix transcriptional regulator | Cro/C1-type HTH DNA-binding domain (6-68) | 0 | 0.0057 | 0.0002 | 0.0003 | 0.9938 |
| **CDR20291_2043** | Hypothetical protein | Hypothetical protein | no significant Pfam matches | 0 | 0.0101 | 0.0001 | 0.001 | 0.9888 |
| **CDR20291_2095** | beta-phosphoglucomutase | beta-phosphoglucomutase | haloacid dehalogenase-like hydrolase (6-189) | 0 | 0.0014 | 0.0001 | 0.0002 | 0.9982 |
| **CDR20291_2190** | putative regulatory protein | BlaI/MecI/CopY family transcriptional regulator | penicillinase repressor (13-118) | 0 | 0.0019 | 0.0002 | 0.0007 | 0.9972 |
| **CDR20291_2192** | transposase (fragment) | IS3 family transposase | no significant Pfam matches | 0 | 0.0136 | 0.0002 | 0.0023 | 0.9839 |
| **CDR20291_2251** | putative membrane protein | hypothetical protein | no significant Pfam matches | 1 | 0.0947 | 0.0006 | 0.004 | 0.9008 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Prediction | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_2405** | putative translation inhibitor endoribonuclease | RidA family protein | endoribonuclease L-PSP (9-125) | 0 | 0.0412 | 0.0085 | 0.0047 | 0.9456 |
| **CDR20291_2416** | conserved hypothetical protein | antibiotic biosynthesis monooxygenase | antibiotic biosynthesis monooxygenase (4-69) | 0 | 0.0023 | 0.0001 | 0.0002 | 0.9974 |
| **CDR20291_2617** | putative membrane protein | threonine/serine exporter family protein | threonine/Serine exporter, ThrE (10-136) | 4 | 0.0187 | 0.0029 | 0.0434 | 0.935 |
| **CDR20291_2644** | PTS system, phosphocarrier protein | HPr family phosphocarrier protein, | PTS HPr component phosphorylation site (2-81) | 0 | 0.0062 | 0.0014 | 0.0007 | 0.9917 |
| **CDR20291_2656** | putative cell wall hydrolase | Peptidoglycan endopeptidase, C40 family peptidase | NlpC/P60 family (126-232), bacterial SH3 domain (51-104) | 1 | 0.9893 | 0.0055 | 0.0033 | 0.0019 |
| **CDR20291_2810** | conserved hypothetical protein | CRISPR-associated endonuclease Cas2 | CRISPR associated protein Cas2 (5-72) | 0 | 0.0181 | 0.0014 | 0.0032 | 0.9774 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Prediction | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_2851** | PTS system, IIA component | PTS sugar transporter subunit IIA | PEP-dependent sugar phosphotransferase system EIIA (6-149) | 0 | 0.0135 | 0.0003 | 0.0006 | 0.9857 |
| **CDR20291_3041** | conserved hypothetical protein | 2'-5' RNA ligase family protein | 2'-5' RNA ligase superfamily (30-141) | 0 | 0.0319 | 0.0005 | 0.0055 | 0.9621 |
| **CDR20291_3099** | conserved hypothetical protein | proline reductase | glycine/sarcosine/betaine reductase component B subunit (5-125) | 0 | 0.0029 | 0.0002 | 0.0003 | 0.9965 |
| **CDR20291_3134** | putative ferrous iron transport protein A | ferrous iron tranport protein A | FeoA domain (1-70) | 0 | 0.0097 | 0.0016 | 0.0097 | 0.979 |
| **CDR20291_3201** | putative protease | DJ-1/PfpI family protein | DJ-1/PfpI family (1-175) | 0 | 0.0279 | 0.0005 | 0.0069 | 0.9647 |
| **CDR20291_3373** | putative phosphonate metabolism protein | phosphonate C-P lyase system protein, PhnG | phosphonate metabolism protein, PhnG (7-142) | 0 | 0.023 | 0.002 | 0.004 | 0.9764 |

| Locus_tag | Annotated Function | BLAST | Pfam Prediction | TM Prediction | Sec Signal Peptide Probability | Tat Signal Peptide Probability | Lipobox Peptide Probability | Other |
|---|---|---|---|---|---|---|---|---|
| **CDR20291_3461** | chloramphenicol o-acetyltransferase | chloramphenicol o-acetyltransferase | chloramphenicol acetyltransferase (7-208) | 0 | 0.0011 | 0.0001 | 0.0002 | 0.9986 |

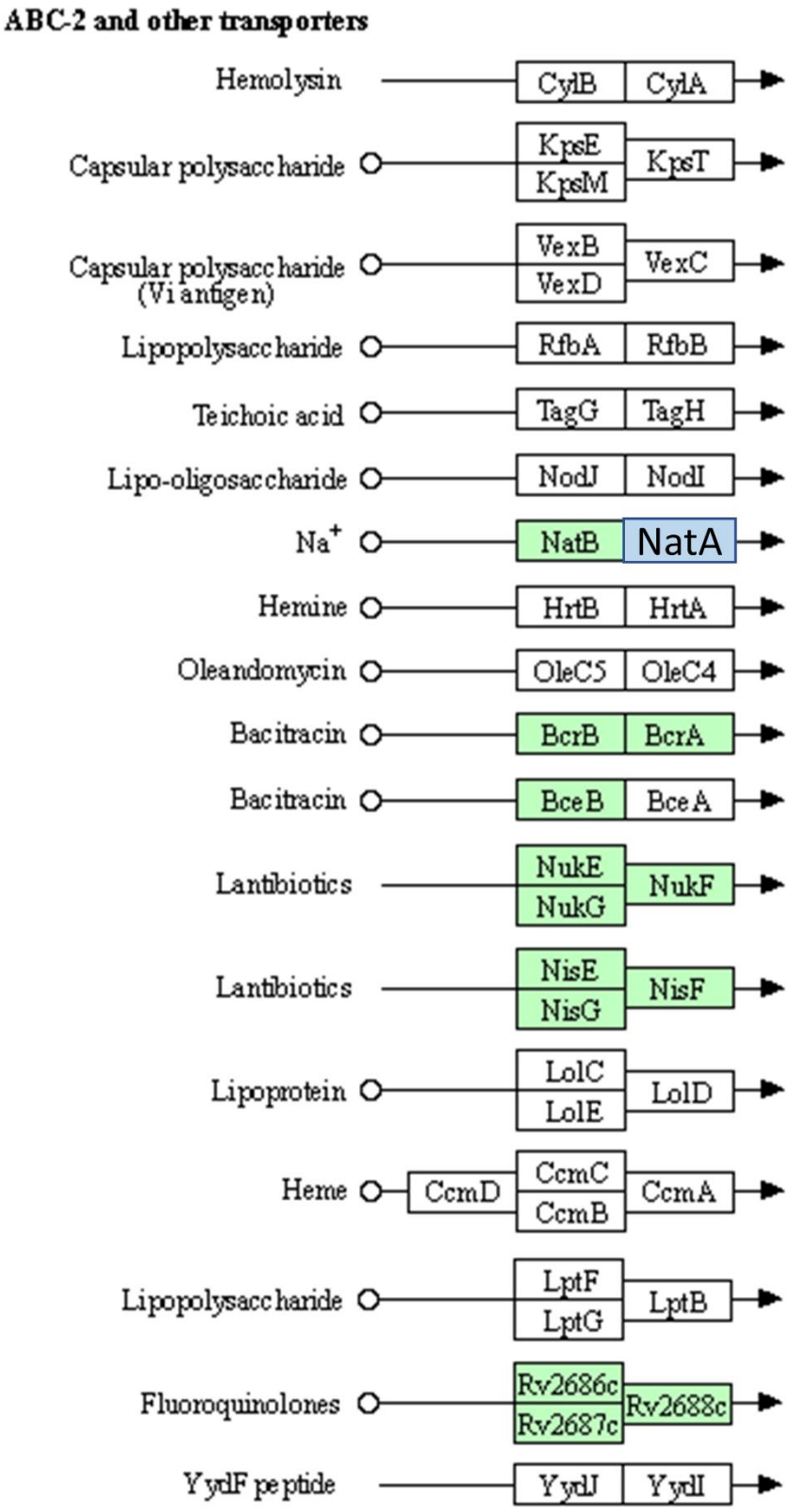



**Fig 4.6. *C. difficile* ABC Transporter KEGG Map.** Blue boxes indicate conditionally essential genes in the absence of an S-layer. Green boxes indicate organism-specific pathways.

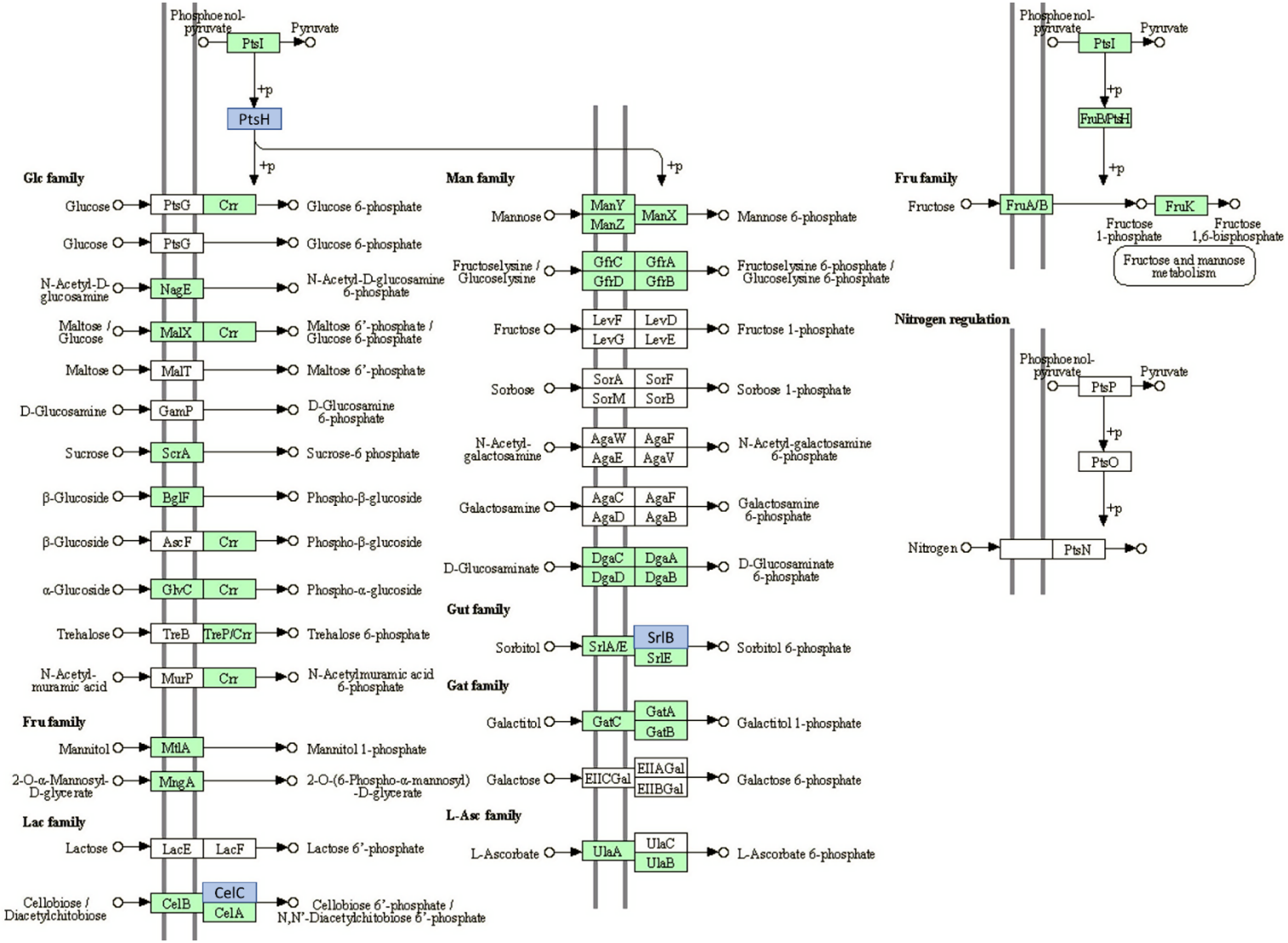



**Fig 4.7. KEGG map of *C. difficile* R20291 phosphotransferase systems.** Blue boxes indicate conditionally essential genes in the absence of SlpA. Green boxes indicate organism-specific pathways.

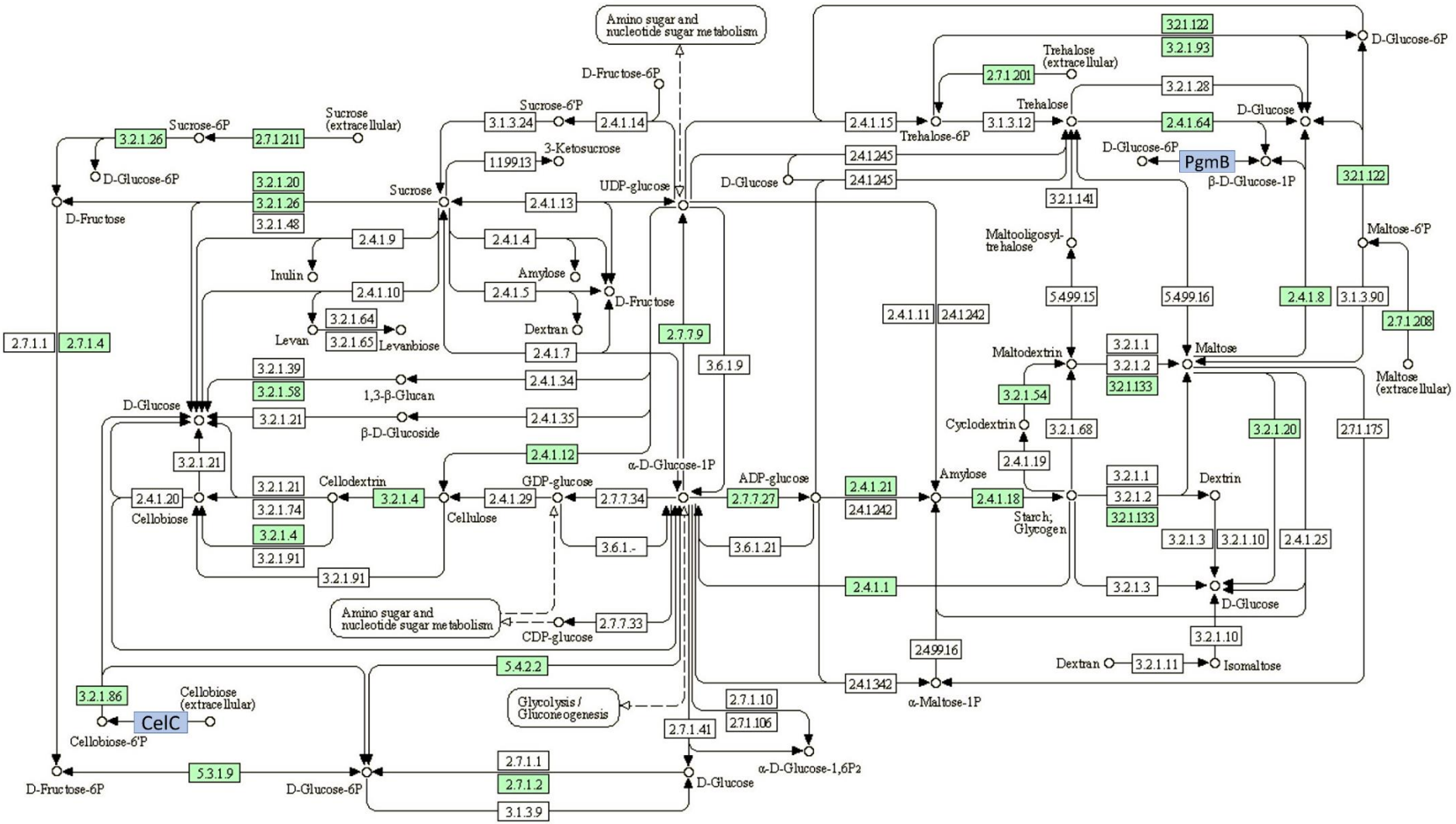



**Fig 4.8. *C. difficile* starch and sucrose metabolic KEGG pathways.** Blue boxes indicate conditionally essential genes in an S-layer null strain. Green boxes indicate organism-specific pathways.

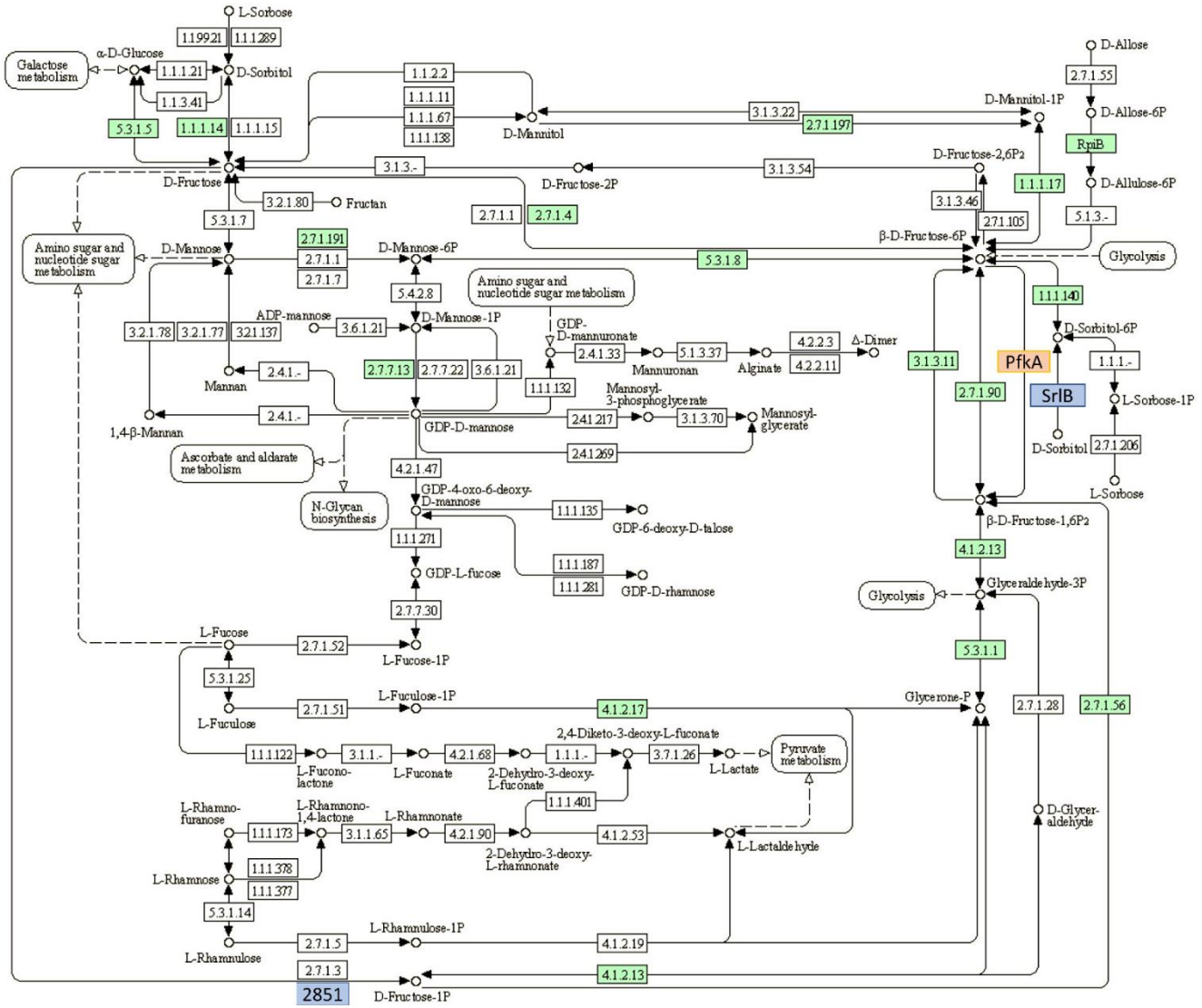



**Fig 4.9. *C. difficile* R20291 fructose and mannose metabolism KEGG pathways.** Blue boxes indicate conditionally essential genes in the absence of an S-layer. Orange boxes indicate conditionally non-essential genes in the absence of an S-layer. Green boxes indicate organism-specific pathways.

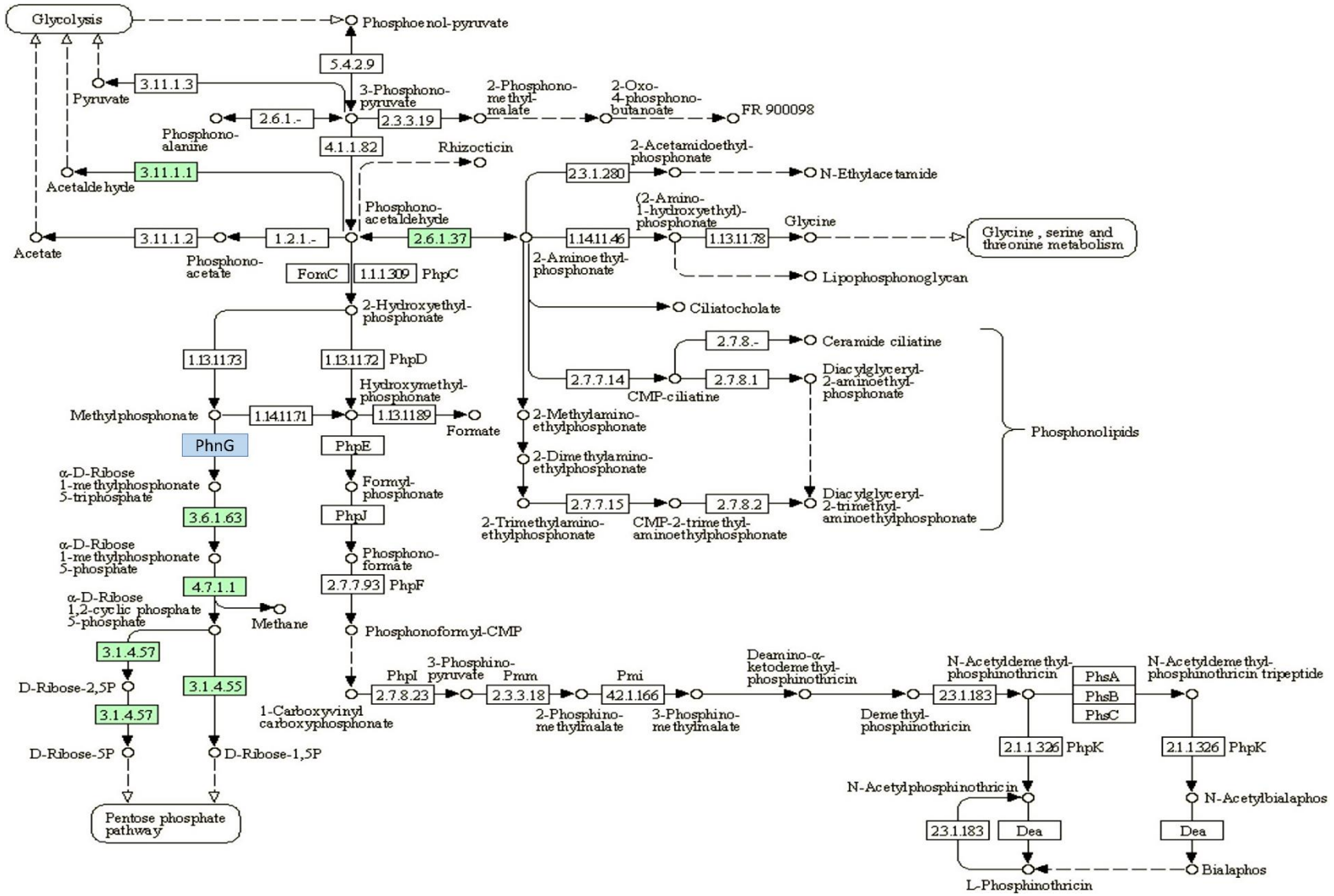



**Fig 4.10. *C. difficile* R20291 phosphonate and phosphinate metabolism KEGG pathway map.** Blue boxes indicate conditionally essential genes identified from the TraDIS analysis. Green boxes indicate organism-specific pathways.

## 4.3 Discussion

The main objective of this chapter was to identify the conditional genome of *C. difficile* lacking the major S-layer protein, SlpA. Libraries constructed in Chapter III were sequenced in biological duplicates with technical duplicates. Insertion indices for all genes were fitted to a bimodal distribution, with the first node containing all the genes with a low number of transposon inserts, the essential genes, and the second node with all the genes that tolerated higher numbers of insertions above the threshold of essentiality. These genes were deemed non-essential. Saturation of our sequenced libraries is not as dense as the published TraDIS dataset for WT *C. difficile* (one insert every 126 bp for the S-layer mutant compared to one insert every 54 bp for the WT). However, by inspecting essentiality of every gene from four sequenced FM2.5 libraries and comparing with the core essential genome of R20291, we have still been able to identify several interesting targets that have become indispensable or vital for growth when an S-layer is lacking **(Table 4.3 and 4.5)**. The next chapter will discuss the development of a CRISPR interference tool for the phenotypic analysis of several targets identified in this study.

20 genes were identified as conditionally non-essential in an S-layer null background. Of these, several metabolic genes were identified, which could be mapped to KEGG pathways, including *cdR20291_0038, cdR20291_2603, cdR20291_2659* and *cdR20291_3535*. We may speculate that in the absence of the S-layer, an energetically costly structure to produce, that metabolic reshufflings occur within cells and several metabolic reactions could become dispensable for growth, for example the conversion of α-D-glucose-1P to UDP-glucoronate through the successive actions of GtaB and RkpK.

As mentioned before, the first mode of transposon insertion indexes contains all genes that did not tolerate transposon inserts. While the most common reason for low numbers of transposon insertions is because the gene product is required for normal growth rates. Unfortunately,

overreliance on the statistical analysis often overestimates the number of essential genes. Often genes won't tolerate insertions not due to the genes being essential for growth but because the transposon is not able to access the DNA due to extreme DNA structure, exclusion by DNA-binding proteins or due to polar effects. With that, we identified 55 targets that are vital for growth in the absence of an S-layer. Interestingly, several genes encoding bacteriophage proteins were identified as conditionally essential. While we do not think that the products of these genes are contributing to bacterial fitness, several other groups have identified phage genes that resisted transposon insertion (Krupovic *et al*., 2006; Willcocks *et al*., 2018). Although, these results may be an artefact as these phage genes also tend to have very short DNA sequences.

As mentioned previously, conditionally non-essential genes represent good targets for disrupting S-layer assembly in the WT strain. To gain further insight into gene function of the conditional genome of *C. difficile*, we performed a computational analysis of all targets identified in the study using several computer programmes: BLAST, the Pfam database, TM pred and SignalP. From this, we identified none of the conditionally non-essential genes encoded transmembrane proteins, nor were any of these proteins secreted **(Table 4.4)**. We speculated that the translocon responsible for S-layer secretion would become conditionally non-essential in the absence of SlpA and that this concept would extend to PSII biosynthesis. Here we present *secA2* (*cdR20291_2681*) and two PSII biosynthetic genes (*cdR20291_2659 and cdR20291_2660*) as dispensable for growth in our sequenced libraries representing potential targets for blocking S-layer secretion and cell surface attachment. Additionally, two more PSII biosynthetic genes (*cdR20291_2661 and cdR20291_2665)* were identified as non-essential in 3 out of 4 sequenced libraries and of ambiguous essentiality in the 4$^{th}$. A higher library density may lead us to conclude if these two genes are conditionally non-essential. Surprisingly, we did not identify *slpA* as conditionally non-essential in FM2.5. There are two potential reasons for this. First, *slpA* may never have been essential and due to high levels of *slpA* expression, this gene does not readily tolerate transposon insertions. Alternatively, *slpA* is an essential gene, however the reason for this remains unclear. However, *cdR20291_2355*, whose gene

product GrpE is an S-layer chaperone, was identified as non-essential for growth *in vitro* (Dr. Peter Oatley, personal communication).

Additionally, beyond the expected changes in essentiality observed with *secA2* and several PSII biosynthesis genes, were changes to essentiality in genes involved in sporulation i.e, *cdR20291_1130*, encoding an SASP located at the core of the spore and *cdR20291_3086,* involved in DPA synthesis. An S-layer mutant displays severe defects in sporulation (Kirk *et al*., 2017a). As such it would be interesting to disrupt these genes and investigate the downstream effects this has on S-layer assembly and the ability to form the transmissible spore particles.

The GTPase Era was identified as dispensable for growth in the absence of SlpA. Although Era is not essential in all bacterial species, depletion of Era protein levels has several phenotypic effects such as defects in cell cycle control, ribosome biogenesis and chromosome segregation (Wood *et al*., 2019). Additionally, for *B. subtilis*, Era-depleted cells are defective in spore formation (Minkovsky *et al*., 2002). *C. difficile* FM2.5 exhibits severe defects in sporulation, potentially uncovering the reasoning why *era* is conditionally non-essential in this strain.

Several hypothetical proteins were identified as either conditionally essential or conditionally non-essential. One such protein was identified as the CRISPR-associated endonuclease Cas2, responsible for spacer integration during infection with foreign nucleic acids (Boudry *et al*., 2015). Moreover, several transcriptional regulators were identified as conditionally essential in the absence of an S-layer. An interesting line to follow would be to see the phenotypic and global effects of depletion of these regulators using the CRISPRi tool developed as part of this thesis.

# Chapter V. CRISPR Interference: a genetic tool for conditional gene repression in *C. difficile*

## 5.1 Introduction

Genome editing changes DNA sequences to a desired sequence in the context of an organism's native genome and can be achieved through insertion, deletion or substitution of sequences. However, tools for manipulation of the *C. difficile* genome are limited. The first to be described, ClosTron, involves inactivation of genes through an insertion of a group II intron. Unfortunately, polar effects can be caused at the site of insertion, and clean knockouts cannot be made using this technique (Heap *et al*., 2007). This system was superseded by a homologous recombination system using *codA* negative selection, which allows deletion, insertion or substitution of sequences (Cartman *et al*., 2012). However, these systems are incompatible with studying the functions of essential genes as loss of function is lethal to the cell. For *C. difficile*, ~ 10% of the genome is essential for bacterial survival and being able to study the functions of these genes will be pivotal to understanding key cellular processes.

Clustered Regularly Interspaced Short Palindromic Repeats Interference (CRISPRi) represents a novel technique with the ability to overcome the existing limitations of studying essential genes (Todor *et al*., 2021). CRISPRi is a powerful tool derived from CRISPR which uses a catalytically inactive variant of Cas9, and a programmable single guideRNA (sgRNA) (Shen, 2019). The dCas9-sgRNA complex is directed to a site complementary to a 20-nucleotide base-pairing region located downstream of a protospacer adjacent motif (PAM). For an active Cas9, a double stranded break (DSB) is introduced through the concerted actions of two conserved nuclease domains, RuvC and HNH-like domains, which nick the DNA. However, by altering two amino acids in these domains, nuclease activity is abolished, resulting in a catalytically dead, dCas9. This can then bind to the target sequence, but fails to introduce a DSB, effectively silencing transcription, either at the promoter,

preventing RNA polymerase access to the promoter, thus blocking transcription initiation, or by preventing transcription elongation (Larson *et al*., 2013).

As discussed in Chapter IV, random mutagenesis of an S-layer mutant identified changes in gene essentiality in the absence of this surface structure **(Table 4.2)**. Unsurprisingly, several PSII biosynthetic genes, required for synthesis of the polymer that attaches the S-layer to the cell surface are essential in wild type *C. difficile,* and were identified as conditionally non-essential in the absence of an S-layer, namely *cdR20291_2659* (*rkpK,* a UDP-glucose 6-dehyrdogenase) and *cdR20291_2660* (*tuaG,* a teichuronic acid biosynthesis glycosyltransferase). Additionally, *cdR20291_2661* (a beta-glycosyltransferase), and *cdR20291_2665* (a polysaccharide polymerase) were identified as conditionally non-essential in 3 out of 4 data sets and of ambiguous essentiality in the 4$^{th}$.

To silence gene expression, an optimised CRISPRi system was developed. To prove system functionality, *secA2* transcription was silenced, as this gene is conditionally non-essential in FM2.5 and silencing of this gene has a well-defined phenotype (Fagan and Fairweather, 2011). This chapter outlines the construction of a CRISPRi system for *C. difficile*, methodology for target design for knock-down, proof of system functionality using a *secA2* knock-down and electron microscopy of thin sections of a *secA2* knock-down to observe the phenotypic effects of gene silencing.

## 5.2 Results

### 5.2.1 Construction of a CRISPR Interference system for *C. difficile*

Making mutants via allelic exchange can be labour intensive and this method cannot be applied to essential genes as loss of function would lead to cell death. Additionally, the inability to generate a deletion of a particular gene does not infer anything about essentiality. Instead, it would be beneficial to develop a CRISPRi system to study gene function of conditionally essential and non-essential genes in FM2.5, identified from the TraDIS dataset **(Table 4.3 and 4.5).**

sgRNAs for CRISPRi are composed of a 20-nucleotide target-specific complementary sequence, a 42-nucleotide dCas9-binding hairpin (dCas9 handle) and a 40-nucleotide transcription terminator derived from *S. pyogenes* (Qi *et al*., 2013) **(Fig 5.1)**.

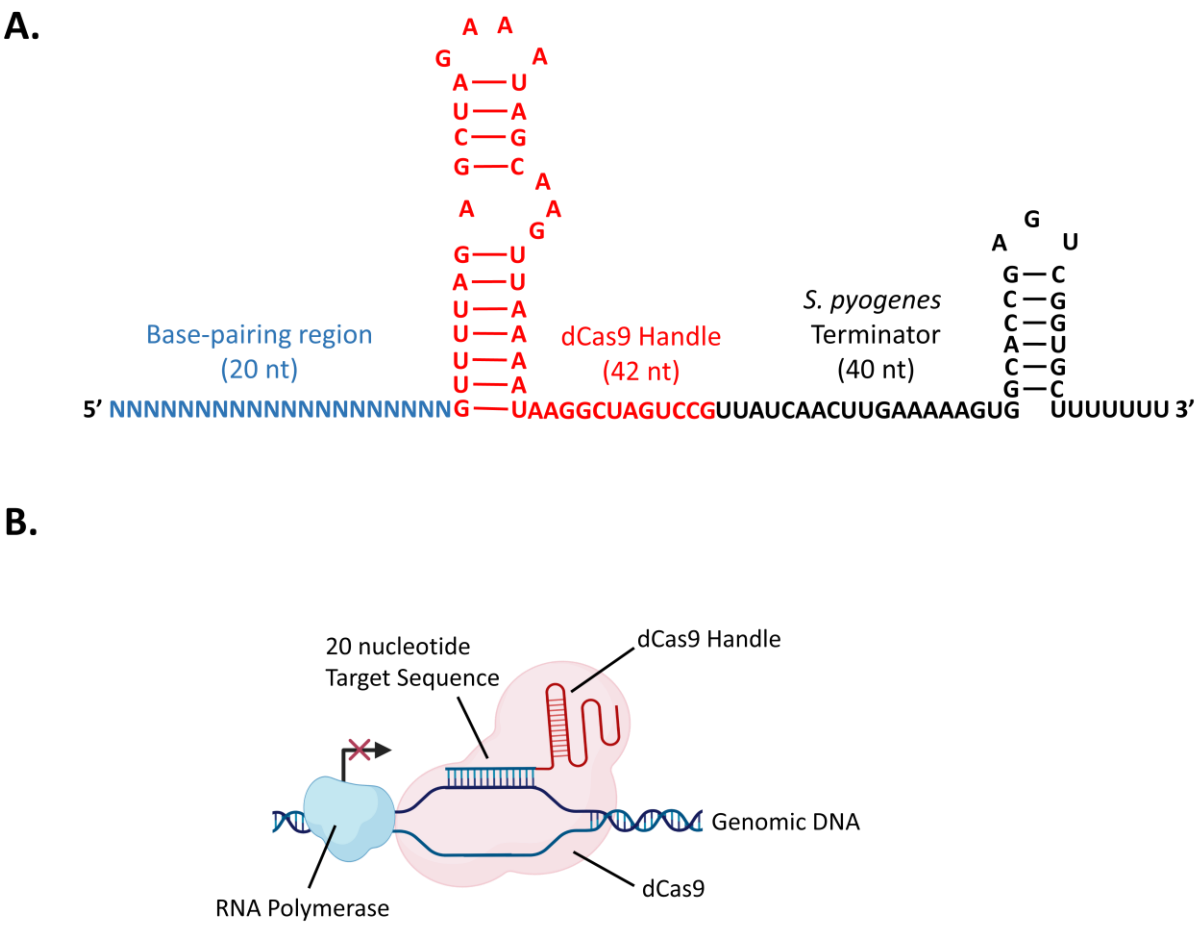
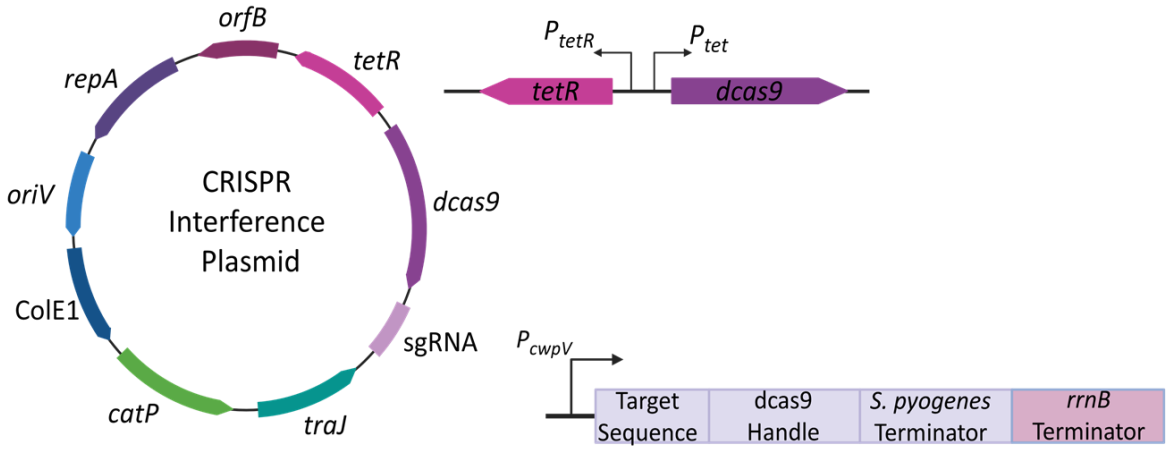



**Fig 5.1. CRISPRi system for transcriptional repression in *C. difficile*. A.** The chimeric sgRNA for *C. difficile* is composed of a 20-nucleotide base pairing region for targeted gene repression, a 42-nucleotide hairpin, the dCas9 handle, required for dCas9 binding, and a 40-nucleotide transcriptional terminator derived from *S. pyogenes*. **B.** dCas9-sgRNA complex binding to the non-template target genomic strand blocks transcription elongation by preventing RNA polymerase access to the target DNA.

Typically, CRISPR and CRISPRi plasmids for *C. difficile* use a 700 bp glutamate dehydrogenase (*gdh*) promoter to drive expression of the sgRNA (McAllister *et al.*, 2017; Muh *et al.*, 2019). Unfortunately, the transcriptional start site of this promoter has not been defined making it a poor choice for precise production of functional RNAs. Use of this promoter in CRISPR will lead to production of an sgRNA of unknown size and with the potential for unintended secondary structure. To overcome this problem, we instead opted to use the *cwpV* promoter. *cwpV* is strongly transcribed

with CwpV accounting for ~15% of the *C. difficile* S-layer (Reynolds *et al*., 2011). Moreover, *cwpV* has defined -10 and -35 sequences and a known +1. Published systems also have potential problems with suboptimal Shine-Dalgarno spacing for *cas9* of 14 bp (McAllister *et al*., 2017; Muh *et al*., 2019). Our system has modified Shine-Dalgarno-ATG for *cas9* spacing of 5-7 bp. The final modification is the introduction of two point mutations in the RuvC and HNH domains of the Cas9 protein, D10A and H840A, respectively, required to abolish the nuclease activity of the protein **(Fig 5.2)**.

**A.**




**B.**

| | **Published CRISPRi Plasmid (Muh, Pannullo et al. 2019)** | **Current CRISPRi System** |
|---|---|---|
| d*cas9* | Distance from d*cas9* to Shine-Dalgarno Sub-optimal | Shine-Dalgarno Spacing Reduced to 7 bp |
| sgRNA | *gdh* Promoter | *cwpV* Promoter |

**Fig 5.2. A CRISPR interference system for *C. difficile*. A.** Schematic representation of a CRISPRi plasmid for *C. difficile*. *dcas9* is under the control of a tetracycline inducible promoter. The tetracycline inducible system is composed of two divergent promoters with overlapping *tet* operator sequences. The gRNA is under the control of the *cwpV* promoter. **B.** Changes made to the published CRISPR plasmid. For the published CRISPR plasmid, *cas9*, was catalytically active and induction of the system would allow DSBs to be made at the target site. Two point mutations in the RuvC and HNH domains of the Cas9 protein, D10A and H840A, were introduced to abolish nuclease activity. Another modification introduced into the system is modified Shine-Dalgarno-ATG spacing for *cas9* of 5-7 bp, reduced from 14 bp, observed in other *C. difficile* CRISPR systems. Lastly, for the published CRISPR and CRISPRi plasmids, the gRNA uses a 700 bp glutamate dehydrogenase (*gdh*) promoter, with no defined transcriptional start site, to drive expression. To avoid unintended secondary structure, this was changed to a *cwpV* promoter with defined -10 and -35 sequences and a known +1.

For CRISPRi vector construction, a DNA sequencing encoding a *pyrE* targeting gRNA was PCR amplified with RF1517/RF1519 and cloned into pRPF185 linearised with BstXI and Bsu361, yielding pSOB008. A catalytically inactive *cas9* with optimised Shine-Dalgarno spacing was synthesised by Genewiz and cloned into the vector pUC-GW-Kan. *dcas9* was PCR amplified from pUC-GW-Kan with primers RF1643/RF1644. *dcas9* was introduced into pSOB008 linearised with BamHI and SacI by Gibson assembly, yielding pSOB009 **(Fig 5.3)**. *dcas9* is under the control of a tetracycline inducible promoter, allowing titratable repression.

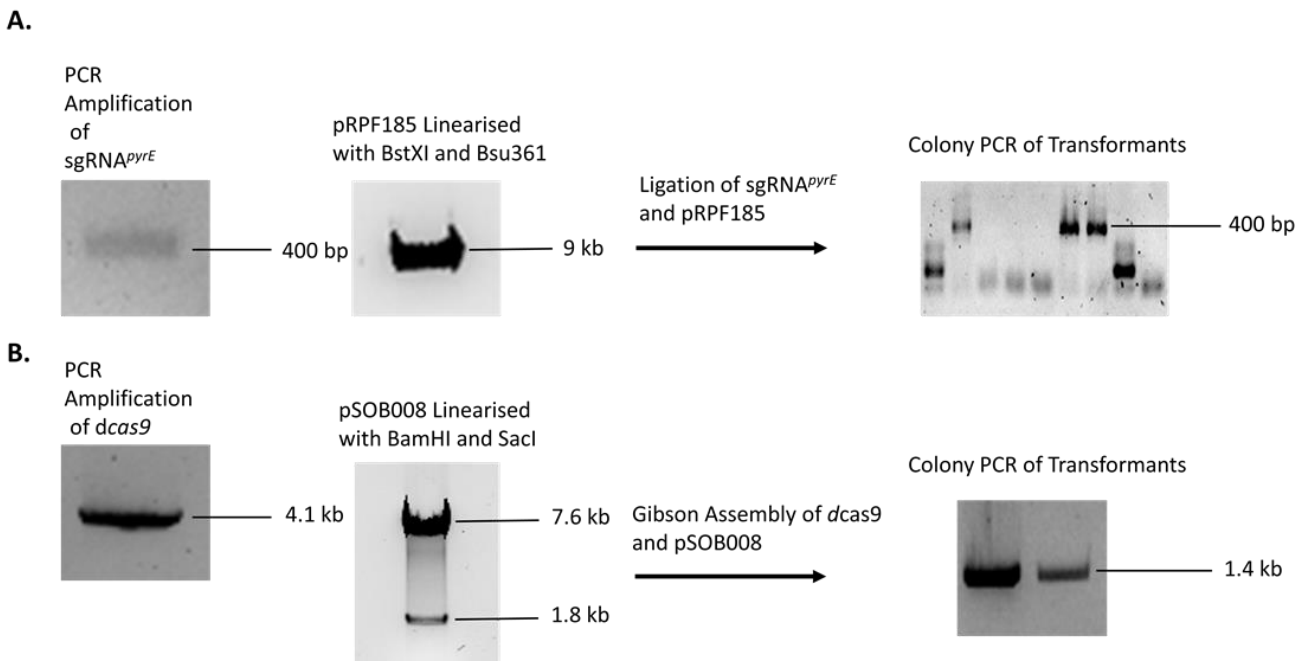



**Fig 5.3. Vector construction of a CRISPRi System for *C. difficile*. A.** sgRNA$^{pyrE}$ cloned into pRPF185 at Bsu361 and BstXI sites through restriction/ligation, yielding pSOB008. Transformants were screened for sgRNA insertion using sgRNA specific primers, RF1517 and RF1519. Expected product size was 400 bp. **B.** *dcas9* cloned under a tetracycline inducible promoter at BamHI/SacI sites in pSOB008 through Gibson Assembly. Transformants were screened for *dcas9* insertion, using two *dcas9* internal screening primers, RF1501 and RF1504, with an expected product size of 1.4 kb.

### 5.2.2 Chimeric single guide RNA design

The published CRISPR system, which our system was constructed from, targets *pyrE*, encoding an orotate phosphoribosyltransferase, required for *de novo* pyrimidine biosynthesis. However, this gene is non-essential in both R20291 and FM2.5. Additionally, the 20-nucleotide target sequence is on the template strand of the DNA. Binding of dcas9-sgRNA complex to this strand has little repressive effect, presumably because RNA polymerase (RNAP) can read through the complex in this orientation (Qi *et al*., 2013). To sterically block access of RNAP to the target site, sequences should be designed on the non-template strand. To prove system functionality, a gRNA was designed to target *secA2*, a conditionally non-essential gene in FM2.5, which has a well-defined

function (Fagan and Fairweather, 2011). sgRNAs were designed for the non-template strand according to Larson *et al*., 2013.

### 5.2.2.1 Workflow for sgRNA target design

First, DNA targeting for CRISPRi requires the presence of a PAM sequence (two to five nucleotides) near the DNA target site. For this CRISPRi system, it is a GG dinucleotide **(Fig 5.4).**

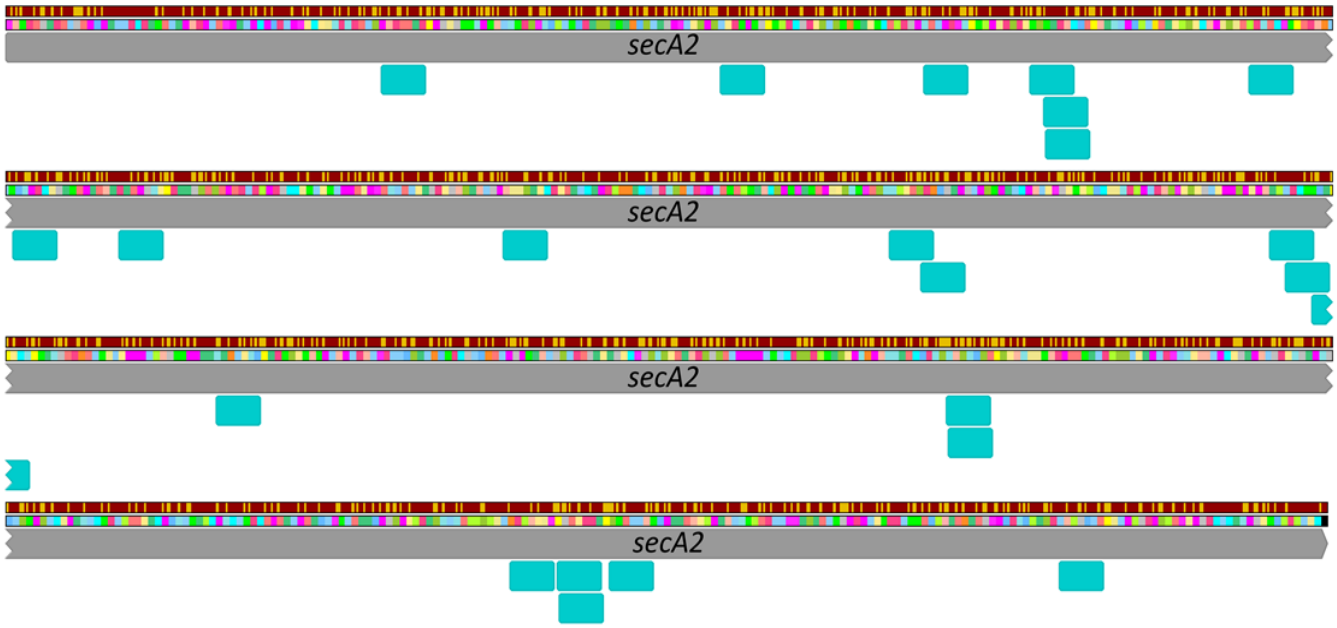



**Fig 5.4. Potential target sequences for knock-down of *secA2*.** DNA targeting of *secA2* by dCas9 requires an NGG protospacer adjacent motif (PAM). Located immediately downstream of the PAM, are the 20 nucleotide dCas9-sgRNA target binding sites (turquoise boxes). For CRISPRi target design, the 20-nucleotide target sequence should be designed such that the dCas9-sgRNA will bind to the non-template strand for optimal repression.

Second, the 20 nucleotides adjacent to the PAM sequence should be a unique binding site, ensuring no off-target effects. Potential *secA2* gRNAs were screened against the R20291 genome using BLAST and any with multiple binding sites were discarded **(Table 5.1).** Moreover, for optimal repression, the target sequence should be positioned closer to the 5' end of the gene. Lastly, the

interaction between dCas9 and the dCas9 handle is crucial. Secondary structure predictions of the target sequence in complex with the dcas9 handle were validated with the RNA-fold server. Target sequences that affected the dCas9 binding RNA structure were discarded **(Fig 5.5).** Of the 23 potential *secA2* target sequences, only one sequence (number 7) met all the criteria for chimeric guide RNA design. A DNA sequence encoding the *secA2* targeting gRNA was PCR amplified with RF1517/RF1519 and cloned into pSOB009 linearised with BstXI and EcoO109I, pSOB010.

**Table 5.1. All potential gRNA sequences for a *secA2* knock-down in *C. difficile*.**

| Potential target sequence | Sequence | Position relative to 5' end of gene | Unique binding site |
|---|---|---|---|
| ***secA2* target sequence 1** | AAAGTTTGGATGATATACTT | Bases 167-186 | No |
| ***secA2* target sequence 2** | AAACTTTAGTTGAGGTAGCT | Bases 317-336 | Yes |
| ***secA2* target sequence 3** | GTGATAAAGAACTTATGCGT | Bases 407-426 | No |
| ***secA2* target sequence 4** | GTAGGTGTAATTTTATCTAA | Bases 454-473 | No |
| ***secA2* target sequence 5** | GTAATTTTATCTAACCAAGA | Bases 460-479 | Yes |
| ***secA2* target sequence 6** | TAATTTTATCTAACCAAGAC | Bases 461-480 | Yes |
| ***secA2* target sequence 7** | ATTTAAGAGATAATATGGTG | Bases 551-570 | Yes |
| ***secA2* target sequence 8** | ACAAAGGGAACTAAACTTTG | Bases 591-610 | Yes |
| ***secA2* target sequence 9** | TAATAGATGAAGCTAGAACT | Bases 638-657 | No |
| ***secA2* Target sequence 10** | TCATTTTTTGGTATAACAAA | Bases 808-827 | No |
| ***secA2* target sequence 11** | AGATATACAGATGGACTTCA | Bases 979-998 | Yes |
| ***secA2* target sequence 12** | ACTTCACCAAGCTATAGAAG | Bases 993-1012 | No |
| ***secA2* target sequence 13** | ATCTATAAATTAAATGTTGT | Bases 1147-1166 | No |
| ***secA2* target sequence 14** | AATTAAATGTTGTCCAAATA | Bases 1154-1173 | Yes |
| ***secA2* target sequence 15** | TCCAAATACCAACTAATAGA | Bases 1166-1185 | No |
| ***secA2* target sequence 16** | GGATACATAAGACTAGACAG | Bases 1268-1287 | Yes |
| ***secA2* target sequence 17** | CGTTCTGGTCGTCAAGGAGA | Bases 1591-1610 | No |
| ***secA2* target sequence 18** | GTTCTGGTCGTCAAGGAGAC | Bases 1592-1611 | No |
| ***secA2* target sequence 19** | ATTTATATAGTACATTTATG | Bases 1985-2004 | No |
| ***secA2* target sequence 20** | TGCAGATACATTACTAATAC | Bases 2006-2025 | Yes |
| ***secA2* target sequence 21** | CTGCAGATACATTACTAATA | Bases 2007-2026 | Yes |
| ***secA2* target sequence 22** | GGTGTAGATAAAAAAAGTGT | Bases 2029-2048 | No |
| ***secA2* target sequence 23** | AATCTTATGCTCAAAAAGAT | Bases 2228-2247 | Yes |

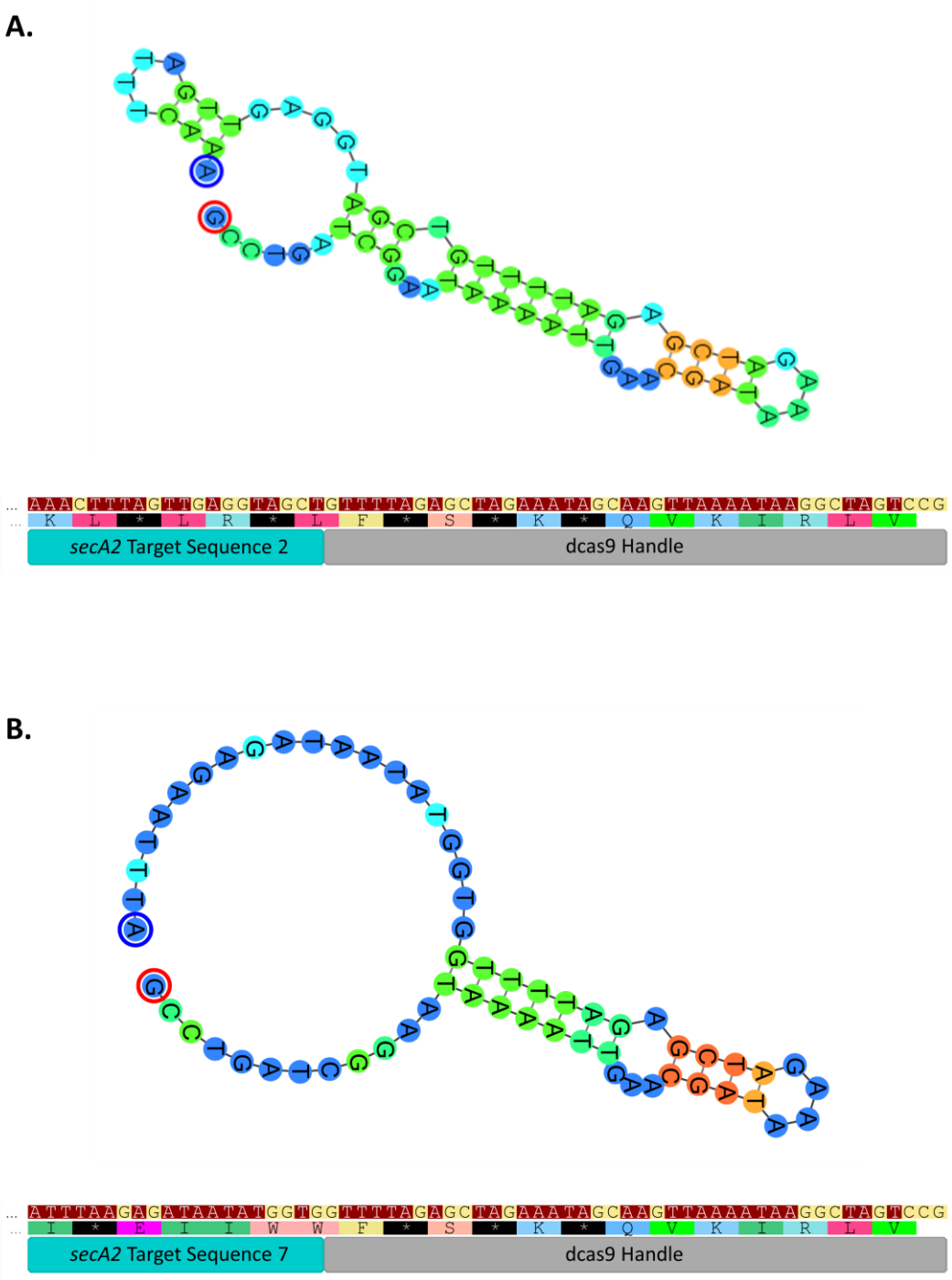



**Fig 5.5. dcas9 handle-sgRNA$^{secA2}$ secondary structure predictions.** The secondary structure of the 20-nucleotide *secA2* target sequence in complex with the dCas9 handle (42-nucleotides) was determined using RNA-fold. **A.** RNA-fold of dCas9 handle-sgRNA$^{secA2\ target\ sequence\ 2}$. The target site disrupts the dCas9 handle hairpin, preventing dCas9 binding. **B.** RNA-fold of dCas9 handle-sgRNA$^{secA2\ target\ sequence\ 7}$. Unlike **A.,** the target sequence of this sgRNA does not hinder hairpin formation.

### 5.2.3 CRISPRi for conditional gene repression

#### 5.2.3.1 Growth analysis of a secA2 knock-down

Induction of the *secA2* targeting CRISPRi system would result in the dCas9-gRNA complex scanning the target DNA for the GG-dinucleotide, annealing at the correct *secA2* target, and blocking RNAP access and transcription elongation. As a result, each cell division would reduce the SecA2 protein copy number. In R20291, this knock-down should be fatal, as depletion of SecA2 will ultimately prevent SlpA secretion (Fagan and Fairweather, 2011). From the FM2.5 TraDIS data, we confirmed that *secA2* is conditionally non-essential in the absence of SlpA, and depletion of *secA2* in this background may cause a growth defect but should not be fatal to cells. At the point where SecA2 has been depleted from the population, growth rate should plateau, and this would be used as a starting point to determine the phenotypic effects of SecA2 depletion. Growth assays were performed in R20291 and FM2.5 by monitoring the optical density of non-induced cells harbouring, pSOB010, with *dcas9* under control of a tetracycline inducible promoter and sgRNA$^{secA2}$ under control of *cwpV*, and cells induced with ATc (20 ng/ml) **(Fig 5.6-5.7)**. O/N cultures were subcultured to 0.01 in TY with and without ATc (20 ng/ml), ensuring cells in a population go through more generations, depleting SecA2 further. R20291 and FM2.5 harbouring pRPF185 were used as a control for induction. pRPF185 is a shuttle vector containing a tetracycline-inducible β-glucuronidase, *gusA* and induction with ATc does not impact growth rate significantly. Another control was R20291 and FM2.5 harbouring pSOB011, a CRISPRi knock-down plasmid whereby the 20-nucleotide target sequence has been deleted.

Both non-induced R20291/pRPF185 and R20291/sgRNA$^{no\ target\ sequence}$ grew similarly and reached the same final optical density **(Fig 5.6)**. Minor ATc toxicity was observed for induced R20291 and FM2.5 strains harbouring pRPF185.

Surprisingly, R20291/sgRNA$^{secA2}$ exhibited a growth pattern similar to FM2.5/pRPF185, reaching comparable maximum $OD_{600nm}$ values. We have previously observed leaky transcription

from the tetracycline inducible promoter when making insertion mutagenesis libraries for *C. difficile* (Chapter III). CRISPRi is highly efficient in bacterial cells and promoter leakage can result in up to 80% repression (Larson *et al*., 2013). Non-induced FM2.5 samples exhibit similar growth patterns, however, there is a stark contrast upon induction of *dcas9* expression **(Fig 5.7)**. For the *secA2* and no target sequence controls, growth rate was severely impacted. Surprisingly, in FM2.5, where *secA2* is conditionally non-essential, the impact of SecA2 depletion is more severe. For the no target sequence controls, growth rates in both R20291 and FM2.5 are slowed, albeit more so in FM2.5.

### 5.2.3.2 Colony morphology of a *secA2* knock-down

Previously, FM2.5 has been shown to form smaller smoother colonies than those of R20291 (Kirk *et al*., 2017b). After 48 h growth on BHI supplemented with thiamphenicol (15 µg/ml) and thiamphenicol and ATc (20 ng/ml) several morphological differences were observed. R20291/pRPF185 formed typical 'rough' colonies and this did not change upon addition of ATc **(Fig 5.6)**. For R20291/sgRNA$^{secA2}$, non-induced and induced samples were smaller and smoother than R20291/pRPF185 colonies. Small smooth colonies were observed for all plasmids and all conditions in FM2.5 **(Fig 5.7)**. The only striking difference was that of induced FM2.5/sgRNA$^{no\ target\ sequence}$, which formed the smallest colonies. The observation that R20291/sgRNA$^{secA2}$ grows like FM2.5/pRPF185 in liquid media in combination with the smooth colonies observed for both non-induced and induced samples lead us to hypothesise that leaky expression of *dcas9* was causing substantial repression of *secA2* in a non-induced background.

### 5.2.3.3 Western blot analysis of a *secA2* knock-down

To test this hypothesis, western blot analysis of FM2.5/sgRNA$^{secA2}$ was carried out. As there are severe growth defects for FM2.5/sgRNA$^{no\ target\ sequence}$, the control was changed for a dCas9-*pyrE* targeting plasmid with 20 nucleotides complementary to the template strand, having little repressive effect upon induction (Larson *et al*., 2013). To sufficiently deplete SecA2 from the population R20291 and FM2.5 harbouring pRPF185, sgRNA$^{pyrE}$ and sgRNA$^{secA2}$ were grown O/N.

Stationary phase cultures were diluted 1:1000, 1:500, 1:250, 1:125, 1:62.5 and 1:31.25 with and without inducer and were grown O/N. The next morning, $OD_{600nm}$ was measured. Induced and non-induced R20291/pRPF185, R20291/sgRNA$^{pyrE}$, FM2.5/pRPF185, FM2.5/sgRNA$^{pyrE}$ grew at all dilutions. However, R20291/sgRNA$^{secA2}$ and FM2.5/sgRNA$^{secA2}$ grew only at 1:62.5 and 1:31.25 dilutions. The 1:62.5 dilutions reached maximum $OD_{600nm}$ of 0.3-0.4, while the 1:31.25 dilutions reached $OD_{600nm}$ of 0.8-1. As the 1:62.5 dilutions had gone through more generations, SecA2 would be sufficiently depleted. 1:62.5 dilutions of FM2.5 harbouring pRPF185, sgRNA$^{pyrE}$ and sgRNA$^{secA2}$ with and without inducer were taken for Western Blot analysis. No differences in SecA2 levels were observed for FM2.5/pRPF185 non-induced and induced samples **(Fig 5.7)**. Similar results were obtained for FM2.5/sgRNA$^{pyrE}$ whereby no differences in SecA2 levels were detected. For non-induced and induced FM2.5/sgRNA$^{secA2}$ no SecA2 was detected confirming leakage of dCas9 and *secA2* repression.


**Fig 5.6. Growth analysis of a *secA2* knock-down in *C. difficile* R20291. A.** Growth assays of a *secA2* knock-down. O/N cultures were subcultured to $OD_{600nm}$ 0.01 in 5 ml pre-reduced TY-thiamphenicol (non-induced) and TY-thiamphenicol-ATc (induced). $OD_{600nm}$ values were monitored hourly. R20291/pRPF185 and R20291/sgRNA$^{no\ target\ sequence}$ functioned as controls. R20291/pRPF185 grew similarly in the presence and absence of inducer, reaching comparable maximum $OD_{600nm}$ values. R20291/sgRNA$^{secA2}$ exhibited slower growth without inducer, and growth rate was further hindered upon induction of *dcas9* expression. Non-induced R20291/sgRNA$^{no\ target\ sequence}$ grew similarly to R20291/pRPF185. However, upon induction with ATc, a growth defect was observed for R20291/sgRNA$^{no\ target\ sequence}$. Shown are the mean, standard deviations, and a linear regression model fitted to the growth data to confirm differences in growth rates. **B.** Colony morphologies for a *secA2* knock-down after 48 h of growth on BHI supplemented with thiamphenicol (non-induced) and BHI-thiamphenicol-ATc (induced). Smaller 'smooth' colonies were observed for R20291/sgRNA$^{secA2}$ for non-induced and induced samples, with the induced colonies being much smaller. R20291/sgRNA$^{no\ target\ sequence}$ maintained the 'rough' edged colonies observed with R20291/pRPF185, although these colonies were smaller than non-induced and induced R20291/pRPF185 colonies. Experiments were carried out in triplicate using biological duplicates.

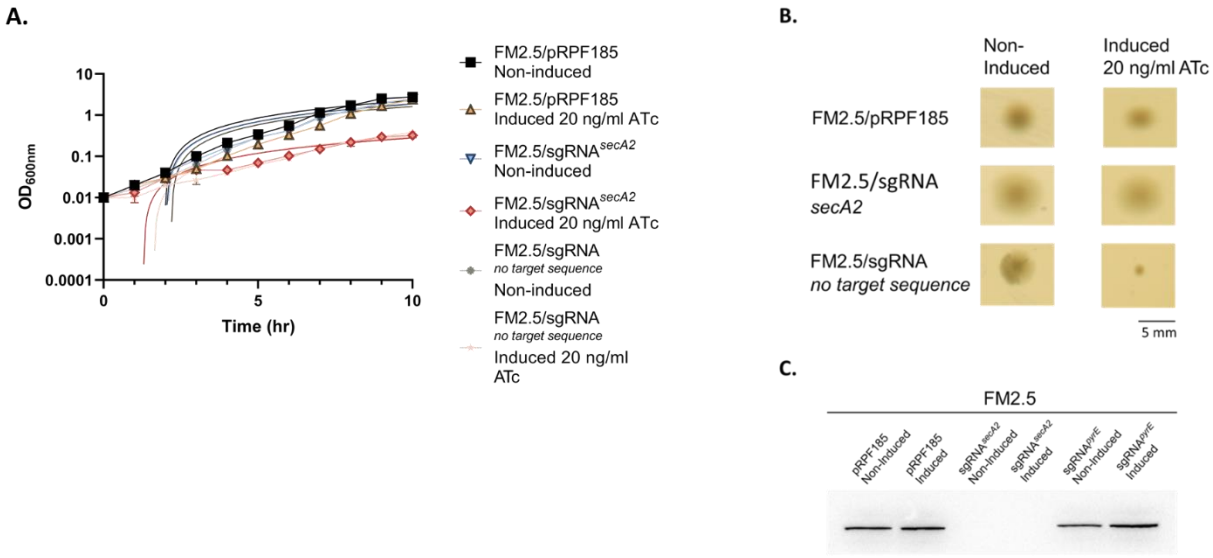



**Fig 5.7. Growth analysis of a *secA2* knock-down in *C. difficile* FM2.5. A.** Growth assays of a *secA2* knock-down in the S-layer mutant. O/N cultures were subcultured to $OD_{600nm}$ 0.01 in 5 ml pre-reduced TY-thiamphenicol (non-induced) and TY-thiamphenicol-ATc (induced). $OD_{600nm}$ values were monitored hourly. FM2.5/pRPF185 and FM2.5/sgRNA$^{no\ target\ sequence}$ functioned as controls. FM2.5/pRPF185 grew similarly in the presence and absence of inducer, reaching comparable maximum $OD_{600nm}$ values. Moderate ATc toxicity was observed for FM2.5/pRPF185 at 20 ng/ml. FM2.5/sgRNA$^{secA2}$ and FM2.5/sgRNA$^{no\ target\ sequence}$ exhibited comparable slow growth patterns without inducer, and growth rate was further hindered upon induction of *dcas9* expression. Shown are the mean, standard deviations, and a linear regression model fitted to the growth data to confirm differences in growth rates.**B.** Colony morphologies for a *secA2* knock-down after 48 h of growth on BHI supplemented with thiamphenicol (non-induced) and BHI-thiamphenicol-ATc (induced). For both non-induced and induced samples, small 'smooth' colonies were observed for all FM2.5 strains. The colonies for FM2.5/sgRNA$^{no\ target\ sequence}$ were much smaller upon induction. **C.** Western blot of *secA2* knock-down in FM2.5. Experiments were carried out in triplicate using biological duplicates.

## 5.3 Discussion

To study gene function of conditionally essential and non-essential genes in FM2.5, we have developed a CRISPRi system for RNA-mediated gene regulation using a catalytically inactive variant of Cas9, dCas9. Co-expression of dCas9 with an sgRNA with a 20 bp target sequence results in specific silencing of the gene of interest, as the dCas9 is defective in cleavage. The CRISPRi plasmid is an *E. coli-C. difficile* shuttle vector that is composed of codon-optimised *dcas9* under the control of a tetracycline-inducible promoter and a constitutively expressed sgRNA, under the control of $P_{cwpV}$. Tuneable repression is achieved by inducing dCas9 expression with anhydrotetracycline. The concentration used for these studies is 20 ng/ml. However, for FM2.5, bacterial growth is affected at these concentrations, as observed with the control plasmid pRPF185. We anticipate that a xylose-inducible promoter system will circumvent the deleterious effects of promoter toxicity. Despite this, CRISPRi represents an attractive tool for studying essential genes as targeting a new gene only requires the cloning of a new sgRNA and a myriad of DNA sequences could be targeted by altering the spacer sequence within the sgRNA, provided a PAM sequence is located near the target sequence (Anzalone *et al*., 2020).

This CRISPRi system was specifically designed to understand gene function of conditionally essential and non-essential genes in the absence of an S-layer. Using the protocol for chimeric sgRNA design outlined in section 5.2.2.1 and aiming to minimise off-target effects, we have designed a system for SecA2 depletion, a conditionally non-essential protein in the absence of SlpA.

However, we have observed that for non-induced WT *C. difficile* targeting SecA2, no SecA2 is detected in these samples. We hypothesise that promoter leakage is resulting in d*cas9* expression and *secA2* silencing, preventing S-layer secretion. This is supplemented by the disappearance of the typical rough colony morphology observed for R20291, and the appearance of the characteristic 'smooth' colonies observed for FM2.5. CRISPRi is highly efficient in bacterial cells and promoter leakage can result in up to 80% repression, when compared to controls harbouring the d*cas9* or

sgRNA alone (Larson *et al*., 2013). In future a switch to employing a xylose-utilisation operon represents a promising potential advantage over tetracycline inducible systems, as no promoter leakage has been identified.

Additional improvements that could be made to the current interference system are the optimisation of the controls plasmids. We used pRPF185, a TetR-regulated plasmid, with induction resulting in GusA expression, to show induction with ATc does not significantly impact growth rate. Additionally, we used a CRISPRi plasmid, whereby the 20-nucleotide targeting sequence of the sgRNA had been deleted. For this plasmid, in both R20291 and FM2.5, we observed severe deleterious effects on growth rate. For this, we hypothesise that dCas9 is lacking target specificity. Normally, dCas9 searches for the NGG PAM sequence along the entire genome. The transient binding to the PAM sequence provides the energy required to unwind the dsDNA immediately upstream of the GG dinucleotide, followed by base-pairing between sgRNA and the seed target sequence. As the sgRNA$^{no\ target\ sequence}$ lacks target specificity, the inability to anneal results in quick release of dCas9, which then samples other sequences. We speculate that the repeat process of dCas9 binding and dissociating from the DNA is a very energetically costly process, owing to the slowed growth rates observed for strains harbouring this plasmid. We also used a *pyrE* targeting sgRNA, whereby the 20-nucleotide target sequence was complementary to the target strand, which for CRISPRi would have little effect upon induction. However, we anticipate that a slight growth defect would be observed, due to the size of the dCas9-sgRNA complex binding at the target, resulting in perhaps sub-optimal transcriptional read-through. Several controls would be more desirable for this CRISPRi system. The first would be a shuttle vector harbouring the TetR-regulated dCas9 alone, showing that with induction, dCas9 is unable to form a heterocomplex with the sgRNA and unable to bind to the genomic DNA. The second would be a pseudo-sgRNA, with 20 nucleotides that does not anneal next to a PAM sequence.

Despite the limitations with our system, CRISPRi is a platform to inexpensive target gene regulation that offers several advantages compared with existing methods for gene silencing, such as RNA interference, which exhibits significant off-target effects and toxicity (Qi *et al.*, 2013). Additionally, the ability to explore the function of essential genes is extremely advantageous. Future work will focus on optimising our system through replacing the TetR-regulated promoter with the xylose inducible system and building the appropriate controls.

# Chapter VI. Phenotypic characterisation of conditionally non-essential genes in the absence of SlpA.

## 6.1 Introduction

As discussed in Chapter IV, random mutagenesis of an S-layer mutant, FM2.5, identified changes in gene essentiality in the absence of this surface structure **(Table 4.3 and 4.5)**. Unsurprisingly, PSII biosynthetic genes are essential in the presence of an S-layer as this is the anionic polymer responsible for S-layer attachment to the *C. difficile* cell wall. However, from the FM2.5 transposon mutagenesis data, several PSII biosynthetic genes were identified as conditionally non-essential in the absence of an S-layer, namely *cdR20291_2659* (*rkpK,* a UDP-glucose 6-dehyrdogenase) and *cdR20291_2660* (*tuaG,* a teichuronic acid biosynthesis glycosyltransferase). Two other PSII biosynthetic genes were recognised as conditionally non-essential in 3 of 4 datasets: *cdR20291_2661* (a beta-glycosyltransferase), and *cdR20291_2665* (a polysaccharide polymerase) **(Fig 6.1).**

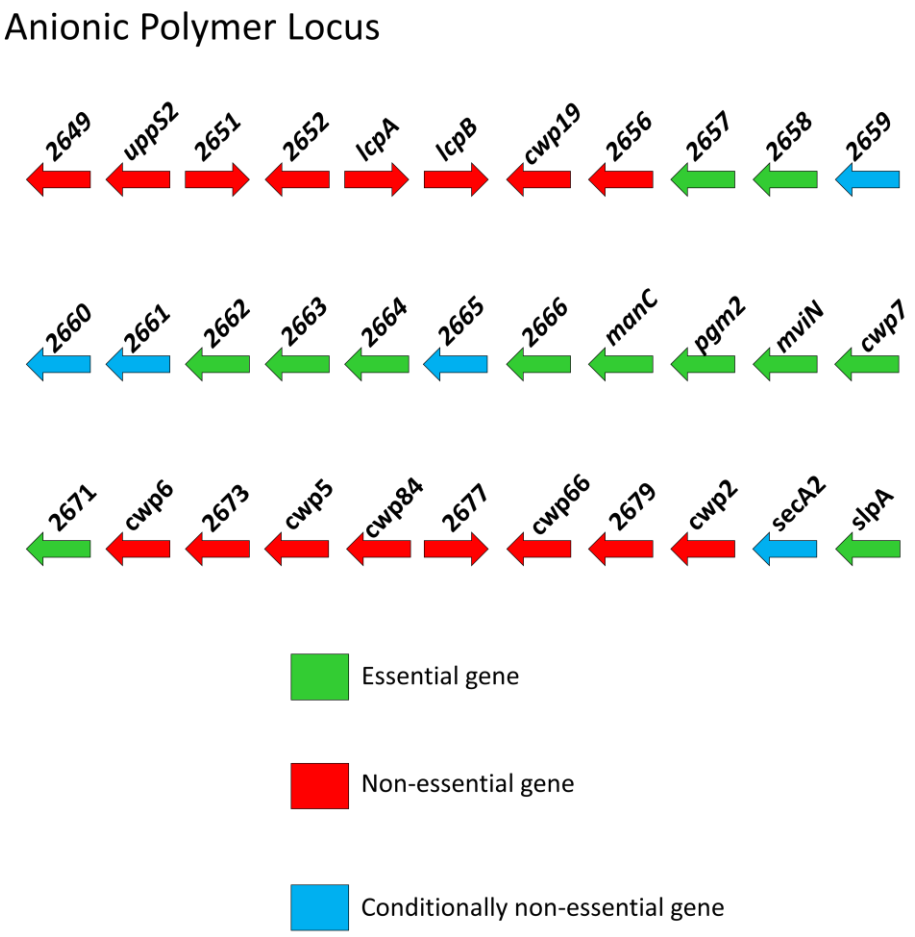



**Fig 6.1 Changes in gene essentiality for the anionic polymer locus in the absence of the *C. difficile* S-layer.** Non-essential genes of the AP locus for the WT, R20291 strain and the S-layer mutant, FM2.5 are indicated by red arrows. Genes that maintained essentiality between the two strains are shown with green arrows. Genes identified as dispensable for growth in the absence of SlpA are indicated by blue arrows. *cdR20291_2661* and *cdR20291_2665* were shown to be conditionally non-essential in 3 of 4 datasets from the TraDIS analysis, and of ambiguous essentiality in the 4$^{th}$, and as such are indicated as conditionally non-essential for the purpose of this diagram.

To study the phenotypic effects of transcriptional silencing of the targets obtained from the TraDIS analysis, a CRISPRi system was developed (Chapter V). Initially, this system was used to compare the effects of SecA2 depletion in a WT and S-layer mutant strain. This gene was identified as conditionally non-essential in the absence of the S-layer and depletion of this protein has a well-established phenotype (Fagan *et al.*, 2011). Use of this system was then extended to other

conditionally non-essential genes, as these have the potential to disrupt S-layer assembly, which is a key virulence determinant for *C. difficile*. The first target chosen for phenotypic studies was *cdR20291_2329* (*era*). Era is a GTPase involved in 30S ribosome subunit biogenesis (Wood *et al*., 2019). Specifically, Era functions as a molecular switch in ribosome biosynthesis regulation, cycling between a GDP-bound inactive state and a GTP-active state. While Era is not essential in all bacterial species, depletion of Era has several phenotypic effects, namely defects in cell cycle control, ribosome biogenesis and chromosome segregation (Wood *et al*., 2019). Additionally, for *B. subtilis*, Era-depleted cells are defective in spore formation (Minkovsky *et al*., 2002). As such, it would be interesting to see if depletion of Era from *C. difficile* shows the same sporulation defect and the fate of the S-layer throughout the growth cycle in Era-depleted cells.

Another target chosen was *cdR20291_2359*, which encodes a poly(A) polymerase, the enzyme responsible for the synthesis of poly(A) tails at the 3' end of RNA. Typically, bacterial mRNAs have short poly(A) tails which are not adenylated, and as such are accessible to exoribonucleases (O'Hara *et al*., 1995). Both model organisms *E. coli* and *B. subtilis* encode two poly(A) polymerases (Cao and Sarkar, 1992; Kalapos *et al*., 1994; Sarkar, 1997). For *E. coli*, addition of the poly(A) tails destabilises RNAs, which is the diametrical opposite of what is observed with eukaryotic mRNAs (Rauhut and Klug, 1999). To date, it is unclear if the function observed in *E. coli* extends to *B. subtilis* or even to *C. difficile*, and what connection this polymerase has to the S-layer. Lastly, *cdR20291_2659* (*rkpK*), *cdR20291_2660* (*tuaG*) and *cdR20291_2661* were chosen. As mentioned previously, these genes are located within the AP locus and are believed to be involved PSII synthesis. As this is the polymer responsible for attachment of all 29 CWPs to the cell wall, it is still surprising that some genes of the locus can tolerate transposon insertions. As such, it would be interesting to study the effects of PSII depletion in the WT and S-layer mutant to observe any differences in the cell surface composition. This chapter outlines sgRNA design and the phenotypic effects of transcriptional silencing of the outlined targets.

## 6.2 Results

### 6.2.1 sgRNA design

All gRNAs were constructed using the methodology described in Chapter V **(Table 6.1)**. Additionally, the sequence used for each CRISPRi knock-down is shown **(Table 6.2)**. DNA sequences encoding each targeting gRNA were synthesised by Genewiz, PCR amplified with RF2028 and RF1519 and cloned into pSOB010 linearised with BstXI and EcoO109I.

**Table 6.1 Workflow of sgRNA target design**

| | Gene orientation | dCas9 binding site | Number of targets with adjacent PAM | Number of unique targets | Number of targets with correct RNA-fold | Gene length (bp) | Position relative to 5' end of gene |
|---|---|---|---|---|---|---|---|
| ***cdR20291_2329*** | 3'-5' | 20 ntd – NGG | 16 | 10 | 1 | 894 | 254-273 |
| ***cdR20291_2359*** | 3'-5' | 20 ntd – NGG | 19 | 12 | 2 | 1,413 | 196-215 |
| ***cdR20291_2659*** | 3'-5' | 20 ntd – NGG | 19 | 8 | 2 | 1,344 | 335-354 |
| ***cdR20291_2660*** | 3'-5' | 20 ntd – NGG | 7 | 5 | 1 | 753 | 320-339 |
| ***cdR20291_2661*** | 3'-5' | 20 ntd – NGG | 13 | 7 | 2 | 909 | 391-410 |

**Table 6.2 sgRNA target sequences for gene silencing**

| | Sequence |
|---|---|
| ***cdR20291_2329*** | ATTGTATTTCTTGTAGTTTG |
| ***cdR20291_2359*** | GTTGTTATATCCCAATCGTT |
| ***cdR20291_2659*** | TGCCTTTATATAATCTAGGT |
| ***cdR20291_2660*** | TTATAGTATTTCTAAGTAAT |
| ***cdR20291_2661*** | GTCCCTATCAAATCTCCACT |

### 6.2.2 Growth analysis of a CRISPRi knock-down

To study the phenotypic effects of protein depletion of our targets, we must first determine the point at which the protein has been sufficiently depleted from the population. Induction of the CRISPRi system would result in dCas9-gRNA complex blocking RNAP access and transcriptional silencing, with each cell division reducing the protein copy number of the target. This manifests as a plateau in growth rate. As the protein copy number and minimum amount needed for function will vary between targets, and perhaps between the WT and S-layer mutant, growth rate for each knock-down was measured individually.

Using *cdR20291_2660* as an example, this knock-down should be fatal in R20291, as evidenced from the R20291 TraDIS data (Dembek *et al*., 2015). However, in FM2.5 *cdR20291_2660* has been identified as conditionally non-essential in the absence of SlpA, and depletion of this protein in this background may impact growth but should not be fatal. To determine the phenotypic effects of CDR20291_2660 depletion, growth assays were performed in R20291 and FM2.5 by monitoring the optical density of non-induced cells harbouring, sgRNA$^{cdR20291_2660}$. Cells were induced with ATc (20 ng/ml) **(Fig 6.3)**. O/N cultures were subcultured to a lower $OD_{600nm}$ than usual (0.01) in TY with and without ATc, ensuring cells in the population went through as many generations as possible in the timeframe of the experiment, depleting CDR20291_2660 further. R20291 and FM2.5 harbouring pRPF185 were used as controls. Another control was R20291 and FM2.5 harbouring sgRNA$^{pyrE}$, a plasmid where the 20-nucleotide targeting sequence for *pyrE* is directed to the template strand, the sub-optimal strand for knock-down.

For R20291, all non-induced strains grew similarly and reached the same final optical density **(Fig 6.2)**. R20291/sgRNA$^{cdR20291_2660}$ exhibited a considerable growth defect, but eventually reached the same $OD_{600nm}$ seen for R20291/pRPF185. Interestingly, no premature growth plateau was observed for induced R20291/sgRNA$^{cdR20291_2660}$. R20291/sgRNA$^{pyrE}$ induced samples exhibited a severe growth defect, reaching final $OD_{600nm}$ of ~2.4, compared to 3.3 observed for R20291/pRPF185

induced samples. The growth defect we observed suggests that even 20 nucleotides directed to the template DNA strand is not an appropriate control. Instead, a shuttle vector harbouring the TetR-regulated dCas9 alone would not be expected to have any impact on growth rate and should be used for future experiments. An additional plasmid could be used which harbours a pseudo-sgRNA, with 20 nucleotides that does not anneal next to a PAM sequence.

For FM2.5/pRPF185, the growth pattern appeared normal, reaching final $OD_{600nm}$ of ~2.8. Non-induced FM2.5/sgRNA$^{pyrE}$ samples exhibit a considerable growth defect, which worsened upon induction of *dcas9* expression **(Fig 6.3)**. FM2.5/sgRNA$^{cdR20291_2660}$ non-induced samples are slow-growing, with final $OD_{600nm}$ values being lower than both pRPF185 and sgRNA$^{pyrE}$ controls. Growth rate for induced samples targeting *cdR20291_2660* is hindered, however, final $OD_{600nm}$ values are higher than those obtained for a *pyrE* knock-down. Again, no premature plateau in growth rate was observed for this knock-down.

Unfortunately, subculturing strains to lower starting $OD_{600nm}$ failed to pinpoint the rate at which the protein of interest is sufficiently depleted from the population. As described in Chapter V, it would be more desirable to dilute stationary phase cultures into fresh TY with and without inducer, and measure the final $OD_{600nm}$ of cultures grown O/N. For induced samples, the dilution with the lowest final $OD_{600nm}$ would be taken for phenotypic analyses. To sufficiently deplete our CRISPRi targets from the population R20291 and FM2.5 harbouring pRPF185, sgRNA$^{pyrE}$, sgRNA$^{cdR20291_2329}$, sgRNA$^{cdR20291_2359}$, sgRNA$^{cdR20291_2659}$, sgRNA$^{cdR20291_2660}$ and sgRNA$^{cdR20291_2661}$, were grown O/N. Stationary phase cultures were diluted 1:1000, 1:500, 1:250, 1:125, 1:62.5 and 1:31.25 with and without inducer and were grown O/N. The next morning, the $OD_{600nm}$ values were measured. Induced and non-induced R20291/pRPF185 and FM2.5/pRPF185 grew at all dilutions. However, all CRISPRi targets grew only at 1:62.5 and 1:31.25 dilutions. The 1:62.5 dilutions reached lower maximum $OD_{600nm}$ than 1:31.25 dilutions, ~0.2-0.4 compared 0.8-1.5. As the 1:62.5 dilutions had gone through more generations, the protein levels of these targets should be sufficiently removed

from the population. As such, 1:62.5 dilutions of all strains with and without inducer were taken for TEM analysis of thin-sections. All thin sectioning experiments were performed at the same time as those shown in Chapter V.

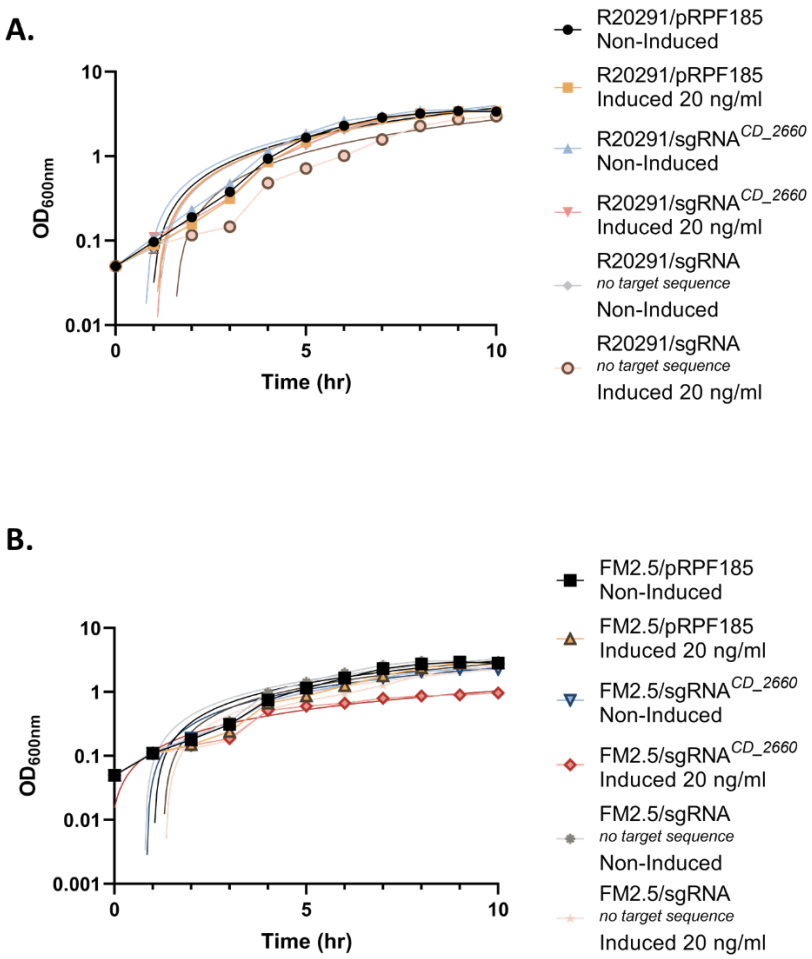



**Fig 6.2. Growth analysis of a *cdR20291_2660* knock-down using CRISPRi.** Growth assays of a *cdR20291_2660* knock-down in R20291 and FM2.5. O/N cultures were subcultured to $OD_{600nm}$ 0.01 in 5 ml pre-reduced TY-thiamphenicol (non-Induced) and TY-thiamphenicol-ATc (induced). $OD_{600nm}$ values were monitored hourly. R20291 and FM2.5 harbouring pRPF185 and sgRNA$^{pyrE}$ functioned as controls. R20291/pRPF185 grew similarly in the presence and absence of inducer, reaching comparable maximum $OD_{600nm}$ values. R20291/sgRNA$^{cdR20291_2660}$ non-induced samples grew similar to the controls, but growth rate was hindered upon induction of *dcas9* expression. Non-induced R20291/sgRNA$^{pyrE}$ grew similarly to R20291/pRPF185. However, upon induction with ATc, a growth defect was observed. FM2.5/pRPF185 grew similarly in the presence and absence of inducer, reaching comparable maximum $OD_{600nm}$ values. Moderate ATc toxicity was observed for FM2.5 pRPF185 at 20 ng/ml. FM2.5/sgRNA$^{cdR20291_2660}$ and FM2.5/sgRNA$^{pyrE}$ exhibited comparable slow growth patterns without inducer. Growth rate was further hindered upon induction of *cdR20291_2660* knock-down. Experiments were carried out in triplicate using biological duplicates. Shown are the mean and standard deviation, and a linear regression model to confirm differences in growth rate.

### 6.2.3 Depletion of CDR20291_2359 from *C. difficile* cells.

As mentioned above, *cdR20291_2359*, encoding a poly(A) polymerase was chosen for phenotypic studies. This enzyme is responsible for the synthesis of poly(A) tails at the 3' end of RNA. This gene was identified as dispensable for growth in the absence of SlpA, however, the connection between this gene and the S-layer is unknown. Depletion of CDR20291_2359 has a plethora of effects on cell surface organisation. R20291/sgRNA$^{cdR20291_2359}$ non-induced samples appear stressed, displaying irregular cell boundaries and some envelope blebbing **(Fig 6.4)**. For induced samples, several cells appear to aggregate, with some cells in the cluster exhibiting a curved cell shape, deviating from the typical rod morphology observed for *C. difficile*. Cell lysis is also observed for these samples**.** For FM2.5/sgRNA$^{cdR20291_2359}$ non-induced samples, some irregular cell boundaries, curvature, misplaced septas and oddly shaped cells are observed. Cell debris was observed for FM2.5 induced samples, which could be part of the cell envelope or the genome. All images of CDR20291_2359 depleted cells for R20291 and FM2.5 are shown in the Appendices.

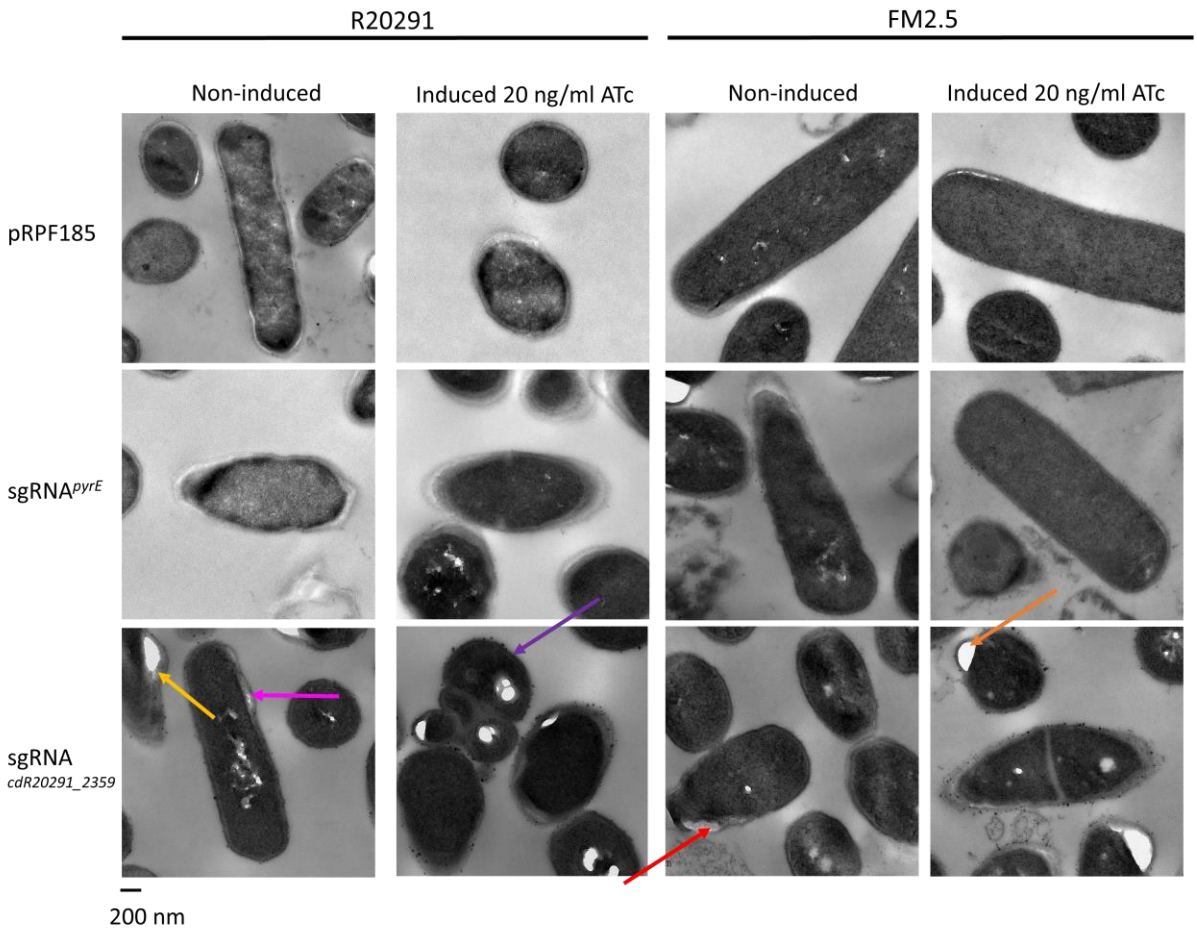



**Fig 6.3. The effects of CDR20291_2359 depletion on cell morphology.** R20291 non-induced samples show some envelope blebbing (yellow arrow) and irregular cell boundaries (pink arrow), indicative of envelope stress**.** For R20291 induced samples, several cells appear to clump, with some cells in the cluster exhibiting a cell curvature, a stark contrast from the healthy long rod shape that is typical for R20291. Additionally, deposits are observed along the outside of the cell, which could correspond to ribosomes (purple arrow). These deposits are also observed in FM2.5 induced samples. Like R20291, some irregular cell boundaries are observed for CDR20291_2359 depletion in the S-layer mutant (red arrow). Upon induction, some cell lysis and membrane blebbing is observed (orange arrow).

### 6.2.4 Morphological effects of depleting RkpK from cells.

*cdR20291_2659, rkpK*, is located within the AP locus and is believed to function in PSII synthesis. RkpK is a UDP glucose 6-dehydrogenase that was identified as conditionally non-essential in the absence of the S-layer and functions with the product of another dispensable gene, *gtaB*, to convert glucose-1P to UDP-glucoronate. For non-induced R20291/sgRNA$^{cdR20291_2659}$, cells appear

longer than typical R20291/pRPF185 cells and possess irregular cell boundaries **(Fig 6.5)**. Moreover, cell curvature and irregular cell shape was observed for induced cells. Depletion of CDR20291_2659 from FM2.5 cells has a myriad of phenotypes. Firstly, for the non-induced samples, cells display this curved morphology when dividing. Moreover, the surface of cells show 'hairy'-like appendages, which could correspond to shedding of cell wall polysaccharide from the surface. It is possible that the polysaccharide is being overproduced, or cells are defective in polysaccharide anchorage, causing this polymer to bulge from the cell surface. Future experiments should study the effects of RkpK depletion in more detail. Additionally, for induced FM2.5 cells, depletion of RkpK from cells results in differences in envelope permeability, displaying a thick border around cells. We speculate that cells are taking up the stain differently in these samples. It is possible that changes to the surface chemistry, namely charge or chemical composition of the cell wall, could affect how the cells are interacting with the stain. All images of RkpK depleted cells for R20291 and FM2.5 are shown in the Appendices.

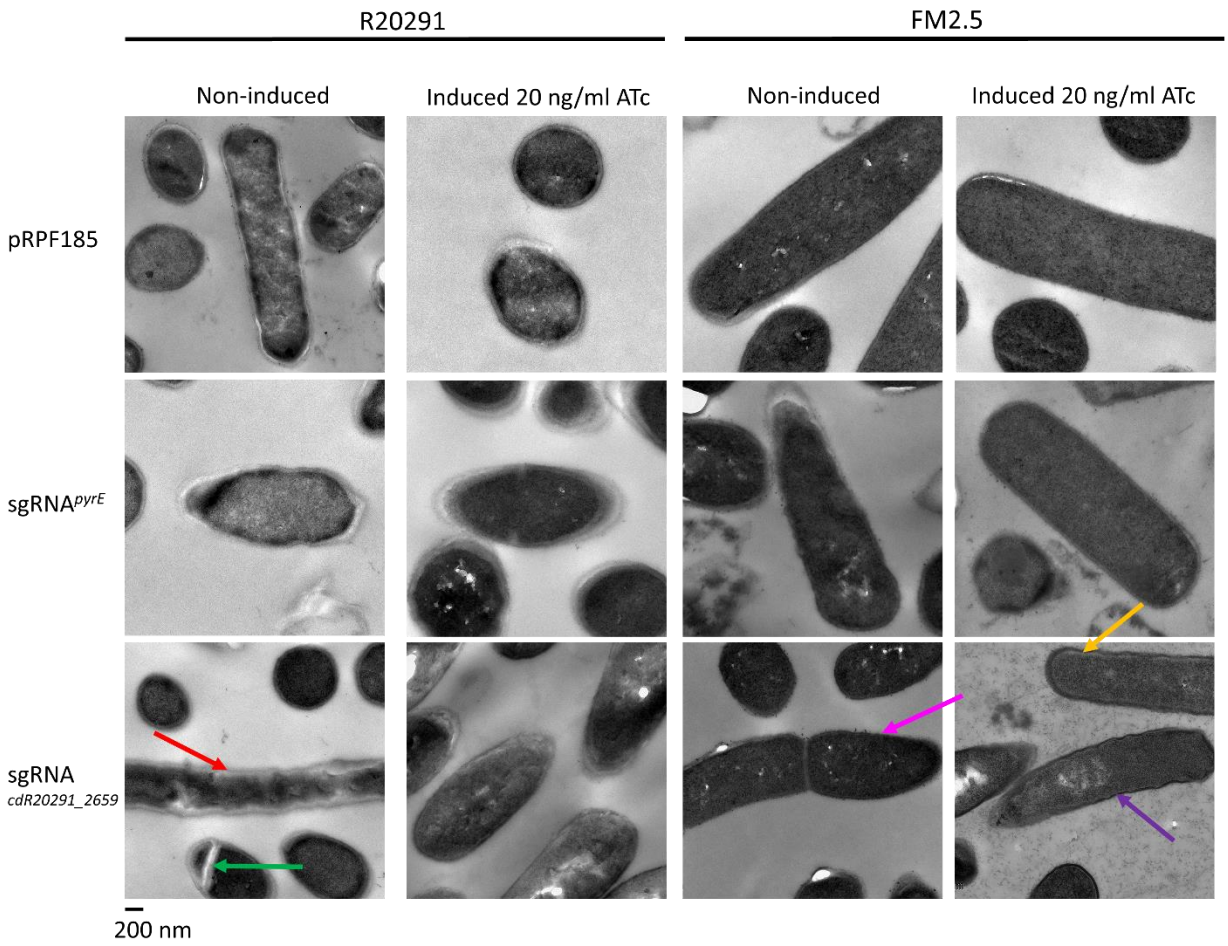



**Fig 6.4. RkpK depletion in a WT and S-layer null mutant.** For R20291 non-induced samples, cell curvature is observed, a phenotype described for cells lacking an S-layer (red arrow). Additionally, mislocalisation of septa is observed (green arrow). RkpK depleted cells in FM2.5 show cell curvature for dividing cells (pink arrow), some wavy morphology (purple arrow) and a thick border around cells (yellow arrow), indicating differences in cell permeability from the WT.

### 6.2.5 TuaG depleted cells of *C. difficile*.

Immediately downstream of *cdR20291_2659,* is *cdR20291_2660* (*tuaG*), the product of which is also believed to function in PSII synthesis. *tuaG* encodes a putative teichuronic acid biosynthesis glycosyltransferase and is conditionally non-essential in the absence of the S-layer. For non-induced R20291/sgRNA$^{cdR20291_2660}$ cells, some misplaced septas are observed along the short cell axis. For FM2.5, cell curvature and lysis is observed. Additionally, TuaG-depleted cells show

mislocalised septa and ‘hairy’-like appendages that protrude from the surface. All images of TuaG depleted cells for R20291 and FM2.5 are shown in the Appendices.

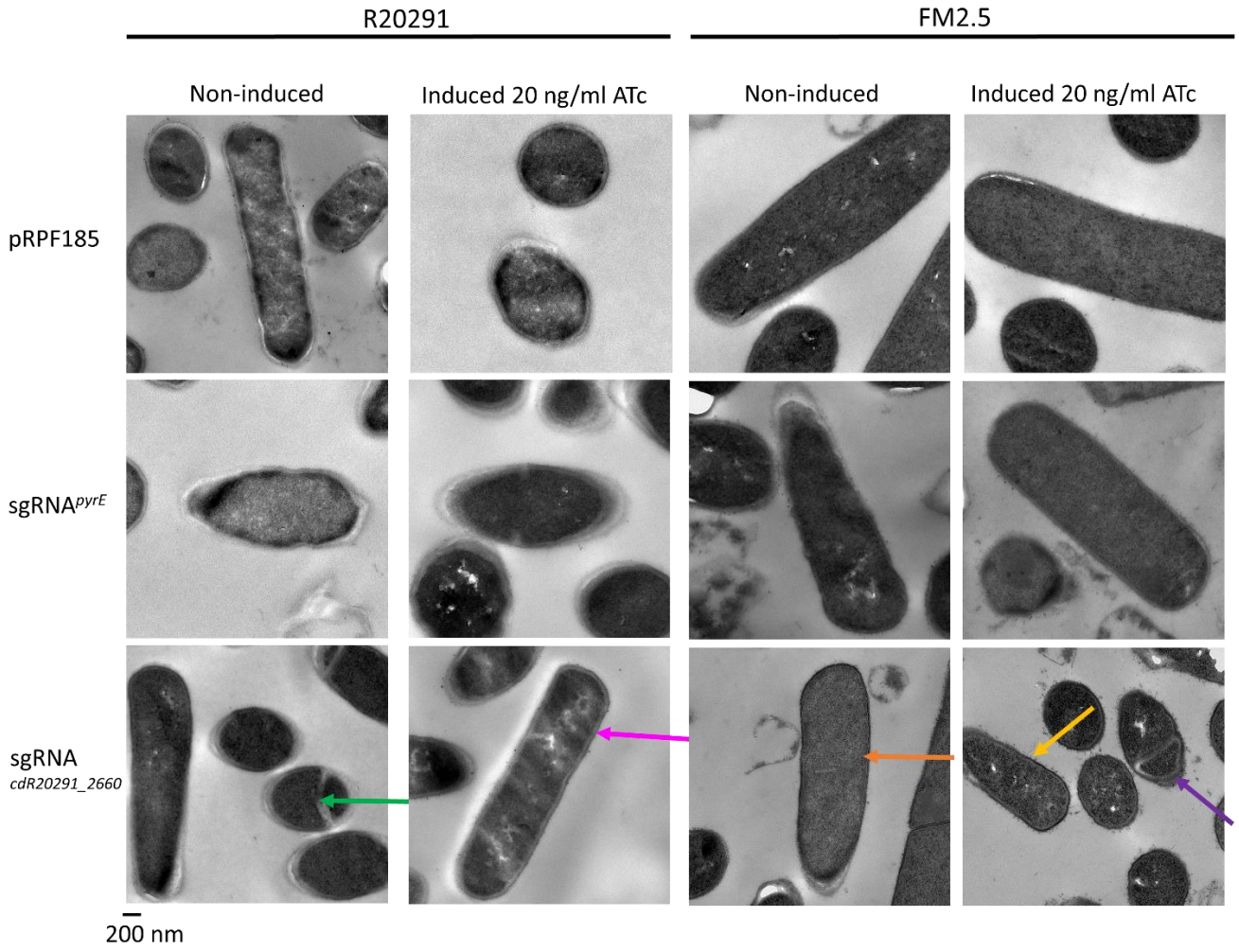



**Fig 6.5. TuaG-depleted cells in a wild-type and S-layer null mutant.** For R20291 non-induced samples, mislocalisation of septa is observed (green arrow). R20291 induced samples appear healthy with rod-shaped cells appearing to have the same width and length as WT cells (pink arrow). However, some appendages protruding from the surface are observed, which could correspond to deposited polysaccharide. For FM2.5, cell curvature (orange arrow) and lysis is observed. TuaG-depleted cells display misolocalised septa (purple arrow) and ‘hairy’-like appendages protruding from the cells surface (yellow arrow).

### 6.2.6 Morphological effects of depleting *cdR20291_2661* from cells.

In R20291, cells depleted of CDR20291_2661 display very irregular cell boundaries. Some cells display a 'wavy' morphology, more dramatic than those observed for other knock-downs, indicating severe envelope stress. For induced samples, a thick border surrounds all cells, with 'hairy'-like appendages protruding from one pole of the cell. Similar to RkpK-depleted cells, these hairy appendages could correspond to deposited cell wall polysaccharide. Some cell lysis is also observed. Similar phenotypes are observed for FM2.5-CDR20291_2661 depleted cells. A thick border that surrounds cells is observed, with some unidentified appendages protruding from the surface. We speculate that the reason for this is that cells depleted of CDR20291_2661 have defects in envelope permeability, taking up the stain for thin-sectioning differently than WT cells. For induced samples, these appendages appear more uniform, completely covering the cell surface.

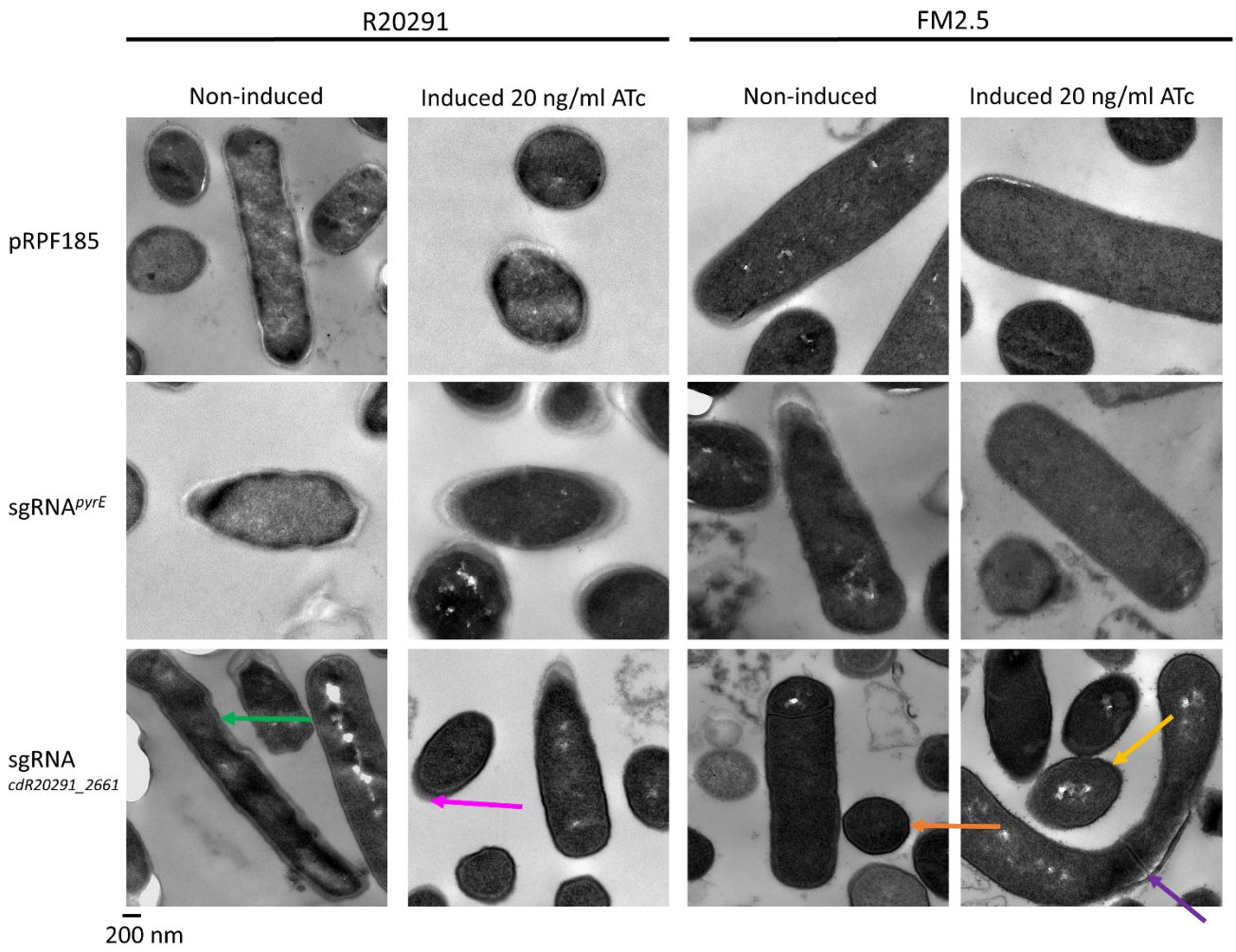



**Fig 6.6. CDR20291_2661-depleted cells in a wild-type and S-layer null mutant.** In R20291, cells depleted of CDR20291_2661 display very irregular cell boundaries, with odd-shaped cells and a 'wavy' morphology (green arrow), indicating severe envelope stress. For induced samples, a thick border surrounds all cells, with 'hairy'-like appendages protruding from one pole of the cell (pink arrow). Some cell lysis is also observed. Similar phenotypes are observed for FM2.5-CDR20291_2661 depleted cells. A thick border that surrounds cells is observed, with some unidentified appendages protruding from the surface (orange arrow). For induced samples, this phenotype appears more dramatic, with appendages completely covering the cell surface (yellow arrow). Additionally, the cell curvature reported for these samples is unlike the curvature previously observed for the S-layer mutant (purple arrow).

## 6.3 Discussion

In this chapter I have demonstrated the utility of a new CRISPRi system, developed for phenotypic studies on essential and non-essential genes in *C. difficile*. Genes shown to be conditionally non-essential in the absence of the S-layer were chosen for analysis as these have the potential to reveal key details of S-layer biogenesis and function while also being potential targets for disruption of S-layer assembly, a marker of pathogenicity *in vivo*. gRNAs were designed for the following conditionally non-essential genes: *cdR20291_2329* (*era*), *cdR20291_2359*, *cdR20291_2659* (*rkpK*), *cdR20291_2660* (*tuaG*) and *cdR20291_2661*.

The *C. difficile* CRISPRi system was used to study the effects of transcriptional silencing on growth rate and thin-sections of cells. Unfortunately, we believe that there is leaky transcription from the $P_{tet}$ promoter as upon induction with ATc, colonies for knock-down were sick and smaller in diameter than the controls, unable to tolerate the same number of cell divisions due to dCas9 blocking RNAP access to the underlying DNA. Even with a *pyrE* knock-down, which does not target the optimal strand for transcription repression, we still observed some growth defects in R20291, and these effects were worse in the S-layer mutant. PyrE is an orotate phosphoribosyltransferase, and loss of this protein results in uracil auxotrophy. Although the dCas9-sgRNA complex binds to the non-template strand, we speculate that due to the size of dCas9, there is some repressive effect, hindering *pyrE* transcription. To test this hypothesis in future, supplementation with uracil should rescue the growth defect observed in R20291 and FM2.5.

Growth rate in liquid media was analysed to allow us to identify the exact point at which protein copies of the target were sufficiently depleted from the population, enabling study of the effects of gene silencing of these targets. This would manifest as a premature plateau in growth rate, which we expected to come before controls entered normal stationary phase. By measuring the effects on growth rate of a *cdR20291_2660* knock-down we observed that *cdR20291_2660* non-induced samples grew similarly to the controls, potentially suggesting that leakiness from the $P_{tet}$

promoter is not happening in all cells at all times. However, future studies with this system should focus on utilising the $P_{xyl}$ inducible system, to eliminate any risk of pre-mature *dcas9* expression. Induction with ATc severely hindered growth of a *cdR20291_2660* knock-down in R20291 and even more so in FM2.5. These findings are substantiated by the zero-growth observed for *cdR20291_2660* knock-downs for FM2.5 on agar. One possibility for this is that PSII is differentially produced in the absence of SlpA, with less required for anchorage of the remaining CWPs. As such, fewer generations might be required to fully deplete these proteins in FM2.5.

Additionally, we did not observe any premature plateau in growth for the *cdR20291_2660* knock-downs, and as such we were unable to determine the exact number of generations required to remove all protein copies of the target from cells. Instead, we diluted stationary phase cultures of R20291 and FM2.5 harbouring all CRISPRi knock-down plasmids into fresh TY with and without inducer. This gave rise to various ranges of final $OD_{600nm}$ values, which were target and strain specific. The lowest $OD_{600nm}$ values obtained for the induced samples, were taken for TEM analysis of thin-sections.

While polyadenylation of RNA 3'-ends was once thought to only occur in eukaryotes, it is now known to play an important role in bacterial RNA metabolism as well (Sarkar, 1997). While RNA 3'-polyadenylation has primarily been described for β and γ- *Proteobacteria*, RNA polyadenylation has also been demonstrated in the *Firmicute, B. subtilis*. The enzyme responsible for polyadenylation is poly(A) polymerase (PAP I), the product of the *pcnB* gene (Liu and Parkinson, 1989). Knowledge of this protein in *C. difficile* is limited, sharing only 47% amino acid sequence with the poly(A) polymerase of *Firmicutes*. Additionally, studies in *E. coli* have shown that addition of the poly(A) tails destabilises RNAs, unlike for eukaryotes, and it is unclear if the function observed in *E. coli* extends to *B. subtilis* or even to *C. difficile*. Also, to date no studies have focussed on the effects of deletion or knock-down of the *pcnB* gene on cell morphology. Here we report that depletion of CDR20291_2359, the *C. difficile* poly(A) polymerase, causes a myriad of morphological defects.

Knock-down of *cdR20291_2359* showed cells with irregular cell boundaries, curvature, misplaced septas. As mentioned previously, we have observed cell curvature for an S-layer mutant, suggesting that depletion of this protein is blocking SlpA from getting to the surface (Dr. Joseph Kirk, personal communication). It is unknown if *slpA* mRNA is polyadenylated, however, in *E. coli* poly(A) adenylation destabilises RNAs. If this protein was to fulfil the same function in *C. difficile* as *E. coli* we speculate that depletion of this protein would give rise to mRNAs that are stable and can be readily translated, which does not appear to be the case. This potentially suggests that this protein does not fulfil the same role in *C. difficile* as *E. coli* and depletion of this protein gives rise to unstable mRNAs, which could be marked for destruction. This potentially has downstream effects such as blocking S-layer from getting to the surface, which could occur at the secretion stage, assembly stage or attachment stage. An interesting line to follow would be to see if depletion of CDR20291_2359 results in accumulation of SlpA in the cytoplasm.

In *B. subtilis* the genes of the *tuaABCDEFGH* operon encode enzymes required for polymerisation of teichuronic acid and synthesis of the precursor, UDP-glucoronate. *tuaD* encodes the UDP-glucose 6-dehydrogenase (Soldo *et al*., 1999). TuaD shares 64% amino acid sequence with RkpK of *C. difficile*. Additionally, *C. difficile* lacks some homologs within the teichuronic acid locus, namely *tuaE* and *tuaF* homologs, which encode proteins involved in repeat unit formation and a protein of unknown function (Soldo *et al.*, 1999). For *C. difficile*, immediately downstream of *rkpK* is *tuaG*, which functions as a putative teichuronic acid biosynthesis glycosyltransferase. This function is also observed for *tuaG* of *B. subtilis* (Soldo *et al.*, 1999). Like RkpK, TuaG of *C. difficile* only shows moderate amino acid similarity to that of *B. anthracis* and *B. cereus*, ~54%. The teichuronic acid operon of *B. subtilis* is part of a Pho regulon and is only produced under conditions of phosphate deprivation, with wall teichoic acids representing the primary cell wall polymer of *B. subtilis* (Devine, 2018). On the other hand, PSII is the primary cell wall polymer for *C. difficile* and is required for attachment of all CWPs to the cell surface. Moreover, the domains responsible for anchorage of CWPs to the cell wall peptidoglycan are different between the two species: SH3 for *B. subtilis* vs.

CWB2 for *C. difficile* (Willing *et al*., 2015). Consequently, it is unlikely that the composition and functionality of these two polymers are similar.

RkpK-depleted cells showed very intriguing phenotypes. Curvature of cells was observed in the WT R20291 strain, which is a phenotype observed for *C. difficile* cells lacking an S-layer. Interestingly, Chu *et. al* reported a curved phenotype for an *mviN* knockdown, encoding the flippase for PSII synthesis. Again, obtaining higher magnification images for these knock-downs would allow us to determine if depletion of RkpK has impacted S-layer assembly, and if PSII has been shed from the cell surface. Additionally, depletion of RkpK from R20291 resulted in mislocalisation of septa, revealing a complex interplay between PSII production, S-layer assembly and cell cycle regulation. Notably, a *lcpB* mutant, encoding a surface anchoring enzyme for PSII attachment to the surface, displays severe growth defects, with cells exhibiting a curved morphology with multiple mislocalised septa. We speculate that depletion of the genes involved in PSII synthesis prevent S-layer attachment to the surface, giving rise to cells with a curved phenotype, which has been reported for FM2.5. Thin-sections of TuaG-depleted cells of R20291 show similar phenotypes to that of RkpK-depleted cells. We observed mislocalised septa as well as irregular cell boundaries, indicating envelope stress in the absence of this protein. Similar to RkpK-depleted cells, we speculate that there is a difference in the cell wall chemical composition, as evidenced by differential cell staining. Like RkpK-depleted cells, hairy-like appendages surround the cell surface of CDR20291_2661-depleted cells. Its plausible that this is deposited cell wall polysaccharide, which is protruding from the cell surface. Although more work is required to specifically identify this appendage. If time had allowed, it would be interesting to study the effects of *cdR20291_2665* depletion. This gene encodes a polysaccharide polymerase, which polymerises PSII hexamers extracellularly. Depletion of CDR20291_2665 could potentially give rise to single hexameric units of PSII, and with this it would be interesting to see if this is sufficient for anchorage of CWPs to the cell surface. Additionally, as several genes involved in PSII synthesis have been identified as conditionally non-essential for growth, one could speculate that the composition of this polymer varies between the WT and S-

layer mutant. Additionally, it would be interesting to explore the effects of GtaB depletion, another conditionally non-essential gene in the absence of SlpA. GtaB, a UTP--glucose-1-phosphate uridylyltransferase is involved in the conversion of glucose-1P to UDP-glucose. Interestingly, the *C. difficile* toxins transfer glucose from UDP-glucose to Rho and Ras GTPases. As such, we may speculate that depletion of GtaB renders TcdA and TcdB non-functional, which could readily be tested. Future experimentation should focus on studying the contribution of individual genes to PSII biosynthesis, and if these genes are differentially expressed between a WT and S-layer null strain.

# Chapter VII. General Discussion

CDI is a potentially fatal intestinal tract infection, accounting for thousands of deaths every year in the UK alone. CDI has a profound impact on the healthcare system and is associated with high economic costs, which is only exacerbated by the high rates of infection recurrence. Despite *C. difficile* being identified in 1935, this pathogen has remained in relative obscurity, with extensive research only being carried out recently. Analysis of the surface of *C. difficile* is likely to reveal key components for pathogenesis, as this is the first point of contact with the host. Studies of the cell envelope could aid in the development of prophylactic agents against this pathogen.

## 7.1 Summary of thesis

The S-layer represents 15% of the total protein content of *C. difficile*, comprising a multifunctional complex required for sporulation, pathogenesis, bacteriophage receptor recognition and resistance to immunity effectors. Production comes at a high metabolic cost to the cell, and this structure is essential for growth *in vitro* (Dembek *et al*., 2015). Exposing *C. difficile* to Avidocin-CDs enabled isolation of spontaneous mutants lacking this layer and isolation of these S-layer mutants afforded us the unique opportunity to dissect the interplay between S-layer biogenesis and *C. difficile* cell biology. Better understanding of essential cellular functions in bacteria is pivotal for the development of more effective antimicrobial agents. Here, we discuss how identifying the conditional genome of *C. difficile* can contribute to our understanding of the pathogen and the potential wider and downstream implications.

In this study, we aimed to define gene essentialities in the absence of SlpA through creation of a large pool of *C. difficile* mutants by insertion mutagenesis (Chapter III). Essential functions, defined as those indispensable for growth and/or survival, are potential targets for new antimicrobial drugs. In this study genes identified as conditionally non-essential in the absence of SlpA could represent targets for prevention of S-layer assembly, a crucial virulence determinant for

CDI. We reported 55 genes that became vital for growth and 20 genes that were dispensable for growth in the absence of SlpA (Chapter IV). The ultimate objective of this work was to study the functions of genes obtained from this initial TraDIS analysis and identify how the functions of these genes are linked to the S-layer and its assembly. However, challenges arise when attempting to unravel essential cellular processes and as such we developed a CRISPRi tool for conditional gene repression of our targets (Chapter V). Lastly, Chapter VI centred around phenotypic analyses of conditionally non-essential genes identified in the TraDIS screens.

## 7.2 The S-layer and cellular metabolism

*C. difficile* adapts to life without an S-layer through metabolic reprogramming of the cell **(Fig 7.1)**. Data from the FM2.5 TraDIS suggest that, in the absence of this surface structure, genes for numerous central metabolic pathways can be rendered dispensable for growth. In fact, metabolic enzymes make up ~ 60% of conditionally non-essential genes and ~ 33% of conditionally essential genes in the absence of SlpA. One such enzyme, encoded by *pfkA*, catalyses phosphorylation of fructose-6-phosphate and is the key-committing step in glycolysis (Phong *et al.*, 2013). In FM2.5, the *pfkA* gene is dispensable for growth. The *C. difficile* genome is predicted to encode three putative PFKs, and PfkA may represent the major PFK, with the other two functioning as minor isoenzymes. This appears to be the case in *E. coli*, which encodes two PFKs, *pfkA* and *pfkB*, the latter accounting for 10% of PFK activity in the cell (Fraenkel, 1986). To see how *C. difficile* adapts without the S-layer, future work should focus on determining if functional redundancy occurs between metabolic genes, and if one of these putative PFKs becomes the sole source of PFK activity in the cell in the absence of the S-layer.

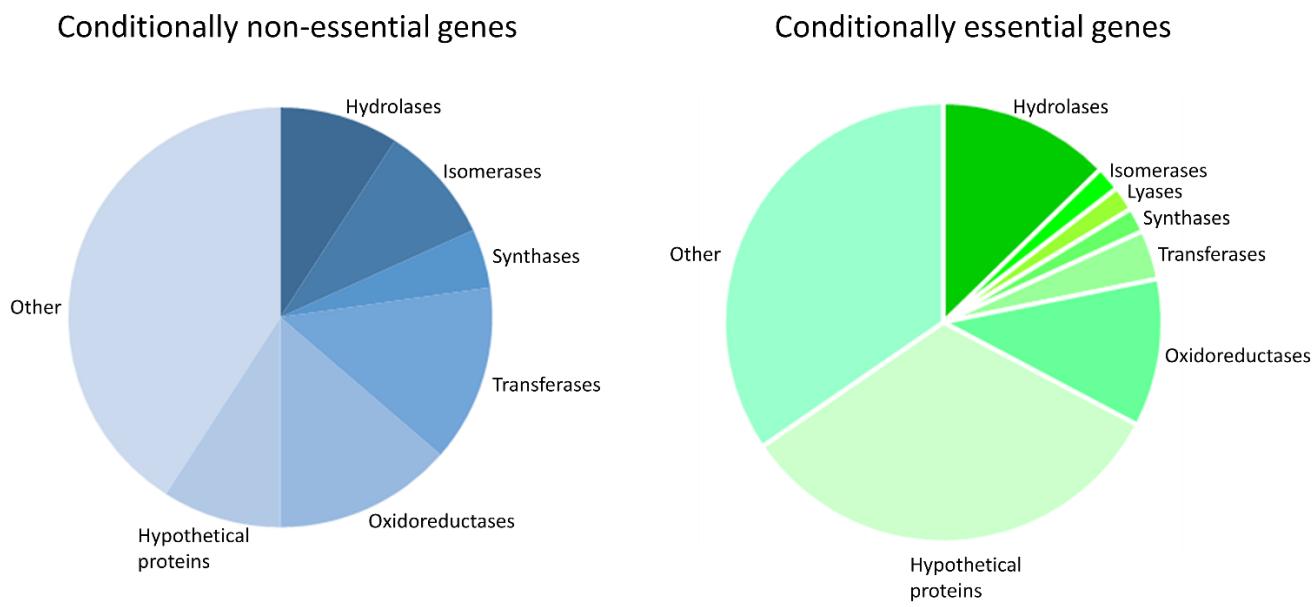



**Fig 7.1 Metabolic reprogramming of *C. difficile* in the absence of an S-layer.** Putative metabolic genes identified from the FM2.5 TraDIS dataset. Of the 20 conditionally non-essential genes, 13 metabolic genes were identified, and of the 55 genes vital for growth in the absence of SlpA, 18 were predicted to encode metabolic enzymes.

## 7.3 S-layer formation

Unlike other bacterial species, the *C. difficile* S-layer provides functions essential for cell viability. The complete layer requires an estimated 590,000 subunits, and to provide this huge amount of protein, *slpA* expression is driven by an extremely powerful promoter (Kirk *et al*., 2017a). When we compare the R20291 and FM2.5 TraDIS datasets, we would expect that the *slpA* gene itself would be identified as conditionally non-essential for growth in the latter, however, this is not the case. There are two potential reasons for this. Firstly, it is possible that *slpA* has never been an essential gene but, due to the high levels of transcription coming from the *slpA* promoter does not tolerate insertions. Alternatively, *slpA* is intrinsically essential even when the S-layer is unable to be assembled on the surface, however, the reason for this remains unclear. This latter possibility seems unlikely but cannot be ruled out. In the absence of SlpA, proteins implicated in S-layer biogenesis have been recognised as dispensable for growth, for example *secA2*, which encodes the ATPase that

targets SlpA for secretion across the membrane. It's possible that this system evolved to cope with the high protein flux across the membrane and without SlpA, *secA2* becomes non-essential. SecA2 is also responsible for CwpV secretion, a CWP which is post-translationally cleaved in a manner analogous to SlpA. In the S-layer deficient strain, CwpV is biased towards the 'ON' orientation. Interestingly, even with this bias towards the 'ON' orientation, *cwpV* is classified as non-essential in both R20291 and FM2.5. Given these changes in CwpV expression it is possible that CwpV could function as a pseudo S-layer. In FM2.5 no secondary suppressor mutations have been identified, however, so changes in CwpV expression are likely due to epigenetic differences, the basis of which is not currently understood. Depletion of SecA2 would allow us to investigate what happens when neither SlpA or CwpV can form an array on the surface and should be a line of enquiry to follow in the future.

### 7.4 S-layer surface attachment

Transposon mutagenesis data from *C. difficile* R20291 and *C. difficile* 630 revealed that all genes thought to be involved in PSII biosynthesis are essential, which is unsurprising as this ligand functions in S-layer anchoring to the cell surface and may well have further functions in the cell envelope. While it is possible that PSII is essential in its own right, for a reason unrelated to the S-layer, TraDIS of FM2.5 revealed several genes believed to be involved in PSII synthesis are dispensable for growth in the absence of SlpA. This lends weight to the hypothesis that PSII essentiality is determined by a functional S-layer. Interestingly, 15 genes of the AP locus are believed to be involved in PSII synthesis, and of these we can only say definitively that two are conditionally non-essential: *cdr20291_2659* and *cdR20291_2660*. We did identify two more genes that were conditionally non-essential in 3 of 4 of our TraDIS datasets. These were *cdR20291_2661* and *cdr20291_2665*. Using CRISPRi, we have successfully been able to deplete CDR20291_2659, CDR20291_2660 and CDR20291_2661 from cells. In R20291, we observed curvature of cells for these knock-downs. Strikingly, this phenotype has been previously reported for FM2.5 and by Chu *et. al*,

who successfully knocked down *mviN* and *pgm2* of the PSII locus. Notably, we observed that the growth defects when knocking down expression of these PSII genes was more severe in FM2.5 than the WT. We speculate that in the absence of SlpA, PSII is not produced to the same levels, as would be required when SlpA is attached to the surface. As such, fewer generations would be required to deplete these proteins from the population, resulting in a more notable growth defect between the WT and S-layer mutant. Future studies should look at the effects depleting these proteins has on the other 28 minor CWPs anchored to the surface via PSII.

As we have now identified some genes that are conditionally non-essential, it is possible that these genes are differentially expressed compared to the WT and that the composition of PSII differs between a WT and S-layer null mutant. Although there may be a core set of genes from the AP locus required to produce a polymer that can still be recognised by the CWPs for attachment, the molecular basis of this interaction is still not fully characterised. Indeed, many of the PSII biosynthetic proteins are still essential in the absence of SlpA, including the mannose-1-phosphateguanylyltransferase, ManC and the phosphomannomutase, Pgm2, responsible for completing PSII synthesis, and the flippase for required for transporting PSII across the membrane. As such, we could speculate that a slightly different polymer is still produced by the S-layer mutant and transported to the cell surface. Additionally, *cdR20291_2665*, encoding the PSII polymerase could be dispensable for growth in the absence of SlpA. Without this gene, only single units of PSII could be produced, and by using CRISPRi we could determine if these single hexameric units are sufficient for CWP anchorage.

## 7.5 Limitations

### Transposon mutagenesis of *C. difficile*

While TraDIS is a powerful tool, this study highlights the importance of obtaining a sufficiently saturated library of mutants, as the library obtained for the S-layer mutant, FM2.5 was less than that of the published R20291 dataset, which had better resolution. With higher resolution

we could discern the essentiality of smaller genes and subject the library to various *in vitro* and *in vivo* conditions. It is conceivable that in time, further improvements could be made to the system, allowing more refined analysis and statistical power. One possible approach, as mentioned previously, would involve generating the library in liquid media rather than on agar, allowing us to generate a much larger number of mutants. Applying the methodology developed in Chapter III, library construction in liquid media should be revisited in the near future. Additionally, as the S-layer null mutant uncovered important roles for SlpA in resistance to lysozyme and LL-37, future experiments should focus on passaging R20291 and FM2.5 libraries through increasing concentrations of these immunity effectors, to identify any additional genes contributing to resistance.

### CRISPRi for *C. difficile*

The development of CRISPRi technologies has overcome limitations associated with studies of essential genes in *C. difficile*. Future developments will improve our system for targeted gene repression. Unfortunately, for our system, d*cas9* expression is under control of a $P_{tet}$ promoter, which is leaky and additionally, minor toxicity of the inducer, ATc, has been observed. As CRISPRi is highly efficient in bacterial cells, leaky expression of the dCas9 protein from an inducible promoter can lead to ~ 80% repression level. We see this for all of our non-induced samples, whereby cell morphologies differ markedly from the strains harbouring the control plasmids. Instead, d*cas9* should instead be placed under the control of the more tightly repressed xylose promoter, allowing control of timing and the extent of gene silencing (Muh *et al*., 2019).

## 7.6 Concluding Remarks

This thesis outlines a project to develop tools for studying the conditional genome of an S-layer null strain of *C. difficile*. In this project we overcame several challenges associated with transposon library generation, and from this we were able to identify the conditional genome of *C. difficile* in the absence of the S-layer, most notably identifying the gene required for SlpA secretion,

and several PSII biosynthesis genes required for SlpA attachment as dispensable for growth. As we continually improve our method for library construction, we are likely to find other genes as dispensable or vital for growth in an S-layer null mutant, and we should be able to ascertain the essentiality of genes identified as ambiguous in this study. Moreover, with the development of a CRISPRi tool for conditional gene repression, we are likely to uncover the functions of the genes identified in this study and how they relate to the S-layer and its assembly. Future studies should focus on improving the tools developed in this thesis for *C. difficile*, and we hope that this will aid in the development of novel therapeutics against this pathogen.

# References


Adams, C.M., Eckenroth, B.E., Putnam, E.E., Doublie, S., and Shen, A. (2013). Structural and functional analysis of the CspB protease required for *Clostridium* spore germination. PLoS Pathog *9*, e1003165.

Ajouz, H., Mukherji, D., and Shamseddine, A. (2014). Secondary bile acids: an underrecognized cause of colon cancer. World J Surg Oncol *12*, 164.

Al-Hinai, M.A., Jones, S.W., and Papoutsakis, E.T. (2015). The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. Microbiol Mol Biol Rev *79*, 19-37.

Alang, N., and Kelly, C.R. (2015). Weight gain after fecal microbiota transplantation. Open Forum Infect Dis *2*, ofv004.

Allen, N.E., Hobbs, J.N., Jr., and Nicas, T.I. (1996). Inhibition of peptidoglycan biosynthesis in vancomycin-susceptible and -resistant bacteria by a semisynthetic glycopeptide antibiotic. Antimicrob Agents Chemother *40*, 2356-2362.

Almagro Armenteros, J.J., Tsirigos, K.D., Sonderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol *37*, 420-423.

Andersen, J.M., Shoup, M., Robinson, C., Britton, R., Olsen, K.E., and Barrangou, R. (2016). CRISPR diversity and microevolution in *Clostridium difficile*. Genome Biol Evol *8*, 2841-2855.

Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol *38*, 824-844.

Arroyo-Olarte, R.D., Bravo Rodriguez, R., and Morales-Rios, E. (2021). Genome editing in bacteria: CRISPR-Cas and beyond. Microorganisms *9*.

Artiguenave, F., Vilagines, R., and Danglot, C. (1997). High-efficiency transposon mutagenesis by electroporation of a *Pseudomonas fluorescens* strain. FEMS Microbiol Lett *153*, 363-369.

Atrih, A., Bacher, G., Allmaier, G., Williamson, M.P., and Foster, S.J. (1999). Analysis of peptidoglycan structure from vegetative cells of *Bacillus subtilis* 168 and role of PBP 5 in peptidoglycan maturation. J Bacteriol *181*, 3956-3966.

Awad, M.M., Johanesen, P.A., Carter, G.P., Rose, E., and Lyras, D. (2014). *Clostridium difficile* virulence factors: Insights into an anaerobic spore-forming pathogen. Gut Microbes *5*, 579-593.

Bachman, M.A., Breen, P., Deornellas, V., Mu, Q., Zhao, L., Wu, W., Cavalcoli, J.D., and Mobley, H.L. (2015). Genome-wide identification of *Klebsiella pneumoniae* fitness genes during lung infection. mBio *6*, e00775.

Baines, S.D., O'Connor, R., Freeman, J., Fawley, W.N., Harmanus, C., Mastrantonio, P., Kuijper, E.J., and Wilcox, M.H. (2008). Emergence of reduced susceptibility to metronidazole in *Clostridium difficile*. J Antimicrob Chemother *62*, 1046-1052.

Bakken, J.S., Borody, T., Brandt, L.J., Brill, J.V., Demarco, D.C., Franzos, M.A., Kelly, C., Khoruts, A., Louie, T., Martinelli, L.P*., et al.* (2011). Treating *Clostridium difficile* infection with fecal microbiota transplantation. Clin Gastroenterol Hepatol *9*, 1044-1049.

Banaei, N., Anikst, V., and Schroeder, L.F. (2015). Burden of *Clostridium difficile* infection in the United States. N Engl J Med *372*, 2368-2369.

Barquist, L., Boinett, C.J., and Cain, A.K. (2013). Approaches to querying bacterial genomes with transposon-insertion sequencing. RNA Biol *10*, 1161-1169.

Barra-Carrasco, J., Olguin-Araneda, V., Plaza-Garrido, A., Miranda-Cardenas, C., Cofre-Araneda, G., Pizarro-Guajardo, M., Sarker, M.R., and Paredes-Sabja, D. (2013). The *Clostridium difficile* exosporium cysteine (CdeC)-rich protein is required for exosporium morphogenesis and coat assembly. J Bacteriol *195*, 3863-3875.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Barrangou, R., and Marraffini, L.A. (2014). CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. Mol Cell *54*, 234-244.

Bartlett, J.G., and Gerding, D.N. (2008). Clinical recognition and diagnosis of *Clostridium difficile* infection. Clin Infect Dis *46 Suppl 1*, S12-18.

Bartlett, J.G., Moon, N., Chang, T.W., Taylor, N., and Onderdonk, A.B. (1978). Role of *Clostridium difficile* in antibiotic-associated pseudomembranous colitis. Gastroenterology *75*, 778-782.

Berg, D.E., Davies, J., Allet, B., and Rochaix, J.D. (1975). Transposition of R factor genes to bacteriophage lambda. Proc Natl Acad Sci U S A *72*, 3628-3632.

Bertolo, L., Boncheff, A.G., Ma, Z., Chen, Y.H., Wakeford, T., Friendship, R.M., Rosseau, J., Weese, J.S., Chu, M., Mallozzi, M*., et al.* (2012). *Clostridium difficile* carbohydrates: glucan in spores, PSII common antigen in cells, immunogenicity of PSII in swine and synthesis of a dual *C. difficile*-ETEC conjugate vaccine. Carbohydr Res *354*, 79-86.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res *41*, 7429-7437.

Blomfield, I.C., Vaughn, V., Rest, R.F., and Eisenstein, B.I. (1991). Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. Mol Microbiol *5*, 1447-1457.

Boudry, P., Semenova, E., Monot, M., Datsenko, K.A., Lopatina, A., Sekulovic, O., Ospina-Bedoya, M., Fortier, L.C., Severinov, K., Dupuy, B*., et al.* (2015). Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. mBio *6*, e01112-01115.

Bradshaw, W.J., Kirby, J.M., Roberts, A.K., Shone, C.C., and Acharya, K.R. (2017). Cwp2 from *Clostridium difficile* exhibits an extended three domain fold and cell adhesion *in vitro*. FEBS J *284*, 2886-2898.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Brown, S., Santa Maria, J.P., Jr., and Walker, S. (2013). Wall teichoic acids of Gram-positive bacteria. Annu Rev Microbiol *67*, 313-336.

Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. Nat Rev Microbiol *2*, 57-65.

Browning, D.F., and Busby, S.J. (2016). Local and global regulation of transcription initiation in bacteria. Nat Rev Microbiol *14*, 638-650.

Cain, A.K., Barquist, L., Goodman, A.L., Paulsen, I.T., Parkhill, J., and van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. Nat Rev Genet *21*, 526-540.

Calabi, E., Calabi, F., Phillips, A.D., and Fairweather, N.F. (2002). Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. Infect Immun *70*, 5770-5778.

Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G., and Fairweather, N. (2001). Molecular characterization of the surface layer proteins from *Clostridium difficile*. Mol Microbiol *40*, 1187-1199.

Cao, G.J., and Sarkar, N. (1992). Identification of the gene for an *Escherichia coli* poly(A) polymerase. Proc Natl Acad Sci U S A *89*, 10380-10384.

Carter, G.P., Douce, G.R., Govind, R., Howarth, P.M., Mackin, K.E., Spencer, J., Buckley, A.M., Antunes, A., Kotsanas, D., Jenkin, G.A*., et al.* (2011). The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of *Clostridium difficile*. PLoS Pathog *7*, e1002317.

Carter, J. (2009). Deaths involving *Clostridium difficile*: England and Wales, 2008. Health Stat Q, 43-47.

Cartman, S.T., Kelly, M.L., Heeg, D., Heap, J.T., and Minton, N.P. (2012). Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype and toxin production. Appl Environ Microbiol *78*, 4683-4690.

Cartman, S.T., and Minton, N.P. (2010). A mariner-based transposon system for in vivo random mutagenesis of *Clostridium difficile*. Appl Environ Microbiol *76*, 1103-1109.

Chandrasekaran, R., and Lacy, D.B. (2017). The role of toxins in *Clostridium difficile* infection. FEMS Microbiol Rev *41*, 723-750.

Chao, M.C., Abel, S., Davis, B.M., and Waldor, M.K. (2016). The design and analysis of transposon insertion sequencing experiments. Nat Rev Microbiol *14*, 119-128.

Charbonneau, A.R.L., Forman, O.P., Cain, A.K., Newland, G., Robinson, C., Boursnell, M., Parkhill, J., Leigh, J.A., Maskell, D.J., and Waller, A.S. (2017). Defining the ABC of gene essentiality in streptococci. BMC Genomics *18*, 426.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J.*, et al.* (2016). Comparison of Cas9 activators in multiple species. Nat Methods *13*, 563-567.

Chen, S., Sun, C., Wang, H., and Wang, J. (2015). The role of Rho GTPases in toxicity of *Clostridium difficile* toxins. Toxins (Basel) *7*, 5254-5267.

Childress, K.O., Edwards, A.N., Nawrocki, K.L., Anderson, S.E., Woods, E.C., and McBride, S.M. (2016). The phosphotransfer protein CD1492 represses sporulation initiation in *Clostridium difficile*. Infect Immun *84*, 3434-3444.

Chiu, C.Y., Sarwal, A., Feinstein, A., and Hennessey, K. (2019). Effective dosage of oral vancomycin in treatment for initial episode of *Clostridioides difficile* infection: A systematic review and meta-analysis. Antibiotics (Basel) *8*.

Cho, J.M., Pardi, D.S., and Khanna, S. (2020). Update on treatment of *Clostridioides difficile* infection. Mayo Clin Proc *95*, 758-769.

Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Coller, J.A., Fero, M.J., McAdams, H.H., and Shapiro, L. (2011). The essential genome of a bacterium. Mol Syst Biol *7*, 528.

Chu, M., Mallozzi, M.J., Roxas, B.P., Bertolo, L., Monteiro, M.A., Agellon, A., Viswanathan, V.K., and Vedantam, G. (2016). A *Clostridium difficile* cell wall glycopolymer locus influences bacterial shape, polysaccharide production and virulence. PLoS Pathog *12*, e1005946.

Collias, D., and Beisel, C.L. (2021). CRISPR technologies and the search for the PAM-free nuclease. Nat Commun *12*, 555.

Coullon, H., Rifflet, A., Wheeler, R., Janoir, C., Boneca, I.G., and Candela, T. (2018). N-Deacetylases required for muramic-delta-lactam production are involved in *Clostridium difficile* sporulation, germination, and heat resistance. J Biol Chem *293*, 18040-18054.

Crawshaw, A.D., Serrano, M., Stanley, W.A., Henriques, A.O., and Salgado, P.S. (2014). A mother cell-to-forespore channel: current understanding and future challenges. FEMS Microbiol Lett *358*, 129-136.

Crobach, M.J., Planche, T., Eckert, C., Barbut, F., Terveer, E.M., Dekkers, O.M., Wilcox, M.H., and Kuijper, E.J. (2016). European Society of Clinical Microbiology and Infectious Diseases: update of the diagnostic guidance document for *Clostridium difficile* infection. Clin Microbiol Infect *22 Suppl 4*, S63-81.

Czepiel, J., Drozdz, M., Pituch, H., Kuijper, E.J., Perucki, W., Mielimonka, A., Goldman, S., Wultanska, D., Garlicki, A., and Biesiada, G. (2019). *Clostridium difficile* infection: review. Eur J Clin Microbiol Infect Dis *38*, 1211-1221.

Dang, T.H., de la Riva, L., Fagan, R.P., Storck, E.M., Heal, W.P., Janoir, C., Fairweather, N.F., and Tate, E.W. (2010). Chemical probes of surface layer biogenesis in *Clostridium difficile*. ACS Chem Biol *5*, 279-285.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun *3*, 945.

de la Riva, L., Willing, S.E., Tate, E.W., and Fairweather, N.F. (2011). Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of *Clostridium difficile*. J Bacteriol *193*, 3276-3285.

Deakin, L.J., Clare, S., Fagan, R.P., Dawson, L.F., Pickard, D.J., West, M.R., Wren, B.W., Fairweather, N.F., Dougan, G., and Lawley, T.D. (2012). The *Clostridium difficile spo0A* gene is a persistence and transmission factor. Infect Immun *80*, 2704-2711.

DeFilipp, Z., Bloom, P.P., Torres Soto, M., Mansour, M.K., Sater, M.R.A., Huntley, M.H., Turbett, S., Chung, R.T., Chen, Y.B., and Hohmann, E.L. (2019). Drug-resistant *E. coli* bacteremia transmitted by fecal microbiota transplant. N Engl J Med *381*, 2043-2050.

Dembek, M., Barquist, L., Boinett, C.J., Cain, A.K., Mayho, M., Lawley, T.D., Fairweather, N.F., and Fagan, R.P. (2015). High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. MBio *6*, e02383.

Dembek, M., Kelly, A., Barwinska-Sendra, A., Tarrant, E., Stanley, W.A., Vollmer, D., Biboy, J., Gray, J., Vollmer, W., and Salgado, P.S. (2018). Peptidoglycan degradation machinery in *Clostridium difficile* forespore engulfment. Mol Microbiol *110*, 390-410.

Dembek, M., Reynolds, C.B., and Fairweather, N.F. (2012). *Clostridium difficile* cell wall protein CwpV undergoes enzyme-independent intramolecular autoproteolysis. J Biol Chem *287*, 1538-1544.

Devine, K.M. (2018). Activation of the PhoPR-mediated response to phosphate limitation Is regulated by wall teichoic acid metabolism in *Bacillus subtilis*. Front Microbiol *9*, 2678.

Di Bella, S., Ascenzi, P., Siarakas, S., Petrosillo, N., and di Masi, A. (2016). *Clostridium difficile* toxins A and B: Insights into pathogenic properties and extraintestinal effects. Toxins (Basel) *8*.

Diaz, O.R., Sayer, C.V., Popham, D.L., and Shen, A. (2018). *Clostridium difficile* lipoprotein GerS is required for cortex modification and thus spore germination. mSphere *3*.

Doublet, P., van Heijenoort, J., Bohin, J.P., and Mengin-Lecreulx, D. (1993). The *murI* gene of *Escherichia coli* is an essential gene that encodes a glutamate racemase activity. J Bacteriol *175*, 2970-2979.

Dupuy, B., Govind, R., Antunes, A., and Matamouros, S. (2008). *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. J Med Microbiol *57*, 685-689.

Duvall, J.R., Bedard, L., Naylor-Olsen, A.M., Manson, A.L., Bittker, J.A., Sun, W., Fitzgerald, M.E., He, Z., Lee, M.D.t., Marie, J.C*., et al.* (2017). Identification of highly specific diversity-oriented synthesis-derived inhibitors of *Clostridium difficile*. ACS Infect Dis *3*, 349-359.

Eiseman, B., Silen, W., Bascom, G.S., and Kauvar, A.J. (1958). Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. Surgery *44*, 854-859.

El Meouche, I., Peltier, J., Monot, M., Soutourina, O., Pestel-Caron, M., Dupuy, B., and Pons, J.L. (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.

Emerson, J.E., Reynolds, C.B., Fagan, R.P., Shaw, H.A., Goulding, D., and Fairweather, N.F. (2009). A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. Mol Microbiol *74*, 541-556.

Eze, P., Balsells, E., Kyaw, M.H., and Nair, H. (2017). Risk factors for *Clostridium difficile* infections - an overview of the evidence base and challenges in data synthesis. J Glob Health *7*, 010417.
Fagan, R.P., Albesa-Jove, D., Qazi, O., Svergun, D.I., Brown, K.A., and Fairweather, N.F. (2009). Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. Mol Microbiol *71*, 1308-1322.

Fagan, R.P., and Fairweather, N.F. (2011). *Clostridium difficile* has two parallel and essential Sec secretion systems. J Biol Chem *286*, 27483-27493.

Fagan, R.P., and Fairweather, N.F. (2014). Biogenesis and functions of bacterial S-layers. Nat Rev Microbiol *12*, 211-222.

Fekety, R., Silva, J., Kauffman, C., Buggy, B., and Deery, H.G. (1989). Treatment of antibiotic-associated *Clostridium difficile* colitis with oral vancomycin: comparison of two dosage regimens. Am J Med *86*, 15-19.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J.*, et al.* (2014). Pfam: the protein families database. Nucleic Acids Res *42*, D222-230.

Finnegan, D.J. (1992). Transposable elements. Curr Opin Genet Dev *2*, 861-867.

Fontana, J., Dong, C., Kiattisewee, C., Chavali, V.P., Tickman, B.I., Carothers, J.M., and Zalatan, J.G. (2020). Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. Nat Commun *11*, 1618.

Fraenkel, D.G. (1986). Mutants in glucose metabolism. Annu Rev Biochem *55*, 317-337.

Francis, M.B., Allen, C.A., Shrestha, R., and Sorg, J.A. (2013). Bile acid recognition by the *Clostridium difficile* germinant receptor, CspC, is important for establishing infection. PLoS Pathog *9*, e1003356.

Francis, M.B., Allen, C.A., and Sorg, J.A. (2015). Spore cortex hydrolysis precedes dipicolinic acid release during *Clostridium difficile* spore germination. J Bacteriol *197*, 2276-2283.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol *31*, 397-405.

Ganeshapillai, J., Vinogradov, E., Rousseau, J., Weese, J.S., and Monteiro, M.A. (2008). *Clostridium difficile* cell-surface polysaccharides composed of pentaglycosyl and hexaglycosyl phosphate repeating units. Carbohydr Res *343*, 703-710.

Garrett, S.C. (2021). Pruning and tending immune memories: Spacer dynamics in the CRISPR array. Front Microbiol *12*, 664299.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A *109*, E2579-2586.

Gawronski, J.D., Wong, S.M., Giannoukos, G., Ward, D.V., and Akerley, B.J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proc Natl Acad Sci U S A *106*, 16422-16427.

Gebhart, D., Williams, S.R., Bishop-Lilly, K.A., Govoni, G.R., Willner, K.M., Butani, A., Sozhamannan, S., Martin, D., Fortier, L.C., and Scholl, D. (2012). Novel high-molecular-weight, R-type bacteriocins of *Clostridium difficile*. J Bacteriol *194*, 6240-6247.

Gleditzsch, D., Pausch, P., Muller-Esparza, H., Ozcan, A., Guo, X., Bange, G., and Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA Biol *16*, 504-517.

Goodall, E.C.A., Robinson, A., Johnston, I.G., Jabbari, S., Turner, K.A., Cunningham, A.F., Lund, P.A., Cole, J.A., and Henderson, I.R. (2018). The Essential Genome of *Escherichia coli* K-12. mBio *9*.

Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe *6*, 279-289.

Goodman, A.L., Wu, M., and Gordon, J.I. (2011). Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. Nat Protoc *6*, 1969-1980.

Goorhuis, A., Bakker, D., Corver, J., Debast, S.B., Harmanus, C., Notermans, D.W., Bergwerff, A.A., Dekker, F.W., and Kuijper, E.J. (2008). Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clin Infect Dis *47*, 1162-1170.

Govind, R., and Dupuy, B. (2012). Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. PLoS Pathog *8*, e1002727.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 172.

Grundling, A., and Schneewind, O. (2007). Genes required for glycolipid synthesis and lipoteichoic acid anchoring in *Staphylococcus aureus*. J Bacteriol *189*, 2521-2530.

Gupta, S., Allen-Vercoe, E., and Petrof, E.O. (2016). Fecal microbiota transplantation: in perspective. Therap Adv Gastroenterol *9*, 229-239.

Hall, A. (2012). Rho family GTPases. Biochem Soc Trans *40*, 1378-1382.

Hall, I.C., and O'Toole, E. (1935). Intestinal flora in new-born infants: with a description of a new pathogenic anaerobe, *Bacillus difficilis*. American Journal of Diseases of Children *49*, 390-402.

Hamer, L., DeZwaan, T.M., Montenegro-Chamorro, M.V., Frank, S.A., and Hamer, J.E. (2001). Recent advances in large-scale transposon mutagenesis. Curr Opin Chem Biol *5*, 67-73.

Hammond, G.A., and Johnson, J.L. (1995). The toxigenic element of *Clostridium difficile* strain VPI 10463. Microb Pathog *19*, 203-213.

Hawkins, J.S., Wong, S., Peters, J.M., Almeida, R., and Qi, L.S. (2015). Targeted transcriptional repression in bacteria using CRISPR Interference (CRISPRi). Methods Mol Biol *1311*, 349-362.

He, L., St John James, M., Radovcic, M., Ivancic-Bace, I., and Bolt, E.L. (2020). Cas3 Protein-A review of a multi-tasking machine. Genes (Basel) *11*.
Heap, J.T., Kuehne, S.A., Ehsaan, M., Cartman, S.T., Cooksley, C.M., Scott, J.C., and Minton, N.P. (2010). The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. J Microbiol Methods *80*, 49-55.

Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P., and Minton, N.P. (2007). The ClosTron: a universal gene knock-out system for the genus *Clostridium*. J Microbiol Methods *70*, 452-464.

Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E., and Holden, D.W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. Science *269*, 400-403.

Hibbing, M.E., Fuqua, C., Parsek, M.R., and Peterson, S.B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol *8*, 15-25.

Ho, H.I., Fang, J.R., Cheung, J., and Wang, H.H. (2020). Programmable CRISPR-Cas transcriptional activation in bacteria. Mol Syst Biol *16*, e9427.

Houwink, A.L. (1953). A macromolecular mono-layer in the cell wall of *Spirillum spec*. Biochim Biophys Acta *10*, 360-366.

Hryckowian, A.J., Pruss, K.M., and Sonnenburg, J.L. (2017). The emerging metabolic view of *Clostridium difficile* pathogenesis. Curr Opin Microbiol *35*, 42-47.

Huang, M., and Hull, C.M. (2017). Sporulation: how to survive on planet Earth (and beyond). Curr Genet *63*, 831-838.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J Bacteriol *169*, 5429-5433.

Janoir, C., Pechine, S., Grosdidier, C., and Collignon, A. (2007). Cwp84, a surface-associated protein of *Clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins. J Bacteriol *189*, 7174-7180.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol *31*, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Res *36*, W5-9.

Johnson, S. (2009). Recurrent *Clostridium difficile* infection: a review of risk factors, treatments, and outcomes. J Infect *58*, 403-410.

Kalapos, M.P., Cao, G.J., Kushner, S.R., and Sarkar, N. (1994). Identification of a second poly(A) polymerase in *Escherichia coli*. Biochem Biophys Res Commun *198*, 459-465.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

Kawata, T., Takeoka, A., Takumi, K., and Masuda, K. (1984). Demonstration and preliminary characterization of a regular array in the cell wall of *Clostridium difficile*. FEMS Microbiology Letters *24*, 323-328.

Keighley, M.R., Burdon, D.W., Arabi, Y., Williams, J.A., Thompson, H., Youngs, D., Johnson, M., Bentley, S., George, R.H., and Mogg, G.A. (1978). Randomised controlled trial of vancomycin for pseudomembranous colitis and postoperative diarrhoea. Br Med J *2*, 1667-1669.

Kieper, S.N., Almendros, C., and Brouns, S.J.J. (2019). Conserved motifs in the CRISPR leader sequence control spacer acquisition levels in Type I-D CRISPR-Cas systems. FEMS Microbiol Lett *366*.

Kirby, J.M., Ahern, H., Roberts, A.K., Kumar, V., Freeman, Z., Acharya, K.R., and Shone, C.C. (2009). Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile*. J Biol Chem *284*, 34666-34673.

Kirk, J.A., Banerji, O., and Fagan, R.P. (2017a). Characteristics of the *Clostridium difficile* cell envelope and its importance in therapeutics. Microb Biotechnol *10*, 76-90.

Kirk, J.A., Gebhart, D., Buckley, A.M., Lok, S., Scholl, D., Douce, G.R., Govoni, G.R., and Fagan, R.P. (2017b). New class of precision antimicrobials redefines role of *Clostridium difficile* S-layer in virulence and viability. Sci Transl Med *9*.

Kleckner, N., Chan, R.K., Tye, B.K., and Botstein, D. (1975). Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. J Mol Biol *97*, 561-575.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol *305*, 567-580.

Krupovic, M., Vilen, H., Bamford, J.K., Kivela, H.M., Aalto, J.M., Savilahti, H., and Bamford, D.H. (2006). Genome characterization of lipid-containing marine bacteriophage PM2 by transposon insertion mutagenesis. J Virol *80*, 9270-9278.

Kuroda, A., Rashid, M.H., and Sekiguchi, J. (1992). Molecular cloning and sequencing of the upstream region of the major *Bacillus subtilis* autolysin gene: a modifier protein exhibiting sequence homology to the major autolysin and the *spoIID* product. J Gen Microbiol *138*, 1067-1076.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

LaFrance, M.E., Farrow, M.A., Chandrasekaran, R., Sheng, J., Rubin, D.H., and Lacy, D.B. (2015). Identification of an epithelial cell receptor responsible for *Clostridium difficile* TcdB-induced cytotoxicity. Proc Natl Acad Sci U S A *112*, 7073-7078.

Langridge, G.C., Phan, M.D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G*., et al.* (2009). Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. Genome Res *19*, 2308-2316.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc *8*, 2180-2196.

Lawler, A.J., Lambert, P.A., and Worthington, T. (2020). A revised understanding of *Clostridioides difficile* spore germination. Trends Microbiol *28*, 744-752.

Le Rhun, A., Escalera-Maurer, A., Bratovic, M., and Charpentier, E. (2019). CRISPR-Cas in *Streptococcus pyogenes*. RNA Biol *16*, 380-389.

Leffler, D.A., and Lamont, J.T. (2015). *Clostridium difficile* Infection. N Engl J Med *373*, 287-288.

Liu, J.D., and Parkinson, J.S. (1989). Genetics and sequence analysis of the *pcnB* locus, an *Escherichia coli* gene involved in plasmid copy number control. J Bacteriol *171*, 1254-1261.

Liu, Y., Wan, X., and Wang, B. (2019). Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. Nat Commun *10*, 3693.

Louie, T.J., Cannon, K., Byrne, B., Emery, J., Ward, L., Eyben, M., and Krulicki, W. (2012). Fidaxomicin preserves the intestinal microbiome during and after treatment of *Clostridium difficile* infection (CDI) and reduces both toxin reexpression and recurrence of CDI. Clin Infect Dis *55 Suppl 2*, S132-142.

Lundqvist, T., Fisher, S.L., Kern, G., Folmer, R.H., Xue, Y., Newton, D.T., Keating, T.A., Alm, R.A., and de Jonge, B.L. (2007). Exploitation of structural and regulatory diversity in glutamate racemases. Nature *447*, 817-822.

Ma, Z., Zhang, G.L., Gadi, M.R., Guo, Y., Wang, P., and Li, L. (2020). *Clostridioides difficile cd2775* encodes a unique mannosyl-1-phosphotransferase for polysaccharide II biosynthesis. ACS Infect Dis *6*, 680-686.

Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet *11*, 181-190.

Martin-Verstraete, I., Peltier, J., and Dupuy, B. (2016). The regulatory networks that control *Clostridium difficile* toxin synthesis. Toxins (Basel) *8*.

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

McClintock (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

McDonald, L.C., Gerding, D.N., Johnson, S., Bakken, J.S., Carroll, K.C., Coffin, S.E., Dubberke, E.R., Garey, K.W., Gould, C.V., Kelly, C*., et al.* (2018). 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). Clin Infect Dis *66*, e1-e48.

McGinn, J., and Marraffini, L.A. (2016). CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. Mol Cell *64*, 616-623.

McKenney, P.T., Driks, A., and Eichenberger, P. (2013). The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. Nat Rev Microbiol *11*, 33-44.

Mills, D.A., Manias, D.A., McKay, L.L., and Dunny, G.M. (1997). Homing of a group II intron from *Lactococcus lactis* subsp. lactis ML3. J Bacteriol *179*, 6107-6111.

Minkovsky, N., Zarimani, A., Chary, V.K., Johnstone, B.H., Powell, B.S., Torrance, P.D., Court, D.L., Simons, R.W., and Piggot, P.J. (2002). Bex, the *Bacillus subtilis* homolog of the essential *Escherichia coli* GTPase Era, is required for normal cell division and spore formation. J Bacteriol *184*, 6389-6394.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol *60*, 174-182.

Monot, M., Eckert, C., Lemire, A., Hamiot, A., Dubois, T., Tessier, C., Dumoulard, B., Hamel, B., Petit, A., Lalande, V*., et al.* (2015). *Clostridium difficile*: New insights into the evolution of the pathogenicity locus. Sci Rep *5*, 15023.

Monteiro, M.A. (2016). The design of a *Clostridium difficile* carbohydrate-based vaccine. Methods Mol Biol *1403*, 397-408.

Muh, U., Pannullo, A.G., Weiss, D.S., and Ellermeier, C.D. (2019). A xylose-inducible expression system and a CRISPR Interference plasmid for targeted knockdown of gene expression in *Clostridioides difficile*. J Bacteriol *201*.

Mullane, K. (2014). Fidaxomicin in *Clostridium difficile* infection: latest evidence and clinical guidance. Ther Adv Chronic Dis *5*, 69-84.

Munoz-Lopez, M., and Garcia-Perez, J.L. (2010). DNA transposons: nature and applications in genomics. Curr Genomics *11*, 115-128.

Na, X., Kim, H., Moyer, M.P., Pothoulakis, C., and LaMont, J.T. (2008). gp96 is a human colonocyte plasma membrane binding protein for Clostridium difficile toxin A. Infect Immun *76*, 2862-2871.

Nagengast, F.M., Grubben, M.J., and van Munster, I.P. (1995). Role of bile acids in colorectal carcinogenesis. Eur J Cancer *31A*, 1067-1070.

Nerber, H.N., and Sorg, J.A. (2021). The small acid-soluble proteins of *Clostridioides difficile* are important for UV resistance and serve as a check point for sporulation. PLoS Pathog *17*, e1009516.

Neumeier, J., and Meister, G. (2020). siRNA specificity: RNAi mechanisms and strategies to reduce off-target effects. Front Plant Sci *11*, 526455.

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

Nidhi, S., Anand, U., Oleksak, P., Tripathi, P., Lal, J.A., Thomas, G., Kuca, K., and Tripathi, V. (2021). Novel CRISPR-Cas systems: An updated review of the current achievements, applications, and future research perspectives. Int J Mol Sci *22*.

O'Hara, E.B., Chekanova, J.A., Ingle, C.A., Kushner, Z.R., Peters, E., and Kushner, S.R. (1995). Polyadenylylation helps regulate mRNA decay in *Escherichia coli*. Proc Natl Acad Sci U S A *92*, 1807-1811.

O'Horo, J.C., Jindai, K., Kunzer, B., and Safdar, N. (2014). Treatment of recurrent *Clostridium difficile* infection: a systematic review. Infection *42*, 43-59.

Oatley, P., Kirk, J.A., Ma, S., Jones, S., and Fagan, R.P. (2020). Spatial organization of *Clostridium difficile* S-layer biogenesis. Sci Rep *10*, 14089.

Ooijevaar, R.E., van Beurden, Y.H., Terveer, E.M., Goorhuis, A., Bauer, M.P., Keller, J.J., Mulder, C.J.J., and Kuijper, E.J. (2018). Update of treatment algorithms for *Clostridium difficile* infection. Clin Microbiol Infect *24*, 452-462.

Papatheodorou, P., Zamboglou, C., Genisyuerek, S., Guttenberg, G., and Aktories, K. (2010). Clostridial glucosylating toxins enter cells via clathrin-mediated endocytosis. PLoS One *5*, e10673.

Paredes-Sabja, D., Shen, A., and Sorg, J.A. (2014). *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. Trends Microbiol *22*, 406-416.

Passmore, C.M., McElnay, J.C., Rainey, E.A., and D'Arcy, P.F. (1988). Metronidazole excretion in human milk and its effect on the suckling neonate. Br J Clin Pharmacol *26*, 45-51.

Peltier, J., Courtin, P., El Meouche, I., Lemee, L., Chapot-Chartier, M.P., and Pons, J.L. (2011). *Clostridium difficile* has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. J Biol Chem *286*, 29053-29062.

Peltier, J., Shaw, H.A., Couchman, E.C., Dawson, L.F., Yu, L., Choudhary, J.S., Kaever, V., Wren, B.W., and Fairweather, N.F. (2015). Cyclic diGMP regulates production of sortase substrates of *Clostridium difficile* and their surface exposure through ZmpI protease-mediated cleavage. J Biol Chem *290*, 24453-24469.

Percy, M.G., and Grundling, A. (2014). Lipoteichoic acid synthesis and function in Gram-positive bacteria. Annu Rev Microbiol *68*, 81-100.

Peters, J.M., Colavin, A., Shi, H., Czarny, T.L., Larson, M.H., Wong, S., Hawkins, J.S., Lu, C.H.S., Koo, B.M., Marta, E.*, et al.* (2016). A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell *165*, 1493-1506.

Peters, J.M., Silvis, M.R., Zhao, D., Hawkins, J.S., Gross, C.A., and Qi, L.S. (2015). Bacterial CRISPR: accomplishments and prospects. Curr Opin Microbiol *27*, 121-126.

Phong, W.Y., Lin, W., Rao, S.P., Dick, T., Alonso, S., and Pethe, K. (2013). Characterization of phosphofructokinase activity in *Mycobacterium tuberculosis* reveals that a functional glycolytic carbon flow is necessary to limit the accumulation of toxic metabolic intermediates under hypoxia. PLoS One *8*, e56037.

Planche, T., and Wilcox, M.H. (2015). Diagnostic pitfalls in *Clostridium difficile* infection. Infect Dis Clin North Am *29*, 63-82.

Popham, D.L. (2002). Specialized peptidoglycan of the bacterial endospore: the inner wall of the lockbox. Cell Mol Life Sci *59*, 426-433.

Popoff, M.R. (2018). *Clostridium difficile* and *Clostridium sordellii* toxins, proinflammatory versus anti-inflammatory response. Toxicon *149*, 54-64.

Poulsen, B.E., Yang, R., Clatworthy, A.E., White, T., Osmulski, S.J., Li, L., Penaranda, C., Lander, E.S., Shoresh, N., and Hung, D.T. (2019). Defining the core essential genome of *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A *116*, 10072-10080.

Poxton, I.R., and Cartmill, T.D. (1982). Immunochemistry of the cell-surface carbohydrate antigens of *Clostridium difficile*. J Gen Microbiol *128*, 1365-1370.
Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., and Wagner, R. (2010). Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol *75*, 1495-1512.

Purdy, D., O'Keeffe, T.A., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., and Minton, N.P. (2002). Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. Mol Microbiol *46*, 439-452.

Pyne, M.E., Bruder, M.R., Moo-Young, M., Chung, D.A., and Chou, C.P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. Sci Rep *6*, 25666.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Rao, K., and Safdar, N. (2016). Fecal microbiota transplantation for the treatment of *Clostridium difficile* infection. J Hosp Med *11*, 56-61.

Rauhut, R., and Klug, G. (1999). mRNA degradation in bacteria. FEMS Microbiol Rev *23*, 353-370.

Reid, C.W., Vinogradov, E., Li, J., Jarrell, H.C., Logan, S.M., and Brisson, J.R. (2012). Structural characterization of surface glycans from *Clostridium difficile*. Carbohydr Res *354*, 65-73.

Reynolds, C.B., Emerson, J.E., de la Riva, L., Fagan, R.P., and Fairweather, N.F. (2011). The *Clostridium difficile* cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. PLoS Pathog *7*, e1002024.

Ridlon, J.M., Harris, S.C., Bhowmik, S., Kang, D.J., and Hylemon, P.B. (2016). Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes *7*, 22-39.

Rineh, A., Kelso, M.J., Vatansever, F., Tegos, G.P., and Hamblin, M.R. (2014). *Clostridium difficile* infection: molecular pathogenesis and novel therapeutics. Expert Rev Anti Infect Ther *12*, 131-150.

Rosenberg, S.M., Shee, C., Frisch, R.L., and Hastings, P.J. (2012). Stress-induced mutation via DNA breaks in *Escherichia coli*: a molecular mechanism with implications for evolution and medicine. Bioessays *34*, 885-892.

Santa Maria, J.P., Jr., Sadaka, A., Moussa, S.H., Brown, S., Zhang, Y.J., Rubin, E.J., Gilmore, M.S., and Walker, S. (2014). Compound-gene interaction mapping reveals distinct roles for *Staphylococcus aureus* teichoic acids. Proc Natl Acad Sci U S A *111*, 12510-12515.

Sara, M., and Sleytr, U.B. (2000). S-Layer proteins. J Bacteriol *182*, 859-868.

Sarkar, N. (1997). Polyadenylation of mRNA in prokaryotes. Annu Rev Biochem *66*, 173-197.

Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeno-Tarraga, A.M., Wang, H.*, et al.* (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet *38*, 779-786.

Sekulovic, O., Ospina Bedoya, M., Fivian-Hughes, A.S., Fairweather, N.F., and Fortier, L.C. (2015). The *Clostridium difficile* cell wall protein CwpV confers phase-variable phage resistance. Mol Microbiol *98*, 329-342.

Setlow, P. (2007). I will survive: DNA protection in bacterial spores. Trends Microbiol *15*, 172-180.

Setlow, P., Wang, S., and Li, Y.Q. (2017). Germination of spores of the orders *Bacillales* and *Clostridiales*. Annu Rev Microbiol *71*, 459-477.

Shen, A. (2012). *Clostridium difficile* toxins: mediators of inflammation. J Innate Immun *4*, 149-158.

Shen, A. (2019). Expanding the *Clostridioides difficile* genetics toolbox. J Bacteriol *201*.

Shen, A., Edwards, A.N., Sarker, M.R., and Paredes-Sabja, D. (2019). Sporulation and germination in Clostridial pathogens. Microbiol Spectr *7*.

Shen, E.P., and Surawicz, C.M. (2008). Current treatment options for severe *Clostridium difficile*-associated disease. Gastroenterol Hepatol (N Y) *4*, 134-139.

Shim, J.K., Johnson, S., Samore, M.H., Bliss, D.Z., and Gerding, D.N. (1998). Primary symptomless colonisation by *Clostridium difficile* and decreased risk of subsequent diarrhoea. Lancet *351*, 633-636.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I.*, et al.* (2017). Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol *15*, 169-182.

Shmakov, S.A., Wolf, Y.I., Savitskaya, E., Severinov, K.V., and Koonin, E.V. (2020). Mapping CRISPR spaceromes reveals vast host-specific viromes of prokaryotes. Commun Biol *3*, 321.

Shrestha, R., Cochran, A.M., and Sorg, J.A. (2019). The requirement for co-germinants during *Clostridium difficile* spore germination is influenced by mutations in *yabG* and *cspA*. PLoS Pathog *15*, e1007681.

Shuman, S., and Glickman, M.S. (2007). Bacterial DNA repair by non-homologous end joining. Nat Rev Microbiol *5*, 852-861.

Silva, J., Jr., Batts, D.H., Fekety, R., Plouffe, J.F., Rifkin, G.D., and Baird, I. (1981). Treatment of *Clostridium difficile* colitis and diarrhea with vancomycin. Am J Med *71*, 815-822.

Sleytr, U.B., Schuster, B., Egelseer, E.M., and Pum, D. (2014). S-layers: principles and applications. FEMS Microbiol Rev *38*, 823-864.

Slimings, C., and Riley, T.V. (2014). Antibiotics and hospital-acquired *Clostridium difficile* infection: update of systematic review and meta-analysis. J Antimicrob Chemother *69*, 881-891.

Smits, W.K., Lyras, D., Lacy, D.B., Wilcox, M.H., and Kuijper, E.J. (2016). *Clostridium difficile* infection. Nat Rev Dis Primers *2*, 16020.

Soldo, B., Lazarevic, V., Pagni, M., and Karamata, D. (1999). Teichuronic acid operon of *Bacillus subtilis* 168. Mol Microbiol *31*, 795-805.

Sorg, J.A., and Sonenshein, A.L. (2008). Bile salts and glycine as cogerminants for *Clostridium difficile* spores. J Bacteriol *190*, 2505-2512.

Sorg, J.A., and Sonenshein, A.L. (2009). Chenodeoxycholate is an inhibitor of *Clostridium difficile* spore germination. J Bacteriol *191*, 1115-1117.

Stabler, R.A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T.D., Sebaihia, M., Quail, M.A., Rose, G.*, et al.* (2009). Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol *10*, R102.

Stefanovic, C., Hager, F.F., and Schaffer, C. (2021). LytR-CpsA-Psr glycopolymer transferases: Essential bricks in Gram-positive bacterial cell wall assembly. Int J Mol Sci *22*.

Sternberg, S.H., and Doudna, J.A. (2015). Expanding the biologist's toolkit with CRISPR-Cas9. Mol Cell *58*, 568-574.

Swick, M.C., Koehler, T.M., and Driks, A. (2016). Surviving between hosts: Sporulation and transmission. Microbiol Spectr *4*.

Tan, K.S., Wee, B.Y., and Song, K.P. (2001). Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*. J Med Microbiol *50*, 613-619.

Terveer, E.M., van Beurden, Y.H., Goorhuis, A., Seegers, J., Bauer, M.P., van Nood, E., Dijkgraaf, M.G.W., Mulder, C.J.J., Vandenbroucke-Grauls, C., Verspaget, H.W.*, et al.* (2017). How to: Establish and run a stool bank. Clin Microbiol Infect *23*, 924-930.

Thompson, S.A. (2002). *Campylobacter* surface-layers (S-layers) and immune evasion. Ann Periodontol *7*, 43-53.

Tieu, J.D., Williams, R.J., 2nd, Skrepnek, G.H., and Gentry, C.A. (2019). Clinical outcomes of fidaxomicin vs oral vancomycin in recurrent *Clostridium difficile* infection. J Clin Pharm Ther *44*, 220-228.

Todor, H., Silvis, M.R., Osadnik, H., and Gross, C.A. (2021). Bacterial CRISPR screens for gene function. Curr Opin Microbiol *59*, 102-109.

Underwood, S., Guan, S., Vijayasubhash, V., Baines, S.D., Graham, L., Lewis, R.J., Wilcox, M.H., and Stephenson, K. (2009). Characterization of the sporulation initiation pathway of *Clostridium difficile* and its role in toxin production. J Bacteriol *191*, 7296-7305.

Urdaneta, V., and Casadesus, J. (2017). Interactions between bacteria and bile salts in the gastrointestinal and hepatobiliary tracts. Front Med (Lausanne) *4*, 163.

van Opijnen, T., Bodi, K.L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods *6*, 767-772.

van Opijnen, T., and Camilli, A. (2010). Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. Curr Protoc Microbiol *Chapter 1*, Unit1E 3.

van Opijnen, T., and Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol *11*, 435-442.

Vincent, C., and Manges, A.R. (2015). Antimicrobial use, human gut microbiota and *Clostridium difficile* colonization and infection. Antibiotics (Basel) *4*, 230-253.

Voth, D.E., and Ballard, J.D. (2005). *Clostridium difficile* toxins: mechanism of action and role in disease. Clin Microbiol Rev *18*, 247-263.

Waligora, A.J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A., and Karjalainen, T. (2001). Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. Infect Immun *69*, 2144-2153.

Wang, N., Ozer, E.A., Mandel, M.J., and Hauser, A.R. (2014). Genome-wide identification of *Acinetobacter baumannii* genes necessary for persistence in the lung. mBio *5*, e01163-01114.

Weidenmaier, C., and Peschel, A. (2008). Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. Nat Rev Microbiol *6*, 276-287.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature *477*, 486-489.

Willcocks, S.J., Stabler, R.A., Atkins, H.S., Oyston, P.F., and Wren, B.W. (2018). High-throughput analysis of *Yersinia pseudotuberculosis* gene essentiality in optimised in vitro conditions, and implications for the speciation of *Yersinia pestis*. BMC Microbiol *18*, 46.

Willing, S.E., Candela, T., Shaw, H.A., Seager, Z., Mesnage, S., Fagan, R.P., and Fairweather, N.F. (2015). *Clostridium difficile* surface proteins are anchored to the cell wall using CWB2 motifs that recognise the anionic polymer PSII. Mol Microbiol *96*, 596-608.

Wong, Y.C., Abd El Ghany, M., Naeem, R., Lee, K.W., Tan, Y.C., Pain, A., and Nathan, S. (2016). Candidate essential genes in *Burkholderia cenocepacia* J2315 identified by genome-wide TraDIS. Front Microbiol *7*, 1288.

Wood, A., Irving, S.E., Bennison, D.J., and Corrigan, R.M. (2019). The (p)ppGpp-binding GTPase Era promotes rRNA processing and cold adaptation in *Staphylococcus aureus*. PLoS Genet *15*, e1008346.

Wydau-Dematteis, S., El Meouche, I., Courtin, P., Hamiot, A., Lai-Kuen, R., Saubamea, B., Fenaille, F., Butel, M.J., Pons, J.L., Dupuy, B*., et al.* (2018). Cwp19 is a novel lytic transglycosylase involved in stationary phase autolysis resulting in toxin release in *Clostridium difficile*. mBio *9*.

Xu, Z., Li, Y., and Yan, A. (2020). Repurposing the native type I-F CRISPR-Cas system in *Pseudomonas aeruginosa* for genome editing. STAR Protoc *1*, 100039.

Yan, F., Wang, W., and Zhang, J. (2019). CRISPR-Cas12 and Cas13: the lesser known siblings of CRISPR-Cas9. Cell Biol Toxicol *35*, 489-492.

Yuan, P., Zhang, H., Cai, C., Zhu, S., Zhou, Y., Yang, X., He, R., Li, C., Guo, S., Li, S*., et al.* (2015). Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for *Clostridium difficile* toxin B. Cell Res *25*, 157-168.

Zar, F.A., Bakkanagari, S.R., Moorthi, K.M., and Davis, M.B. (2007). A comparison of vancomycin and metronidazole for the treatment of *Clostridium difficile*-associated diarrhea, stratified by disease severity. Clin Infect Dis *45*, 302-307.

Zhanel, G.G., Walkty, A.J., and Karlowsky, J.A. (2015). Fidaxomicin: A novel agent for the treatment of *Clostridium difficile* infection. Can J Infect Dis Med Microbiol *26*, 305-312.

Zhu, D., Bullock, J., He, Y., and Sun, X. (2019). Cwp22, a novel peptidoglycan cross-linking enzyme, plays pleiotropic roles in *Clostridioides difficile*. Environ Microbiol *21*, 3076-3090.

# Appendices

## Appendix I

**Table A1. *Clostridium difficile* strains used in this study.**

| Strain | Description | Source |
|---|---|---|
| **R20291** | *C. difficile* U.K. epidemic strain ribotype 027 responsible for the 2004 Stoke-Mandeville outbreak. | Lab Stocks (Stabler *et al*., 2009) |
| **FM2.5** | *C. difficile* diffocin-resistant mutant containing a point mutation in *slpA* early in the gene, resulting in a frame shift and loss of function. | Avidbiotics |

## Appendix II

**Table A2. *Escherichia coli* strains used in this study.**

| Strain | Description | Source |
|---|---|---|
| **Neb5α** | Chemically competent *E. coli* K-12 strain for routine cloning and plasmid propagation. | New England Biolabs |
| **CA434** | Strain HB101 harbouring the IncPβ conjugative plasmid R702. | Lab Stocks (Purdy *et al*., 2002) |

## Appendix III

**Table A3. Plasmids used in this study.**

| Plasmid Name | Description | Source |
|---|---|---|
| **pRPF185** | Shuttle vector containing a tetracycline-inducible *gusA.* | Lab Stocks (Fagan and Fairweather, 2011) |
| **pRPF215** | Shuttle vector containing a tetracycline-inducible mariner transposase and an *ermB* transposon. | Lab Stocks (Dembek *et al*., 2015) |
| **pSOB008** | pRPF185 in which a guide RNA targeting *pyrE* has been added. | This Study |
| **pSOB009** | pSOB008 in which a codon-optimised d*cas9* has been added. | This Study |
| **pSOB010** | pSOB009 in which the *pyrE* gRNA has been replaced with a gRNA targeting *secA2.* | This Study |

| | | |
|---|---|---|
| **pSOB011** | pSOB009 in which the *pyrE* gRNA has been replaced by a gRNA lacking the 20 nucleotide base-pairing region. | This Study |
| **pSOB012** | pSOB009 in which the *pyrE* gRNA has been replaced with a gRNA targeting *cdR20291_2329.* | This Study |
| **pSOB013** | pSOB009 in which the *pyrE* gRNA has been replaced with a gRNA targeting *cdR20291_2359.* | This Study |
| **pSOB014** | pSOB009 in which the *pyrE* gRNA has been replaced with a gRNA targeting *cdR20291_2659.* | This Study |
| **pSOB015** | pSOB009 in which the *pyrE* gRNA has been replaced with a gRNA targeting *cdR20291_2660.* | This Study |
| **pSOB016** | pSOB009 in which the *pyrE* gRNA has been replaced with a gRNA targeting *cdR20291_2661.* | This Study |

**Appendix IV**

**Table A4. Primers used in this study.**

| Name | Sequence 5'-3' | Restriction Site |
|---|---|---|
| **RF1501** | GCCATTTCTATAACTATATTTTCTGG | |
| **RF1504** | GTGATATACTTAGAGTAAATACAG | |
| **RF1517** | GATCCCAGTGTGCTGGGTTTCAAATGTTGGT AACTTTGAC | BstXI |
| **RF1519** | GATCCCTAAGGATAAAACGAAAGGCCCAG TCTTTCGACTGAGCCTTTCG | Bsu361 |
| **RF1520** | GAAAGTTACACGTTACTAAAGGCA TAAAAATAAGAAGCCTGCAAATGC | |
| **RF1521** | AATGATACGGCGACCACCGAGATCTACACTC TTTCCCTACACGACGCTCTTCCGATCTCGTAC GGGCTTCTTATTTTTATGTGTTAGACCGGGGA CTTATCAGC | |
| **RF1522** | GACTGGAGTTCAGACGTGTGCTCTTCCGATC | |
| **RF1523** | AATGATACGGCGACCACCGAGATCTACACTC TTTCCCTACACGACGCTCTTCCGATCTGCTAG CTGGCTTCTTATTTTTATGTGTTAGACCGGGG ACTTATCAGC | |
| **RF1629** | AATGATACGGCGACCACCGAGATCTACACTCT TTCCCTACACGACGCTCTTCCGATCTTACGTAG GCTTCTTATTTTTATGTGTTAGACCG GGGACTTATCAGC | |
| **RF1630** | AATGATACGGCGACCACCGAGATCTACACTCTT TCCCTACACGACGCTCTTCCGATCTTAGCTAGG GCTTCTTATTTTTATGTGTTAGACCGGGGACTTA TCAGC | |
| **RF1640** | ACACTCTTTCCCTACACGACGCTCTTCCGATCTG GCTTCTTATTTTTATGTGTTAGACCGGGGACTTA TCAGC | |
| **RF1643** | TTCTATTTAAAGTTTTATTAAAACTTATAGGATC CGCTTTTAATCACC | SacI |
| **RF1644** | GCTTGATCGTAGCGTTAACAGATCTGAGCTAGC TCAATAATACTAGGAGG | BamHI |
| **RF2027** | GATCAGGTCCCTAGCGCAGATAAAACG | EcoO109I |
| **RF2028** | GATCAGGTCCCTAGCGCAGCCTAAGGATAAAAC | EcoO109I |
| **NebNEXT Index 11 Primer for Illumina** | CAAGCAGAAGACGGCATACGAGATTACAAGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATC | |

**Appendix V**

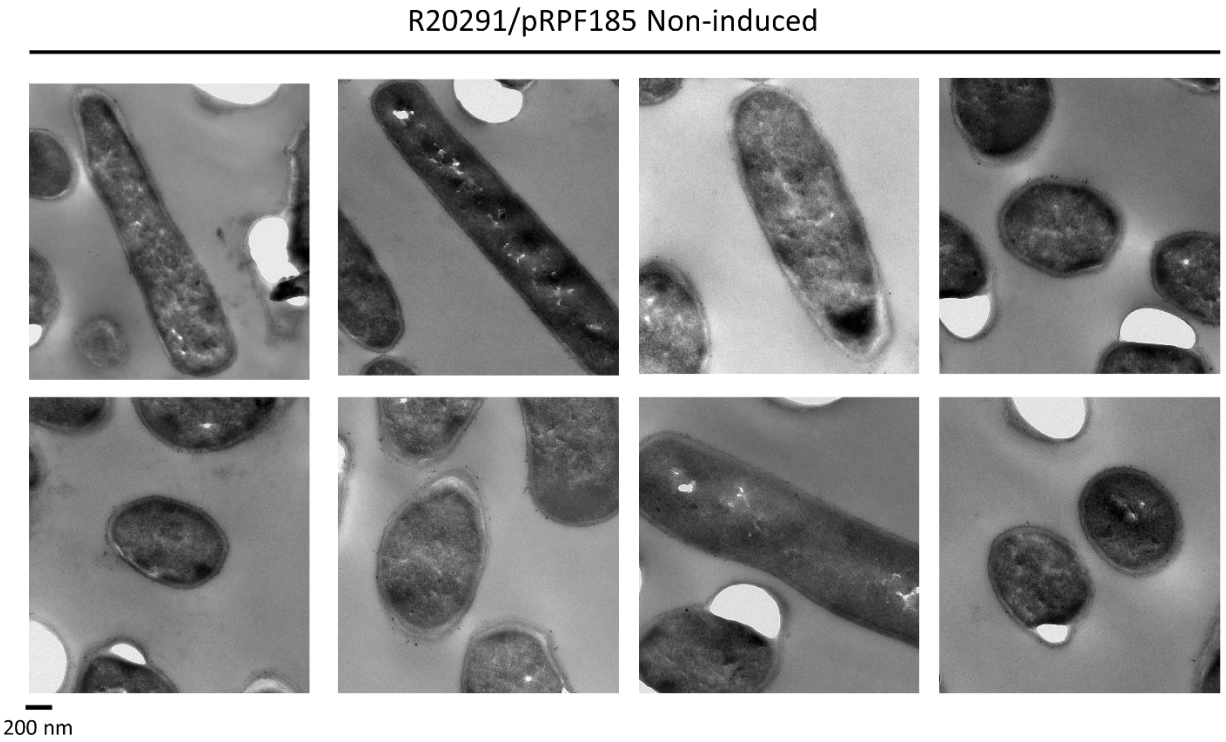
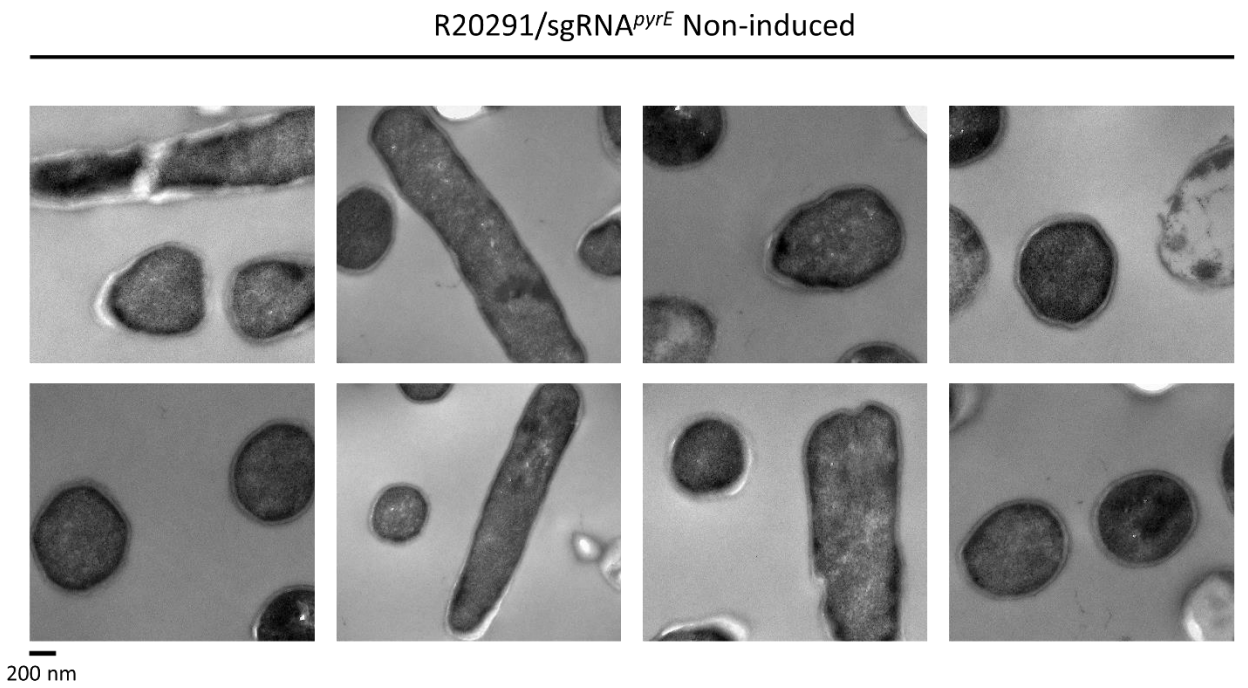



**Fig. 6S1. All images obtained from TEM analysis of non-induced and induced stationary phase cultures of R20291 harbouring pRPF185.**

**Appendix VI**



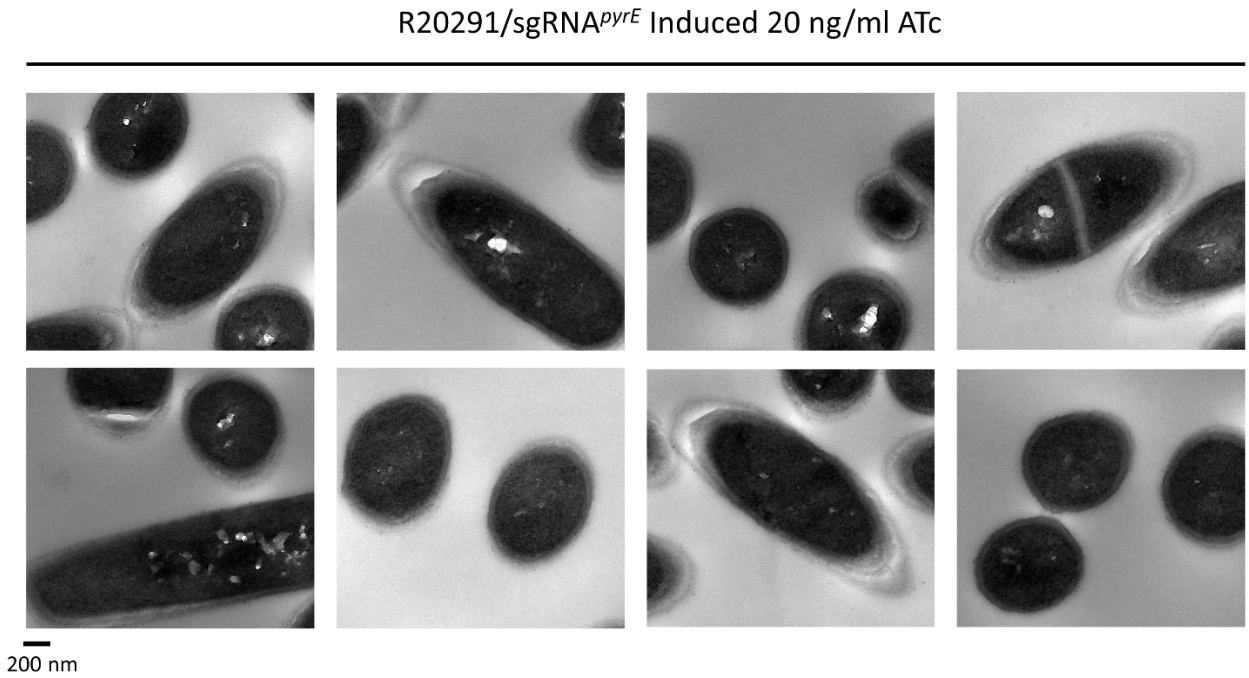




**Fig. 6S2. All images obtained from TEM analysis of non-induced and induced stationary phase cultures of R20291 harbouring sgRNA$^{pyrE}$.**

**Appendix VII**

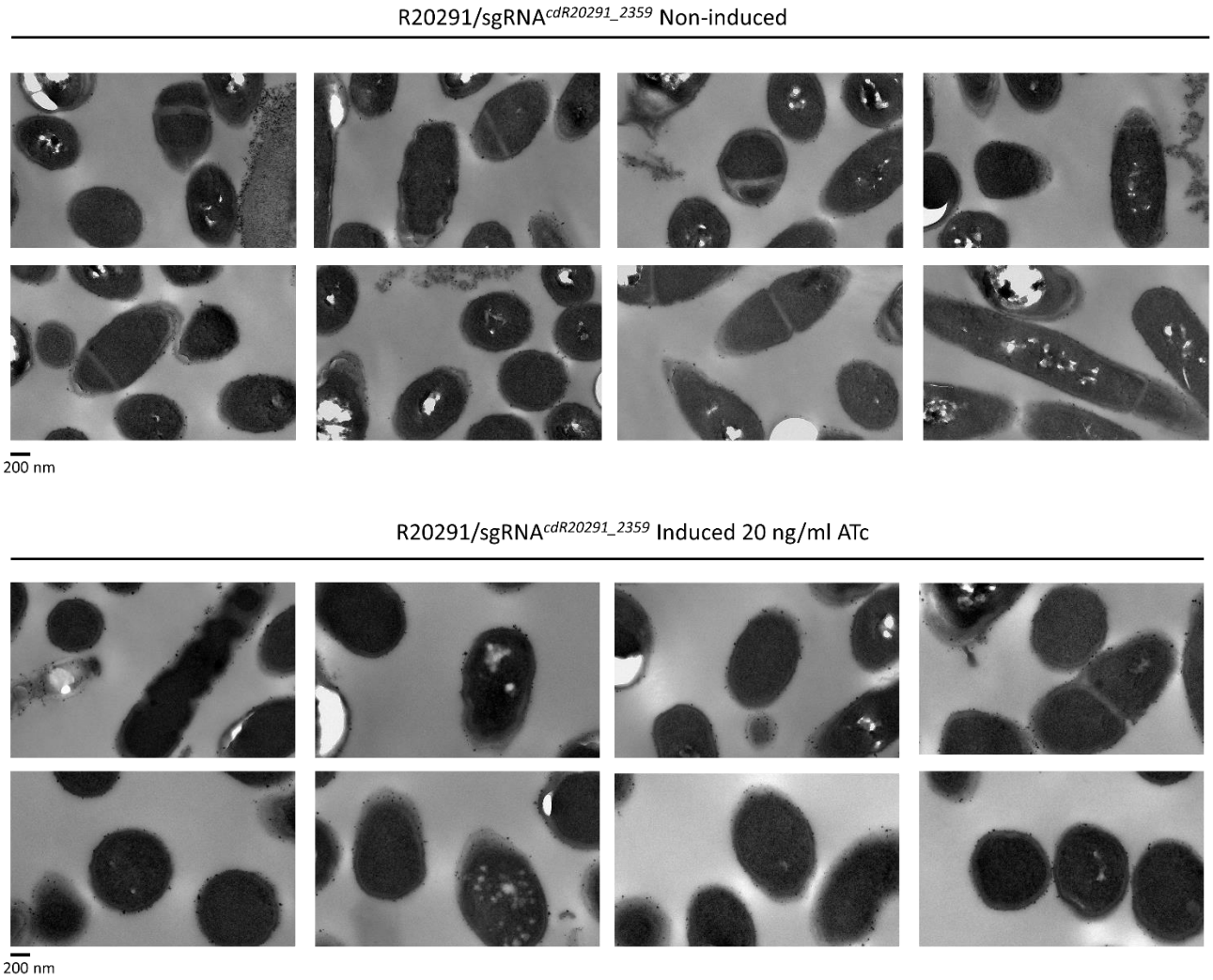



**Fig. 6S3. CDR20291_2359-depleted cells in R20291.** All images obtained from TEM analysis of non-induced and induced stationary phase cultures of R20291 harbouring sgRNA$^{cdR20291_2359}$.

**Appendix VIII**

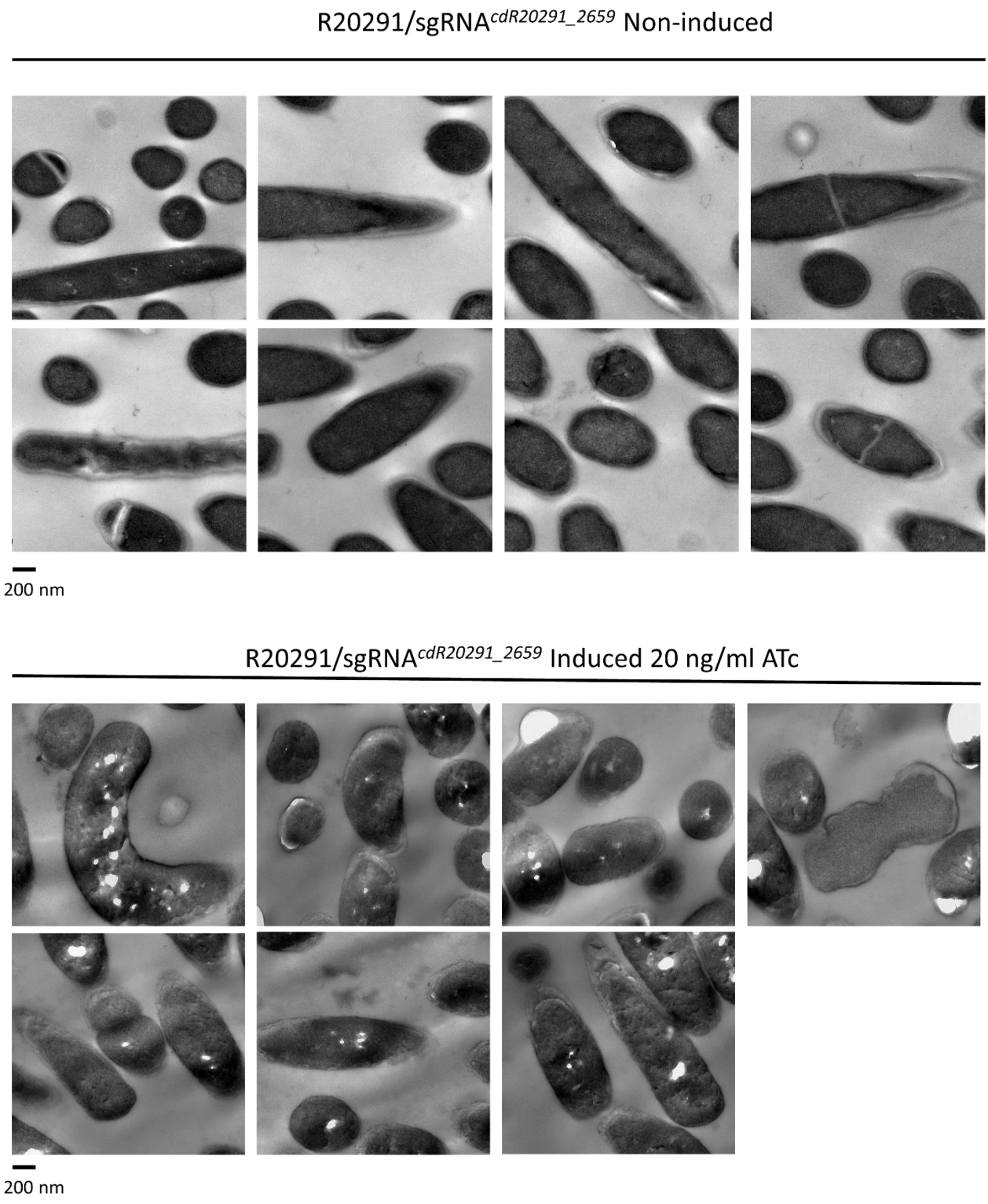



**Fig. 6S4. RkpK-depleted cells in R20291.** All images obtained from TEM analysis of non-induced and induced stationary phase cultures of R20291 harbouring sgRNA$^{cdR20291_2659}$.

**Appendix IX**

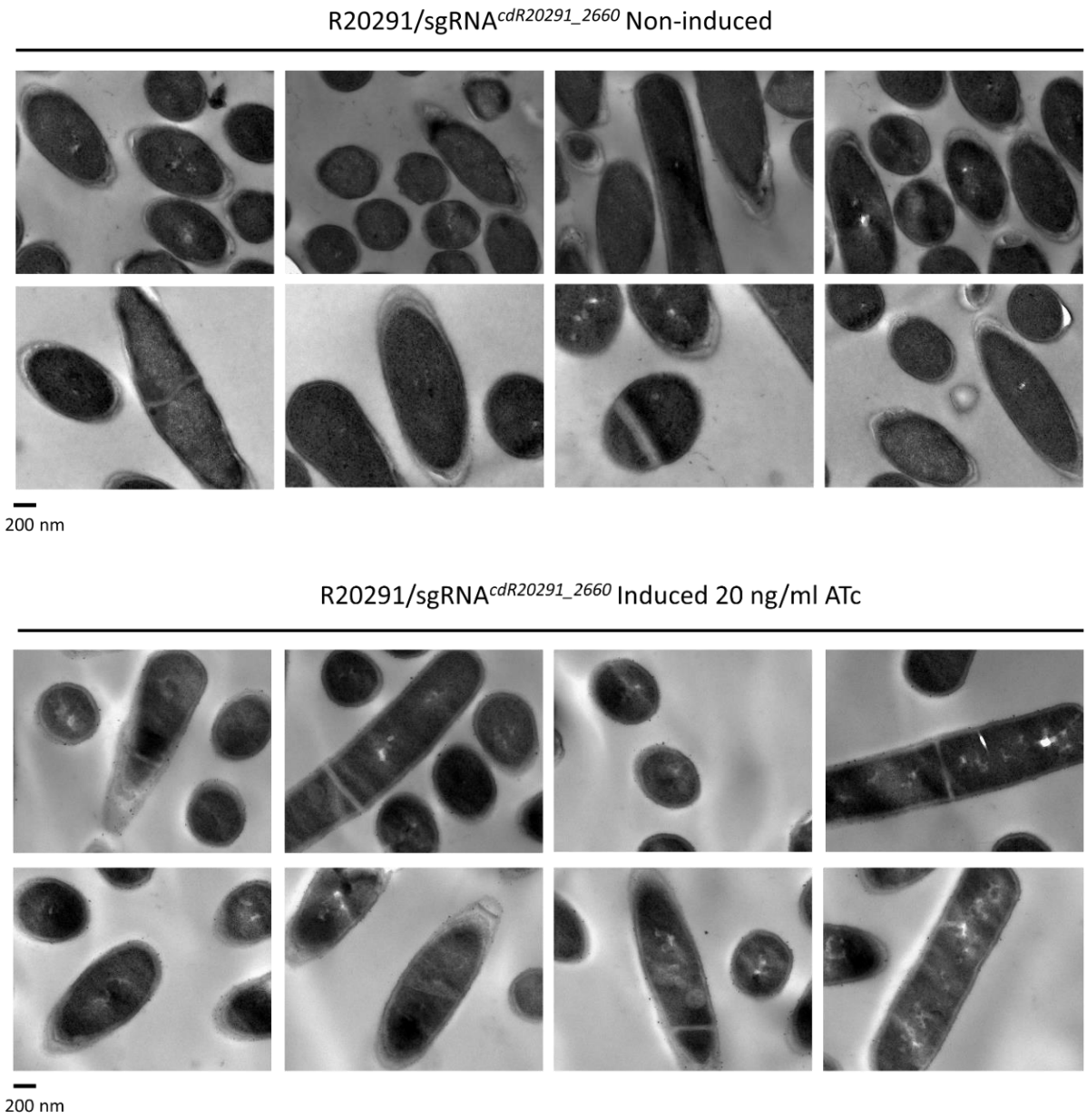
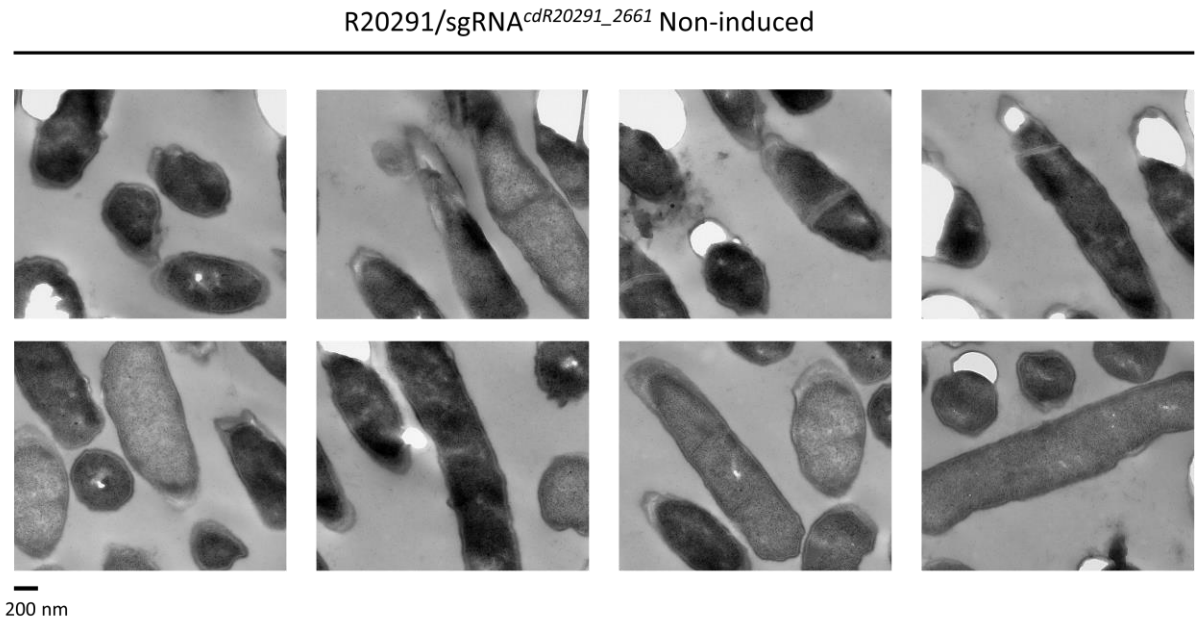



**Fig. 6S5. TuaG-depleted cells in R20291.** All images obtained from TEM analysis of non-induced and induced stationary phase cultures of R20291 harbouring sgRNA$^{cdR20291_2660}$.

**Appendix X**



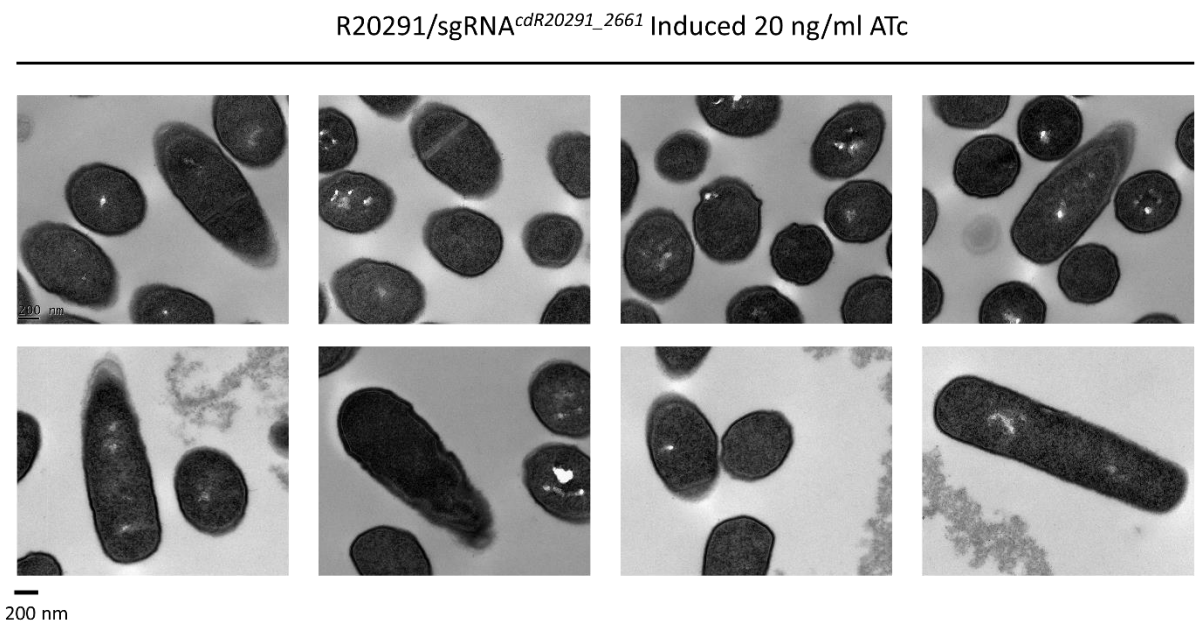




**Fig. 6S6. CDR20291_2661-depleted cells in R20291.** All images obtained from TEM analysis of non-induced and induced stationary phase cultures of R20291 harbouring sgRNA$^{cdR20291_2661}$.

## Appendix XI

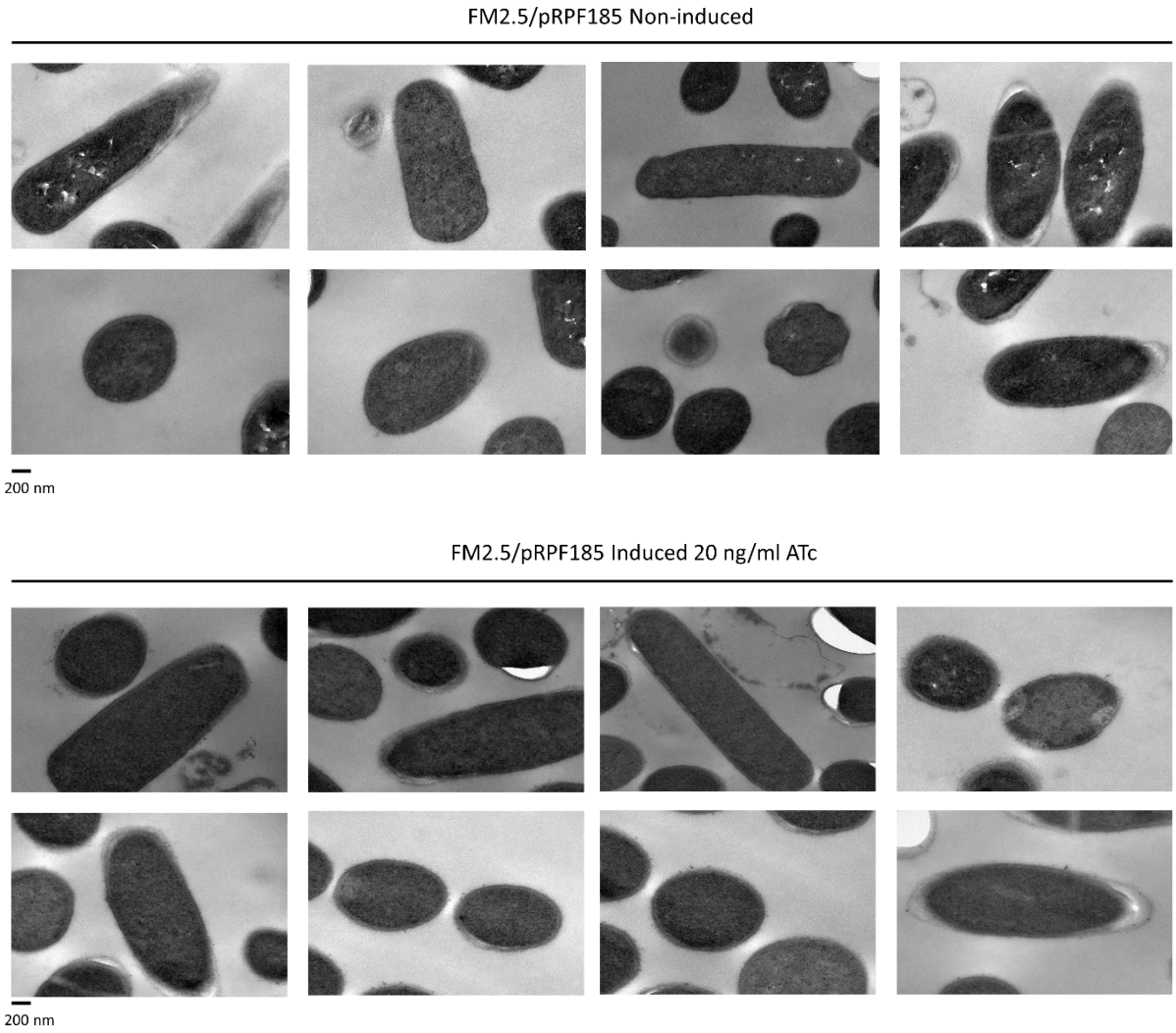



**Fig. 6S7. All images obtained from TEM analysis of non-induced and induced stationary phase cultures of FM2.5 harbouring pRPF185.**

**Appendix XII**

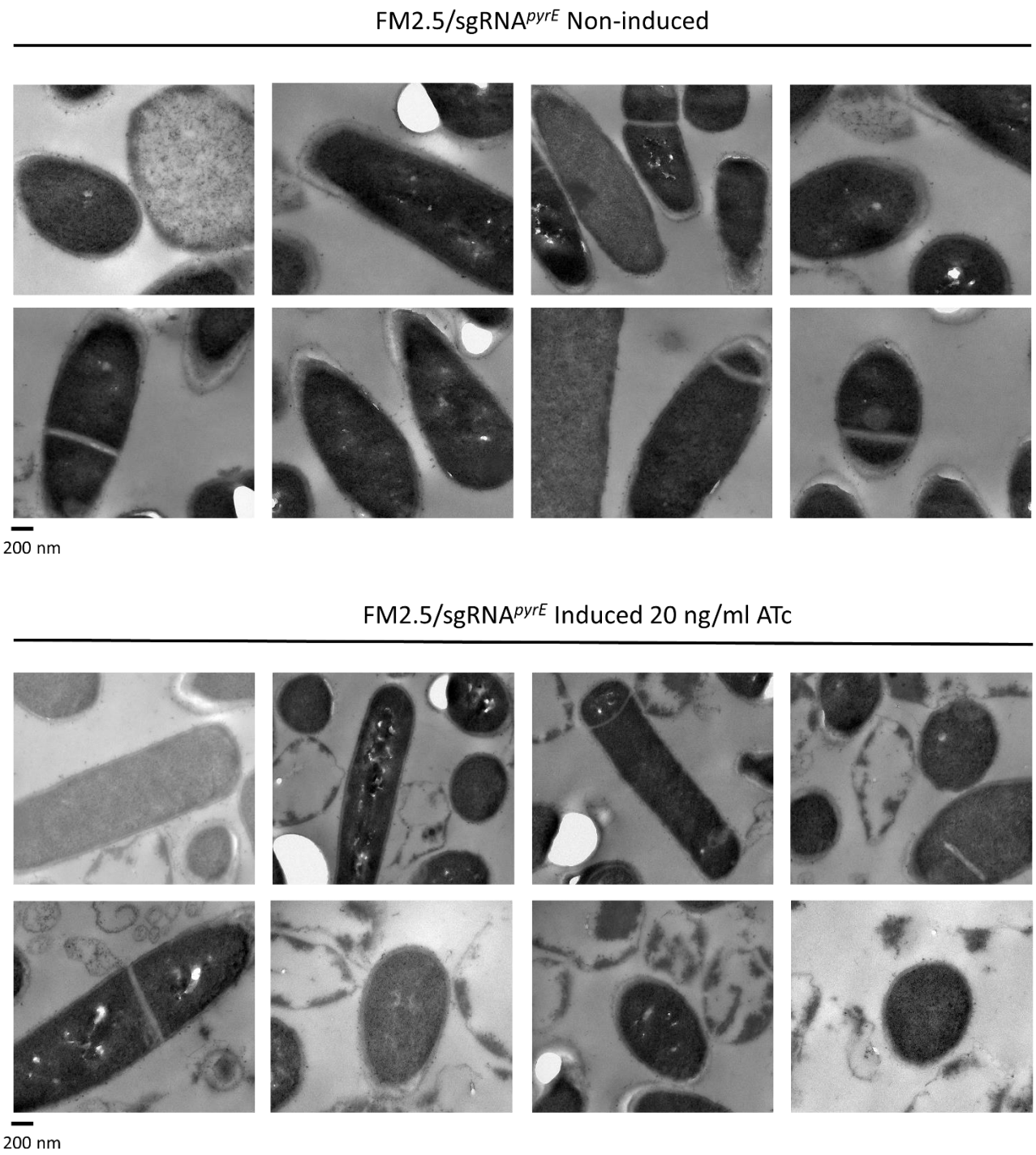



**Fig. 6S8. All images obtained from TEM analysis of non-induced and induced stationary phase cultures of FM2.5 harbouring sgRNA$^{pyrE}$.**

**Appendix XIII**

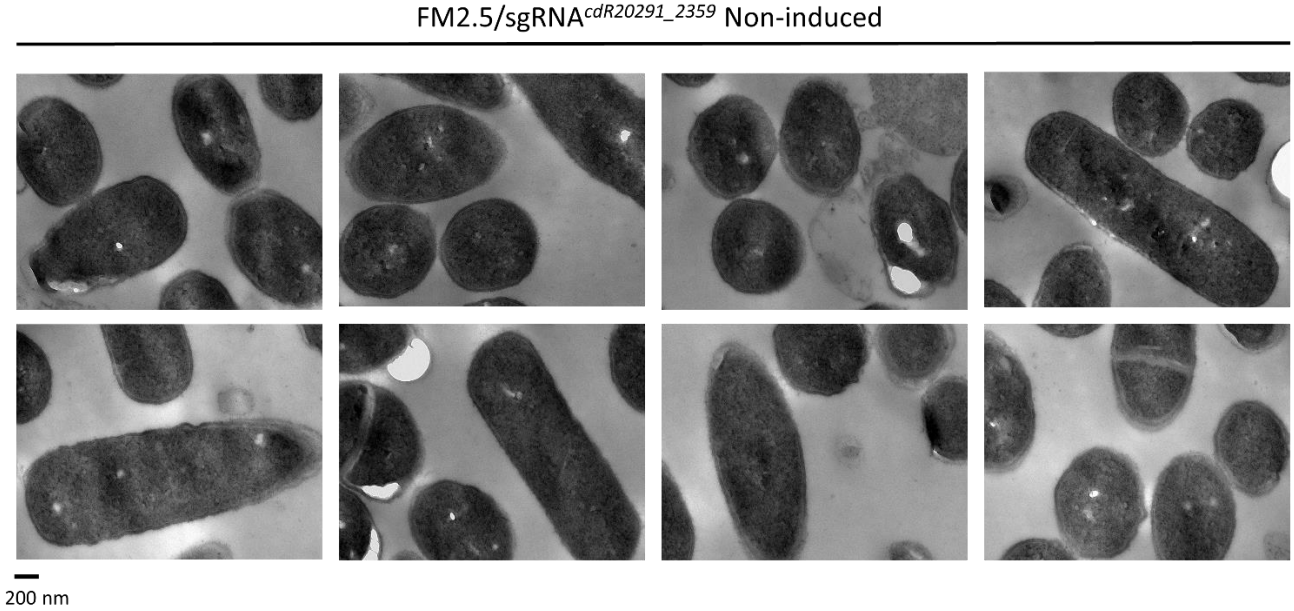



**Fig. 6S9. CDR20291_2359-depleted cells in FM2.5.** All images obtained from TEM analysis of non-induced and induced stationary phase cultures of FM2.5 harbouring sgRNA$^{cdR20291_2359}$.

**Appendix XIV**

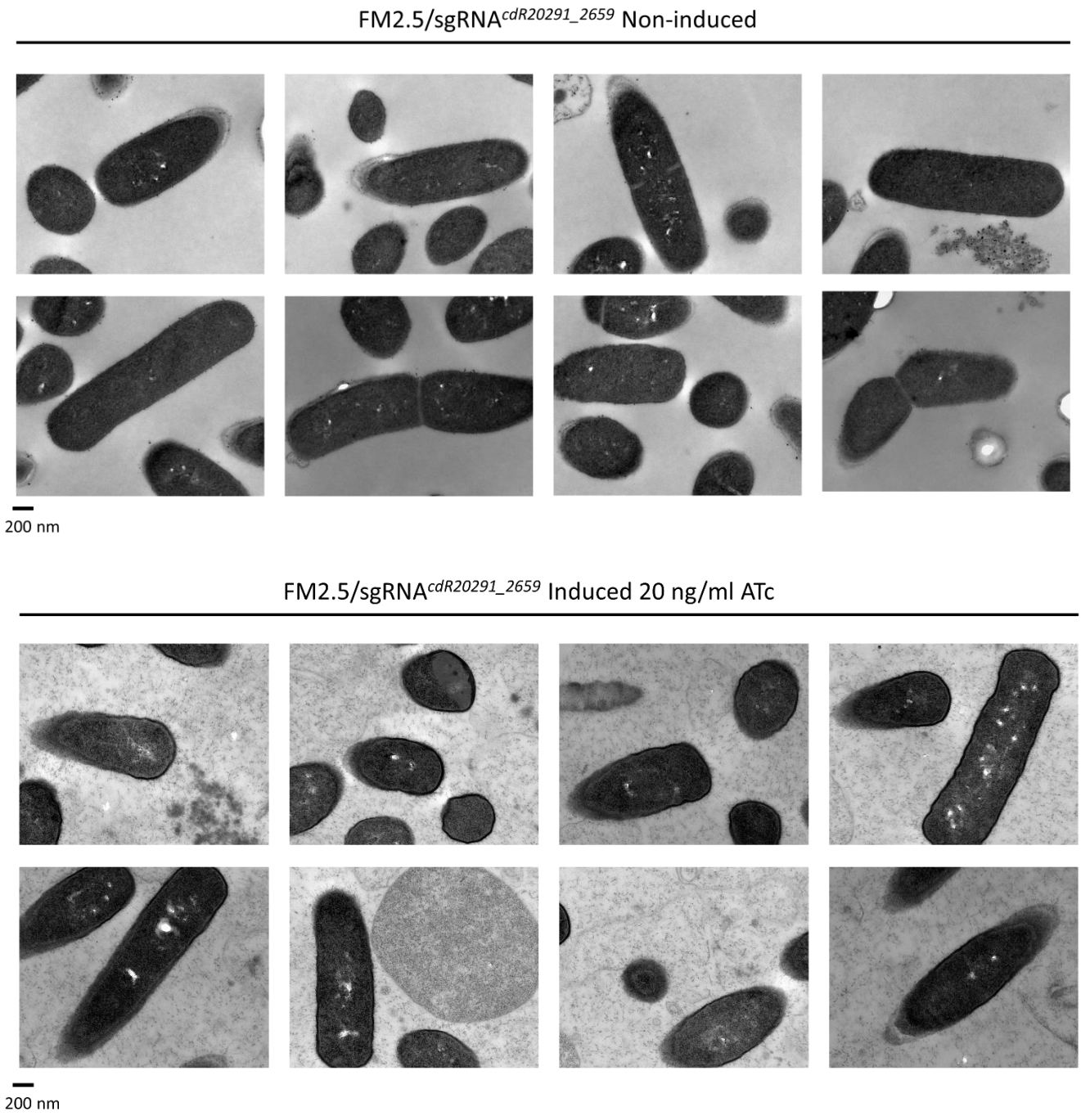



**Fig. 6S10. RkpK-depleted cells in FM2.5.** All images obtained from TEM analysis of non-induced and induced stationary phase cultures of FM2.5 harbouring sgRNA$^{cdR20291_2659}$.

**Appendix XV**

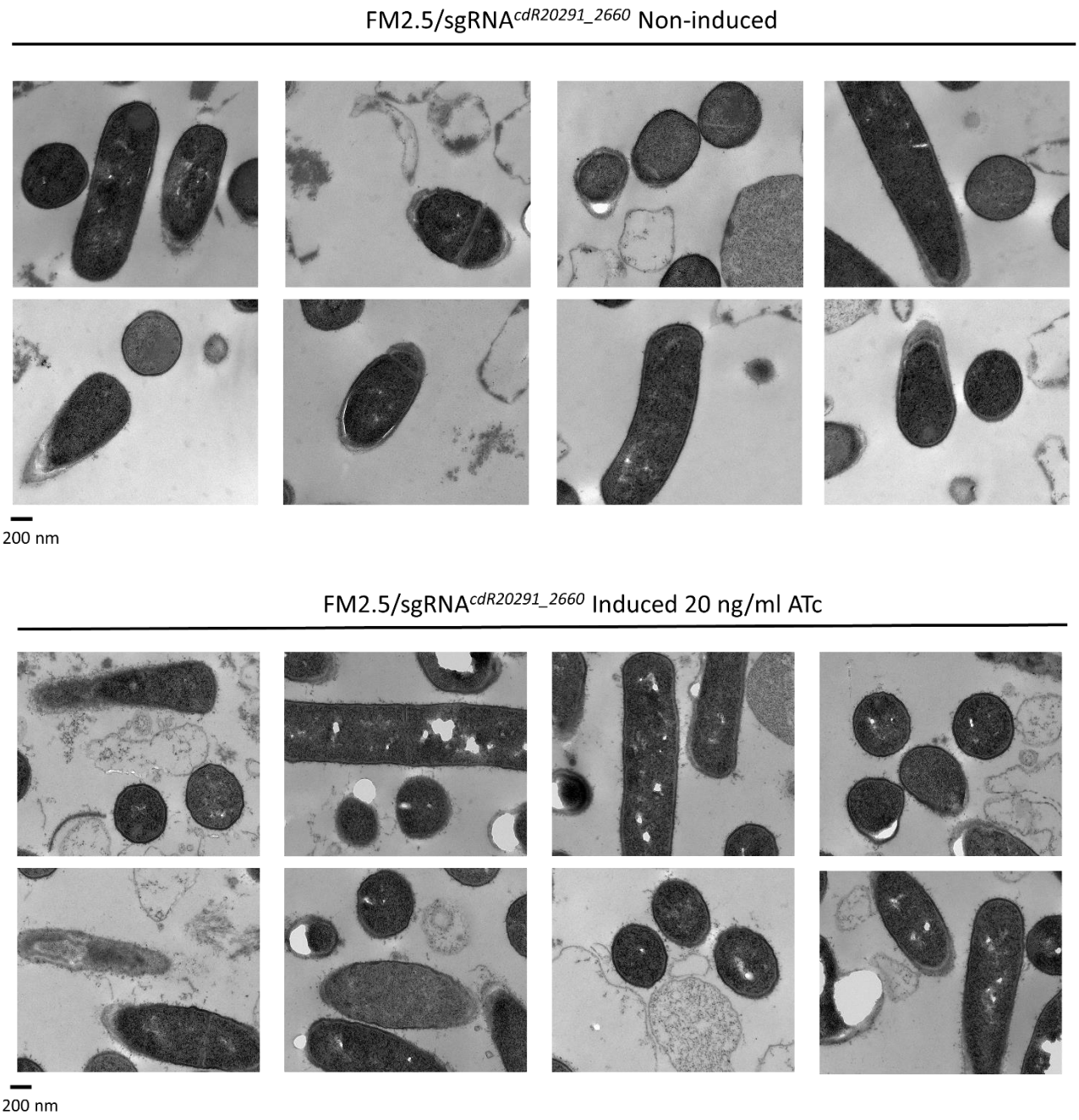
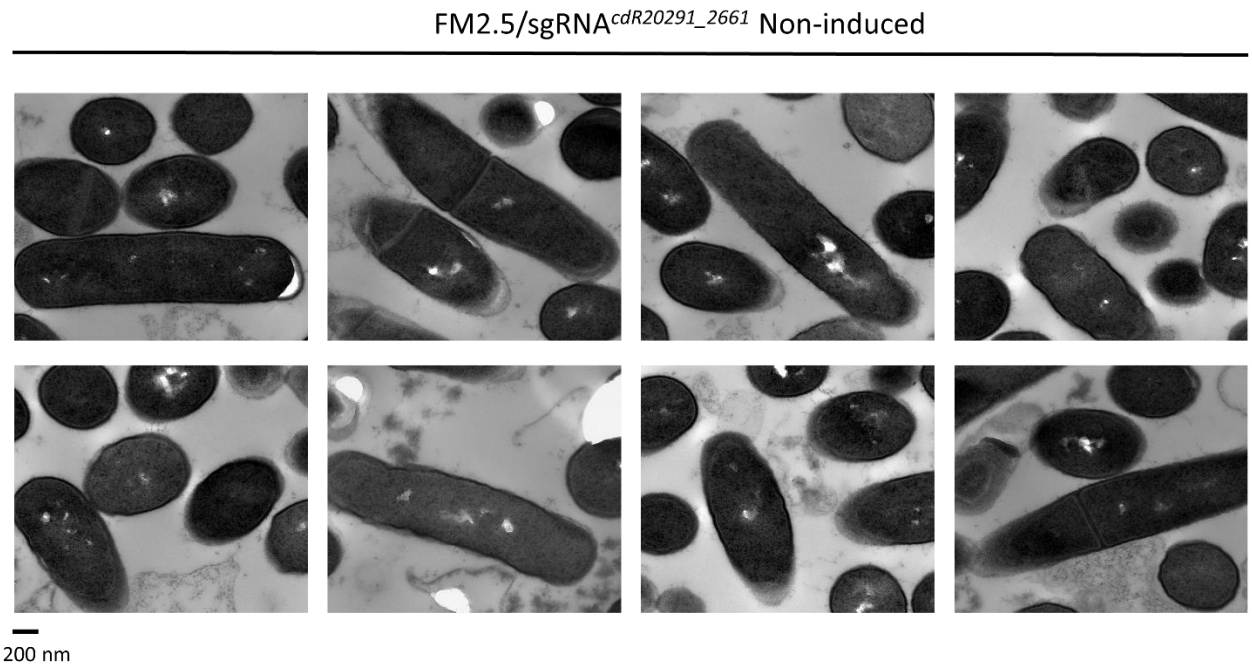



**Fig. 6S11. TuaG-depleted cells in FM2.5**. All images obtained from TEM analysis of non-induced and induced stationary phase cultures of FM2.5 harbouring sgRNA$^{cdR20291_2660}$.

**Appendix XVI**



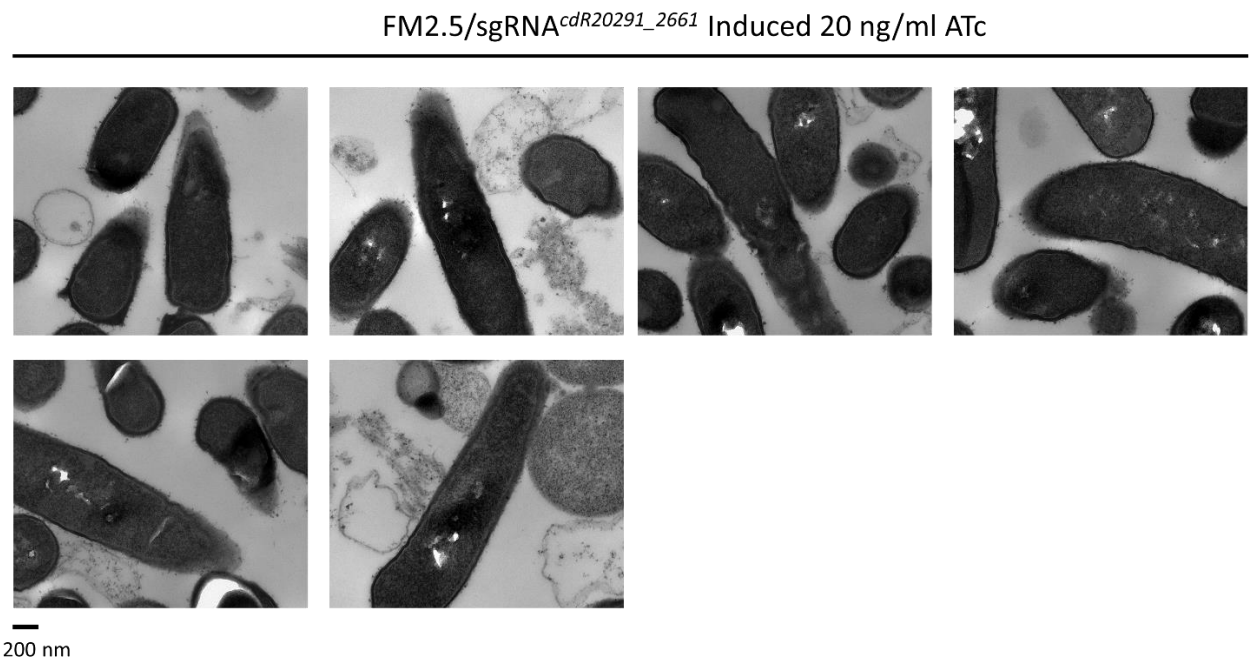




**Fig. 6S12. CDR20291_2661-depleted cells in FM2.5.** All images obtained from TEM analysis of non-induced and induced stationary phase cultures of FM2.5 harbouring sgRNA$^{cdR20291_2661}$.